Measure and Integral

An Introduction to Real Analysis

Second Edition

Richard L. Wheeden
Antoni Zygmund

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Measure and Integral

An Introduction to Real Analysis

Second Edition

PURE AND APPLIED MATHEMATICS

A Program of Monographs, Textbooks, and Lecture Notes

EXECUTIVE EDITORS

Earl J. Taft
Rutgers University
Piscataway, New Jersey

Zuhair Nashed
University of Central Florida
Orlando, Florida

EDITORIAL BOARD

Jane Cronin
Rutgers University

S. Kobayashi
University of California, Berkeley

Marvin Marcus
University of California, Santa Barbara

W. S. Massey
Yale University

Anil Nerode
Cornell University

Freddy van Oystaeyen
University of Antwerp, Belgium

Donald Passman
University of Wisconsin, Madison

Fred S. Roberts
Rutgers University

David L. Russell
Virginia Polytechnic Institute and State University

Walter Schempp
Universität Siegen

MONOGRAPHS AND TEXTBOOKS IN PURE AND APPLIED MATHEMATICS

Recent Titles

John Dauns and Yiqiang Zhou, Classes of Modules (2006)

N. K. Govil, H. N. Mhaskar, Ram N. Mohapatra, Zuhair Nashed, and J. Szabados, Frontiers in Interpolation and Approximation (2006)

Luca Lorenzi and Marcello Bertoldi, Analytical Methods for Markov Semigroups (2006)

M. A. Al-Gwaiz and S. A. Elsanousi, Elements of Real Analysis (2006)

Theodore G. Faticoni, Direct Sum Decompositions of Torsion-Free Finite Rank Groups (2007)

R. Sivaramakrishnan, Certain Number-Theoretic Episodes in Algebra (2006)

Aderemi Kuku, Representation Theory and Higher Algebraic K-Theory (2006)

Robert Piziak and P. L. Odell, Matrix Theory: From Generalized Inverses to Jordan Form (2007)

Norman L. Johnson, Vikram Jha, and Mauro Biliotti, Handbook of Finite Translation Planes (2007)

Lieven Le Bruyn, Noncommutative Geometry and Cayley-smooth Orders (2008)

Fritz Schwarz, Algorithmic Lie Theory for Solving Ordinary Differential Equations (2008)

Jane Cronin, Ordinary Differential Equations: Introduction and Qualitative Theory, Third Edition (2008)

Su Gao, Invariant Descriptive Set Theory (2009)

Christopher Apelian and Steve Surace, Real and Complex Analysis (2010)

Norman L. Johnson, Combinatorics of Spreads and Parallelisms (2010)

Lawrence Narici and Edward Beckenstein, Topological Vector Spaces, Second Edition (2010)

Moshe Sniedovich, Dynamic Programming: Foundations and Principles, Second Edition (2010)

Drumi D. Bainov and Snezhana G. Hristova, Differential Equations with Maxima (2011)

Willi Freeden, Metaharmonic Lattice Point Theory (2011)

Murray R. Bremner, Lattice Basis Reduction: An Introduction to the LLL Algorithm and Its Applications (2011)

Clifford Bergman, Universal Algebra: Fundamentals and Selected Topics (2011)

A. A. Martynyuk and Yu. A. Martynyuk-Chernienko, Uncertain Dynamical Systems: Stability and Motion Control (2012)

Washek F. Pfeffer, The Divergence Theorem and Sets of Finite Perimeter (2012)

Willi Freeden and Christian Gerhards, Geomathematically Oriented Potential Theory (2013)

Anatoly Martynyuk, Larisa Chernetskaya, and Vladislav Martynyuk, Weakly Connected Nonlinear Systems: Boundedness and Stability of Motion (2013)

Franz Halter-Koch, Quadratic Irrationals: An Introduction to Classical Number Theory (2013)

Łukasz Piasecki, Classification of Lipschitz Mappings (2014)

Andy R. Magid, The Separable Galois Theory of Commutative Rings, Second Edition (2015)

Measure and Integral

An Introduction to Real Analysis

Second Edition

Richard L. Wheeden

Rutgers University

Antoni Zygmund

Late of the University of Chicago

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20150318

International Standard Book Number-13: 978-1-4987-0290-4 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To our families

Richard L. Wheeden and Antoni Zygmund

To my wife, Sharon, whose love and support have been vital to me during the revision process and throughout our marriage.

Richard L. Wheeden

Contents

Preface to the Second Edition xiii
Preface to the First Edition xv
Authors xvii

1. Preliminaries 1
1.1 Points and Sets in $\mathbf{R}^n$ 1
1.2 $\mathbf{R}^n$ as a Metric Space 2
1.3 Open and Closed Sets in $\mathbf{R}^n$, and Special Sets 5
1.4 Compact Sets and the Heine–Borel Theorem 9
1.5 Functions 10
1.6 Continuous Functions and Transformations 11
1.7 The Riemann Integral 12
Exercises 13

2. Functions of Bounded Variation and the Riemann–Stieltjes Integral 17
2.1 Functions of Bounded Variation 17
2.2 Rectifiable Curves 25
2.3 The Riemann–Stieltjes Integral 26
2.4 Further Results about Riemann–Stieltjes Integrals 34
Exercises 37

3. Lebesgue Measure and Outer Measure 41
3.1 Lebesgue Outer Measure and the Cantor Set 41
3.2 Lebesgue Measurable Sets 45
3.3 Two Properties of Lebesgue Measure 50
3.4 Characterizations of Measurability 52
3.5 Lipschitz Transformations of $\mathbf{R}^n$ 54
3.6 A Nonmeasurable Set 57
Exercises 58

4. Lebesgue Measurable Functions 63
4.1 Elementary Properties of Measurable Functions 64
4.2 Semicontinuous Functions 69
4.3 Properties of Measurable Functions and Theorems of Egorov and Lusin 71
4.4 Convergence in Measure 73
Exercises 76

5. The Lebesgue Integral 81
5.1 Definition of the Integral of a Nonnegative Function 81
5.2 Properties of the Integral 84
5.3 The Integral of an Arbitrary Measurable f 90
5.4 Relation between Riemann–Stieltjes and Lebesgue Integrals, and the L^p Spaces, $0 < p < \infty$ 97
5.5 Riemann and Lebesgue Integrals 106
Exercises 108

6. Repeated Integration 113
6.1 Fubini's Theorem 113
6.2 Tonelli's Theorem 117
6.3 Applications of Fubini's Theorem 120
Exercises 125

7. Differentiation 129
7.1 The Indefinite Integral 129
7.2 Lebesgue's Differentiation Theorem 131
7.3 Vitali Covering Lemma 142
7.4 Differentiation of Monotone Functions 144
7.5 Absolutely Continuous and Singular Functions 149
7.6 Convex Functions 154
7.7 The Differential in $\mathbf{R^n}$ 160
Exercises 176

8. L^p Classes 183
8.1 Definition of L^p 183
8.2 Hölder's Inequality and Minkowski's Inequality 185
8.3 Classes l^p 189
8.4 Banach and Metric Space Properties 191
8.5 The Space L^2 and Orthogonality 195
8.6 Fourier Series and Parseval's Formula 199
8.7 Hilbert Spaces 203
Exercises 205

9. Approximations of the Identity and Maximal Functions 213
9.1 Convolutions 213
9.2 Approximations of the Identity 217
9.3 The Hardy–Littlewood Maximal Function 226
9.4 The Marcinkiewicz Integral 230
Exercises 233

10. Abstract Integration 237
10.1 Additive Set Functions and Measures 237
10.2 Measurable Functions and Integration 244
10.3 Absolutely Continuous and Singular Set Functions and Measures 253
10.4 The Dual Space of L^p 263
10.5 Relative Differentiation of Measures 266
Exercises 273

11. Outer Measure and Measure 279
11.1 Constructing Measures from Outer Measures 279
11.2 Metric Outer Measures 282
11.3 Lebesgue–Stieltjes Measure 285
11.4 Hausdorff Measure 289
11.5 The Carathéodory–Hahn Extension Theorem 293
Exercises 298

12. A Few Facts from Harmonic Analysis 301
12.1 Trigonometric Fourier Series 301
12.2 Theorems about Fourier Coefficients 308
12.3 Convergence of $S[f]$ and $\widetilde{S}[f]$ 314
12.4 Divergence of Fourier Series 321
12.5 Summability of Sequences and Series 324
12.6 Summability of $S[f]$ and $\widetilde{S}[f]$ by the Method of the Arithmetic Mean 332
12.7 Summability of $S[f]$ by Abel Means 346
12.8 Existence of $\widetilde{f}$ 351
12.9 Properties of $\widetilde{f}$ for $f \in L^p, 1 < p < \infty$ 359
12.10 Application of Conjugate Functions to Partial Sums of $S[f]$ 363
Exercises 364

13. The Fourier Transform 371
13.1 The Fourier Transform on L^1 372
13.2 The Fourier Transform on L^2 392
13.3 The Hilbert Transform on L^2 399
13.4 The Fourier Transform on L^p, $1 < p < 2$ 403
Exercises 406

14. Fractional Integration 415
14.1 Subrepresentation Formulas and Fractional Integrals 415
14.2 L^1, L^1 Poincaré Estimates, the Subrepresentation Formula, and Hölder Classes 421
14.3 Norm Estimates for I_α 432

14.4 Exponential Integrability of $I_\alpha f$ 436
14.5 Bounded Mean Oscillation 444
Exercises 452

15. Weak Derivatives and Poincaré–Sobolev Estimates 461
15.1 Weak Derivatives 461
15.2 Smooth Approximation and Sobolev Spaces 471
15.3 Poincaré–Sobolev Estimates 478
Exercises 496

Notations 501

Index 505

Preface to the Second Edition

The first edition of this textbook was written more than 35 years ago. In the interim, applications of the theory of Lebesgue measure and integration and the rudiments of harmonic analysis have soared.

This second edition does not attempt to summarize the enormous recent expansion of the field. Instead, it develops some important topics not treated in the first edition. These include the Rademacher–Stepanov theorem, added in Chapter 7, and Chapters 13 through 15, which treat the Fourier transform, fractional integrals (or Riesz potentials), and first-order Poincaré–Sobolev estimates, respectively. The setting in the new chapters is the classical Euclidean one. However, many of the methods in Chapters 14 and 15 can be extended to more general geometric situations as well as to measures that are more general than Lebesgue measure.

Chapter 13 studies the Fourier transform of functions in the spaces L^1, L^2, and L^p, $1 < p < 2$. As an application of the L^2 theory in the one-dimensional case, the Hilbert transform is shown to be a bounded operator on L^2.

Chapter 14 studies fractional integration and some topics related to mean oscillation properties of functions, including the classes of Hölder continuous functions and the space of functions of bounded mean oscillation. Motivation for studying fractional integration is provided by a subrepresentation formula, which in higher dimensions plays a role roughly similar to the one played by the fundamental theorem of calculus in one dimension.

In Chapter 15, the norm estimates derived in Chapter 14 for fractional integral operators are applied to obtain local and global first-order Poincaré–Sobolev inequalities, including endpoint cases. The notion of weak (distributional) partial derivatives is considered in advance. The subrepresentation formula derived in Chapter 14 for smooth functions is extended to functions with a weak gradient, and the formula plays a critical role.

In this second edition, some changes have been made in the 12 chapters that were in the first edition, such as the addition of many new exercises. The biggest change in the text itself is the addition of Section 7.7, proving the existence of a tangent plane to the graph of a Lipschitz function of several variables. However, the order of the presentation in Chapters 1 through 12 has been purposely retained. For example, new exercises are generally placed after those that were already present in the first edition. The numbering of equations, theorems, etc., has been retained from the first edition in the printed version; however, please note that these items have been renumbered according to the publisher's current style in all electronic versions of this edition.

In the years since the book was originally published, many readers have sent suggestions to me for ways to improve the presentation of the material.

I am very grateful for many helpful comments and have incorporated a large number of them. I thank the following people in this regard: James Bennett, Earl Berkson, Bernard Bialecki, Sagun Chanillo, Richard Gundy, Max Jodeit, Russell John, Edward Lotkowski, Umberto Neri, Roger Nussbaum, Eugene Speer, and Jason Tedor. Thanks also go to Luc Nguyen for collecting and typing course notes related to Chapters 14 and 15.

I especially thank my friend and student Edward Lotkowski. At his own suggestion, he proofread most of the second edition (as well as the first one). His comments have been thoughtful and informed, and they have improved the clarity and content of the book in many places.

I thank Senior Editor Robert Ross for helping in many ways, especially for his continued support in making the printed version of the book simpler to read by allowing flexibility in formatting styles. I also acknowledge the help and energy of my project editor, Todd Perry.

My coauthor, teacher, and friend Antoni Zygmund passed away in 1992. The first edition of this book evolved from notes that I took as a graduate student in his real variables course at the University of Chicago. His influence is profound both in the book and in the development of harmonic analysis. I take this opportunity to express my thanks for his nuturing support. In doing so, I also speak for many others whom he helped.

Richard L. Wheeden

Preface to the First Edition

The modern theory of measure and integration was created, primarily through the work of Lebesgue, at the turn of twentieth century. Although the basic ideas are by now well established, there are ever-widening applications that have made the theory one of the central parts of mathematical analysis. However, different applications require different emphasis on various aspects of the theory. For example, certain facts are of primary interest for real and complex analysis, others for functional analysis, and still others for probability and statistics. This text is written from the point of view of real variables and treats the theory primarily as modern calculus.

The book presupposes that the reader has a feeling for rigor and some knowledge of elementary facts from calculus. Some material that is no doubt familiar to many readers has been included; its inclusion seemed desirable in order to make the presentation clear and self-contained.

The approach of the book is to develop the theory of measure and integration first in the simple setting of Euclidean space. In this case, there is a rich theory having a close relation to familiar facts from calculus and generalizing those facts. Later on, we introduce a more general treatment based on abstract notions characterized by axioms and with less geometric content. We have chosen this approach purposely, even though it leads to some repetition, since considering a special case first usually helps in developing a better understanding of the general situation. Anyway, we all "learn by repetition."

The outline of the book is as follows: Chapter 1 is primarily a collection of various background information, including elementary definitions and results that will be taken for granted later in the book; the reader should already be familiar with most of this material. Very few proofs are given in Chapter 1. Actual presentation of the theory begins in Chapter 2, which treats notions associated with functions of bounded variation, such as the Riemann–Stieltjes integral. Strictly speaking, a reading of Chapter 2 could be postponed until Chapter 5, where we use the Riemann–Stieltjes integral as a way of representing the Lebesgue integral.

Chapter 3 deals with Lebesgue measure in Euclidean space, via the notion of outer measure. Chapter 4 gives the theory of measurable functions, and Chapter 5 considers the Lebesgue integral, again in Euclidean space. In Chapter 6, we study repeated integration, the central result being Fubini's theorem. Chapter 7 treats the process that is the inverse of integration, namely, differentiation. Here, we consider the differentiation of integrals treated as set functions, as well as the differentiation of real-valued functions of a single variable, such as the differentiability of monotone functions. Chapters 3 through 7 complete the treatment of the general theory of integration in Euclidean spaces.

In Chapters 8 and 9, we consider special classes of functions, like L^2 and L^p, and special results for these classes, such as the behavior of convolution operators, the Hardy–Littlewood maximal function, and the integral of Marcinkiewicz.

In Chapters 10 and 11, we give an abstract treatment of Lebesgue measure and integration. Here, there are several possible approaches. We have chosen to start with an abstract definition of measure and develop the theory of integration following the pattern of earlier chapters. This is done in Chapter 10. It is natural to ask how such abstract measures actually arise. This question is answered to some extent in Chapter 11, where we use the notion of abstract outer measure to construct some specific examples of measures.

Chapter 12 plays a special role and can be read immediately after Chapter 9. It deals with an application of the Lebesgue integral to a specific branch of analysis—harmonic analysis. This is a very broad field, and we consider only a few problems indicative of the role that Lebesgue integration plays in applications. Harmonic analysis also happens to be a field whose development had a great impact on the theory of integration.

At the end of each chapter, we list a number of problems as exercises, sometimes with parenthetical hints at solutions. Some relatively important results are given in the exercises, but as a rule, the text does not require facts that have appeared earlier only as exercises.

We would like to express our thanks to the Departments of Mathematics of Rutgers University and the University of Chicago, and in particular to Professor William H. Meyer for the friendly help he offered us during the preparation of the manuscript. Special thanks also go to Joanne Darken and Dr. Edward Lotkowski, both of whom proofread almost the entire manuscript and offered many helpful comments, and to Michele Ginouves for her help with the cover design. Finally, thanks to Annette Roselli, our typist, for an excellent job.

Richard L. Wheeden
Antoni Zygmund

Authors

Richard L. Wheeden is Distinguished Professor of Mathematics at Rutgers University, New Brunswick, New Jersey. His primary research interests lie in the fields of classical harmonic analysis and partial differential equations, and he is the author or coauthor of more than a hundred research articles. After earning his PhD from the University of Chicago (1965), he held an instructorship there (1965–1966) and a National Science Foundation (NSF) postdoctoral fellowship at the Institute for Advanced Study (1966–1967).

Antoni Zygmund was Professor of Mathematics at the University of Chicago, Illinois. He was earlier a professor at Mt. Holyoke College and the University of Pennsylvania. His years at the University of Chicago began in 1947, and in 1964, he was appointed Gustavus F. and Ann M. Swift Distinguished Service Professor there. He published extensively in many branches of analysis, including Fourier series, singular integrals, and differential equations. He is the author of the classical treatise *Trigonometric Series* and a coauthor (with S. Saks) of *Analytic Functions*. He was elected to the National Academy of Sciences in Washington, DC (1961) as well as to a number of foreign academies.

1

Preliminaries

This book is devoted to Lebesgue integration and related topics, a basic part of modern analysis. There are classical and abstract approaches to the integral, and we have chosen the classical one, postponing a more abstract treatment until later in the book. The classical approach is based on the theory of measure (while in some modern treatments, the integral is introduced as a linear functional). Measure can be defined and studied in various spaces, but we will primarily consider n-dimensional Euclidean space, $\mathbf{R^n}$. A prerequisite, undertaken in this chapter, is a review of elementary notions about $\mathbf{R^n}$. We have not attempted to present these in a thorough manner, but only to list some of the definitions and notation that will be used throughout the book and state some background facts that a reader should know. We assume a knowledge of various properties of the real line $\mathbf{R^1}$ and of functions defined on $\mathbf{R^1}$ and leave as exercises the proofs of many facts that are either similar to or derivable from their one-dimensional analogues.

1.1 Points and Sets in $\mathbf{R^n}$

Let n be a positive integer. By *n-dimensional Euclidean space* $\mathbf{R^n}$, we mean the collection of all n-tuples $\mathbf{x} = (x_1, \ldots, x_n)$ of real numbers $x_k, -\infty < x_k < +\infty, k = 1, \ldots, n$. If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ are points of $\mathbf{R^n}$, we say that $\mathbf{x} = \mathbf{y}$ if $x_k = y_k$ for $1 \leq k \leq n$. $\mathbf{R^n}$ is a vector space over the reals if for $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, and $\alpha \in \mathbf{R^1}$ we define $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n)$ and $\alpha\mathbf{x} = (\alpha x_1, \ldots, \alpha x_n)$. The point each of whose coordinates is zero is called the *origin* and denoted $0 = (0, \ldots, 0)$ or $\mathbf{0} = (0, \ldots, 0)$. By the *vector* emanating from $\mathbf{x}$ and terminating at $\mathbf{y}$, we mean the line segment connecting $\mathbf{x}$ and $\mathbf{y}$, directed from $\mathbf{x}$ to $\mathbf{y}$. The points of this segment are of the form $(1 - t)\,\mathbf{x} + t\mathbf{y}, 0 \leq t \leq 1$. We will identify vectors that have equal length and direction. We will also identify $\mathbf{x}$ with the vector emanating from 0 and terminating at $\mathbf{x}$.

If E is a set of points of $\mathbf{R^n}$, we use the notation $CE = \mathbf{R^n} - E$ for the *complement* of E. The complement of $\mathbf{R^n}$ is the *empty set* $\emptyset$. If $\mathscr{F} = \{E\}$ is a family of subsets of $\mathbf{R^n}$, the *union* and *intersection* of the sets E in $\mathscr{F}$ are defined, respectively, by

$$\bigcup_{E\in\mathscr{F}} E = \left\{\mathbf{x} : \mathbf{x} \in E \text{ for some } E \in \mathscr{F}\right\},$$
$$\bigcap_{E\in\mathscr{F}} E = \left\{\mathbf{x} : \mathbf{x} \in E \text{ for all } E \in \mathscr{F}\right\}.$$

(Here, and systematically in the following, we use the notation $\{\mathbf{x} : \ldots\}$ to denote the set of points $\mathbf{x}$ that satisfy....)

If $\mathscr{F}$ is countable (i.e., finite or countably infinite), it will be called a *sequence of sets* and denoted $\mathscr{F} = \{E_k : k = 1, 2, \ldots\}$. The corresponding union and intersection will be written $\bigcup_k E_k$ and $\bigcap_k E_k$. A sequence $\{E_k\}$ of sets is said to *increase* to $\bigcup_k E_k$ if $E_k \subset E_{k+1}$ for all k and to *decrease* to $\bigcap_k E_k$ if $E_k \supset E_{k+1}$ for all k; we use the notations $E_k \nearrow \bigcup_k E_k$ and $E_k \searrow \bigcap_k E_k$ to denote these two possibilities. If $\{E_k\}_{k=1}^{\infty}$ is a sequence of sets, we define

$$\limsup E_k = \bigcap_{j=1}^{\infty}\left(\bigcup_{k=j}^{\infty} E_k\right), \quad \liminf E_k = \bigcup_{j=1}^{\infty}\left(\bigcap_{k=j}^{\infty} E_k\right), \tag{1.1}$$

noting that the sets $U_j = \bigcup_{k=j}^{\infty} E_k$ and $V_j = \bigcap_{k=j}^{\infty} E_k$ satisfy $U_j \searrow \limsup E_k$ and $V_j \nearrow \liminf E_k$. We leave it as a simple exercise to verify that $\limsup E_k$ consists of those points of $\mathbf{R^n}$ that belong to infinitely many E_k and $\liminf E_k$ of those that belong to all E_k for $k \geq k_0$ (where k_0 may vary from point to point). Thus, $\liminf E_k \subset \limsup E_k$.

If E_1 and E_2 are two sets, we define $E_1 - E_2$ by $E_1 - E_2 = E_1 \cap CE_2$ and call it the *difference* of E_1 and E_2 or the *relative complement* of E_2 in E_1. We will often have occasion to use the *De Morgan laws*, which govern relations between complements, unions, and intersections; these state that

$$C\left(\bigcup_{E\in\mathscr{F}} E\right) = \bigcap_{E\in\mathscr{F}} CE, \qquad C\left(\bigcap_{E\in\mathscr{F}} E\right) = \bigcup_{E\in\mathscr{F}} CE,$$

and are easily verified. The set-theoretic notions discussed earlier are not confined to $\mathbf{R^n}$ and hold for subsets of an arbitrary set S.

1.2 $\mathbf{R^n}$ as a Metric Space

$\mathbf{R^n}$ also has, of course, a metric space structure. If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, we define their *inner* (*dot*) *product* by

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k y_k.$$

We have $\mathbf{x}\cdot\mathbf{y}=\mathbf{y}\cdot\mathbf{x}$, $\alpha\mathbf{x}\cdot\mathbf{y}=\alpha(\mathbf{x}\cdot\mathbf{y})$ for real α and $\mathbf{x}\cdot(\mathbf{y}+\mathbf{z})=\mathbf{x}\cdot\mathbf{y}+\mathbf{x}\cdot\mathbf{z}$. Noting that $\mathbf{x}\cdot\mathbf{x}\geq 0$, we define the *absolute value* of $\mathbf{x}$, or the *length* of $\mathbf{x}$, by

$$|\mathbf{x}| = (\mathbf{x}\cdot\mathbf{x})^{1/2} = \left(\sum_{k=1}^{n} x_k^2\right)^{1/2}.$$

We will use this notation regardless of the dimension n. Thus, if $x\in\mathbf{R^1}$, $|x|$ means the usual one-dimensional absolute value of x. Then in any dimension, $|\mathbf{x}|$ has the following properties:

(i) $|\mathbf{x}|\geq 0$ and $|\mathbf{x}|=0$ if and only if $\mathbf{x}=0$

(ii) $|\alpha\mathbf{x}|=|\alpha||\mathbf{x}|$ for $\alpha\in\mathbf{R^1}$

(iii) $|\mathbf{x}+\mathbf{y}|\leq|\mathbf{x}|+|\mathbf{y}|$ (the *triangle inequality*)

To verify (iii), observe that if we square both sides of the inequality, (iii) is equivalent to showing that $(\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})\leq|\mathbf{x}|^2+2|\mathbf{x}||\mathbf{y}|+|\mathbf{y}|^2$. Since $(\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})=|\mathbf{x}|^2+2(\mathbf{x}\cdot\mathbf{y})+|\mathbf{y}|^2$, the problem reduces to showing that $\mathbf{x}\cdot\mathbf{y}\leq|\mathbf{x}||\mathbf{y}|$, that is, that

$$\sum_{k=1}^{n} x_k y_k \leq \left(\sum_{k=1}^{n} x_k^2\right)^{1/2}\left(\sum_{k=1}^{n} y_k^2\right)^{1/2}. \tag{1.2}$$

This important inequality is called the *Schwarz* (or *Cauchy–Schwarz*) *inequality* and can be proved as follows. For $\alpha,\beta\in\mathbf{R^1}$, the fact that $(\alpha-\beta)^2\geq 0$ gives $\alpha\beta\leq\frac{1}{2}\alpha^2+\frac{1}{2}\beta^2$. Therefore, $\sum_{k=1}^{n} x_k y_k\leq\sum_{k=1}^{n}\left(\frac{1}{2}x_k^2+\frac{1}{2}y_k^2\right)=\frac{1}{2}(|\mathbf{x}|^2+|\mathbf{y}|^2)$. Inequality (1.2) follows immediately if $|\mathbf{x}|=|\mathbf{y}|=1$, since then $\sum_{k=1}^{n}\mathbf{x}_k\mathbf{y}_k\leq\frac{1}{2}(1+1)=1=|\mathbf{x}||\mathbf{y}|$. Moreover, (1.2) is obvious if either $|\mathbf{x}|=0$ or $|\mathbf{y}|=0$ since then both sides must be zero. Finally, if $|\mathbf{x}|>0$ and $|\mathbf{y}|>0$, let $x_k'=x_k/|\mathbf{x}|$, $y_k'=y_k/|\mathbf{y}|$, $\mathbf{x}'=(x_1',\ldots,x_n')=\mathbf{x}/|\mathbf{x}|$, and $\mathbf{y}'=(y_1',\ldots,y_n')=\mathbf{y}/|\mathbf{y}|$. Then $|\mathbf{x}'|=|\mathbf{y}'|=1$, so that by the case already proved, $\sum_{k=1}^{n}x_k'y_k'\leq 1$; that is, $\sum_{k=1}^{n}x_k y_k\leq|\mathbf{x}||\mathbf{y}|$, as claimed.

If we now define the *distance between two points* $\mathbf{x}$ and $\mathbf{y}$ by $d(\mathbf{x},\mathbf{y})=|\mathbf{x}-\mathbf{y}|$, we immediately obtain the characteristic metric space properties:

(i) $d(\mathbf{x},\mathbf{y})=d(\mathbf{y},\mathbf{x})$

(ii) $d(\mathbf{x},\mathbf{y})\geq 0$, and $d(\mathbf{x},\mathbf{y})=0$ if and only if $\mathbf{x}=\mathbf{y}$

(iii) $d(\mathbf{x},\mathbf{y})\leq d(\mathbf{x},\mathbf{z})+d(\mathbf{z},\mathbf{y})$

We have used the symbol x_k to denote the kth coordinate of $\mathbf{x}$. When no confusion should arise, we will also use $\{\mathbf{x}_k\}$ to denote a *sequence of points* of $\mathbf{R^n}$. If $\mathbf{x}\in\mathbf{R^n}$, we say that a sequence $\{\mathbf{x}_k\}$ *converges* to $\mathbf{x}$, or that $\mathbf{x}$ is the *limit point* of $\{\mathbf{x}_k\}$, if $|\mathbf{x}-\mathbf{x}_k|\to 0$ as $k\to\infty$. We denote this by writing either

$\mathbf{x} = \lim_{k\to\infty} \mathbf{x}_k$ or $\mathbf{x}_k \to \mathbf{x}$ as $k \to \infty$. A point $\mathbf{x} \in \mathbf{R^n}$ is called a *limit point of a set* E if it is the limit point of a sequence of distinct points of E. A point $\mathbf{x} \in E$ is called an *isolated point* of E if it is not the limit of any sequence in E (excluding the trivial sequence $\{\mathbf{x}_k\}$ where $\mathbf{x}_k = \mathbf{x}$ for all k). It follows that $\mathbf{x}$ is isolated if and only if there is a $\delta > 0$ such that $|\mathrm{x} - \mathrm{y}| > \delta$ for every $\mathbf{y} \in E$, $\mathbf{y} \neq \mathbf{x}$.

For sequences $\{\mathbf{x}_k\}$ in $\mathbf{R^1}$, we will write $\lim_{k\to\infty} x_k = +\infty$, or $x_k \to +\infty$ as $k \to \infty$, if given $M > 0$ there is an integer K such that $x_k \geq M$ whenever $k \geq K$. A similar definition holds for $\lim_{k\to\infty} x_k = -\infty$.

A sequence $\{\mathbf{x}_k\}$ in $\mathbf{R^n}$ is called a *Cauchy sequence* if given $\varepsilon > 0$ there is an integer K such that $|\mathbf{x}_k - \mathbf{x}_j| < \varepsilon$ for all $k, j \geq K$. We leave it as an exercise to prove that $\mathbf{R^n}$ is a *complete metric space*, that is, that every Cauchy sequence in $\mathbf{R^n}$ converges to a point of $\mathbf{R^n}$.

A set $E \subset E_1$ is said to be *dense* in E_1 if for every $\mathbf{x}_1 \in E_1$ and $\varepsilon > 0$ there is a point $\mathbf{x} \in E$ such that $0 < |\mathbf{x} - \mathbf{x}_1| < \varepsilon$. Thus, E is dense in E_1 if every point of E_1 is a limit point of E. If $E = E_1$, we say E is *dense in itself*. As an example, the set of points of $\mathbf{R^n}$ each of whose coordinates is a rational number is dense in $\mathbf{R^n}$. Since this set is also countable, it follows that $\mathbf{R^n}$ is *separable*, by which we mean that $\mathbf{R^n}$ has a countable dense subset.

For nonempty subsets E of $\mathbf{R^1}$, we use the standard notations $\sup E$ and $\inf E$ for the *supremum* (*least upper bound*) and *infimum* (*greatest lower bound*) of E. In case $\sup E$ belongs to E, it will be called $\max E$; similarly, $\inf E$ will be called $\min E$ if it belongs to E.

If $\{a_k\}_{k=1}^{\infty}$ is a sequence of points in $\mathbf{R^1}$, let $b_j = \sup_{k\geq j} a_k$ and $c_j = \inf_{k\geq j} a_k, j = 1, 2, \ldots$. Then $-\infty \leq c_j \leq b_j \leq +\infty$, and $\{b_j\}$ and $\{c_j\}$ are monotone decreasing and increasing, respectively; that is, $b_j \geq b_{j+1}$ and $c_j \leq c_{j+1}$. Define $\limsup_{k\to\infty} a_k$ and $\liminf_{k\to\infty} a_k$ by

$$\begin{aligned} \limsup_{k\to\infty} a_k &= \lim_{j\to\infty} b_j = \lim_{j\to\infty} \left\{ \sup_{k\geq j} a_k \right\}, \\ \liminf_{k\to\infty} a_k &= \lim_{j\to\infty} c_j = \lim_{j\to\infty} \left\{ \inf_{k\geq j} a_k \right\}. \end{aligned} \tag{1.3}$$

We leave it as an exercise to show that $-\infty \leq \liminf_{k\to\infty} a_k \leq \limsup_{k\to\infty} a_k \leq +\infty$ and that the following characterizations hold.

Theorem 1.4

(a) $L = \limsup_{k\to\infty} a_k$ *if and only if* (i) *there is a subsequence* $\{a_{k_j}\}$ *of* $\{a_k\}$ *that converges to* L *and* (ii) *if* $L' > L$, *there is an integer* K *such that* $a_k < L'$ *for* $k \geq K$.

(b) $l = \liminf_{k\to\infty} a_k$ *if and only if* (i) *there is a subsequence* $\{a_{k_j}\}$ *of* $\{a_k\}$ *that converges to* l *and* (ii) *if* $l' < l$, *there is an integer* K *such that* $a_k > l'$ *for* $k \geq K$.

Thus, when they are finite, $\limsup_{k\to\infty} a_k$ and $\liminf_{k\to\infty} a_k$ are the largest and smallest limit points of $\{a_k\}$, respectively. We leave it as a problem to show that $\{a_k\}$ converges to $a, -\infty \le a \le +\infty$, if and only if $\limsup_{k\to\infty} a_k = \liminf_{k\to\infty} a_k = a$.

We can also use the metric on $\mathbf{R^n}$ to define the *diameter of a set* E by letting

$$\delta(E) = \operatorname{diam} E = \sup\left\{\left|\mathbf{x}-\mathbf{y}\right| : \mathbf{x},\mathbf{y} \in E\right\}.$$

If the diameter of E is finite, E is said to be *bounded*. Equivalently, E is bounded if there is a finite constant M such that $|\mathbf{x}| \le M$ for all $\mathbf{x} \in E$. If E_1 and E_2 are two sets, the *distance between* E_1 *and* E_2 is defined by

$$d(E_1, E_2) = \inf\{|\mathbf{x}-\mathbf{y}| : \mathbf{x} \in E_1, \mathbf{y} \in E_2\}.$$

1.3 Open and Closed Sets in $\mathbf{R^n}$, and Special Sets

For $\mathbf{x} \in \mathbf{R^n}$ and $\delta > 0$, the set

$$B(\mathbf{x};\delta) = \{\mathbf{y} : |\mathbf{x}-\mathbf{y}| < \delta\}$$

is called the *open ball with center* $\mathbf{x}$ *and radius* δ. A point $\mathbf{x}$ of a set E is called an *interior point* of E if there exists $\delta > 0$ such that $B(\mathbf{x};\delta) \subset E$. The collection of all interior points of E is called the *interior* of E and denoted E°. A set E is said to be *open* if $E = E^\circ$; that is, E is open if for each $\mathbf{x} \in E$ there exists $\delta > 0$ such that $B(\mathbf{x};\delta) \subset E$. The empty set $\emptyset$ is open by convention. The whole space $\mathbf{R^n}$ is clearly open, and we leave it as an exercise to prove that $B(\mathbf{x};\delta)$ is open. We will generally denote open sets by the letter G.

A set E is called *closed* if CE is open. Note that $\emptyset$ and $\mathbf{R^n}$ are closed. Closed sets will generally be denoted by the letter F. The union of a set E and all its limit points is called the *closure* of E and written $\bar{E}$. By the *boundary* of E, we mean the set $\bar{E} - E^\circ$. We leave it to the reader to prove the following facts.

Theorem 1.5

(i) $\overline{B(\mathbf{x};\delta)} = \left\{\mathbf{y} : \left|\mathbf{x}-\mathbf{y}\right| \le \delta\right\}$.

(ii) *E is closed if and only if $E = \bar{E}$; that is, E is closed if and only if it contains all its limit points.*

(iii) *$\bar{E}$ is closed, and $\bar{E}$ is the smallest closed set containing E; that is, if F is closed and $E \subset F$, then $\bar{E} \subset F$.*

The open subsets of $\mathbf{R^n}$ satisfy the properties listed in the next theorem.

Theorem 1.6

(i) *The union of any number of open sets is open.*
(ii) *The intersection of a finite number of open sets is open.*

Verification is left to the reader. Using the De Morgan laws, we obtain the following equivalent statements.

Theorem 1.7

(i) *The intersection of any number of closed sets is closed.*
(ii) *The union of a finite number of closed sets is closed.*

A subset E_1 of E is said to be *relatively open with respect to* E if it can be written $E_1 = E \cap G$ for some open set G. Similarly, E_1 is *relatively closed with respect to* E if $E_1 = E \cap F$ for some closed F. Note that the relative complement of a relatively open set is relatively closed. A useful alternate characterization of relatively closed is as follows.

Theorem 1.8 *A set $E_1 \subset E$ is relatively closed with respect to E if and only if $E_1 = E \cap \bar{E}_1$, that is, if and only if every limit point of E_1 that lies in E is in E_1.*

The proof is left as an exercise.

Consider a collection $\{A\}$ of sets A. Then a set is said to be of *type* A_δ if it can be written as a countable intersection of sets A and to be of *type* A_σ if it can be written as a countable union of sets A. Thus, "δ" stands for intersection and "σ" for union. The most common uses of this notation are G_δ and F_σ, where $\{G\}$ denotes the open sets in $\mathbf{R^n}$ and $\{F\}$ the closed sets. Hence, H is of *type* G_δ if

$$H = \bigcap_k G_k,\ G_k \text{ open},$$

and H is of *type* F_σ if

$$H = \bigcup_k F_k,\ F_k \text{ closed}.$$

The complement of a G_δ set is an F_σ set, and vice versa. A G_δ (F_σ) set is of course not generally open (closed); in fact, any closed (open) set in $\mathbf{R^n}$ is of

type G_δ (F_σ): see Exercise l(j). These two special types of sets will be very useful later in the measure approximation of general sets.

Another special type of set that we will have occasion to use is a *perfect set*, by that we mean a closed set C each of whose points is a limit point of C. Thus, a perfect set is a closed set that is dense in itself. One particular property of perfect sets we will use is stated in the following theorem. The proof is postponed until Section 1.4.

Theorem 1.9 *A perfect set is uncountable.*

Other special sets that will be important are n-dimensional intervals. When $n=1$ and $a<b$, we will use the usual notations $[a,b]=\{x:a\le x\le b\}$, $(a,b)=\{x:a<x<b\}$, $[a,b)=\{x:a\le x<b\}$, and $(a,b]=\{x:a<x\le b\}$ for *closed, open, and partly open intervals*. Whenever we use just the word interval, we generally mean closed interval. An *n-dimensional interval* I is a subset of $\mathbf{R^n}$ of the form $I=\{\mathbf{x}=(x_1,\ldots,x_n)\ :\ a_k\le x_k\le b_k, k=1,\ldots,n\}$, where $a_k<b_k, k=1,\ldots,n$. An interval is thus closed, and we say it has edges parallel to the coordinate axes. If the edge lengths b_k-a_k are all equal, I will be called an *n-dimensional cube* with edges parallel to the coordinate axes. Cubes will usually be denoted by the letter Q. Two intervals I_1 and I_2 are said to be *nonoverlapping* if their interiors are disjoint, that is, if the most they have in common is some part of their boundaries. A set equal to an interval minus some part of its boundary will be called a *partly open interval*. By definition, the *volume* $v(I)$ of the interval $I=\{(x_1,\ldots,x_n):a_k\le x_k\le b_k, k=1,\ldots,n\}$ is

$$v(I)=\prod_{k=1}^{n}(b_k-a_k).$$

Somewhat more generally, if $\{\mathbf{e}_k\}_{k=1}^n$ is any given set of n vectors emanating from a point in $\mathbf{R^n}$, we will consider the closed *parallelepiped*

$$P=\{\mathbf{x}:\mathbf{x}=\sum_{k=1}^{n}t_k\mathbf{e}_k, 0\le t_k\le 1\}.$$

Note that the edges of P are parallel translates of the $\mathbf{e}_k$. Thus, P is an interval if the $\mathbf{e}_k$ are parallel to the coordinate axes. The *volume* $v(P)$ of P is *by definition* the absolute value of the $n\times n$ determinant having $\mathbf{e}_1,\ldots,\mathbf{e}_n$ as rows.* In case P is an interval, this definition agrees with the one given earlier. A linear transformation T of $\mathbf{R^n}$ transforms a parallelpiped P into a parallelpiped

* See, for example, G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, 3rd ed., Macmillan, New York, 1965, Theorem 8, p. 290.

P' with volume $v(P') = |\det T| v(P)$.* In particular, a rotation of axes in $\mathbf{R^n}$ (which is an orthogonal linear transformation) does not change the volume of a parallelepiped. We will assume basic facts about volume: for example, if N is finite and P is a parallelepiped with $P \subset \bigcup_1^N I_k$, then $v(P) \leq \sum_1^N v(I_k)$, and if $\{I_k\}_1^N$ are nonoverlapping intervals contained in a parallelepiped P, then $\sum_1^N v(I_k) \leq v(P)$.

We shall use the notion of interval to obtain a basic decomposition of open sets in $\mathbf{R^n}$. We consider first the case $n = 1$, which is somewhat simpler than $n > 1$.

Theorem 1.10 *Every open set in $\mathbf{R^1}$ can be written as a countable union of disjoint open intervals.*

Proof. Let G be an open set in $\mathbf{R^1}$. For $x \in G$, let I_x denote the maximal open interval containing x which is in G; that is, I_x is the union of all open intervals that contain x and that lie in G. If $x, x' \in G$ and $x \neq x'$, then I_x and $I_{x'}$ must either be disjoint or identical, since if they intersect, their union is an open interval containing x and x'. Clearly, $G = \bigcup_{x \in G} I_x$. Since each I_x contains a rational number, the number of distinct I_x must be countable, and the theorem follows.

The construction used in this proof fails in $\mathbf{R^n}$ if $n > 1$, since the union of (overlapping) intervals is not generally an interval. The theorem itself fails when $n > 1$, as is easily seen by considering any open ball. As a substitute, we have the following useful result.

Theorem 1.11 *Every open set in $\mathbf{R^n}, n \geq 1$, can be written as a countable union of nonoverlapping (closed) cubes. It can also be written as a countable union of disjoint partly open cubes.*

Proof. Consider the lattice of points of $\mathbf{R^n}$ with integral coordinates and the corresponding net K_0 of cubes with edge length 1 and vertices at these lattice points. Bisecting each edge of a cube in K_0, we obtain from it 2^n subcubes of edge length $\frac{1}{2}$. The total collection of these subcubes for every cube in K_0 forms a net K_1 of cubes. If we continue bisecting, we obtain finer and finer nets K_j of cubes such that each cube in K_j has edge length 2^{-j} and is the union of 2^n nonoverlapping cubes in K_{j+1}.

Now let G be any open set in $\mathbf{R^n}$. Let S_0 be the collection of all cubes in K_0 that lie entirely in G. Let S_1 be those cubes in K_1 that lie in G but that are not

* See, for example, G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, 3rd ed., Macmillan, New York, 1965, Theorem 9, p. 290.

subcubes of any cube in S_0. More generally, for $j \geq 1$, let S_j be the cubes in K_j that lie in G but that are not subcubes of any cube in $S_0, \ldots, S_{j-1}$. If S denotes the total collection of cubes from all the S_j, then S is countable since each K_j is countable, and the cubes in S are nonoverlapping by construction. Moreover, since G is open and the cubes in K_j become arbitrarily small as $j \to \infty$, each point of G will eventually be caught in a cube in some S_j. Hence, $G = \bigcup_{Q \in S} Q$, which proves the first statement. The proof of the second statement is left to the reader.

The collection $\{Q\colon Q \in K_j, j = 1, 2, \ldots\}$ constructed above is called a family of dyadic cubes. In general, by *dyadic cubes*, we mean the family of cubes obtained from repeated bisection of any initial net of cubes in $\mathbf{R^n}$. Note that the family of dyadic cubes used in the proof of Theorem 1.11 could be replaced by one in which the initial net consists of cubes of any fixed edge length.

It follows from Theorem 1.10 that any closed set in $\mathbf{R^1}$ can be constructed by deleting a countable number of open disjoint intervals from $\mathbf{R^1}$. A perfect set results by removing the intervals in such a way as to create no isolated points; thus, we would not remove any two open intervals with a common endpoint.

1.4 Compact Sets and the Heine–Borel Theorem

By a *cover* of a set E, we mean a family $\mathscr{F}$ of sets A such that $E \subset \bigcup_{A \in F} A$. A *subcover* $\mathscr{F}_1$ of a cover $\mathscr{F}$ is a cover with the property that $A_1 \in \mathscr{F}$ whenever $A_1 \in \mathscr{F}_1$. A cover $\mathscr{F}$ is called an *open cover* if each set in $\mathscr{F}$ is open. We say E is *compact* if every open cover of E has a finite subcover. Two equivalent statements, whose proofs are left as exercises, are as follows.

Theorem 1.12

(i) (*The Heine–Borel theorem*) *A set $E \subset \mathbf{R^n}$ is compact if and only if it is closed and bounded.*

(ii) *A set $E \subset \mathbf{R^n}$ is compact if and only if every sequence of points of E has a subsequence that converges to a point of E.*

We leave it as an exercise to show that the distance between two nonempty, compact, disjoint sets is positive and that the intersection of a countable sequence of decreasing, nonempty, compact sets is nonempty. Thus, a nested sequence of closed intervals has a nonempty intersection. See also Exercise 12.

With these facts, we can now prove Theorem 1.9.

Proof of Theorem 1.9. Let C be a perfect set in $\mathbf{R^n}$, and suppose that C is countable: $C = \{\mathbf{c}_k\}_{k=1}^{\infty}$. Let $C_k = C - \{\mathbf{c}_k\}, k \geq 1$. Given $\mathbf{x}_1 \in C_1$, let Q_1 be a (closed) cube with center $\mathbf{x}_1$ such that $\mathbf{c}_1 \notin Q_1$. Then $Q_1 \cap C$ is compact (closed and bounded) and not empty. Since $\mathbf{x}_1 \in C$ and C is perfect, $\mathbf{x}_1$ is a limit point of C and so also of C_2. It follows that $C_2 \cap Q_1^\circ$ is not empty. Let $\mathbf{x}_2 \in C_2 \cap Q_1^\circ$ and choose a cube Q_2 with center $\mathbf{x}_2$ such that $Q_2 \subset Q_1$ and $\mathbf{c}_2 \notin Q_2$. Then $Q_2 \cap C$ is a compact, nonempty subset of $Q_1 \cap C$. Continuing in this way, we obtain a decreasing sequence $Q_k \cap C$ of compact, nonempty sets such that $\mathbf{c}_k \notin Q_k$. It follows that $\bigcap_k (Q_k \cap C)$ is a nonempty subset of C that contains no $\mathbf{c}_k$. This contradiction proves that C must be uncountable and establishes the theorem.

1.5 Functions

By a function $f = f(\mathbf{x})$ defined for $\mathbf{x}$ in a set $E \subset \mathbf{R^n}$, we will always mean a *real-valued* function, unless explicitly stated otherwise. By *real-valued*, we generally mean *extended real-valued*, that is, f may take the values $\pm\infty$; if $|f(\mathbf{x})| < +\infty$ for all $\mathbf{x} \in E$, we say f is *finite* (or *finite-valued*) on E. A finite function f is said to be *bounded* on E if there is a finite number M such that $|f(\mathbf{x})| \leq M$ for $\mathbf{x} \in E$; that is, f is bounded on E if $\sup_{x \in E} |f(\mathbf{x})|$ is finite. A sequence $\{f_k\}$ of functions is said to be *uniformly bounded* on E if there is a finite M such that $|f_k(\mathbf{x})| \leq M$ for $\mathbf{x} \in E$ and all k.

By the *support* of f, we mean the closure of the set where f is not zero. Thus, the support of a function is always closed. It follows that a function defined in $\mathbf{R^n}$ has *compact support* if and only if it vanishes outside some bounded set.

A function f defined on an interval I in $\mathbf{R^1}$ is called *monotone increasing* (*decreasing*) if $f(x) \leq f(y)$ $[f(x) \geq f(y)]$ whenever $x < y$ and $x, y \in I$. By *strictly* monotone increasing (decreasing), we mean that $f(x) < f(y)$ $[f(x) > f(y)]$ if $x < y$ and $x, y \in I$.

Let f be defined on $E \subset \mathbf{R^n}$ and let $\mathbf{x}_0$ be a limit point of E. Let $B'(\mathbf{x}_0; \delta) = B(\mathbf{x}_0; \delta) - \{\mathbf{x}_0\}$ denote the punctured ball with center $\mathbf{x}_0$ and radius δ, and let

$$M(\mathbf{x}_0; \delta) = \sup_{\mathbf{x} \in B'(\mathbf{x}_0; \delta) \cap E} f(\mathbf{x}), \qquad m(\mathbf{x}_0; \delta) = \inf_{\mathbf{x} \in B'(\mathbf{x}_0; \delta) \cap E} f(\mathbf{x}).$$

As $\delta \searrow 0$, $M(\mathbf{x}_0; \delta)$ decreases and $m(\mathbf{x}_0; \delta)$ increases, and we define

$$\begin{aligned} \limsup_{\mathbf{x} \to \mathbf{x}_0; \mathbf{x} \in E} f(\mathbf{x}) &= \lim_{\delta \to 0} M(\mathbf{x}_0; \delta), \\ \liminf_{\mathbf{x} \to \mathbf{x}_0; \mathbf{x} \in E} f(\mathbf{x}) &= \lim_{\delta \to 0} m(\mathbf{x}_0; \delta). \end{aligned} \tag{1.13}$$

We leave it as an exercise to show that the following characterizations are valid.

Theorem 1.14

(a) $M = \limsup_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\in E} f(\mathbf{x})$ *if and only if* (i) *there exists* $\{\mathbf{x}_k\}$ *in* $E-\{\mathbf{x}_0\}$ *such that* $\mathbf{x}_k \to \mathbf{x}_0$ *and* $f(\mathbf{x}_k) \to M$ *and* (ii) *if* $M' > M$, *there exists* $\delta > 0$ *such that* $f(\mathbf{x}) < M'$ *for* $\mathbf{x} \in B'(\mathbf{x}_0; \delta) \cap E$.

(b) $m = \liminf_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\in E} f(\mathbf{x})$ *if and only if* (i) *there exists* $\{\mathbf{x}_k\}$ *in* $E-\{\mathbf{x}_0\}$ *such that* $\mathbf{x}_k \to \mathbf{x}_0$ *and* $f(\mathbf{x}_k) \to m$ *and* (ii) *if* $m' < m$, *there exists* $\delta > 0$ *such that* $f(\mathbf{x}) > m'$ *for* $\mathbf{x} \in B'(\mathbf{x}_0; \delta) \cap E$.

We also define $\limsup_{|\mathbf{x}|\to\infty;\mathbf{x}\in E} f(\mathbf{x})$ and $\liminf_{|\mathbf{x}|\to\infty;\mathbf{x}\in E} f(\mathbf{x})$. For example, $M = \limsup_{|\mathbf{x}|\to\infty;\mathbf{x}\in E} f(\mathbf{x})$ means (i) there exist $\{\mathbf{x}_k\}$ in E such that $|\mathbf{x}_k| \to \infty$ and $f(\mathbf{x}_k) \to M$ and (ii) if $M' > M$, there exists N such that $f(\mathbf{x}) < M'$ if $|\mathbf{x}| > N$ and $\mathbf{x} \in E$. These notions should not be confused with $\limsup_{k\to\infty} f_k(\mathbf{x})$ and $\liminf_{k\to\infty} f_k(\mathbf{x})$, which denote the lim sup and lim inf of the sequence $\{f_k(\mathbf{x})\}$.

1.6 Continuous Functions and Transformations

A function f defined in a neighborhood of $\mathbf{x}_0$ is said to be *continuous* at $\mathbf{x}_0$ if $f(\mathbf{x}_0)$ is finite and $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$. If f is not continuous at $\mathbf{x}_0$, it follows that unless $f(\mathbf{x}_0)$ is infinite, either $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})$ does not exist or is different from $f(\mathbf{x}_0)$.

For functions on $\mathbf{R}^1$, we will use the notation

$$f(x_0+) = \lim_{x\to x_0; x>x_0} f(x) \quad \text{and} \quad f(x_0-) = \lim_{x\to x_0; x<x_0} f(x)$$

for the *right-* and *left-hand limits* of f at x_0, when they exist. If $f(x_0+)$, $f(x_0-)$, and $f(x_0)$ exist and are finite, but f is not continuous at x_0, then either $f(x_0+) \neq f(x_0-)$ or $f(x_0+) = f(x_0-) \neq f(x_0)$. In the first case, x_0 is called a *jump discontinuity* of f and in the second, a *removable discontinuity* of f (since by changing the value of f at x_0, we can make it continuous there). Such discontinuities are said to be of the *first kind*, as distinguished from those of the *second kind*, for which either $f(x_0+)$ or $f(x_0-)$ does not exist or for which $f(x_0+)$, $f(x_0-)$ or $f(x_0)$ is infinite.

If f is defined only in a set E containing $\mathbf{x}_0$, $E \subset \mathbf{R}^n$, then f is said to be *continuous at* $\mathbf{x}_0$ *relative to* E if $f(\mathbf{x}_0)$ is finite and either $\mathbf{x}_0$ is an isolated point of E or $\mathbf{x}_0$ is a limit point of E and $\lim_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\in E} f(\mathbf{x}) = f(\mathbf{x}_0)$. If $E_1 \subset E$, a function

is said to be *continuous in E_1 relative to E* if it is continuous relative to E at every point of E_1. The proofs of the following basic facts are left as exercises.

Theorem 1.15 *Let E be a compact set in $\mathbf{R^n}$ and f be continuous in E relative to E. Then the following are true:*

(i) *f is bounded on E; that is, $\sup_{\mathbf{x}\in E} |f(\mathbf{x})| < \infty$.*

(ii) *f attains its supremum and infimum on E; that is, there exist $\mathbf{x}_1, \mathbf{x}_2 \in E$ such that $f(\mathbf{x}_1) = \sup_{\mathbf{x}\in E} f(\mathbf{x}), f(\mathbf{x}_2) = \inf_{\mathbf{x}\in E} f(\mathbf{x})$.*

(iii) *f is uniformly continuous on E relative to E; that is, given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$ if $|\mathbf{x} - \mathbf{y}| < \delta$ and $\mathbf{x}, \mathbf{y} \in E$.*

A sequence of functions $\{f_k\}$ defined on E is said to *converge uniformly* on E to a finite f if given $\varepsilon > 0$, there exists K such that $|f_k(\mathbf{x}) - f(\mathbf{x})| < \varepsilon$ for $k \geq K$ and $\mathbf{x} \in E$. We will use the following fact, whose proof is again left to the reader.

Theorem 1.16 *Let $\{f_k\}$ be a sequence of functions defined on E that are continuous in E relative to E and that converge uniformly on E to a finite f. Then f is continuous in E relative to E.*

A *transformation T of a set $E \subset \mathbf{R^n}$ into $\mathbf{R^n}$* is a mapping $\mathbf{y} = T\mathbf{x}$ that carries points $\mathbf{x} \in E$ into points $\mathbf{y} \in \mathbf{R^n}$. If $\mathbf{y} = (y_1, \ldots, y_n)$, then T can be identified with the collection of coordinate functions $y_k = f_k(\mathbf{x})$, $k = 1, \ldots, n$, which are induced by T. The *image of* E under T is the set $\{\mathbf{y}\colon \mathbf{y} = T\mathbf{x}$ for some $\mathbf{x} \in E\}$. T is continuous at $\mathbf{x}_0 \in E$ relative to E (by which we mean $\lim_{\mathbf{x}\to\mathbf{x}_0;\mathbf{x}\in E} T\mathbf{x} = T\mathbf{x}_0$) if and only if each f_k is continuous at $\mathbf{x}_0$ relative to E. We will use the following result in Chapter 3.

Theorem 1.17 *Let $\mathbf{y} = T\mathbf{x}$ be a transformation of $\mathbf{R^n}$ that is continuous in E relative to E. If E is compact, then so is its image TE.*

1.7 The Riemann Integral

We shall see that the Lebesgue integral is more general than the Riemann integral, in the sense that whenever the Riemann integral of a function exists, then so does its Lebesgue integral, and the two are equal (Theorem 5.52).

The Riemann integral is nonetheless useful, its significance being simplicity and computability.

If f is defined and bounded on an interval $I = \{\mathbf{x} : \mathbf{x} = (x_1, \ldots, x_n), a_k \le x_k \le b_k, k = 1, \ldots, n\}$ in $\mathbf{R^n}$, its Riemann integral will be denoted by

$$(R) \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \text{ or } (R) \int_I f(\mathbf{x})\, d\mathbf{x} \tag{1.18}$$

and is defined as follows. Partition I into a finite collection Γ of nonoverlapping intervals, $\Gamma = \{I_k\}_{k=1}^N$, and define the *norm* $|\Gamma|$ of Γ by $|\Gamma| = \max_k (\operatorname{diam} I_k)$. Select a point $\boldsymbol{\xi}_k$ in I_k for $k \ge 1$, and let

$$\begin{aligned} R_\Gamma &= R_\Gamma(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N) = \sum_{k=1}^N f(\boldsymbol{\xi}_k) v(I_k), \\ U_\Gamma &= \sum_{k=1}^N [\sup_{\mathbf{x} \in I_k} f(\mathbf{x})] v(I_k), \quad L_\Gamma = \sum_{k=1}^N [\inf_{\mathbf{x} \in I_k} f(\mathbf{x})] v(I_k). \end{aligned} \tag{1.19}$$

We then define the Riemann integral by saying that $A = (R)\int_I f(\mathbf{x})\, d\mathbf{x}$ if $\lim_{|\Gamma| \to 0} R_\Gamma$ exists and equals A; that is, if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|A - R_\Gamma| < \varepsilon$ for any Γ and any chosen $\{\boldsymbol{\xi}_k\}$, provided only that $|\Gamma| < \delta$. This definition is actually equivalent to the statement that

$$\inf_\Gamma U_\Gamma = \sup_\Gamma L_\Gamma = A. \tag{1.20}$$

The integral of course exists if f is continuous on I. Proofs of these facts are left as exercises; the treatment given in Chapter 2 for the Riemann–Stieltjes integrals should serve as a review for many facts about Riemann integrals. See also Theorem 5.54.

Exercises

1. Prove the following facts, which were left earlier as exercises.

(a) For a sequence of sets $\{E_k\}$, $\limsup E_k$ consists of those points that belong to infinitely many E_k, and $\liminf E_k$ consists of those points that belong to all E_k from some k on.

(b) The De Morgan laws.

(c) Every Cauchy sequence in $\mathbf{R^n}$ converges to a point of $\mathbf{R^n}$. (This can be deduced from its analogue in $\mathbf{R^1}$ by noting that the entries in a

given coordinate position of the points in a Cauchy sequence in $\mathbf{R^n}$ form a Cauchy sequence in $\mathbf{R^1}$.)

(d) Theorem 1.4.

(e) A sequence $\{a_k\}$ in $\mathbf{R^1}$ converges to a, $-\infty \leq a \leq +\infty$, if and only if $\limsup_{k\to\infty} a_k = \liminf_{k\to\infty} a_k = a$.

(f) $B(\mathbf{x}; \delta)$ is open.

(g) Theorem 1.5.

(h) Theorems 1.6 and 1.7.

(i) Theorem 1.8.

(j) Any closed (open) set in $\mathbf{R^n}$ is of type G_δ (F_σ). (If F is closed, consider the sets $\{\mathbf{x} : \text{dist}(\mathbf{x}, F) < (1/k)\}$, $k = 1, 2, \ldots$.)

(k) Theorem 1.12.

(l) The distance between two nonempty, compact, disjoint sets in $\mathbf{R^n}$ is positive. See also Exercise 12.

(m) The intersection of a countable sequence of decreasing, nonempty, compact sets is nonempty.

(n) Theorem 1.14.

(o) Theorem 1.15.

(p) Theorem 1.16.

(q) Theorem 1.17.

(r) The Riemann integral $A = (R)\int_I f(\mathbf{x})\, d\mathbf{x}$ of a bounded f over an interval I exists if and only if $\inf_\Gamma U_\Gamma = \sup_\Gamma L_\Gamma = A$.

(s) If f is continuous on an interval I, then $(R)\int_I f(\mathbf{x})\, d\mathbf{x}$ exists.

2. Find $\limsup E_k$ and $\liminf E_k$ if $E_k = [-(1/k), 1]$ for k odd and $E_k = [-1, (1/k)]$ for k even.

3. (a) Show that $C(\limsup E_k) = \liminf CE_k$.

(b) Show that if $E_k \nearrow E$ or $E_k \searrow E$, then $\limsup E_k = \liminf E_k = E$.

4. (a) Show that $\limsup_{k\to\infty}(-a_k) = -\liminf_{k\to\infty} a_k$.

(b) Show that $\limsup_{k\to\infty}(a_k + b_k) \leq \limsup_{k\to\infty} a_k + \limsup_{k\to\infty} b_k$, provided that the expression on the right does not have the form $\infty + (-\infty)$ or $-\infty + \infty$.

(c) If $\{a_k\}$ and $\{b_k\}$ are nonnegative, bounded sequences, show that $\limsup_{k\to\infty}(a_k b_k) \leq (\limsup_{k\to\infty} a_k)(\limsup_{k\to\infty} b_k)$.

(d) Give examples for which the inequalities in parts (b) and (c) are not equalities. Show that if either $\{a_k\}$ or $\{b_k\}$ converges, equality holds in (b) and (c).

5. Find analogues of the statements in Exercise 4 for $\limsup_{\mathbf{x}\to\mathbf{x}_0; \mathbf{x}\in E} f(\mathbf{x})$.

6. Compare $\limsup_{k\to\infty} a_k$ and $\limsup(-\infty, a_k)$.

7. Show that $E_1^\circ \cap E_2^\circ = (E_1 \cap E_2)^\circ$ and $E_1^\circ \cup E_2^\circ \subset (E_1 \cup E_2)^\circ$. Give an example when $E_1^\circ \cup E_2^\circ \neq (E_1 \cup E_2)^\circ$.

8. Let E be a set in $\mathbf{R^n}$ that is relatively open with respect to an interval I. Show that E can be written as a countable union of nonoverlapping intervals.

9. Prove that any closed subset of a compact set is compact.

10. Let $\{\mathbf{x}_k\}$ be a bounded infinite sequence in $\mathbf{R^n}$. Show that $\{\mathbf{x}_k\}$ has a limit point. (This is the Bolzano–Weierstrass theorem in $\mathbf{R^n}$.)

11. Give an example of a decreasing sequence of nonempty closed sets in $\mathbf{R^n}$ whose intersection is empty.

12. (a) Give an example of two disjoint, nonempty, closed sets E_1 and E_2 in $\mathbf{R^n}$ for which $d(E_1, E_2) = 0$.

(b) Let E_1, E_2 be nonempty sets in $\mathbf{R^n}$ with E_1 closed and E_2 compact. Show that there are points $x_1 \in E_1$ and $x_2 \in E_2$ such that $d(E_1, E_2) = |x_1 - x_2|$. Deduce that $d(E_1, E_2)$ is positive if such E_1, E_2 are disjoint.

13. If f is defined and uniformly continuous on E, show there is a function $\bar{f}$ defined and continuous on $\bar{E}$ such that $\bar{f} = f$ on E.

14. If f is defined and uniformly continuous on a bounded set E, show that f is bounded on E.

15. Show that a bounded f is Riemann integrable on I if and only if given $\varepsilon > 0$, there is a partition Γ of I such that $0 \leq U_\Gamma - L_\Gamma < \varepsilon$. (Exercise 1(r) may be helpful.)

16. If $\{f_k\}$ is a sequence of bounded, Riemann integrable functions on an interval I that converges uniformly on I to f, show that f is Riemann integrable on I and that

$$(R)\int_I f_k(\mathbf{x})\,d\mathbf{x} \to (R)\int_I f(\mathbf{x})\,d\mathbf{x}.$$

17. Let f be a finite function on $\mathbf{R^n}$ and define

$$\omega(\delta) = \sup\{|f(\mathbf{x}) - f(\mathbf{y})| : |\mathbf{x} - \mathbf{y}| \leq \delta\},$$

$\delta > 0$, to be the *modulus of continuity* of f. Show that $\omega(\delta)$ decreases as δ decreases to 0 and that f is uniformly continuous if and only if $\omega(\delta) \to 0$ as $\delta \to 0$.

18. Let F be a closed subset of $(-\infty, +\infty)$, and let f be continuous relative to F. Show that there is a continuous function g on $(-\infty, +\infty)$ which equals f in F. If $|f(x)| \leq M$ for $x \in F$, show that g can be chosen so that $|g(x)| \leq M$ for $-\infty < x < +\infty$. (This is the Tietze extension theorem for the real line.)

19. Prove the following special case of the Baire category theorem: the intersection of a countable number of open dense sets in $\mathbf{R^1}$ is dense in $\mathbf{R^1}$.

20. Show that the irrational numbers form a set of type G_δ, but that the rational numbers do not. (For the second part, it is possible to argue by contradiction, using the first part and the result in Exercise 19.)

21. Construct a set in $\mathbf{R^1}$ that is neither of type G_δ nor of type F_σ. (Consider the union of the negative rationals and the positive irrationals, and use facts from Exercise 20.)

22. For an integer $k = 1, \ldots, n$ and a real number α, consider the hyperplane $H = \{\mathbf{x} = (x_1, \ldots, x_n) : x_k = \alpha\}$. Show that for every $\varepsilon > 0$, there is a collection $\{Q_j\}_{j=1}^{\infty}$ of cubes in $\mathbf{R^n}$ with edges parallel to the coordinate axes such that $H \subset \bigcup Q_j$ and $\sum v(Q_j) < \varepsilon$. (Using the terminology of Chapter 3, it follows that H has outermeasure 0 in $\mathbf{R^n}$.)

2

Functions of Bounded Variation and the Riemann–Stieltjes Integral

In the chapters ahead, we will study the Lebesgue integral. In this chapter, we introduce the Riemann–Stieltjes integral and, as a natural preliminary step, study functions of bounded variation. The justification for doing so is that Lebesgue integration is intimately connected with Riemann–Stieltjes integration, although this is not apparent from the definitions. We shall see in Theorem 5.43 that Lebesgue integrals can be represented as Riemann–Stieltjes integrals.

2.1 Functions of Bounded Variation

Let $f(x)$ be a real-valued function that is defined and finite for all x in a closed bounded interval $a \leq x \leq b$. Let

$$\Gamma = \{x_0, x_1, \ldots, x_m\}$$

be a *partition* of $[a, b]$; that is, Γ is a collection of points x_i, $i = 0, 1, \ldots, m$, satisfying $x_0 = a$, $x_m = b$, and $x_{i-1} < x_i$ for $i = 1, \ldots, m$. With each partition Γ, we associate the sum

$$S_\Gamma = S_\Gamma[f; a, b] = \sum_{i=1}^{m} |f(x_i) - f(x_{i-1})|.$$

The *variation of f over* $[a, b]$ is defined as

$$V = V[f; a, b] = \sup_\Gamma S_\Gamma,$$

where the supremum is taken over all partitions Γ of $[a, b]$. The variation $V[f; a, b]$ will sometimes also be denoted by $V[a, b]$ or $V(f)$. Since $0 \leq S_\Gamma < +\infty$, we have $0 \leq V \leq +\infty$. If $V < +\infty$, f is said to be of *bounded variation on* $[a, b]$; if $V = +\infty$, f is of *unbounded variation on* $[a, b]$.

We list several simple examples.

Example 1 Suppose f is monotone in $[a,b]$. Then, clearly, each S_Γ equals $|f(b)-f(a)|$, and therefore $V=|f(b)-f(a)|$.

Example 2 Suppose the graph of f can be split into a finite number of monotone arcs; that is, suppose $[a,b]=\bigcup_{i=1}^{k}[a_i,a_{i+1}]$ and f is monotone in each $[a_i,a_{i+1}]$. Then $V=\sum_{i=1}^{k}|f(a_{i+1})-f(a_i)|$. To see this, we use the result of Example 1 and the fact, to be proved in Theorem 2.2, that $V=V[a,b]=\sum_{i=1}^{k}V[a_i,a_{i+1}]$.

Example 3 Let f be defined by $f(x)=0$ when $x\neq 0$ and $f(0)=1$, and let $[a,b]$ be any interval containing 0 in its interior. Then S_Γ is either 2 or 0, depending on whether or not $x=0$ is a partitioning point of Γ. Thus, $V[a,b]=2$.

If $\Gamma=\{x_0,x_1,\ldots,x_m\}$ is a partition of $[a,b]$, let $|\Gamma|$, called the *norm of* Γ, be defined as the length of a longest subinterval of Γ:

$$|\Gamma|=\max_i\left(x_i-x_{i-1}\right).$$

If f is continuous on $[a,b]$ and $\{\Gamma_j\}$ is a sequence of partitions of $[a,b]$ with $|\Gamma_j|\to 0$, we shall see in Theorem 2.9 that $V=\lim_{j\to\infty}S_{\Gamma_j}$. Example 3 shows that this equality may fail for functions that are discontinuous even at a single point: if we take f and $[a,b]$ as in Example 3 and choose the Γ_j such that $x=0$ is never a partitioning point, then $\lim S_{\Gamma_j}=0$, while if we choose the Γ_j such that $x=0$ alternately is and is not a partitioning point, then $\lim S_{\Gamma_j}$ does not exist. See also Exercise 20.

Example 4 Let f be the *Dirichlet function*, defined by $f(x)=1$ for rational x and $f(x)=0$ for irrational x. Then, clearly, $V[a,b]=+\infty$ for any interval $[a,b]$.

Example 5 A function that is continuous on an interval is not necessarily of bounded variation on the interval. To see this, let $\{a_j\}$ and $\{d_j\}$, $j=1,2,\ldots$, be two monotone decreasing sequences in $(0,1]$ with $a_1=1$, $\lim_{j\to\infty}a_j=\lim_{j\to\infty}d_j=0$ and $\sum d_j=+\infty$. Construct a continuous f as follows. On each subinterval $[a_{j+1},a_j]$, the graph of f consists of the sides of the isosceles triangle with base $\left[a_{j+1},a_j\right]$ and height d_j. Thus, $f(a_j)=0$, and if m_j denotes the midpoint of $\left[a_{j+1},a_j\right]$, then $f(m_j)=d_j$. If we further define $f(0)=0$, then f is continuous on [0,1]. Taking Γ_k to be the partition defined by the points 0, $\left\{a_j\right\}_{j=1}^{k+1}$, and $\left\{m_j\right\}_{j=1}^{k}$, we see that $S_{\Gamma_k}=2\sum_{j=1}^{k}d_j$. Hence, $V[f;0,1]=+\infty$. See also Exercise 1.

We mention here that there exist functions that are continuous on an interval but that are not of bounded variation on any subinterval. See Exercise 26 of Chapter 3.

Example 6 A function f defined on $[a,b]$ is said to satisfy a *Lipschitz condition* on $[a,b]$, or to be a *Lipschitz function* on $[a,b]$, if there is a constant C such that

$$|f(x) - f(y)| \leq C|x - y| \text{ for all } x, y \in [a,b].$$

Such a function is clearly of bounded variation, with $V[f;a,b] \leq C(b-a)$. For example, if f has a continuous derivative on $[a,b]$, or even just a bounded derivative, then (by the mean-value theorem) f satisfies a Lipschitz condition on $[a,b]$.

For more examples of functions of bounded variation, see the exercises at the end of the chapter.

In the next two theorems, we summarize some of the simplest properties of functions of bounded variation. The proof of the first theorem is left as an exercise.

Theorem 2.1

(i) *If f is of bounded variation on $[a,b]$, then f is bounded on $[a,b]$.*

(ii) *Let f and g be of bounded variation on $[a,b]$. Then cf (for any real constant c), $f+g$, and fg are of bounded variation on $[a,b]$. Moreover, f/g is of bounded variation on $[a,b]$ if there exists an $\varepsilon > 0$ such that $|g(x)| \geq \varepsilon$ for x ε $[a,b]$.*

Before stating the second result, we note that if $\bar{\Gamma}$ is a *refinement* of Γ, that is, if $\bar{\Gamma}$ contains all the partitioning points of Γ plus some additional points, then $S_\Gamma \leq S_{\bar{\Gamma}}$. This follows from the triangle inequality and is most easily seen in the case when $\bar{\Gamma}$ consists of all the points of Γ plus one additional point. The case of general $\bar{\Gamma}$ can be reduced to this simple case by adding one point at a time to Γ.

Theorem 2.2

(i) *If $[a',b']$ is a subinterval of $[a,b]$, then $V[a',b'] \leq V[a,b]$; that is, variation increases with interval.*

(ii) *If $a < c < b$, then $V[a,b] = V[a,c] + V[c,b]$; that is, variation is additive on adjacent intervals.*

Proof. (i) This follows easily from (ii), as the reader can check. A simple direct proof based on adjoining the points a, b to generic partitions of $[a', b']$ can also be given.

(ii) Let $I = [a, b], I_1 = [a, c], I_2 = [c, b], V = V[a, b]$, $V_1 = V[a, c]$, and $V_2 = V[c, b]$. If Γ_1 and Γ_2 are any partitions of I_1 and I_2, respectively, then $\Gamma = \Gamma_1 \cup \Gamma_2$ is one of I, and $S_\Gamma[I] = S_{\Gamma_1}[I_1] + S_{\Gamma_2}[I_2]$. Thus, $S_{\Gamma_1}[I_1] + S_{\Gamma_2}[I_2] \leq V$. Therefore, taking the supremum over Γ_1 and Γ_2 separately, we obtain $V_1 + V_2 \leq V$.

To show the opposite inequality, let Γ be any partition of I, and let $\bar{\Gamma}$ be Γ with c adjoined. Then $S_\Gamma[I] \leq S_{\bar{\Gamma}}[I]$, and $\bar{\Gamma}$ splits into partitions Γ_1 of I_1 and Γ_2 of I_2. Thus, we have

$$S_\Gamma[I] \leq S_{\bar{\Gamma}}[I] = S_{\Gamma_1}[I_1] + S_{\Gamma_2}[I_2] \leq V_1 + V_2.$$

Therefore, $V \leq V_1 + V_2$, which completes the proof of (ii).

For any real number x, define

$$x^+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0, \end{cases} \qquad x^- = \begin{cases} 0 & \text{if } x > 0 \\ -x & \text{if } x \leq 0. \end{cases}$$

These are called the *positive* and *negative parts of* x, respectively, and satisfy the relations

$$x^+, x^- \geq 0; \quad |x| = x^+ + x^-; \quad x = x^+ - x^-. \tag{2.3}$$

Given a finite function f on $[a, b]$ and a partition $\Gamma = \{x_i\}_{i=0}^m$ of $[a, b]$, define

$$P_\Gamma = P_\Gamma[f; a, b] = \sum_{i=1}^m [f(x_i) - f(x_{i-1})]^+,$$

$$N_\Gamma = N_\Gamma[f; a, b] = \sum_{i=1}^m [f(x_i) - f(x_{i-1})]^-.$$

Thus, P_Γ is the sum of the positive terms of S_Γ, and $-N_\Gamma$ is the sum of the negative terms of S_Γ. In particular, by (2.3), $P_\Gamma \geq 0$, $N_\Gamma \geq 0$,

$$P_\Gamma + N_\Gamma = S_\Gamma, \tag{2.4}$$

$$P_\Gamma - N_\Gamma = f(b) - f(a). \tag{2.5}$$

The positive variation P and the *negative variation* N of f are defined by

$$P = P[f; a, b] = \sup_{\Gamma} P_{\Gamma},$$
$$N = N[f; a, b] = \sup_{\Gamma} N_{\Gamma}.$$

Thus, $0 \le P, N \le +\infty$.

Theorem 2.6 *If any one of P, N, or V is finite, then all three are finite. Moreover, we then have*

$$P + N = V, \quad P - N = f(b) - f(a),$$

or equivalently

$$P = \frac{1}{2}\left[V + f(b) - f(a)\right], \quad N = \frac{1}{2}\left[V - f(b) + f(a)\right].$$

Proof. By (2.4), $P_\Gamma + N_\Gamma \le V$, and therefore, since P_Γ and N_Γ are nonnegative, $P \le V$ and $N \le V$. In particular, P and N are finite if V is. By (2.4) again, $S_\Gamma \le P + N$ and therefore $V \le P + N$. If either P or N is finite, so is the other by (2.5), and therefore so is V. This gives the first part of the theorem.

Now choose a sequence of partitions Γ_k so that $P_{\Gamma_k} \to P$. Let us show that $N_{\Gamma_k} \to N$ and $P - [f(b) - f(a)] = N$. By (2.5), $N_{\Gamma_k} = P_{\Gamma_k} - [f(b) - f(a)] \to P - [f(b) - f(a)]$, and since $N_{\Gamma_k} \le N$, it follows that $P - [f(b) - f(a)] \le N$. If $P - [f(b) - f(a)] < N$, there is, by definition of N, a partition Γ with $N_\Gamma > P - [f(b) - f(a)]$. Then $P_\Gamma = N_\Gamma + [f(b) - f(a)] > P$, which is impossible. Hence, $P - [f(b) - f(a)] = N$ and $N_{\Gamma_k} \to N$. If N is finite, it follows that $P - N = f(b) - f(a)$. Letting $k \to \infty$ in the inequality $P_{\Gamma_k} + N_{\Gamma_k} \le V$ gives $P + N \le V$. Since $V \le P + N$ was shown earlier, we have $V = P + N$, and the theorem follows.

Corollary 2.7 (Jordan's Theorem) *A function f is of bounded variation on* $[a, b]$ *if and only if it can be written as the difference of two bounded increasing functions on* $[a, b]$.

Proof. Suppose $f = f_1 - f_2$, where f_1 and f_2 are bounded and increasing on $[a, b]$. Then f_1 and f_2 are of bounded variation on $[a, b]$, and therefore, by Theorem 2.1(ii), so is f.

Conversely, suppose f is of bounded variation on $[a, b]$. By Theorem 2.2(i), f is of bounded variation on every interval $[a, x]$, $a \le x \le b$. Let $P(x)$ and $N(x)$ denote the positive and negative variations of f on $[a, x]$, respectively.

By the analogue of Theorem 2.2(i) for P and N (see Exercise 3), it follows that $P(x)$ and $N(x)$ are bounded and increasing on $[a, b]$. Moreover, by Theorem 2.6 applied to $[a, x]$, $f(x) = [P(x) + f(a)] - N(x)$ when $a \leq x \leq b$. Since $P(x)$ is bounded and increasing, so is $P(x) + f(a)$, and the corollary follows.

Note that since the negative of an increasing function is decreasing, Corollary 2.7 may be rephrased to say that f is of bounded variation if and only if it is the sum of a bounded increasing function and a bounded decreasing function.

We remark here that there exist continuous functions of bounded variation that are not monotone in any subinterval. See Exercise 27 of Chapter 3.

In the next theorem, we consider a continuity property of functions of bounded variation. We recall from Chapter 1 that a discontinuity is said to be of the first kind if it is either a jump or a removable discontinuity.

Theorem 2.8 *Every function of bounded variation has at most a countable number of discontinuities, and they are all of the first kind.*

Here, if f is of bounded variation on $[a, b]$, we can clarify what it means to say that f has a discontinuity of the first kind at the endpoints a, b by extending the definition of f outside $[a, b]$ by setting $f(x) = f(a)$ if $x < a$ and $f(x) = f(b)$ if $x > b$ and then using the usual notion.

Proof. Let f be of bounded variation on $[a, b]$. Suppose first that f is bounded and increasing on $[a, b]$. Then the only discontinuities of f are of the first kind; in fact, they are all jump discontinuities. If D denotes the set of all discontinuities of f, then $D = \bigcup_{k=1}^{\infty}\{x : f(x+) - f(x-) \geq 1/k\}$. Since f is bounded, each set on the right is finite (or empty); therefore, D is countable. The general case follows from this by using Corollary 2.7. Note that removable discontinuities may arise by subtracting monotone functions; for example, consider the function f in Example 3 and the corresponding monotone functions $P(x)$ and $N(x)$. See also Exercise 25.

We now discuss a property of the variation of a continuous function. See also Exercise 20.

Theorem 2.9 *If f is continuous on $[a, b]$, then $V = \lim_{|\Gamma| \to 0} S_\Gamma$; that is, given M satisfying $M < V$, there exists $\delta > 0$ such that $S_\Gamma > M$ for any partition Γ of $[a, b]$ with $|\Gamma| < \delta$.*

Proof. We remind the reader of the discussion following Example 3. Given M with $M < V$, we must find $\delta > 0$ so that $S_\Gamma > M$ if $|\Gamma| < \delta$. Select $\mu > 0$ such

that $M+\mu < V$, and choose a fixed partition $\bar{\Gamma} = \{\bar{x}_j\}_{j=0}^{k}$ such that $S_{\bar{\Gamma}} > M+\mu$. Using the uniform continuity of f on $[a,b]$, pick $\eta > 0$ such that

(i) $|f(x) - f(x')| < \mu/[2(k+1)]$ if $|x - x'| < \eta$.

Now let Γ be any partition that satisfies

(ii) $|\Gamma| < \eta$,
(iii) $|\Gamma| < \min_j \left(\bar{x}_j - \bar{x}_{j-1}\right)$.

We claim that $S_\Gamma > M$, from which the theorem will follow by choosing δ to be the smaller of η and $\min_j(\bar{x}_j - \bar{x}_{j-1})$. Write $\Gamma = \{x_i\}_{i=0}^{m}$ and

$$S_\Gamma = \sum_{i=1}^{m} |f(x_i) - f(x_{i-1})| = \sum{}' + \sum{}'',$$

where $\sum''$ is extended over all i such that (x_{i-1}, x_i) contains some $\bar{x}_j$. By (iii), any (x_{i-1}, x_i) can contain at most one $\bar{x}_j$, and therefore the number of terms of $\sum''$ is at most $k+1$. Let $\Gamma \cup \bar{\Gamma}$ denote the partition formed by the union of the points of Γ and $\bar{\Gamma}$. Then $\Gamma \cup \bar{\Gamma}$ is a refinement of both Γ and $\bar{\Gamma}$. Moreover, $S_{\Gamma\cup\bar{\Gamma}} = \sum' + \sum'''$, where Σ''' is obtained from Σ'' by replacing each term by $|f(x_i) - f(\bar{x}_j)| + |f(\bar{x}_j) - f(x_{i-1})|$, $\bar{x}_j$ being the point of $\bar{\Gamma}$ in (x_{i-1}, x_i). By (i) and (ii), each of these two terms is less than $\mu/[2(k+1)]$, and therefore

$$\sum{}''' < 2(k+1)\frac{\mu}{2(k+1)} = \mu.$$

Hence,

$$\sum{}' = S_{\Gamma\cup\bar{\Gamma}} - \sum{}''' > S_{\Gamma\cup\bar{\Gamma}} - \mu,$$

so that $S_\Gamma > S_{\Gamma\cup\bar{\Gamma}} - \mu$. Since $\Gamma \cup \bar{\Gamma}$ is a refinement of $\bar{\Gamma}$, $S_{\Gamma\cup\bar{\Gamma}} \geq S_{\bar{\Gamma}}$. This gives $S_\Gamma > S_{\bar{\Gamma}} - \mu > M$ and completes the proof.

Corollary 2.10 *If f has a continuous derivative f' on $[a,b]$, then*

$$V = \int_a^b |f'|\,dx, \quad P = \int_a^b \{f'\}^+\,dx, \quad N = \int_a^b \{f'\}^-\,dx.$$

Proof. By the mean-value theorem,

$$S_\Gamma = \sum_{i=1}^{m} |f(x_i) - f(x_{i-1})| = \sum_{i=1}^{m} |f'(\xi_i)|(x_i - x_{i-1})$$

for appropriate $\xi_i \in (x_{i-1}, x_i), i = 1, \ldots, m$. Hence, by Theorem 2.9,

$$V = \lim_{|\Gamma|\to 0} S_\Gamma = \lim_{|\Gamma|\to 0} \sum_{i=1}^{m} |f'(\xi_i)|(x_i - x_{i-1}) = \int_a^b |f'(x)|\, dx,$$

by definition of the Riemann integral. Moreover, by Theorem 2.6,

$$\begin{aligned} P &= \frac{1}{2}[V + f(b) - f(a)] = \frac{1}{2}\left[\int_a^b |f'(x)|\, dx + \int_a^b f'(x)\, dx\right] \\ &= \frac{1}{2}\int_a^b \big[|f'(x)| + f'(x)\big] dx = \int_a^b [f'(x)]^+\, dx. \end{aligned}$$

The formula for N follows similarly from the fact that

$$N = \frac{1}{2}[V - f(b) + f(a)].$$

For an extension of Corollary 2.10, see Theorem 7.31. See also Exercise 22 of Chapter 7.

In passing, we note that there are notions of bounded variation for open or partly open intervals, as well as for infinite intervals. Suppose, for example, that (a, b) is a bounded open interval. Let $[a', b'] \subset (a, b)$, and define $V^\circ(a, b) = \lim V[a', b']$ as $a' \to a$ and $b' \to b$. If $V^\circ(a, b) < +\infty$, we say f is of *bounded variation on* (a, b). Similarly, if f is defined on $(-\infty, +\infty)$, let $V(-\infty, +\infty) = \lim V[a, b]$ as $a \to -\infty$ and $b \to +\infty$. Analogous definitions hold for $[a, b)$, $(a, +\infty)$, $[a, +\infty)$, etc. See Exercise 8.

We may also consider the notion of bounded variation for complex-valued f defined on an interval. The definition is the same as for the real-valued case, and we leave it to the reader to show that a complex-valued f is of bounded variation if and only if both its real and imaginary parts are as well.

2.2 Rectifiable Curves

As an application of the notion of bounded variation, we shall discuss its relation to rectifiable curves (initially, those in the plane). A *curve* C in the plane is two finite real-valued parametric equations

$$C : \begin{cases} x = \varphi(t) \\ y = \psi(t) \end{cases}, \quad a \le t \le b. \tag{2.11}$$

The *graph of* C is $\{(x, y) : x = \phi(t), y = \psi(t),\ a \le t \le b\}$. The graph may have self-intersections and is not necessarily continuous or bounded. We think of the curve itself as the mapping of $[a, b]$ onto the graph.

Let $\Gamma = \{a = t_0 < t_1 < \cdots < t_m = b\}$ be a partition of $[a, b]$, and consider the corresponding points $P_i = (\phi(t_i), \psi(t_i)), i = 0, 1, \ldots, m$, on the graph of C. Draw the polygonal (broken) line connecting P_0 to P_1, P_1 to $P_2, \ldots, P_{m-1}$ to P_m in order, and let

$$l(\Gamma) = \sum_{i=1}^{m} \left(\left[\phi\left(t_i\right) - \phi\left(t_{i-1}\right)\right]^2 + \left[\psi\left(t_i\right) - \psi\left(t_{i-1}\right)\right]^2 \right)^{1/2}$$

denote its length. The *length* L of C is defined by the equation

$$L = L(C) = \sup_{\Gamma} l(\Gamma). \tag{2.12}$$

Thus, $0 \le L \le +\infty$. If the graph of C is discontinuous, then as we move along the graph, the length of every missing segment will contribute to L. Moreover, the possibility that the graph may be traversed more than once, that is, that the mapping $t \to (\phi(t), \psi(t)), a \le t \le b$, may not be one-to-one, will add to L.

We say C is *rectifiable* if $L < +\infty$.

Theorem 2.13 *Let C be a curve defined by (2.11). Then C is rectifiable if and only if both* ϕ *and* ψ *are of bounded variation. Moreover,*

$$V(\phi), V(\psi) \le L \le V(\phi) + V(\psi).$$

Proof. We will use the simple inequalities

$$|a|, |b| \le \left(a^2 + b^2\right)^{1/2} \le |a| + |b|$$

for real a and b. Thus, if C is rectifiable and $\Gamma = \{t_i\}$ is any partition of $[a, b]$, the inequality

$$l(\Gamma) = \sum \left(\left[\phi(t_i) - \phi\left(t_{i-1}\right) \right]^2 + \left[\psi(t_i) - \psi\left(t_{i-1}\right) \right]^2 \right)^{1/2} \leq L$$

implies $\sum |\phi(t_i) - \phi\left(t_{i-1}\right)| \leq L$ and $\sum |\psi(t_i) - \psi\left(t_{i-1}\right)| \leq L$. Hence, $V(\phi)$, $V(\psi) \leq L$. On the other hand, for any C,

$$l(\Gamma) \leq \sum |\phi(t_i) - \phi\left(t_{i-1}\right)| + \sum |\psi(t_i) - \psi\left(t_{i-1}\right)| \leq V(\phi) + V(\psi).$$

Hence, $L \leq V(\phi) + V(\psi)$, which completes the proof.

It follows that if $\phi(t)$ is any bounded function that is not of bounded variation on $[a, b]$ (see Example 5 and Exercise 1), then the curve given by $x = y = \phi(t), a \leq t \leq b$, is not rectifiable, even though its graph lies in a finite segment of the line $y = x$. Thus, the length of the graph of a curve is not necessarily the same as the length of the curve.

In the special case that C is given by a function $y = f(x)$, Theorem 2.13 reduces to the simple statement that C is rectifiable if and only if f is of bounded variation.

Curves in $\mathbf{R^n}$ can be treated similarly, and we shall be brief. By a *curve* C in $\mathbf{R^n}$, we mean a system $x_1 = \phi_1(t), \ldots, x_n = \phi_n(t)$, for t in some $[a, b]$. We consider a partition $\Gamma = \{t_i\}_{i=0}^m$ of $[a, b]$ and the length $l(\Gamma)$ of the corresponding polygonal line:

$$l(\Gamma) = \sum_{i=1}^{m} P_{i-1}P_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \left[\phi_j(t_i) - \phi_j\left(t_{i-1}\right) \right]^2 \right)^{1/2}.$$

The quantity $L = \sup l(\Gamma)$ is called the *length* of C, and if $L < +\infty$, C is said to be *rectifiable*. As seen from the definition of $l(\Gamma)$, exactly as in the case $n = 2$, C is rectifiable if and only if each ϕ_j is of bounded variation.

2.3 The Riemann–Stieltjes Integral

Let f and ϕ be two functions that are defined and finite on a finite interval $[a, b]$. If $\Gamma = \{a = x_0 < x_1 < \cdots < x_m = b\}$ is a partition of $[a, b]$, we arbitrarily select intermediate points $\{\xi_i\}_{i=1}^m$ satisfying $x_{i-1} \leq \xi_i \leq x_i$ and write

$$R_\Gamma = \sum_{i=1}^{m} f(\xi_i) \left[\phi(x_i) - \phi\left(x_{i-1}\right) \right]. \tag{2.14}$$

R_Γ is called a *Riemann–Stieltjes sum* for Γ and of course depends on the points ξ_i, the functions f and ϕ, and the interval $[a, b]$, although we shall usually not display this dependence in our notation.

If

$$I = \lim_{|\Gamma| \to 0} R_\Gamma \tag{2.15}$$

exists and is finite, that is, if given $\varepsilon > 0$ there is a $\delta > 0$ such that $|I - R_\Gamma| < \varepsilon$ for any Γ satisfying $|\Gamma| < \delta$ and for any choice of intermediate points, then I is called *the Riemann–Stieltjes integral of f with respect to ϕ on* $[a, b]$ and denoted

$$I = \int_a^b f(x)\, d\phi(x) = \int_a^b f\, d\phi.$$

A necessary and sufficient condition for the existence of $\int_a^b f\, d\phi$ is the following Cauchy criterion: given $\varepsilon > 0$, there exists $\delta > 0$ such that $|R_\Gamma - R_{\Gamma'}| < \varepsilon$ if $|\Gamma|, |\Gamma'| < \delta$. See Exercise 11.

We list four preliminary remarks about this integral.

1. If $\phi(x) = x$, $\int_a^b f\, d\phi$ is clearly just the Riemann integral $\int_a^b f\, dx$. In this case, Theorem 5.54 in Chapter 5 gives a necessary and sufficient condition on f for the existence of the integral.
2. If f is continuous on $[a, b]$ and ϕ is continuously differentiable on $[a, b]$, then $\int_a^b f\, d\phi = \int_a^b f\phi'\, dx$. (See also Theorem 7.32.) In fact, by the mean-value theorem,

$$R_\Gamma = \sum f(\xi_i)\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right] = \sum f(\xi_i)\, \phi'(\eta_i)\left(x_i - x_{i-1}\right),$$

   with $x_{i-1} \le \xi_i, \eta_i \le x_i$. Using the uniform continuity of ϕ', we obtain $\lim_{|\Gamma| \to 0} R_\Gamma = \int_a^b f\phi'\, dx$.
3. Let $\phi(x)$ be a *step function*; that is, suppose there are points $a = \alpha_0 < \alpha_1 < \cdots < \alpha_m = b$ such that ϕ is constant on each interval (α_{i-1}, α_i). Let

$$\phi(\alpha_i+) = \lim_{x \to \alpha_i+} \phi(x), \quad i = 0, 1, \ldots, m-1,$$

   and

$$\phi(\alpha_i-) = \lim_{x \to \alpha_i-} \phi(x), \quad i = 1, \ldots, m,$$

denote the limits from the right and left at α_i, and let $d_i = \phi(\alpha_i+) - \phi(\alpha_i-), i = 1, \ldots, m-1$, $d_0 = \phi(\alpha_0+) - \phi(\alpha_0)$, and $d_m = \phi(\alpha_m) - \phi(\alpha_m-)$ denote the jumps of ϕ. Then, for continuous f,

$$\int_a^b f\,d\phi = \sum_{i=0}^{m} f(\alpha_i)d_i.$$

The existence of the integral can be verified directly or by appealing to Theorem 2.24. See also Exercise 29.

For example, if $\phi(x)$ is chosen to be the *Heaviside function* $H(x)$ defined by $H(x) = 0$ when $x < 0$ and $H(x) = 1$ when $x \geq 0$, then

$$\int_{-1}^{1} f\,dH = f(0)$$

if f is continuous at $x = 0$.

4. In definition (2.15), no condition other than finiteness is imposed on f or ϕ, but we shall see later that the most important applications occur when ϕ is monotone or, more generally, of bounded variation. We note now that if $\int_a^b f d\phi$ exists, then f and ϕ have no common points of discontinuity. To prove this, suppose that both f and ϕ are discontinuous at $\bar{x}, a < \bar{x} < b$. Suppose first that the discontinuity of ϕ is not removable. Then there is a fixed $\eta > 0$ such that, for every $\varepsilon > 0$, there exist points $\bar{x}_1$ and $\bar{x}_2$ with $\bar{x} - \frac{\varepsilon}{2} < \bar{x}_1 < \bar{x} < \bar{x}_2 < \bar{x} + \frac{\varepsilon}{2}$ and $|\phi(\bar{x}_2) - \phi(\bar{x}_1)| > \eta$. Let $\Gamma = \{x_i\}$ be a partition of $[a, b]$ with $|\Gamma| < \varepsilon$ such that $x_{i_0-1} = \bar{x}_1$ and $x_{i_0} = \bar{x}_2$ for some i_0. Choose a point $\xi_i \in [x_{i-1}, x_i]$ for $i \neq i_0$ and two different points ξ_{i_0} and ξ'_{i_0} in $[x_{i_0-1}, x_{i_0}]$. Let R_Γ be the Riemann–Stieltjes sum using ξ_i in each $[x_{i-1}, x_i]$, and let R'_Γ be the sum using ξ_i in $[x_{i-1}, x_i]$ for $i \neq i_0$ and ξ'_{i_0} in $[x_{i_0-1}, x_{i_0}]$. Then, clearly,

$$\begin{aligned} |R_\Gamma - R'_\Gamma| &= |f(\xi_{i_0}) - f(\xi'_{i_0})|\,|\phi(x_{i_0}) - \phi(x_{i_0-1})| \\ &> \eta |f(\xi_{i_0}) - f(\xi'_{i_0})|. \end{aligned}$$

Since f is discontinuous at $\bar{x}$, we can choose ξ_{i_0} and ξ'_{i_0} subject to the restrictions earlier and such that $|f(\xi_{i_0}) - f(\xi'_{i_0})| > \mu$ for some $\mu > 0$ independent of ε. It follows that $|R_\Gamma - R'_\Gamma|$ exceeds a positive constant independent of ε, contradicting the assumption that $R_\Gamma - R'_\Gamma \to 0$ as $|\Gamma|, |\Gamma'| \to 0$.

If the discontinuity of ϕ at $\bar{x}$ is removable, a similar argument can be given. The main difference is that we consider Γ with $\bar{x}$ as a partitioning

point x_{i_0} and argue for either $[x_{i_0-1}, \bar{x}]$ or $[\bar{x}, x_{i_0+1}]$, depending on the nature of the discontinuity of f at $\bar{x}$. The arguments in the case where $\bar{x}$ is either a or b are similar.

In the theorem that follows, we list some simple properties of the Riemann–Stieltjes integral. The proofs are left as an exercise.

Theorem 2.16

(i) *If* $\int_a^b f\,d\phi$ *exists, then so do* $\int_a^b cf\,d\phi$ *and* $\int_a^b f\,d(c\phi)$ *for any constant* c*, and*

$$\int_a^b cf\,d\phi = \int_a^b f\,d(c\phi) = c\int_a^b f\,d\phi.$$

(ii) *If* $\int_a^b f_1\,d\phi$ *and* $\int_a^b f_2\,d\phi$ *both exist, so does* $\int_a^b (f_1+f_2)\,d\phi$*, and*

$$\int_a^b (f_1+f_2)\,d\phi = \int_a^b f_1\,d\phi + \int_a^b f_2\,d\phi.$$

(iii) *If* $\int_a^b f\,d\phi_1$ *and* $\int_a^b f\,d\phi_2$ *exist, so does* $\int_a^b f\,d(\phi_1+\phi_2)$*, and*

$$\int_a^b f\,d(\phi_1+\phi_2) = \int_a^b f\,d\phi_1 + \int_a^b f\,d\phi_2.$$

The additivity of the integral with respect to intervals is given by the following result. See also Exercise 14.

Theorem 2.17 *If* $\int_a^b f\,d\phi$ *exists and* $a < c < b$*, then* $\int_a^c f\,d\phi$ *and* $\int_c^b f\,d\phi$ *both exist and*

$$\int_a^b f\,d\phi = \int_a^c f\,d\phi + \int_c^b f\,d\phi.$$

Proof. In the proof, $R_\Gamma[a,b]$ will denote a Riemann–Stieltjes sum corresponding to a partition Γ of $[a,b]$. To show that $\int_a^c f\,d\phi$ exists, it is enough to show

that given $\varepsilon > 0$, there exists $\delta > 0$ so that if Γ_1 and Γ_2 are partitions of $[a, c]$ with $|\Gamma_1|, |\Gamma_2| < \delta$, then

$$|R_{\Gamma_1}[a, c] - R_{\Gamma_2}[a, c]| < \varepsilon. \tag{2.18}$$

Since $\int_a^b f\, d\phi$ exists, there is a $\delta > 0$ so that for any partitions Γ_1' and Γ_2' of $[a, b]$ with $|\Gamma_1'|, |\Gamma_2'| < \delta$, we have

$$|R_{\Gamma_1'}[a, b] - R_{\Gamma_2'}[a, b]| < \varepsilon. \tag{2.19}$$

Let Γ_1 and Γ_2 be partitions of $[a, c]$ with given sets of intermediate points. Complete Γ_1 and Γ_2 to partitions Γ_1' and Γ_2' of $[a, b]$ by adjoining the same points of $[c, b]$; that is, let Γ' be a partition of $[c, b]$ and let $\Gamma_1' = \Gamma_1 \cup \Gamma'$, $\Gamma_2' = \Gamma_2 \cup \Gamma'$. Select a set of intermediate points in $[c, b]$ for Γ', and let the intermediate points of Γ_1' and Γ_2' consist of these together with the sets for Γ_1 and Γ_2, respectively. Then

$$\begin{aligned} R_{\Gamma_1'}[a, b] &= R_{\Gamma_1}[a, c] + R_{\Gamma'}[c, b] \\ R_{\Gamma_2'}[a, b] &= R_{\Gamma_2}[a, c] + R_{\Gamma'}[c, b]. \end{aligned} \tag{2.20}$$

If we now assume that $|\Gamma_1|, |\Gamma_2| < \delta$ and choose Γ' so that $|\Gamma'| < \delta$, then $|\Gamma_1'|, |\Gamma_2'| < \delta$, and (2.18) follows from (2.19) by subtracting the equations in (2.20).

The proof of the existence of $\int_c^b f\, d\phi$ is similar. The fact that

$$\int_a^b f\, d\phi = \int_a^c f\, d\phi + \int_c^b f\, d\phi$$

follows from (2.20). This completes the proof. See also Exercise 19.

The next result is the very useful formula for *integration by parts*.

Theorem 2.21 *If $\int_a^b f\, d\phi$ exists, then so does $\int_a^b \phi\, df$, and*

$$\int_a^b f\, d\phi = [f(b)\phi(b) - f(a)\phi(a)] - \int_a^b \phi\, df.$$

Proof. Let $\Gamma = \{a = x_0 < x_1 < \cdots < x_m = b\}$ and $x_{i-1} \le \xi_i \le x_i$. Then

$$\begin{aligned} R_\Gamma &= \sum_{i=1}^{m} f(\xi_i)\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right] = \sum_{i=1}^{m} f(\xi_i)\,\phi(x_i) - \sum_{i=1}^{m} f(\xi_i)\,\phi\left(x_{i-1}\right) \\ &= \sum_{i=1}^{m} f(\xi_i)\,\phi(x_i) - \sum_{i=0}^{m-1} f\left(\xi_{i+1}\right)\phi(x_i) \\ &= -\sum_{i=1}^{m-1} \phi(x_i)\left[f\left(\xi_{i+1}\right) - f(\xi_i)\right] + f(\xi_m)\,\phi(b) - f(\xi_1)\,\phi(a) \end{aligned}$$

since $x_m = b$ and $x_0 = a$. If we subtract and add $\phi(a)[f(\xi_1) - f(a)] + \phi(b)[f(b) - f(\xi_m)]$ on the right side of the last equality and cancel like terms, we obtain $R_\Gamma = -T_\Gamma + [f(b)\phi(b) - f(a)\phi(a)]$, where

$$T_\Gamma = \sum_{i=1}^{m-1} \phi(x_i)\left[f\left(\xi_{i+1}\right) - f(\xi_i)\right] + \phi(a)\left[f(\xi_1) - f(a)\right] + \phi(b)\left[f(b) - f(\xi_m)\right].$$

Since the ξ_i straddle the x_i (successive ξ_i's may be equal), T_Γ is a Riemann–Stieltjes sum for $\int_a^b \phi\, df$. Observing that the roles of ϕ and f can be interchanged, and taking the limit as $|\Gamma| \to 0$, we see that $\int_a^b f\, d\phi$ exists if and only if $\int_a^b \phi\, df$ exists and that $\int_a^b f\, d\phi = [f(b)\phi(b) - f(a)\phi(a)] - \int_a^b \phi\, df$. This completes the proof.

Now let f be bounded and ϕ be monotone increasing on $[a, b]$. If $\Gamma = \{x_i\}_{i=0}^m$, let

$$\begin{aligned} m_i &= \inf_{x_{i-1} \le x \le x_i} f(x), \quad M_i = \sup_{x_{i-1} \le x \le x_i} f(x), \\ L_\Gamma &= \sum_{i=1}^{m} m_i\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right], \\ U_\Gamma &= \sum_{i=1}^{m} M_i\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right]. \end{aligned}$$

Since $-\infty < m_i \le M_i < +\infty$ and $\phi(x_i) - \phi(x_{i-1}) \ge 0$, we see that

$$L_\Gamma \le R_\Gamma \le U_\Gamma. \tag{2.22}$$

L_Γ and U_Γ are called the *lower* and *upper Riemann–Stieltjes sums for* Γ, respectively. The behavior of L_Γ and U_Γ is somewhat more predictable than that of R_Γ, as we now show.

Lemma 2.23 *Let f be bounded and ϕ be increasing on $[a, b]$.*

(i) *If Γ' is a refinement of Γ, then $L_{\Gamma'} \geq L_\Gamma$ and $U_{\Gamma'} \leq U_\Gamma$.*
(ii) *If Γ_1 and Γ_2 are any two partitions, then $L_{\Gamma_1} \leq U_{\Gamma_2}$.*

Proof. To see (i) for upper sums, suppose that Γ' has only one point x' not in Γ. If x' lies between x_{i-1} and x_i of Γ, then $\sup_{[x_{i-1},x']} f(x), \sup_{[x',x_i]} f(x) \leq M_i$, so that

$$\sup_{[x_{i-1},x']} f(x)\left[\phi(x') - \phi\left(x_{i-1}\right)\right] + \sup_{[x',x_i]} f(x)\left[\phi\left(x_i\right) - \phi\left(x'\right)\right] \leq M_i\left[\phi\left(x_i\right) - \phi\left(x_{i-1}\right)\right].$$

Hence, $U_{\Gamma'} \leq U_\Gamma$. Since Γ' can be obtained by adding one point at a time to Γ, an extension of this reasoning proves (i) for upper sums. The argument for lower sums is similar.

To show (ii), note that $\Gamma_1 \cup \Gamma_2$ is a refinement of both Γ_1 and Γ_2. Hence, by part (i) and (2.22), we obtain $L_{\Gamma_1} \leq L_{\Gamma_1 \cup \Gamma_2} \leq U_{\Gamma_1 \cup \Gamma_2} \leq U_{\Gamma_2}$, which completes the proof.

We now come to an important result that gives sufficient conditions for the existence of $\int_a^b f\, d\phi$. See also Exercise 23.

Theorem 2.24 *If f is continuous on $[a, b]$ and ϕ is of bounded variation on $[a, b]$, then $\int_a^b f\, d\phi$ exists. Moreover,*

$$\left|\int_a^b f\, d\phi\right| \leq \left(\sup_{[a,b]} |f|\right) V[\phi; a, b].$$

Proof. To prove the existence, we may suppose by Corollary 2.7 and Theorem 2.16(iii) that ϕ is monotone increasing. Then, by (2.22), $L_\Gamma \leq R_\Gamma \leq U_\Gamma$, and it is enough to show that $\lim_{|\Gamma| \to 0} L_\Gamma$ and $\lim_{|\Gamma| \to 0} U_\Gamma$ exist and are equal. This is clear if ϕ is constant on $[a, b]$. If ϕ is not constant, let $\Gamma = \{x_i\}$ and note that

given $\varepsilon > 0$, the uniform continuity of f implies there exists $\delta > 0$ such that if $|\Gamma| < \delta$, then $M_i - m_i < \varepsilon/[\phi(b) - \phi(a)]$. Hence, if $|\Gamma| < \delta$,

$$0 \leq U_\Gamma - L_\Gamma = \sum (M_i - m_i) \left[\phi(x_i) - \phi\left(x_{i-1}\right)\right] < \varepsilon. \tag{2.25}$$

Therefore,

$$\lim_{|\Gamma| \to 0} (U_\Gamma - L_\Gamma) = 0, \tag{2.26}$$

and it is enough to show that $\lim_{|\Gamma|\to 0} U_\Gamma$ exists. This is immediate since otherwise there would exist an $\varepsilon > 0$ and two sequences of partitions, $\{\Gamma'_k\}$ and $\{\Gamma''_k\}$, with norms tending to zero such that $U_{\Gamma'_k} - U_{\Gamma''_k} > \varepsilon$. In view of (2.26), we would then have, for k large enough, $L_{\Gamma'_k} - U_{\Gamma''_k} > \varepsilon/2 > 0$, contradicting the fact that $L_{\Gamma'} \leq U_{\Gamma''}$ for any Γ' and Γ'' (Lemma 2.23).

To complete the proof, note that the inequality $|\int_a^b f\, d\phi| \leq (\sup_{(a,b)} |f|)$ $V[\phi; a, b]$ follows from a similar one for R_Γ by taking the limit.

Combining Theorems 2.21 and 2.24, we see that $\int_a^b f\, d\phi$ exists if either f or ϕ is continuous and the other is of bounded variation.

Theorem 2.27 (Mean-Value Theorem) *If f is continuous on $[a, b]$ and ϕ is bounded and increasing on $[a, b]$, there exists $\xi, \in [a, b]$ such that*

$$\int_a^b f\, d\phi = f(\xi)[\phi(b) - \phi(a)].$$

Proof. Since ϕ is increasing, we have

$$\left(\min_{[a,b]} f\right)[\phi(b) - \phi(a)] \leq R_\Gamma \leq \left(\max_{[a,b]} f\right)[\phi(b) - \phi(a)]$$

for any R_Γ. Since $\int_a^b f\, d\phi$ exists (see Theorem 2.24), it must satisfy

$$\left(\min_{[a,b]} f\right)[\phi(b) - \phi(a)] \leq \int_a^b f\, d\phi \leq \left(\max_{[a,b]} f\right)[\phi(b) - \phi(a)].$$

The result now follows immediately from the intermediate value theorem for continuous functions.

In passing, we note that Riemann–Stieltjes integrals can also be defined, in the improper sense, on open or partly open bounded intervals and on infinite intervals. If f and ϕ are defined on (a,b), for example, let $a < a' < b' < b$ and define

$$\int_a^b f\,d\phi = \lim_{\substack{a'\to a\\ b'\to b}} \int_{a'}^{b'} f\,d\phi,$$

provided the limit exists in the sense that it is independent of how $a' \to a$ and $b' \to b$. Similarly, let

$$\int_{-\infty}^{+\infty} f\,d\phi = \lim_{\substack{a\to -\infty\\ b\to +\infty}} \int_a^b f\,d\phi$$

if the limit exists. Analogous definitions can be given for $[a,b)$, $(a,+\infty)$, $[a,+\infty)$, etc. See Exercise 24.

2.4 Further Results about Riemann–Stieltjes Integrals

We will discuss a variant of the definition of $\int_a^b f\,d\phi$ in the case where f is bounded and ϕ is increasing. Note that it then follows from part (ii) of Lemma 2.23 that $-\infty < \sup_\Gamma L_\Gamma \le \inf_\Gamma U_\Gamma < +\infty$. It is natural to ask if the existence of $\int_a^b f\,d\phi$ in this case is equivalent to the statement that

$$\sup_\Gamma L_\Gamma = \inf_\Gamma U_\Gamma, \tag{2.28}$$

which we know to be an equivalent definition in the case of Riemann integrals (see (1.20)). Unfortunately, the answer in general is *no*, as the following example shows. Let $[a,b] = [-1,1]$ and

$$f(x) = \begin{cases} 0 & \text{if } -1 \le x < 0 \\ 1 & \text{if } \quad 0 \le x \le 1, \end{cases}$$

$$\phi(x) = \begin{cases} 0 & \text{if } -1 \le x \le 0 \\ 1 & \text{if } \quad 0 < x \le 1. \end{cases}$$

Since f and ϕ have a common discontinuity, $\int_{-1}^1 f\,d\phi$ does not exist. In fact, if Γ straddles 0, that is, if $x_{i_0-1} < 0 < x_{i_0}$ for some i_0, then $R_\Gamma = f(\xi_{i_0})$ for $x_{i_0-1} \le \xi_{i_0} \le x_{i_0}$. Hence, R_Γ may be 0 or 1 and thus cannot have a limit. On the other

hand, it is easy to check that $U_\Gamma = 1$ for any Γ and that $L_\Gamma = 0$ if Γ straddles 0 and $L_\Gamma = 1$ otherwise. Hence, neither $\lim_{|\Gamma|\to 0} R_\Gamma$ nor $\lim_{|\Gamma|\to 0} L_\Gamma$ exists, but $\sup_\Gamma L_\Gamma = \inf_\Gamma U_\Gamma = 1$.

In the following two theorems, we explore relations between (2.15) and (2.28).

Theorem 2.29 *Let f be bounded and ϕ be monotone increasing on $[a, b]$. If $\int_a^b f\, d\phi$ exists, then $\lim_{|\Gamma|\to 0} L_\Gamma$ and $\lim_{|\Gamma|\to 0} U_\Gamma$ exist, and*

$$\lim_{|\Gamma|\to 0} L_\Gamma = \lim_{|\Gamma|\to 0} U_\Gamma = \sup_\Gamma L_\Gamma = \inf_\Gamma U_\Gamma = \int_a^b f\, d\phi.$$

Proof. We may assume that ϕ is not constant on $[a, b]$ since the result is obvious otherwise. Let $I = \int_a^b f\, d\phi$. By hypothesis, given $\varepsilon > 0$, there is a $\delta > 0$ such that $|I - R_\Gamma| < \varepsilon$ for any R_Γ with $|\Gamma| < \delta$. Given $\Gamma = \{x_i\}_{i=0}^m$ with $|\Gamma| < \delta$, choose ξ_i and η_i in $[x_{i-1}, x_i]$, $i = 1, \ldots, m$, such that

$$0 \le M_i - f(\xi_i) < \frac{\varepsilon}{\phi(b) - \phi(a)} \quad \text{and} \quad 0 \le f(\eta_i) - m_i < \frac{\varepsilon}{\phi(b) - \phi(a)}.$$

Let $R'_\Gamma = \sum f(\xi_i)\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right]$ and $R''_\Gamma = \sum f(\eta_i)\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right]$. Then $|I - R'_\Gamma| < \varepsilon$, $|I - R''_\Gamma| < \varepsilon$,

$$0 \le U_\Gamma - R'_\Gamma \le \sum \frac{\varepsilon}{\phi(b) - \phi(a)}\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right] = \varepsilon,$$

and

$$0 \le R''_\Gamma - L_\Gamma \le \sum \frac{\varepsilon}{\phi(b) - \phi(a)}\left[\phi(x_i) - \phi\left(x_{i-1}\right)\right] = \varepsilon.$$

Combining inequalities, we obtain

$$|U_\Gamma - I| \le |U_\Gamma - R'_\Gamma| + |R'_\Gamma - I| < \varepsilon + \varepsilon = 2\varepsilon$$

and

$$|L_\Gamma - I| \le |L_\Gamma - R''_\Gamma| + |R''_\Gamma - I| < \varepsilon + \varepsilon = 2\varepsilon.$$

Hence, $\lim_{|\Gamma|\to 0} U_\Gamma = \lim_{|\Gamma|\to 0} L_\Gamma = I$. Since, by Lemma 2.23, $L_\Gamma \le \sup_\Gamma L_\Gamma \le \inf_\Gamma U_\Gamma \le U_\Gamma$, the theorem follows.

Theorem 2.30 *Let f be bounded, and let ϕ be monotone increasing and continuous on $[a,b]$. Then $\lim_{|\Gamma|\to 0} L_\Gamma$ and $\lim_{|\Gamma|\to 0} U_\Gamma$ exist, and*

$$\lim_{|\Gamma|\to 0} L_\Gamma = \sup_\Gamma L_\Gamma, \quad \lim_{|\Gamma|\to 0} U_\Gamma = \inf_\Gamma U_\Gamma.$$

In particular, if in addition $\sup_\Gamma L_\Gamma = \inf_\Gamma U_\Gamma$, then $\int_a^b f\,d\phi$ exists, and

$$\sup_\Gamma L_\Gamma = \inf_\Gamma U_\Gamma = \int_a^b f\,d\phi.$$

Proof. The proof is similar to that of Theorem 2.9. It is enough to show that $\lim_{|\Gamma|\to 0} L_\Gamma = \sup_\Gamma L_\Gamma$ and $\lim_{|\Gamma|\to 0} U_\Gamma = \inf_\Gamma U_\Gamma$ since the last assertion of the theorem will then follow by (2.22). We will give the argument for the upper sums; the one for the lower sums is similar. Let $\inf_\Gamma U_\Gamma = U$. Given $\varepsilon > 0$, we must find $\delta > 0$ such that $U_\Gamma < U + \varepsilon$ if $|\Gamma| < \delta$. Choose $\bar\Gamma = \{\bar x_j\}_{j=0}^k$ such that $U_{\bar\Gamma} < U + \frac{\varepsilon}{2}$, and let $M = \sup_{[a,b]} |f|$. By the uniform continuity of ϕ, there exists $\eta > 0$ such that

$$|\phi(x) - \phi(x')| < \frac{\varepsilon}{4(k+1)M}$$

if $|x - x'| < \eta$.

Now let $\Gamma = \{x_i\}_{i=0}^m$ be any partition for which $|\Gamma| < \eta$ and $|\Gamma| < \min_j(\bar x_j - \bar x_{j-1})$. It is enough to show that $U_\Gamma < U + \varepsilon$. Write

$$U_\Gamma = \sum_{i=1}^m M_i\left[\phi(x_i) - \phi(x_{i-1})\right] = \sum{}' + \sum{}'',$$

where $\sum'$ is as in the proof of Theorem 2.9. Then $U_{\Gamma\cup\bar\Gamma} = \sum' + \sum'''$, where $\sum'''$ is obtained from $\sum''$ by replacing each of the terms $M_i[\phi(x_i) - \phi(x_{i-1})]$ by

$$\sup_{[x_{i-1},\bar x_j]} f(x)\left[\phi(\bar x_j) - \phi(x_{i-1})\right] + \sup_{[\bar x_j, x_i]} f(x)\left[\phi(x_i) - \phi(\bar x_j)\right], \tag{2.31}$$

$\bar{x}_j$ being the point of $\bar{\Gamma}$ in (x_{i-1}, x_i). Hence, $U_\Gamma - U_{\Gamma\cup\bar{\Gamma}} = \sum'' - \sum'''$. At least one of $\sup_{[x_{i-1},\bar{x}_j]} f$ and $\sup_{[\bar{x}_j,x_i]} f$ equals M_i. If it is the first, the difference between $M_i[\phi(x_i) - \phi(x_{i-1})]$ and (2.31) is easily seen to be

$$(M_i - \sup_{[\bar{x}_j,x_i]} f)[\phi(x_i) - \phi(\bar{x}_j)].$$

If it is the second, the difference is

$$(M_i - \sup_{[x_{i-1},\bar{x}_j]} f)[\phi(\bar{x}_j) - \phi(x_{i-1})].$$

In either case, the difference is at most $2M\varepsilon/[4(k+1)M] = \varepsilon/[2(k+1)]$ in absolute value. Hence, $U_\Gamma - U_{\Gamma\cup\bar{\Gamma}} \leq (k+1)\varepsilon/[2(k+1)] = \frac{1}{2}\varepsilon$. Moreover, $U_{\Gamma\cup\bar{\Gamma}} \leq U_{\bar{\Gamma}} < U + \frac{1}{2}\varepsilon$. Therefore, $U_\Gamma < U + \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = U + \varepsilon$, and the theorem follows.

Exercises

1. Let $f(x) = x \sin(1/x)$ for $0 < x \leq 1$ and $f(0) = 0$. Show that f is bounded and continuous on $[0,1]$, but that $V[f;0,1] = +\infty$.
2. Prove Theorem 2.1.
3. If $[a',b']$ is a subinterval of $[a,b]$, show that $P[a',b'] \leq P[a,b]$ and $N[a',b'] \leq N[a,b]$.
4. Let $\{f_k\}$ be a sequence of functions of bounded variation on $[a,b]$. If $V[f_k;a,b] \leq M < +\infty$ for all k and if $f_k \to f$ pointwise on $[a,b]$, show that f is of bounded variation and that $V[f;a,b] \leq M$. Give an example of a convergent sequence of functions of bounded variation whose limit is not of bounded variation.
5. Suppose f is finite on $[a,b]$ and of bounded variation on every interval $[a+\varepsilon,b]$, $\varepsilon > 0$, with $V[f;a+\varepsilon,b] \leq M < +\infty$. Show that $V[f;a,b] < +\infty$. Is $V[f;a,b] \leq M$? If not, what additional assumption will make it so?
6. Let $f(x) = x^2 \sin(1/x)$ for $0 < x \leq 1$ and $f(0) = 0$. Show that $V[f;0,1] < +\infty$. (Examine the graph of f, or use Exercise 5 and Corollary 2.10.)
7. Suppose f is of bounded variation on $[a,b]$. If f is continuous at a point $\bar{x}$, show that $V(x), P(x)$, and $N(x)$ are also continuous at $\bar{x}$. In particular, if f is continuous on $[a,b]$, then so are $V(x), P(x)$, and $N(x)$. (If $\Gamma = \{x_i\}$, note

that $V[x_{i-1}, x_i] - |f(x_{i-1}) - f(x_i)| \leq V[a,b] - S_\Gamma$. Recall that S_Γ increases when Γ is refined.)

8. The main results about functions of bounded variation on a closed bounded interval remain true for open or partly open intervals and for infinite intervals. Prove, for example, that if f is of bounded variation on $(-\infty, +\infty)$, then f is the difference of two increasing bounded functions.

9. Let C be a curve with parametric equations $x = \phi(t)$ and $y = \psi(t)$, $a \leq t \leq b$.

(a) If ϕ and ψ are of bounded variation and continuous, show that $L = \lim_{|\Gamma| \to 0} l(\Gamma)$.

(b) If ϕ and ψ are continuously differentiable, show that $L = \int_a^b ([\phi'(t)]^2 + [\psi'(t)]^2)^{1/2} dt$.

10. If $\lambda_1 < \lambda_2 < \cdots < \lambda_m$ is a finite sequence and $-\infty < s < +\infty$, write $\sum_k a_k e^{-s\lambda_k}$ as a Riemann–Stieltjes integral. (Take $f(x) = e^{-sx}$, ϕ to be an appropriate step function and $[a,b]$ to contain all the λ_k in its interior.)

11. Show that $\int_a^b f\, d\phi$ exists if and only if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|R_\Gamma - R_{\Gamma'}| < \varepsilon$ if $|\Gamma|, |\Gamma'| < \delta$.

12. Prove that the conclusion of Theorem 2.30 is valid if the assumption that ϕ is continuous is replaced by the assumption that f and ϕ have no common discontinuities. (Instead of the uniform continuity of ϕ, use the fact that either f or ϕ is continuous at each point $\bar{x}_j$ of $\bar{\Gamma}$.)

13. Prove Theorem 2.16.

14. Give an example that shows that for $a < c < b$, $\int_a^c f\, d\phi$ and $\int_c^b f\, d\phi$ may both exist but $\int_a^b f\, d\phi$ may not. Compare Theorem 2.17. (Take $[a,b] = [-1,1]$, $c = 0$, and f and ϕ as in the example following (2.28).)

15. Suppose f is continuous and ϕ is of bounded variation on $[a,b]$. Show that the function $\psi(x) = \int_a^x f\, d\phi$ is of bounded variation on $[a,b]$. If g is continuous on $[a,b]$, show that $\int_a^b g\, d\psi = \int_a^b gf\, d\phi$.

16. Suppose that ϕ is of bounded variation on $[a,b]$ and that f is bounded and continuous except for a finite number of jump discontinuities in $[a,b]$. If ϕ is continuous at each discontinuity of f, show that $\int_a^b f\, d\phi$ exists.

17. If ϕ is of bounded variation on $(-\infty, +\infty)$, f is continuous on $(-\infty, +\infty)$, and $\lim_{|x| \to +\infty} f(x) = 0$, show that $\int_{-\infty}^{+\infty} f\, d\phi$ exists.

18. Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$ be a power series. Show that if $\sum |a_k| < +\infty$, then $f(z)$ is of bounded variation on every radius of the circle $|z| = 1$. (If, e.g., the radius is $0 \leq x \leq 1$ and the a_k are real, then $f(x) = \sum a_k^+ x^k - \sum a_k^- x^k$.)

19. Let f and ϕ be finite functions on $[a,b]$. If $[a', b'] \subset [a,b]$ and $\int_a^b f\, d\phi$ exists, show that $\int_{a'}^{b'} f\, d\phi$ exists.

20. Let f be finite on $[a,b]$. If $\lim_{|\Gamma|\to 0} S_\Gamma[f;a,b]$ exists, show that it equals $V[f;a,b]$.

21. If $V[\phi;a,b] = +\infty$, show that there is a point $x_0 \in [a,b]$ such that either $V[\phi;I] = +\infty$ for every subinterval I of $[a,b]$ having x_0 as left-hand endpoint or $V[\phi;I] = +\infty$ for every subinterval I of $[a,b]$ having x_0 as right-hand endpoint. (Note that for any x_0, a', b' such that $a \le a' < x_0 < b' \le b$, $V[\phi;a',b'] < +\infty$ if both $V[\phi;a',x_0], V[\phi;x_0,b'] < +\infty$.)

22. If $V[\phi;a,b] = +\infty$, show that there exist $x_0 \in [a,b]$ and a monotone sequence $\{x_k\}_1^\infty$ in $[a,b]$ such that $x_k \to x_0$ and $\sum_{k=1}^\infty |\phi(x_{k+1}) - \phi(x_k)| = +\infty$. (Use the results in Exercises 21 and 5.)

23. If $V[\phi;a,b] = +\infty$, show that there is a continuous f on $[a,b]$ such that $\int_a^b f\,d\phi$ does not exist. (Use the result in Exercise 22 together with the following fact: if $\{\alpha_k\}$ is a sequence of positive numbers with $\sum \alpha_k = +\infty$, then there is a sequence $\{\varepsilon_k\}$ of positive numbers with $\varepsilon_k \to 0$ and $\sum \varepsilon_k \alpha_k = +\infty$.)

24. Let f be continuous and ϕ be of bounded variation on $[a,b]$, and recall that the Riemann–Stieltjes integral $\int_a^b f\,d\phi$ then exists by Theorem 2.24. Show that $\lim_{\varepsilon\to 0+} \int_a^{a+\varepsilon} f\,d\phi = 0$ if and only if either $f(a) = 0$ or ϕ is continuous at a. Deduce that the formula $\int_a^b f\,d\phi = \lim_{\varepsilon\to 0+} \int_{a+\varepsilon}^b f\,d\phi$ may not hold.

25. Construct a bounded nondecreasing function on $(-\infty,\infty)$ which is continuous at every irrational number and discontinuous at every rational number. (Let $\{r_k\}_{k=1}^\infty$ be an enumeration of the rational numbers and consider the function $f(x) = \sum_{k:r_k\le x} 2^{-k}$.)

26. Let $f(x) = \sin x$. Sketch the graphs of its variations $P(x)$, $N(x)$, $V(x)$ for $0 \le x \le 2\pi$, and use Corollary 2.10 to find explicit formulas for these variations on $[0,2\pi]$.

27. If f is an even function on $[-1,1]$, verify the formulas $V[f;-1,1] = 2P[f;-1,1]$, $V[f;0,1] = P[f;-1,1]$ and $P[f;0,1] = N[f;-1,0]$.

28. Let f be Lipschitz continuous on an interval $[a,b]$. Show that $V[f;a,b]$ is strictly less than the length of the graph of f over $[a,b]$.

29. Use Theorem 2.17 to verify the formula for $\int_a^b f\,d\phi$ given in remark 3 in Section 2.3.

30. Let f and ϕ be real-valued functions on $[a,b]$.

(a) If $\int_a^b f\,d\phi$ exists and ϕ is not constant on any subinterval of $[a,b]$, show that f is bounded on $[a,b]$.

(b) If $\int_a^b f\,d\phi$ exists and ϕ is increasing, show that $L_\Gamma \le \int_a^b f\,d\phi \le U_\Gamma$ for every partition Γ of $[a,b]$, where L_Γ and U_Γ are the corresponding lower and upper sums.

31. Show that for any real-valued function f on $[a, b]$, the Riemann–Stieltjes integral $\int_a^b df$ exists and equals $f(b) - f(a)$. If f' exists and is Riemann integrable on $[a, b]$, show that $\int_a^b f'(x)\,dx = \int_a^b df$, and consequently $\int_a^b f'(x)\,dx = f(b) - f(a)$.

32. Let f be a function of bounded variation on an interval $[a, b]$. Show that f is Riemann integrable on $[a, b]$. (This follows from Theorems 5.54 and 2.8, but it can be derived solely from the results in this chapter.)

3

Lebesgue Measure and Outer Measure

In this chapter, we will define and study the Lebesgue measure of sets in $\mathbf{R^n}$. This will be the foundation for the theory of integration to be developed later. We will base the presentation on the notion of the outer measure of a set.

3.1 Lebesgue Outer Measure and the Cantor Set

We consider closed n-dimensional intervals $I = \{\mathbf{x}\colon a_j \le x_j \le b_j, j = 1, \ldots, n\}$ and their volumes $v(I) = \prod_{j=1}^{n}(b_j - a_j)$. (See p. 7 in Section 1.3.) To define the outer measure of an arbitrary subset E of $\mathbf{R^n}$, cover E by a *countable* collection S of intervals I_k, and let

$$\sigma(S) = \sum_{I_k \in S} v(I_k).$$

The *Lebesgue outer measure (or exterior measure) of* E, denoted $|E|_e$, is defined by

$$|E|_e = \inf\ \sigma(S), \tag{3.1}$$

where the infimum is taken over all such covers S of E. Thus, $0 \le |E|_e \le +\infty$.

Theorem 3.2 *For an interval I, $|I|_e = v(I)$.*

Proof. Since I covers itself, we have $|I|_e \le v(I)$. To show the opposite inequality, suppose that $S = \{I_k\}_{k=1}^{\infty}$ is a cover of I. Given $\varepsilon > 0$, let I_k^* be an interval containing I_k in its interior such that $v(I_k^*) \le (1+\varepsilon)v(I_k)$. Then $I \subset \bigcup_k (I_k^*)^\circ$, and since I is closed and bounded, the Heine–Borel Theorem 1.9 implies there is an integer N such that $I \subset \bigcup_{k=1}^{N} I_k^*$. Therefore, $v(I) \le \sum_{k=1}^{N} v(I_k^*)$ by a basic property of the volume of intervals (see p. 8 in Section 1.3). Hence, $v(I) \le (1+\varepsilon)\sum_{k=1}^{N} v(I_k) \le (1+\varepsilon)\sigma(S)$. Since ε can be chosen arbitrarily small, it follows that $v(I) \le \sigma(S)$ and, therefore, that $v(I) \le |I|_e$. This completes the proof.

Note that the boundary of any interval has outer measure zero. Compare Exercise 22 of Chapter 1.

The following two theorems state simple but basic properties of outer measure.

Theorem 3.3 *If $E_1 \subset E_2$, then $|E_1|_e \leq |E_2|_e$.*

The proof follows immediately from the fact that any cover of E_2 is also a cover of E_1.

Theorem 3.4 *If $E = \bigcup E_k$ is a countable union of sets, then $|E|_e \leq \sum |E_k|_e$.*

Proof. We may assume that $|E_k|_e < +\infty$ for each $k = 1, 2, \ldots$, since otherwise the conclusion is obvious. Fix $\varepsilon > 0$. Given k, choose intervals $I_j^{(k)}$ such that $E_k \subset \bigcup_j I_j^{(k)}$ and $\sum_j v(I_j^{(k)}) < |E_k|_e + \varepsilon 2^{-k}$. Since $E \subset \bigcup_{j,k} I_j^{(k)}$, we have $|E|_e \leq \sum_{j,k} v(I_j^{(k)}) = \sum_k \sum_j v(I_j^{(k)})$. Therefore,

$$|E|_e \leq \sum_k (|E_k|_e + \varepsilon 2^{-k}) = \sum_k |E_k|_e + \varepsilon,$$

and the result follows by letting $\varepsilon \to 0$.

We see in particular that any subset of a set with outer measure zero has outer measure zero and that the countable union of sets with outer measure zero has outer measure zero. Since any set consisting of a single point clearly has outer measure zero, it follows that *any countable subset* of $\mathbf{R^n}$ *has outer measure zero*. For example, the set consisting of all points each of whose coordinates is rational has outer measure zero, even though it is dense in $\mathbf{R^n}$.

There are sets with outer measure zero that are not countable. As an illustration, we will construct a subset of the real line with outer measure zero that is perfect, and therefore uncountable, by Theorem 1.9. Variants of the construction and analogues for $\mathbf{R^n}, n > 1$, are given in the exercises.

Consider the closed interval $[0, 1]$. The first stage of the construction is to subdivide $[0, 1]$ into thirds and remove the interior of the middle third; that is, remove the open interval $\left(\frac{1}{3}, \frac{2}{3}\right)$. Each successive step of the construction is essentially the same. Thus, at the second stage, we subdivide each of the remaining two intervals $\left[0, \frac{1}{3}\right]$ and $\left[\frac{2}{3}, 1\right]$ into thirds and remove the interiors, $\left(\frac{1}{9}, \frac{2}{9}\right)$ and $\left(\frac{7}{9}, \frac{8}{9}\right)$, of their middle thirds. We continue the construction for each of the remaining intervals. The sets removed in the first three successive stages are indicated in the following illustration by darkened intervals:

The subset of $[0,1]$ that remains after infinitely many such operations is called the *Cantor set* C: thus, if C_k denotes the union of the intervals left at the kth stage, then

$$C = \bigcap_{k=1}^{\infty} C_k. \tag{3.5}$$

Since each C_k is closed, it follows from Theorem 1.7 that C is closed. Note that C_k consists of 2^k closed disjoint intervals, each of length 3^{-k}, and that C contains the endpoints of all these intervals. Any point of C belongs to an interval in C_k for every k and is therefore a limit point of the endpoints of the intervals. This proves that C is perfect. Finally, since C is covered by the intervals in any C_k, we have $|C|_e \leq 2^k 3^{-k}$ for each k. Therefore, $|C|_e = 0$.

We now introduce a function associated with the Cantor set that will be useful later. If $D_k = [0,1] - C_k$, then D_k consists of the $2^k - 1$ open intervals I_j^k (ordered from left to right as j proceeds from $j = 1$ to $j = 2^k - 1$) removed in the first k stages of construction of the Cantor set. Let f_k be the continuous function on $[0,1]$ which satisfies $f_k(0) = 0$, $f_k(1) = 1$, $f_k(x) = j2^{-k}$ on I_j^k, $j = 1, \ldots, 2^k - 1$, and which is linear on each interval of C_k. The graphs of f_1 and f_2 are shown in the following illustration:

By construction, each f_k is monotone increasing, $f_{k+1} = f_k$ on $I_j^k, j = 1, \ldots, 2^k - 1$, and $|f_k - f_{k+1}| < 2^{-k}$. Hence, $\sum(f_k - f_{k+1})$ converges uniformly on $[0,1]$, and therefore $\{f_k\}$ converges uniformly on $[0,1]$. Let $f = \lim_{k\to\infty} f_k$. Then $f(0) = 0, f(1) = 1, f$ is monotone increasing and continuous on $[0,1]$, and f is constant on every interval removed in constructing C. This f is called the *Cantor–Lebesgue function*. Its graph is sometimes called the *Devil's staircase*.

The next two theorems give useful relations between the outer measure of a set and the outer measures of open sets and G_δ sets (see p. 6 in Section 1.3) that contain it.

Theorem 3.6 *Let $E \subset \mathbf{R^n}$. Then given $\varepsilon > 0$, there exists an open set G such that $E \subset G$ and $|G|_e \le |E|_e + \varepsilon$. Hence,*

$$|E|_e = \inf |G|_e, \tag{3.7}$$

where the infimum is taken over all open sets G containing E.

Proof. Given $\varepsilon > 0$, choose intervals I_k with $E \subset \bigcup_{k=1}^\infty I_k$ and $\sum_{k=1}^\infty v(I_k) \le |E|_e + \frac{1}{2}\varepsilon$. Let I_k^* be an interval containing I_k in its interior $(I_k^*)^\circ$ such that $v(I_k^*) \le v(I_k) + \varepsilon 2^{-k-1}$. If $G = \bigcup (I_k^*)^\circ$, then G is open and contains E. Furthermore,

$$|G|_e \le \sum_{k=1}^\infty v(I_k^*) \le \sum_{k=1}^\infty v(I_k) + \varepsilon \sum_{k=1}^\infty 2^{-k-1} \le |E|_e + \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = |E|_e + \varepsilon.$$

This completes the proof.

Theorem 3.8 *If $E \subset \mathbf{R^n}$, there exists a set H of type G_δ such that $E \subset H$ and $|E|_e = |H|_e$.*

Proof. By Theorem 3.6, there is for every positive integer k an open set $G_k \supset E$ such that $|G_k|_e \le |E|_e + 1/k$. If $H = \bigcap_{k=1}^\infty G_k$, then H is of type G_δ and $H \supset E$. Moreover, for every k, $|E|_e \le |H|_e \le |G_k|_e \le |E|_e + 1/k$. Thus, $|E|_e = |H|_e$. Note that each $|G_k|_e < \infty$ if $|E|_e < \infty$.

The significance of Theorem 3.8 is that the most general set in $\mathbf{R^n}$ can be included in a set of relatively simple type, namely, G_δ, with the same outer measure.

In defining the notion of outer measure, we used intervals I with edges parallel to the coordinate axes. The question arises whether the outer measure of a set depends on the position of the (orthogonal) coordinate axes.

The answer is *no*, and to prove this, we will simultaneously consider the usual coordinate system in $\mathbf{R^n}$ and a fixed rotation of this system. Notions pertaining to the rotated system will be denoted by primes. Thus, I' denotes an interval with edges parallel to the rotated coordinate axes, and $|E|'_e$ denotes the outer measure of a subset E relative to these rotated intervals:

$$|E|'_e = \inf \sum v(I'_k), \tag{3.9}$$

where the infimum is taken over all coverings of E by rotated intervals I'_k. The volume of an interval is clearly unchanged by rotation. (See p. 8 in Section 1.3.)

Theorem 3.10 $|E|'_e = |E|_e$ *for every* $E \subset \mathbf{R^n}$.

Proof. We first claim that given I' and $\varepsilon > 0$, there exist $\{I_l\}$ such that $I' \subset \bigcup I_l$, and $\sum v(I_l) \leq v(I') + \varepsilon$. To see this, let I'_1 be an interval containing I' in its interior such that $v(I'_1) \leq v(I') + \varepsilon$. By Theorem 1.11, the interior of I'_1 can be written as a union of nonoverlapping intervals I_l. Hence, $I' \subset \bigcup I_l$. Moreover, since the I_l are nonoverlapping and $\bigcup_{l=1}^N I_l \subset I'_1$ for every positive integer N, we have $\sum_{l=1}^N v(I_l) \leq v(I'_1)$ by a property of volume listed on p. 8 in Section 1.3. Therefore, $\sum_{l=1}^\infty v(I_l) \leq v(I'_1) \leq v(I') + \varepsilon$, which proves the claim. A parallel result is that given I and $\varepsilon > 0$, there exist $\{I'_l\}$ such that $I \subset \bigcup I'_l$ and $\sum v(I'_l) \leq v(I) + \varepsilon$.

Let E be any subset of $\mathbf{R^n}$. Given ε, choose $\{I_k : k = 1, 2, \ldots\}$ such that $E \subset \bigcup I_k$ and $\sum v(I_k) \leq |E|_e + \frac{1}{2}\varepsilon$. For each k, choose $\{I'_{k,l}\}$ such that $I_k \subset \bigcup_l I'_{k,l}$ and $\sum_l v(I'_{k,l}) \leq v(I_k) + \varepsilon 2^{-k-1}$. Thus,

$$\sum_{k,l} v(I'_{k,l}) \leq \sum_k v(I_k) + \frac{1}{2}\varepsilon \leq |E|_e + \varepsilon.$$

Since $E \subset \bigcup_{k,l} I'_{k,l}$, we obtain $|E|'_e \leq |E|_e + \varepsilon$. Hence, $|E|'_e \leq |E|_e$, and by symmetry, $|E|'_e = |E|_e$.

For related results, see Exercise 22.

3.2 Lebesgue Measurable Sets

A subset E of $\mathbf{R^n}$ is said to be *Lebesgue measurable*, or simply *measurable*, if given $\varepsilon > 0$, there exists an open set G such that

$$E \subset G \text{ and } |G - E|_e < \varepsilon.$$

If E is measurable, its outer measure is called its *Lebesgue measure*, or simply its *measure*, and denoted $|E|$:

$$|E| = |E|_e, \text{ for measurable } E. \tag{3.11}$$

The condition that E be measurable should not be confused with the conclusion of Theorem 3.6, which states that there exists an open G containing E such that $|G|_e \leq |E|_e + \varepsilon$. In general, since $G = E \bigcup (G - E)$ when $E \subset G$, we only have $|G|_e \leq |E|_e + |G - E|_e$, and we cannot conclude from $|G|_e \leq |E|_e + \varepsilon$ that $|G - E|_e < \varepsilon$.

We now list two simple examples of measurable sets. A nonmeasurable set will be constructed in Theorem 3.38.

Example 1 Every open set is measurable.
This is immediate from the definition.

Example 2 Every set of outer measure zero is measurable.
Suppose that $|E|_e = 0$. Then given $\varepsilon > 0$, there is by Theorem 3.6 an open G containing E with $|G| < |E|_e + \varepsilon = \varepsilon$. Hence,

$$|G - E|_e \leq |G| < \varepsilon.$$

Theorem 3.12 *The union $E = \bigcup E_k$ of a countable number of measurable sets is measurable, and*

$$|E| \leq \sum |E_k|.$$

Proof. Let $\varepsilon > 0$. For each $k = 1, 2, \ldots$, choose an open set G_k such that $E_k \subset G_k$ and $|G_k - E_k|_e < \varepsilon 2^{-k}$. Then $G = \bigcup G_k$ is open and $E \subset G$. Moreover, since $G - E \subset \bigcup (G_k - E_k)$, we have

$$|G - E|_e \leq \left| \bigcup (G_k - E_k) \right|_e \leq \sum |G_k - E_k|_e < \varepsilon.$$

This proves that E is measurable. The fact that $|E| \leq \sum |E_k|$ follows from Theorem 3.4.

Corollary 3.13 *An interval I is measurable, and $|I| = v(I)$.*

Proof. I is the union of its interior and its boundary. Since its boundary has measure zero, the fact that I is measurable follows from Theorem 3.12 and the results of Examples 1 and 2. Theorem 3.2 shows that $|I| = v(I)$.

Theorem 3.14 *Every closed set F is measurable.*

In order to prove this, we will use Theorem 1.11 and the next two lemmas.

Lemma 3.15 *If $\{I_k\}_{k=1}^N$ is a finite collection of nonoverlapping intervals, then $\bigcup I_k$ is measurable and $|\bigcup I_k| = \sum |I_k|$.*

Proof. The fact that $\bigcup I_k$ is measurable follows from Corollary 3.13. The rest of the lemma is a minor extension of Theorem 3.2, and its proof is left as an exercise. The reader should note the important role played by the Heine–Borel theorem in the proof.

We recall from Chapter 1 that the distance between two sets E_1 and E_2 is defined as $d(E_1, E_2) = \inf\{|\mathbf{x}_1 - \mathbf{x}_2| : \mathbf{x}_1 \in E_1, \mathbf{x}_2 \in E_2\}$.

Lemma 3.16 *If $d(E_1, E_2) > 0$, then $|E_1 \cup E_2|_e = |E_1|_e + |E_2|_e$.*

Proof. By Theorem 3.4, $|E_1 \cup E_2|_e \leq |E_1|_e + |E_2|_e$. To prove the opposite inequality, suppose $\varepsilon > 0$, and choose intervals $\{I_k\}$ such that $E_1 \cup E_2 \subset \bigcup I_k$ and $\sum |I_k| \leq |E_1 \cup E_2|_e + \varepsilon$. We may assume that the diameter of each I_k is less than $d(E_1, E_2)$. (Otherwise, we divide each I_k into a finite number of nonoverlapping subintervals with this property and apply Lemma 3.15.) Hence, $\{I_k\}$ splits into two subsequences $\{I'_k\}$ and $\{I''_k\}$, the first of which covers E_1 and the second, E_2. Clearly,

$$|E_1|_e + |E_2|_e \leq \sum |I'_k| + \sum |I''_k| = \sum |I_k| \leq |E_1 \cup E_2|_e + \varepsilon.$$

Therefore, $|E_1|_e + |E_2|_e \leq |E_1 \cup E_2|_e$, which completes the proof.

A special case of this result will be used in the proof of Theorem 3.14. If E_1 and E_2 are compact and disjoint, then $d(E_1, E_2) > 0$ by Exercise 1(l) of Chapter 1; therefore, $|E_1 \cup E_2|_e = |E_1|_e + |E_2|_e$ if E_1 and E_2 are compact and disjoint.

Proof of Theorem 3.14. Suppose first that F is compact. Given $\varepsilon > 0$, choose an open set G such that $F \subset G$ and $|G| < |F|_e + \varepsilon$. Since $G - F$ is open, Theorem 1.11

implies there are nonoverlapping closed intervals $I_k, k = 1, 2, \ldots$, such that $G - F = \bigcup I_k$. Therefore, $|G - F|_e \leq \sum |I_k|$, and it suffices to show that $\sum |I_k| \leq \varepsilon$. We have $G = F \cup (\bigcup I_k) \supset F \cup (\bigcup_{k=1}^N I_k)$ for every finite number $\{I_k\}_{k=1}^N$ of the I_k. Therefore,

$$|G| \geq \left| F \cup \left(\bigcup_{k=1}^N I_k \right) \right|_e = |F|_e + \left| \left(\bigcup_{k=1}^N I_k \right) \right|_e$$

by Lemma 3.16, F and $\bigcup_{k=1}^N I_k$ being disjoint and compact. Since $|\bigcup_{k=1}^N I_k|_e = \sum_{k=1}^N |I_k|$ by Lemma 3.15, we obtain $\sum_{k=1}^N |I_k| \leq |G| - |F|_e < \varepsilon$ for every N, so that $\sum |I_k| \leq \varepsilon$, as desired. This proves the result in the case when F is compact.

To complete the proof, let F be any closed subset of $\mathbf{R^n}$ and write $F = \bigcup F_k$, where $F_k = F \cap \{\mathbf{x} : |\mathbf{x}| \leq k\}$, $k = 1, 2, \ldots$. Each F_k is compact and, therefore, measurable. Hence, F is measurable by Theorem 3.12.

Theorem 3.17 *The complement of a measurable set is measurable.*

Proof. Let E be measurable. For each positive integer k, choose an open set G_k such that $E \subset G_k$ and $|G_k - E|_e < 1/k$. Since $\mathrm{C}G_k$ is closed, it is measurable by Theorem 3.14. Let $H = \bigcup_k \mathrm{C}G_k$. Then H is measurable and $H \subset \mathrm{C}E$. Write $\mathrm{C}E = H \cup Z$, where $Z = \mathrm{C}E - H$. Then $Z \subset \mathrm{C}E - \mathrm{C}G_k = G_k - E$, and therefore $|Z|_e < 1/k$ for every k. Hence, $|Z|_e = 0$ and, in particular, Z is measurable. Thus, $\mathrm{C}E$ is measurable since it is the union of two measurable sets.

The following two theorems are corollaries of Theorems 3.12 and 3.17.

Theorem 3.18 *The intersection $E = \bigcap_k E_k$ of a countable number of measurable sets is measurable.*

Proof. Since E_k is measurable, $\mathrm{C}E_k$ is measurable by Theorem 3.17. However, $\mathrm{C}E = \mathrm{C}(\bigcap_k E_k) = \bigcup_k \mathrm{C}E_k$, and hence $\mathrm{C}E$ is measurable. Therefore, by another application of Theorem 3.17, E is measurable.

Theorem 3.19 *If E_1 and E_2 are measurable, then $E_1 - E_2$ is measurable.*

Proof. Since $E_1 - E_2 = E_1 \cap \mathrm{C}E_2$, the result follows from Theorems 3.17 and 3.18.

As a consequence of Theorems 3.12, 3.17, 3.18, and 3.19 it follows that the class of measurable subsets of $\mathbf{R^n}$ is closed under the set-theoretic operations

of taking complements, countable unions, and countable intersections. Such a class of sets is called a σ-algebra; that is, a nonempty collection Σ of subsets E of some universal set U is called a *σ-algebra of sets* if it satisfies the following two conditions:

(i) $CE \in \Sigma$ if $E \in \Sigma$.

(ii) $\bigcup_k E_k \in \Sigma$ if $E_k \in \Sigma, k = 1, 2, \ldots$.

Note that it follows from the relation $C(\bigcap_k E_k) = \bigcup_k CE_k$ that $\bigcap_k E_k \in \Sigma$ if Σ is a σ-algebra and $E_k \in \Sigma, k = 1, 2, \ldots$. Moreover, it is easy to see that if Σ is a σ-algebra, then the universal set U and the empty set $\emptyset$ belong to Σ.

The following result is just a reformulation of Theorems 3.12 and 3.17.

Theorem 3.20 *The collection of measurable subsets of* $\mathbf{R^n}$ *is a σ-algebra.*

Note, for example, that if $\{E_k\}_{k=1}^{\infty}$ are measurable, then $\limsup E_k$ and $\liminf E_k$ are measurable since

$$\limsup E_k = \bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} E_k \quad \text{and} \quad \liminf E_k = \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} E_k.$$

If $\mathscr{C}_1$ and $\mathscr{C}_2$ are two collections of sets, we say that $\mathscr{C}_1$ is contained in $\mathscr{C}_2$ if every set in $\mathscr{C}_1$ is also in $\mathscr{C}_2$. If $\mathscr{F}$ is a family of σ-algebras Σ, we define $\bigcap_{\Sigma \in \mathscr{F}} \Sigma$ to be the collection of all sets E that belong to every Σ in $\mathscr{F}$. It is easy to check that $\bigcap_{\Sigma \in \mathscr{F}} \Sigma$ is itself a σ-algebra and is contained in every Σ in $\mathscr{F}$.

Given a collection $\mathscr{C}$ of subsets of $\mathbf{R^n}$, consider the family $\mathscr{F}$ of all σ-algebras that contain $\mathscr{C}$, and let $\mathscr{E} = \bigcap_{\Sigma \in \mathscr{F}} \Sigma$. Then $\mathscr{E}$ is the *smallest σ-algebra containing* $\mathscr{C}$; that is, $\mathscr{E}$ is a σ-algebra containing $\mathscr{C}$, and if Σ is any other σ-algebra containing $\mathscr{C}$, then Σ contains $\mathscr{E}$.

The smallest σ-algebra of subsets of $\mathbf{R^n}$ containing all the open subsets of $\mathbf{R^n}$ is called the *Borel σ-algebra* $\mathscr{B}$ of $\mathbf{R^n}$, and the sets in $\mathscr{B}$ are called *Borel subsets of* $\mathbf{R^n}$. Sets of type $F_\sigma, G_\delta, F_{\sigma\delta}, G_{\delta\sigma}$ (see p. 6 in Section 1.3), etc., are Borel sets.

Theorem 3.21 *Every Borel set is measurable.*

Proof. Let $\mathscr{M}$ be the collection of measurable subsets of $\mathbf{R^n}$. By Theorem 3.20, $\mathscr{M}$ is a σ-algebra. Since every open set belongs to $\mathscr{M}$, and $\mathscr{B}$ is the smallest σ-algebra containing the open sets, $\mathscr{B}$ is contained in $\mathscr{M}$.

The converse of Theorem 3.21 is false: see Exercise 31.

3.3 Two Properties of Lebesgue Measure

The next two theorems give properties of Lebesgue measure that are of fundamental importance. To prove them, we first need the following characterization of measurability in terms of closed sets.

Lemma 3.22 *A set E in $\mathbf{R^n}$ is measurable if and only if given $\varepsilon > 0$, there exists a closed set $F \subset E$ such that $|E - F|_e < \varepsilon$.*

Proof. E is measurable if and only if $\mathrm{C}E$ is measurable, that is, if and only if given $\varepsilon > 0$, there exists an open G such that $\mathrm{C}E \subset G$ and $|G - \mathrm{C}E|_e < \varepsilon$. Such G exists if and only if the set $F = \mathrm{C}G$ is closed, $F \subset E$, and $|E - F|_e < \varepsilon$ (since $G - \mathrm{C}E = E - F$).

Theorem 3.23 *If $\{E_k\}$ is a countable collection of disjoint measurable sets, then*

$$|\bigcup_k E_k| = \sum_k |E_k|.$$

Proof. First, suppose each E_k is bounded. Given $\varepsilon > 0$ and $k = 1, 2, \ldots$, use Lemma 3.22 to choose a closed $F_k \subset E_k$ with $|E_k - F_k| < \varepsilon 2^{-k}$. Then $|E_k| \leq |F_k| + \varepsilon 2^{-k}$ by Theorem 3.12. Since the E_k are bounded and disjoint, the F_k are compact and disjoint. Therefore, by Lemma 3.16, $|\bigcup_{k=1}^m F_k| = \sum_{k=1}^m |F_k|$ for every finite number $\{F_k\}_{k=1}^m$ of the F_k. The fact that $\bigcup_{k=1}^m F_k \subset \bigcup_k E_k$ then implies that $\sum_{k=1}^m |F_k| \leq |\bigcup_k E_k|$. Hence,

$$|\bigcup_k E_k| \geq \sum_k |F_k| \geq \sum_k (|E_k| - \varepsilon 2^{-k}) = \sum_k |E_k| - \varepsilon,$$

so that $|\bigcup_k E_k| \geq \sum_k |E_k|$. Since the opposite inequality is always true, the theorem follows in this case.

For the general case, let $I_j, j = 1, 2, \ldots$, be a sequence of intervals increasing to $\mathbf{R^n}$, and define $S_1 = I_1$ and $S_j = I_j - I_{j-1}$ for $j \geq 2$. Then the sets $E_{k,j} = E_k \cap S_j, k, j = 1, 2, \ldots$, are bounded, disjoint, and measurable; $E_k = \bigcup_j E_{k,j}$ and $\bigcup_k E_k = \bigcup_{k,j} E_{k,j}$. Therefore, by the case already established, we have

$$|\bigcup_k E_k| = |\bigcup_{k,j} E_{k,j}| = \sum_{k,j} |E_{k,j}| = \sum_k \left(\sum_j |E_{k,j}| \right) = \sum_k |E_k|.$$

Corollary 3.24 *If $\{I_k\}$ is a sequence of nonoverlapping intervals, then $|\bigcup I_k| = \sum |I_k|$.*

Proof. It is clear that $|\bigcup I_k| \leq \sum |I_k|$. On the other hand, the $(I_k)^\circ$ being disjoint, we have $|\bigcup I_k| \geq |\bigcup (I_k)^\circ| = \sum |(I_k)^\circ| = \sum |I_k|$. Thus, $|\bigcup I_k| = \sum |I_k|$. An alternate proof not using Theorem 3.23 can be obtained from Lemma 3.15.

Corollary 3.25 *Suppose E_1 and E_2 are measurable, $E_2 \subset E_1$, and $|E_2| < +\infty$. Then $|E_1 - E_2| = |E_1| - |E_2|$.*

Proof. Since $E_1 = E_2 \cup (E_1 - E_2)$, Theorem 3.23 gives $|E_1| = |E_2| + |E_1 - E_2|$. The corollary now follows from the assumption that $|E_2| < +\infty$.

The second basic property of Lebesgue measure concerns its behavior for a monotone sequence of sets.

Theorem 3.26 *Let $\{E_k\}_{k=1}^\infty$ be a sequence of measurable sets.*

(i) *If $E_k \nearrow E$, then $\lim_{k\to\infty} |E_k| = |E|$.*
(ii) *If $E_k \searrow E$ and $|E_k| < +\infty$ for some k, then $\lim_{k\to\infty} |E_k| = |E|$.*

Proof. (i) We may assume that $|E_k| < +\infty$ for all k; otherwise, both $\lim_{k\to\infty} |E_k|$ and $|E|$ are infinite. Write

$$E = E_1 \cup (E_2 - E_1) \cup \cdots \cup (E_k - E_{k-1}) \cup \cdots .$$

Since the terms in this union are measurable and disjoint, we have by Theorem 3.23 that

$$|E| = |E_1| + |E_2 - E_1| + \cdots + |E_k - E_{k-1}| + \cdots .$$

By Corollary 3.25,

$$|E| = |E_1| + (|E_2| - |E_1|) + \cdots + (|E_k| - |E_{k-1}|) + \cdots = \lim_{k\to\infty} |E_k|.$$

(ii) We may clearly assume that $|E_1| < +\infty$. Write

$$E_1 = E \cup (E_1 - E_2) \cup \cdots \cup (E_k - E_{k+1}) \cup \cdots .$$

Since the terms on the right are disjoint measurable sets, and since each E_k has finite measure, we have

$$\begin{aligned}|E_1| &= |E| + (|E_1| - |E_2|) + \cdots + (|E_k| - |E_{k+1}|) + \cdots \\ &= |E| + |E_1| - \lim_{k\to\infty} |E_k|.\end{aligned}$$

Therefore, $|E| = \lim_{k\to\infty} |E_k|$, which completes the proof.

The restriction in (ii) that $|E_k| < +\infty$ for some k is necessary, as the following example shows. Let E_k be the complement of the ball with center 0 and radius k. Then $|E_k| = +\infty$ for all k and $E_k \searrow \emptyset$, the empty set. Therefore, $\lim_{k\to\infty} |E_k| = +\infty$, while $|\emptyset| = 0$.

Although we are interested almost exclusively in measurable sets, proving the measurability of a given set is occasionally difficult in practice, and it may be desirable to apply theorems about outer measure. A particularly useful result is the following modification of part (i) of Theorem 3.26. The corresponding modification of part (ii) of Theorem 3.26 does not generally hold; see Exercise 21.

Theorem 3.27 *If $E_k \nearrow E$, then $\lim_{k\to\infty} |E_k|_e = |E|_e$.*

Proof. For each k, let H_k be a measurable set (e.g., a set of type G_δ) such that $E_k \subset H_k$ and $|H_k| = |E_k|_e$. For $m = 1, 2, \ldots$, let $V_m = \bigcap_{k=m}^{\infty} H_k$. Since the V_m are measurable and increase to $V = \bigcup V_m$, it follows from Theorem 3.26 that $\lim_{m\to\infty} |V_m| = |V|$. Since $E_m \subset V_m \subset H_m$, we have $|E_m|_e \le |V_m| \le |H_m| = |E_m|_e$. Hence, $|V_m| = |E_m|_e$ and $\lim_{m\to\infty} |E_m|_e = |V|$. However, $V = \bigcup V_m \supset \bigcup E_m = E$, and therefore, $\lim_{m\to\infty} |E_m|_e \ge |E|_e$. The opposite inequality is obvious since $E_m \subset E$, and the theorem follows.

3.4 Characterizations of Measurability

Lemma 3.22 characterizes measurability in terms of closed subsets of a set. The next three theorems give some other characterizations. The first one states that the most general measurable set differs from a Borel set by a set of measure zero.

Theorem 3.28

(i) *E is measurable if and only if $E = H - Z$, where H is of type G_δ and $|Z| = 0$.*

(ii) *E is measurable if and only if $E = H \cup Z$, where H is of type F_σ and $|Z| = 0$.*

Proof. If E has the representation expressed in either (i) or (ii), it is measurable since H and Z are.

Conversely, to prove the necessity in (i), suppose that E is measurable. For each $k = 1, 2, \ldots$, choose an open set G_k such that $E \subset G_k$ and $|G_k - E| < 1/k$. Set $H = \bigcap_k G_k$. Then H is of type G_δ, $E \subset H$, and $H - E \subset G_k - E$ for every k. Hence, $|H - E| = 0$, and (i) is proved.

The necessity of (ii) follows from that of (i) by taking complements: if E is measurable, so is CE, and therefore $CE = \bigcap G_k - Z$, where the G_k are open and $|Z| = 0$. Hence, $E = (\bigcup CG_k) \cup Z$, which completes the proof.

Theorem 3.29 *Suppose that $|E|_e < +\infty$. Then E is measurable if and only if given $\varepsilon > 0$, $E = (S \cup N_1) - N_2$, where S is a finite union of nonoverlapping intervals and $|N_1|_e, |N_2|_e < \varepsilon$.*

The proof is left as an exercise.

Our final characterization of measurability states that the measurable sets are those that split every set (measurable or not) into pieces that are additive with respect to outer measure. This characterization will be used in Chapter 11 to construct measures in abstract spaces.

Theorem 3.30 (Carathéodory) *A set E is measurable if and only if for every set A*

$$|A|_e = |A \cap E|_e + |A - E|_e.$$

Proof. Suppose that E is measurable. Given A, let H be a set of type G_δ such that $A \subset H$ and $|A|_e = |H|$. Since $H = (H \cap E) \cup (H - E)$, and since $H \cap E$ and $H - E$ are measurable and disjoint, $|H| = |H \cap E| + |H - E|$. Therefore, $|A|_e = |H \cap E| + |H - E| \geq |A \cap E|_e + |A - E|_e$. Since the opposite inequality is clearly true, we obtain $|A|_e = |A \cap E|_e + |A - E|_e$.

Conversely, suppose that E satisfies the stated condition for every A. In case $|E|_e < +\infty$, choose a G_δ set H such that $E \subset H$ and $|H| = |E|_e$. Then $H = E \cup (H - E)$, and by hypothesis, $|H| = |E|_e + |H - E|_e$. Therefore, $|H - E|_e = 0$; so the set $Z = H - E$ is measurable, and consequently, E is measurable.

In case $|E|_e = +\infty$, let B_k be the ball with center 0 and radius $k, k = 1, 2, \ldots$, and let $E_k = E \cap B_k$. Then each E_k has finite outer measure and $E = \bigcup E_k$. Let H_k be a set of type G_δ containing E_k with $|H_k| = |E_k|_e$. By hypothesis, $|H_k| = |H_k \cap E|_e + |H_k - E|_e \geq |E_k|_e + |H_k - E|_e$. Therefore, $|H_k - E| = 0$. Let $H = \bigcup H_k$. Then H is measurable, $E \subset H$, and $H - E = \bigcup (H_k - E)$. In particular, $H - E$ has measure zero, and since $E = H - (H - E)$, E is measurable. This completes the proof.

As a special case of Carathéodory's theorem, we obtain the following result.

Corollary 3.31 *If E is a measurable subset of a set A, then $|A|_e = |E| + |A - E|_e$. Hence, if $|E| < +\infty$, $|A - E|_e = |A|_e - |E|$.*

We can now prove a stronger version of Theorem 3.8.

Theorem 3.32 *Let E be a subset of $\mathbf{R^n}$. There exists a set H of type G_δ such that $E \subset H$ and for any measurable set M, $|E \cap M|_e = |H \cap M|$.*

If $M = \mathbf{R^n}$, this reduces to Theorem 3.8.

Proof. Consider first the case when $|E|_e < +\infty$. Let H be a set of type G_δ such that $E \subset H$ and $|E|_e = |H|$. If M is measurable, then by Carathéodory's theorem $|E|_e = |E \cap M|_e + |E - M|_e$ and $|H| = |H \cap M| + |H - M|$. Therefore,

$$|E \cap M|_e + |E - M|_e = |H \cap M| + |H - M|.$$

Since all these terms are finite, and since the inclusion $E - M \subset H - M$ implies that $|E - M|_e \leq |H - M|$, we have $|E \cap M|_e \geq |H \cap M|$. The opposite inequality is also true since $E \cap M \subset H \cap M$, and the theorem follows in this case.

In case $|E|_e = +\infty$, we write $E = \bigcup E_k$ with $|E_k|_e < +\infty$ and $E_k \nearrow E$. For example, E_k could be the intersection of E with the ball of radius k centered at the origin. By the case already considered, there is a set U_k of type G_δ such that $E_k \subset U_k$ and $|E_k \cap M|_e = |U_k \cap M|$ for any measurable M. Let $H_k = \bigcap_{m=k}^{\infty} U_m$. Then H_k is measurable (in fact, H_k is of type G_δ), $H_k \nearrow H = \bigcup H_k$, and $E_k \subset H_k \subset U_k$. Hence, $E_k \cap M \subset H_k \cap M \subset U_k \cap M$, and therefore, $|E_k \cap M|_e = |H_k \cap M|$ for measurable M. Since $E_k \nearrow E$ and $H_k \nearrow H$, we have $E_k \cap M \nearrow E \cap M$ and $H_k \cap M \nearrow H \cap M$. By Theorem 3.27, $|E \cap M|_e = |H \cap M|$ for measurable M.

Note that our set H is of type $G_{\delta\sigma}$. To obtain a set of type G_δ, use Theorem 3.28 to write $H = H_1 - Z$, where H_1 is of type G_δ and $|Z| = 0$. Then $E \subset H_1$. Moreover, $H_1 \cap M = (H \cap M) \cup (Z \cap M)$, and since $|Z| = 0$, we have $|H_1 \cap M| = |H \cap M|$, so that $|E \cap M|_e = |H_1 \cap M|$. This completes the proof.

3.5 Lipschitz Transformations of $\mathbf{R^n}$

Theorem 3.10 shows that the notion of outer measure is independent of the orientation of the coordinate axes. Since measurability and measure are

defined in terms of outer measure, it follows that these too are independent of rotation of the axes. We wish to study the effect of other transformations of $\mathbf{R^n}$ on the class of measurable sets; that is, we seek a condition on a transformation T: $\mathbf{R^n} \to \mathbf{R^n}$ such that the image $TE = \{\mathbf{y} : \mathbf{y} = T\mathbf{x}, \mathbf{x} \in E\}$ of every measurable set E is measurable. We note that a continuous transformation may not preserve measurability: see Exercise 17 of this chapter, as well as Chapter 7, Exercise 10.

A transformation $\mathbf{y} = T\mathbf{x}$ of $\mathbf{R^n}$ into itself is called a *Lipschitz transformation* if there is a constant c such that

$$|\mathbf{Tx} - \mathbf{Tx}'| \leq c|\mathbf{x} - \mathbf{x}'|.$$

The smallest such constant c, namely, the number

$$c = \sup_{\mathbf{x},\mathbf{x}';\mathbf{x}\neq\mathbf{x}'} \frac{|T\mathbf{x} - T\mathbf{x}'|}{|\mathbf{x} - \mathbf{x}'|},$$

is called the *Lipschitz constant of* T. If $y_j = f_j(\mathbf{x}), j = 1, \ldots, n$, are the coordinate functions representing T (see p. 12 in Section 1.7), it follows that T is Lipschitz if and only if each f_j satisfies a Lipschitz condition $|f_j(\mathbf{x}) - f_j(\mathbf{x}')| \leq c_j|\mathbf{x} - \mathbf{x}'|$. For example, a linear transformation of $\mathbf{R^n}$ is clearly a Lipschitz transformation; see Exercise 29. More generally, a mapping $y_j = f_j(\mathbf{x}), j = 1, \ldots, n$, for which each f_j has bounded first partial derivatives in $\mathbf{R^n}$ is a Lipschitz mapping.

Theorem 3.33 *If* $\mathbf{y} = T\mathbf{x}$ *is a Lipschitz transformation of* $\mathbf{R^n}$, *then* T *maps measurable sets into measurable sets.*

Proof. We will first show that a continuous transformation sends sets of type F_σ into sets of type F_σ. A continuous T maps compact sets into compact sets by Theorem 1.17; therefore, since any closed set can be written as a countable union of compact sets, T maps closed sets into sets of type F_σ. Here, we have used the relation $T\left(\bigcup E_k\right) = \bigcup TE_k$, which holds for any T and $\{E_k\}$. It follows that a continuous T preserves the class of F_σ sets.

We will next show that a Lipschitz transformation T sends sets of measure zero into sets of measure zero. Since $|T\mathbf{x} - T\mathbf{x}'| \leq c|\mathbf{x} - \mathbf{x}'|$, the image of a set with diameter d has diameter at most cd. It follows (see Exercise 28) by covering with cubes that there is a constant c' such that $|TI| \leq c'|I|$ for any interval I; note that TI is measurable since I is closed. If $|Z| = 0$ and $\varepsilon > 0$, choose intervals $\{I_k\}$ covering Z such that $\sum |I_k| < \varepsilon$. Since $TZ \subset \bigcup TI_k$, we have $|TZ|_e \leq \sum |TI_k| \leq c' \sum |I_k| < c'\varepsilon$. Hence, $|TZ| = 0$.

If E is a measurable set, we use Theorem 3.28 to write $E = H \cup Z$, where H is of type F_σ and $|Z| = 0$. Since $TE = TH \cup TZ$, TE is measurable as the union of measurable sets. This completes the proof.

For an extension of Theorem 3.33 in case $n = 1$, see Exercise 10 of Chapter 7.

In the special case that T is a linear transformation of $\mathbf{R^n}$, we will derive a formula for the measure of TE. By elementary linear algebra, such T has a unique $n \times n$ matrix representation with respect to every given basis of $\mathbf{R^n}$, and the value of the determinant of every matrix representation of T is the same. The common value is denoted det T and called the determinant of T. For the sake of definiteness, we may think of T as identified with its matrix representation with respect to the standard orthonormal basis of unit vectors along the coordinate axes. Then $T\mathbf{x}$ is the vector resulting from the matrix action of T on $\mathbf{x}$.

If P is a parallelepiped (see p. 7 in Section 1.3), arguments like those used in proving Theorems 3.2 and 3.10 show that

$$|P| = v(P) \tag{3.34}$$

(see Exercise 16). Hence, by p. 7 in Section 1.3, $|P|$ is the absolute value of the $n \times n$ determinant whose rows are the edges of P.

Theorem 3.35 *Let T be a linear transformation of $\mathbf{R^n}$, and let E be measurable. Then $|TE| = \delta|E|$, where δ is the absolute value of the determinant of T.**

Proof. Let $\delta = |\det T|$. Then for any interval I, $|TI| = \delta|I|$ by p. 8 in Section 1.3. If E is any measurable subset of $\mathbf{R^n}$ and $\varepsilon > 0$, choose intervals $\{I_k\}$ covering E with $\sum |I_k| \leq |E| + \varepsilon$. Then $|TE| \leq \sum |TI_k| = \delta \sum |I_k| \leq \delta(|E| + \varepsilon)$. Note here that if $\delta = 0$, then $|TI_k| = 0$ for every k and consequently the first inequality $|TE| \leq \sum |TI_k|$ yields $|TE| = 0$ even when $|E| = \infty$. Therefore,

$$|TE| \leq \delta|E|, \tag{3.36}$$

where we interpret $0 \cdot \infty$ as 0. To see that $|TE| = \delta|E|$, we may assume that $\delta > 0$ by (3.36). Then T has an inverse T^{-1} that is also linear, and $E = T^{-1}(TE)$. Therefore, by (3.36) applied to T^{-1} and the set TE,

$$|E| \leq |\det (T^{-1})|\, |TE| = \delta^{-1}|TE|,$$

or $|TE| \geq \delta|E|$. Hence, by (3.36), $\delta|E| = |TE|$. We leave it as an exercise to show that $|TE|_e = \delta|E|_e$ for any set E.

* Here $0 \cdot \infty$ should be interpreted as 0.

The following special case of Theorem 3.35 deserves mention: if E is a measurable set in $\mathbf{R^n}$ and λ is a real number, then the set λE defined by $\lambda E = \{\lambda x : x \in E\}$ is measurable with measure $|\lambda E| = |\lambda|^n |E|$. In fact, λE is the image of E under the matrix $\lambda \mathcal{I}$, where $\mathcal{I}$ is the identity matrix on $\mathbf{R^n}$.

3.6 A Nonmeasurable Set

We will now construct a nonmeasurable subset of $\mathbf{R^1}$; the construction in $\mathbf{R^n}, n > 1$, is similar. We will need the axiom of choice in the following form.

Zermelo's Axiom: Consider a family of arbitrary nonempty disjoint sets indexed by a set A, $\{E_\alpha \colon \alpha \in A\}$. Then there exists a set consisting of exactly one element from each E_α, $\alpha \in A$.

We also need the following lemma.

Lemma 3.37 *Let E be a measurable subset of $\mathbf{R^1}$ with $|E| > 0$. Then the set of differences $\{d : d = x - y, x \in E, y \in E\}$ contains an interval centered at the origin.*

Proof. Given $\varepsilon > 0$, since $|E| > 0$, there exists an open set G such that $E \subset G$ and $|G| \leq (1 + \varepsilon)|E|$. By Theorem 1.11, G can be written as a union of nonoverlapping intervals, $G = \bigcup I_k$. Letting $E_k = E \cap I_k$, it follows that $E = \bigcup E_k$, that the E_k are measurable, and that two different E_k's have at most one point in common. Therefore, $|G| = \sum |I_k|$ and $|E| = \sum |E_k|$. Since $|G| \leq (1 + \varepsilon)|E|$, we must have $|I_{k_0}| \leq (1+\varepsilon)|E_{k_0}|$ for some k_0. Choosing $\varepsilon = \frac{1}{3}$ and letting I and $\mathscr{E}$ denote the sets I_{k_0} and E_{k_0}, respectively, we have $\mathscr{E} \subset I$ and $|\mathscr{E}| \geq \frac{3}{4}|I|$. We claim that if $\mathscr{E}$ is translated by any number d satisfying $|d| < \frac{1}{2}|I|$, the translated set $\mathscr{E}_d$ has points in common with $\mathscr{E}$. Otherwise, since $\mathscr{E} \cup \mathscr{E}_d$ is contained in an interval of length $|I| + |d|$, we would have $|\mathscr{E}| + |\mathscr{E}_d| \leq |I| + |d|$, or $2|\mathscr{E}| \leq |I| + |d|$. [Here, we have used the fact that $\mathscr{E}_d$ is measurable and $|\mathscr{E}_d| = |\mathscr{E}|$ (see Exercise 18).] However, the last inequality is false if $|d| < \frac{1}{2}|I|$ since $|\mathscr{E}| \geq \frac{3}{4}|I|$. This proves the claim and thus the lemma. See Exercise 30 in Chapter 7 for a related fact.

Theorem 3.38 (Vitali) *There exist nonmeasurable sets.*

Proof. We define an equivalence relation on the real line by saying that x and y are equivalent if $x - y$ is rational. The equivalence classes then have the form $E_x = \{x + r : r \text{ is rational}\}$. Two classes E_x and E_y are either identical or disjoint;

therefore, one equivalence class consists of all the rational numbers, and the other distinct classes are disjoint sets of irrational numbers. The number of distinct equivalence classes is uncountable since each class is countable but the union of all the classes is uncountable (this union being the entire line).

Using Zermelo's axiom, let E be a set consisting of exactly one element from each distinct equivalence class. Since any two points of E must differ by an irrational number, the set $\{d : d = x - y, x \in E, y \in E\}$ cannot contain an interval. By Lemma 3.37, it follows that either E is not measurable or $|E| = 0$. Since the union of the translates of E by every rational number is all of $\mathbf{R^1}$, $\mathbf{R^1}$ would have measure zero if E did. We conclude that E is not measurable.

Corollary 3.39 *Any set in* $\mathbf{R^1}$ *with positive outer measure contains a nonmeasurable set.*

Proof. Let A satisfy $|A|_e > 0$, and let E be the nonmeasurable set of Theorem 3.38. For rational r, let E_r denote the translate of E by r. Then the E_r are disjoint and $\bigcup E_r = (-\infty, +\infty)$. Thus, $A = \bigcup (A \cap E_r)$ and $|A|_e \leq \sum |A \cap E_r|_e$. If $A \cap E_r$ is measurable, then $|A \cap E_r| = 0$ by Lemma 3.37, using the fact that for every r, $\{x - y : x, y \in A \cap E_r\}$ is a subset of $\{x - y : x, y \in E\}$ and so contains no interval. Since $|A|_e > 0$, it follows that there is some r such that $A \cap E_r$ is not measurable. This completes the proof.

Exercises

1. There is an analogue for bases different from 10 of the usual decimal expansion of a number. If b is an integer larger than 1 and $0 \leq x \leq 1$, show that there exist integral coefficients $c_k, 0 \leq c_k < b$, such that $x = \sum_{k=1}^{\infty} c_k b^{-k}$. Show that this expansion is unique unless $x = cb^{-k}, c = 1, \ldots, b^k - 1$, in which case there are two expansions.
2. When $b = 3$ in Exercise 1, the expansion is called the *triadic* or *ternary expansion of* x.
   (a) Show that the Cantor set C consists of all x such that x has *some* triadic expansion for which every c_k is either 0 or 2.
   (b) Let $f(x)$ be the Cantor–Lebesgue function: see p. 43 in Section 3.1. Show that if $x \in C$ and $x = \sum c_k 3^{-k}$, where each c_k is either 0 or 2, then $f(x) = \sum (\frac{1}{2} c_k) 2^{-k}$.
3. Construct a two-dimensional Cantor set in the unit square $\{(x, y) : 0 \leq x, y \leq 1\}$ as follows. Subdivide the square into nine equal parts and keep only the four closed corner squares, removing the remaining region (which forms a cross). Then repeat this process in a suitably scaled

version for the remaining squares, ad infinitum. Show that the resulting set is perfect, has plane measure zero, and equals $C \times C$.

4. Construct a subset of $[0,1]$ in the same manner as the Cantor set by removing from each remaining interval a subinterval of relative length $\theta, 0 < \theta < 1$. Show that the resulting set is perfect and has measure zero.

5. Construct a subset of $[0,1]$ in the same manner as the Cantor set, except that at the kth stage, each interval removed has length $\delta 3^{-k}, 0 < \delta < 1$. Show that the resulting set is perfect, has measure $1 - \delta$, and contains no intervals.

6. Construct a Cantor-type subset of $[0,1]$ by removing from each interval remaining at the kth stage a subinterval of relative length $\theta_k, 0 < \theta_k < 1$. Show that the remainder has measure zero if and only if $\sum \theta_k = +\infty$. (Use the fact that for $a_k > 0, \prod_{k=1}^{\infty} a_k$ converges, in the sense that $\lim_{N\to\infty} \prod_{k=1}^{N} a_k$ is finite and not zero, if and only if $\sum_{k=1}^{\infty} \log a_k$ converges.)

7. Prove Lemma 3.15.

8. Show that the Borel σ-algebra $\mathscr{B}$ in $\mathbf{R^n}$ is the smallest σ-algebra containing the closed sets in $\mathbf{R^n}$.

9. If $\{E_k\}_{k=1}^{\infty}$ is a sequence of sets with $\sum |E_k|_e < +\infty$, show that $\limsup E_k$ (and so also $\liminf E_k$) has measure zero.

10. If E_1 and E_2 are measurable, show that $|E_1 \cup E_2| + |E_1 \cap E_2| = |E_1| + |E_2|$.

11. Prove Theorem 3.29. (For the sufficiency, pick open sets G and G_1 with $S \subset G$, $N_1 \subset G_1$, $|G - S| < \varepsilon$, and $|G_1| < |N_1|_e + \varepsilon < 2\varepsilon$. Estimate $|(G \cup G_1) - E|_e$.)

12. If E_1 and E_2 are measurable subsets of $\mathbf{R^1}$, show that $E_1 \times E_2$ is a measurable subset of $\mathbf{R^2}$ and $|E_1 \times E_2| = |E_1||E_2|$. (Interpret $0 \cdot \infty$ as 0.) (Hint: Use a characterization of measurability.)

13. Motivated by (3.7), define the *inner measure* of E by $|E|_i = \sup |F|$, where the supremum is taken over all closed subsets F of E. Show that (i) $|E|_i \leq |E|_e$, and (ii) if $|E|_e < +\infty$, then E is measurable if and only if $|E|_i = |E|_e$. (Use Lemma 3.22.)

14. Show that the conclusion of part (ii) of Exercise 13 is false if $|E|_e = +\infty$.

15. If E is measurable and A is any subset of E, show that $|E| = |A|_i + |E - A|_e$. (See Exercise 13 for the definition of $|A|_i$.) As a consequence, using Exercise 13, show that if $A \subset [0,1]$ and $|A|_e + |[0,1] - A|_e = 1$, then A is measurable.

16. Prove (3.34).

17. Give an example that shows that the image of a measurable set under a continuous transformation may not be measurable. (Consider the Cantor–Lebesgue function and the pre-image of an appropriate nonmeasurable subset of its range.) See also Exercise 10 of Chapter 7.

18. Prove that outer measure is *translation invariant*; that is, if $E_{\mathbf{h}} = \{\mathbf{x} + \mathbf{h} : \mathbf{x} \in E\}$ is the translate of E by $\mathbf{h}$, $\mathbf{h} \in \mathbf{R^n}$, show that $|E_{\mathbf{h}}|_e = |E|_e$. If E is measurable, show that $E_{\mathbf{h}}$ is also measurable. (This fact was used in proving Lemma 3.37.)

19. Carry out the details of the construction of a nonmeasurable subset of $\mathbf{R^n}, n > 1$.

20. Show that there exist disjoint $E_1, E_2, \ldots, E_k, \ldots$ such that $|\bigcup E_k|_e < \sum |E_k|_e$ with strict inequality. (Let E be a nonmeasurable subset of $[0, 1]$ whose rational translates are disjoint. Consider the translates of E by all rational numbers $r, 0 < r < 1$, and use Exercise 18.)

21. Show that there exist sets $E_1, E_2, \ldots, E_k, \ldots$ such that $E_k \searrow E, |E_k|_e < +\infty$, and $\lim_{k\to\infty} |E_k|_e > |E|_e$ with strict inequality.

22. (a) Show that the outer measure of a set is unchanged if in the definition of outer measure we use coverings of the set by cubes with edges parallel to the coordinate axes instead of coverings by intervals.

(b) Show that outer measure is also unchanged if coverings by parallelepipeds with a fixed orientation (i.e., with edges parallel to a fixed set of n linearly independent vectors) are used rather than coverings by intervals.

23. Let Z be a subset of $\mathbf{R^1}$ with measure zero. Show that the set $\{x^2 : x \in Z)$ also has measure zero.

24. Let $0.\alpha_1\alpha_2\ldots$ be the dyadic development of any x in $[0, 1]$, that is, $x = \alpha_1 2^{-1} + \alpha_2 2^{-2} + \cdots$ with $\alpha_i = 0, 1$. Let $k_1, k_2, \ldots$ be a fixed permutation of the positive integers $1, 2, \ldots$, and consider the transformation T that sends $x = 0.\alpha_1\alpha_2\cdots$ to $Tx = 0.\alpha_{k_1}\alpha_{k_2}\ldots$. If E is a measurable subset of $[0, 1]$, show that its image TE is also measurable and that $|TE| = |E|$. (Consider first the special cases of E a dyadic interval $[s2^{-k}, (s+1)2^{-k}]$, $s = 0, 1, \ldots, 2^k - 1$, and then of E an open set [which is a countable union of nonoverlapping dyadic intervals]. Also show that if E has small measure, then so has TE.)

25. Construct a measurable subset E of $[0, 1]$ such that for every subinterval I, both $E \cap I$ and $I - E$ have positive measure. (Take a Cantor-type subset of $[0, 1]$ with positive measure [see Exercise 5], and on each subinterval of the complement of this set, construct another such set, and so on. The measures can be arranged so that the union of all the sets has the desired property.) See also Exercise 21 of Chapter 4.

26. Construct a continuous function on $[0, 1]$, which is not of bounded variation on any subinterval. (The construction follows the pattern of the Cantor–Lebesgue function with some modifications. At the first stage, for example, make the corresponding function increase to 2/3 (rather than 1/2) in (0, 1/3), then make it decrease by 1/3 in (1/3, 2/3), and then increase again 2/3 in (2/3, 1). The construction at other stages is

similar, depending on whether the preceding function was increasing or decreasing in the subinterval under consideration.)

27. Construct a continuous function of bounded variation on $[0,1]$ which is not monotone in any subinterval. (The construction is like that in the preceding exercise, except that the approximating functions are less steep. For example, at the first stage, let the function increase to $1/2+\varepsilon$, then decrease by 2ε, and then increase again by $1/2+\varepsilon$. Choose the ε's at each stage so that their sum converges.)

28. Prove the following assertion that is made in the proof of Theorem 3.33: If $T : \mathbf{R^n} \to \mathbf{R^n}$ is a Lipschitz transformation, then there is a constant $c' > 0$ such that $|TI| \le c'|I|$ for every interval I. (Consider first the case when I is a cube Q, noting that TQ is contained in a cube with edge length c diam Q where c is the Lipschitz constant of T. The case of general I can then be deduced by applying Theorem 1.11 to the interiors J° of intervals J with $I \subset J^\circ$.)

29. Let $T : \mathbf{R^n} \to \mathbf{R^n}$ be a linear transformation.

(a) If T has matrix representation (t_{ij}) and $t = \left(\sum_{i,j} t_{ij}^2\right)^{1/2}$, show that $|Tx - Ty| \le t|x-y|$ for all $x, y \in \mathbf{R^n}$. (Use (1.2).) The number t is called the *Hilbert–Schmidt norm* of (t_{ij}).

(b) Prove the fact mentioned at the end of the proof of Theorem 3.35 that $|TE|_e = \delta|E|_e$ for every $E \subset \mathbf{R^n}$, where $\delta = |\det T|$.

30. Let $f : \mathbf{R^n} \to \mathbf{R^1}$ be continuous. Show that the inverse image $f^{-1}(B)$ of a Borel set B is a Borel set; see p. 64 in Section 4.1 for the definition of the inverse image of a set. (The collection of sets $\{E : f^{-1}(E)$ is a Borel set$\}$ is a σ-algebra and contains all open sets in $\mathbf{R^1}$; cf. Exercise 10 of Chapter 4 and Corollary 4.15.) See also Exercise 22 of Chapter 4.

31. Construct a Lebesgue measurable set that is not Borel measurable. (If f is the Cantor–Lebesgue function, then the function $g(x) = x + f(x)$ is strictly increasing and continuous on $[0,1]$. Consider the set $g^{-1}(E)$ for an appropriate $E \subset g(C)$, where C is the Cantor set.)

32. Let E be a set in $\mathbf{R^n}$ with $|E|_e > 0$ and let θ satisfy $0 < \theta < 1$. Show that there is a set $E_\theta \subset E$ with $|E_\theta|_e = \theta|E|_e$ and that E_θ can be chosen to be measurable if E is measurable. (If $Q(r)$ denotes the cube with edge length r centered at the origin, $0 < r < \infty$, then $|E \cap Q(r)|_e$ is a continuous monotone function of r.)

33. Let E be a measurable set with $0 < |E| \le \infty$. Show that there are infinite collections $\{A_j\}, \{B_j\}$ of measurable subsets of E with the following properties: $0 < |A_j|, |B_j| < \infty$, $A_j \cap A_l = \emptyset$ if $j \ne l$, $B_{j+1} \subset B_j$, and $|B_j| \to 0$. See the end of Section 8.5 for an application. (This can be proved in many ways, for example, by using the Exercise 32.)

34. Let E and Z be sets in $\mathbf{R^n}$ and $|Z| = 0$. If $E \cup Z$ is measurable, show that E is measurable and that $|E| = |E \cup Z|$. See Exercise 2 of Chapter 10.

4

Lebesgue Measurable Functions

We will use the concept of Lebesgue measure to introduce a rich class of functions and a method of integrating these functions. In this chapter, we describe the class of functions.

Let E be a measurable set in $\mathbf{R^n}$. Let f be a real-valued (in the usual extended sense) function defined on E, that is, $-\infty \leq f(\mathbf{x}) \leq +\infty$, $\mathbf{x} \in E$. Then, f is called a *Lebesgue measurable function on* E, or simply a *measurable function*, if for every finite a, the set

$$\{\mathbf{x} \in E : f(\mathbf{x}) > a\}$$

is a measurable subset of $\mathbf{R^n}$. In what follows, we shall often use the abbreviation $\{f > a\}$ for $\{\mathbf{x} \in E : f(\mathbf{x}) > a\}$. Note that the definition of measurability of a function on a set E presupposes that E is measurable. Since

$$E = \{f = -\infty\} \cup \left(\bigcup_{k=1}^{\infty} \{f > -k\} \right),$$

the measurability of E implies that of $\{f = -\infty\}$ if we assume that f is measurable.

As a varies, the behavior of the sets $\{f > a\}$ describes how the values of f are distributed. Intuitively, it is clear that the smoother f is, the smaller the variety of such sets will be. For example, if $E = \mathbf{R^n}$ and f is continuous in $\mathbf{R^n}$, then $\{f > a\}$ is always open. A function f defined on a Borel set E is said to be *Borel measurable* if $\{f > a\}$ is a Borel set for every a. Thus, every Borel measurable function is measurable. See also Exercise 24.

One further comment will be helpful later. Let $\mathscr{M}$ denote the class of measurable subsets of $\mathbf{R^n}$. Much of the development of measurable functions given in this chapter depends only on the σ-algebra structure of $\mathscr{M}$ and the properties of Lebesgue measure. Thus, a measurable function inherits its elementary properties from those of measurable sets. It is reasonable to expect, therefore, that many of the methods of this chapter can be used to develop results in more general settings, for spaces other than $\mathbf{R^n}$ and σ-algebras other than $\mathscr{M}$. This will be done in Chapter 10. To save too much repetition there, it will be helpful if the reader notices which properties of $\mathscr{M}$ and Lebesgue measure are used in the proofs here. These will be discussed at the end of the chapter.

4.1 Elementary Properties of Measurable Functions

Theorem 4.1 *Let f be a real-valued function defined on a measurable set E. Then f is measurable if and only if any of the following statements holds for every finite a:*

(i) $\{f \geq a\}$ *is measurable.*
(ii) $\{f < a\}$ *is measurable.*
(iii) $\{f \leq a\}$ *is measurable.*

Proof. Since $\{f \geq a\} = \bigcap_{k=1}^{\infty}\{f > a - 1/k\}$, the measurability of f implies (i). Since $\{f < a\}$ is the complement of $\{f \geq a\}$, it follows that (i) implies (ii). Since $\{f \leq a\} = \bigcap_{k=1}^{\infty}\{f < a+1/k\}$, we see that (ii) implies (iii). Finally, since $\{f > a\}$ is the complement of $\{f \leq a\}$, it follows that f is measurable if (iii) holds.

The proof of the following is left as an exercise.

Corollary 4.2 *Let f be defined on a measurable set E. If f is measurable, then $\{f > -\infty\}$, $\{f < +\infty\}$, $\{f = +\infty\}$, $\{a \leq f \leq b\}$, $\{f = a\}$, etc., are all measurable. Moreover, for any f, if either $\{f = +\infty\}$ or $\{f = -\infty\}$ is measurable, then f is measurable if $\{a < f < +\infty\}$ is measurable for every finite a.*

Also, observe that $\{a < f \leq b\} = \{f > a\} - \{f > b\}$.

Let f be defined in E. If S is a subset of $\mathbf{R^1}$, the *inverse image of S* under f is defined by

$$f^{-1}(S) = \{\mathbf{x} \in E : f(\mathbf{x}) \in S\}.$$

Theorem 4.3 *Let f be defined on a measurable set E. If f is measurable, then for every open set G in $\mathbf{R^1}$, the inverse image $f^{-l}(G)$ is a measurable subset of $\mathbf{R^n}$. Conversely, f is measurable if $f^{-1}(G)$ is measurable for every open set G in $\mathbf{R^n}$ and either $\{f = +\infty\}$ or $\{f = -\infty\}$ is measurable.*

Proof. Suppose that f is measurable and let G be any open subset of $\mathbf{R^1}$. By Theorem 1.10, G can be written as $G = \bigcup_k (a_k, b_k)$. But $f^{-1}((a_k, b_k))$ equals $\{a_k < f < b_k\}$ and is therefore measurable. Since $f^{-1}(G) = \bigcup_k f^{-1}((a_k, b_k))$, it follows that $f^{-1}(G)$ is measurable too. To prove the converse, note that if $G = (a, +\infty)$, then $f^{-1}(G) = \{a < f < +\infty\}$ and apply the second part of Corollary 4.2.

This result shows that a *finite* function f defined on a measurable set is measurable if and only if $f^{-1}(G)$ is measurable for every open $G \subset \mathbf{R^1}$. Similarly, a finite f defined on a Borel set is Borel measurable if and only if $f^{-1}(G)$ is Borel measurable for every open $G \subset \mathbf{R^1}$.

Theorem 4.4 *Let A be a dense subset of $\mathbf{R^1}$. Then f is measurable if $\{f > a\}$ is measurable for all $a \in A$.*

Proof. Given any real a, choose a sequence $\{a_k\}$ in A that converges to a from above: $a_k \in A, a_k \geq a, \lim_{k\to\infty} a_k = a$. Then $\{f > a\} = \bigcup\{f > a_k\}$, and the theorem follows.

A property is said to hold *almost everywhere in* E or, in abbreviated form, *a.e.*, if it holds in E except in some subset of E with measure zero. For example, the statement "$f = 0$ a.e. in E" means that $f(\mathbf{x}) = 0$ in E, with the possible exception of those $\mathbf{x}$ in some subset Z of E with $|Z| = 0$.

The next few theorems give some simple properties of the class of measurable functions.

Theorem 4.5 *If f is measurable and if $g = f$ a.e., then g is measurable and $|\{g > a\}| = |\{f > a\}|$.*

Proof. If $Z = \{f \neq g\}$, then $|Z| = 0$ and $\{g > a\} \cup Z = \{f > a\} \cup Z$. Therefore, f being measurable, $\{g > a\} \cup Z$ is measurable, and since this differs from $\{g > a\}$ by a set of measure zero, g is measurable (cf. Exercise 34 of Chapter 3 and Exercise 2 of Chapter 10). Finally,

$$|\{g > a\}| = |\{g > a\} \cup Z| = |\{f > a\} \cup Z| = |\{f > a\}|.$$

In view of the previous theorem, it is natural to extend the definition of measurability to include functions that are defined only a.e. in E, by saying that such an f is measurable on E if it is measurable on the subset of E where it is defined. Note also that if f is measurable on E, then it is measurable on any measurable $E_1 \subset E$ since $\{\mathbf{x} \in E_1 : f(\mathbf{x}) > a\} = \{\mathbf{x} \in E : f(\mathbf{x}) > a\} \cap E_1$.

If ϕ and f are finite measurable functions defined on $\mathbf{R^1}$ and $\mathbf{R^n}$, respectively, their composition $\phi(f(\mathbf{x}))$ may not be measurable (see Exercise 5). If ϕ is continuous, however, we have the following result.

Theorem 4.6 *Let ϕ be continuous on $\mathbf{R^1}$ and let f be finite a.e. in E, so that, in particular, $\phi(f)$ is defined a.e. in E. Then $\phi(f)$ is measurable if f is.*

Proof. We may assume that f is finite everywhere in E. We will use the familiar fact that since ϕ is continuous, the inverse image $\phi^{-1}(G)$ of an open set G is open (cf. Corollary 4.15 and Exercise 10). By Theorem 4.3, it is enough to show that for every open G in $\mathbf{R^1}$, $\{\mathbf{x} : \phi(f(\mathbf{x})) \in G\}$ is measurable. However, $\{\mathbf{x} : \phi(f(\mathbf{x})) \in G\} = f^{-1}(\phi^{-1}(G))$, and since $\phi^{-1}(G)$ is open and f is measurable, $f^{-1}(\phi^{-1}(G))$ is measurable by Theorem 4.3. See also Exercise 22(b).

Remark: The cases that arise most frequently are $\phi(t) = |t|$, $|t|^p (p > 0)$, e^{ct}, etc. Thus,

$$|f|, \ |f|^p (p > 0), \ e^{cf}$$

are measurable if f is measurable (even if we do not assume that f is finite a.e., as is easily seen). Another special case worth mentioning is that of

$$f^+ = \max\{f, 0\}, \quad f^- = -\min\{f, 0\}.$$

It is enough to observe that the functions x^+ and x^- are continuous.

Theorem 4.7 *If f and g are measurable, then $\{f > g\}$ is measurable.*

Proof. Let $\{r_k\}$ be the rational numbers. Then,

$$\{f > g\} = \bigcup_k \{f > r_k > g\} = \bigcup_k (\{f > r_k\} \cap \{g < r_k\}),$$

and the theorem follows.

Theorem 4.8 *If f is measurable and λ is any real number, then $f + \lambda$ and λf are measurable.*

The proof is left as an exercise. We interpret $0 \cdot \pm\infty$ to be 0.

Note that the sum $f + g$ of two functions f and g is well defined wherever it is not of the form $+\infty + (-\infty)$ or $-\infty + \infty$. In the next theorem, we assume for simplicity that $f + g$ is well defined everywhere. See Exercise 6 for extensions.

Theorem 4.9 *If f and g are measurable, so is $f + g$.*

Proof. Since g is measurable, so is $a - g$ for any real a, by Theorem 4.8. Since $\{f + g > a\} = \{f > a - g\}$, the result follows from Theorem 4.7.

A corollary of Theorems 4.8 and 4.9 is that a finite linear combination $\lambda_1 f_1 + \cdots + \lambda_N f_N$ of measurable functions $f_1, \ldots, f_N$ is measurable provided it is well defined. Thus, the class of measurable functions on a set E that are finite a.e. in E forms a vector space; here, we identify measurable functions that are equal a.e.

In the theorem that follows, we consider products of functions. In addition to the familiar conventions about products of infinities, we adopt as usual the convention that $0 \cdot \pm\infty = \pm\infty \cdot 0 = 0$. Also, if $-\infty \le \alpha \le +\infty$, we interpret $\alpha/(\pm\infty) = \alpha \cdot 0 = 0$.

Theorem 4.10 *If f and g are measurable, so is fg. If $g \neq 0$ a.e., then f/g is measurable.*

Proof. By Theorem 4.6 and the remark following it, $f^2 (= |f|^2)$ is measurable if f is. Hence, if f and g are measurable and finite, the formula $fg = [(f+g)^2 - (f-g)^2]/4$ implies that fg is measurable. The proof when f and g can be infinite and the proof of the second statement of the theorem are left as exercises.

Theorem 4.11 *If $\{f_k(\mathbf{x})\}_{k=1}^{\infty}$ is a sequence of measurable functions, then $\sup_k f_k(\mathbf{x})$ and $\inf_k f_k(\mathbf{x})$ are measurable.*

Proof. Since $\inf_k f_k = -\sup_k(-f_k)$, it is enough to prove the result for $\sup_k f_k$. But this follows easily from the fact that $\{\sup_k f_k > a\} = \bigcup_k \{f_k > a\}$.

As a special case of the preceding theorem, we see that if $f_1, \ldots, f_N$ are measurable, then so are $\max_k f_k$ and $\min_k f_k$. In particular, if f is measurable, then so are $f^+ = \max\{f, 0\}$ and $f^- = -\min\{f, 0\}$, a fact we have already observed in the remark following Theorem 4.6.

Theorem 4.12 *If $\{f_k\}$ is a sequence of measurable functions, then $\limsup_{k\to\infty} f_k$ and $\liminf_{k\to\infty} f_k$ are measurable. In particular, if $\lim_{k\to\infty} f_k$ exists a.e., it is measurable.*

Proof. Since

$$\limsup_{k\to\infty} f_k = \inf_j \{\sup_{k\ge j} f_k\}, \quad \liminf_{k\to\infty} f_k = \sup_j \{\inf_{k\ge j} f_k\},$$

the first statement is a corollary of Theorem 4.11. The second statement then follows by Theorem 4.5 since wherever $\lim_{k\to\infty} f_k$ exists, it equals $\limsup_{k\to\infty} f_k$.

The *characteristic function*, or *indicator function*, $\chi_A(\mathbf{x})$, of a set A is defined by

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

Clearly, χ_A is measurable if and only if A is measurable. χ_A is an example of what is called a simple function on $\mathbf{R^n}$: a *simple function on a set* $E \subset \mathbf{R^n}$ is one that is defined on E and assumes only a finite number of values on E, all of which are finite. If f is a simple function on E taking (distinct) values $a_1, \ldots, a_N$ on (disjoint) subsets $E_1, \ldots, E_N$ of E, $E = \bigcup_{k=1}^N E_k$, then

$$f(\mathbf{x}) = \sum_{k=1}^{N} a_k \chi_{E_k}(\mathbf{x}), \quad \mathbf{x} \in E.$$

We leave it as an exercise to show that such an f is measurable if and only if $E_1, \ldots, E_N$ are measurable.

Theorem 4.13

(i) *Every function f can be written as the limit of a sequence $\{f_k\}$ of simple functions.*

(ii) *If $f \geq 0$, the sequence can be chosen to increase to f, that is, chosen such that $f_k \leq f_{k+1}$ for every k.*

(iii) *If the function f in either* (i) *or* (ii) *is measurable, then the f_k can be chosen to be measurable.*

Proof. We will prove (ii) first. Thus, suppose that $f \geq 0$. For each $k, k = 1, 2, \ldots$, subdivide the values of f that fall in $[0, k]$ by partitioning $[0, k]$ into subintervals $[(j-1)2^{-k}, j2^{-k}], j = 1, \ldots, k2^k$. Let

$$f_k(\mathbf{x}) = \begin{cases} \dfrac{j-1}{2^k} & \text{if } \dfrac{j-1}{2^k} \leq f(\mathbf{x}) < \dfrac{j}{2^k},\ j = 1, \ldots, k2^k \\ k & \text{if } f(\mathbf{x}) \geq k. \end{cases}$$

Each f_k is a simple function defined everywhere in the domain of f. Clearly, $f_k \leq f_{k+1}$ since in passing from f_k to f_{k+1}, each subinterval $[(j-1)2^{-k}, j2^{-k}]$ is divided in half. Moreover, $f_k \to f$ since $0 \leq f - f_k \leq 2^{-k}$ for sufficiently large k wherever f is finite, and $f_k = k \to +\infty$ wherever $f = +\infty$. This proves (ii).

To prove (i), apply the result of (ii) to each of the nonnegative functions f^+ and f^-, obtaining increasing sequences $\{f_k'\}$ and $\{f_k''\}$ of simple functions such that $f_k' \to f^+$ and $f_k'' \to f^-$. Then $f_k' - f_k''$ is simple and $f_k' - f_k'' \to f^+ - f^- = f$.

Finally, it is enough to prove (iii) for $f \geq 0$ since otherwise we may consider f^+ and f^-. In this case, however,

$$f_k = \sum_{j=1}^{k2^k} \frac{j-1}{2^k} \chi_{\{(j-1)/2^k \leq f < j/2^k\}} + k\chi_{\{f \geq k\}}.$$

This is measurable if f is since all the sets involved are measurable.

Note that if f is bounded, the simple functions earlier will converge uniformly to f.

4.2 Semicontinuous Functions

We now study classes of functions f whose continuity properties on a set can be characterized by the topological nature of $\{f > a\}$ or $\{f < a\}$. Measurability of such functions will consequently be easy to establish. We will encounter a particularly important example when we study the Hardy–Littlewood maximal function in Chapter 7.

Let f be defined on E, and let $\mathbf{x}_0$ be a limit point of E that lies in E. Then f is said to be *upper semicontinuous at* $\mathbf{x}_0$ if

$$\limsup_{\mathbf{x} \to \mathbf{x}_0; \mathbf{x} \in E} f(\mathbf{x}) \leq f(\mathbf{x}_0).$$

We will usually abbreviate this by saying that f is *usc at* $\mathbf{x}_0$. Note that if $f(\mathbf{x}_0) = +\infty$, then f is automatically usc at $\mathbf{x}_0$; otherwise, the statement that f is usc at $\mathbf{x}_0$ means that given $M > f(\mathbf{x}_0)$, there exists $\delta > 0$ such that $f(\mathbf{x}) < M$ for all $\mathbf{x} \in E$ that lie in the ball $|\mathbf{x} - \mathbf{x}_0| < \delta$. Intuitively, this means that near $\mathbf{x}_0$, the values of f do not exceed $f(\mathbf{x}_0)$ by a fixed amount.

Similarly, f is said to be *lower semicontinuous* at $\mathbf{x}_0$, or *lsc at* $\mathbf{x}_0$, if

$$\liminf_{\mathbf{x} \to \mathbf{x}_0; \mathbf{x} \in E} f(\mathbf{x}) \geq f(\mathbf{x}_0).$$

Thus, if $f(\mathbf{x}_0) = -\infty$, f is lsc at $\mathbf{x}_0$, while if $f(\mathbf{x}_0) > -\infty$, the definition amounts to saying that given $m < f(\mathbf{x}_0)$, there exists $\delta > 0$ such that $f(\mathbf{x}) > m$ if $\mathbf{x} \in E$ and $|\mathbf{x} - \mathbf{x}_0| < \delta$. Equivalently, f is lsc at $\mathbf{x}_0$ if and only if $-f$ is usc at $\mathbf{x}_0$.

It follows that f is continuous at $\mathbf{x}_0$ if and only if $|f(\mathbf{x}_0)| < +\infty$ and f is both usc and lsc at $\mathbf{x}_0$. As simple examples of functions that are usc everywhere in $\mathbf{R}^1$ but not continuous at some $\mathbf{x}_0$, we have

$$u_1(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} < \mathbf{x}_0 \\ 1 & \text{if } \mathbf{x} \geq \mathbf{x}_0, \end{cases} \qquad u_2(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \neq \mathbf{x}_0 \\ 1 & \text{if } \mathbf{x} = \mathbf{x}_0. \end{cases}$$

Hence, $-u_1$ and $-u_2$ are lsc everywhere in $\mathbf{R}^1$. The Dirichlet function of Example 4 in Chapter 2 is usc at the rational numbers and lsc at the irrationals.

A function defined on E is called *usc* (*lsc, continuous*) *relative to* E if it is usc (lsc, continuous) at every limit point of E that is in E. The next theorem characterizes functions that are semicontinuous relative to a set.

Theorem 4.14

(i) *A function f is usc relative to E if and only if $\{\mathbf{x} \in E : f(\mathbf{x}) \geq a\}$ is relatively closed [equivalently, $\{\mathbf{x} \in E : f(\mathbf{x}) < a\}$ is relatively open] for all finite a.*

(ii) *A function f is lsc relative to E if and only if $\{\mathbf{x} \in E : f(\mathbf{x}) \leq a\}$ is relatively closed [equivalently, $\{\mathbf{x} \in E : f(\mathbf{x}) > a\}$ is relatively open] for all finite a.*

Proof. Statements (i) and (ii) are equivalent since f is usc if and only if $-f$ is lsc. It is therefore enough to prove (i). Suppose first that f is usc relative to E. Given a, let $\mathbf{x}_0$ be a limit point of $\{\mathbf{x} \in E : f(\mathbf{x}) \geq a\}$ that is in E. Then there exist $\mathbf{x}_k \in E$ such that $\mathbf{x}_k \to \mathbf{x}_0$ and $f(\mathbf{x}_k) \geq a$. Since f is usc at $\mathbf{x}_0$, we have $f(\mathbf{x}_0) \geq \limsup_{k\to\infty} f(\mathbf{x}_k)$. Therefore, $f(\mathbf{x}_0) \geq a$, so that $\mathbf{x}_0 \in \{\mathbf{x} \in E : f(\mathbf{x}) \geq a\}$. This shows that $\{\mathbf{x} \in E : f(\mathbf{x}) \geq a\}$ is relatively closed.

Conversely, let $\mathbf{x}_0$ be a limit point of E that is in E. If f is not usc at $\mathbf{x}_0$, then $f(\mathbf{x}_0) < +\infty$, and there exist M and $\{\mathbf{x}_k\}$ such that $f(\mathbf{x}_0) < M$, $\mathbf{x}_k \in E, \mathbf{x}_k \to \mathbf{x}_0$, and $f(\mathbf{x}_k) \geq M$. Hence, $\{\mathbf{x} \in E : f(\mathbf{x}) \geq M\}$ is not relatively closed since it does not contain all its limit points that are in E.

Corollary 4.15 *A finite function f is continuous relative to E if and only if all sets of the form $\{\mathbf{x} \in E : f(\mathbf{x}) \geq a\}$ and $\{\mathbf{x} \in E : f(\mathbf{x}) \leq a\}$ are relatively closed [or, equivalently, all $\{\mathbf{x} \in E : f(\mathbf{x}) > a\}$ and $\{\mathbf{x} \in E : f(\mathbf{x}) < a\}$ are relatively open] for finite a.*

Corollary 4.16 *Let E be measurable, and let f be defined on E. If f is usc (lsc, continuous) relative to E, then f is measurable.*

Proof. Let f be usc relative to E. Since $\{\mathbf{x} \in E : f(\mathbf{x}) \geq a\}$ is relatively closed, it is the intersection of E with a closed set. Hence, it is measurable, and the result follows from Theorem 4.1.

The previous three results deserve special attention in certain cases. Suppose, for example, that $E = \mathbf{R}^n$ and f is usc everywhere in $\mathbf{R}^n$. Since $\{f > a\} = \bigcup_{k=1}^{\infty}\{f \geq a + 1/k\}$, it follows from Theorem 4.14 that $\{f > a\}$

is of type F_σ. Since an F_σ set is a Borel set, we see that a function that is usc (similarly, lsc or continuous) at every point of $\mathbf{R^n}$ is Borel measurable.

4.3 Properties of Measurable Functions and Theorems of Egorov and Lusin

Our next theorem states in effect that if a sequence of measurable functions converges at each point of a set E, then, with the exception of a subset of E with arbitrarily small measure, the sequence actually converges *uniformly.* This remarkable result cannot hold, at least in the form just stated, without some further restrictions. For example, if $E = \mathbf{R^n}$ and $f_k = \chi_{\{\mathbf{x}:|\mathbf{x}|<k\}}$, then f_k converges to 1 everywhere but does not converge uniformly outside any bounded set. Again, if the f_k are finite but the limit f is infinite in a set of positive measure, then $|f_k - f|$ is also infinite in this set. The difficulties in these examples can be easily overcome: the missing ingredient in the first case is that $|E| < +\infty$ and in the second that $|f| < +\infty$ a.e. Adding these restrictions, we obtain the following basic result.

Theorem 4.17 (Egorov's Theorem) *Suppose that $\{f_k\}$ is a sequence of measurable functions that converges almost everywhere in a set E of finite measure to a finite limit f. Then given $\varepsilon > 0$, there is a closed subset F of E such that $|E - F| < \varepsilon$ and $\{f_k\}$ converge uniformly to f on F.*

In order to prove this, we need a preliminary result that is interesting in its own right.

Lemma 4.18 *Under the same hypothesis as in Egorov's theorem, given $\varepsilon, \eta > 0$, there is a closed subset F of E and an integer K such that $|E - F| < \eta$ and $|f(\mathbf{x}) - f_k(\mathbf{x})| < \varepsilon$ for $\mathbf{x} \in F$ and $k > K$.*

Proof. Fix $\varepsilon, \eta > 0$. For each m, let $E_m = \{|f - f_k| < \varepsilon \text{ for all } k > m\}$. Thus, $E_m = \bigcap_{k>m}\{|f - f_k| < \varepsilon\}$, so that E_m is measurable. Clearly, $E_m \subset E_{m+1}$. Moreover, since $f_k \to f$ a.e. in E and f is finite, $E_m \nearrow E - Z, |Z| = 0$. Hence, by Theorem 3.26, $|E_m| \to |E - Z| = |E|$. Since $|E| < +\infty$, it follows that $|E - E_m| \to 0$. Choose m_0 so that $|E - E_{m_0}| < \frac{1}{2}\eta$, and let F be a closed subset of E_{m_0} with $|E_{m_0} - F| < \frac{1}{2}\eta$. Then $|E - F| < \eta$, and $|f - f_k| < \varepsilon$ in F if $k > m_0$.

Proof of Egorov's theorem. Given $\varepsilon > 0$, use Lemma 4.18 to select closed $F_m \subset E, m \geq 1$, and integers $K_{m,\varepsilon}$ such that $|E - F_m| < \varepsilon 2^{-m}$ and $|f - f_k| < 1/m$

in F_m if $k > K_{m,\varepsilon}$. The set $F = \bigcap_m F_m$ is closed, and since $F \subset F_m$ for all m, f_k converges uniformly to f on F. Finally, $E - F = E - \bigcap F_m = \bigcup(E - F_m)$ and, therefore, $|E - F| \leq \sum |E - F_m| < \varepsilon$. This completes the proof.

See Exercises 13 and 14 for an analogue of Egorov's theorem in the continuous parameter case, that is, in the case when $f_y(\mathbf{x}) \to f(\mathbf{x})$ as $y \to y_0$.

We have observed that a continuous function is measurable. Our next result, Lusin's theorem, gives a continuity property that characterizes measurable functions. In order to state the result, we first make the following definition. A function f defined on a measurable set E has *property* $\mathscr{C}$ on E if given $\varepsilon > 0$, there is a closed set $F \subset E$ such that

(i) $|E - F| < \varepsilon$

(ii) f is continuous relative to F

We recall that condition (ii) means that if $\mathbf{x}_0$ and $\{\mathbf{x}_k\}$ belong to F and $\mathbf{x}_k \to \mathbf{x}_0$, then $f(\mathbf{x}_0)$ is finite and $f(\mathbf{x}_k) \to f(\mathbf{x}_0)$. In case F is bounded (and, therefore, compact), (ii) implies that the restriction of f to F is uniformly continuous (Theorem 1.15).

Lemma 4.19 *A simple measurable function has property $\mathscr{C}$.*

Proof. Suppose that f is a simple measurable function on E, taking distinct values $a_1, \ldots, a_N$ on measurable subsets $E_1, \ldots, E_N$. Given $\varepsilon > 0$, choose closed $F_j \subset E_j$ with $|E_j - F_j| < \varepsilon/N$. Then the set $F = \bigcup_{j=1}^N F_j$ is closed, and since $E - F = \bigcup E_j - \bigcup F_j \subset \bigcup(E_j - F_j)$, we have $|E - F| \leq \sum |E_j - F_j| < \varepsilon$. It remains only to show that f is continuous on F. Note that each F_j is relatively open in F, in fact, $F_j = F \cap C\big(\bigcup_{k \neq j} F_k\big)$, so the only points of F in a small neighborhood of any point of F_j are points of F_j itself. The continuity on F of f follows from this since f is constant on each F_j.

Property $\mathscr{C}$ is actually equivalent to measurability, as we now show.

Theorem 4.20 (Lusin's Theorem) *Let f be defined and finite on a measurable set E. Then f is measurable if and only if it has property $\mathscr{C}$ on E.*

Proof. If f is measurable, then by Theorem 4.13, there exist simple measurable $f_k, k = 1, 2, \ldots$, which converge to f. By Lemma 4.19, each f_k has property $\mathscr{C}$, so given $\varepsilon > 0$, there exist closed $F_k \subset E$ such that $|E - F_k| < \varepsilon 2^{-k-1}$ and f_k is continuous relative to F_k. Assuming for the moment that $|E| < +\infty$, we see by Egorov's theorem that there is a closed $F_0 \subset E$ with $|E - F_0| < \frac{1}{2}\varepsilon$ such that

$\{f_k\}$ converges uniformly to f on F_0. If $F = F_0 \cap (\bigcap_{k=1}^{\infty} F_k)$, then F is closed, each f_k is continuous relative to F, and $\{f_k\}$ converges uniformly to f on F. Hence, f is continuous relative to F by Theorem 1.16. Since

$$|E - F| \leq |E - F_0| + \sum_{k=1}^{\infty} |E - F_k| < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon,$$

it follows that f has property $\mathscr{C}$ on E. This proves the necessity of property $\mathscr{C}$ for measurability if $|E| < +\infty$.

If $|E| = +\infty$, write $E = \bigcup_{k=1}^{\infty} E_k$, where E_k is the part of E in the ring $\{\mathbf{x} : k - 1 \leq |\mathbf{x}| < k\}$. Since $|E_k| < +\infty$, we may select closed $F_k \subset E_k$ such that $|E_k - F_k| < \varepsilon 2^{-k}$ and f is continuous relative to F_k. If $F = \bigcup_{k=1}^{\infty} F_k$, it follows that $|E - F| \leq \sum |E_k - F_k| < \varepsilon$ and that f is continuous relative to F. Different F_k lie in different rings, and the distance $d(F_k, F_l)$ between F_k and F_l is positive if $k \neq l$. A simple argument shows that F is closed, and therefore f has property $\mathscr{C}$ on E.

Conversely, suppose that f has property $\mathscr{C}$ on E. For each $k, k = 1, 2, \ldots$, choose a closed $F_k \subset E$ such that $|E - F_k| < 1/k$ and the restriction of f to F_k is continuous. If $H = \bigcup_{k=1}^{\infty} F_k$, then $H \subset E$ and the set $Z = E - H$ has measure zero. We have

$$\begin{aligned}\{\mathbf{x} \in E : f(\mathbf{x}) > a\} &= \{\mathbf{x} \in H : f(\mathbf{x}) > a\} \cup \{\mathbf{x} \in Z : f(\mathbf{x}) > a\} \\ &= \bigcup_{k=1}^{\infty} \{\mathbf{x} \in F_k : f(\mathbf{x}) > a\} \cup \{\mathbf{x} \in Z : f(\mathbf{x}) > a\}.\end{aligned}$$

Since $\{\mathbf{x} \in Z : f(\mathbf{x}) > a\}$ has measure zero, the measurability of f will follow from that of each $\{\mathbf{x} \in F_k : f(\mathbf{x}) > a\}$. However, since f is continuous relative to F_k and F_k is measurable, $\{\mathbf{x} \in F_k : f(\mathbf{x}) > a\}$ is measurable by Corollary 4.16. This completes the proof of Lusin's theorem.

4.4 Convergence in Measure

Let f and $\{f_k\}$ be measurable functions that are defined and finite a.e. in a set E. Then $\{f_k\}$ is said to *converge in measure* on E to f if for every $\varepsilon > 0$,

$$\lim_{k \to \infty} |\{\mathbf{x} \in E : |f(\mathbf{x}) - f_k(\mathbf{x})| > \varepsilon\}| = 0.$$

We will indicate convergence in measure by writing

$$f_k \xrightarrow{m} f.$$

This concept has many useful applications in analysis. Here, we will discuss its relation to ordinary pointwise convergence; the first result is basically a reformulation of Lemma 4.18.

Theorem 4.21 *Let f and $f_k, k = 1, 2, \ldots$, be measurable and finite a.e. in E. If $f_k \to f$ a.e. on E and $|E| < +\infty$, then $f_k \xrightarrow{m} f$ on E.*

Proof. Given $\varepsilon, \eta > 0$, let F and K be as defined in Lemma 4.18. Then if $k > K$, $\{\mathbf{x} \in E : |f(\mathbf{x}) - f_k(\mathbf{x})| > \varepsilon\} \subset E - F$, and since $|E - F| < \eta$, the result follows.

We recall that this conclusion may not hold if $|E| = +\infty$, as shown by the example $E = \mathbf{R^n}$, $f_k = \chi_{\{\mathbf{x}:|\mathbf{x}|<k\}}$, and $f = 1$.

Convergence in measure does not imply pointwise convergence a.e., even for sets of finite measure. To see this, take n =1 and let $\{I_k\}$ be a sequence of subintervals of [0,1] satisfying the following conditions:

(i) Each point of [0,1] belongs to infinitely many I_k.
(ii) $\lim_{k\to\infty} |I_k| = 0$.

For example, let the first interval be [0,1], the next two be the two halves of [0,1], and the next four be the four quarters, and so on. Then if $f_k = \chi_{I_k}$, we have $f_k \xrightarrow{m} 0$, while f_k diverges at every point of [0,1].

There is, however, the following partial converse to Theorem 4.21.

Theorem 4.22 *If $f_k \xrightarrow{m} f$ on E, there is a subsequence f_{k_j} such that $f_{k_j} \to f$ a.e. in E.*

Proof. Since $f_k \xrightarrow{m} f$, given $j = 1, 2, \ldots$ there exists k_j such that

$$\left| \left\{ |f - f_k| > \frac{1}{j} \right\} \right| < \frac{1}{2^j}$$

for $k \geq k_j$. We may assume that $k_j \nearrow$. Let $E_j = \{|f - f_{k_j}| > 1/j\}$, and $H_m = \bigcup_{j=m}^{\infty} E_j$. Then $|E_j| < 2^{-j}, |H_m| \leq \sum_{j=m}^{\infty} 2^{-j} = 2^{-m+1}$ and $|f - f_{k_j}| \leq 1/j$ in $E - E_j$. Thus, if $j \geq m$, $|f - f_{k_j}| \leq 1/j$ in $E - H_m$, so that $f_{k_j} \to f$ in $E - H_m$. Then $f_{k_j} \to f$ in $\bigcup(E - H_m) = E - \bigcap H_m$. Since $|H_m| \to 0$, it follows that $|\bigcap H_m| = 0$ and $f_{k_j} \to f$ a.e. in E. This completes the proof.

The next theorem gives a Cauchy criterion for convergence in measure.

Theorem 4.23 *A necessary and sufficient condition that $\{f_k\}$ converge in measure on E is that for each $\varepsilon > 0$,*

$$\lim_{k,l\to\infty} |\{\mathbf{x} \in E : |f_k(\mathbf{x}) - f_l(\mathbf{x})| > \varepsilon\}| = 0.$$

Proof. The necessity follows from the formula

$$\{|f_k - f_l| > \varepsilon\} \subset \left\{|f_k - f| > \frac{1}{2}\varepsilon\right\} \cup \left\{|f_l - f| > \frac{1}{2}\varepsilon\right\}$$

and the fact that the measures of the sets on the right tend to zero as $k, l \to \infty$ if $f_k \xrightarrow{m} f$.

To prove the converse, choose $N_j, j = 1, 2, \ldots$, so that if $k, l \geq N_j$, then $|\{|f_k - f_l| > 2^{-j}\}| < 2^{-j}$. We may assume that $N_j \nearrow$. Then $|f_{N_{j+1}} - f_{N_j}| \leq 2^{-j}$ except for a set E_j, $|E_j| < 2^{-j}$. Let $H_i = \bigcup_{j=i}^{\infty} E_j$, $i = 1, 2, \ldots$. Then

$$|f_{N_{j+1}}(\mathbf{x}) - f_{N_j}(\mathbf{x})| \leq 2^{-j} \text{ for } j \geq i \text{ and } \mathbf{x} \notin H_i.$$

It follows that $\sum(f_{N_{j+1}} - f_{N_j})$ converges uniformly outside H_i for every i and, therefore, that $\{f_{Nj}\}$ converges uniformly outside every H_i. Since $|H_i| \leq \sum_{j\geq i} 2^{-j} = 2^{-i+1}$, we obtain that $\{f_{Nj}\}$ converges a.e. in E and, letting $f = \lim f_{N_j}$, that $f_{N_j} \xrightarrow{m} f$ on E. In order to show that $f_k \xrightarrow{m} f$ on E, note that

$$\{|f_k - f| > \varepsilon\} \subset \left\{|f_k - f_{N_j}| > \frac{1}{2}\varepsilon\right\} \cup \left\{|f_{N_j} - f| > \frac{1}{2}\varepsilon\right\}$$

for any N_j. To show that the measure of the set on the left is less than a prescribed $\eta > 0$ for all sufficiently large k, select N_j so that the first term on the right has measure less than $\frac{1}{2}\eta$ for all large k (here, we use the Cauchy condition) and so that the measure of the second term on the right is also less than $\frac{1}{2}\eta$. This completes the proof.

As pointed out at the beginning of the chapter, many of the results of the chapter depend on only a few basic properties of Lebesgue measurable sets and Lebesgue measure. This is especially true for the elementary properties of measurable functions (Corollary 4.2, Theorems 4.1 and 4.3 through 4.12) and the section about convergence in measure, which use only the fact that the class of measurable subsets of $\mathbf{R^n}$ is a σ-algebra and, in Theorem 4.5, that subsets of a set of measure zero are measurable. Egorov's theorem uses two additional facts: the fundamental result Theorem 3.26 concerning monotone sequences of sets and Lemma 3.22 about the approximability of measurable

sets by closed sets. Actually, even Lemma 3.22, which is a topological property of Lebesgue measure, is not needed in the proof of Egorov's theorem if instead of requiring that F be closed, we merely require that it be measurable.

The rest of the chapter, namely, the material on semicontinuous functions and Lusin's theorem, uses somewhat more restrictive topological properties of $\mathbf{R^n}$ and Lebesgue measure. For example, about $\mathbf{R^n}$, we have used metric properties, and about Lebesgue measure, we have used Lemma 3.22 (e.g., in Lemma 4.19) and the fact that Borel sets are measurable (e.g., in Corollary 4.16).

Exercises

1. Prove Corollary 4.2 and Theorem 4.8.
2. Let f be a simple function, taking its distinct values on disjoint sets $E_1, \ldots, E_N$. Show that f is measurable if and only if $E_1, \ldots, E_N$ are measurable.
3. Theorem 4.3 can be used to define measurability for vector-valued (e.g., complex-valued) functions. Suppose, for example, that f and g are real-valued and finite in $\mathbf{R^n}$, and let $F(\mathbf{x}) = (f(\mathbf{x}), g(\mathbf{x}))$. Then F is said to be measurable if $F^{-1}(G)$ is measurable for every open $G \subset \mathbf{R^2}$. Prove that F is measurable if and only if both f and g are measurable in $\mathbf{R^n}$.
4. Let f be defined and measurable in $\mathbf{R^n}$. If T is a nonsingular linear transformation of $\mathbf{R^n}$, show that $f(T\mathbf{x})$ is measurable. (If $E_1 = \{\mathbf{x} : f(\mathbf{x}) > a\}$ and $E_2 = \{\mathbf{x} : f(T\mathbf{x}) > a\}$, show that $E_2 = T^{-1}E_1$.)
5. Give an example to show that $\phi(f(\mathbf{x}))$ may not be measurable if ϕ and f are measurable and finite. (Let F be the Cantor–Lebesgue function and let f be its inverse, suitably defined. Let ϕ be the characteristic function of a set of measure zero whose image under F is not measurable.) Show that the same may be true even if f is continuous. (Let $g(x) = x + F(x)$, where F is the Cantor–Lebesgue function, and consider $f = g^{-1}$.) Cf. Exercise 22.
6. Let f and g be measurable functions on E:
   (a) If f and g are finite a.e. in E, show that $f + g$ is measurable no matter how we define it at the points when it has the form $+\infty + (-\infty)$ or $-\infty + \infty$.
   (b) Show that fg is measurable without restriction on the finiteness of f and g. Show that $f + g$ is measurable if it is defined to have the same value at every point where it has the form $+\infty + (-\infty)$ or $-\infty + \infty$. (Note that a function h defined on E is measurable if and only if both $\{h = +\infty\}$ and $\{h = -\infty\}$ are measurable and the restriction of h to the subset of E where h is finite is measurable.)

7. Let f be usc and less than $+\infty$ on a compact set E. Show that f is bounded above on E. Show also that f assumes its maximum on E, that is, that there exists $\mathbf{x}_0 \in E$ such that $f(\mathbf{x}_0) \geq f(\mathbf{x})$ for all $\mathbf{x} \in E$.

8. (a) Let f and g be two functions that are usc at $\mathbf{x}_0$. Show that $f + g$ is usc at $\mathbf{x}_0$. Is $f - g$ usc at $\mathbf{x}_0$? When is fg usc at $\mathbf{x}_0$?

(b) If $\{f_k\}$ is a sequence of functions that are usc at $\mathbf{x}_0$, show that $\inf_k f_k(\mathbf{x})$ is usc at $\mathbf{x}_0$.

(c) If $\{f_k\}$ is a sequence of functions that are usc at $\mathbf{x}_0$ and converge uniformly near $\mathbf{x}_0$, show that $\lim f_k$ is usc at $\mathbf{x}_0$.

9. (a) Show that the limit of a decreasing (increasing) sequence of functions usc (lsc) at $\mathbf{x}_0$ is usc (lsc) at $\mathbf{x}_0$. In particular, the limit of a decreasing (increasing) sequence of functions continuous at $\mathbf{x}_0$ is usc (lsc) at $\mathbf{x}_0$.

(b) Let f be usc and less than $+\infty$ on $[a, b]$. Show that there exist continuous f_k on $[a, b]$ such that $f_k \searrow f$. (First show that there are usc step functions $f_k \searrow f$.)

10. (a) If f is defined and continuous on E, show that $\{a < f < b\}$ is relatively open and that $\{a \leq f \leq b\}$ and $\{f = a\}$ are relatively closed.

(b) Let f be a finite function on $\mathbf{R^n}$. Show that f is continuous on $\mathbf{R^n}$ if and only if $f^{-1}(G)$ is open for every open G in $\mathbf{R^1}$, or if and only if $f^{-1}(F)$ is closed for every closed F in $\mathbf{R^1}$.

11. Let f be defined on $\mathbf{R^n}$ and let $B(\mathbf{x})$ denote the open ball $\{\mathbf{y} : |\mathbf{x} - \mathbf{y}| < r\}$ with center $\mathbf{x}$ and fixed radius r. Show that the function $g(\mathbf{x}) = \sup\{f(\mathbf{y}) : \mathbf{y} \in B(\mathbf{x})\}$ is lsc and that the function $h(\mathbf{x}) = \inf\{f(\mathbf{y}) : \mathbf{y} \in B(\mathbf{x})\}$ is usc on $\mathbf{R^n}$. Is the same true for the *closed* ball $\{\mathbf{y} : |\mathbf{x} - \mathbf{y}| \leq r\}$?

12. If $f(x), x \in \mathbf{R^1}$, is continuous at almost every point of an interval $[a, b]$, show that f is measurable on $[a, b]$. Generalize this to functions defined in $\mathbf{R^n}$. (For a constructive proof, use the subintervals of a sequence of partitions to define a sequence of simple measurable functions converging to f a.e. in $[a, b]$. Use Theorem 4.12. See also the proof of Theorem 5.54.)

13. One difficulty encountered in trying to extend the proof of Egorov's theorem to the continuous parameter case $f_y(\mathbf{x}) \to f(\mathbf{x})$ as $y \to y_0$ is showing that the analogues of the sets E_m in Lemma 4.18 are measurable. This difficulty can often be overcome in individual cases. Suppose, for example, that $f(x, y)$ is defined and continuous in the square $0 \leq x \leq 1, 0 < y \leq 1$ and that $f(x) = \lim_{y \to 0} f(x, y)$ exists and is finite for x in a measurable subset E of [0,1]. Show that if ε and δ satisfy $0 < \varepsilon, \delta < 1$, the set $E_{\varepsilon,\delta} = \{x \in E : |f(x, y) - f(x)| \leq \varepsilon \text{ for all } y < \delta\}$ is measurable. (If $y_k, k = 1, 2, \ldots$, is a dense subset of $(0, \delta)$, show that $E_{\varepsilon,\delta} = \bigcap_k \{x \in E : |f(x, y_k) - f(x)| \leq \varepsilon\}$.)

14. Let $f(x, y)$ be as in Exercise 13. Show that given $\varepsilon > 0$, there exists a closed $F \subset E$ with $|E - F| < \varepsilon$ such that $f(x, y)$ converges uniformly for $x \in F$ to $f(x)$ as $y \to 0$. (Follow the proof of Egorov's theorem, using the

sets $E_{\varepsilon,1/m}$ defined in Exercise 13 in place of the sets E_m in the proof of Lemma 4.18.)

15. Let $\{f_k\}$ be a sequence of measurable functions defined on a measurable E with $|E| < +\infty$. If $|f_k(\mathbf{x})| \leq M_{\mathbf{x}} < +\infty$ for all k for each $\mathbf{x} \in E$, show that given $\varepsilon > 0$, there is a closed $F \subset E$ and a finite M such that $|E - F| < \varepsilon$ and $|f_k(\mathbf{x})| \leq M$ for all k and all $\mathbf{x} \in F$.

16. Prove that $f_k \xrightarrow{m} f$ on E if and only if given $\varepsilon > 0$, there exists K such that $|\{|f - f_k| > \varepsilon\}| < \varepsilon$ if $k > K$. Give an analogous Cauchy criterion.

17. Suppose that $f_k \xrightarrow{m} f$ and $g_k \xrightarrow{m} g$ on E. Show that $f_k + g_k \xrightarrow{m} f + g$ on E and, if $|E| < +\infty$, that $f_k g_k \xrightarrow{m} fg$ on E. If, in addition, $g_k \to g$ on $E, g \neq 0$ a.e., and $|E| < +\infty$, show that $f_k/g_k \xrightarrow{m} f/g$ on E. (For the product $f_k g_k$, write $f_k g_k - fg = (f_k - f)(g_k - g) + f(g_k - g) + g(f_k - f)$. Consider each term separately, using the fact that a function that is finite on $E, |E| < +\infty$ is bounded outside a subset of E with small measure.)

18. If f is measurable on E, define $\omega_f(a) = |\{f > a\}|$ for $-\infty < a < +\infty$. If $f_k \nearrow f$, show that $\omega_{f_k} \nearrow \omega_f$. If $f_k \xrightarrow{m} f$, show that $\omega_{f_k} \to \omega_f$ at each point of continuity of ω_f. (For the second part, show that if $f_k \xrightarrow{m} f$, then $\limsup_{k\to\infty} \omega_{f_k}(a) \leq \omega_f(a - \varepsilon)$ and $\liminf_{k\to\infty} \omega_{f_k}(a) \geq \omega_f(a + \varepsilon)$ for every $\varepsilon > 0$.)

19. Let $f(x, y)$ be a function defined on the unit square $0 \leq x \leq 1, 0 \leq y \leq 1$ which is continuous in each variable separately. Show that f is a measurable function of (x, y). Is the same true if f is only assumed to be continuous in x for each y?

20. If f is measurable and finite a.e. on $[a, b]$, show that given $\varepsilon > 0$, there is a continuous g on $[a, b]$ such that $|\{x : f(x) \neq g(x)\}| < \varepsilon$. (See Exercise 18 of Chapter 1.) Formulate and prove a similar result in $\mathbf{R^n}$ by combining Lusin's theorem with the Tietze extension theorem.

21. Show that the necessity part of Lusin's theorem is not true for $\varepsilon = 0$, that is, find a measurable set E and a finite measurable function f on E such that f is not continuous relative to $E - Z$ for any Z with $|Z| = 0$. (Consider, e.g., χ_E for the set E in Exercise 25 of Chapter 3.)

22. (a) Show that if f is measurable and B is a Borel set in $\mathbf{R^1}$, then $f^{-1}(B)$ is measurable. (Recall that the Borel sets form the smallest σ-algebra that contains the open sets. Consider the collection of sets $\{E : f^{-1}(E)$ is measurable$\}$.) Cf. Exercise 30 of Chapter 3.

(b) If ϕ is a Borel measurable function on $\mathbf{R^1}$ and f is finite and measurable on $\mathbf{R^n}$, show that $\phi(f(\mathbf{x}))$ is measurable on $\mathbf{R^n}$. Cf. Theorem 4.6 and Exercise 5.

23. Let $\{f_k\}_{k=1}^{\infty}$ be a sequence of measurable functions defined on a measurable set E. Show that the sets $\{\mathbf{x} : \lim f_k(\mathbf{x})$ exists and is finite$\}$, $\{\mathbf{x} : \lim f_k(\mathbf{x}) = +\infty\}$, $\{\mathbf{x} : \lim f_k(\mathbf{x}) = -\infty\}$ are measurable.

24. Let f be a (Lebesgue) measurable function defined on a (Lebesgue) measurable set E. Show that there is a Borel set $H \subset E$ and a Borel measurable function h on H such that $|E| = |H|$ and $f = h$ on H. (This can be deduced from Lusin's theorem.) In case E is a Borel set, and consequently the exceptional set $Z = E - H$ is also a Borel set, show that f can be redefined on Z so that the resulting function is Borel measurable on E.

5

The Lebesgue Integral

5.1 Definition of the Integral of a Nonnegative Function

There are several equivalent ways to define the Lebesgue integral and develop its main properties. The approach we have chosen is based on the notion that the integral of a nonnegative f should represent the volume of the region under the graph of f.

We start then with a *nonnegative* function f, $0 \le f \le +\infty$, defined on a measurable subset E of $\mathbf{R^n}$. Let

$$\Gamma(f,E) = \left\{ \big(\mathbf{x}, f(\mathbf{x})\big) \in \mathbf{R^{n+1}} : \mathbf{x} \in E, f(\mathbf{x}) < +\infty \right\},$$

$$R(f,E) = \Big\{ \big(\mathbf{x}, y\big) \in \mathbf{R^{n+1}} : \mathbf{x} \in E, 0 \le y \le f(\mathbf{x}) \text{ if } f(\mathbf{x}) < +\infty,$$
$$\text{and} \quad 0 \le y < +\infty \quad \text{if } f(\mathbf{x}) = +\infty \Big\}.$$

$\Gamma(f,E)$ is called the *graph of f over E* and $R(f,E)$ the *region under f over E*.

If $R(f,E)$ is measurable (as a subset of $\mathbf{R^{n+1}}$), its measure $|R(f,E)|_{(n+1)}$ is called the *Lebesgue integral of f over E*, and we write

$$|R(f,E)|_{(n+1)} = \int_E f(\mathbf{x})\, d\mathbf{x}.$$

Usually, one of the abbreviations

$$\int_E f\, d\mathbf{x} \quad \text{or} \quad \int_E f$$

is used, and at times the lengthy notation

$$\int \cdots \int_E f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

is convenient. We stress that the definition applies only to nonnegative f; a definition for functions that are not nonnegative will be given in Section 5.3.

We also note that the existence of the integral is equivalent to the measurability of $R(f, E)$ and does not require the finiteness of $|R(f, E)|_{(n+1)}$. The next theorem is of basic importance.

Theorem 5.1 *Let f be a nonnegative function defined on a measurable set E. Then $\int_E f$ exists if and only if f is measurable.*

We will show here only that the integral exists if f is measurable, postponing a proof of the converse until Theorem 6.11. We need several lemmas, the first of which proves the theorem for functions that are constant on E. In this case, $R(f, E)$ is a *cylinder set*; that is, it has one of the forms $\{(\mathbf{x}, y) : \mathbf{x} \in E, 0 \le y \le a\}$, $0 \le a < +\infty$, or $\{(\mathbf{x}, y) : \mathbf{x} \in E, 0 \le y < +\infty\}$.

Lemma 5.2 *Let E be a subset of $\mathbf{R^n}$, $0 \le a \le +\infty$, and define $E_a = \{(\mathbf{x}, y) : \mathbf{x} \in E,\ 0 \le y \le a\}$ for finite a and $E_\infty = \{(\mathbf{x}, y) : \mathbf{x} \in E,\ 0 \le y < +\infty\}$. If E is measurable (as a subset of $\mathbf{R^n}$), then E_a is measurable (as a subset of $\mathbf{R^{n+1}}$) and $|E_a|_{(n+1)} = a|E|_{(n)}$.**

Proof. The result follows from a series of simple observations. First, assume that a is finite. If $|E| = 0$ or if E is an interval that is either closed, partly open, or open, the result is clear. Next, if E is an open set, then by Theorem 1.11, it can be written as a disjoint union of partly open intervals, $E = \bigcup I_k$. Therefore, $E_a = \bigcup I_{k,a}$, and since $I_{k,a}$ are measurable and disjoint, E_a is measurable and $|E_a| = \sum \left|I_{k,a}\right| = \sum a\,|I_k| = a|E|$.

Let E be of type G_δ, $E = \bigcap_{k=1}^{\infty} G_k$, with $|G_1| < +\infty$. We may assume that $G_k \searrow E$ by writing $E = G_1 \cap (G_1 \cap G_2) \cap (G_1 \cap G_2 \cap G_3) \cap \cdots$. Therefore, by Theorem 3.26, $|G_k| \to |E|$ as $k \to \infty$. Moreover, $G_{k,a}$ is measurable, $\left|G_{k,a}\right| = a\,|G_k|$, and $G_{k,a} \searrow E_a$. Therefore, E_a is measurable and $|E_a| = \lim_{k\to\infty} \left|G_{k,a}\right| = a \lim_{k\to\infty} |G_k| = a|E|$.

If E is any measurable set with $|E| < +\infty$, then by Theorem 3.28, $E = H - Z$, where $|Z| = 0$, H is a set of type G_δ, $H = \bigcap_{k=1}^{\infty} G_k$, and $|G_1| < +\infty$. Since $E_a = H_a - Z_a$, we see that E_a is measurable and $|E_a| = |H_a| = a|H| = a|E|$. Finally, if $|E| = +\infty$, the result follows by writing E as the countable union of disjoint measurable sets with finite measure. This completes the proof in case a is finite. If $a = +\infty$, choose a sequence $\{a_k\}$ of finite numbers increasing to $+\infty$. The conclusion then follows easily from the fact that $E_{a_k} \nearrow E_\infty$.

As is easily seen (e.g., by using the same proof as above), the conclusion of Lemma 5.2 holds with E_a replaced by $\{(\mathbf{x}, y) : \mathbf{x} \in E, 0 \le y < a\}$, $0 \le a < +\infty$.

* Here and in the following text, $0 \cdot \infty$ and $\infty \cdot 0$ should be interpreted as 0.

Lemma 5.3 *If f is a nonnegative measurable function on E, $0 \le |E| \le +\infty$, then $\Gamma(f,E)$ has measure zero.*

Proof. Given $\varepsilon > 0$ and $k = 0, 1, \ldots$, let $E_k = \{k\varepsilon \le f < (k+1)\varepsilon\}$. The E_k are disjoint and measurable, and their union is the subset of E where f is finite. Hence, $\Gamma(f,E) = \bigcup \Gamma\left(f, E_k\right)$. Since $\left|\Gamma(f,E_k)\right|_e \le \varepsilon\,|E_k|$ by Lemma 5.2, we obtain

$$|\Gamma(f,E)|_e \le \sum \left|\Gamma\left(f,E_k\right)\right|_e \le \varepsilon \sum |E_k| \le \varepsilon |E|.$$

If $|E| < +\infty$, this implies that $\Gamma(f,E)$ has measure zero. If $|E| = +\infty$, write E as the countable union of disjoint measurable sets with finite measure. Then $\Gamma(f,E)$ is the countable union of sets of measure zero, and the lemma follows.

Proof of the sufficiency in Theorem 5.1. Let f be nonnegative and measurable on E. We must show that $R(f,E)$ is measurable. By Theorem 4.13, there exist simple measurable $f_k \nearrow f$. Therefore, $R\left(f_k,E\right) \cup \Gamma(f,E) \nearrow R(f,E)$, and since $\Gamma(f,E)$ is measurable (with measure zero), it is enough to show that each $R\left(f_k,E\right)$ is measurable. Fix k and suppose that the distinct values of f_k are $a_1, \ldots, a_N$, taken on measurable sets $E_1, \ldots, E_N$, respectively. Then $R\left(f_k,E\right) = \bigcup_{j=1}^N E_{j,a_j}$. Therefore, $R\left(f_k,E\right)$ is measurable by Lemma 5.2, and the proof is complete.

Corollary 5.4 *If f is a nonnegative measurable function, taking constant values $a_1, a_2, \ldots$ (possibly $+\infty$) on disjoint sets $E_1, E_2, \ldots$, respectively, and if $E = \bigcup E_j$, then*

$$\int_E f = \sum_j a_j \left|E_j\right|.$$

Proof. Clearly $R(f,E) = \bigcup_j E_{j,a_j}$. Since the E_j are measurable and disjoint, so are the E_{j,a_j}. Therefore, $\int_E f = \sum \left|E_{j,a_j}\right|$, and the corollary follows from the fact that $\left|E_{j,a_j}\right| = a_j \left|E_j\right|$.

Note that Corollary 5.4 applies in particular to nonnegative simple measurable functions.

5.2 Properties of the Integral

Theorem 5.5

(i) *If f and g are measurable and if $0 \le g \le f$ on E, then $\int_E g \le \int_E f$. In particular, $\int_E (\inf f) \le \int_E f$.*

(ii) *If f is nonnegative and measurable on E and if $\int_E f$ is finite, then $f < +\infty$ a.e. in E.*

(iii) *Let E_1 and E_2 be measurable and $E_1 \subset E_2$. If f is nonnegative and measurable on E_2, then $\int_{E_1} f \le \int_{E_2} f$.*

Proof. Parts (i) and (iii) follow from the relations $R(g,E) \subset R(f,E)$ and $R\left(f,E_1\right) \subset R\left(f,E_2\right)$, respectively. To prove (ii), we may assume that $|E| > 0$. If $f = +\infty$ in a subset E_1 of E with positive measure, then by (iii) and (i), we have $\int_E f \ge \int_{E_1} f \ge \int_{E_1} a = a\,|E_1|$, no matter how large a is. This contradicts the finiteness of $\int_E f$.

Theorem 5.6 (Monotone Convergence Theorem for Nonnegative Functions) *If $\{f_k\}$ is a sequence of nonnegative measurable functions such that $f_k \nearrow f$ on E, then*

$$\int_E f_k \to \int_E f.$$

Proof. By Theorem 4.12, f is measurable. Since $R\left(f_k,E\right) \cup \Gamma(f,E) \nearrow R(f,E)$ and $\Gamma(f,E)$ has measure zero, the result follows from Theorem 3.26.

Theorem 5.7 *Suppose that f is nonnegative and measurable on E and that E is the countable union of disjoint measurable sets E_j, $E = \bigcup E_j$. Then*

$$\int_E f = \sum \int_{E_j} f.$$

Proof. The sets $R\left(f,E_j\right)$ are disjoint and measurable. Since $R(f,E) = \bigcup R\left(f,E_j\right)$, the result follows from Theorem 3.23.

The next four theorems are corollaries of the results just proved. The first one provides an alternate definition of the integral that will be useful in

Chapter 10 as a motivation for defining integration with respect to abstract measures.

Theorem 5.8 *Let f be nonnegative and measurable on E. Then*

$$\int_E f = \sup \sum_j \left[\inf_{\mathbf{x} \in E_j} f(\mathbf{x}) \right] |E_j|,$$

where the supremum is taken over all decompositions $E = \bigcup_j E_j$ of E into the union of a finite number of disjoint measurable sets E_j.

The reader will observe that the formula resembles the definition of the Riemann integral if the E_j are taken to be subintervals. We note however that the roles of sup and inf cannot be interchanged; see Exercise 25.

Proof. If $E = \bigcup_{j=1}^N E_j$ is such a decomposition, consider the measurable function g taking values $a_j = \inf_{\mathbf{y} \in E_j} f(\mathbf{y})$ on E_j, $j = 1, \ldots, N$. Since $0 \le g \le f$, we have by Corollary 5.4 and Theorem 5.5 that $\sum_{j=1}^N a_j |E_j| \le \int_E f$. Therefore,

$$\sup \sum_j \left(\inf_{E_j} f\right) |E_j| \le \int_E f.$$

To prove the opposite inequality, consider for $k = 1, 2, \ldots,$ the sets $\left\{E_j^{(k)}\right\}, j = 0, 1, \ldots, k2^k$, defined by

$$E_j^{(k)} = \left\{ \frac{j-1}{2^k} \le f < \frac{j}{2^k} \right\}, \quad j = 1, \ldots, k2^k; \quad E_0^{(k)} = \{f \ge k\},$$

and the corresponding measurable functions

$$f_k = \sum_j \left(\inf_{E_j^{(k)}} f\right) \chi_{E_j^{(k)}}.$$

(Compare the simple functions in Theorem 4.13(ii).) Then $0 \le f_k \nearrow f$, and by the monotone convergence theorem, $\int_E f_k \to \int_E f$. Since $\int_E f_k = \sum_j \left(\inf_{E_j^{(k)}} f\right) \left|E_j^{(k)}\right|$ by Corollary 5.4, it follows that

$$\sup \sum_j (\inf_{E_j} f)\, |E_j| \geq \int_E f,$$

which completes the proof.

Theorem 5.9 *Let f be nonnegative on E. If $|E| = 0$, then $\int_E f = 0$.*

This can be proved in many ways; for example, it follows immediately from the last result. Measurability of f is automatic since $|E| = 0$.

We can now slightly strengthen the statement of part (i) of Theorem 5.5.

Theorem 5.10 *If f and g are nonnegative and measurable on E and if $g \leq f$ a.e. in E, then $\int_E g \leq \int_E f$.*

In particular, if f and g are nonnegative and measurable on E and if $f = g$ a.e. in E, then $\int_E f = \int_E g$.

Proof. Write $E = A \cup Z$, where A and Z are disjoint and $Z = \{g > f\}$. Then $|Z| = 0$. Therefore, by Theorems 5.7 and 5.9, $\int_E f = \int_A f + \int_Z f = \int_A f$. Since the same is true for g, and since $f \geq g$ everywhere on A, the result follows.

In defining $\int_E f$, we assumed that f was defined everywhere in E. In view of Theorem 5.10, $\int_E f$ is unchanged if we modify f in a set of measure zero. Hence, we may consider integrals $\int_E f$ where f is defined only a.e. in E, by completing the definition of f arbitrarily in the set Z of measure zero where it is undefined. As a result of Theorem 5.9, this amounts to defining $\int_E f$ to be $\int_{E-Z} f$. Similarly, we may extend the definition of the integral and the results earlier to measurable functions that are nonnegative only a.e. in E.

Theorem 5.11 *Let f be nonnegative and measurable on E. Then $\int_E f = 0$ if and only if $f = 0$ a.e. in E.*

Proof. If $f = 0$ a.e. in E, then $\int_E f = 0$ by Theorem 5.10. Conversely, suppose that f is nonnegative and measurable on E and that $\int_E f = 0$. For $\alpha > 0$, we have by Corollary 5.4 and Theorem 5.5 that

$$\alpha |\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}| \leq \int_{\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}} f \leq \int_E f = 0.$$

Therefore, $\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}$ has measure zero for every $\alpha > 0$. Since the set where $f > 0$ is the union of those where $f > 1/k$, it follows that $f = 0$ a.e. in E.

This proof also establishes the following useful result.

Corollary 5.12 ***(Tchebyshev's Inequality)*** *Let f be nonnegative and measurable on E. If $\alpha > 0$, then*

$$|\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}| \leq \frac{1}{\alpha} \int_E f.$$

The significance of Tchebyshev's inequality is that it estimates the size of f in terms of the integral of f.

The next two theorems establish the linear properties of the integral for nonnegative functions.

Theorem 5.13 *If f is nonnegative and measurable, and if c is any nonnegative constant, then*

$$\int_E cf = c \int_E f.$$

Proof. If f is simple, then so is cf, and the theorem follows in this case from the formula for integrating simple functions (see Corollary 5.4). For arbitrary measurable $f \geq 0$, choose simple measurable f_k with $0 \leq f_k \nearrow f$. Then $0 \leq cf_k \nearrow cf$ and

$$\int_E cf = \lim_{k\to\infty} \int_E cf_k = \lim_{k\to\infty} c \int_E f_k = c \int_E f.$$

Theorem 5.14 *If f and g are nonnegative and measurable, then*

$$\int_E (f+g) = \int_E f + \int_E g.$$

Proof. First, suppose that f and g are simple: $f = \sum_{i=1}^N a_i \chi_{A_i}$ and $g = \sum_{j=1}^M b_j \chi_{B_j}$, where $E = \bigcup_i A_i = \bigcup_j B_j$ and all A_i and B_j are measurable. Then, $f+g$ is also simple, taking values $a_i + b_j$ on $A_i \cap B_j$: $f+g = \sum_{i,j} (a_i + b_j)\, \chi_{A_i \cap B_j}$. Thus,

$$\begin{aligned}\int_E (f+g) &= \sum_{i,j} (a_i + b_j)\, |A_i \cap B_j| = \sum_i a_i \sum_j |A_i \cap B_j| + \sum_j b_j \sum_i |A_i \cap B_j| \\ &= \sum a_i\, |A_i| + \sum b_j\, |B_j| = \int_E f + \int_E g.\end{aligned}$$

For general nonnegative measurable f and g, choose simple measurable f_k and g_k such that $0 \le f_k \nearrow f$ and $0 \le g_k \nearrow g$. Then $f_k + g_k$ is simple and $0 \le f_k + g_k \nearrow f + g$. Therefore,

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

which completes the proof.

Corollary 5.15 *Suppose that f and ϕ are measurable on E, $0 \le f \le \phi$, and $\int_E f$ is finite. Then*

$$\int_E (\phi - f) = \int_E \phi - \int_E f.$$

Proof. By Theorem 5.14, we have $\int_E f + \int_E (\phi - f) = \int_E \phi$. Since $\int_E f$ is finite, the result follows by subtraction.

Theorem 5.16 *If f_k, $k = 1, 2, \ldots$, are nonnegative and measurable, then*

$$\int_E \left(\sum_{k=1}^{\infty} f_k \right) = \sum_{k=1}^{\infty} \int_E f_k.$$

Proof. The functions F_N defined by $F_N = \sum_{k=1}^{N} f_k$ are nonnegative and measurable and increase to $\sum_{k=1}^{\infty} f_k$. Hence,

$$\int_E \left(\sum_{k=1}^{\infty} f_k \right) = \lim_{N\to\infty} \int_E F_N = \lim_{N\to\infty} \sum_{k=1}^{N} \int_E f_k = \sum_{k=1}^{\infty} \int_E f_k.$$

Note that the preceding theorem is essentially a corollary of the Monotone Convergence Theorem 5.6. Conversely, Theorem 5.6 can be deduced from this result. Verification is left to the reader.

The monotone convergence theorem gives a sufficient condition for interchanging the operations of integration and passage to the limit: $\int_E \lim f_k = \lim \int_E f_k$. It is an important problem to find other conditions under which this is true. First, we show that some restriction other than the mere convergence of f_k to f is necessary. Let E be the interval $[0, 1]$, and for $k = 1, 2, \ldots$, let f_k be defined as follows: when $0 \le x \le 1/k$, the graph of f_k consists of the sides of the isosceles triangle with altitude k and base $[0, 1/k]$; when $1/k \le x \le 1$, $f_k(x) = 0$.

Clearly, $f_k \to 0$ on $[0,1]$, but $\int_0^1 f_k = \frac{1}{2}(1/k)(k) = \frac{1}{2}$ for all k. Hence, $\int_0^1 \lim f_k < \lim \int_0^1 f_k$. For any nonnegative measurable $\{f_k\}$ such that $f_k \to f$ and $\int f_k$ converges, the fact that $\int f \leq \lim \int f_k$ is a consequence of the next theorem.

Theorem 5.17 (Fatou's Lemma) *If $\{f_k\}$ is a sequence of nonnegative measurable functions on E, then*

$$\int_E \left(\liminf_{k\to\infty} f_k\right) \leq \liminf_{k\to\infty} \int_E f_k.$$

Proof. First, note that the integral on the left exists since its integrand is nonnegative and measurable. Next, let $g_k = \inf\{f_k, f_{k+1}, \ldots\}$ for each k. Then, $g_k \nearrow \liminf f_k$ and $0 \leq g_k \leq f_k$. Therefore, by Theorems 5.6 and 5.10,

$$\int_E g_k \to \int_E (\liminf f_k), \qquad \int_E g_k \leq \int_E f_k,$$

so that

$$\int_E (\liminf f_k) = \lim \int_E g_k \leq \liminf \int_E f_k.$$

Corollary 5.18 *Let $f_k, k = 1, 2, \ldots$, be nonnegative and measurable on E, and let $f_k \to f$ a.e. in E. If $\int_E f_k \leq M$ for all k, then $\int_E f \leq M$.*

Proof. By Fatou's lemma, $\int_E \left(\liminf f_k\right) \leq M$. Since $\liminf f_k = \lim f_k = f$ a.e. in E, the conclusion follows.

We now prove a basic result about term-by-term integration of convergent sequences.

Theorem 5.19 (Lebesgue's Dominated Convergence Theorem for Nonnegative Functions) *Let $\{f_k\}$ be a sequence of nonnegative measurable functions on E such that $f_k \to f$ a.e. in E. If there exists a measurable function ϕ such that $f_k \le \phi$ a.e. for all k and if $\int_E \phi$ is finite, then*

$$\int_E f_k \to \int_E f.$$

Proof. By Fatou's lemma,

$$\int_E f = \int_E \liminf f_k \le \liminf \int_E f_k,$$

and the theorem will follow if we show that

$$\int_E f \ge \limsup \int_E f_k.$$

To prove this inequality, apply Fatou's lemma to the nonnegative functions $\phi - f_k$, obtaining

$$\int_E \liminf (\phi - f_k) \le \liminf \int_E (\phi - f_k).$$

Since $f_k \to f$ a.e., the integrand on the left equals $\phi - f$ a.e., so that the integral on the left is $\int_E \phi - \int_E f$ by Corollary 5.15. The right-hand side equals

$$\liminf \left(\int_E \phi - \int_E f_k \right) = \int_E \phi - \limsup \int_E f_k.$$

Combining formulas and cancelling $\int_E \phi$, we obtain the inequality $\int_E f \ge \limsup \int_E f_k$, as desired.

5.3 The Integral of an Arbitrary Measurable f

Let f be any measurable function defined on a set E. Then $f = f^+ - f^-$ and, by the comments following Theorem 4.11, f^+ and f^- are measurable. Therefore, the integrals $\int_E f^+(\mathbf{x})\,d\mathbf{x}$ and $\int_E f^-(\mathbf{x})\,d\mathbf{x}$ exist and are nonnegative, possibly having value $+\infty$. *Provided at least one of these integrals is finite*, we define

$$\int_E f(\mathbf{x})\,d\mathbf{x} = \int_E f^+(\mathbf{x})\,d\mathbf{x} - \int_E f^-(\mathbf{x})\,d\mathbf{x}$$

and say that the integral $\int_E f(\mathbf{x})\,d\mathbf{x}$ *exists*. If $f \geq 0$, then $f = f^+$, and this definition agrees with the previous one. As in the case when $f \geq 0$, we will use the abbreviations $\int_E f\,d\mathbf{x}$ and $\int_E f$.

The definition clearly applies if f is defined only a.e. in E, as in the case when $f \geq 0$ (see p. 86 in Section 5.2). For the sake of simplicity, we shall usually assume that f is defined everywhere in E.

If $\int_E f$ exists then, of course, $-\infty \leq \int_E f \leq +\infty$. If $\int_E f$ *exists and is finite*, we say that f is *Lebesgue integrable*, or simply *integrable*, on E and write $f \in L(E)$. Thus,

$$L(E) = \left\{ f : \int_E f \text{ is finite} \right\}.$$

If $\int_E f$ exists, then

$$\left| \int_E f \right| \leq \int_E f^+ + \int_E f^- = \int_E \left(f^+ + f^-\right),$$

by Theorem 5.14. Since $f^+ + f^- = |f|$, we obtain the inequality

$$\left| \int_E f\,d\mathbf{x} \right| \leq \int_E |f|\,d\mathbf{x}. \tag{5.20}$$

Theorem 5.21 *Let f be measurable on E. Then f is integrable over E if and only if $|f|$ is.*

Proof. If $|f| \in L(E)$ then $f^+, f^- \in L(E)$, and consequently $\int_E f$ exists and is finite. If $f \in L(E)$, then the difference $\int_E f^+ - \int_E f^-$ is finite, and therefore, since at least one of $\int_E f^+$ or $\int_E f^-$ is finite, both must be finite. Hence, their sum is finite. Since this sum is $\int_E \left(f^+ + f^-\right) = \int_E |f|$, it follows that $|f| \in L(E)$.

The simple properties of $\int_E f$ for general f follow from the results already established for $f \geq 0$. As a first example, we have the following theorem.

Theorem 5.22 *If $f \in L(E)$, then f is finite a.e. in E.*

Proof. If $f \in L(E)$, then $|f| \in L(E)$, and the result follows from Theorem 5.5(ii).

Theorem 5.23

(i) *If both* $\int_E f$ *and* $\int_E g$ *exist and if* $f \leq g$ *a.e. in* E*, then* $\int_E f \leq \int_E g$*. Also, if* f *and* g *are functions with* $f = g$ *a.e. in* E *and if* $\int_E f$ *exists, then* $\int_E g$ *exists and* $\int_E f = \int_E g$*.*

(ii) *If* $\int_{E_2} f$ *exists and* E_1 *is a measurable subset of* E_2*, then* $\int_{E_1} f$ *exists.*

Proof. (i) The fact that $f \leq g$ a.e. implies that $0 \leq f^+ \leq g^+$ and $0 \leq g^- \leq f^-$ a.e. in E. By Theorem 5.10, we then have $\int_E f^+ \leq \int_E g^+$ and $\int_E f^- \geq \int_E g^-$, and the first part of (i) follows by subtracting these inequalities. The proof of the second part of (i) is similar; note that measurability of f is equivalent to that of g when $f = g$ a.e.

(ii) If $\int_{E_2} f$ exists, at least one of $\int_{E_2} f^+$ or $\int_{E_2} f^-$ is finite. If $E_1 \subset E_2$, then by Theorem 5.5(iii), at least one of $\int_{E_1} f^+$ or $\int_{E_1} f^-$ is finite. Therefore, $\int_{E_1} f$ exists.

Theorem 5.24 *If* $\int_E f$ *exists and* $E = \bigcup_k E_k$ *is the countable union of disjoint measurable sets* E_k*, then*

$$\int_E f = \sum_k \int_{E_k} f.$$

Proof. Each $\int_{E_k} f$ exists by Theorem 5.23(ii). We have

$$\int_E f = \int_E f^+ - \int_E f^- = \sum \int_{E_k} f^+ - \sum \int_{E_k} f^-$$

by Theorem 5.7. Since at least one of these sums is finite, we obtain

$$\int_E f = \sum \left(\int_{E_k} f^+ - \int_{E_k} f^- \right) = \sum \int_{E_k} f.$$

Theorem 5.25 *If* $|E| = 0$ *or if* $f = 0$ *a.e. in* E*, then* $\int_E f = 0$*.*

Proof. The theorem follows by applying Theorem 5.9 or 5.11 to f^+ and f^-.

The next few results deal with linearity properties of the integral.

Lemma 5.26 *If* $\int_E f$ *is defined, then so is* $\int_E (-f)$*, and* $\int_E (-f) = -\int_E f$*.*

Proof. Since $(-f)^+ = f^-$ and $(-f)^- = f^+$, and at least one of $\int_E f^-$ or $\int_E f^+$ is finite, we have $\int_E(-f) = \int_E f^- - \int_E f^+ = -\int_E f$.

Theorem 5.27 *If $\int_E f$ exists and c is any real constant, then $\int_E(cf)$ exists and*

$$\int_E (cf) = c\int_E f.$$

Proof. If $c \geq 0$, $(cf)^+ = cf^+$ and $(cf)^- = cf^-$. Therefore, by Theorem 5.13, $\int_E(cf)^+ = c\int_E f^+$ and $\int_E(cf)^- = c\int_E f^-$. It follows that $\int_E(cf)$ exists and $\int_E(cf) = c\left(\int_E f^+ - \int_E f^-\right) = c\int_E f$. If $c = -1$, the theorem reduces to Lemma 5.26. For any $c \leq 0$, we have $c = (-1)(|c|)$, and the result follows from the cases $c = -1$ and $c \geq 0$.

Theorem 5.28 *If $f, g \in L(E)$, then $f + g \in L(E)$ and*

$$\int_E (f+g) = \int_E f + \int_E g.$$

Proof. Since $|f+g| \leq |f| + |g|$, we have from Theorems 5.23(i) and 5.14 that

$$\int_E |f+g| \leq \int_E (|f| + |g|) = \int_E |f| + \int_E |g| < +\infty.$$

Hence, $f + g \in L(E)$.

To prove the rest of the theorem, first note that if $g = 0$ everywhere in E, then the formula $\int_E(f+g) = \int_E f + \int_E g$ is obvious by Theorem 5.25. Thus, by writing $E = (E \cap \{g = 0\}) \cup (E \cap \{g \neq 0\})$ and using Theorem 5.24 for each of the three integrals in the formula, it is enough to prove the formula with E replaced by $E \cap \{g \neq 0\}$. Hence, because of similar considerations for f, it suffices to prove the formula under the extra assumption that f and g never vanish on E. To do so, we begin by considering the following six cases, in which each inequality is assumed to hold everywhere in E: (1) $f > 0$, $g > 0$ (so that $f + g > 0$); (2) $f > 0$, $g < 0$, $f + g \geq 0$; (3) $f > 0$, $g < 0$, $f + g < 0$; (4) $f < 0$, $g > 0$, $f + g \geq 0$; (5) $f < 0$, $g > 0$, $f + g < 0$; (6) $f < 0$, $g < 0$ (so that $f + g < 0$). Note that these possibilities are mutually exclusive. The result in case 1 is just Theorem 5.14. Cases 2–6 are all similar, and we shall consider only case 2. Then $f > 0$, $-g > 0$, $f + g \geq 0$, and since by Theorem 5.22 each function is finite a.e., $(f+g) + (-g) = f$ a.e. Hence, by Theorems 5.23(i) and 5.14, we have $\int_E(f+g) + \int_E(-g) = \int_E f$. The result in case 2 now follows from Lemma 5.26 and the fact that all the integrals involved are finite.

For arbitrary f and g in $L(E)$ that never vanish in E, we subdivide E into at most six measurable sets, $E_1, \ldots, E_6$, where possibilities (1), ..., (6) hold, respectively. Since E_i and E_j are disjoint for $i \neq j$, we have

$$\int_E (f+g) = \sum_{j=1}^{6} \int_{E_j} (f+g) = \sum_{j=1}^{6} \left(\int_{E_j} f + \int_{E_j} g \right) = \int_E f + \int_E g.$$

This completes the proof.

It follows that if $f_k \in L(E), k = 1, \ldots, N$, and if a_k are real constants, then $\sum_{k=1}^{N} a_k f_k \in L(E)$ and

$$\int_E \left(\sum_{k=1}^{N} a_k f_k \right) = \sum_{k=1}^{N} a_k \int_E f_k.$$

Corollary 5.29 *Let f and ϕ be measurable on E, $f \geq \phi$ a.e., and $\phi \in L(E)$. Then,*

$$\int_E (f - \phi) = \int_E f - \int_E \phi.$$

Proof. First, note that $\int_E f$ exists since $f^- \leq \phi^-$ a.e. implies that $\int_E f^-$ is finite. Next, $\int_E (f - \phi)$ exists since $f - \phi \geq 0$ a.e. If $f \in L(E)$, the corollary follows from Theorem 5.28. If $f \notin L(E)$, then since $f^- \in L(E)$, we must have $\int_E f = +\infty$. The fact that $\phi \in L(E)$ implies that $f - \phi \notin L(E)$, so that $\int_E (f - \phi) = +\infty$ since $f - \phi \geq 0$ a.e. This completes the proof.

In Chapter 8, we will study conditions on f and g that imply that $fg \in L(E)$. For now, we have the following simple result.

Theorem 5.30 *If $f \in L(E)$, g is measurable on E, and there exists a finite constant M such that $|g| \leq M$ a.e. in E, then $fg \in L(E)$ and $\int_E |fg| \leq M \int_E |f|$.*

Proof. Since $|fg| \leq M|f|$ a.e. in E, we have by Theorems 5.10 and 5.27 that $\int_E |fg| \leq \int_E M|f| = M \int_E |f|$. Hence, $fg \in L(E)$.

Corollary 5.31 *If $f \in L(E)$, $f \geq 0$ a.e., and there exist finite constants α and β such that $\alpha \leq g \leq \beta$ a.e. in E, then*

$$\alpha \int_E f \leq \int_E fg \leq \beta \int_E f.$$

Proof. By Theorem 5.30, $fg \in L(E)$. Since $f \geq 0$ a.e., we have $\alpha f \leq fg \leq \beta f$ a.e. in E, and the conclusion follows by integrating.

We now study conditions that imply that $\int_E f_k \to \int_E f$ if $f_k \to f$ in E. Most of the results are extensions of those we derived for nonnegative functions.

Theorem 5.32 (Monotone Convergence Theorem) *Let $\{f_k\}$ be a sequence of measurable functions on E:*

(i) *If $f_k \nearrow f$ a.e. on E and there exists $\phi \in L(E)$ such that $f_k \geq \phi$ a.e. on E for all k, then $\int_E f_k \to \int_E f$.*
(ii) *If $f_k \searrow f$ a.e. on E and there exists $\phi \in L(E)$ such that $f_k \leq \phi$ a.e. on E for all k, then $\int_E f_k \to \int_E f$.*

Proof. To prove (i), we may assume by Theorem 5.25 that $f_k \nearrow f$ and $f_k \geq \phi$ everywhere on E. Then $0 \leq f_k - \phi \nearrow f - \phi$ on E, so that by Theorem 5.6, $\int_E (f_k - \phi) \to \int_E (f - \phi)$. Therefore, by Corollary 5.29, $\int_E f_k - \int_E \phi \to \int_E f - \int_E \phi$, and since $\phi \in L(E)$, the result follows.

We can deduce (ii) from (i) by considering the functions $-f_k$. Details are left to the reader.

Theorem 5.33 (Uniform Convergence Theorem) *Let $f_k \in L(E)$ for $k = 1, 2, \ldots$, and let $\{f_k\}$ converge uniformly to f on E, $|E| < +\infty$. Then $f \in L(E)$ and $\int_E f_k \to \int_E f$.*

Proof. Since $|f| \leq |f_k| + |f - f_k|$ and $\{f_k\}$ converge uniformly to f on E, we have $|f| \leq |f_k| + 1$ on E if k is sufficiently large. Since $|E| < +\infty$, it follows that $f \in L(E)$. From Theorem 5.28 and (5.20), we obtain

$$\left| \int_E f - \int_E f_k \right| = \left| \int_E (f - f_k) \right| \leq \int_E |f - f_k| .$$

The last integral is bounded by $\left(\sup_{\mathbf{x} \in E} |f(\mathbf{x}) - f_k(\mathbf{x})|\right) |E|$, which by hypothesis tends to 0 as $k \to \infty$. This proves the theorem.

Theorem 5.34 (Fatou's Lemma) *Let $\{f_k\}$ be a sequence of measurable functions on E. If there exists $\phi \in L(E)$ such that $f_k \geq \phi$ a.e. on E for all k, then*

$$\int_E \left(\liminf_{k \to \infty} f_k\right) \leq \liminf_{k - \infty} \int_E f_k.$$

Proof. The result follows by first applying Theorem 5.17 to the sequence $\{f_k - \phi\}$ of nonnegative functions, and then using Corollary 5.29. Details are left to the reader.

Corollary 5.35 *Let $\{f_k\}$ be a sequence of measurable functions on E. If there exists $\phi \in L(E)$ such that $f_k \leq \phi$ a.e. on E for all k, then*

$$\int_E \left(\limsup_{k\to\infty} f_k\right) \geq \limsup_{k-\infty} \int_E f_k.$$

Proof. This follows from Fatou's lemma since $-f_k \geq -\phi$ a.e. and $\liminf\left(-f_k\right) = -\limsup\left(f_k\right)$.

Theorem 5.36 (Lebesgue's Dominated Convergence Theorem) *Let $\{f_k\}$ be a sequence of measurable functions on E such that $f_k \to f$ a.e. in E. If there exists $\phi \in L(E)$ such that $|f_k| \leq \phi$ a.e. in E for all k, then $\int_E f_k \to \int_E f$.*

Proof. By hypothesis, $-\phi \leq f_k \leq \phi$ a.e. in E. Therefore, $0 \leq f_k + \phi \leq 2\phi$ a.e. in E. Since $2\phi \in L(E)$, we conclude from Theorem 5.19 that $\int_E \left(f_k + \phi\right) \to \int_E (f + \phi)$. Since ϕ, f, and all the f_k are integrable on E, the result follows from Theorem 5.28.

See Exercises 23 and 26 for two useful variants of Theorem 5.36, one about weakening the assumption that $|f_k| \leq \phi$ and the other about replacing the hypothesis of pointwise convergence of $\{f_k\}$ by convergence in measure.

The following special case of the dominated convergence theorem is often useful if E has finite measure.

Corollary 5.37 (Bounded Convergence Theorem) *Let $\{f_k\}$ be a sequence of measurable functions on E such that $f_k \to f$ a.e. in E. If $|E| < +\infty$ and there is a finite constant M such that $|f_k| \leq M$ a.e. in E, then $\int_E f_k \to \int_E f$.*

In later chapters, we will consider the integrals of complex-valued functions. Here, we mention only the definition. (See p. 183 in Section 8.1 for some further remarks.) If $f = f_1 + if_2$ with f_1 and f_2 real-valued, we define

$$\int_E f = \int_E f_1 + i \int_E f_2,$$

provided the integrals on the right exist and are finite. (For the measurability of such f, see Exercise 3 Chapter 4.) Many basic properties of the ordinary integral are valid in this case.

5.4 Relation between Riemann–Stieltjes and Lebesgue Integrals, and the L^p Spaces, $0 < p < \infty$

It turns out that there is a remarkably simple and useful representation of Lebesgue integrals over subsets of $\mathbf{R^n}$ in terms of Riemann–Stieltjes integrals (over subsets of $\mathbf{R^1}$, of course). In order to establish this relation, we must first study the function

$$\omega(\alpha) = \omega_{f,E}(\alpha) = |\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}|,$$

where f is a measurable function on E and $-\infty < \alpha < +\infty$. We call $\omega_{f,E}$ *the distribution function of f on E.*

Some properties of ω were given in Exercise 18 of Chapter 4. Clearly, it is not affected by changing f in a set of measure zero, and it is decreasing. As $\alpha \nearrow +\infty$,

$$\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\} \searrow \{\mathbf{x} \in E : f(\mathbf{x}) = +\infty\};$$

hence, assuming that f is finite a.e. in E, by Theorem 3.26(ii),

$$\lim_{\alpha \to +\infty} \omega(\alpha) = 0,$$

unless $\omega(\alpha) \equiv +\infty$. Similarly,

$$\lim_{\alpha \to -\infty} \omega(\alpha) = |E|.$$

We will assume from now on that $|E| < +\infty$. This insures that ω is bounded, that $\lim_{\alpha \to +\infty} \omega(\alpha) = 0$, and that ω is of bounded variation on $(-\infty, +\infty)$ with variation equal to $|E|$. The assumption is made only to simplify the properties of ω, and is not entirely necessary (see, e.g., Exercise 16); in fact, the case $|E| = +\infty$ is often important.

In the following results, we assume that f is a measurable function that is finite a.e. in E, $|E| < +\infty$, and we write

$$\omega(\alpha) = \omega_{f,E}(\alpha), \qquad \{f > a\} = \{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}, \text{ etc.}$$

Lemma 5.38 *If $\alpha < \beta$, then $|\{\alpha < f \le \beta\}| = \omega(\alpha) - \omega(\beta)$.*

Proof. We have $\{f > \beta\} \subset \{f > \alpha\}$ and $\{a < f \leq \beta\} = \{f > \alpha\} - \{f > \beta\}$. Since $|\{f > \beta\}| < +\infty$, the lemma follows from Corollary 3.25.

Given α, let

$$\omega(\alpha+) = \lim_{\varepsilon \searrow 0} \omega(\alpha + \varepsilon), \qquad \omega(\alpha-) = \lim_{\varepsilon \searrow 0} \omega(\alpha - \varepsilon)$$

denote the limits of ω from the right and left at α.

Lemma 5.39

(a) $\omega(\alpha+) = \omega(\alpha)$; *that is, ω is continuous from the right.*
(b) $\omega(\alpha-) = |\{f \geq \alpha\}|$.

Proof. If $\varepsilon_k \searrow 0$, then $\{f > \alpha + \varepsilon_k\} \nearrow \{f > \alpha\}$ and $\{f > \alpha - \varepsilon_k\} \searrow \{f \geq \alpha\}$. Since these sets have finite measures, it follows from Theorem 3.26 that $\omega(\alpha + \varepsilon_k) \to \omega(\alpha)$ and $\omega(\alpha - \varepsilon_k) \to |\{f \geq \alpha\}|$. This completes the proof.

We now know that ω is a decreasing function that is continuous from the right. It may have jump discontinuities, with jumps $\omega(\alpha-) - \omega(\alpha)$, and intervals of constancy. These possibilities are characterized by the behavior of f stated in the following result.

Corollary 5.40

(a) $\omega(\alpha-) - \omega(\alpha) = |\{f = \alpha\}|$; *in particular, ω is continuous at α if and only if* $|\{f = \alpha\}| = 0$.
(b) ω *is constant in an open interval* (α, β) *if and only if* $|\{\alpha < f < \beta\}| = 0$, *that is, if and only if f takes almost no values between α and β.*

Proof. Since $|\{f \geq \alpha\}| = |\{f > \alpha\}| + |\{f = \alpha\}|$, part (a) follows immediately from Lemma 5.39(b). To prove part (b), note that $|\{\alpha < f < \beta\}| = |\{f > \alpha\}| - |\{f \geq \beta\}| = \omega(\alpha) - \omega(\beta-)$. This is zero if and only if ω is constant in the half-open interval $[\alpha, \beta)$. However, since ω is continuous from the right, it is constant in (α, β) if and only if it is constant in $[\alpha, \beta)$.

The rest of the theorems in this section give relations between Lebesgue and Riemann–Stieltjes integrals. As always, f is measurable and finite a.e. in E, $|E| < +\infty$ and $\omega = \omega_{f,E}$.

Theorem 5.41 *If $a < f(\mathbf{x}) \leq b$ (a and b finite) for all $\mathbf{x} \in E$, then*

$$\int_E f = -\int_a^b \alpha \, d\omega(\alpha).$$

Proof. The Lebesgue integral on the left exists and is finite since f is bounded and $|E| < +\infty$. The Riemann–Stieltjes integral on the right exists by Theorem 2.24. To show that they are equal, partition the interval $[a, b]$ by $a = \alpha_0 < \alpha_1 < \cdots < \alpha_k = b$ and let $E_j = \left\{\alpha_{j-1} < f \leq \alpha_j\right\}$. The E_j are disjoint and $E = \bigcup_{j=1}^k E_j$. Hence, $\int_E f = \sum_{j=1}^k \int_{Ej} f$ and, therefore,

$$\sum_{j=1}^k \alpha_{j-1} \left|E_j\right| \leq \int_E f \leq \sum_{j=1}^k \alpha_j \left|E_j\right|.$$

By Lemma 5.38, $\left|E_j\right| = \omega\left(\alpha_{j-1}\right) - \omega\left(\alpha_j\right) = -\left[\omega\left(\alpha_j\right) - \omega\left(\alpha_{j-1}\right)\right]$. Hence, these sums are Riemann–Stieltjes sums for $-\int_a^b \alpha \, d\,\omega(\alpha)$. Since these sums must converge to $-\int_a^b \alpha \, d\,\omega(\alpha)$ as the norm of the partition tends to zero, the conclusion follows.

Theorem 5.41 can be extended to the case when f is not bounded on E as follows.

Theorem 5.42 *Let f be any measurable function on E, and let $E_{ab} = \{\mathbf{x} \in E : a < f(\mathbf{x}) \leq b\}$ (a and b finite). Then,*

$$\int_{E_{ab}} f = -\int_a^b \alpha \, d\omega(\alpha).$$

Proof. Let $\omega_{ab}(\alpha) = \left|\{\mathbf{x} \in E_{ab} : f(\mathbf{x}) > \alpha\}\right|$. Then ω_{ab} is the distribution function of f on E_{ab}. By Theorem 5.41, we have $\int_{E_{ab}} f = -\int_a^b \alpha \, d\omega_{ab}(\alpha)$. We claim that the last expression equals $-\int_a^b \alpha \, d\omega(\alpha)$. By taking limits of Riemann–Stieltjes sums that approximate the integrals, we only need to show that $\omega_{ab}(\alpha) - \omega_{ab}(\beta) = \omega(\alpha) - \omega(\beta)$ for $a \leq \alpha < \beta \leq b$. By Lemma 5.38, this is equivalent to showing that $\left|\left\{\mathbf{x} \in E_{ab} : \alpha < f(\mathbf{x}) \leq \beta\right\}\right| = |\{\mathbf{x} \in E : \alpha < f(\mathbf{x}) \leq \beta\}|$ for such α and β. However, by the definition of E_{ab} and the restrictions on α and β, $\left\{\mathbf{x} \in E_{ab} : \alpha < f(\mathbf{x}) \leq \beta\right\} = \{\mathbf{x} \in E : \alpha < f(\mathbf{x}) \leq \beta\}$. This proves the claim and the theorem too.

In both Theorems 5.41 and 5.42, the integrals of f are extended over sets where f is bounded. This restriction is removed in the next theorem, where we define (see p. 34 in Section 2.4)

$$\int_{-\infty}^{+\infty} \alpha\, d\omega(\alpha) = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b \alpha\, d\omega(\alpha),$$

if the limit exists.

Theorem 5.43 *If either $\int_E f$ or $\int_{-\infty}^{+\infty} \alpha\, d\omega(\alpha)$ exists and is finite, then the other exists and is finite, and*

$$\int_E f = -\int_{-\infty}^{+\infty} \alpha\, d\omega(\alpha).$$

Proof. By Theorem 5.42, $\int_{E_{ab}} f = -\int_a^b \alpha\, d\omega(\alpha)$. If $f \in L(E)$, then $\int_{E_{ab}} f \to \int_E f$ as $a \to -\infty$, $b \to +\infty$ since this holds for both f^+ and f^-. Therefore, $\lim_{a\to-\infty,\, b\to+\infty} \left[-\int_a^b \alpha\, d\omega(\alpha)\right]$ exists and equals $\int_E f$, which proves half of the theorem.

Now suppose that $\int_{-\infty}^{+\infty} \alpha\, d\omega(\alpha)$ exists and is finite. Then $\int_0^\infty \alpha\, d\omega(\alpha)$ is finite, and we claim that $\int_E f^+ = -\int_0^\infty \alpha\, d\omega(\alpha)$. By Theorem 5.42, for $b > 0$, $\int_{E_{0b}} f = -\int_0^b \alpha\, d\omega(\alpha)$. Therefore, as $b \to +\infty$, $\int_{E_{0b}} f \to -\int_0^\infty \alpha\, d\omega(\alpha)$. On the other hand, as b increases to $+\infty$, $E_{0b} \nearrow \{0 < f < +\infty\}$. Therefore,

$$\int_{E_{0b}} f = \int_{E_{0b}} f^+ \to \int_{\{0<f<+\infty\}} f^+ = \int_E f^+,$$

due to our standing assumption that f is finite a.e. in E, and the claim follows. A similar argument, using the sets E_{a0} with $a \to -\infty$, shows that $\int_E f^- = \int_{-\infty}^0 \alpha\, d\omega(\alpha)$. Since all the integrals are finite, it follows that $\int_E f = \int_E f^+ - \int_E f^- = -\int_{-\infty}^{+\infty} \alpha\, d\omega(\alpha)$.

Two measurable functions f and g defined on E are said to be *equimeasurable*, or *equidistributed*, if

$$\omega_{f,E}(\alpha) = \omega_{g,E}(\alpha) \text{ for all } \alpha.$$

In case $n = 1$, simple examples of equimeasurable functions are x and $-x$ on $[-1, 1]$, or x and $1-x$ on $[0, 1]$. If $n > 1$ and f is measurable on $\mathbf{R^n}$, then $f(x)$ and

$f(Tx)$ are equimeasurable on $\{\mathbf{x} : |\mathbf{x}| < 1\}$ for any orthogonal transformation T of $\mathbf{R^n}$ (recall Theorem 3.35). See also Exercise 27.

We may intuitively think of two equimeasurable functions as being *rearrangements* of each other. For such functions, we have

$$|\{a < f \leq b\}| = |\{a < g \leq b\}|, \ |\{f = a\}| = |\{g = a\}|, \text{ etc.}$$

We also have the following immediate corollary of Theorem 5.43.

Corollary 5.44 *If f and g are equimeasurable on E and $f \in L(E)$, then $g \in L(E)$ and*

$$\int_E f = \int_E g.$$

The method used to derive Theorems 5.41 through 5.43 illustrates a basic difference between Riemann and Lebesgue integrals. The Riemann integral is defined by a limiting process whose initial step involves partitioning the *domain* of f. On the other hand, we saw in the proof of Theorem 5.41 that the Lebesgue integral can be obtained from a process that partitions the *range* of f. In order to define this process more clearly, let f be a nonnegative measurable function that is finite a.e. in E, $|E| < +\infty$. Let $\Gamma = \{0 = \alpha_0 < \alpha_1 < \cdots\}$ be a partition of the positive ordinate axis by a countable number of points $\alpha_k \to +\infty$, and let $|\Gamma| = \sup_k (\alpha_{k+1} - \alpha_k)$. Let $E_k = \{\alpha_k \leq f < \alpha_{k+1}\}$ and $Z = \{f = +\infty\}$. Then the E_k are measurable and disjoint, $|Z| = 0$ and $E = \left(\bigcup E_k\right) \cup Z$, so that $|E| = \sum |E_k|$. Let

$$s_\Gamma = \sum \alpha_k |E_k|, \qquad S_\Gamma = \sum \alpha_{k+1} |E_k|.$$

Theorem 5.45 *Let f be a nonnegative measurable function that is finite a.e. in E, $|E| < +\infty$. Then*

$$\int_E f = \lim_{|\Gamma| \to 0} s_\Gamma = \lim_{|\Gamma| \to 0} S_\Gamma.$$

Proof. We may assume that f is finite everywhere, since changing it in a set of measure zero does not affect the expressions above. Given Γ with $|\Gamma| < +\infty$, define functions ϕ_Γ and ψ_Γ by setting $\phi_\Gamma = \alpha_k$ in E_k and $\psi_\Gamma = \alpha_{k+1}$ in $E_k, k = 0, 1, \ldots$. Then $0 \leq \phi_\Gamma \leq f \leq \psi_\Gamma$, $\int_E \phi_\Gamma = s_\Gamma$, and $\int_E \psi_\Gamma = S_\Gamma$. Hence,

$$s_\Gamma \leq \int_E f \leq S_\Gamma.$$

If $s_\Gamma < +\infty$, then

$$0 \le S_\Gamma - s_\Gamma = \sum \left(\alpha_{k+1} - \alpha_k\right) |E_k| \le |\Gamma||E|,$$

so that $S_\Gamma < +\infty$ for the same Γ, and therefore $\int_E f < +\infty$. Then s_Γ and S_Γ are finite for all Γ with $|\Gamma| < +\infty$, and $S_\Gamma - s_\Gamma \to 0$ as $|\Gamma| \to 0$. The conclusion of the theorem now follows easily in case $\int_E f < +\infty$. On the other hand, if $\int_E f = +\infty$, then all $S_\Gamma = +\infty$, and therefore also all $s_\Gamma = +\infty$, which completes the proof.

Theorem 5.45 is the origin of an anecdote that compares the methods that Lebesgue and Riemann might have used to count coins. The story goes that Lebesgue would have been a better bank teller. To see why, imagine coins placed at various points along the x-axis (there may be coins of equal value at different points), and think of $f(x)$ as the value of the coin at x. Suppose that we want to determine the total value of all the coins. In Lebesgue's method, partitioning the ordinate axis and forming the sets E_k corresponds to sorting the coins according to value; computing $|E_k|$ corresponds to counting the number with a given value. Thus, $\sum \alpha_k |E_k| \equiv \int f$ represents the total value. Riemann's method is less efficient; it approximates the total by arbitrarily grouping the coins (partitioning the x-axis) and then summing the products of the number of coins in a given group by the value of any chosen coin in the group.

The relation between Lebesgue and Riemann–Stieltjes integrals can be extended in a useful way to give Riemann–Stieltjes representations for integrals of the form $\int_E \phi(f)$, where f and E are subject to the usual restrictions (see p. 97 in Section 5.4), and ϕ is assumed to be continuous. This last assumption assures the measurability of $\phi(f)$ by Theorem 4.6.

Theorem 5.46 *If $a < f \le b$ (a and b finite) in E and ϕ is continuous on $[a, b]$, then*

$$\int_E \phi(f) = -\int_a^b \phi(\alpha) d\omega(\alpha).$$

Proof. Since ϕ is bounded and E (as always) has finite measure, we see that $\phi(f) \in L(E)$. Since ϕ is continuous, the Riemann–Stieltjes integral exists by Theorem 2.24. Write f as the limit of simple measurable f_k with $a < f_k \le b$ as follows: for $k = 1, 2, \ldots$, let $a = \alpha_0^{(k)} < \alpha_1^{(k)} < \cdots < \alpha_{m_k}^{(k)} = b$ be partitions of $[a, b]$ with norms tending to zero, and let $f_k(\mathbf{x}) = \alpha_j^{(k)}$ when $\alpha_{j-1}^{(k)} < f(\mathbf{x}) \le \alpha_j^{(k)}$. Then $\phi\left(f_k\right) \to \phi(f)$ in E. Since the $\phi\left(f_k\right)$ are uniformly

bounded and $|E| < +\infty$, it follows from the bounded convergence theorem that $\int_E \phi\left(f_k\right) \to \int_E \phi(f)$. However, $\phi\left(f_k\right)$ is simple, taking values $\phi\left(\alpha_j^{(k)}\right)$ on $\left\{\alpha_{j-1}^{(k)} < f \le \alpha_j^{(k)}\right\}$. Therefore, by Lemma 5.38,

$$\int_E \phi\left(f_k\right) = -\sum_j \phi\left(\alpha_j^{(k)}\right)\left[\omega\left(\alpha_j^{(k)}\right) - \omega\left(\alpha_{j-1}^{(k)}\right)\right],$$

so that as $k \to \infty$, $\int_E \phi\left(f_k\right) \to -\int_a^b \phi(\alpha)\, d\omega(\alpha)$. This completes the proof.

In the next theorem, let

$$\int_{-\infty}^{+\infty} \phi(\alpha)\, d\omega(\alpha) = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b \phi(\alpha)\, d\omega(\alpha),$$

if the limit exists (cf. p. 34 in Section 3.4).

Theorem 5.47 *Let ϕ be continuous on $(-\infty, +\infty)$. If $\phi(f) \in L(E)$, then $\int_{-\infty}^{+\infty} \phi(\alpha)\, d\omega(\alpha)$ exists and*

$$\int_E \phi(f) = -\int_{-\infty}^{+\infty} \phi(\alpha)\, d\omega(\alpha).$$

Proof. Since the proof is similar to that of part of Theorem 5.43, we shall be brief. For finite a and b, $a < b$, let E_{ab} and ω_{ab} be as in Theorem 5.42. By Theorem 5.46, $\int_{E_{ab}} \phi(f) = -\int_a^b \phi(\alpha)\, d\omega_{ab}(\alpha)$. Therefore, as in the proof of Theorem 5.42, $\int_{E_{ab}} \phi(f) = -\int_a^b \phi(\alpha) d\omega(\alpha)$. The result now follows by letting $a \to -\infty$ and $b \to +\infty$.

We remark that if ϕ is continuous *and nonnegative*, then the equality

$$\int_E \phi(f) = -\int_{-\infty}^{+\infty} \phi(\alpha)\, d\omega(\alpha)$$

holds without restriction on the finiteness of either side. To see this, simply let $a \to -\infty$, $b \to +\infty$ in the equation $\int_{E_{ab}} \phi(f) = -\int_a^b \phi(\alpha)\, d\omega(\alpha)$.

Thus, for any continuous ϕ, we have

$$\int_E |\phi(f)| = -\int_{-\infty}^{+\infty} |\phi(\alpha)|\, d\omega(\alpha).$$

Taking $\phi(\alpha) = |\alpha|^p$, $0 < p < \infty$, it follows that

$$\int_E |f|^p = -\int_{-\infty}^{+\infty} |\alpha|^p\, d\omega(\alpha).$$

If f is nonnegative, we obtain

$$\int_E f^p = -\int_0^{\infty} \alpha^p\, d\omega(\alpha). \tag{5.48}$$

Hence, for any measurable f,

$$\int_E |f|^p = -\int_0^{\infty} \alpha^p\, d\omega_{|f|}(\alpha).$$

Given $\phi \geq 0$, let $L_\phi(E)$ denote the class of measurable f such that $\phi(f) \in L(E)$. If $\phi(\alpha) = |\alpha|^p$, $0 < p < \infty$, the standard notation is

$$L^p(E) = \left\{ f : \int_E |f|^p < +\infty \right\},\ 0 < p < \infty.$$

Note that $L^1(E) = L(E)$. We will systematically study the L^p classes in Chapter 8. For now, we only want to complete (5.48) by integrating its right side by parts. To do so, we will borrow some facts (Theorems 5.52 and 5.54) from the Section 5.5.

First, note that for measurable f, there is an L^p version of *Tchebyshev's inequality*:

$$\omega(\alpha) \leq \frac{1}{\alpha^p} \int_{\{f > \alpha\}} f^p, \quad \alpha > 0. \tag{5.49}$$

The proof is left as an exercise. Hence, if f is in $L^p(E)$, then $\alpha^p \omega(\alpha)$ remains bounded as $\alpha \to +\infty$. A stronger result is actually true.

Lemma 5.50 *If $0 < p < \infty$ and $f \in L^p(E)$, then*

$$\lim_{\alpha \to +\infty} \alpha^p \omega(\alpha) = 0.$$

Proof. This will be a corollary of (5.49) if we show that $\int_{\{f>\alpha\}} f^p \to 0$ as $\alpha \to +\infty$. We may suppose that α runs through a sequence $\alpha_k \to +\infty$. Let $f_k = f$ wherever $f > \alpha_k$ and $f_k = 0$ elsewhere. Then $\int_{\{f>\alpha_k\}} f^p = \int_E f_k^p$. Since f is finite a.e., $f_k \to 0$ a.e. Moreover, $0 \le f_k^p \le |f|^p \in L(E)$, and the result follows from the dominated convergence theorem.

In the next theorem, we use Lemma 5.50 to integrate the Riemann–Stieltjes integral in (5.48) by parts.

Theorem 5.51 *If $0 < p < \infty$, $f \ge 0$, and $f \in L^p(E)$, then*

$$\int_E f^p = -\int_0^\infty \alpha^p \, d\omega(\alpha) = p \int_0^\infty \alpha^{p-1} \, \omega(\alpha) d\alpha,$$

where the last integral may be interpreted as either a Lebesgue or an improper Riemann integral.

Proof. The first equality is just (5.48). For the second, if $0 < a < b < +\infty$, we have

$$-\int_a^b \alpha^p \, d\omega(\alpha) = -b^p \omega(b) + a^p \omega(a) + p \int_a^b \alpha^{p-1} \omega(\alpha) \, d\alpha,$$

by Theorem 2.21 and the fact that α^p is continuously differentiable on $[a, b]$. Here, together with Theorem 2.21, we use the fact that for partitions $\Gamma = \{\alpha_k\}$ of $[a, b]$ and intermediate points $\{\beta_k\}$ satisfying $\alpha_k < \beta_k < \alpha_{k+1}$ and $\alpha_{k+1}^p - \alpha_k^p = p\beta_k^{p-1}\left(\alpha_{k+1} - \alpha_k\right)$, we have

$$\begin{aligned}\int_a^b \omega(\alpha) \, d\left(\alpha^p\right) &= \lim_{|\Gamma| \to 0} \sum_k \omega\left(\beta_k\right) p\beta_k^{p-1}\left(\alpha_{k+1} - \alpha_k\right) \\ &= p \int_a^b \alpha^{p-1} \omega(\alpha) d\alpha.\end{aligned}$$

The last integral exists in the Riemann sense by Theorem 5.54 since the number of discontinuities of ω is at most countable. By Theorem 5.52, the last integral may also be interpreted in the Lebesgue sense.

Now let $a \to 0$ and $b \to +\infty$. Then $b^p\omega(b) \to 0$ by Lemma 5.50, $a^p\omega(a) \to 0$ since $|E| < +\infty$ (see also Exercise 14), and the theorem follows. Note that in case $1 \le p < \infty$, the proof works with $a = 0$ and no other changes.

For an extension of Theorem 5.51, see Exercise 16. See also Exercise 5 of Chapter 6. In practice, the representation of $\int_E |f|^p$ as a Riemann–Stieltjes integral provides a powerful tool for determining whether or not $f \in L^p(E)$.

5.5 Riemann and Lebesgue Integrals

We now study a relation between Lebesgue and Riemann integrals over finite intervals $[a, b]$ in $\mathbf{R^1}$ and give a characterization of those bounded functions that are Riemann integrable. The Lebesgue integral $\int_{[a,b]} f$ will be denoted by $\int_a^b f$ and the Riemann integral by $(R)\int_a^b f$.

Theorem 5.52 *Let f be a bounded function that is Riemann integrable on $[a, b]$. Then $f \in L[a, b]$ and*

$$\int_a^b f = (R)\int_a^b f.$$

Proof. Let $\{\Gamma_k\}$ be a sequence of partitions of $[a, b]$ with norms tending to zero. For each k, define two simple functions as follows: if $x_1^{(k)} < x_2^{(k)} < \cdots$ are the partitioning points of Γ_k, let $l_k(x)$ and $u_k(x)$ be defined in each semiopen interval $[x_i^{(k)}, x_{i+1}^{(k)})$ as the inf and sup of f on $\left[x_i^{(k)}, x_{i+1}^{(k)}\right]$, respectively. Then l_k and u_k are uniformly bounded and measurable in $[a, b)$, and if L_k and U_k denote the lower and upper Riemann sums of f corresponding to Γ_k, we have

$$\int_a^b l_k = L_k, \qquad \int_a^b u_k = U_k.$$

Note also that $l_k \le f \le u_k$ on the half-open interval $[a, b)$ and, if we assume that Γ_{k+1} is a refinement of Γ_k, that $l_k \nearrow$ and $u_k \searrow$. Let $l = \lim_{k\to\infty} l_k$ and $u = \lim_{k\to\infty} u_k$. Then l and u are measurable, $l \le f \le u$ on $[a, b)$, and, by the bounded convergence theorem, $L_k \to \int_a^b l$ and $U_k \to \int_a^b u$. But since

f is Riemann integrable, L_k and U_k both converge to $(R)\int_a^b f$ by Theorem 2.29. Therefore,

$$(R)\int_a^b f = \int_a^b l = \int_a^b u.$$

Since $u - l \geq 0$, Theorem 5.11 implies that $l = f = u$ a.e. in $[a, b]$. Therefore, f is measurable and $(R)\int_a^b f = \int_a^b f$, which completes the proof.

Theorem 5.52 says that any function that is Riemann integrable is also Lebesgue integrable and that the two integrals are equal. There are, of course, bounded functions that are Lebesgue integrable but not Riemann integrable. One such is the Dirichlet function defined for $0 \leq x \leq 1$ by letting $f(x) = 1$ if x is rational and $f(x) = 0$ if x is irrational. Since $f = 0$ except for a subset of [0,1] of measure zero, its Lebesgue integral is 0. On the other hand, its Riemann integral does not exist since every upper Riemann sum is 1 and every lower Riemann sum is 0.

The practical value of Theorem 5.52 is that it allows us to compute the Lebesgue integral of Riemann integrable (e.g., continuous) functions.

Using the monotone convergence theorem, we can easily extend Theorem 5.52 to include *improper* Riemann integrals of nonnegative functions. Special as it is, the following result is useful in applications.

Theorem 5.53 *Let f be nonnegative on a finite interval $[a, b]$ and Riemann integrable (so, in particular, bounded) over every subinterval $[a + \varepsilon, b]$, $\varepsilon > 0$. Define the improper Riemann integral*

$$I = \lim_{\varepsilon \to 0} (R) \int_{a+\varepsilon}^{b} f,$$

$0 \leq I \leq +\infty$. *Then f is measurable on $[a, b]$ and*

$$\int_a^b f = I.$$

Proof. Observe that by Theorem 5.52, for every $\varepsilon > 0$, f is measurable on $[a+\varepsilon, b]$ and $\int_{a+\varepsilon}^b f = (R)\int_{a+\varepsilon}^b f$. Measurability of f on $[a, b]$ follows easily from Theorem 4.12 by letting $\varepsilon \to 0$. The formula $\int_a^b f = I$ follows similarly from the monotone convergence theorem.

Similar results hold for improper Riemann integrals over infinite intervals. For example, if $a \in (-\infty, \infty)$ and f is nonnegative on $[a, \infty)$ and Riemann integrable on every $[a, N]$, $a < N < \infty$, then f is measurable on $[a, \infty)$ and $\int_a^\infty f = \lim_{N\to\infty} (R) \int_a^N f$. Verification is left to the reader; note that since $(R) \int_a^N f$ increases with N, the limit exists but may be $+\infty$.

We note in passing that the finiteness of the improper Riemann integral of an f that is not nonnegative does not in general imply that f is integrable (see Exercise 7).

Our final result is a characterization of those bounded functions that are Riemann integrable.

Theorem 5.54 *A bounded function is Riemann integrable on* $[a, b]$ *if and only if it is continuous a.e. in* $[a, b]$.

Proof. Suppose that f is bounded and Riemann integrable. Let Γ_k, l_k, u_k, etc., be as in the proof of Theorem 5.52. Let Z be the set of measure zero outside which $l = f = u$. We claim that if x is not a partitioning point of any Γ_k and if $x \notin Z$, then f is continuous at x. In fact, if f is not continuous at x and x is never a partitioning point, there exists $\varepsilon > 0$, depending on x but not on k, such that $u_k(x) - l_k(x) \geq \varepsilon$. This implies that $u(x) - l(x) \geq \varepsilon$, which is impossible if $x \notin Z$. Therefore, f is continuous a.e. in $[a, b]$.

To prove the converse, let f be a bounded function that is continuous a.e. in $[a, b]$. Let $\{\Gamma'_k\}$ be any sequence of partitions with norms tending to zero, and define the corresponding l'_k, u'_k, L'_k, and U'_k as in Theorem 5.52. Note that $\{l'_k\}$ and $\{u'_k\}$ may not be monotone since Γ'_{k+1} may not be a refinement of Γ'_k. However, by the continuity of f, both l'_k and u'_k converge a.e. to f. Hence, by the bounded convergence theorem, $\int_a^b l'_k$ and $\int_a^b u'_k$ both converge to $\int_a^b f$. Since $L'_k = \int_a^b l'_k$ and $U'_k = \int_a^b u'_k$, it follows that the upper and lower Riemann sums converge to the same limit. Therefore, f is Riemann integrable.

Exercises

1. If f is a simple measurable function (not necessarily nonnegative) taking values a_j on E_j, $j = 1, 2, \ldots, N$, show that $\int_E f = \sum_{j=1}^N a_j |E_j|$. (Use Theorem 5.24.)
2. Show that the conclusions of Theorem 5.32 are not generally true without the assumption that $\phi \in L(E)$. (In part (ii), for example, take $f_k = \chi_{(k,\infty)}$.)

   Show that Theorem 5.33 fails without the assumption that $|E| < \infty$.

3. Let $\{f_k\}$ be a sequence of nonnegative measurable functions defined on E. If $f_k \to f$ and $f_k \le f$ a.e. on E, show that $\int_E f_k \to \int_E f$.

4. If $f \in L(0,1)$, show that $x^k f(x) \in L(0,1)$ for $k = 1, 2, \ldots$, and that $\int_0^1 x^k f(x)\,dx \to 0$.

5. Use Egorov's theorem to prove the bounded convergence theorem.

6. Let $f(x,y)$, $0 \le x, y \le 1$, satisfy the following conditions: for each x, $f(x,y)$ is an integrable function of y, and $(\partial f(x,y)/\partial x)$ is a bounded function of (x,y). Show that $(\partial f(x,y)/\partial x)$ is a measurable function of y for each x and

$$\frac{d}{dx}\int_0^1 f(x,y)\,dy = \int_0^1 \frac{\partial}{\partial x} f(x,y)\,dy.$$

7. Give an example of an f that is not integrable, but whose improper Riemann integral exists and is finite.

8. Prove (5.49).

9. If $p > 0$ and $\int_E |f - f_k|^p \to 0$ as $k \to \infty$, show that $f_k \xrightarrow{m} f$ on E (and thus that there is a subsequence $f_{k_j} \to f$ a.e. in E).

10. If $p > 0$, $\int_E |f - f_k|^p \to 0$, and $\int_E |f_k|^p \le M$ for all k, show that $\int_E |f|^p \le M$.

11. For which $p > 0$ does $1/x \in L^p(0,1)$? $L^p(1,\infty)$? $L^p(0,\infty)$?

12. Give an example of a bounded continuous f on $(0,\infty)$ such that $\lim_{x\to\infty} f(x) = 0$ but $f \notin L^p(0,\infty)$ for any $p > 0$.

13. (a) Let $\{f_k\}$ be a sequence of measurable functions on E. Show that $\sum f_k$ converges absolutely a.e. in E if $\sum \int_E |f_k| < +\infty$. (Use Theorems 5.16 and 5.22.)

(b) If $\{r_k\}$ denotes the rational numbers in [0,1] and $\{a_k\}$ satisfies $\sum |a_k| < +\infty$, show that $\sum a_k |x - r_k|^{-1/2}$ converges absolutely a.e. in [0,1].

14. Prove the following result (which is obvious if $|E| < +\infty$), describing the behavior of $a^p\omega(a)$ as $a \to 0+$. If $f \in L^p(E)$, then $\lim_{a\to 0+} a^p\omega(a) = 0$. (If $f \ge 0$, $\varepsilon > 0$, choose $\delta > 0$ so that $\int_{\{f\le\delta\}} f^p < \varepsilon$. Thus, $a^p[\omega(a) - \omega(\delta)] \le \int_{\{a<f\le\delta\}} f^p < \varepsilon$ for $0 < a < \delta$. Now let $a \to 0$.)

15. Suppose that f is nonnegative and measurable on E and that ω is finite on $(0,\infty)$. If $\int_0^\infty \alpha^{p-1}\omega(\alpha)d\alpha$ is finite, show that $\lim_{a\to 0+} a^p\omega(a) = \lim_{b\to+\infty} b^p\omega(b) = 0$. (Consider $\int_{a/2}^a$ and $\int_{b/2}^b$.)

16. Suppose that f is nonnegative and measurable on E and that ω is finite on $(0,\infty)$. Show that Theorem 5.51 holds without any further restrictions (i.e., f need not be in $L^p(E)$ and $|E|$ need not be finite) if we interpret $\int_0^\infty \alpha^p d\omega(\alpha) = \lim_{\substack{a\to 0+\\ b\to 0+\infty}} \int_b^a$. (For the first part, use the sets E_{ab} to obtain the relation $\int_E f^p = -\int_0^\infty \alpha^p\,d\omega(\alpha)$. If either $\int_0^\infty \alpha^p\,d\omega(\alpha)$ or

$\int_0^\infty \alpha^{p-1}\,\omega(\alpha)d\alpha$ is finite, use Lemma 5.50 and the results of Exercises 14 or 15 to integrate by parts.)

17. If $f \geq 0$ and $\omega(\alpha) \leq c(1+\alpha)^{-p}$ for all $\alpha > 0$, show that $f \in L^r$, $0 < r < p$.

18. If $f \geq 0$, show that $f \in L^p$ if and only if $\sum_{k=-\infty}^{+\infty} 2^{kp}\omega\left(2^k\right) < +\infty$. (Use Exercise 16.)

19. Derive analogues of Theorems 5.52 and 5.54 for integrals over intervals in $\mathbf{R^n}$, $n > 1$.

20. Let $\mathbf{y} = T\mathbf{x}$ be a nonsingular linear transformation of $\mathbf{R^n}$. If $\int_E f(\mathbf{y})\,d\mathbf{y}$ exists, show that

$$\int_E f(\mathbf{y})\,d\mathbf{y} = |\det T| \int_{T^{-1}E} f(T\mathbf{x})\,d\mathbf{x}.$$

(The case when $f = \chi_{E_1}$, $E_1 \subset E$, follows from integrating the formula $\chi_{E_1}(T\mathbf{x}) = \chi_{T^{-1}E_1}(\mathbf{x})$ over $T^{-1}E$ and then applying Theorem 3.35.)

21. If $\int_A f = 0$ for every measurable subset A of a measurable set E, show that $f = 0$ a.e. in E.

22. Show that the conclusion of the Lebesgue dominated convergence theorem can be strengthened to $\int_E |f_k - f| \to 0$.

23. Prove the following fact, sometimes referred to as the *Sequential (or Generalized) Version of the Lebesgue Dominated Convergence Theorem*. Let $\{f_k\}$ and $\{\phi_k\}$ be sequences of measurable functions on E satisfying $f_k \to f$ a.e. in E, $\phi_k \to \phi$ a.e. in E, and $|f_k| \leq \phi_k$ a.e. in E. If $\phi \in L(E)$ and $\int_E \phi_k \to \int_E \phi$, then $\int_E |f_k - f| \to 0$. (In case $f = 0$ and all $f_k \geq 0$, apply Fatou's lemma to $\{\phi_k - f_k\}$.) An application is given in Exercise 12 of Chapter 8; for example, if $f_k \geq 0$, $f_k \to f$ a.e. in E, $f \in L(E)$, and $\int_E f_k \to \int_E f$, then $\int_E |f_k - f| \to 0$.

24. A measurable function f on E is said to belong to *weak* $L^p(E)$, $0 < p < \infty$, if there is a constant $A \geq 0$ such that $\omega_{|f|}(\alpha) \leq A\alpha^{-p}$ for all $\alpha > 0$ (cf. (7.8) in case $p = 1$):

(a) Show that if $f \in L^p(E)$, then f belongs to weak $L^p(E)$, but that the converse is generally false.

(b) Show that if $1 < p < r < \infty$ and f belongs to both weak $L^1(E)$ and weak $L^r(E)$, then $f \in L^p(E)$.

(c) Show that if f belongs to weak $L^1(E)$ and f is bounded on E, then $f \in L^p(E)$ for all $1 < p < \infty$.

25. Give an example to show that the analogue of Theorem 5.8 with the roles of sup and inf interchanged is false.

26. Prove the following variant of Lebesgue's dominated convergence theorem: if $\{f_k\}$ satisfies $f_k \xrightarrow{m} f$ on E and $|f_k| \leq \phi \in L(E)$, then $f \in L(E)$ and

$\int_E f_k \to \int_E f$. (Show that every subsequence of $\{f_k\}$ has a subsequence $\{f_{k_j}\}$ such that $\int_E f_{k_j} \to \int_E f$.)

27. The notion of equimeasurability of functions can be extended to different sets E_1 and E_2, even in different dimensions, by saying that two measurable functions f_1, f_2 defined on E_1, E_2, respectively, are equimeasurable if

$$\left|\{\mathbf{x} \in E_1 : f_1(\mathbf{x}) > \alpha\}\right| = \left|\{\mathbf{y} \in E_2 : f_2(\mathbf{y}) > \alpha\}\right| \quad \text{for all } \alpha.$$

(a) Show that if f is measurable and finite a.e. in E and ω_f is strictly decreasing and continuous, then f and the inverse function of ω_f are equimeasurable (on E and $(0, |E|)$, respectively).

(b) Let f be measurable and finite a.e. in E, and suppose that ω_f is finite. Define $f^*(t) = \inf\{\alpha > 0 : \omega_f(\alpha) \le t\}$, $t > 0$. Show that f and f^* are equimeasurable on E and $(0, \infty)$, respectively. (The function f^* is called the *nonincreasing rearrangement* of f.)

28. Let E be a measurable set in $\mathbf{R^n}$ with $|E| < \infty$. Suppose that $f > 0$ a.e. in E and $f, \log f \in L^1(E)$. Prove that

$$\lim_{p \to 0+} \left(\frac{1}{|E|} \int_E f^p \right)^{1/p} = \exp\left(\frac{1}{|E|} \int_E \log f \right).$$

(Start by using Theorem 5.36 to show that $\int_E f^p \to |E|$ as $p \to 0+$. Note that $\left(f^p - 1\right)/p \to \log f$.)

29. Let f be measurable, nonnegative, and finite a.e. in a set E. Prove that for any nonnegative constant c,

$$\int_E e^{cf(\mathbf{x})} d\mathbf{x} = |E| + c \int_0^\infty e^{c\alpha} \omega_f(\alpha)\, d\alpha.$$

Deduce that $e^{cf} \in L(E)$ if $|E| < \infty$ and there exist constants C_1 and c_1 such that $c_1 > c$ and $\omega_f(\alpha) \le C_1 e^{-c_1 \alpha}$ for all $\alpha > 0$. We will study such an exponential integrability property in Section 14.5.

6

Repeated Integration

Let $f(x,y)$ be defined in a rectangle

$$I = \left\{(x,y) : a \le x \le b,\ c \le y \le d\right\}.$$

If f is continuous, we have the classical formula

$$\iint_I f(x,y)\,dxdy = \int_a^b \left[\int_c^d f(x,y)\,dy\right] dx,$$

and there is an analogous formula for functions of n variables.

Sections 6.1 and 6.2 extend this and related results on repeated integration to the case of Lebesgue integrable functions. Section 6.3 contains some applications.

6.1 Fubini's Theorem

We shall use the following notation. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a point of an n-dimensional interval I_1,

$$I_1 = \{\mathbf{x} = (x_1, \ldots, x_n) : a_i \le x_i \le b_i,\ i = 1, \ldots, n\},$$

and let $\mathbf{y}$ be a point of an m-dimensional interval I_2,

$$I_2 = \left\{\mathbf{y} = (y_1, \ldots, y_m) : c_j \le y_j \le d_j, j = 1, \ldots, m\right\}.$$

Here, I_1 and I_2 may also be partly open or unbounded, such as all of $\mathbf{R^n}$ and $\mathbf{R}^m$, respectively. The Cartesian product $I = I_1 \times I_2$ is contained in $\mathbf{R^{n+m}}$ and consists of points $(x_1, \ldots, x_n, y_1, \ldots, y_m)$. We shall denote such points by $(\mathbf{x},\mathbf{y})$. A function $f(x_1, \ldots, x_n, y_1, \ldots, y_m)$ defined in I will be written $f(\mathbf{x},\mathbf{y})$, and its integral $\int_I f$ will be denoted by $\iint_I f(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y}$.

Theorem 6.1 (Fubini's Theorem) *Let $f(\mathbf{x},\mathbf{y}) \in L(I)$, $I = I_1 \times I_2$. Then*

(i) *For almost every $\mathbf{x} \in I_1$, $f(\mathbf{x},\mathbf{y})$ is measurable and integrable on I_2 as a function of $\mathbf{y}$;*

(ii) *As a function of $\mathbf{x}$, $\int_{I_2} f(\mathbf{x},\mathbf{y})\,d\mathbf{y}$ is measurable and integrable on I_1, and*

$$\iint\limits_I f(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = \int\limits_{I_1}\left[\int\limits_{I_2} f(\mathbf{x},\mathbf{y})\,d\mathbf{y}\right]d\mathbf{x}.$$

Setting $f = 0$ outside I, we see that it is enough to prove the theorem when $I_1 = \mathbf{R^n}$, $I_2 = \mathbf{R}^m$, and $I = \mathbf{R^{n+m}}$. For simplicity, we then drop I_1, I_2, and I from the notation and write $\int f(\mathbf{x},\mathbf{y})\,d\mathbf{x}$ for $\int_{I_1} f(\mathbf{x},\mathbf{y})\,d\mathbf{x}$, $L(d\mathbf{x})$ for $L(I_1)$, $L(d\mathbf{x}d\mathbf{y})$ for $L(I)$, etc.

We will prove the theorem by considering a series of special cases. The first two lemmas below will help in passing from one case to the next. In these lemmas, we say that a function f in $L(d\mathbf{x}d\mathbf{y})$ for which Fubini's theorem is true has *property* $\mathscr{F}$.

Lemma 6.2 *A finite linear combination of functions with property $\mathscr{F}$ has property $\mathscr{F}$.*

This follows immediately from Theorems 4.9 and 5.28.

Lemma 6.3 *Let $f_1, f_2, \ldots, f_k, \ldots$ have property $\mathscr{F}$. If $f_k \nearrow f$ or $f_k \searrow f$, and if $f \in L(d\mathbf{x}d\mathbf{y})$, then f has property $\mathscr{F}$.*

Proof. We will concentrate on integrability properties, leaving questions of measurability to the reader. Changing signs if necessary, we may assume that $f_k \nearrow f$. For each k, there exists by hypothesis a set Z_k in $\mathbf{R^n}$ with measure zero such that $f_k(\mathbf{x},\mathbf{y}) \in L(d\mathbf{y})$ if $\mathbf{x} \notin Z_k$. Let $Z = \bigcup_k Z_k$, so that Z has $\mathbf{R^n}$-measure zero. If $\mathbf{x} \notin Z$, then $f_k(\mathbf{x},\mathbf{y}) \in L(d\mathbf{y})$ for all k, and therefore, by the monotone convergence theorem applied to $\{f_k(\mathbf{x},\mathbf{y})\}$ as functions of $\mathbf{y}$,

$$h_k(\mathbf{x}) = \int f_k(\mathbf{x},\mathbf{y})\,d\mathbf{y} \nearrow h(\mathbf{x}) = \int f(\mathbf{x},\mathbf{y})\,d\mathbf{y} \qquad (\mathbf{x} \notin Z).$$

By assumption, we have $h_k(\mathbf{x}) \in L(d\mathbf{x})$, $f_k \in L(d\mathbf{x}d\mathbf{y})$, and $\iint f_k(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = \int h_k(\mathbf{x})\,d\mathbf{x}$. Therefore, another application of the monotone convergence theorem gives $\iint f(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = \int h(\mathbf{x})\,d\mathbf{x}$. Since $f \in L(d\mathbf{x}d\mathbf{y})$, it follows that $h \in L(d\mathbf{x})$, which implies that h is finite a.e. This completes the proof.

The next three lemmas prove special cases of Fubini's theorem.

Lemma 6.4 *If E is a set of type G_δ, namely, $E = \bigcap_{k=1}^{\infty} G_k$, and if G_1 has finite measure, then χ_E has property $\mathscr{F}$.*

Proof. *Case 1.* Suppose that E is a bounded open interval in $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$: $E = J_1 \times J_2$, where J_1 and J_2 are bounded open intervals in $\mathbf{R}^{\mathbf{n}}$ and $\mathbf{R}^{\boldsymbol{m}}$, respectively. Then $|E| = |J_1||J_2|$, where $|J_1|$ and $|J_2|$ denote the measures of J_1 and J_2 in $\mathbf{R}^{\mathbf{n}}$ and $\mathbf{R}^{\boldsymbol{m}}$. For every $\mathbf{x}$, $\chi_E(\mathbf{x},\mathbf{y})$ is clearly measurable as a function of $\mathbf{y}$. If $h(\mathbf{x}) = \int \chi_E(\mathbf{x},\mathbf{y})\,d\mathbf{y}$, then $h(\mathbf{x}) = |J_2|$ for $\mathbf{x} \in J_1$, and $h(\mathbf{x}) = 0$ otherwise. Therefore, $\int h(\mathbf{x})\,d\mathbf{x} = |J_1||J_2|$. But also, $\iint \chi_E(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = |E| = |J_1||J_2|$, and the lemma is proved in this case.

Case 2. Suppose that E is any set (of type G_δ or not) on the boundary of an interval in $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$. Then for almost every $\mathbf{x}$, the set $\{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in E\}$ has $\mathbf{R}^m$-measure zero. Therefore, if $h(\mathbf{x}) = \int \chi_E(\mathbf{x},\mathbf{y})\,d\mathbf{y}$, it follows that $h(\mathbf{x}) = 0$ a.e. Hence, $\int h(\mathbf{x})\,d\mathbf{x} = 0$. But also, $\iint \chi_E(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = |E| = 0$.

Case 3. Suppose next that E is a partly open interval in $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$. Then E is the union of its interior and a subset of its boundary. It follows from cases 1 and 2 and Lemma 6.2 that χ_E has property $\mathscr{F}$.

Case 4. Let E be an open set in $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$ with finite measure. Write $E = \bigcup I_j$, where the I_j are disjoint, partly open intervals. If $E_k = \bigcup_{j=1}^{k} I_j$, then $\chi_{E_k} = \sum_{j=1}^{k} \chi_{I_j}$, so that χ_{E_k} has property $\mathscr{F}$ by case 3 and Lemma 6.2. Since $\chi_{E_k} \nearrow \chi_E$, χ_E has property $\mathscr{F}$ by Lemma 6.3.

Case 5. Let E satisfy the hypothesis of Lemma 6.4. We may assume that $G_k \searrow E$ by considering the open sets G_1, $G_1 \cap G_2$, $G_1 \cap G_2 \cap G_3$, etc. Then $\chi_{G_k} \searrow \chi_E$, and the lemma follows from case 4 and Lemma 6.3.

Lemma 6.5 *If Z is a subset of $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$ with measure zero, then χ_Z has property $\mathscr{F}$. Hence, for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$, the set $\{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in Z\}$ has $\mathbf{R}^{\boldsymbol{m}}$-measure zero.*

Proof. Using Theorem 3.8, select a set H of type G_δ such that $Z \subset H$ and $|H| = 0$. If $H = \bigcap G_k$, we may assume that G_1 has finite measure, so that by Lemma 6.4,

$$\int \left[\int \chi_H(\mathbf{x},\mathbf{y})\,d\mathbf{y} \right] d\mathbf{x} = \iint \chi_H(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = 0.$$

Therefore, by Theorem 5.11, $|\{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in H\}| = \int \chi_H(\mathbf{x},\mathbf{y})\,d\mathbf{y} = 0$ for almost every $\mathbf{x}$. If $\{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in H)$ has $\mathbf{R}^{\boldsymbol{m}}$-measure zero, so does $\{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in Z\}$ since $Z \subset H$. It follows that for almost every $\mathbf{x}$, $\chi_Z(\mathbf{x},\mathbf{y})$ is measurable in $\mathbf{y}$ and $\int \chi_Z(\mathbf{x},\mathbf{y})\,d\mathbf{y} = 0$. Hence, $\int [\int \chi_Z(\mathbf{x},\mathbf{y})\,d\mathbf{y}]\,d\mathbf{x} = 0$, which proves the lemma since $\iint \chi_Z(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = |Z| = 0$.

Lemma 6.6 *Let $E \subset \mathbf{R}^{\mathbf{n}+\mathbf{m}}$. If E is measurable with finite measure, then χ_E has property $\mathscr{F}$.*

Proof. Using Theorem 3.28, write $E = H - Z$, where H is of type G_δ and Z has measure zero. If $H = \bigcap G_k$, choose G_1 with finite measure (see the proof of Theorem 3.28). Since $\chi_E = \chi_H - \chi_Z$, the result follows from Lemmas 6.2, 6.4, and 6.5.

Proof of Fubini's theorem. We must show that every $f \in L(d\mathbf{x}d\mathbf{y})$ has property $\mathscr{F}$. Since $f = f^+ - f^-$, we may assume by Lemma 6.2 that $f \geq 0$. Then, by Theorem 4.13, there are simple measurable $f_k \nearrow f, f_k \geq 0$. Each $f_k \in L(d\mathbf{x}d\mathbf{y})$, and by Lemma 6.3, it is enough to show that these have property $\mathscr{F}$. Hence, we may assume that f is simple and integrable, say $f = \sum_{j=1}^{N} v_j \chi_{E_j}$. Since each E_j for which $v_j \neq 0$ must have finite measure, the result follows from Lemmas 6.2 and 6.6.

If $f \in L(\mathbf{R}^{\mathbf{n}+\mathbf{m}})$, then by Fubini's theorem, $f(\mathbf{x},\mathbf{y})$ is a measurable function of $\mathbf{y}$ for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$. We now show that the same conclusion holds if f is merely measurable.

Theorem 6.7 *Let $f(\mathbf{x},\mathbf{y})$ be a measurable function on $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$. Then for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$, $f(\mathbf{x},\mathbf{y})$ is a measurable function of $\mathbf{y} \in \mathbf{R}^{m}$.*

In particular, if E is a measurable subset of $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$, then the set

$$E_{\mathbf{x}} = \{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in E\}$$

is measurable in $\mathbf{R}^{\mathbf{m}}$ for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$.

Proof. Note that if f is the characteristic function χ_E of a measurable $E \subset \mathbf{R}^{\mathbf{n}+\mathbf{m}}$, then the two statements of the theorem are equivalent. To prove the result in this case, write $E = H \cup Z$, where H is of type F_σ in $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$ and $|Z|_{n+m} = 0$. Then $E_{\mathbf{x}} = H_{\mathbf{x}} \cup Z_{\mathbf{x}}$, $H_{\mathbf{x}}$ is of type F_σ in $\mathbf{R}^{m}$, and for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$, $|Z_{\mathbf{x}}|_m = 0$ by Lemma 6.5. Therefore, $E_{\mathbf{x}}$ is measurable for almost every $\mathbf{x}$.

If f is any measurable function on $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$, consider the set $E(a) = \{(\mathbf{x},\mathbf{y}) : f(\mathbf{x},\mathbf{y}) > a\}$. Since $E(a)$ is measurable in $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$, the set $E(a)_{\mathbf{x}} = \{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in E(a)\}$ is measurable in $\mathbf{R}^{m}$ for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$. The exceptional set of $\mathbf{R}^{\mathbf{n}}$-measure zero depends on a. The union Z of these exceptional sets for all rational a still has $\mathbf{R}^{\mathbf{n}}$-measure zero. If $\mathbf{x} \notin Z$, then $\{\mathbf{y} : f(\mathbf{x},\mathbf{y}) > a\}$ is measurable for all rational a and so for all a by Theorem 4.4. This completes the proof.

We will now extend Fubini's theorem to functions defined on measurable subsets of $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$.

Theorem 6.8 *Let $f(\mathbf{x}, \mathbf{y})$ be a measurable function defined on a measurable subset E of $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$, and let $E_{\mathbf{x}} = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in E\}$.*

(i) *For almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$, $f(\mathbf{x}, \mathbf{y})$ is a measurable function of $\mathbf{y}$ on $E_{\mathbf{x}}$.*

(ii) *If $f(\mathbf{x}, \mathbf{y}) \in L(E)$, then for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$, $f(\mathbf{x}, \mathbf{y})$ is integrable on $E_{\mathbf{x}}$ with respect to $\mathbf{y}$; moreover, $\int_{E_{\mathbf{x}}} f(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$ is an integrable function of $\mathbf{x}$ and*

$$\iint\limits_{E} f(\mathbf{x}, \mathbf{y})\, d\mathbf{x} d\mathbf{y} = \int\limits_{\mathbf{R}^{\mathbf{n}}} \left[\int\limits_{E_{\mathbf{x}}} f(\mathbf{x}, \mathbf{y})\, d\mathbf{y} \right] d\mathbf{x}.$$

Proof. Let $\bar{f}$ be the function equal to f in E and to zero elsewhere in $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$. Since f is measurable on E, $\bar{f}$ is measurable on $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$. Therefore, by Theorem 6.7, $\bar{f}(\mathbf{x}, \mathbf{y})$ is a measurable function of $\mathbf{y}$ for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$. Since $E_{\mathbf{x}}$ is measurable for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$, it follows that $f(\mathbf{x}, \mathbf{y})$ is measurable on almost every $E_{\mathbf{x}}$. This proves (i).

If $f \in L(E)$, then $\bar{f} \in L(\mathbf{R}^{\mathbf{n}+\mathbf{m}})$ and

$$\iint\limits_{E} f(\mathbf{x}, \mathbf{y})\, d\mathbf{x} d\mathbf{y} = \iint\limits_{\mathbf{R}^{\mathbf{n}+\mathbf{m}}} \bar{f}(\mathbf{x}, \mathbf{y})\, d\mathbf{x} d\mathbf{y} = \int\limits_{\mathbf{R}^{\mathbf{n}}} \left[\int\limits_{\mathbf{R}^{m}} \bar{f}(\mathbf{x}, \mathbf{y})\, d\mathbf{y} \right] d\mathbf{x}.$$

Since $E_{\mathbf{x}}$ is measurable for almost every $\mathbf{x}$, we obtain by Theorem 5.24 that $\int_{\mathbf{R}^m} \bar{f}(\mathbf{x}, \mathbf{y})\, d\mathbf{y} = \int_{E_{\mathbf{x}}} f(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$ for almost every $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$. Part (ii) follows by combining equalities.

6.2 Tonelli's Theorem

By Fubini's theorem, the finiteness of a multiple integral implies that of the corresponding iterated integrals. The converse is not true, even if all the iterated integrals are equal, as shown by the following example.

Example 6.9 Let $n = m = 1$ and let I be the unit square and $\{I_k\}$ be the infinite sequence of subsquares shown in the following illustration:

Subdivide each I_k into four equal subsquares by lines parallel to the x- and y-*axes*.

For each k, let $f = 1/|I_k|$ on the interiors of $I_k^{(1)}$ and $I_k^{(3)}$ and let $f = -1/|I_k|$ on the interiors of $I_k^{(2)}$ and $I_k^{(4)}$. Let $f = 0$ on the rest of I, that is, outside $\bigcup I_k$ and on the boundaries of all the subsquares. Clearly, $\int_0^1 f(x,y)\,dy = 0$ for all x, and $\int_0^1 f(x,y)\,dx = 0$ for all y. Therefore,

$$\int_0^1 \left[\int_0^1 f(x,y)\,dy \right] dx = \int_0^1 \left[\int_0^1 f(x,y)\,dx \right] dy = 0.$$

However,

$$\iint_I f^+(x,y)\,dxdy = \sum_k \iint_{I_k} f^+(x,y)\,dxdy = \sum_k \frac{1}{2} = +\infty.$$

Similarly, $\iint_I f^- \, dxdy = +\infty$. Hence, finiteness of the iterated integrals of f does not in general imply either the existence of the multiple integral of f or the finiteness of the multiple integral of $|f|$. However, for nonnegative f, we have the following basic result.

Theorem 6.10 (Tonelli's Theorem) *Let $f(\mathbf{x}, \mathbf{y})$ be nonnegative and measurable on an interval $I = I_1 \times I_2$ of $\mathbf{R}^{\mathbf{n}+\mathbf{m}}$. Then, for almost every $\mathbf{x} \in I_1$, $f(\mathbf{x}, \mathbf{y})$ is a measurable function of $\mathbf{y}$ on I_2. Moreover, as a function of $\mathbf{x}$, $\int_{I_2} f(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}$ is measurable on I_1, and*

$$\iint_I f(\mathbf{x}, \mathbf{y}) \, d\mathbf{x}d\mathbf{y} = \int_{I_1} \left[\int_{I_2} f(\mathbf{x}, \mathbf{y}) \, d\mathbf{y} \right] d\mathbf{x}.$$

Proof. This is actually a corollary of Fubini's theorem. For $k = 1, 2, \ldots$, let $f_k(\mathbf{x}, \mathbf{y}) = 0$ if $|(\mathbf{x}, \mathbf{y})| > k$ and $f_k(\mathbf{x}, \mathbf{y}) = \min\{k, f(\mathbf{x}, \mathbf{y})\}$ if $|(\mathbf{x}, \mathbf{y})| \leq k$. Then $f_k \geq 0$, $f_k \nearrow f$ on I, and $f_k \in L(I)$ (f_k is bounded and vanishes outside a compact set). Hence, Fubini's theorem applies to each f_k. The statement concerning the measurability of $\int_{I_2} f(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}$ then follows from its analogue for f_k; in fact, by the monotone convergence theorem, $\int_{I_2} f_k(\mathbf{x}, \mathbf{y}) \, d\mathbf{y} \nearrow \int_{I_2} f(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}$. (The measurability of $f(\mathbf{x},\mathbf{y})$ as a function of $\mathbf{y}$ was proved in Theorem 6.8.) By the monotone convergence theorem again,

$$\iint_I f_k(\mathbf{x}, \mathbf{y}) \, d\mathbf{x}d\mathbf{y} \to \iint_I f(\mathbf{x}, \mathbf{y}) \, d\mathbf{x}d\mathbf{y}, \quad \text{and}$$

$$\int_{I_1} \left[\int_{I_2} f_k(\mathbf{x}, \mathbf{y}) \, d\mathbf{y} \right] d\mathbf{x} \to \int_{I_1} \left[\int_{I_2} f(\mathbf{x}, \mathbf{y}) \, d\mathbf{y} \right] d\mathbf{x}.$$

Since $f_k \in L(I)$, the left-hand sides in the last two limits are equal. Therefore, so are the right-hand sides, and the theorem follows.

An extension of Tonelli's theorem to functions defined over arbitrary measurable sets E is straightforward.

Since the roles of $\mathbf{x}$ and $\mathbf{y}$ can be interchanged above, it follows that if f is nonnegative and measurable, then

$$\iint_I f(\mathbf{x}, \mathbf{y}) \, d\mathbf{x}d\mathbf{y} = \int_{I_1} \left[\int_{I_2} f(\mathbf{x}, \mathbf{y}) \, d\mathbf{y} \right] d\mathbf{x} = \int_{I_2} \left[\int_{I_1} f(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \right] d\mathbf{y}.$$

In particular, we obtain the important fact that *for $f \geq 0$, the finiteness of any one of Fubini's three integrals implies that of the other two.* Hence, for any measurable

f, the finiteness of one of these integrals for $|f|$ implies that f is integrable and that all three Fubini integrals of f are equal.

An easy consequence of Tonelli's theorem is that the conclusion

$$\iint_I f(\mathbf{x},\mathbf{y})\,d\mathbf{x}d\mathbf{y} = \int_{I_1}\left[\int_{I_2} f(\mathbf{x},\mathbf{y})\,d\mathbf{y}\right]d\mathbf{x}$$

of Fubini's theorem (including the existence and measurability of the inner integral on the right-hand side) holds for measurable f even if $\iint_I f = \pm\infty$ (i.e., it holds if $\iint_I f$ merely *exists*). In fact, if $\iint_I f = +\infty$, we have $\iint_I f^+ = +\infty$ and $f^- \in L(I)$. By Tonelli's theorem,

$$\iint_I f^+ = \int_{I_1}\left(\int_{I_2} f^+\,d\mathbf{y}\right)d\mathbf{x}, \qquad \iint_I f^- = \int_{I_1}\left(\int_{I_2} f^-\,d\mathbf{y}\right)d\mathbf{x}.$$

Since $\iint_I f^-$ is finite, the desired formula follows by subtraction.

6.3 Applications of Fubini's Theorem

We shall derive several important results as corollaries of Fubini's and Tonelli's theorems. The first one is the necessity of the condition in Theorem 5.1. Using the notation of Chapter 5, we will prove the following result.

Theorem 6.11 *Let f be a nonnegative function defined on a measurable set $E \subset \mathbf{R^n}$. If $R(f,E)$, the region under f over E, is a measurable subset of $\mathbf{R^{n+1}}$, then f is measurable.*

Proof. For $0 \le y < +\infty$, we have

$$\{\mathbf{x} \in E : f(\mathbf{x}) \ge y\} = \{\mathbf{x} : (\mathbf{x}, y) \in R(f,E)\}.$$

Since $R(f,E)$ is measurable, it follows from Theorem 6.7 that $\{\mathbf{x} \in E : f(\mathbf{x}) \ge y\}$ is measurable (in $\mathbf{R^n}$) for almost all (linear measure) such y. In particular, $\{\mathbf{x} \in E : f(\mathbf{x}) \ge y\}$ is measurable for all y in a dense subset of $(0,\infty)$. If y is negative, then $\{\mathbf{x} \in E : f(\mathbf{x}) \ge y\} = E$, which is measurable. We conclude that f is measurable (cf. Theorem 4.4).

As a second application of Fubini's theorem, we will prove a result about the convolution of two functions. More general results of this kind will be proved

in Chapter 9. If f and g are measurable in $\mathbf{R^n}$, their *convolution* $(f * g)(\mathbf{x})$ is defined by

$$(f * g)(\mathbf{x}) = \int_{\mathbf{R^n}} f(\mathbf{x} - \mathbf{t}) g(\mathbf{t})\, d\mathbf{t},$$

provided the integral exists.

We first claim that $f * g = g * f$, that is, that

$$\int_{\mathbf{R^n}} f(\mathbf{x} - \mathbf{t}) g(\mathbf{t})\, d\mathbf{t} = \int_{\mathbf{R^n}} f(\mathbf{t}) g(\mathbf{x} - \mathbf{t})\, d\mathbf{t}. \tag{6.12}$$

This is actually a special case of results dealing with changes of variable. In this simple case, however, it amounts to the statement that if $\mathbf{x} \in \mathbf{R^n}$, then

$$\int_{\mathbf{R^n}} r(\mathbf{t})\, d\mathbf{t} = \int_{\mathbf{R^n}} r(\mathbf{x} - \mathbf{t})\, d\mathbf{t} \tag{6.13}$$

when $r(\mathbf{t}) = f(\mathbf{x} - \mathbf{t}) g(\mathbf{t})$. For fixed $\mathbf{x}$, $\mathbf{x} - \mathbf{t}$ ranges over $\mathbf{R^n}$ as $\mathbf{t}$ does. Therefore, for any measurable $r \geq 0$, (6.13) follows from the geometric interpretation of the integral. (See Theorem 5.1 and the definition of the integral of a nonnegative function.) For any measurable r, it follows by writing $r = r^+ - r^-$. (For the effect of a *linear* change of variables, see Exercise 20 of Chapter 5, and see Exercise 18 of Chapter 3 for the effect of translations. Measurability of $r(\mathbf{x} - \mathbf{t})$ as a function of $\mathbf{t}$ can then be deduced from that of $r(\mathbf{t})$.)

The result we wish to prove for convolutions is the following.

Theorem 6.14 *If $f \in L(\mathbf{R^n})$ and $g \in L(\mathbf{R^n})$, then $(f * g)(\mathbf{x})$ exists for almost every $\mathbf{x} \in \mathbf{R^n}$ and is measurable. Moreover, $f * g \in L(\mathbf{R^n})$ and*

$$\int_{\mathbf{R^n}} |f * g|\, d\mathbf{x} \leq \left(\int_{\mathbf{R^n}} |f|\, d\mathbf{x} \right) \left(\int_{\mathbf{R^n}} |g|\, d\mathbf{x} \right),$$

$$\int_{\mathbf{R^n}} (f * g)\, d\mathbf{x} = \left(\int_{\mathbf{R^n}} f\, d\mathbf{x} \right) \left(\int_{\mathbf{R^n}} g\, d\mathbf{x} \right).$$

In order to prove this, we need a lemma.

Lemma 6.15 *If $f(\mathbf{x})$ is measurable in $\mathbf{R^n}$, then the function $F(\mathbf{x}, \mathbf{t}) = f(\mathbf{x} - \mathbf{t})$ is measurable in $\mathbf{R^n} \times \mathbf{R^n} = \mathbf{R^{2n}}$.*

Proof. Let $F_1(\mathbf{x}, \mathbf{t}) = f(\mathbf{x})$. Since f is measurable, it follows as in Lemma 5.2 (or by n-fold iteration of the conclusion of Lemma 5.2) that $F_1(\mathbf{x}, \mathbf{t})$ is measurable in $\mathbf{R}^{2\mathbf{n}}$: in fact, the set $\{(\mathbf{x}, \mathbf{t}) : F_1(\mathbf{x}, \mathbf{t}) > a\}$, which equals $\{(\mathbf{x}, \mathbf{t}) : f(\mathbf{x}) > a, \mathbf{t} \in \mathbf{R}^{\mathbf{n}}\}$, is a cylinder type set with measurable base $\{\mathbf{x} : f(\mathbf{x}) > a)$ in $\mathbf{R}^{\mathbf{n}}$. For $(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \mathbf{R}^{2\mathbf{n}}$, consider the transformation $\mathbf{x} = \boldsymbol{\xi} - \boldsymbol{\eta}$, $\mathbf{t} = \boldsymbol{\xi} + \boldsymbol{\eta}$. This is a nonsingular linear transformation of $\mathbf{R}^{2\mathbf{n}}$, and therefore, by Theorem 3.33 (see Exercise 4 of Chapter 4), the function F_2 defined by $F_2(\boldsymbol{\xi}, \boldsymbol{\eta}) = F_1(\boldsymbol{\xi} - \boldsymbol{\eta}, \boldsymbol{\xi} + \boldsymbol{\eta})$ is measurable in $\mathbf{R}^{2\mathbf{n}}$. Since $F_2(\boldsymbol{\xi}, \boldsymbol{\eta}) = f(\boldsymbol{\xi} - \boldsymbol{\eta})$, the lemma follows.

Proof of Theorem 6.14 Suppose first that both f and g are nonnegative on $\mathbf{R}^{\mathbf{n}}$. By Lemma 6.15, $f(\mathbf{x} - \mathbf{t})g(\mathbf{t})$ is measurable on $\mathbf{R}^{\mathbf{n}} \times \mathbf{R}^{\mathbf{n}}$ since it is the product of two such functions. Hence, the integral

$$I = \iint f(\mathbf{x} - \mathbf{t})g(\mathbf{t})\, d\mathbf{t} d\mathbf{x}$$

is well defined. By Tonelli's theorem and (6.13),

$$\begin{aligned} I &= \int \left[\int f(\mathbf{x} - \mathbf{t})g(\mathbf{t})\, d\mathbf{t}\right] d\mathbf{x} \\ &= \int g(\mathbf{t}) \left[\int f(\mathbf{x} - \mathbf{t})\, d\mathbf{x}\right] d\mathbf{t} = \left[\int f(\mathbf{x})\, d\mathbf{x}\right] \left[\int g(\mathbf{t})\, d\mathbf{t}\right]. \end{aligned}$$

The first of these equations can be written $I = \int (f * g)(\mathbf{x})\, d\mathbf{x}$, where measurability of $f * g$ is guaranteed by Tonelli's theorem, so that

$$\int (f * g)(\mathbf{x})\, d\mathbf{x} = \left[\int f(\mathbf{x})\, d\mathbf{x}\right] \left[\int g(\mathbf{x})\, d\mathbf{x}\right].$$

This proves the theorem for $f \geq 0$ and $g \geq 0$. For general $f, g \in L(\mathbf{R}^{\mathbf{n}})$, it follows that

$$\iint |f(\mathbf{x} - \mathbf{t})|\, |g(\mathbf{t})|\, d\mathbf{t} d\mathbf{x} = \int \left(|f| * |g|\right) d\mathbf{x} = \left(\int |f|\, d\mathbf{x}\right) \left(\int |g|\, d\mathbf{x}\right) < \infty.$$

Hence, $f(\mathbf{x} - \mathbf{t})g(\mathbf{t}) \in L(d\mathbf{t} d\mathbf{x})$. By Fubini's theorem, $\int f(\mathbf{x} - \mathbf{t})g(\mathbf{t})\, d\mathbf{t}$ exists for a.e. $\mathbf{x}$ and is measurable and integrable; also,

$$\iint f(\mathbf{x} - \mathbf{t})g(\mathbf{t})\, d\mathbf{t} d\mathbf{x} = \int \left[\int f(\mathbf{x} - \mathbf{t})g(\mathbf{t})\, d\mathbf{t}\right] d\mathbf{x} = \left[\int f(\mathbf{x})\, d\mathbf{x}\right] \left[\int g(\mathbf{x})\, d\mathbf{x}\right].$$

In particular, $f * g$ is well-defined a.e. and measurable (even integrable), and

$$\int (f * g) = \left(\int f\right) \left(\int g\right).$$

Finally, since $|f*g| \leq |f|*|g|$ wherever $f*g$ exists, we obtain by integration that

$$\int |f*g| \leq \int (|f|*|g|) = \left(\int |f|\right)\left(\int |g|\right),$$

and the proof is complete.

See Exercise 21 of Chapter 9 for a criterion which guarantees measurability of $f*g$.

In the proof of Theorem 6.14, we showed the following useful fact.

Corollary 6.16 *If f and g are nonnegative and measurable on $\mathbf{R^n}$, then $f*g$ is measurable on $\mathbf{R^n}$ and*

$$\int_{\mathbf{R^n}} (f*g)\, d\mathbf{x} = \left(\int_{\mathbf{R^n}} f\, d\mathbf{x}\right)\left(\int_{\mathbf{R^n}} g\, d\mathbf{x}\right).$$

Our final application of Fubini's theorem is an important result due to Marcinkiewicz concerning the structure of closed sets. For simplicity, we will restrict our attention to the one-dimensional case. Given a closed subset F of $\mathbf{R^1}$ and a point x, let

$$\delta(x) = \delta(x, F) = \min\left\{|x-y| : y \in F\right\}$$

denote the *distance of x from F*. Thus, $\delta(x) = 0$ if and only if $x \in F$. By Theorem 1.10, the complement of F is a union $\bigcup_k (a_k, b_k)$ of disjoint open intervals. At most, two of these intervals can be infinite. The graph of $\delta(x)$ is thus an irregular sawtooth curve: over any finite interval $[a_k, b_k]$, the graph is the sides of the isosceles triangle with base $[a_k, b_k]$ and altitude $\frac{1}{2}(b_k - a_k)$; outside the terminal points of F, the graph is linear. If we move from a point x to a point y, the distance from F cannot increase by more than $|x-y|$. Hence,

$$\left|\delta(x) - \delta(y)\right| \leq |x-y|;$$

that is, δ satisfies a Lipschitz condition.

We shall prove the following theorem. (See also Exercises 7 through 9 and Theorem 9.19.) The result will be used in the proof of Theorem 12.67.

Theorem 6.17 (Marcinkiewicz) Let F be a closed subset of a bounded open interval (a, b), and let $\delta(x) = \delta(x, F)$ be the corresponding distance function. Then, given $\lambda > 0$, the integral

$$M_\lambda(x) = M_\lambda(x;F) = \int_a^b \frac{\delta^\lambda(y)}{|x-y|^{1+\lambda}}\,dy$$

is finite a.e. in F. Moreover, $M_\lambda \in L(F)$ and

$$\int_F M_\lambda \le 2\lambda^{-1}|G|,$$

where $G = (a,b) - F$.

Before the proof, we note that what makes the finiteness of $M_\lambda(x)$ remarkable is the singular behavior of $\delta^\lambda(y)/|x-y|^{1+\lambda}$ as $y \to x$. Since $\delta(y) \to \delta(x)$ as $y \to x$, it follows that $M_\lambda(x) = +\infty$ if $x \notin F$ (see Exercise 9). If $x \in F$, then $\delta(x) = 0$, but the mere Lipschitz character of δ in the estimate

$$\frac{\delta^\lambda(y)}{|x-y|^{1+\lambda}} = \frac{|\delta(y) - \delta(x)|^\lambda}{|x-y|^{1+\lambda}} \le \frac{|x-y|^\lambda}{|x-y|^{1+\lambda}} = \frac{1}{|x-y|}$$

is not enough to imply that $M_\lambda(x)$ is finite since $\int_a^b dy/|x-y| = +\infty$. The key to the convergence of M_λ at a point x is the fact that $\delta(y)$ vanishes not only at x but also at every $y \in F$. Thus, roughly speaking, the finiteness of $M_\lambda(x)$ means that F is *very dense* near x. In this regard, see also Exercise 9(b).

Proof. Measurability of M_λ follows from Corollary 6.16. Since $\delta = 0$ in F, integration in the integral defining M_λ can be restricted to the set $G = (a,b) - F$ without changing M_λ. Thus,

$$\int_F M_\lambda(x)\,dx = \int_G \delta^\lambda(y)\left(\int_F \frac{dx}{|x-y|^{1+\lambda}}\right)dy,$$

the change in the order of integration being justified by Tonelli's theorem. To estimate the inner integral, fix $y \in G$ and note that for any $x \in F$, we have $|x-y| \ge \delta(y) > 0$. Thus,

$$\int_F \frac{dx}{|x-y|^{1+\lambda}} \le \int_{|x-y|\ge\delta(y)} \frac{dx}{|x-y|^{1+\lambda}} = 2\int_{\delta(y)}^\infty \frac{dt}{t^{1+\lambda}} = 2\lambda^{-1}\delta(y)^{-\lambda}.$$

In particular,

$$\int_F M_\lambda(x)\,dx \le \int_G \delta^\lambda(y)\left[2\lambda^{-1}\delta(y)^{-\lambda}\right]dy = 2\lambda^{-1}|G| < +\infty.$$

Exercises

1. (a) Let E be a measurable subset of $\mathbf{R^2}$ such that for almost every $x \in \mathbf{R^1}$, $\{y\colon (x,y) \in E)$ has $\mathbf{R^1}$-measure zero. Show that E has measure zero and that for almost every $y \in \mathbf{R^1}$, $\{x\colon (x,y) \in E\}$ has measure zero.

(b) Let $f(x,y)$ be nonnegative and measurable in $\mathbf{R^2}$. Suppose that for almost every $x \in \mathbf{R^1}, f(x,y)$ is finite for almost every y. Show that for almost every $y \in \mathbf{R^1}, f(x,y)$ is finite for almost every x.

2. If f and g are measurable in $\mathbf{R^n}$, show that the function $h(\mathbf{x},\mathbf{y}) = f(\mathbf{x})g(\mathbf{y})$ is measurable in $\mathbf{R^n}\times\mathbf{R^n}$. Deduce that if E_1 and E_2 are measurable subsets of $\mathbf{R^n}$, then their Cartesian product $E_1 \times E_2 = \{(\mathbf{x},\mathbf{y}) : \mathbf{x} \in E_1, \mathbf{y} \in E_2\}$ is measurable in $\mathbf{R^n} \times \mathbf{R^n}$, and $|E_1 \times E_2| = |E_1||E_2|$. As usual in measure theory, $0 \cdot \infty$ and $\infty \cdot 0$ are interpreted as 0.

3. Let f be measurable and finite a.e. on [0,1]. If $f(x) - f(y)$ is integrable over the square $0 \le x \le 1, 0 \le y \le 1$, show that $f \in L[0,1]$.

4. Let f be measurable and periodic with period 1: $f(t+1) = f(t)$. Suppose that there is a finite c such that

$$\int_0^1 \left|f(a+t) - f(b+t)\right| dt \le c$$

for all a and b. Show that $f \in L[0,1]$. (Set $a = x, b = -x$, integrate with respect to x, and make the change of variables $\xi = x + t$, $\eta = -x + t$.)

5. (a) If f is nonnegative and measurable on E and $\omega(y) = |\{\mathbf{x} \in E : f(\mathbf{x}) > y\}|$, $y > 0$, use Tonelli's theorem to prove that $\int_E f = \int_0^\infty \omega(y)\,dy$. (By definition of the integral, $\int_E f = \left|R(f,E)\right| = \iint_{R(f,E)} d\mathbf{x}dy$. Use the observation in the proof of Theorem 6.11 that $\{\mathbf{x} \in E : f(\mathbf{x}) \ge y\} = \{\mathbf{x} : (x,y) \in R(f,E)\}$, and recall that $\omega(y) = |\{\mathbf{x} \in E : f(\mathbf{x}) \ge y\}|$ unless y is a point of discontinuity of ω.)

(b) Deduce from this special case the general formula

$$\int_E f^p = p\int_0^\infty y^{p-1}\omega(y)\,dy \quad \left(f \ge 0,\ 0 < p < \infty\right).$$

6. For $f \in L(\mathbf{R^1})$, define the *Fourier transform* $\widehat{f}$ of f by

$$\widehat{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(t) e^{-ixt} dt \qquad \left(x \in \mathbf{R^1}\right).$$

(For a complex-valued function $F = F_0 + iF_1$ whose real and imaginary parts F_0 and F_1 are integrable, we define $\int F = \int F_0 + i \int F_1$.) Show that if f and g belong to $L(\mathbf{R^1})$, then

$$\widehat{(f * g)}(x) = 2\pi \widehat{f}(x) \widehat{g}(x).$$

7. Let F be a closed subset of $\mathbf{R^1}$ and let $\delta(x) = \delta(x, F)$ be the corresponding distance function. If $\lambda > 0$ and f is nonnegative and integrable over the complement of F, prove that the function

$$\int_{\mathbf{R^1}} \frac{\delta^{\lambda}(y) f(y)}{|x-y|^{1+\lambda}} dy$$

is integrable over F and so is finite a.e. in F. (In case $f = \chi_{(a,b)}$, this reduces to Theorem 6.17.)

8. Under the hypotheses of Theorem 6.17 and assuming that $b - a < 1$, prove that the function

$$M_0(x) = \int_a^b \left[\log \frac{1}{\delta(y)}\right]^{-1} |x-y|^{-1} dy$$

is finite a.e. in F.

9. (a) Show that $M_\lambda(x; F) = +\infty$ if $x \notin F, \lambda > 0$.

(b) Let $F = [c, d]$ be a closed subinterval of a bounded open interval $(a, b) \subset \mathbf{R^1}$, and let M_λ be the corresponding Marcinkiewicz integral, $\lambda > 0$. Show that M_λ is finite for every $x \in (c, d)$ and that $M_\lambda(c) = M_\lambda(d) = \infty$. Show also that $\int_F M_\lambda \leq \lambda^{-1}|G|$, where $G = (a, b) - [c, d]$.

10. Let v_n be the volume of the unit ball in $\mathbf{R^n}$. Show by using Fubini's theorem that

$$v_n = 2v_{n-1} \int_0^1 \left(1 - t^2\right)^{(n-1)/2} dt.$$

(We also observe that by setting $w = t^2$, the integral is a multiple of a classical β-function and so can be expressed in terms of the Γ-function: $\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt$, $s > 0$.)

11. Use Fubini's theorem to prove that

$$\int_{\mathbf{R}^n} e^{-|\mathbf{x}|^2} d\mathbf{x} = \pi^{n/2}.$$

(For $n = 1$, write $\left(\int_{-\infty}^{+\infty} e^{-x^2} dx\right)^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2-y^2} dxdy$ and use polar coordinates. For $n > 1$, use the formula $e^{-|\mathbf{x}|^2} = e^{-x_1^2} \dots e^{-x_n^2}$ and Fubini's theorem to reduce to the case $n = 1$.)

12. (a) Give an example that shows that the projection onto the x-axis of a measurable subset of the plane may not be linearly measurable.

(b) Show that if E is either an open or closed set in the plane, then its projection onto the x-axis is linearly measurable.

(For (b), the projection of an open set is open, and the projection of a compact set is compact. Any closed set is a countable union of compact sets.)

13. Let $f \in L(-\infty, \infty)$, and let $h > 0$ be fixed. Prove that

$$\int_{-\infty}^{\infty} \left(\frac{1}{2h} \int_{x-h}^{x+h} f(y)\, dy \right) dx = \int_{-\infty}^{\infty} f(x)\, dx.$$

7

Differentiation

The main results in this chapter deal with questions of differentiability. A variety of topics is considered, but for the most part, the results are related to the analogue for Lebesgue integrals of the fundamental theorem of calculus and to the differentiability a.e. of functions that are Lipschitz continuous.

7.1 The Indefinite Integral

If f is a Riemann integrable function on an interval $[a, b]$ in $\mathbf{R^1}$, then the familiar definition of its indefinite integral is

$$F(x) = \int_a^x f(y)\, dy, \quad a \le x \le b.$$

The fundamental theorem of calculus asserts that $F' = f$ if f is continuous. We will study an analogue of this result for Lebesgue integrable f and higher dimensions.

We must first find an appropriate definition of the indefinite integral. In two dimensions, for example, we might choose

$$F(x_1, x_2) = \int_{a_1}^{x_1} \int_{a_2}^{x_2} f(y_1, y_2)\, dy_1 dy_2.$$

It turns out, however, to be better to abandon the notion that the indefinite integral be a function of point and adopt the idea that it be a function of set. Thus, given $f \in L(A)$, where A is a measurable subset of $\mathbf{R^n}$, we define the *indefinite integral of f* to be the function

$$F(E) = \int_E f,$$

where E is any measurable subset of A.

F is an example of a *set function*, by which we mean any real-valued function F defined on a σ-algebra Σ of measurable sets such that

(i) $F(E)$ is finite for every $E \in \Sigma$.

(ii) F is *countably additive*; that is, if $E = \bigcup_k E_k$ is a union of disjoint $E_k \in \Sigma$, then

$$F(E) = \sum_k F(E_k).$$

By Theorems 5.5 and 5.24, the indefinite integral of an $f \in L(A)$ satisfies (i) and (ii) for the σ-algebra of measurable subsets of A. We shall systematically study set functions in Chapter 10.

We now discuss a continuity property of the indefinite integral. Recall (from p. 5 in Section 1.3) that the diameter of a set E is the value

$$\sup\{|\mathbf{x} - \mathbf{y}| : \mathbf{x}, \mathbf{y} \in E\}.$$

A set function $F(E)$ is called *continuous* if $F(E)$ tends to zero as the diameter of E tends to zero; that is, $F(E)$ is continuous if, given $\varepsilon > 0$, there exists $\delta > 0$ such that $|F(E)| < \varepsilon$ whenever the diameter of E is less than δ. An example of a set function that is *not* continuous can be obtained by setting $F(E) = 1$ for any measurable E that contains the origin, and $F(E) = 0$ otherwise.

A set function F is called *absolutely continuous with respect to Lebesgue measure*, or simply *absolutely continuous*, if $F(E)$ tends to zero as the *measure* of E tends to zero. Thus, F is absolutely continuous if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|F(E)| < \varepsilon$ whenever the measure of E is less than δ.

A set function that is absolutely continuous is clearly continuous. The converse, however, is false, as shown by the following example. Let A be the unit square in $\mathbf{R}^2$, let D be a diagonal of A, and consider the σ-algebra of measurable subsets E of A for which $E \cap D$ is linearly measurable. For such E, let $F(E)$ be the linear measure of $E \cap D$. Then F is a continuous set function. However, it is not absolutely continuous since there are sets E containing a fixed segment of D whose $\mathbf{R}^2$-measures are arbitrarily small.

Theorem 7.1 *If $f \in L(A)$, its indefinite integral is absolutely continuous.*

Proof. We may assume that $f \geq 0$ by considering f^+ and f^-. Fix k and write $f = g + h$, where $g = f$ whenever $f \leq k$ and $g = k$ otherwise. Given $\varepsilon > 0$, we may choose k so large that $0 \leq \int_A h < \frac{1}{2}\varepsilon$ and, *a fortiori*, $0 \leq \int_E h < \frac{1}{2}\varepsilon$ for every measurable $E \subset A$. On the other hand, since $0 \leq g \leq k$, we have $0 \leq \int_E g \leq k|E| < \frac{1}{2}\varepsilon$ if $|E|$ is small enough. Thus,

$$0 \le \int_E f = \int_E g + \int_E h < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon$$

if $|E|$ is small enough. This completes the proof.

We remark here that Theorem 7.1 has the following converse: If $F(E)$ is a set function that is absolutely continuous with respect to Lebesgue measure, then there exists an integrable f such that $F(E) = \int_E f$ for measurable E. A proof of this fact, known as the Radon–Nikodym theorem, is given in Chapter 10.

In the case of the real line, there is an alternate notion, also termed absolute continuity, which pertains to ordinary functions. This notion and its relation to the integral $\int_a^x f(y)\,dy$ are discussed in Section 7.5.

7.2 Lebesgue's Differentiation Theorem

We now come to a fundamental theorem of Lebesgue concerning differentiation of the indefinite integral. For $f \in L(\mathbf{R^n})$, let F be the indefinite integral of f, and let Q denote an n-dimensional cube with edges parallel to the coordinate axes. Given $\mathbf{x}$, we consider those Q centered at $\mathbf{x}$ and ask whether the average

$$\frac{F(Q)}{|Q|} = \frac{1}{|Q|}\int_Q f(\mathbf{y})\,d\mathbf{y}$$

converges to $f(\mathbf{x})$ as Q contracts to $\mathbf{x}$. If this is the case, we write

$$\lim_{Q \searrow x} \frac{F(Q)}{|Q|} = f(\mathbf{x})$$

and say that the indefinite integral of f is differentiable at $\mathbf{x}$ with derivative $f(\mathbf{x})$. In case $n = 1$, the question is whether

$$\lim_{h\to 0} \frac{1}{2h}\int_{x-h}^{x+h} f(y)\,dy = f(x),$$

which we shall later see is essentially equivalent to

$$\lim_{h\to 0} \frac{1}{h}\int_{x}^{x+h} f(y)\,dy = f(x),$$

that is, to $d/dx \int_a^x f(y)\,dy = f(x)$.

Since f can be changed arbitrarily in a set of measure zero without affecting its indefinite integral, the best we can hope for is that F is differentiable to f almost everywhere. This is actually the case.

Theorem 7.2 (Lebesgue's Differentiation Theorem) *If $f \in L(\mathbf{R^n})$, its indefinite integral is differentiable with derivative $f(\mathbf{x})$ at almost every $\mathbf{x} \in \mathbf{R^n}$.*

The proof of this basic result is difficult and requires several new ideas with wide applications. One of them is to consider the function

$$f^*(\mathbf{x}) = \sup \frac{1}{|Q|} \int_Q |f(\mathbf{y})|\, d\mathbf{y},$$

where the sup is taken over all Q with center $\mathbf{x}$. This function plays an important role in analysis.

Let us first observe that the theorem is easy to prove for continuous functions. In fact, if f is continuous at $\mathbf{x}$ and Q is a cube with center $\mathbf{x}$, then

$$\begin{aligned} \left| \frac{1}{|Q|} \int_Q f(\mathbf{y}) d\mathbf{y} - f(\mathbf{x}) \right| &= \left| \frac{1}{|Q|} \int_Q [f(\mathbf{y}) - f(\mathbf{x})]\, d\mathbf{y} \right| \\ &\le \frac{1}{|Q|} \int_Q |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y} \le \sup_{\mathbf{y} \in Q} |f(\mathbf{y}) - f(\mathbf{x})|, \end{aligned}$$

which tends to zero as Q shrinks to $\mathbf{x}$.

The strategy of the proof is to approximate a given $f \in L(\mathbf{R^n})$ by continuous functions C_k. This approximation is stated in Lemma 7.3 and is *global* in nature. Hence, it will be necessary to find a way to control the *local* behavior (i.e., the averages) of $f - C_k$ by this global estimate. This step is carried out in Lemma 7.9 and consists of estimating the size of f^* in terms of $\int |f|$. Lemma 7.4 is a crucial covering lemma used to prove Lemma 7.9.

Lemma 7.3 *If $f \in L(\mathbf{R^n})$, there exists a sequence $\{C_k\}$ of continuous functions with compact support such that*

$$\int_{\mathbf{R^n}} |f - C_k|\, d\mathbf{x} \to 0 \quad \text{as } k \to \infty.$$

Proof. If f is an integrable function for which the conclusion holds, we will say that f has property $\mathscr{A}$. We will prove the lemma by considering a

series of special cases. To help in passing from one case to the next, we first show that

(1) A finite linear combination of functions with property $\mathscr{A}$ has property $\mathscr{A}$.
(2) If $\{f_k\}$ is a sequence of functions with property $\mathscr{A}$, and if $\int_{\mathbf{R}^n} |f - f_k| \to 0$, then f has property $\mathscr{A}$.

To prove (1), it is enough to show that any constant multiple, af, of a function with property $\mathscr{A}$ has property $\mathscr{A}$ and that the sum, $f_1 + f_2$, of two functions with property $\mathscr{A}$ has property $\mathscr{A}$. These facts follow easily from the relations

$$\int |af - aC| = |a| \int |f - C|,$$
$$\int |(f_1 + f_2) - (C_1 + C_2)| \leq \int |f_1 - C_1| + \int |f_2 - C_2|.$$

To prove (2), let $\{f_k\}$ and f satisfy the hypotheses of (2). First note that since f_k is integrable and $\int |f| \leq \int |f - f_k| + \int |f_k|$, it follows that f is integrable. Next, given $\varepsilon > 0$, choose k_0 so that $\int |f - f_{k_0}| < \varepsilon/2$. Then choose a continuous C with compact support such that $\int |f_{k_0} - C| < \varepsilon/2$. Since

$$\int |f - C| \leq \int |f - f_{k_0}| + \int |f_{k_0} - C| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

we see that f has property $\mathscr{A}$.

To prove the lemma, let $f \in L(\mathbf{R}^n)$. Writing $f = f^+ - f^-$, we may assume by (1) that $f \geq 0$. Then, by Theorem 4.13, there exist nonnegative simple $f_k \nearrow f$. Thus, $f_k \in L(\mathbf{R}^n)$ and $\int |f - f_k| \to 0$, so that by (2), we may suppose that f is an integrable simple function. Hence, by (1), we may assume that $f = \chi_E$ for a set E with $|E| < +\infty$. Given $\varepsilon > 0$, choose an open G such that $E \subset G$ and $|G - E| < \varepsilon$. Then

$$\int |\chi_G - \chi_E| = |G - E| < \varepsilon,$$

so we may assume that $f = \chi_G$ for an open G with $|G| < +\infty$. Using Theorem 1.11, write $G = \bigcup I_k$, where the I_k are disjoint, partly open intervals. If we let f_N be the characteristic function of $\bigcup_{k=1}^N I_k$, we obtain

$$\int |f - f_N| = \sum_{k=N+1}^{\infty} |I_k| \to 0$$

since $\sum_{k=1}^{\infty} |I_k| = |G| < +\infty$. By (2), it is thus enough to show that each f_N has property $\mathscr{A}$. But f_N is the sum of $\chi_{I_k}, k = 1, \ldots, N$, so it suffices by (1) to show

that the characteristic function of any partly open interval I has property $\mathscr{A}$. This is practically self-evident: if I^* denotes an interval that contains the closure of I in its interior and that satisfies $|I^* - I| < \varepsilon$, then for any continuous $C, 0 \leq C \leq 1$, which is 1 in I and 0 outside I^*, we have

$$\int |\chi_I - C| \leq \left|I^* - I\right| < \varepsilon.$$

This completes the proof of Lemma 7.3. The proof also shows that the functions $\{C_k\}$ can be chosen to be finite linear combinations of characteristic functions of intervals (i.e., step functions) instead of continuous functions.

The lemma that follows is a preliminary version of a covering lemma due to Vitali (Theorem 7.17) and has many applications.

Lemma 7.4 (Simple Vitali Lemma) *Let E be a subset of $\mathbf{R^n}$ with $|E|_e < +\infty$, and let K be a collection of cubes Q covering E. Then there exist a positive constant β, depending only on n, and a finite number of disjoint cubes $Q_1, \ldots, Q_N$ in K such that*

$$\sum_{j=1}^{N} \left|Q_j\right| \geq \beta |E|_e.$$

Proof. We will index the size of a cube $Q \in K$ by writing $Q = Q(t)$, where t is the edge length of Q. Let $K_1 = K$ and

$$t_1^* = \sup \{t : Q = Q(t) \in K_1\}.$$

If $t_1^* = +\infty$, then K_1 contains a sequence of cubes Q with $|Q| \to +\infty$. In this case, given $\beta > 0$, we simply choose one $Q \in K_1$ with $|Q| \geq \beta |E|_e$. If $t_1^* < +\infty$, the idea is still to pick a relatively large cube: choose $Q_1 = Q_1(t_1) \in K_1$ such that $t_1 > \frac{1}{2} t_1^*$. Now split $K_1 = K_2 \cup K_2'$, where K_2 consists of those cubes in K_1 that are disjoint from Q_1, and K_2' of those that intersect Q_1. Let Q_1^* denote the cube concentric with Q_1 whose edge length is $5t_1$. Thus, $\left|Q_1^*\right| = 5^n |Q_1|$, and since $2t_1 > t_1^*$, every cube in K_2' is contained in Q_1^*.

Starting with $j = 2$, continue this selection process for $j = 2, 3, \ldots$, by letting

$$t_j^* = \sup \left\{t : Q = Q(t) \in K_j\right\},$$

choosing a cube $Q_j = Q_j\left(t_j\right) \in K_j$ with $t_j > \frac{1}{2} t_j^*$, and splitting $K_j = K_{j+1} \cup K_{j+1}'$, where K_{j+1} consists of all those cubes of K_j that are disjoint from Q_j. If K_{j+1} is empty, the process ends. We have $t_j^* \geq t_{j+1}^*$; moreover, for each j, the $Q_1, \ldots, Q_j$

are disjoint from one another and from every cube in K_{j+1}, and every cube in K'_{j+1} is contained in the cube Q_j^* concentric with Q_j whose edge length is $5t_j$. Note that $\left|Q_j^*\right| = 5^n \left|Q_j\right|$.

Consider the sequence $t_1^* \geq t_2^* \geq \cdots$. If some K_{N+1} is empty (i.e., if $t_j^* = 0$ for $j \geq N+1$), then since

$$K_1 = K_2 \cup K'_2 = \cdots = K_{N+1} \cup K'_{N+1} \cup \cdots \cup K'_2,$$

and E is covered by the cubes in K_1, it follows that E is covered by the cubes in $K'_{N+1} \cup \cdots \cup K'_2$. Hence, $E \subset \bigcup_{j=1}^N Q_j^*$, so that

$$|E|_e \leq \sum_{j=1}^N \left|Q_j^*\right| = 5^n \sum_{j=1}^N \left|Q_j\right|.$$

This proves the lemma with $\beta = 5^{-n}$.

On the other hand, if no t_j^* is zero, then either there exists a $\delta > 0$ such that $t_j^* \geq \delta$ for all j, or $t_j^* \to 0$. In the first case, $t_j \geq \frac{1}{2}\delta$ for all j and, therefore, $\sum_{j=1}^N \left|Q_j\right| \to +\infty$ as $N \to \infty$. Given any $\beta > 0$, the lemma follows in this case by choosing N sufficiently large.

Finally, if $t_j^* \to 0$, we claim that every cube in K_1 is contained in $\bigcup_j Q_j^*$. Otherwise, there would be a cube $Q = Q(t)$ not intersecting any Q_j. Since this cube would belong to every K_j, t would satisfy $t \leq t_j^*$ for every j and, therefore, $t = 0$. This contradiction establishes the claim. Since E is covered by the cubes in K_1, it follows that

$$|E|_e \leq \sum_j \left|Q_j^*\right| = 5^n \sum_j \left|Q_j\right|.$$

Hence, given β with $0 < \beta < 5^{-n}$, there exists an N such that $\sum_{j=1}^N \left|Q_j\right| \geq \beta |E|_e$. This completes the proof.

We stress that Lemma 7.4 does not presuppose the measurability of E and that the proof can be shortened if E is measurable. In fact, if E is measurable, we can suppose it is closed and bounded (see, e.g., Lemma 3.22). Hence, assuming as we may that the cubes in K are open (by slightly enlarging each cube concentrically), it follows from the Heine–Borel Theorem 1.12 that E can be covered by a finite number of cubes. For Q_1, we then choose the largest cube; similarly, in subsequent steps, we take Q_j to be the largest cube disjoint from $Q_1, \ldots, Q_{j-1}$. Thus, $E \subset \bigcup Q_j^*$, and the lemma follows.

See Exercise 18 for a more set-theoretic version of Lemma 7.4.

Before stating the final lemma, we make a definition and a few remarks. If f is defined on $\mathbf{R^n}$ and integrable over every cube Q, let

$$f^*(\mathbf{x}) = \sup \frac{1}{|Q|} \int_Q |f(\mathbf{y})|\, d\mathbf{y}, \tag{7.5}$$

where the supremum is taken over all Q with edges parallel to the coordinate axes and center $\mathbf{x}$. The function f^*, called the *Hardy–Littlewood maximal function of f*, is a gauge of the size of the averages of $|f|$ around $\mathbf{x}$. It clearly satisfies the following:

$$\begin{aligned} &\text{(i)} && 0 \le f^*(\mathbf{x}) \le +\infty, \\ &\text{(ii)} && (f+g)^*(\mathbf{x}) \le f^*(\mathbf{x}) + g^*(\mathbf{x}), \\ &\text{(iii)} && (cf)^*(\mathbf{x}) = |c| f^*(\mathbf{x}). \end{aligned} \tag{7.6}$$

If $f^*(\mathbf{x}_0) > \alpha$ for some $\mathbf{x}_0 \in \mathbf{R^n}$ and $\alpha > 0$, it follows from the absolute continuity of the indefinite integral that $f^*(\mathbf{x}) > \alpha$ for all $\mathbf{x}$ near $\mathbf{x}_0$. Hence, according to Theorem 4.14, f^* is lower semicontinuous in $\mathbf{R^n}$. In particular, it is measurable.

We now investigate the size of f^*. For any measurable E,

$$\chi_E^*(\mathbf{x}) = \sup \left\{ \frac{|E \cap Q|}{|Q|} : Q \text{ has center } \mathbf{x} \right\}.$$

If E is bounded and $Q^{\mathbf{x}}$ denotes the smallest cube with center $\mathbf{x}$ containing E, then

$$\frac{|E \cap Q^{\mathbf{x}}|}{|Q^{\mathbf{x}}|} = \frac{|E|}{|Q^{\mathbf{x}}|}.$$

It follows that there are positive constants c_1 and c_2 such that

$$c_1 \frac{|E|}{|\mathbf{x}|^n} \le \chi_E^*(\mathbf{x}) \le c_2 \frac{|E|}{|\mathbf{x}|^n} \quad \text{for large } |\mathbf{x}|. \tag{7.7}$$

In particular, if $|E| > 0$, χ_E^* is not integrable over $\mathbf{R^n}$. We leave it as an exercise to show that for any measurable f that is different from zero on a set of positive measure, there is a positive constant c such that

$$f^*(\mathbf{x}) \ge \frac{c}{|\mathbf{x}|^n} \text{ for } |\mathbf{x}| \ge 1.$$

Therefore, *f^* is never integrable over $\{\mathbf{x} : |\mathbf{x}| \ge 1\}$ unless $f = 0$* a.e. This failure of integrability of f^* is due to its size when $|\mathbf{x}|$ is large. On the other hand,

f^* is not integrable over bounded sets for *some* $f \in L(\mathbf{R^n})$. For example, in case $n = 1$, the function

$$f(x) = \frac{1}{|x|\ (\log|x|)^2}\,\chi_{\{x:|x|<1/2\}}(x)$$

belongs to $L(\mathbf{R^1})$, but for all x with $0 < |x| < 1/4$,

$$f^*(x) \geq \frac{1}{2|x|}\int_0^{2|x|} \frac{dy}{y\,(\log y)^2} = \frac{1}{2|x|\,|\log 2|x||},$$

and consequently, f^* is not integrable over any neighborhood of the origin.

However, we will see later that f^* is integrable over bounded sets if $f \in L^p(\mathbf{R^n})$ for some $p > 1$, or even if $|f|\left(1 + \log^+|f|\right) \in L^1(\mathbf{R^n})$; see Theorem 9.16 and Exercise 22 of Chapter 9.

To find a way to estimate the size of f^* when $f \in L(\mathbf{R^n})$, recall that by Tchebyshev's inequality,

$$\left|\left\{\mathbf{x} \in \mathbf{R^n} : \left|f(\mathbf{x})\right| > \alpha\right\}\right| \leq \frac{1}{\alpha}\int_{\mathbf{R^n}} |f(\mathbf{x})|\,d\mathbf{x}, \quad \alpha > 0.$$

Hence, if $f \in L(\mathbf{R^n})$, there is a constant c independent of α such that

$$\left\{|\mathbf{x} \in \mathbf{R^n} : |f(\mathbf{x})| > \alpha\right\}| \leq \frac{c}{\alpha}, \quad \alpha > 0. \tag{7.8}$$

Any measurable f, integrable or not, for which (7.8) is valid is said to belong to *weak* $L(\mathbf{R^n})$. Thus, any function in $L(\mathbf{R^n})$ is also in weak $L(\mathbf{R^n})$. The function $|\mathbf{x}|^{-n}$ is an example of a function in weak $L(\mathbf{R^n})$, which is not in $L(\mathbf{R^n})$ (see also Exercise 24 in Chapter 5).

Lemma 7.9 (Hardy–Littlewood) *If $f \in L(\mathbf{R^n})$, then f^* belongs to weak $L(\mathbf{R^n})$. Moreover, there is a constant c independent of f and α such that*

$$\left|\left\{\mathbf{x} \in \mathbf{R^n} : f^*(\mathbf{x}) > \alpha\right\}\right| \leq \frac{c}{\alpha}\int_{\mathbf{R^n}} |f|, \quad \alpha > 0.$$

Proof. Fix $\alpha > 0$ and let

$$E = \left\{f^* > \alpha\right\}.$$

If $\mathbf{x} \in E$, then by the definitions of E and f^*, there is a cube $Q_{\mathbf{x}}$ with center $\mathbf{x}$ such that $|Q_{\mathbf{x}}|^{-1} \int_{Q_{\mathbf{x}}} |f| > \alpha$. Equivalently,

$$|Q_{\mathbf{x}}| < \frac{1}{\alpha} \int_{Q_{\mathbf{x}}} |f|.$$

The collection of such $Q_{\mathbf{x}}$ covers E. For $k = 1, 2, \ldots$, the sets E_k defined by $E_k = E \cap \{\mathbf{x} : |\mathbf{x}| < k\}$ are also covered and have finite measure. By Lemma 7.4 applied to each E_k, there exist $\beta > 0$ (depending only on n) and a finite number of points $\left\{\mathbf{x}_j^{(k)}\right\}_j \subset E$ such that the cubes $Q_{\mathbf{x}_j^{(k)}}$ are disjoint in j (for each k) and $|E_k| < \beta^{-1} \sum_j \left|Q_{\mathbf{x}_j^{(k)}}\right|$. Therefore,

$$|E_k| < \frac{1}{\beta} \sum_j \frac{1}{\alpha} \int_{Q_{\mathbf{x}_j^{(k)}}} |f| = \frac{1}{\beta\alpha} \int_{\bigcup_j Q_{\mathbf{x}_j^{(k)}}} |f| \leq \frac{1}{\beta\alpha} \int_{\mathbf{R^n}} |f|.$$

Since $E_k \nearrow E$ as $k \to \infty$, it follows from Theorem 3.26 that

$$|E| \leq \frac{1}{\beta\alpha} \int_{\mathbf{R^n}} |f|,$$

which proves the lemma with $c = \beta^{-1}$.

Proof of Lebesgue's theorem. Given $f \in L(\mathbf{R^n})$, there exists by Lemma 7.3 a sequence of continuous integrable C_k such that $\int_{\mathbf{R^n}} |f - C_k| \to 0$. Let $F(Q) = \int_Q f$ and $F_k(Q) = \int_Q C_k$. Then for any k,

$$\limsup_{Q \searrow \mathbf{x}} \left| \frac{F(Q)}{|Q|} - f(\mathbf{x}) \right| \leq \limsup_{Q \searrow \mathbf{x}} \left| \frac{F(Q)}{|Q|} - \frac{F_k(Q)}{|Q|} \right|$$
$$+ \limsup_{Q \searrow \mathbf{x}} \left| \frac{F_k(Q)}{|Q|} - C_k(\mathbf{x}) \right| + \left| C_k(\mathbf{x}) - f(\mathbf{x}) \right|,$$

where the lim sup is taken for cubes with center $\mathbf{x}$ that shrink to $\mathbf{x}$. Since C_k is continuous, the second term on the right is zero. Moreover,

$$\left| \frac{F(Q)}{|Q|} - \frac{F_k(Q)}{|Q|} \right| \leq \frac{1}{|Q|} \int_Q |f - C_k| \leq (f - C_k)^* (\mathbf{x}),$$

and therefore, the first term on the right is majorized by $\left(f - C_k\right)^* (\mathbf{x})$. Hence, for every k,

$$\limsup_{Q \searrow \mathbf{x}} \left| \frac{F(Q)}{|Q|} - f(\mathbf{x}) \right| \leq \left(f - C_k\right)^* (\mathbf{x}) + \left| f(\mathbf{x}) - C_k(\mathbf{x}) \right|. \tag{7.10}$$

Given $\varepsilon > 0$, let E_ε be the set on which the left side of (7.10) exceeds ε. By (7.10),

$$E_\varepsilon \subset \left\{ \mathbf{x} \colon \left(f - C_k\right)^* (\mathbf{x}) > \frac{\varepsilon}{2} \right\} \cup \left\{ \mathbf{x} \colon \left| f(\mathbf{x}) - C_k(\mathbf{x}) \right| > \frac{\varepsilon}{2} \right\}.$$

Applying Lemma 7.9 to the first set on the right and Tchebyshev's inequality to the second, we obtain

$$|E_\varepsilon|_e \leq c \left(\frac{\varepsilon}{2}\right)^{-1} \int_{\mathbf{R}^n} |f - C_k| + \left(\frac{\varepsilon}{2}\right)^{-1} \int_{\mathbf{R}^n} |f - C_k|.$$

Since c is independent of k, it follows by letting $k \to \infty$ that $|E_\varepsilon|_e = 0$. Let E be the set where the left side of (7.10) is positive. Then $E = \bigcup_k E_{\varepsilon_k}$ for any sequence $\varepsilon_k \to 0$, and therefore $|E| = 0$. This means that $\lim_{Q \searrow \mathbf{x}} F(Q)/|Q|$ exists and equals $f(\mathbf{x})$ for almost every $\mathbf{x}$, which completes the proof.

We now list several extensions and corollaries of Lebesgue's theorem.

(I) A measurable function f defined on $\mathbf{R}^n$ is said to be *locally integrable on* $\mathbf{R}^n$ if it is integrable over every bounded measurable subset of $\mathbf{R}^n$. This is easily seen to be equivalent to the integrability of f over every compact set.

Theorem 7.11 *The conclusion of Lebesgue's theorem is valid if, instead of being integrable, f is locally integrable on* $\mathbf{R}^n$.

Proof. It is enough to show that the conclusion holds a.e. in every open ball. Fix a ball and replace f by zero outside it. This new function is integrable over $\mathbf{R}^n$, its integral is differentiable a.e., and since differentiability is a local property, the initial function f is differentiable a.e. in the ball. This completes the proof.

(II) For any measurable E, note that

$$\frac{1}{|Q|} \int_Q \chi_E = \frac{|E \cap Q|}{|Q|}.$$

By Theorem 7.11, the left-hand side tends to $\chi_E(\mathbf{x})$ a.e. as $Q \searrow \mathbf{x}$; that is,

$$\lim_{Q \searrow \mathbf{x}} \frac{|E \cap Q|}{|Q|} = \chi_E(\mathbf{x}) \quad \text{a.e.} \tag{7.12}$$

A point $\mathbf{x}$ for which this limit is 1 is called a *point of density* of E, and a point for which it is zero is called a *point of dispersion* of E. Since

$$\frac{|Q \cap E|}{|Q|} + \frac{|Q \cap CE|}{|Q|} = \frac{|Q|}{|Q|} = 1,$$

every point of density of E is a point of dispersion of CE, and vice versa. Formula (7.12) can be restated as follows.

Theorem 7.13 *Let E be a measurable set. Then almost every point of E is a point of density of E.*

Thus, roughly speaking, a set *clusters* around almost all of its points.

(III) The formula $\lim_{Q \searrow \mathbf{x}} (1/|Q|) \int_Q f(\mathbf{y}) d\mathbf{y} = f(\mathbf{x})$ can be written

$$\lim_{Q \searrow \mathbf{x}} \frac{1}{|Q|} \int_Q [f(\mathbf{y}) - f(\mathbf{x})]\, d\mathbf{y} = 0$$

and is valid for almost every $\mathbf{x}$ if f is locally integrable. A point $\mathbf{x}$ at which the stronger statement

$$\lim_{Q \searrow \mathbf{x}} \frac{1}{|Q|} \int_Q |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y} = 0 \tag{7.14}$$

is valid is called a *Lebesgue point of f*, and the collection of all such points is called the *Lebesgue set of f.*

Theorem 7.15 *Let f be locally integrable in $\mathbf{R^n}$. Then almost every point of $\mathbf{R^n}$ is a Lebesgue point of f; that is, there exists a set Z (depending on f) of measure zero such that (7.14) holds for $\mathbf{x} \notin Z$.*

Proof. Let $\{r_k\}$ be the rational numbers, and let Z_k be the set where the formula

$$\lim_{Q \searrow \mathbf{x}} \frac{1}{|Q|} \int_Q \left| f(\mathbf{y}) - r_k \right| d\mathbf{y} = \left| f(\mathbf{x}) - r_k \right|$$

is *not* valid. Since $\left|f(\mathbf{y}) - r_k\right|$ is locally integrable, we have $|Z_k| = 0$. Let $Z = \bigcup Z_k$; then $|Z| = 0$. For any Q, $\mathbf{x}$, and r_k,

$$\begin{aligned}\frac{1}{|Q|}\int_Q |f(\mathbf{y}) - f(\mathbf{x})|\,d\mathbf{y} &\le \frac{1}{|Q|}\int_Q \left|f(\mathbf{y}) - r_k\right|\,d\mathbf{y} + \frac{1}{|Q|}\int_Q \left|f(\mathbf{x}) - r_k\right|\,d\mathbf{y}\\ &= \frac{1}{|Q|}\int_Q \left|f(\mathbf{y}) - r_k\right|\,d\mathbf{y} + \left|f(\mathbf{x}) - r_k\right|.\end{aligned}$$

Therefore, if $\mathbf{x} \notin Z$,

$$\limsup_{Q \searrow \mathbf{x}} \frac{1}{|Q|}\int_Q |f(\mathbf{y}) - f(\mathbf{x})|\,d\mathbf{y} \le 2|f(\mathbf{x}) - r_k|$$

for every r_k. For an $\mathbf{x}$ at which $f(\mathbf{x})$ is finite (in particular, almost everywhere), we can choose r_k such that $\left|f(\mathbf{x}) - r_k\right|$ is arbitrarily small. This shows that the left side of the last formula is zero a.e. and completes the proof.

(IV) So far, the sets contracting to $\mathbf{x}$ have been cubes centered at $\mathbf{x}$ with edges parallel to the coordinate axes. Many other sets can be used. A family $\{S\}$ of measurable sets is said to *shrink regularly to* $\mathbf{x}$ provided

(i) The diameters of the sets S tend to zero.

(ii) If Q is the smallest cube with center $\mathbf{x}$ containing S, there is a constant k independent of S such that

$$|Q| \le k|S|.$$

The sets S need not contain $\mathbf{x}$.

Theorem 7.16 *Let f be locally integrable in $\mathbf{R^n}$. Then at every point $\mathbf{x}$ of the Lebesgue set of f (in particular, almost everywhere),*

$$\frac{1}{|S|}\int_S |f(\mathbf{y}) - f(\mathbf{x})|\,d\mathbf{y} \to 0$$

for any family $\{S\}$ that shrinks regularly to $\mathbf{x}$. Thus, also

$$\frac{1}{|S|}\int_S f(\mathbf{y})\,d\mathbf{y} \to f(\mathbf{x}) \quad \textit{a.e.}$$

Proof. If $S \subset Q$, we have

$$\int_S |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y} \le \int_Q |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y}.$$

Hence, if $\{S\}$ shrinks regularly to $\mathbf{x}$ and Q is the least cube with center $\mathbf{x}$ containing S, then

$$\begin{aligned}\frac{1}{|S|}\int_S |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y} &\le \frac{|Q|}{|S|}\frac{1}{|Q|}\int_Q |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y}\\ &\le k\frac{1}{|Q|}\int_Q |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y}.\end{aligned}$$

If $\mathbf{x}$ is a Lebesgue point of f, the last expression tends to zero, and the theorem follows.

In particular, for functions of a single variable, we obtain (cf. p. 131 in Section 7.2)

$$\lim_{h\to 0}\frac{1}{h}\int_x^{x+h} f(y)\, dy = f(x) \quad \text{a.e.}$$

7.3 Vitali Covering Lemma

The theorem that follows is a refinement of the simple Vitali Lemma 7.4. Given a set and a collection of cubes, we will now assume that each point of the set is covered not just by a single cube in the collection but by a sequence of cubes in the collection with diameters tending to zero. In this case, it turns out that we can cover *almost all* points of the set by a sequence of disjoint cubes in the collection. The result will be essentially a corollary of Lemma 7.4.

A family K of cubes is said to cover a set E *in the Vitali sense* if for every $\mathbf{x} \in E$ and $\eta > 0$, there is a cube in K containing $\mathbf{x}$ whose diameter is less than η.

Theorem 7.17 (Vitali Covering Lemma) *Suppose that E is covered in the Vitali sense by a family K of cubes and that $0 < |E|_e < +\infty$. Then, given $\varepsilon > 0$, there is a sequence $\{Q_j\}$ of disjoint cubes in K such that*

$$\left|E - \bigcup_j Q_j\right| = 0 \quad \text{and} \quad \sum_j |Q_j| < (1+\varepsilon)|E|_e.$$

Proof. The second relation is automatically satisfied if we choose an open set G containing E with $|G| < (1+\varepsilon)|E|_e$ and consider only those Q in K which lie in G.

By Lemma 7.4, there exist a constant $\beta, 0 < \beta < 1$, depending only on the dimension, and disjoint $Q_1, \ldots, Q_{N_1}$ in K such that $\sum_{j=1}^{N_1} |Q_j| > \beta|E|_e$. Therefore,

$$\left|E - \bigcup_{j=1}^{N_1} Q_j\right|_e \leq \left|G - \bigcup_{j=1}^{N_1} Q_j\right| = |G| - \sum_{j=1}^{N_1} |Q_j| < |E|_e(1+\varepsilon-\beta).$$

Hence, by considering from the start only those ε with $0 < \varepsilon < \beta/2$, we have

$$\left|E - \bigcup_{j=1}^{N_1} Q_j\right|_e < |E|_e\left(1 - \frac{\beta}{2}\right).$$

Thus, the part of E not covered by the cubes obtained from the simple Vitali lemma has outer measure less than $|E|_e(1-\beta/2)$. We now repeat the process for the set $E_1 = E - \bigcup_{j=1}^{N_1} Q_j$, which is still covered in the Vitali sense by those cubes in K that are disjoint from $Q_1, \ldots, Q_{N_1}$. We obtain $Q_{N_1+1}, \ldots, Q_{N_2}$, disjoint from each other and from $Q_1, \ldots, Q_{N_1}$, such that

$$\left|E - \bigcup_{j=1}^{N_2} Q_j\right|_e = \left|E_1 - \bigcup_{j=N_1+1}^{N_2} Q_j\right|_e < |E_1|_e\left(1 - \frac{\beta}{2}\right) < |E|_e\left(1 - \frac{\beta}{2}\right)^2.$$

Continuing in this way, we obtain at the mth stage disjoint $Q_1, \ldots, Q_{N_m}$ in K such that

$$\left|E - \bigcup_{j=1}^{N_m} Q_j\right|_e < |E|_e\left(1 - \frac{\beta}{2}\right)^m.$$

Since $(1-\beta/2)^m \to 0$ as $m \to \infty$, there is a sequence of disjoint cubes in K with the desired properties.

Corollary 7.18 *Suppose that E is covered in the Vitali sense by a family K of cubes and $0 < |E|_e < +\infty$. Then, given $\varepsilon > 0$, there is a finite collection $Q_1, \ldots, Q_N$ of disjoint cubes in K such that*

$$\left|E - \bigcup_{j=1}^{N} Q_j\right|_e < \varepsilon \quad \text{and} \quad \sum_{j=1}^{N} \left|Q_j\right| < (1+\varepsilon)|E|_e.$$

This is part of the proof of Vitali's lemma.

Note by Carathéodory's theorem that

$$|E|_e = \left|E - \bigcup_{j=1}^{N} Q_j\right|_e + \left|E \cap \bigcup_{j=1}^{N} Q_j\right|_e,$$

so that for E and $\{Q_j\}_{j=1}^{N}$ as in Corollary 7.18, we have

$$|E|_e - \varepsilon < \left|E \cap \bigcup_{j=1}^{N} Q_j\right|_e. \tag{7.19}$$

In particular,

$$|E|_e - \varepsilon < \sum_{j=1}^{N} \left|Q_j\right|. \tag{7.20}$$

7.4 Differentiation of Monotone Functions

As an application of Vitali's covering lemma, we will prove a basic result concerning the differentiability of monotone functions on $\mathbf{R}^1$. If $f(x)$, $x \in \mathbf{R}^1$, is a real-valued function defined and finite in a neighborhood of x_0, consider the four *Dini numbers* (or *derivates*)

$$D_1 f(x_0) = \limsup_{h \to 0+} \frac{f(x_0+h) - f(x_0)}{h},$$

$$D_2 f(x_0) = \liminf_{h \to 0+} \frac{f(x_0+h) - f(x_0)}{h},$$

$$D_3 f(x_0) = \limsup_{h \to 0-} \frac{f(x_0+h) - f(x_0)}{h},$$

$$D_4 f(x_0) = \liminf_{h \to 0-} \frac{f(x_0+h) - f(x_0)}{h}.$$

Clearly, $D_2 f \leq D_1 f$ and $D_4 f \leq D_3 f$. If all four Dini numbers are equal, that is, if $\lim_{h\to 0}[f(x_0+h)-f(x_0)]/h$ exists, finite or infinite, we say that f *has a derivative at* x_0 and call the common value the *derivative* $f'(x_0)$ at x_0. Thus, $-\infty \leq f'(x_0) \leq +\infty$ if $f'(x_0)$ exists.

Theorem 7.21 *Let f be monotone increasing and finite on an open interval $(a,b) \subset \mathbf{R}^1$. Then f has a measurable, nonnegative, finite derivative f' almost everywhere in (a,b). Moreover,*

$$0 \leq \int_a^b f' \leq f(b-) - f(a+). \tag{7.22}$$

Proof. We may assume that (a,b) is finite; the general case follows from this by passage to the limit. We will show that the set $\{x \in (a,b) : D_1 f(x) > D_4 f(x)\}$ has measure zero. A similar argument will apply to any two Dini numbers of f. It is enough to show that each set

$$A_{r,s} = \{x \in (a,b) : D_1 f(x) > r > s > D_4 f(x)\}$$

has measure zero since the original set is the union of these over rational r and s. We may assume $r, s > 0$ since f is increasing.

Fix r and s with $r > s > 0$, write $A = A_{r,s}$, and suppose that $|A|_e > 0$. If $x \in A$, the fact that $D_4 f(x) < s$ implies the existence of arbitrarily small $h > 0$ such that

$$\frac{f(x-h)-f(x)}{-h} < s.$$

By Corollary 7.18 and (7.19), given $\varepsilon > 0$, there exist disjoint intervals $[x_j - h_j, x_j]$, $j = 1, \ldots, N$, such that

(i) $f(x_j) - f(x_j - h_j) < s h_j$, $j = 1, \ldots, N$,

(ii) $\left| A \cap \bigcup_{j=1}^N [x_j - h_j, x_j] \right|_e > |A|_e - \varepsilon$,

(iii) $\sum_{j=1}^N h_j < (1+\varepsilon)|A|_e$.

Combining (i) and (iii), we obtain

(iv) $\sum_{j=1}^N [f(x_j) - f(x_j - h_j)] < s(1+\varepsilon)|A|_e$.

Let $B = A \cap \bigcup_{j=1}^N [x_j - h_j, x_j]$. By (ii), $|B|_e > |A|_e - \varepsilon$. For every $y \in B$ that is not an endpoint of some $[x_j - h_j, x_j]$, the fact that $D_1 f(y) > r$ implies that there

exist arbitrarily small $k > 0$ such that $[y, y+k]$ lies in some $\left[x_j - h_j, x_j\right]$ and

$$\frac{f(y+k) - f(y)}{k} > r.$$

Hence, by Corollary 7.18 and (7.20), there exist disjoint $[y_i, y_i + k_i]$, $i = 1, \ldots, M$, such that

(v) Each $[y_i, y_i + k_i]$ lies in some $\left[x_j - h_j, x_j\right]$,

(vi) $f\left(y_i + k_i\right) - f\left(y_i\right) > rk_i$, $i = 1, \ldots, M$,

(vii) $\sum_{i=1}^{M} k_i > |B|_e - \varepsilon > |A|_e - 2\varepsilon$ [by (ii)].

Therefore,

(viii) $\sum_{i=1}^{M} \left[f\left(y_i + k_i\right) - f\left(y_i\right)\right] > r\left(|A|_e - 2\varepsilon\right)$.

Since f is increasing, it follows from (v) that

$$\sum_{i=1}^{M} \left[f\left(y_i + k_i\right) - f\left(y_i\right)\right] \leq \sum_{j=1}^{N} \left[f\left(x_j\right) - f\left(x_j - h_j\right)\right].$$

Combining this inequality with (iv) and (viii), we obtain $r\left(|A|_e - 2\varepsilon\right) < s(1+\varepsilon)|A|_e$. Since $\varepsilon > 0$ is arbitrary, this gives $r \leq s$, which is a contradiction. Hence, $|A|_e = 0$.

Since an analogous argument applies to any two Dini numbers, it follows that $f'(x)$ exists for almost every x in (a, b). Extend the definition of f to (a, ∞) by setting $f(x) = f(b-)$ for $x \geq b$, and let

$$f_k(x) = \frac{f(x+h) - f(x)}{h}, \quad h = \frac{1}{k},$$

for $x \in (a, b)$ and $k = 1, 2, \ldots$. Each f_k is nonnegative and measurable, and $f_k \to f'$ a.e. in (a, b). Hence, f' is measurable on (a, b), and by Fatou's lemma,

$$\int_a^b f' \leq \liminf_{k \to \infty} \int_a^b f_k.$$

If $f(b-)$ is finite (otherwise, (7.22) is obvious), we have

$$\begin{aligned}\int_a^b f_k &= \frac{1}{h} \int_{a+h}^{b+h} f - \frac{1}{h} \int_a^b f \\ &= \frac{1}{h} \int_b^{b+h} f - \frac{1}{h} \int_a^{a+h} f = f(b-) - \frac{1}{h} \int_a^{a+h} f.\end{aligned}$$

Since $f(a+) \le (1/h)\int_a^{a+h} f \le f(a+h)$, we obtain

$$\int_a^b f' \le \liminf_{k\to\infty} \int_a^b f_k = \lim_{k\to\infty} \int_a^b f_k = f(b-) - f(a+),$$

which proves (7.22). It remains to show that f' is finite a.e. in (a,b). This follows from (7.22) if $f(b-) - f(a+) < \infty$ since then $f' \in L(a,b)$. Since f is finite on (a,b), it follows in any case by applying (7.22) to a sequence of intervals (a_k, b_k) increasing to (a,b), $a < a_k < b_k < b$, and the proof is complete.

The inequality in formula (7.22) cannot in general be replaced by equality, even if f is continuous on $[a,b]$. To see this, let f be the Cantor–Lebesgue function on $[0,1]$ (p. 43 in Section 3.1). Then f is continuous and monotone increasing on $[0,1]$, $f(0) = 0$, $f(1) = 1$, and since f is constant on every interval removed in constructing the Cantor set, $f' = 0$ a.e. Thus, $\int_0^1 f' = 0$, while $f(1) - f(0) = 1$. We shall return in Theorem 7.29 to the question of equality in (7.22).

By Corollary 2.7, any function of bounded variation can be written as the difference of two bounded monotone increasing functions. Hence, we obtain the following result (see also Exercise 9).

Corollary 7.23 *If f is of bounded variation on $[a,b]$, then f' exists a.e. in $[a,b]$, and $f' \in L[a,b]$.*

If f is of bounded variation on $[a,b]$ and $V(x)$ denotes its (total) variation on $[a,x]$, $a \le x \le b$, then V is monotone increasing by Theorem 2.2. The next result gives an important relation between V' and f'.

Theorem 7.24 *If f is of bounded variation on $[a,b]$ and $V(x)$ is the variation of f on $[a,x]$, $a \le x \le b$, then*

$$V'(x) = |f'(x)| \text{ for a.e. } x \in [a,b].$$

We will prove this with the help of the next lemma, which is of independent interest.

Lemma 7.25 (Fubini) *Let $\{f_k\}$ be a sequence of monotone increasing functions on $[a,b]$. If the series $s(x) = \sum f_k(x)$ converges on $[a,b]$ (equivalently, if $s(a)$ and $s(b)$ are finite), then*

$$s'(x) = \sum f_k'(x) \text{ a.e. in } [a,b].$$

In particular, $f_k' \to 0$ a.e. in $[a,b]$.

Proof. Let $s_m = \sum_{k=1}^{m} f_k$ and $r_m = \sum_{k=m+1}^{\infty} f_k$. Then s_m and r_m are monotone increasing functions and $s = s_m + r_m$. With the exception of a set Z_m, $|Z_m| = 0$, these three functions together with $f_1, \ldots, f_m$ are differentiable and $s' = s'_m + r'_m$. In particular, $s' \geq s'_m = \sum_{k=1}^{m} f'_k$ except in Z_m. It follows that

$$\sum_{k=1}^{\infty} f'_k \leq s'$$

except in $Z = \bigcup_{m=1}^{\infty} Z_m$, $|Z| = 0$.

To prove that in the last inequality we actually have equality a.e., it is enough to show that $r'_m \to 0$ a.e. for m running through a sequence of values $m_1 < m_2 < \cdots$. Select $\{m_j\}$ increasing so rapidly that $\sum r_{m_j}(x)$ converges at both $x = a$ and $x = b$. This implies the convergence of $\sum \{r_{m_j}(b) - r_{m_j}(a)\}$ and also, in view of the monotonicity of r_{m_j}, of $\sum \{r_{m_j}(b-) - r_{m_j}(a+)\}$. By (7.22), we have

$$0 \leq \int_a^b \sum r'_{m_j} = \sum \int_a^b r'_{m_j} \leq \sum \{r_{m_j}(b-) - r_{m_j}(a+)\}.$$

Thus, $\sum r'_{m_j}$ is integrable over (a, b), and therefore, it is finite a.e. in (a, b). Thus, $r'_{m_j} \to 0$ a.e., and the proof is complete.

Proof of Theorem 7.24. Let f be of bounded variation on $[a, b]$ and let $V(x) = V[a, x]$ be its variation on $[a, x]$, $a \leq x \leq b$. Then $V(a) = 0$ and $V(b)$ is the variation of f on $[a, b]$. Select a sequence $\left\{\Gamma_k : \Gamma_k = \left\{x_j^k\right\}\right\}$ of partitions of $[a, b]$ such that $0 \leq V(b) - S_{\Gamma_k} < 2^{-k}$, where

$$S_{\Gamma_k} = \sum_j \left| f\left(x_j^k\right) - f\left(x_{j-1}^k\right) \right|.$$

For each k, define a function f_k on $[a, b]$ as follows: if $x \in \left[x_{j-1}^k, x_j^k\right]$, let

$$f_k(x) = \begin{cases} f(x) + c_j^k \text{ if } f\left(x_j^k\right) \geq f\left(x_{j-1}^k\right), \\ -f(x) + c_j^k \text{ if } f\left(x_j^k\right) < f\left(x_{j-1}^k\right), \end{cases}$$

where the c_j^k are constants chosen so that $f_k(a) = 0$ and f_k is well-defined (i.e., single-valued) at x_j^k for every j. Then, for all k and j,

$$f_k(x_j^k) - f_k(x_{j-1}^k) = \left| f(x_j^k) - f(x_{j-1}^k) \right|,$$

so that

$$S_{\Gamma_k} = \sum_j \left[f_k(x_j^k) - f_k(x_{j-1}^k) \right] = f_k(b).$$

Hence, for any k, we obtain $0 \le V(b) - f_k(b) < 2^{-k}$.

We will show that each $V(x) - f_k(x)$ is an increasing function of x. This amounts to showing that if $x < y$, then $f_k(y) - f_k(x) \le V(y) - V(x)$. If x and y both belong to the same partitioning interval of Γ_k, then $f_k(y) - f_k(x) \le |f(y) - f(x)|$, and therefore,

$$f_k(y) - f_k(x) \le V[x, y] = V(y) - V(x).$$

In the general case, if $x_l^k, x_{l+1}^k, \ldots, x_m^k$ are the points of Γ_k between x and y, the result follows by adding the inequalities for the intervals $\left[x, x_l^k\right], \left[x_l^k, x_{l+1}^k\right], \ldots, \left[x_m^k, y\right]$. Since $V(a) = f_k(a) = 0$ and $V(b) - f_k(b) < 2^{-k}$, it follows that $0 \le V(x) - f_k(x) < 2^{-k}$ for all $x \in [a, b]$. Hence,

$$\sum [V(x) - f_k(x)] < \sum 2^{-k} < +\infty$$

for $x \in [a, b]$, so that by Lemma 7.25 the series $\sum \left[V'(x) - f_k'(x)\right]$ converges a.e. in $[a, b]$. Hence, $f_k' \to V'$ a.e. However, $\left|f_k'\right| = \left|f'\right|$ a.e., so that $\left|f'\right| = \left|V'\right|$ a.e. The theorem now follows from the fact that V' is nonnegative wherever it exists.

7.5 Absolutely Continuous and Singular Functions

We now turn to the question of equality in formula (7.22). As we know, the Cantor–Lebesgue function is an example of an increasing continuous f whose derivative is integrable on $[0, 1]$, but for which $\int_0^1 f' \ne f(1) - f(0)$.

In Section 7.1, the notion of absolute continuity of set functions was defined. We now introduce a related notion for functions of a single variable. A finite function f on a finite interval $[a, b]$ is said to be *absolutely continuous*

on $[a,b]$ if given $\varepsilon > 0$, there exists $\delta > 0$ such that for any collection $\{[a_i, b_i]\}$ (finite or not) of nonoverlapping subintervals of $[a,b]$,

$$\sum \left| f(b_i) - f(a_i) \right| < \varepsilon \text{ if } \sum (b_i - a_i) < \delta.$$

For example, if g is integrable on $[a,b]$ and $f(x) = \int_a^x g$ for $a \le x \le b$, then

$$\sum \left| f(b_i) - f(a_i) \right| \le \int\limits_{\bigcup [a_i, b_i]} |g|$$

for any nonoverlapping $[a_i, b_i]$. By Theorem 7.1, $\int_E |g|$ is an absolutely continuous *set function*, and therefore, f is an absolutely continuous function on $[a,b]$. One of the main results proved below (Theorem 7.29) is that every absolutely continuous function f has the form $f(x) = f(a) + \int_a^x g$ for an integrable g.

Another example of an absolutely continuous function is any f that satisfies a Lipschitz condition:

$$|f(x) - f(y)| \le C|x - y| \quad \text{for all } x, y \in [a,b] \tag{7.26}$$

and some constant $C \ge 0$.

On the other hand, the Cantor–Lebesgue function f is an example of a continuous function that is not absolutely continuous, since if $C_k = \bigcup_j [a_j^k, b_j^k]$ denotes the intervals remaining at the kth stage of construction of the Cantor set, then $|C_k| \to 0$, while

$$\sum_j \left[f\left(b_j^k\right) - f\left(a_j^k\right) \right] = 1$$

for every k.

It is simple to see that a linear combination of absolutely continuous functions is absolutely continuous and that an absolutely continuous function is continuous. Moreover, if f is absolutely continuous on $[a,b]$, then f' exists a.e. in $[a,b]$ and $f' \in L[a,b]$. This follows immediately from Corollary 7.23 and the next theorem.

Theorem 7.27 *If f is absolutely continuous on $[a,b]$, then it is of bounded variation on $[a,b]$.*

Proof. Choose δ so that $\sum \left| f(b_i) - f(a_i) \right| \le 1$ for any collection of nonoverlapping intervals with $\sum (b_i - a_i) \le \delta$. Then the variation of f over any

subinterval of $[a, b]$ with length less than δ is at most 1. Hence, if we split $[a, b]$ into N intervals each with length less than δ, then $V[a, b] \leq N$.

A function f for which f' is zero a.e. in $[a, b]$ is said to be *singular* on $[a, b]$. The Cantor–Lebesgue function is an example of a nonconstant, singular function on $[0, 1]$.

Theorem 7.28 *If f is both absolutely continuous and singular on $[a, b]$, then it is constant on $[a, b]$.*

Proof. It is enough to show that $f(a) = f(b)$ since this result applied to any subinterval proves that f is constant. Let E be the subset of (a, b) where $f' = 0$, so that $|E| = b - a$. Given $\varepsilon > 0$ and $x \in E$, we have $[x, x + h] \subset (a, b)$ and $|f(x + h) - f(x)| < \varepsilon h$ for all sufficiently small $h > 0$. Let δ be the number corresponding to ε in the definition of the absolute continuity of f. By Corollary 7.18 and (7.20), there exist disjoint $Q_j = \left[x_j, x_j + h_j\right]$, $j = 1, \ldots, N$, in (a, b) such that

(i) $\left|f\left(x_j + h_j\right) - f\left(x_j\right)\right| < \varepsilon h_j$,

(ii) $\sum_{j=1}^{N} \left|Q_j\right| > (b - a) - \delta$.

By (i),

$$\sum_{j=1}^{N} \left|f\left(x_j + h_j\right) - f\left(x_j\right)\right| < \varepsilon \sum_{j=1}^{N} \left|Q_j\right| \leq \varepsilon(b - a).$$

Moreover, since by (ii) the total length of the complementary intervals is less than δ, the sum of the absolute values of the increments of f over them is less than ε. Thus, the sum of the absolute values of the increments of f over the Q_j and the complementary intervals is less than $\varepsilon(b-a)+\varepsilon$. Hence, $|f(b)-f(a)| < \varepsilon(b - a) + \varepsilon$, so that $f(b) = f(a)$. This completes the proof.

Theorem 7.29 *A function f is absolutely continuous on $[a, b]$ if and only if f' exists a.e. in $[a, b]$, f' is integrable on $[a, b]$, and*

$$f(x) - f(a) = \int_a^x f', \quad a \leq x \leq b.$$

Proof. We have already observed (see p. 150 in Section 7.5) that any function of the form $G(x) = \int_a^x g$, $g \in L[a,b]$, is absolutely continuous on $[a,b]$. Hence, the sufficiency part of the theorem follows.

Conversely, if f is absolutely continuous, let $F(x) = \int_a^x f'$. F is well-defined by virtue of Theorem 7.27 and Corollary 7.23. Moreover, by Theorem 7.16, $F' = f'$ a.e. in $[a,b]$. It follows that $F(x) - f(x)$ is both absolutely continuous and singular on $[a,b]$. Hence, by Theorem 7.28, we obtain $F(x) - f(x) = F(a) - f(a)$ for $x \in [a,b]$. Since $F(a) = 0$, the proof is complete.

Theorem 7.30 *If f is of bounded variation on $[a,b]$, then f can be written $f = g + h$, where g is absolutely continuous on $[a,b]$ and h is singular on $[a,b]$. Moreover, g and h are unique up to additive constants.*

Proof. Let $g(x) = \int_a^x f'$ and $h = f - g$. Then $h' = f' - g' = f' - f' = 0$ a.e. in $[a,b]$, so that h is singular, and the formula $f = g + h$ gives the desired decomposition. If $f = g_1 + h_1$ is another such decomposition, then $g - g_1 = h_1 - h$. Since $g - g_1$ is absolutely continuous and $h_1 - h$ is singular, it follows from Theorem 7.28 that $g - g_1 = h_1 - h =$ constant, which completes the proof.

In view of Theorems 7.30 and 7.27, it is natural to ask if every bounded singular function is of bounded variation, as is the case, for example, with the Cantor–Lebesgue function. The answer is *no*, and an example is the characteristic function χ_C of the Cantor set C. In fact, since χ_C is zero on every (open) interval in the complement of C, then $(\chi_C)' = 0$ there, and so $(\chi_C)' = 0$ a.e. in $[0,1]$. However, χ_C does not belong to $BV[0,1]$ since it equals 1 at the endpoints of every interval removed during the construction of C.

The next theorem, which is an extension of Corollary 2.10, gives formulas for the variations of an absolutely continuous function.

Theorem 7.31 *Let f be absolutely continuous on $[a,b]$ and let $V(x)$, $P(x)$ and $N(x)$ denote its total, positive, and negative variations on $[a,x]$, $a \le x \le b$. Then V, P, and N are absolutely continuous on $[a,b]$, and*

$$V(x) = \int_a^x |f'|, \quad P(x) = \int_a^x \{f'\}^+, \quad \text{and} \quad N(x) = \int_a^x \{f'\}^- .$$

Proof. We will first show that V is absolutely continuous. If $[\alpha,\beta]$ is a subinterval of $[a,b]$ and $\Gamma = \{x_k\}$ is a partition of $[\alpha,\beta]$, then

$$V(\beta) - V(\alpha) = V[\alpha, \beta] = \sup_{\Gamma} \sum \left| f(x_k) - f\left(x_{k-1}\right) \right|$$
$$= \sup_{\Gamma} \sum \left| \int_{x_{k-1}}^{x_k} f' \right| \leq \int_{\alpha}^{\beta} |f'|.$$

Hence, if $\{[\alpha_i, \beta_i]\}$ is a collection of nonoverlapping subintervals of $[a, b]$, then

$$\sum [V(\beta_i) - V(\alpha_i)] \leq \int_{\bigcup [\alpha_i, \beta_i]} |f'|.$$

From this inequality and Theorem 7.1, it follows that V is absolutely continuous on $[a, b]$. Therefore, by Theorem 7.29 and the fact that $V(a) = 0$, we have $V(x) = \int_a^x V'$. Since $V' = |f'|$ a.e. by Theorem 7.24, we obtain $V(x) = \int_a^x |f'|$.

The fact that P and N are absolutely continuous and the formulas for P and N now follow from the relations $V(x) = \int_a^x |f'|$, $f(x) = f(a) + \int_a^x f'$, $P(x) = \frac{1}{2}[V(x) + f(x) - f(a)]$, and $N(x) = \frac{1}{2}[V(x) - f(x) + f(a)]$. (See Theorem 2.6.) This completes the proof.

On p. 27 in Section 2.3, we proved that if g is continuous on $[a, b]$ and f is continuously differentiable on $[a, b]$, then

$$\int_a^b g\, df = \int_a^b g f' dx.$$

This is a special case of the first part of the following theorem.

Theorem 7.32 (Integration by Parts)

(i) *If g is continuous on $[a, b]$ and f is absolutely continuous on $[a, b]$, then*

$$\int_a^b g\, df = \int_a^b g f'\, dx.$$

(ii) *If both f and g are absolutely continuous on $[a, b]$, then*

$$\int_a^b g f' dx = g(b)f(b) - g(a)f(a) - \int_a^b g' f\, dx.$$

Proof. To prove (i), first note that the integrals in the conclusion exist and are finite by Theorems 2.24 and 5.30. Let $\Gamma = \{x_i\}$ be a partition of $[a, b]$ with norm $|\Gamma|$. Then

$$\int_a^b gf'dx = \sum \int_{x_{i-1}}^{x_i} gf'dx = \sum g\left(x_{i-1}\right) \int_{x_{i-1}}^{x_i} f'(x)\,dx$$
$$+ \sum \int_{x_{i-1}}^{x_i} \left[g(x) - g\left(x_{i-1}\right)\right] f'(x)\,dx.$$

The first term on the right equals $\sum g\left(x_{i-1}\right)\left[f\left(x_i\right) - f\left(x_{i-1}\right)\right]$, which converges to $\int_a^b g\,df$ as $|\Gamma| \to 0$. The second term on the right is majorized in absolute value by

$$\left[\sup_{|x-y|\le|\Gamma|} \left|g(x) - g(y)\right|\right] \sum \int_{x_{i-1}}^{x_i} \left|f'\right| dx = \left[\sup_{|x-y|\le|\Gamma|} \left|g(x) - g(y)\right|\right] \int_a^b |f'|\,dx.$$

Since g is uniformly continuous on $[a, b]$, the last expression tends to 0 as $|\Gamma| \to 0$. This proves (i).

We can easily deduce (ii) from (i) by using Theorem 2.21. In fact, if f and g are absolutely continuous, then

$$\int_a^b gf'dx = \int_a^b g\,df = g(b)f(b) - g(a)f(a) - \int_a^b f\,dg$$
$$= g(b)f(b) - g(a)f(a) - \int_a^b fg'dx.$$

This proves (ii).

For a generalization of part (i) of the theorem, see Lemma 9.14.

7.6 Convex Functions

Let ϕ be defined and finite on an open interval (a, b), possibly of infinite length. Then ϕ is said to be *convex* in (a, b) if for every $[x_1, x_2]$ in (a, b), the graph of ϕ on $[x_1, x_2]$ lies on or below the line segment connecting the points $(x_1, \phi(x_1))$ and $(x_2, \phi(x_2))$ of the graph of ϕ.

Let $y = L(x)$ be the equation of this line segment. Then since every $x \in [x_1, x_2]$ can be written $x = \theta x_1 + (1-\theta)x_2$ for appropriate $\theta, 0 \le \theta \le 1$, the condition for convexity is

$$\phi\left(\theta x_1 + (1-\theta)x_2\right) \le L\left(\theta x_1 + (1-\theta)x_2\right)$$

for $[x_1, x_2] \subset (a, b)$ and $0 \le \theta \le 1$. Since

$$L\left(\theta x_1 + (1-\theta)\, x_2\right) = \theta L\left(x_1\right) + (1-\theta)L\left(x_2\right) = \theta\phi\left(x_1\right) + (1-\theta)\phi\left(x_2\right),$$

it follows that ϕ is convex in (a, b) if and only if

$$\phi\left(\theta x_1 + (1-\theta)x_2\right) \le \theta\phi\left(x_1\right) + (1-\theta)\phi\left(x_2\right) \tag{7.33}$$

for $a < x_1 < x_2 < b, 0 \le \theta \le 1$. Equivalently, ϕ is convex in (a, b) if and only if

$$\phi\left(\frac{p_1 x_1 + p_2 x_2}{p_1 + p_2}\right) \le \frac{p_1\phi\left(x_1\right) + p_2\phi\left(x_2\right)}{p_1 + p_2} \tag{7.34}$$

for $a < x_1 < x_2 < b,\ p_1 \ge 0,\ p_2 \ge 0, \mathrm{p}_1 + p_2 > 0$.

Theorem 7.35 (Jensen's Inequality) *Let ϕ be convex in (a, b). Let $\left\{x_j\right\}_{j=1}^N$ be points of (a, b) and $\left\{p_j\right\}_{j=1}^N$ satisfy $p_j \ge 0$ and $\sum p_j > 0$. Then*

$$\phi\left(\frac{\sum p_j x_j}{\sum p_j}\right) \le \frac{\sum p_j \phi\left(x_j\right)}{\sum p_j}.$$

The proof follows by repeated application of (7.34).

The slope of the chord connecting two points $(\alpha, \phi(\alpha))$ and $(\beta, \phi(\beta))$ of the graph of ϕ is $[\phi(\beta) - \phi(\alpha)]/(\beta - \alpha)$. We leave it as an exercise to verify the following simple relations between the convexity of ϕ and the slopes of its chords: if ϕ is convex in (a, b), then for all x, x_1, x_2 with $a < x_1 < x < x_2 < b$,

$$\frac{\phi(x) - \phi(x_1)}{x - x_1} \leq \frac{\phi(x_2) - \phi(x_1)}{x_2 - x_1} \leq \frac{\phi(x_2) - \phi(x)}{x_2 - x};$$

conversely, if for every such x, x_1, x_2, either one of these two inequalities holds, then ϕ is convex in (a, b).

Theorem 7.36

(i) *If ϕ_1 and ϕ_2 are convex in (a, b), then $\phi_1 + \phi_2$ is convex in (a, b).*

(ii) *If ϕ is convex in (a, b) and c is a positive constant, then $c\phi$ is convex in (a, b).*

(iii) *If $\phi_k, k = 1, 2, \ldots,$ are convex in (a, b) and $\phi_k \to \phi$ in (a, b), then ϕ is convex in (a, b).*

The proof is left as an exercise.

Theorem 7.37 *If ϕ' exists and is monotone increasing in (a, b), then ϕ is convex in (a, b). In particular, if ϕ'' exists and is nonnegative in (a, b), then ϕ is convex in (a, b).*

Proof. We shall use the following inequality: If $b_1, b_2 > 0$, then

$$\min\left\{\frac{a_1}{b_1}, \frac{a_2}{b_2}\right\} \leq \frac{a_1 + a_2}{b_1 + b_2} \leq \max\left\{\frac{a_1}{b_1}, \frac{a_2}{b_2}\right\}. \tag{7.38}$$

To prove the first inequality in (7.38), let $m = \min\{a_1/b_1, a_2/b_2\}$. Then $mb_1 \leq a_1$ and $mb_2 \leq a_2$, so that $m(b_1 + b_2) \leq a_1 + a_2$. This is the desired result. The second inequality is proved similarly.

To prove that ϕ is convex, we will show that if $a < x_1 < x < x_2 < b$, then

$$\frac{\phi(x_2) - \phi(x_1)}{x_2 - x_1} \geq \frac{\phi(x) - \phi(x_1)}{x - x_1}.$$

Write

$$\frac{\phi(x_2) - \phi(x_1)}{x_2 - x_1} = \frac{[\phi(x_2) - \phi(x)] + [\phi(x) - \phi(x_1)]}{[x_2 - x] + [x - x_1]}.$$

Since ϕ' exists in (a,b), the mean-value theorem implies that there are $\xi_1 \in (x_1, x)$ and $\xi_2 \in (x, x_2)$ such that

$$\frac{\phi(x) - \phi(x_1)}{x - x_1} = \phi'(\xi_1), \quad \frac{\phi(x_2) - \phi(x)}{x_2 - x} = \phi'(\xi_2).$$

Since ϕ' is increasing, $\phi'(\xi_1) \leq \phi'(\xi_2)$, and it follows from the first inequality in (7.38) that

$$\frac{\phi(x_2) - \phi(x_1)}{x_2 - x_1} \geq \frac{\phi(x) - \phi(x_1)}{x - x_1}.$$

This completes the proof.

As a corollary of Theorem 7.37, we see that

$$\begin{aligned} &\text{(i)} && x^p \text{ is convex in } (0,\infty) \text{ if } p \geq 1 \text{ or if } p \leq 0 \\ &\text{(ii)} && e^{ax} \text{ is convex in } (-\infty,\infty) \\ &\text{(iii)} && \log(1/x) = -\log x \text{ is convex in } (0,\infty) \end{aligned} \tag{7.39}$$

Theorem 7.40 *If ϕ is convex in (a,b), then ϕ is continuous in (a,b). Moreover, ϕ' exists except at most in a countable set and is monotone increasing.*

Proof. Since ϕ is convex, the slope $[\phi(x+h)-\phi(x)]/h, h > 0$, decreases with h. Hence, the derivative on the right,

$$D^+\phi(x) = \lim_{h \to 0+} \frac{\phi(x+h) - \phi(x)}{h},$$

exists and is distinct from $+\infty$ in (a,b). Similarly, the derivative on the left,

$$D^-\phi(x) = \lim_{h \to 0+} \frac{\phi(x) - \phi(x-h)}{h},$$

exists and is distinct from $-\infty$ in (a,b). Since $[\phi(x) - \phi(x-h)]/h \leq [\phi(x+h) - \phi(x)]/h$, $h > 0$, we obtain

$$-\infty < D^-\phi(x) \leq D^+\phi(x) < +\infty. \tag{7.41}$$

This shows in particular that ϕ is continuous in (a,b). We next claim that

$$D^+\phi(y) \leq D^-\phi(x) \text{ if } a < y < x < b. \tag{7.42}$$

In fact, if $y < x$, then as seen from the discussion earlier in the proof, we have

$$D^+\phi(y) \leq \frac{\phi(x) - \phi(y)}{x - y} \leq D^-\phi(x).$$

This proves the claim. We therefore obtain $D^+\phi(y) \leq D^-\phi(x) \leq D^+\phi(x)$ if $y < x$, which shows that $D^+\phi$ is monotone increasing. Similarly, $D^-\phi$ is monotone increasing.

To complete the proof of the theorem, note that (cf. Theorem 2.8) $D^+\phi$ can have at most a countable number of discontinuities since it is monotone and finite on (a, b). If x is a point of continuity of $D^+\phi$, then letting $y \to x-$ in the last inequalities, we obtain $D^+\phi(x) = D^-\phi(x)$. Therefore, ϕ' exists at every point of continuity of $D^+\phi$, and the theorem follows.

Theorem 7.43 *If ϕ is convex in (a, b), then it satisfies a Lipschitz condition on every closed subinterval of (a, b). In particular, if $a < x_1 < x_2 < b$, we have*

$$\phi(x_2) - \phi(x_1) = \int_{x_1}^{x_2} \phi'.$$

Proof. Let $[x_1, x_2]$ be a closed subinterval of (a, b) and let $x_1 \leq y < x \leq x_2$. Then as before

$$D^+\phi(y) \leq \frac{\phi(x) - \phi(y)}{x - y} \leq D^-\phi(x),$$

so that, since $D^+\phi$ and $D^-\phi$ are monotone increasing,

$$D^+\phi(x_1) \leq \frac{\phi(x) - \phi(y)}{x - y} \leq D^-\phi(x_2).$$

Hence, $|\phi(x) - \phi(y)| \leq C|x - y|$, where C is the larger of $\left|D^+\phi(x_1)\right|$ and $\left|D^-\phi(x_2)\right|$. This shows that ϕ satisfies a Lipschitz condition on $[x_1, x_2]$, and the rest of the theorem follows since Lipschitz functions of a single variable are absolutely continuous.

The next result is a useful version of Jensen's inequality for integrals. We shall need the notion of a *supporting line*: If ϕ is convex on (a, b) and $x_0 \in (a, b)$, a supporting line at x_0 is a line through $(x_0, \phi(x_0))$ that lies on or below the graph of ϕ on (a, b). It follows from the discussion preceding (7.41) that any line through $(x_0, \phi(x_0))$ whose slope m satisfies $D^-\phi(x_0) \leq m \leq D^+\phi(x_0)$ is a supporting line at x_0.

Theorem 7.44 (Jensen's Integral Inequality) *Let f and p be measurable functions finite a.e. on a measurable set $A \subset \mathbf{R^n}$. Suppose that fp and p are integrable on A, that $p \geq 0$, and that $\int_A p > 0$. If ϕ is convex in an interval containing the range of f, then*

$$\phi\left(\frac{\int_A fp}{\int_A p}\right) \leq \frac{\int_A \phi(f)\, p}{\int_A p}.$$

Proof. By hypothesis, f is finite a.e. in A. Choose (a, b), $-\infty \leq a < b \leq +\infty$, such that ϕ is convex in (a, b) and $a < f(x) < b$ for every x at which $f(x)$ is finite. The number γ defined by

$$\gamma = \frac{\int_A fp}{\int_A p}$$

is finite and satisfies $a < \gamma < b$. If m is the slope of a supporting line at γ and $a < t < b$, then $\phi(\gamma) + m(t - \gamma) \leq \phi(t)$. Hence, for almost every $x \in A$,

$$\phi(\gamma) + m[f(x) - \gamma] \leq \phi(f(x)).$$

Multiplying both sides of this inequality by $p(x)$ and integrating the result with respect to x, we obtain

$$\phi(\gamma) \int_A p + m\left(\int_A fp - \gamma \int_A p\right) \leq \int_A \phi(f)\, p.$$

Here the existence of $\int_A \phi(f)\, p$ follows from the integrability of p and fp. (The continuity of ϕ implies that $\phi(f)$ is measurable.) Since $\int_A fp - \gamma \int_A p = 0$, the last inequality reduces to

$$\phi(\gamma) \int_A p \leq \int_A \phi(f)\, p,$$

which is the desired result.

In passing, we mention that a function f is called *concave in* (a, b) if $-f$ is convex on (a, b). Properties of concave functions are easily deduced from those of convex functions.

7.7 The Differential in $\mathbf{R^n}$

The principal fact to be proved in this section is the Rademacher–Stepanov theorem stating that a function that is locally Lipschitz continuous in an open set in $\mathbf{R^n}$, $n > 1$, has a first differential (or tangent plane) almost everywhere in the set. The result is somewhat surprising in view of the fact that the graph of a Lipschitz function may have corners and edges. The analogous result in case $n = 1$ follows by combining Theorem 7.27 and Corollary 7.23. We will also derive a result in which the assumption of Lipschitz continuity is replaced by a weaker condition involving the second difference of a function, and we will study a simple way to extend Lipschitz continuous functions on subsets of $\mathbf{R^n}$ to all of $\mathbf{R^n}$.

We begin by defining the notion of the first differential of a function. A finite real-valued function f defined in a neighborhood of a point $\mathbf{x} \in \mathbf{R^n}$, $n \geq 1$, is said to have a *total first differential* at $\mathbf{x}$ if there exists $\mathbf{A} = (a_1, \ldots, a_n) \in \mathbf{R^n}$ such that

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \mathbf{A} \cdot \mathbf{h} = o(|\mathbf{h}|) \quad \text{as } \mathbf{h} \to \mathbf{0}. \tag{7.45a}$$

Here, $\mathbf{A} \cdot \mathbf{h}$ denotes the dot product of $\mathbf{A}$ and $\mathbf{h}$, that is,

$$\mathbf{A} \cdot \mathbf{h} = \sum_{i=1}^{n} a_i h_i, \quad \mathbf{h} = (h_1, \ldots h_n),$$

and the notation "$o(|\mathbf{h}|)$ as $\mathbf{h} \to \mathbf{0}$" in (7.45a) means that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \mathbf{A} \cdot \mathbf{h}}{|\mathbf{h}|} = 0.$$

(In general, for any finite real-valued function $F(\mathbf{h})$ defined in a deleted neighborhood of the origin, the notation $F(\mathbf{h}) = o(|\mathbf{h}|)$ as $|\mathbf{h}| \to 0$ means that $\lim_{\mathbf{h} \to \mathbf{0}} F(\mathbf{h})/|\mathbf{h}| = 0$.)

Sometimes, we will just say that such an f has a *first differential* at $\mathbf{x}$, or simply a *differential* at $\mathbf{x}$. When $n = 1$, (7.45a) means only that f has a first derivative at x.

If f satisfies (7.45a), then f is clearly continuous at $\mathbf{x}$. By choosing $\mathbf{h} = (0, \ldots, 0, h_i, 0, \ldots, 0)$, $i = 1, \ldots, n$, in (7.45a), we also have

$$\lim_{h_i \to 0} \frac{f\left(x_1, \ldots, x_{i-1}, x_i + h_i, x_{i+1}, \ldots, x_n\right) - f\left(x_1, \ldots, x_n\right)}{h_i} = a_i.$$

Thus, $\mathbf{A}$ is unique if it exists, and a function f that has a differential at $\mathbf{x}$ has (first) partial derivatives $\partial f/\partial x_i$ at $\mathbf{x}$, $i = 1, \ldots, n$, and

$$\mathbf{A} = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right).$$

Consequently, (7.45a) is the same as

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}) h_i + o(|\mathbf{h}|) \quad \text{as } \mathbf{h} = (h_1, \ldots, h_n) \to \mathbf{0}. \tag{7.45b}$$

If f has a differential at $\mathbf{x}$, then the graph of the linear function $L_{\mathbf{x}}(\mathbf{y})$ defined by

$$L_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}) \left(y_i - x_i\right), \quad \mathbf{y} = \left(y_1, \ldots, y_n\right) \in \mathbf{R^n}, \tag{7.46}$$

is called the *tangent plane* (or *tangent line* if $n = 1$) to f at $\mathbf{x}$. By choosing $\mathbf{h} = \mathbf{y} - \mathbf{x}$ in (7.45b), we obtain

$$f(\mathbf{y}) = L_{\mathbf{x}}(\mathbf{y}) + o(|\mathbf{y} - \mathbf{x}|) \quad \text{as } \mathbf{y} \to \mathbf{x}. \tag{7.47}$$

When $n > 1$, a standard alternate terminology for (7.45b) is to say that f *has a tangent plane at* $\mathbf{x}$.

If f is any function whose partial derivatives $\partial f/\partial x_i$ all exist at $\mathbf{x}$, we write

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right)$$

and call $\nabla f(\mathbf{x})$ the *gradient of* f *at* $\mathbf{x}$.

When $n > 1$, the mere existence of ∇f at a point $\mathbf{x}$ does not imply that f has a tangent plane at $\mathbf{x}$ even if f is continuous at $\mathbf{x}$. For example, in case $n = 2$, the function

$$f(x_1, x_2) = \begin{cases} x_1 x_2^2 / \left[x_1^2 + x_2^2\right] & \text{if } (x_1, x_2) \neq (0,0) \\ 0 \quad \text{if } (x_1, x_2) = (0,0) \end{cases} \tag{7.48}$$

is continuous and has first partial derivatives everywhere in $\mathbf{R^2}$, but f does not have a tangent plane at $(0,0)$ since $f(0,0) = 0$, $\nabla f(0,0) = (0,0)$, and $f(h_1, h_1)/h_1 = 1/2$ if $h_1 \neq 0$. We leave it as an exercise to show that f satisfies the Lipschitz condition

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq C\,|\mathbf{x} - \mathbf{y}|, \quad \mathbf{x}, \mathbf{y} \in \mathbf{R^2},$$

for some constant C that is independent of $\mathbf{x}$ and $\mathbf{y}$.

We now define a weaker notion of differentiability that will play a role in the proof of the Rademacher–Stepanov theorem. Let $\mathbf{x} \in \mathbf{R^n}$, $n \geq 1$, be a finite real-valued function defined in a measurable set that contains $\mathbf{x}$ and has $\mathbf{x}$ as a point of density. Equivalently, suppose that $f(\mathbf{x} + \mathbf{h})$ is defined and finite for all $\mathbf{h}$ in a measurable set $H \subset \mathbf{R^n}$ containing $\mathbf{0}$ and having $\mathbf{0}$ as a point of density. We say that f has an *approximate first differential* at $\mathbf{x}$, or simply an *approximate differential* at $\mathbf{x}$, if there exists $\mathbf{A} \in \mathbf{R^n}$, depending on $\mathbf{x}$, such that

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{A} \cdot \mathbf{h} + o(|\mathbf{h}|), \quad \mathbf{h} \in H, \quad \mathbf{h} \to \mathbf{0}. \tag{7.49}$$

When $n = 1$, the standard terminology for (7.49) is that f has an *approximate derivative* at the point in question.

The vector $\mathbf{A}$ in (7.49) is unique in the following sense. If f satisfies (7.49) as well as

$$f\left(\mathbf{x} + \mathbf{h}'\right) = f(\mathbf{x}) + \mathbf{A}' \cdot \mathbf{h}' + o\left(\left|\mathbf{h}'\right|\right), \quad \mathbf{h}' \in H', \quad \mathbf{h}' \to \mathbf{0}$$

for some $\mathbf{A}'$ and some measurable set H' that has $\mathbf{0}$ as a point of density, then $\mathbf{A}' = \mathbf{A}$ (see Exercise 27).

A function that has a total differential at a point clearly has an approximate differential there, but the converse is false even if the function is continuous. For example, let $f(x_1, x_2)$ be a function on $\mathbf{R^2}$ with the properties

(i) $f(x_1, x_2) = 0$ if $|x_2| \geq x_1^2$,

(ii) $f(x_1, 0) = x_1$ for all $x_1 \in (-\infty, \infty)$,

(iii) f is continuous on $\mathbf{R^2}$.

Then f has an approximate differential at $(0, 0)$ since it satisfies (7.49) at $\mathbf{x} = (0, 0)$ with $f(0, 0) = 0$ and $\mathbf{A} = (0, 0)$ for the set $H = \left\{(x_1, x_2) : |x_2| \geq x_1^2\right\}$. However, f does not have a total differential at $(0, 0)$ since $f(0, 0) = 0$, $\nabla f(0, 0) = (1, 0)$, and, for example, the estimate

$$f\left(x_1, x_1^2\right) = x_1 + o\left(\left\{x_1^2 + x_1^4\right\}^{1/2}\right) \quad \text{as } x_1 \to 0$$

is false due to the fact that $f\left(x_1, x_1^2\right) = 0$ for all x_1.

We leave it as an exercise to show that the function f in the example above cannot be redefined in any set of measure zero so that the resulting function has a total differential at $(0, 0)$.

On the other hand, the next theorem shows that a Lipschitz continuous function (see the following definition) has a total differential at every point where it has an approximate differential.

A finite real-valued function f defined in an open set $G \subset \mathbf{R^n}$ is said to be *Lipschitz continuous* in G if there is a constant C such that

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq C|\mathbf{x} - \mathbf{y}| \quad \text{for all } \mathbf{x}, \mathbf{y} \in G.$$

We then write $f \in Lip(G)$. Similarly, for any finite f defined on G, we write $f \in Lip_{loc}(G)$ and say that f is *locally Lipschitz continuous* in G if for every compact set $K \subset G$, there is a constant C_K such that

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq C_K|\mathbf{x} - \mathbf{y}| \quad \text{for all } \mathbf{x}, \mathbf{y} \in K.$$

If $f \in Lip(G)$, then $f \in Lip_{loc}(G)$, but the converse is false. For example, in $\mathbf{R^2}$, the function $f(x_1, x_2) = (1 - x_1)^{-1}$ is locally Lipschitz continuous on the open unit ball centered at $(0, 0)$, but it is not Lipschitz continuous there. In fact, for all (x_1, x_2), (y_1, y_2) in the ball, we have

$$\left|f(x_1, x_2) - f(y_1, y_2)\right| = \frac{|x_1 - y_1|}{(1 - x_1)(1 - y_1)}.$$

We leave it as an exercise to construct a function that is bounded, uniformly continuous and locally Lipschitz continuous on the open unit ball in $\mathbf{R^2}$ but not Lipschitz continuous there.

Theorem 7.50 *Let $\mathbf{x}_0 \in \mathbf{R^n}$ and f be a function that is Lipschitz continuous in a neighborhood of $\mathbf{x}_0$. If f has an approximate differential at $\mathbf{x}_0$, then f has a total differential at $\mathbf{x}_0$.*

Proof. Suppose that f is Lipschitz continuous near $\mathbf{x}_0$ and that

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \mathbf{A} \cdot \mathbf{h} + o(|\mathbf{h}|), \quad \mathbf{h} \to \mathbf{0},\ \mathbf{h} \in H,$$

for some $\mathbf{A}$ and some measurable set $H \subset \mathbf{R^n}$ that has $\mathbf{0}$ as a point of density. We will prove the theorem by showing that the same asymptotic formula holds for all $\mathbf{h} \to \mathbf{0}$ without the restriction that $\mathbf{h} \in H$.

By considering the function $g(\mathbf{x}) = f(\mathbf{x}_0 + \mathbf{x}) - f(\mathbf{x}_0) - \mathbf{A} \cdot \mathbf{x}$, we may assume that $\mathbf{x}_0 = \mathbf{0}$, $f(\mathbf{0}) = 0$, and $\mathbf{A} = \mathbf{0}$, that is, we may assume that f is Lipschitz continuous near $\mathbf{0}$ and

$$\lim_{\mathbf{h} \to \mathbf{0}, \mathbf{h} \in H} \frac{f(\mathbf{h})}{|\mathbf{h}|} = 0. \tag{7.51}$$

Our goal is then to prove that $f(\mathbf{h})/|\mathbf{h}| \to 0$ as $|\mathbf{h}| \to 0$.

For any $\mathbf{y} \in \mathbf{R^n}$ and $t > 0$, let

$$B(\mathbf{y}; t) = \left\{\mathbf{x} \in \mathbf{R^n} : |\mathbf{x} - \mathbf{y}| < t\right\}$$

denote the open ball with center $\mathbf{y}$ and radius t. Choose $C, r > 0$ such that

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq C|\mathbf{x} - \mathbf{y}| \quad \text{if } \mathbf{x}, \mathbf{y} \in B(\mathbf{0}; r).$$

Fix $\varepsilon > 0$ and let β, γ satisfy

$$0 < \beta < 1, \quad 0 < \gamma < 1 - \beta, \quad \gamma + 1 - \beta < \frac{\varepsilon}{2C}. \tag{7.52}$$

Note that when ε is small, then β is near 1 and γ is small.

For any $\mathbf{h} \in \mathbf{R^n} - \{\mathbf{0}\}$, let

$$D_{\mathbf{h}} = B(\beta\, \mathbf{h}; \gamma\, |\mathbf{h}|).$$

The triangle inequality and (7.52) imply that

$$D_{\mathbf{h}} \subset B(\mathbf{0}; |\mathbf{h}|) \cap B\left(\mathbf{h}; \frac{\varepsilon}{2C}|\mathbf{h}|\right).$$

Now choose α such that $1 - \gamma^n < \alpha < 1$. Since $\mathbf{0}$ is a point of density of H, there exists $\delta > 0$ such that if $0 < |\mathbf{h}| < \delta$ then

$$\begin{aligned}
\alpha|B(\mathbf{0}; |\mathbf{h}|)| &\leq |B(\mathbf{0}; |\mathbf{h}|) \cap H| \\
&\leq |D_{\mathbf{h}} \cap H| + |B(\mathbf{0}; |\mathbf{h}|) - D_{\mathbf{h}}| \\
&= |D_{\mathbf{h}} \cap H| + |B(\mathbf{0}; |\mathbf{h}|)| - |D_{\mathbf{h}}| \\
&= |D_{\mathbf{h}} \cap H| + \left(1 - \gamma^n\right) |B(\mathbf{0}; |\mathbf{h}|)|.
\end{aligned}$$

Since $\alpha > 1 - \gamma^n$, it follows that $|D_{\mathbf{h}} \cap H| > 0$ and in particular that $D_{\mathbf{h}} \cap H$ is not empty. Thus, for every $\mathbf{h}$ with $0 < |\mathbf{h}| < \delta$, there is a point $\mathbf{h}_1 \in H$ such that $|\mathbf{h}_1| < |\mathbf{h}|$ and $|\mathbf{h} - \mathbf{h}_1| < \{\varepsilon/(2C)\}|\mathbf{h}|$.

We may also choose δ so small that $\delta < r$ and, by (7.51), so that $|f(\tilde{\mathbf{h}})| < (\varepsilon/2)|\tilde{\mathbf{h}}|$ if $\tilde{\mathbf{h}} \in H$ and $|\tilde{\mathbf{h}}| < \delta$. Note that δ depends ultimately only on ε, H, f, and n. If $0 < |\mathbf{h}| < \delta$ and $\mathbf{h}_1$ is chosen as above, we obtain

$$\begin{aligned}|f(\mathbf{h})| &\leq \left|f(\mathbf{h}) - f(\mathbf{h}_1)\right| + \left|f(\mathbf{h}_1)\right| \\ &\leq C\,|\mathbf{h} - \mathbf{h}_1| + \frac{\varepsilon}{2}\,|\mathbf{h}_1| \\ &\leq C\frac{\varepsilon}{2C}|\mathbf{h}| + \frac{\varepsilon}{2}|\mathbf{h}| = \varepsilon|\mathbf{h}|,\end{aligned}$$

which completes the proof.

The next theorem is the main result of this section.

Theorem 7.53 (Rademacher–Stepanov) *Let G be an open set in $\mathbf{R^n}$, $n > 1$, and let $f \in Lip_{loc}(G)$. Then f has a total first differential a.e. in G. The partial derivatives $\partial f/\partial x_i$, $i = 1, \ldots, n$, are measurable in G and bounded a.e. in every compact subset of G. Furthermore, if $f \in Lip(G)$, then all $\partial f/\partial x_i$ are bounded a.e. in G.*

The proof will use Theorem 7.50 together with Lusin's theorem and the next four lemmas. It will also use the one-dimensional version of Theorem 7.53, which is included in Corollary 7.23.

We begin by showing that the derivative (from the right) of a Lipschitz function of one variable can be defined in terms of a certain sequential (as opposed to ordinary) limit of difference quotients, a fact that will be useful in proving measurability of the partial derivatives in Theorem 7.53.

Lemma 7.54 *Let f be Lipschitz continuous in a half-open interval $[a, b)$ in $\mathbf{R^1}$ and let $k = 1, 2, \ldots$. If*

$$\lim_{k\to\infty} \frac{f(a + 1/k) - f(a)}{1/k}$$

exists, then so does

$$\lim_{h\to 0+} \frac{f(a + h) - f(a)}{h},$$

and the two limits are the same.

Proof. Denote

$$L = \lim_{k\to\infty} \frac{f(a + 1/k) - f(a)}{1/k}.$$

Note that L is finite since f is Lipschitz continuous in $[a, b)$. Given $\varepsilon > 0$, choose a positive integer K_ε such that

$$\left|\frac{f(a+1/k)-f(a)}{1/k} - L\right| < \varepsilon \quad \text{if } k \geq K_\varepsilon.$$

Let h satisfy $0 < h < 1/K_\varepsilon$ and choose $k \geq K_\varepsilon$ such that $1/(k+1) \leq h < 1/k$. Then by the triangle inequality,

$$\begin{aligned}
\left|\frac{f(a+h)-f(a)}{h} - L\right| &\leq \left|\frac{f(a+h)-f(a)}{h} - \frac{f(a+1/k)-f(a)}{1/k}\right| + \varepsilon \\
&\leq \left|\frac{[f(a+h)-f(a)] - [f(a+1/k)-f(a)]}{h}\right| \\
&\quad + \left|f(a+1/k)-f(a)\right| \left(\frac{1}{h} - k\right) + \varepsilon \\
&= I + II + \varepsilon,
\end{aligned}$$

say. Since $0 < (1/h) - k \leq 1$, we have

$$II \leq |f(a+1/k) - f(a)| \to 0 \quad \text{as } k \to \infty.$$

Also,

$$I = \left|\frac{f(a+h)-f(a+1/k)}{h}\right|,$$

and therefore, by the Lipschitz character of f, there is a constant C independent of h and k such that

$$I \leq \frac{C|h-(1/k)|}{h} \leq C\,\frac{1/[k(k+1)]}{1/(k+1)} = C\,\frac{1}{k}.$$

Combining estimates, we obtain

$$\lim_{h \to 0+} \frac{f(a+h)-f(a)}{h} = L,$$

and the proof of the lemma is complete.

Lemma 7.55 *Let G be an open set in $\mathbf{R^n}$, $n > 1$. If $f \in Lip_{loc}(G)$, then f has measurable first partial derivatives $\partial f/\partial x_i$ a.e. in G, $i = 1, \ldots, n$.*

Proof. Fix G and f as in the hypothesis. Consider, for example, the case $i = 1$ and denote points of $\mathbf{R^n}$ by

$$(x, \mathbf{y}), \quad \text{with } x \in (-\infty, \infty) \text{ and } \mathbf{y} \in \mathbf{R^{n-1}}. \tag{7.56}$$

Also, denote the difference quotient of f in the first variable by

$$D_h f(x, \mathbf{y}) = \frac{f(x+h, \mathbf{y}) - f(x, \mathbf{y})}{h}, \quad h \neq 0, \tag{7.57}$$

provided $(x, \mathbf{y}), (x+h, \mathbf{y}) \in G$.

Let E be the set defined by

$$E = \left\{ (x, \mathbf{y}) \in G : \frac{\partial f}{\partial x}(x, \mathbf{y}) = \lim_{h \to 0} D_h f(x, \mathbf{y}) \text{ exists} \right\}.$$

The derivative $\partial f / \partial x$ is automatically finite in E since f is locally Lipschitz continuous in G. Our goal is to show that E is measurable in $\mathbf{R^n}$, that $|G - E| = 0$, and that $(\partial f / \partial x)(x, \mathbf{y})$ is a measurable function on E. Let

$$D^+ f(x, \mathbf{y}) = \lim_{h \to 0+} D_h f(x, \mathbf{y}) \quad \text{and} \quad D^- f(x, \mathbf{y}) = \lim_{h \to 0-} D_h f(x, \mathbf{y})$$

at any $(x, \mathbf{y}) \in G$ where these limits exist, and define sets E^+ and E^- by

$$E^+ = \left\{ (x, \mathbf{y}) \in G : D^+ f(x, \mathbf{y}) \text{ exists} \right\},$$
$$E^- = \left\{ (x, \mathbf{y}) \in G : D^- f(x, \mathbf{y}) \text{ exists} \right\}.$$

Note that

$$E = \left\{ (x, \mathbf{y}) \in E^+ \cap E^- : D^+ f(x, \mathbf{y}) = D^- f(x, \mathbf{y}) \right\}. \tag{7.58}$$

Moreover, by Lemma 7.54, E^+ can be expressed in terms of a sequential limit:

$$E^+ = \left\{ (x, \mathbf{y}) \in G : \lim_{k \to \infty} D_{1/k} f(x, \mathbf{y}) \text{ exists} \right\},$$

and by similar reasoning,

$$E^- = \left\{ (x, \mathbf{y}) \in G : \lim_{k \to \infty} D_{-1/k} f(x, \mathbf{y}) \text{ exists} \right\}.$$

In order to have a difference quotient that is well-defined for **all** $h \neq 0$ when $(x, \mathbf{y}) \in G$, we extend f to be zero outside G. Thus, let

$$\bar{f} = \begin{cases} f \text{ in } G \\ 0 \text{ in } \mathbf{R^n} - G. \end{cases}$$

Then $\bar{f}$ is measurable in $\mathbf{R^n}$, and consequently, $D_h\bar{f}$ is defined and measurable in $\mathbf{R^n}$ for all $h \neq 0$. In particular, both $D_{1/k}\bar{f}$ and $D_{-1/k}\bar{f}$ are measurable in $\mathbf{R^n}$ for all $k = 1, 2, \ldots$. Since G is open, for every $(x, \mathbf{y}) \in G$, we have $D_{1/k}\bar{f}(x, \mathbf{y}) = D_{1/k}f(x, \mathbf{y})$ for all large k. Therefore,

$$E^+ = \left\{ (x, \mathbf{y}) \in G : \lim_{k \to \infty} D_{1/k}\bar{f}(x, \mathbf{y}) \text{ exists} \right\},$$

and a similar representation holds for E^- with $D_{1/k}$ replaced by $D_{-1/k}$. It follows that E^+ and E^- are measurable in $\mathbf{R^n}$ (cf. Exercise 23 in Chapter 4) and that D^+f is measurable on E^+ and D^-f is measurable on E^-. Then (7.58) implies that E is a measurable set in $\mathbf{R^n}$ and $\partial f/\partial x$ is a measurable function on E.

Since G is open, for every $\mathbf{y} \in \mathbf{R^{n-1}}$, the one-dimensional set $G_\mathbf{y}$ defined by

$$G_\mathbf{y} = \left\{ x \in (-\infty, \infty) : (x, \mathbf{y}) \in G \right\}$$

is open (possibly empty) in $\mathbf{R^1}$. Also, $f(x, \mathbf{y})$ considered as a function of x is locally Lipschitz continuous in $G_\mathbf{y}$ for every $\mathbf{y}$. Hence, by Corollary 7.23, $(\partial f/\partial x)(x, \mathbf{y})$ exists for a.e. (linear measure) $x \in G_\mathbf{y}$. Defining $E_\mathbf{y}$ in the same manner as $G_\mathbf{y}$, we have from Tonelli's theorem and the measurability of E that for a.e. $\mathbf{y} \in \mathbf{R^{n-1}}$, $E_\mathbf{y}$ is linearly measurable, $|E_\mathbf{y}| = |G_\mathbf{y}|$ (linear measure again), and

$$|E| = \int_{\mathbf{R^{n-1}}} |E_\mathbf{y}| \, d\mathbf{y} = \int_{\mathbf{R^{n-1}}} |G_\mathbf{y}| \, d\mathbf{y} = |G|.$$

In case G has finite measure, it follows that $|G - E| = 0$, and we have accomplished our goal. The same is true if $|G| = \infty$ by intersecting G and E with a sequence of open balls increasing to $\mathbf{R^n}$. Finally, a similar argument holds for every coordinate x_i, $i = 1, \ldots, n$, and the proof of Lemma 7.55 is complete.

Before proceeding, we remark that if f is any measurable function in an open set G and if $\partial f/\partial x_i$ exists a.e. in a measurable set $E \subset G$ for some i, then $\partial f/\partial x_i$ is measurable in E. If $i = 1$, for example, this follows by considering the extension $\bar{f}$ in the proof of Lemma 7.55 and using the fact that $D_{h_k}\bar{f}$ is measurable in $\mathbf{R^n}$ and converges a.e. in E to $\partial f/\partial x_1$ for any fixed

sequence $h_k \to 0$. While this fact is generally useful, we did not use it in the proof of Lemma 7.55 because the existence of the first partial derivatives of f a.e. in G was not guaranteed in advance.

The next lemma establishes a fact about uniform convergence of difference quotients. Its proof is similar in spirit to the proof of Egorov's theorem (see also Exercise 13 of Chapter 4). We will continue to use the notations in (7.56) and (7.57) for points in $\mathbf{R^n}$ and the difference quotient in the first variable.

Lemma 7.59 *Let G be an open set in $\mathbf{R^n}$, $n > 1$, and E be a measurable subset of G with $|E| < \infty$. Let f be a continuous function on G such that $\partial f/\partial x$ exists and is finite in E. Then given $\varepsilon > 0$, there exist a closed set $F \subset E$ and $\delta > 0$ such that $|E - F| < \varepsilon$, the set $\{(x+h, \mathbf{y}) : (x, \mathbf{y}) \in F, |h| \le \delta\}$ is contained in G, and $D_h f(x, \mathbf{y})$ converges uniformly to $(\partial f/\partial x)(x, \mathbf{y})$ in F as $h \to 0$. A similar result holds for the difference quotient of f in each of the other coordinate variables.*

Proof. Fix G, E, and f as in the hypothesis. For $m, k = 1, 2, \ldots$, let

$$E_{m,k} = \left\{(x, \mathbf{y}) \in E : (x+h, \mathbf{y}) \in G \text{ and } \left|D_h f(x, \mathbf{y}) - \frac{\partial f}{\partial x}(x, \mathbf{y})\right| \le \frac{1}{m} \text{ if } 0 < |h| \le \frac{1}{k}\right\}.$$

To see why each $E_{m,k}$ is measurable, first define

$$G(r) = \{(x, \mathbf{y}) \in G : (x+s, \mathbf{y}) \in G \text{ if } |s| \le |r|\}, \quad r \in \mathbf{R^1} - \{0\}.$$

Thus, $G(r)$ consists of all points $(x, \mathbf{y})$ of G such that the closed line segment of length $2|r|$ centered at $(x, \mathbf{y})$ and parallel to the x-axis lies in G. Since the distance from any compact line segment in G to the complement of G is positive (cf. Exercise 12(b) of Chapter 1), it follows that $G(r)$ is open and therefore that the set $E(r) = G(r) \cap E$ is measurable. Next, using the continuity of f on G, and letting $\mathcal{Q}$ denote the collection of all rational numbers, we can represent $E_{m,k}$ as a countable intersection as follows:

$$E_{m,k} = \bigcap_{\substack{r \in \mathcal{Q} \\ 0 < |r| \le 1/k}} \left\{(x, \mathbf{y}) \in E(r) : \left|D_r f(x, \mathbf{y}) - \frac{\partial f}{\partial x}(x, \mathbf{y})\right| \le \frac{1}{m}\right\}.$$

Consequently, $E_{m,k}$ is measurable since each set in the intersection is measurable; in fact, by the remark following the proof of Lemma 7.55, $\partial f/\partial x$ is measurable on E, and so also on $E(r)$, and $D_r f$ is measurable (even continuous) on $E(r)$.

Since $\partial f/\partial x$ exists and is finite in E, then for each m, we have $E_{m,k} \nearrow E$ as $k \nearrow \infty$. Hence, for each m,

$$\lim_{k\to\infty} \left|E_{m,k}\right| = |E|,$$

and $\left|E - E_{m,k}\right| \to 0$ as $k \to \infty$ since $|E| < \infty$. It follows that for every $\varepsilon > 0$ and $m = 1, 2, \ldots$, there is a measurable set $H_m = H_m^\varepsilon \subset E$ and an index $k_m = k_m^\varepsilon$ such that $|E - H_m| < \varepsilon 2^{-m-1}$ and

$$\left|D_h f(x, \mathbf{y}) - \frac{\partial f}{\partial x}(x, \mathbf{y})\right| \le \frac{1}{m} \quad \text{if } (x, \mathbf{y}) \in H_m \text{ and } 0 < |h| \le \frac{1}{k_m}.$$

Let $H = \bigcap_1^\infty H_m$ and $\delta = 1/k_1$. Note that if $(x, \mathbf{y}) \in H$ and $|h| \le \delta$ then $(x + h, \mathbf{y}) \in G$. Also, $D_h f$ converges uniformly to $\partial f/\partial x$ in H as $h \to 0$, and

$$|E - H| \le \sum_1^\infty \frac{\varepsilon}{2^{m+1}} = \frac{\varepsilon}{2}.$$

Now choose a closed set $F = F^\varepsilon \subset H$ with $|H - F| < \varepsilon/2$. Then $|E - F| < \varepsilon$, $(x + h, \mathbf{y}) \in G$ if $(x, \mathbf{y}) \in H$ and $|h| \le \delta$, and $D_h f$ converges uniformly in F as $h \to 0$. This proves Lemma 7.59.

The final fact that we will use to prove Theorem 7.53 is given in the next lemma.

Lemma 7.60 *Let G be an open set in $\mathbf{R^n}$, $n > 1$, and E be a measurable subset of G. Let f be a continuous function on G all of whose first partial derivatives exist and are finite in E. Then f has an approximate differential a.e. in E.*

Proof. The proof is by induction on n. Let G, E, and f satisfy the hypothesis. Fix $n \ge 2$ and assume that the result is true for dimension $n - 1$. In case $n = 2$, this inductive assumption is true since a function of a single variable has an approximate derivative at every point where it has a derivative.

We will again use the notation $(x, \mathbf{y})$ and $D_h f(x, \mathbf{y})$ in (7.56) and (7.57). Without loss of generality, we may assume that $|E| < \infty$. Let $\varepsilon > 0$ and choose δ and F as in Lemma 7.59. Then $(x + h, \mathbf{y}) \in G$ if $(x, \mathbf{y}) \in F$ and $|h| \le \delta$, $|E - F| < \varepsilon$, and $D_h f$ converges uniformly to $\partial f/\partial x$ in F. We may also assume that $|F| > 0$ and, by Lusin's theorem, that $\partial f/\partial x$ is continuous on F relative to F. Fix any $x_0 \in (-\infty, \infty)$ for which the set F_{x_0} defined by

$$F_{x_0} = \left\{\mathbf{y} \in \mathbf{R^{n-1}} : (x_0, \mathbf{y}) \in F\right\}$$

has positive $(n-1)$-dimensional measure. Note that a.e. (linear measure) x_0 such that F_{x_0} is not empty has this property by Fubini's theorem. Similarly, define

$$G_{x_0} = \left\{\mathbf{y} \in \mathbf{R^{n-1}} : (x_0, \mathbf{y}) \in G\right\}, \quad E_{x_0} = \left\{\mathbf{y} \in \mathbf{R^{n-1}} : (x_0, \mathbf{y}) \in E\right\}.$$

Then G_{x_0} is open in $\mathbf{R^{n-1}}$, and E_{x_0} is measurable in $\mathbf{R^{n-1}}$ for a.e. x_0. Also, $f(x_0, \mathbf{y})$ considered as a function of $\mathbf{y}$ is continuous in G_{x_0}, and since f has finite first partial derivatives a.e. in E (by hypothesis), we may assume that $f(x_0, \mathbf{y})$ has finite first partial derivatives with respect to $\mathbf{y}$ for a.e. $\mathbf{y} \in E_{x_0}$. Thus, by our inductive hypothesis, $f(x_0, \mathbf{y})$ has an approximate differential in $\mathbf{y}$ a.e. ($(n-1)$-dimensional measure) in E_{x_0} and so a.e. in F_{x_0}.

Now fix any $\mathbf{y}_0 \in F_{x_0}$ such that $\mathbf{y}_0$ is an $(n-1)$-dimensional point of density of F_{x_0} and such that $f(x_0, \mathbf{y})$ has a differential in $\mathbf{y}$ at $\mathbf{y} = \mathbf{y}_0$, that is, so that

$$f(x_0, \mathbf{y}) = f(x_0, \mathbf{y}_0) + \tilde{\mathbf{A}}(x_0, \mathbf{y}_0) \cdot (\mathbf{y} - \mathbf{y}_0) + o(|\mathbf{y} - \mathbf{y}_0|) \quad \text{as } \mathbf{y} \to \mathbf{y}_0, \tag{7.61}$$

where $\tilde{\mathbf{A}}(x_0, \mathbf{y}_0)$ is a vector in $\mathbf{R^{n-1}}$. Almost every point of F_{x_0} has these properties.

We claim that f has an approximate differential (relative to $\mathbf{R^n}$) at $(x_0, \mathbf{y}_0)$. With δ as above, let $H \subset \mathbf{R^n}$ be the Cartesian product $[x_0 - \delta, x_0 + \delta] \times F_{x_0}$:

$$H = \left\{(x, \mathbf{y}) \in \mathbf{R^n} : |x - x_0| \le \delta, \mathbf{y} \in F_{x_0}\right\}.$$

Then $H \subset G$, and $(x_0, \mathbf{y}_0)$ is an n-dimensional point of density of H since $\mathbf{y}_0$ is an $(n-1)$-dimensional point of density of F_{x_0}. Let $(x, \mathbf{y}) \in H$ and write

$$\mathbf{A}(x_0, \mathbf{y}_0) = \left(\frac{\partial f}{\partial x}(x_0, \mathbf{y}_0), \tilde{\mathbf{A}}(x_0, \mathbf{y}_0)\right).$$

Then $\mathbf{A}(x_0, \mathbf{y}_0)$ is a vector in $\mathbf{R^n}$ and

$$\begin{aligned}
&f(x, \mathbf{y}) - f(x_0, \mathbf{y}_0) - \mathbf{A}(x_0, \mathbf{y}_0) \cdot (x - x_0, \mathbf{y} - \mathbf{y}_0) \\
&\quad = [f(x, \mathbf{y}) - f(x_0, \mathbf{y})] + [f(x_0, \mathbf{y}) - f(x_0, \mathbf{y}_0) - \mathbf{A}(x_0, \mathbf{y}_0) \cdot (x - x_0, \mathbf{y} - \mathbf{y}_0)] \\
&\quad = \left[f(x, \mathbf{y}) - f(x_0, \mathbf{y}) - \frac{\partial f}{\partial x}(x_0, \mathbf{y})(x - x_0)\right] \\
&\qquad + \left[\frac{\partial f}{\partial x}(x_0, \mathbf{y}) - \frac{\partial f}{\partial x}(x_0, \mathbf{y}_0)\right](x - x_0) \\
&\qquad + \left[f(x_0, \mathbf{y}) - f(x_0, \mathbf{y}_0) - \tilde{\mathbf{A}}(x_0, \mathbf{y}_0) \cdot (\mathbf{y} - \mathbf{y}_0)\right] \\
&\quad = I + II + III,
\end{aligned}$$

say. Our claim will be proved by showing that each of $|I|$, $|II|$, and $|III|$ has size $o\left(|x-x_0|+\left|\mathbf{y}-\mathbf{y}_0\right|\right)$ if $(x,\mathbf{y})\in H$ and $(x,\mathbf{y})\to\left(x_0,\mathbf{y}_0\right)$.

Let $\eta>0$. Since $D_h f$ converges uniformly to $\partial f/\partial x$ in F, there exists $\nu>0$ such that $\left|D_h f-\partial f/\partial x\right|<\eta$ in F if $0<|h|<\nu$. We may assume that $\nu<\delta$. Hence, since

$$I=\left[\left(D_{x-x_0}f\right)\left(x_0,\mathbf{y}\right)-\frac{\partial f}{\partial x}\left(x_0,\mathbf{y}\right)\right](x-x_0),$$

we obtain

$$|I|<\eta\,|x-x_0| \quad \text{if } |x-x_0|<\nu \text{ and } (x,\mathbf{y})\in H.$$

Also, if $(x,\mathbf{y})\in H$, then $\left(x_0,\mathbf{y}\right)\in F$, and consequently by using the continuity of $\partial f/\partial x$ on F relative to F, we see that there exists $\nu'>0$ independent of $(x,\mathbf{y})$ such that

$$|II|<\eta\,|x-x_0| \quad \text{if } \left|\mathbf{y}-\mathbf{y}_0\right|<\nu' \text{ and } (x,\mathbf{y})\in H.$$

Finally, by (7.61), we have $|III|<\eta\left|\mathbf{y}-\mathbf{y}_0\right|$ when $\left|\mathbf{y}-\mathbf{y}_0\right|$ is sufficiently small. Our claim follows by combining estimates.

Since a.e. point in F shares the properties of the point $\left(x_0,\mathbf{y}_0\right)$ above, we conclude that f has an approximate differential relative to $\mathbf{R^n}$ a.e. in F. Recall that the value of $\varepsilon>0$ can be arbitrarily small and that the set $F=F^{\varepsilon}$ satisfies $F\subset E$ and $|E-F|<\varepsilon$. Now let $\varepsilon\to 0$ through a sequence $\{\varepsilon_m\}$. Then f has an approximate differential a.e. in the set $\bigcup_m F^{\varepsilon_m}$, which is a subset of E of full measure. This completes the proof of the Lemma 7.60.

Proof of Theorem 7.53. Let G be an open set in $\mathbf{R^n}, n>1$, and $f\in Lip_{loc}(G)$. By Lemma 7.55, f has measurable first partial derivatives a.e. in G. The fact that these derivatives are bounded a.e. on every compact subset of G follows from their definitions as limits of difference quotients since $f\in Lip_{loc}(G)$. In fact, if $\left\{G_j : j=1,2,\ldots\right\}$ is a sequence of bounded open sets with closures contained in G such that $G_j\nearrow G$, for example,

$$G_j=\left\{\mathbf{x}\in G : |\mathbf{x}|<j \text{ and } d(\mathbf{x},CG)>1/j\right\}, \quad j=1,2,\ldots,$$

then $f\in Lip\left(G_j\right)$ for each j, and every compact set in G lies in some G_j. Clearly, the first partial derivatives of a function that is Lipschitz continuous in an open set are bounded a.e. in that set.

Thus, it remains only to show that f has a total first differential a.e. in G. However, this is an immediate consequence of Lemma 7.60 and Theorem 7.50, which completes the proof of the Rademacher–Stepanov theorem.

An examination of the proof of Theorem 7.53 shows that the assumption of local Lipschitz continuity of f is used in two key ways, first in order to show that f has measurable partial derivatives a.e. in G and again in order to conclude that f has a total differential wherever it has an approximate one (Theorem 7.50). Note that the part of the proof showing that f has an approximate differential a.e. in G (Lemmas 7.59 and 7.60) does not require Lipschitz continuity, although it uses continuity of f in G and relies on the existence of the first partial derivatives of f a.e. in G. In order to transfer some of the proof technique to other situations, we can bypass the first way that Lipschitz continuity is used by simply *assuming* that all $\partial f/\partial x_i$ exist and are finite. With this assumption, the next result gives a condition different from Lipschitz continuity that implies total differentiability. The condition is stated as follows in terms of the size of the second difference of f.

We say that a function f that is defined and finite in a neighborhood of a point $\mathbf{x} \in \mathbf{R^n}$ is *smooth at* $\mathbf{x}$ if

$$f(\mathbf{x}+\mathbf{h}) + f(\mathbf{x}-\mathbf{h}) - 2f(\mathbf{x}) = o(|\mathbf{h}|) \quad \text{as } |\mathbf{h}| \to 0. \tag{7.62}$$

The intuitive reason for calling such an f smooth at $\mathbf{x}$ arises from the one-dimensional situation, noting that

$$\frac{f(x+h)+f(x-h)-2f(x)}{h} = \frac{f(x+h)-f(x)}{h} - \frac{f(x-h)-f(x)}{-h}, \quad h \neq 0,$$

and consequently, if f is smooth and has a finite derivative from either side at x, then it has a derivative from both sides at x, and the two are the same. Thus, the graph of a function that is smooth at a point cannot have a corner there.

Theorem 7.63 *Let G be an open set in $\mathbf{R^n}$, $n > 1$, and E be a measurable subset of G. Let f be a continuous function on G such that*

(i) *f has finite partial derivatives $\partial f/\partial x_i$ in E, $i = 1, \ldots, n$,*
(ii) *f satisfies the smoothness condition (7.62) at every $\mathbf{x} \in E$.*

Then each $\partial f/\partial x_i$ is measurable on E, and f has a total first differential a.e. in E.

Proof. We will be brief. Fix G, E, and f as in the hypothesis. The measurability on E of the derivatives $\partial f/\partial x_i$ follows from the remark after the proof of

Lemma 7.55. It remains to show that f has a total first differential a.e. in E. For $m, k = 1, 2, \ldots$, define

$$E_{m,k} = \left\{ \mathbf{x} \in E : \frac{|f(\mathbf{x}+\mathbf{h}) + f(\mathbf{x}-\mathbf{h}) - 2f(\mathbf{x})|}{|\mathbf{h}|} \leq \frac{1}{m} \text{ if } 0 < |\mathbf{h}| \leq \frac{1}{k} \right\}.$$

By (7.62), for every m, $E_{m,k} \nearrow E$ as $k \nearrow \infty$. Every $E_{m,k}$ is measurable since the continuity of f in G gives

$$E_{m,k} = \bigcap_{\substack{\mathbf{r} \in \mathcal{Q}^{\mathbf{n}} \\ 0 < |\mathbf{r}| \leq 1/k}} \left\{ (x, \mathbf{y}) \in E : \frac{|f(\mathbf{x}+\mathbf{r}) + f(\mathbf{x}-\mathbf{r}) - 2f(\mathbf{x})|}{|\mathbf{r}|} \leq \frac{1}{m} \right\},$$

where $\mathcal{Q}^{\mathbf{n}}$ denotes the (countable) collection of all points of $\mathbf{R}^{\mathbf{n}}$ with rational coordinates. Assuming as we may that $|E| < \infty$, we obtain that $\left|E - E_{m,k}\right| \to 0$ as $k \to \infty$ for every m. Given $\varepsilon > 0$, it follows (cf. the proof of Lemma 7.59) that there is a measurable set $H \subset E$ with $|E - H| < \varepsilon$ such that

$$\frac{|f(\mathbf{x}+\mathbf{h}) + f(\mathbf{x}-\mathbf{h}) - 2f(\mathbf{x})|}{|\mathbf{h}|} \to 0 \text{ uniformly in } H \text{ as } |\mathbf{h}| \to 0.$$

To complete the proof of the theorem, it is enough to show that f has a total differential a.e. in H. By Lemma 7.60, f has an approximate differential a.e. in E and therefore also a.e. in H. Let $\mathbf{x}_0 \in H$ be a point of density of H where f has an approximate differential. It suffices to prove that f has a total differential at $\mathbf{x}_0$. By the definition of an approximate differential, there is a vector $\mathbf{A}_0 \in \mathbf{R}^{\mathbf{n}}$ and a measurable set, which we may assume is the same as the set H, having $\mathbf{x}_0$ as a point of density such that

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \mathbf{A}_0 \cdot \mathbf{h} = o(|\mathbf{h}|), \quad \mathbf{h} \in H, \text{ as } |\mathbf{h}| \to 0.$$

Without loss of generality, we may assume that $\mathbf{x}_0 = \mathbf{0}$, $f(\mathbf{x}_0) = 0$, and $\mathbf{A}_0 = \mathbf{0}$. Thus, given $\eta > 0$, there exists $\rho_0 > 0$ such that both

$$\sup_{\substack{\mathbf{x},\mathbf{h} \\ \mathbf{x} \in H, |\mathbf{h}| \leq \rho}} |f(\mathbf{x}+\mathbf{h}) + f(\mathbf{x}-\mathbf{h}) - 2f(\mathbf{x})| \leq \eta\rho \quad \text{if } 0 \leq \rho \leq \rho_0$$

and

$$|f(\mathbf{h})| \leq \eta|\mathbf{h}| \quad \text{if } \mathbf{h} \in H \text{ and } 0 \leq |\mathbf{h}| \leq \rho_0.$$

If ρ is sufficiently small, then every $\mathbf{u} \in \mathbf{R}^{\mathbf{n}}$ satisfying $|\mathbf{u}| < \rho$ can be expressed in the form $\mathbf{u} = \mathbf{x} + \mathbf{h}$ with $\mathbf{x} - \mathbf{h}, \mathbf{x} \in H$ and $|\mathbf{x}|, |\mathbf{h}| < \rho$ (see Exercise 30).

Combining estimates, we then obtain that for all sufficiently small $\rho > 0$ and all $\mathbf{u}$ with $|\mathbf{u}| < \rho$,

$$\begin{aligned}|f(\mathbf{u})| = |f(\mathbf{x}+\mathbf{h})| &\le |f(\mathbf{x}+\mathbf{h}) + f(\mathbf{x}-\mathbf{h}) - 2f(\mathbf{x})| + |f(\mathbf{x}-\mathbf{h})| + 2|f(\mathbf{x})| \\ &\le \eta\rho + \eta\, 2\rho + 2\eta\rho = 5\eta\rho,\end{aligned}$$

and the theorem follows.

In passing, we will derive an interesting result about extending a Lipschitz function from a set to the entire space in such a way that the extended function remains Lipschitz continuous. The result is not limited to functions defined in open sets. If E is any set in $\mathbf{R^n}$ and f is a finite real-valued function defined on E, then f is said to be *Lipschitz continuous on E* if there is a constant C such that

$$|f(\mathbf{x}) - f(\mathbf{y})| \le C|\mathbf{x}-\mathbf{y}|, \quad \mathbf{x}, \mathbf{y} \in E.$$

The smallest such C, namely, the constant

$$C_{f,E} = \sup_{\substack{\mathbf{x},\mathbf{y}\in E \\ \mathbf{x}\neq\mathbf{y}}} \frac{|f(\mathbf{x}) - f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|},$$

is called the *Lipschitz constant of f on E*.

We have the following extension result for such functions.

Theorem 7.64 *Let f be Lipschitz continuous on a set $E \subset \mathbf{R^n}$. Then f can be extended to $\mathbf{R^n}$ as a Lipschitz function with the same Lipschitz constant, that is, there is a function $f_1 \in Lip\,(\mathbf{R^n})$ such that $f_1 = f$ on E and $C_{f_1,\mathbf{R^n}} = C_{f,E}$.*

Proof. Part of the proof will be left as an exercise. Let f and E be as in the hypothesis and denote $C = C_{f,E}$. Note that if $\mathbf{y}, \mathbf{y}_0 \in E$ and $\mathbf{x} \in \mathbf{R^n}$, then $f(\mathbf{y}_0) - C|\mathbf{x}-\mathbf{y}_0| \le f(\mathbf{y}) + C|\mathbf{x}-\mathbf{y}|$ since $f(\mathbf{y}_0) - f(\mathbf{y}) \le C|\mathbf{y}_0 - \mathbf{y}| \le C(|\mathbf{x}-\mathbf{y}_0| + |\mathbf{x}-\mathbf{y}|)$. Therefore, the conical functions $\gamma_{\mathbf{y}}(\mathbf{x})$ defined for $\mathbf{y} \in E$ and $\mathbf{x} \in \mathbf{R^n}$ by

$$\gamma_{\mathbf{y}}(\mathbf{x}) = \gamma_{\mathbf{y},f}(\mathbf{x}) = f(\mathbf{y}) + C|\mathbf{x}-\mathbf{y}|$$

satisfy

$$\inf_{\mathbf{y}\in E} \gamma_{\mathbf{y}}(\mathbf{x}) \ge f(\mathbf{y}_0) - C|\mathbf{x}-\mathbf{y}_0| > -\infty \qquad (\mathbf{y}_0 \in E)$$

for all $\mathbf{x} \in \mathbf{R^n}$. Also, for any $\mathbf{y}_0 \in E$, $\inf_{\mathbf{y}\in E} \gamma_{\mathbf{y}}(\mathbf{x}) \leq \gamma_{\mathbf{y}_0}(\mathbf{x}) < +\infty$. Define

$$f_1(\mathbf{x}) = \inf_{\mathbf{y}\in E} \gamma_{\mathbf{y}}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n}. \tag{7.65}$$

Then (see Exercise 32) $f_1(\mathbf{x}) = f(\mathbf{x})$ if $\mathbf{x} \in E$, $f_1 \in Lip\,(\mathbf{R^n})$ and $C_{f_1,\mathbf{R^n}} = C = C_{f,E}$.

Exercises

1. Let f be measurable in $\mathbf{R^n}$ and different from zero in some set of positive measure. Show that there is a positive constant c such that $f^*(\mathbf{x}) \geq c|\mathbf{x}|^{-n}$ for $|\mathbf{x}| \geq 1$.
2. Let $\phi(\mathbf{x}), \mathbf{x} \in \mathbf{R^n}$, be a bounded measurable function such that $\phi(\mathbf{x}) = 0$ for $|\mathbf{x}| \geq 1$ and $\int \phi = 1$. For $\varepsilon > 0$, let $\phi_\varepsilon(\mathbf{x}) = \varepsilon^{-n}\phi(\mathbf{x}/\varepsilon)$. ($\phi_\varepsilon$ is called an *approximation to the identity.*) If $f \in L(\mathbf{R^n})$, show that

$$\lim_{\varepsilon \to 0} \left(f * \phi_\varepsilon\right)(x) = f(\mathbf{x})$$

in the Lebesgue set of f. (Note that $\int \phi_\varepsilon = 1$, $\varepsilon > 0$, so that

$$\left(f * \phi_\varepsilon\right)(\mathbf{x}) - f(\mathbf{x}) = \int [f(\mathbf{x} - \mathbf{y}) - f(\mathbf{x})]\phi_\varepsilon(\mathbf{y})\, d\mathbf{y}.$$

Use Theorem 7.16.)
3. Show that the conclusion of Lemma 7.4 remains true for the case of two dimensions if instead of being squares, the sets Q covering E are rectangles with x-dimension equal to h and y-dimension equal to h^2. (Of course, h varies with Q and the rectangles have edges parallel to the coordinate axes.) Show that the conclusion of Theorem 7.2 remains valid for $n = 2$ if the cubes Q are replaced by rectangles of this type that are centered at $\mathbf{x}$ and with $h \to 0$.

   Show that the same conclusions are valid for rectangles whose x-dimension is h and whose y-dimension is any fixed increasing function of h. Generalize this to higher dimensions.
4. If E_1 and E_2 are measurable subsets of $\mathbf{R^1}$ with $|E_1| > 0$ and $|E_2| > 0$, prove that the set $\{x : x = x_1 - x_2, x_1 \in E_1, x_2 \in E_2\}$ contains an interval. (cf. Lemma 3.37.)

5. Let f be of bounded variation on $[a, b]$. If $f = g + h$, where g is absolutely continuous and h is singular, show that

$$\int_a^b \phi \, df = \int_a^b \phi f' \, dx + \int_a^b \phi \, dh$$

for any continuous ϕ.

6. Show that if $\alpha > 0$, then x^α is absolutely continuous on every bounded subinterval of $[0, \infty)$.

7. Prove that f is absolutely continuous on $[a, b]$ if and only if given $\varepsilon > 0$, there exists $\delta > 0$ such that $\left|\sum [f(b_i) - f(a_i)]\right| < \varepsilon$ for every *finite* collection $\{[a_i, b_i]\}$ of nonoverlapping subintervals of $[a, b]$ with $\sum (b_i - a_i) < \delta$.

8. Prove the following converse of Theorem 7.31: If f is of bounded variation on $[a, b]$, and if the function $V(x) = V[a, x]$ is absolutely continuous on $[a, b]$, then f is absolutely continuous on $[a, b]$.

9. If f is of bounded variation on $[a, b]$, show that

$$\int_a^b |f'| \leq V[a, b].$$

Show that if equality holds in this inequality, then f is absolutely continuous on $[a, b]$. (For the second part, use Theorems 2.2(ii) and 7.24 to show that $V(x)$ is absolutely continuous and then use the result of Exercise 8.)

10. (a) Show that if f is absolutely continuous on $[a, b]$ and Z is a subset of $[a, b]$ of measure zero, then the image set defined by $f(Z) = \{w : w = f(z), z \in Z\}$ also has measure zero. Deduce that the image under f of any measurable subset of $[a, b]$ is measurable. (Compare Theorem 3.33.) (Hint: use the fact that the image of an interval $[a_i, b_i]$ is an interval of length at most $V(b_i) - V(a_i)$.)

(b) Give an example of a strictly increasing Lipschitz continuous function f and a set Z with measure 0 such that $f^{-1}(Z)$ does not have measure 0 (and consequently, f^{-1} is not absolutely continuous). (Let $f^{-1}(x) = x + C(x)$ on $[0, 1]$, where $C(x)$ is the Cantor–Lebesgue function.)

11. Prove the following result concerning *changes of variable.* Let $g(t)$ be monotone increasing and absolutely continuous on $[\alpha, \beta]$ and let f be integrable on $[a, b]$, $a = g(\alpha)$, $b = g(\beta)$. Then $f(g(t))g'(t)$ is measurable and integrable on $[\alpha, \beta]$, and

$$\int_a^b f(x) \, dx = \int_\alpha^\beta f(g(t))g'(t) \, dt.$$

(Consider the cases when f is the characteristic function of an interval, an open set, etc.)

12. Use Jensen's inequality to prove that if $a, b \geq 0$, $p, q > 1$, $(1/p) + (1/q) = 1$, then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

More generally, show that

$$a_1 \cdots a_N \leq \sum_{j=1}^{N} \frac{a_j^{p_j}}{p_j},$$

where $a_j \geq 0$, $p_j > 1$, $\sum_{j=1}^{N}(1/p_j) = 1$. (Write $a_j = e^{x_j/p_j}$ and use the convexity of e^x.)

13. Prove Theorem 7.36.

14. Prove that ϕ is convex on (a, b) if and only if it is continuous and

$$\phi\left(\frac{x_1 + x_2}{2}\right) \leq \frac{\phi(x_1) + \phi(x_2)}{2}$$

for $x_1, x_2 \in (a, b)$.

15. Theorem 7.43 shows that a convex function is the indefinite integral of a monotone increasing function. Prove the converse: If $\phi(x) = \int_a^x f(t)dt + \phi(a)$ in (a, b) and f is monotone increasing, then ϕ is convex in (a, b). (Use Exercise 14.)

16. Show that the formula

$$\int_{-\infty}^{+\infty} fg' = -\int_{-\infty}^{+\infty} f'g$$

for integration by parts may not hold if f is of bounded variation on $(-\infty, +\infty)$ and g is infinitely differentiable with compact support. (Let f be the Cantor–Lebesgue function on $[0, 1]$, and let $f = 0$ elsewhere.)

17. A sequence $\{\phi_k\}$ of set functions is said to be *uniformly absolutely continuous* if given $\varepsilon > 0$, there exists $\delta > 0$ such that if E satisfies $|E| < \delta$, then $|\phi_k(E)| < \varepsilon$ for all k. If $\{f_k\}$ is a sequence of integrable functions on $(0, 1)$ which converges pointwise a.e. to an integrable f, show that $\int_0^1 |f - f_k| \to 0$ if and only if the indefinite integrals of the f_k are uniformly absolutely continuous. (cf. Exercise 23 of Chapter 10.)

18. Prove the following set-theoretic result related to the simple Vitali covering lemma. If $\mathscr{C} = \{Q\}$ is a collection of cubes all contained in a fixed bounded set in $\mathbf{R^n}$, then there is a countable subcollection $\{Q_k\}$ of disjoint cubes in $\mathscr{C}$ such that every $Q \in \mathscr{C}$ is contained in some Q_k^*, where Q_k^* denotes the cube concentric with Q_k of edge length 5 times that of Q_k.

Deduce the measure-theoretic consequence (cf. Lemma 7.4) that if a set E is covered by such a collection $\mathscr{C}$ of cubes, then there exist $\beta > 0$, depending only on n, and a finite number of disjoint cubes $Q_1, \ldots, Q_N$ in $\mathscr{C}$ such that $\beta|E|_e \leq \sum_{k=1}^{N} |Q_k|$.

Formulate analogues of these facts for a collection of balls in $\mathbf{R^n}$.

19. Use Exercise 18 to prove Lemma 7.9.

20. (a) Let $f(\mathbf{x})$ be defined for all $\mathbf{x} \in \mathbf{R^n}$ by $f(\mathbf{x}) = 0$ if every coordinate of $\mathbf{x}$ is rational, and $f(\mathbf{x}) = 1$ otherwise. Describe the set of all $\mathbf{x}$ at which $\frac{1}{|Q|}\int_Q f$ has a limit as $Q \searrow x$ and describe all Lebesgue points of f.

(b) Give an example of a bounded function f on $(-\infty, \infty)$ with the following properties: f is continuous except at a single point x_0; $(d/dx)\int_0^x f = f(x)$ for all x (in particular when $x = x_0$); x_0 is not a Lebesgue point of f.

21. For $\mathbf{x} \in \mathbf{R^n}$ and $0 < \alpha < n$, define $f(\mathbf{x}) = |\mathbf{x}|^{-\alpha}\chi_{\{|\mathbf{x}|<1\}}(\mathbf{x})$. Show that its maximal function $f^*(\mathbf{x})$ is bounded both above and below by positive constants (depending only on α and n) times $(|\mathbf{x}|^\alpha + |\mathbf{x}|^n)^{-1}$.

22. In order to better understand Theorem 7.24 and its proof on an intuitive level, it may be helpful to sketch the graphs of a particularly simple continuously differentiable function $f(x)$ and its variation $V(x)$ on the same set of axes, and then to compare f' and V'. Draw both graphs for $f(x) = \sin x$ on $[0, 2\pi]$ or more generally for any f defined on a closed interval of finite length whose graph is a finite union of smooth monotone arcs.

23. Show that a convex function on (a, b) cannot attain a maximum on (a, b) unless it is constant. Is the same true for a local maximum?

24. Suppose that ϕ is continuous on (a, b), ϕ' exists and is finite on (a, b) except at a single point ξ, and ϕ' is increasing on $(a, \xi) \cup (\xi, b)$. Show that ϕ is convex on (a, b).

25. Let $1 \leq p < \infty$. Show that $x^p \log(1 + x)$ and $x^p\left(1 + \log^+ x\right)$ are convex on $(0, \infty)$. (For the second function, Exercise 24 may be helpful.) See also Exercise 28 in Chapter 8.

26. Show that the function defined in (7.48) is Lipschitz continuous on $\mathbf{R^2}$.

27. Verify the uniqueness of $\mathbf{A}$ in (7.49) in the sense described immediately after (7.49). (Show that if $\mathbf{0}$ is a point of density of a measurable set $E \subset \mathbf{R^n}$ and $\mathbf{B}$ is a nonzero vector in $\mathbf{R^n}$, then there is a sequence $\{\mathbf{h}_j\} \subset E$ such that $|\mathbf{h}_j| \to 0$, $|\mathbf{h}_j| \neq 0$, and the angle between $\mathbf{h}_j$ and $\mathbf{B}$ is bounded away from $\pi/2$ uniformly in j.)

28. Let $f \in Lip\,(\mathbf{R^n})$ and T be a nonsingular linear transformation of $\mathbf{R^n}$. Show that the function $f \circ T$ defined by $(f \circ T)(\mathbf{x}) = f(T\mathbf{x})$ belongs to $Lip\,(\mathbf{R^n})$ and that for a.e. $\mathbf{x} \in \mathbf{R^n}$, $\nabla(f \circ T)(\mathbf{x}) = T^* \left((\nabla f)(T\mathbf{x})\right)$, where T^* denotes the adjoint (transpose) of T.

29. (a) In one dimension, give an example of a Lipschitz function f that satisfies the smoothness condition $f\,(x_0 + h) + f\,(x_0 - h) - 2f\,(x_0) = o(|h|)$ as $h \to 0$ at a point x_0 but for which $f'(x_0)$ does not exist.

(b) Show that any (finite) f that satisfies the smoothness condition in part (a) at a point x_0 where f has a local extremum is differentiable at x_0 and $f'\,(x_0) = 0$. (Express the ratio $[f\,(x_0 + h) + f\,(x_0 - h) - 2f\,(x_0)]\,/h$ in terms of right- and left-hand difference quotients of f at x_0.)

(c) Suppose that a function f satisfies the smoothness condition in part (a) for every x in an interval (a, b) of the real line and that f is also continuous in (a, b). Show that f' exists and is finite in a dense subset of (a, b). (If $[a', b'] \subset (a, b)$, apply the result in part (b) to the function $f(x) - L(x)$ where L is the chord equal to f at $x = a', b'$.)

(d) Prove the following generalization of the classical mean value theorem: if f is continuous in an interval $[a, b] \subset (-\infty, \infty)$ and smooth (in the sense of part (a)) in (a, b), then there exists $c \in (a, b)$ such that $f(b) - f(a) = f'(c)\,(b - a)$.

30. Let H be a measurable set in $\mathbf{R^n}$ that has $\mathbf{0}$ as a point of density. For $\rho > 0$, let H_ρ be the intersection of H with the ball B_ρ of radius ρ centered at $\mathbf{0}$.

(a) Show that the set $\left\{2\mathbf{x} - \mathbf{y} : \mathbf{x}, \mathbf{y} \in H_\rho\right\}$ covers B_ρ if ρ is sufficiently small. Deduce that the set $\left\{\mathbf{x} + \mathbf{h} : \mathbf{x}, \mathbf{x} - \mathbf{h} \in H_\rho, \mathbf{h} \in \mathbf{R^n}\right\}$ covers B_ρ if ρ is small. (For the first part, compare Lemma 3.37. Recall from Theorem 3.35 that if E is a measurable set in $\mathbf{R^n}$, then the set $2E$ defined by $2E = \{2\mathbf{x} : \mathbf{x} \in E\}$ has measure $|2E| = 2^n|E|$. Note that $2H_\rho$ and the set obtained by translating H_ρ by any fixed point in B_ρ are both subsets of $B_{2\rho}$.)

(b) While the result in part (a) is adequate for the purpose of proving Theorem 7.63, it can be improved and generalized. If r, s are real numbers with $0 < |s| \le |r|$, show that the set $\left\{r\mathbf{x} + s\mathbf{y} : \mathbf{x}, \mathbf{y} \in H_\rho\right\}$ covers $B_{|r|\rho}$ if ρ is sufficiently small. (Argue as for part (a), but consider only translations of an appropriate subset of $-sH_\rho$.)

31. Let f be a finite measurable function on $\mathbf{R^n}$ that satisfies $f(\mathbf{x} + \mathbf{h}) + f(\mathbf{x} - \mathbf{h}) - 2f(\mathbf{x}) = O(1)$ uniformly in $\mathbf{x}, \mathbf{h}$ for $|\mathbf{h}| < \delta$, that is, there are positive constants A, δ such that $|f(\mathbf{x} + \mathbf{h}) + f(\mathbf{x} - \mathbf{h}) - 2f(\mathbf{x})| \le A$ if $\mathbf{x}, \mathbf{h} \in \mathbf{R^n}$ and $|\mathbf{h}| < \delta$. Show that f is bounded on every bounded set in $\mathbf{R^n}$. (First show that f is bounded on some ball in $\mathbf{R^n}$. To do so, note that there is a measurable set E with $|E| > 0$ on which f is bounded. Pick a point of density of E and apply Exercise 30.) Generalize the result to functions f that satisfy $f(\mathbf{x} + a\mathbf{y}) + f(\mathbf{x} + b\mathbf{y}) - 2f(\mathbf{x}) = O(1)$ for fixed real $a, b \neq 0$.

32. Complete the proof of the extension Theorem 7.64 by verifying the properties listed in the last line of its proof for the function f_1 defined in (7.65).

33. Let f be defined and Lipschitz continuous on a measurable set $E \subset \mathbf{R^n}$. Show that f has an approximate differential relative to E at almost every point of E.

8

L^p Classes

8.1 Definition of L^p

If E is a measurable subset of $\mathbf{R^n}$ and p satisfies $0 < p < \infty$, then $L^p(E)$ denotes the collection of measurable f for which $\int_E |f|^p$ is finite, that is,

$$L^p(E) = \left\{ f : \int_E |f|^p < +\infty \right\}, \quad 0 < p < \infty.$$

Here, f may be complex-valued (see Exercise 3 of Chapter 4 for the definition of measurability of vector-valued functions). In this case, if $f = f_1 + if_2$ for measurable real-valued f_1 and f_2, we have $|f|^2 = f_1^2 + f_2^2$, so that

$$|f_1|, |f_2| \le |f| \le |f_1| + |f_2|.$$

It follows that $f \in L^p(E)$ if and only if both $f_1, f_2 \in L^p(E)$. See Exercise 1.

We shall write

$$||f||_{p,E} = \left(\int_E |f|^p \right)^{1/p}, \quad 0 < p < \infty;$$

thus, $L^p(E)$ is the class of measurable f for which $||f||_{p,E}$ is finite. Whenever it is clear from context what E is, we will write L^p for $L^p(E)$ and $||f||_p$ for $||f||_{p,E}$. Note that $L^1 = L$.

In order to define $L^\infty(E)$, let f be real-valued and measurable on a set E of positive measure. Define the *essential supremum* of f on E as follows: If $|\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}| > 0$ for all real α, let $\text{ess}_E \sup f = +\infty$; otherwise, let

$$\operatorname*{ess\,sup}_E f = \inf\{\alpha : |\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}| = 0\}.$$

Since the distribution function $\omega(\alpha) = |\{\mathbf{x} \in E : f(\mathbf{x}) > \alpha\}|$ is continuous from the right (see Lemma 5.39), it follows that $\omega(\text{ess}_E \sup f) = 0$ if $\text{ess}_E \sup f$

is finite. Therefore, $\text{ess}_E \sup f$ is the smallest number M, $-\infty \leq M \leq +\infty$, such that $f(\mathbf{x}) \leq M$ except for a subset of E of measure zero.

In the definition of $\text{ess}_E \sup f$, we have assumed that $|E| \neq 0$. If the same definition were used when $|E| = 0$, then $\text{ess}_E \sup f$ would be $-\infty$ for every real-valued f defined on E, resulting in incorrect or awkward statements of results such as Theorem 8.1. We will avoid technical difficulty of this type by adopting the convention that $\text{ess}_E \sup f$ is 0 when $|E| = 0$. This may be considered an analogue of the fact that $\int_E f = 0$ when $|E| = 0$. In practice, we will use the convention only when $f \geq 0$ and $|E| = 0$.

A real- or complex-valued measurable f is said to be *essentially bounded*, or simply *bounded*, on E if $\text{ess}_E \sup |f|$ is finite. By convention, if $|E| = 0$, then every function is essentially bounded on E and has essential supremum equal to 0. The class of all functions that are essentially bounded on E is denoted by $L^\infty(E)$. Clearly, f belongs to $L^\infty(E)$ if and only if its real and imaginary parts do. We shall write

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

Thus, $||f||_\infty$ is the smallest M such that $|f| \leq M$ a.e. in E, and

$$L^\infty = L^\infty(E) = \{f : ||f||_\infty < +\infty\}.$$

The following theorem gives some motivation for this notation.

Theorem 8.1 *If $|E| < +\infty$, then $||f||_\infty = \lim_{p\to\infty} ||f||_p$.*

Proof. We may assume that $|E| > 0$ since $\|f\|_\infty$ and $\|f\|_p$ are both 0 if $|E| = 0$. Let $M = \|f\|_\infty$. If $M' < M$, then the set $A = \{\mathbf{x} \in E : |f(\mathbf{x})| > M'\}$ has positive measure. Moreover, $||f||_p \geq \left(\int_A |f|^p\right)^{1/p} \geq M'|A|^{1/p}$. Since $|A|^{1/p} \to 1$ as $p \to \infty$, it follows that $\liminf_{p\to\infty} ||f||_p \geq M'$ and therefore that $\liminf_{p\to\infty} ||f||_p \geq M$. However, we also have $||f||_p \leq \left(\int_E M^p\right)^{1/p} = M|E|^{1/p}$. Hence, $\limsup_{p\to\infty} ||f||_p \leq M$, which completes the proof.

Remark: This result may fail if $|E| = +\infty$. Consider, for example, the constant function $f(x) = c$, $c \neq 0$, in $(0, \infty)$. Clearly, $f \in L^\infty$ but $f \notin L^p$ for $0 < p < \infty$. See also Exercise 26.

We will now study some basic properties of the L^p classes.

Theorem 8.2 *If $0 < p_1 < p_2 \leq \infty$ and $|E| < +\infty$, then $L^{p_2} \subset L^{p_1}$.*

Proof. Write $E = E_1 \cup E_2$, E_1 being the set where $|f| \leq 1$ and E_2 the set where $|f| > 1$. Then

$$\int_E |f|^p = \int_{E_1} |f|^p + \int_{E_2} |f|^p, \quad 0 < p < \infty.$$

The first term on the right is majorized by $|E_1|$; the second increases with p since its integrand exceeds 1. It follows that if $f \in L^{p_2}, p_2 < \infty$, then $f \in L^{p_1}$, $p_1 < p_2$. If $p_1 < p_2 = \infty$ and $f \in L^\infty$ then f is a bounded function on a set of finite measure and so belongs to L^{p_1}.

Remarks

(i) In Theorem 8.2, the hypothesis that E have finite measure cannot be omitted: for example, x^{-1/p_1} belongs to $L^{p_2}(1, \infty)$ if $p_2 > p_1$ but does not belong to $L^{p_1}(1, \infty)$. Again, any nonzero constant is in L^∞, but is not in $L^{p_1}(E)$ if $|E| = +\infty$ and $p_1 < \infty$.

(ii) A function may belong to all L^{p_1} with $p_1 < p_2$ and yet not belong to L^{p_2}. In fact, if $p_2 < \infty$, x^{-1/p_2} belongs to $L^{p_1}(0, 1)$, $p_1 < p_2$, but does not belong to $L^{p_2}(0, 1)$; $\log(1/x)$ is in $L^{p_1}(0, 1)$ for $p_1 < \infty$, but is not in $L^\infty(0, 1)$.

(iii) We leave it to the reader to show that any function that is bounded on E ($|E| < +\infty$ or not) and that belongs to $L^{p_1}(E)$ also belongs to $L^{p_2}(E), p_2 > p_1$.

The next theorem states that the L^p classes are vector (i.e., linear) spaces. Its proof is left as an exercise.

Theorem 8.3 *If $f, g \in L^p(E), p > 0$, then $f + g \in L^p(E)$ and $cf \in L^p(E)$ for any constant c.*

8.2 Hölder's Inequality and Minkowski's Inequality

In order to discuss the integrability of the product of two functions, we will use the following basic result.

Theorem 8.4 (Young's Inequality) *Let $y = \phi(x)$ be continuous, real-valued, and strictly increasing for $x \geq 0$ and let $\phi(0) = 0$. If $x = \psi(y)$ is the inverse of ϕ, then for $a, b > 0$,*

$$ab \leq \int_0^a \phi(x)\,dx + \int_0^b \psi(y)\,dy.$$

Equality holds if and only if $b = \phi(a)$.

Proof. A geometric proof is immediate if we interpret each term as an area and remember that the graph of ϕ also serves as that of ψ if we interchange the x- and y-axes. Equality holds if and only if the point (a, b) lies on the graph of ϕ.

If $\phi(x) = x^\alpha, \alpha > 0$, then $\psi(y) = y^{1/\alpha}$, and Young's inequality becomes $ab \leq a^{1+\alpha}/(1+\alpha) + b^{1+1/\alpha}/(1+1/\alpha)$. Setting $p = \alpha + 1$ and $p' = 1 + 1/\alpha$, we obtain

$$ab \leq \frac{a^p}{p} + \frac{b^{p'}}{p'} \text{ if } a, b \geq 0, 1 < p < \infty, \text{ and } \frac{1}{p} + \frac{1}{p'} = 1. \tag{8.5}$$

Two numbers p and p' that satisfy $1/p + 1/p' = 1, p, p' > 1$, are called *conjugate exponents*. Note that $p' = p/(p-1)$, and that 2 is self-conjugate. We will adopt the conventions that $p' = \infty$ if $p = 1$, and $p' = 1$ if $p = \infty$.

Theorem 8.6 (Hölder's Inequality) *If* $1 \leq p \leq \infty$ *and* $1/p + 1/p' = 1$, *then* $||fg||_1 \leq ||f||_p ||g||_{p'}$; *that is,*

$$\int_E |fg| \leq \left(\int_E |f|^p\right)^{1/p} \left(\int_E |g|^{p'}\right)^{1/p'}, \quad 1 < p < \infty;$$

$$\int_E |fg| \leq \left(\operatorname*{ess\,sup}_E |f|\right) \int_E |g|.$$

Proof. The last inequality, which corresponds to the case $p = \infty$, is obvious. Let us suppose then that $1 < p < \infty$. In case $||f||_p = ||g||_{p'} = 1$, (8.5) implies that

$$\int_E |fg| \le \int_E \left(\frac{|f|^p}{p} + \frac{|g|^{p'}}{p'} \right) = \frac{||f||_p^p}{p} + \frac{||g||_{p'}^{p'}}{p'}$$
$$= \frac{1}{p} + \frac{1}{p'} = 1 = ||f||_p ||g||_{p'}.$$

For the general case, we may assume that neither $||f||_p$ nor $||g||_{p'}$ is zero; otherwise, fg is zero a.e. in E, and the result is immediate. We may also assume that neither $||f||_p$ nor $||g||_{p'}$ is infinite. If we set $f_1 = f/||f||_p$ and $g_1 = g/||g||_{p'}$, then $||f_1||_p = ||g_1||_{p'} = 1$. Therefore, by the case already considered, we have $\int_E |f_1 g_1| \le 1$; that is, $\int_E |fg| \le ||f||_p ||g||_{p'}$, as desired.

See Exercise 4 concerning the case of equality in Hölder's inequality.

The case $p = p' = 2$ of Hölder's inequality is a classical inequality:

Corollary 8.7 (Schwarz's Inequality)

$$\int_E |fg| \le \left(\int_E |f|^2 \right)^{1/2} \left(\int_E |g|^2 \right)^{1/2}.$$

The theorem that follows is usually referred to as the *converse of Hölder's inequality* (see also Exercise 15 in Chapter 10).

Theorem 8.8 *Let f be real-valued and measurable on E, and let $1 \le p \le \infty$ and $1/p + 1/p' = 1$. Then*

$$||f||_p = \sup \int_E fg, \tag{8.9}$$

where the supremum is taken over all real-valued g such that $\|g\|_{p'} \le 1$ and $\int_E fg$ exists.

Proof. That the left-hand side of (8.9) majorizes the right-hand side follows from Hölder's inequality. To show the opposite inequality, let us consider first the case of $f \ge 0$, $1 < p < \infty$.

If $||f||_p = 0$, then $f = 0$ a.e. in E, and the result is obvious. If $0 < ||f||_p < +\infty$, we may further assume that $||f||_p = 1$ by dividing both sides of (8.9)

by $||f||_p$. Now let $g = f^{p/p'}$. It is easy to verify that $||g||_{p'} = 1$ and $\int_E fg = 1$, which completes the proof in this case.

If $||f||_p = +\infty$, define functions f_k on E by setting

$$f_k(\mathbf{x}) = 0 \text{ if } |\mathbf{x}| > k, \quad f_k(\mathbf{x}) = \min\{f(\mathbf{x}), k\} \text{ if } |\mathbf{x}| \le k.$$

Then each f_k belongs to L^p and $||f_k||_p \to ||f||_p = +\infty$. By the case already considered, we have $||f_k||_p = \int_E f_k g_k$ for some $g_k \ge 0$ with $||g_k||_{p'} = 1$. Since $f \ge f_k$, it follows that

$$\int_E f g_k \ge \int_E f_k g_k \to +\infty.$$

This shows that

$$\sup_{||g||_{p'}=1} \int_E fg = +\infty = ||f||_p.$$

To dispose of the restriction $f \ge 0$, apply the result above to $|f|$. Thus, there exists g_k with $||g_k||_{p'} = 1$ such that

$$||f||_p = \lim \int_E |f|\, g_k = \lim \int_E f\, \tilde{g}_k,$$

where $\tilde{g}_k = g_k\,(\operatorname{sign} f)$. (By sign x, we mean the function equal to +1 for $x > 0$ and to -1 for $x < 0$.) Since $\left\|\tilde{g}_k\right\|_{p'} = 1$, the result follows.

The cases $p = 1$ and ∞ are left as exercises.

We leave it to the reader to check that (8.9) is true if the supremum is taken only over those real-valued g with $||g||_{p'} = 1$ for which $\int_E fg$ exists. Also, if $1 \le p \le \infty$, then a measurable function f belongs to $L^p(E)$ if $fg \in L^1(E)$ for every $g \in L^{p'}(E)$, $1/p + 1/p' = 1$. See Exercise 2.

We have already observed that the sum of two L^p functions is again in L^p. The next theorem gives a more specific result when $1 \le p \le \infty$.

Theorem 8.10 (Minkowski's Inequality) *If $1 \le p \le \infty$, then $\|f + g\|_p \le \|f\|_p + \|g\|_p$; that is,*

$$\left(\int_E |f+g|^p\right)^{1/p} \le \left(\int_E |f|^p\right)^{1/p} + \left(\int_E |g|^p\right)^{1/p}, \quad 1 \le p < \infty,$$

$$\operatorname*{ess\,sup}_E |f+g| \le \operatorname*{ess\,sup}_E |f| + \operatorname*{ess\,sup}_E |g|.$$

Proof. If $p = 1$, the result is obvious. If $p = \infty$, we have $|f| \le ||f||_\infty$ a.e. in E and $|g| \le ||g||_\infty$ a.e. in E. Therefore, $|f+g| \le ||f||_\infty + ||g||_\infty$ a.e. in E, so that $||f+g||_\infty \le ||f||_\infty + ||g||_\infty$.

For $1 < p < \infty$,

$$\begin{aligned} ||f+g||_p^p &= \int_E |f+g|^p = \int_E |f+g|^{p-1}|f+g| \\ &\le \int_E |f+g|^{p-1}|f| + \int_E |f+g|^{p-1}|g|. \end{aligned}$$

In the last integral, apply Hölder's inequality to $|f+g|^{p-1}$ and $|g|$ with exponents $p' = p/(p-1)$ and p, respectively. This gives

$$\int_E |f+g|^{p-1}|g| \le ||f+g||_p^{p-1}||g||_p.$$

Since a similar result holds for $\int_E |f+g|^{p-1}|f|$, we obtain $||f+g||_p^p \le ||f+g||_p^{p-1}(||f||_p + ||g||_p)$, and the theorem follows by dividing both sides by $||f+g||_p^{p-1}$. (Note that if $||f+g||_p = 0$, there is nothing to prove; if $||f+g||_p = +\infty$, then either $\|f\|_p = +\infty$ or $\|g\|_p = +\infty$ by Theorem 8.3, and the result is obvious again.)

See Exercise 27(a) concerning a version of Minkowski's inequality for infinite series.

Remark: Minkowski's inequality fails for $0 < p < 1$. To see this, take $E = (0,1)$, $f = \chi_{(0,\frac{1}{2})}$, and $g = \chi_{(\frac{1}{2},1)}$. Then $||f+g||_p = 1$, while $||f||_p + ||g||_p = 2^{-1/p} + 2^{-1/p} = 2^{1-1/p} < 1$. See also (8.17) and Exercise 27(b).

8.3 Classes l^p

Let $a = \{a_k\}$ be a sequence of real or complex numbers, and let

$$||a||_p = \left(\sum_k |a_k|^p\right)^{1/p}, \ 0 < p < \infty; \quad ||a||_\infty = \sup_k |a_k|.$$

Then a is said to belong to l^p, $0 < p < \infty$, if $||a||_p < +\infty$, and to belong to l^∞ if $||a||_\infty < +\infty$. We will sometimes also denote $||a||_p$ by $||a||_{l^p}$.

Let us show that if $0 < p_1 < p_2 \leq \infty$, then $l^{p_1} \subset l^{p_2}$. (The opposite inclusion holds for $L^p(E)$, $|E| < +\infty$, by Theorem 8.2.) For $p_2 = \infty$, this is clear, and for $p_2 < \infty$, it follows from the fact that if $|a_k| \leq 1$, then $|a_k|^{p_2} \leq |a_k|^{p_1}$. Moreover (see Exercise 31),

$$||a||_p \leq ||a||_1 \text{ if } 1 \leq p \leq \infty, \quad \text{and } ||a||_p \leq ||a||_q \text{ if } 0 < q \leq p \leq \infty.$$

An example of a sequence that is in l^{p_2} for a given $p_2 < \infty$ but that is not in l^{p_1} for $p_1 < p_2$ is $\left\{\left(1/k\log^2 k\right)^{1/p_2} : k \geq 2\right\}$. Any constant sequence $\{a_k\}$, $a_k = c \neq 0$, belongs to l^∞ but not to l^p for $p < \infty$. The same is true for $\{1/\log k : k \geq 2\}$, whose terms even tend to zero.

Theorem 8.11 *If $a = \{a_k\}$ belongs to l^p for some $p < \infty$, then $\lim_{p\to\infty} \|a\|_p = \|a\|_\infty$.*

Proof. If $a \in l^{p_0}$, then $a \in l^p$ for $p_0 \leq p \leq \infty$. Since $|a_k| \to 0$, there is a largest $|a_k|$, say $|a_{k_0}|$. Thus, $||a||_\infty = |a_{k_0}|$. Write $\sum |a_k|^p = |a_{k_0}|^p \sum |a_k/a_{k_0}|^p$. Since $|a_k/a_{k_0}| \leq 1$, we see that $\sum |a_k/a_{k_0}|^p$ decreases (and so is bounded) as $p \nearrow \infty$. Hence, there is a constant $c > 0$ such that $|a_{k_0}|^p \leq ||a||_p^p \leq c|a_{k_0}|^p$ for all large p. Since $c^{1/p} \to 1$ as $p \to \infty$, the theorem follows.

The next two results are analogues for series of Hölder's and Minkowski's inequalities. Their proofs are left as exercises. If $a = \{a_k\}$ and $b = \{b_k\}$, we use the notation

$$ab = \{a_k b_k\}, \quad a + b = \{a_k + b_k\}, \quad \text{etc.}$$

Theorem 8.12 (Hölder's Inequality) *Suppose that $1 \leq p \leq \infty$, $1/p + 1/p' = 1$, $a = \{a_k\}$, $b = \{b_k\}$, and $ab = \{a_k b_k\}$. Then $||ab||_1 \leq ||a||_p ||b||_{p'}$; that is,*

$$\sum |a_k b_k| \leq \left(\sum |a_k|^p\right)^{1/p} \left(\sum |b_k|^{p'}\right)^{1/p'}, \quad 1 < p < \infty;$$
$$\sum |a_k b_k| \leq (\sup |a_k|) \sum |b_k|.$$

Theorem 8.13 (Minkowski's Inequality) *Suppose that $1 \leq p \leq \infty$, $a = \{a_k\}$, $b = \{b_k\}$, and $a + b = \{a_k + b_k\}$. Then $||a + b||_p \leq ||a||_p + ||b||_p$; that is,*

$$\left(\sum |a_k + b_k|^p\right)^{1/p} \leq \left(\sum |a_k|^p\right)^{1/p} + \left(\sum |b_k|^p\right)^{1/p}, \quad 1 \leq p < \infty;$$
$$\sup |a_k + b_k| \leq \sup |a_k| + \sup |b_k|.$$

Even though Minkowski's inequality fails when $p < 1$ (see Exercise 3), l^p is still a vector space for $0 < p < 1$; that is, $a + b \in l^p$ and $\alpha a = \{\alpha a_k\} \in l^p$ if $a, b \in l^p$ and α is any constant.

8.4 Banach and Metric Space Properties

We now define a notion that incorporates the main properties of L^p and l^p when $p \geq 1$. A set X is called a *Banach space over the complex numbers* if it satisfies the following three conditions:

(B_1) X is a *linear space* over the complex numbers $\mathbf{C}$; that is, if $x, y \in X$ and $\alpha \in \mathbf{C}$, then $x + y \in X$ and $\alpha x \in X$.

(B_2) X is a *normed space*; that is, for every $x \in X$ there is a nonnegative (finite) number $||x||$ such that
(a) $||x|| = 0$ if and only if x is the zero element of X,
(b) $||\alpha x|| = |\alpha|||x||$ for $\alpha \in \mathbf{C}$ and $x \in X$,
(c) $\|x + y\| \leq \|x\| + \|y\|$.
If these conditions are fulfilled, $||x||$ is called the *norm* of x.

(B_3) X is *complete* with respect to its norm; that is, every Cauchy sequence in X converges in X, or if $||x_k - x_m|| \to 0$ as $k, m \to \infty$, then there is an $x \in X$ such that $||x_k - x|| \to 0$.

A set X that satisfies (B_1) and (B_2), but not necessarily (B_3), is called a *normed linear space* over the complex numbers. A sequence $\{x_k\}$ such that $||x_k - x|| \to 0$ as $k \to \infty$ is said to *converge in norm* to x.

Restricting the scalars α in (B_1) and (B_2) to be real numbers, we obtain definitions for a Banach space over the real numbers and for a normed linear space over the real numbers. Unless specifically stated to the contrary, we will take the scalar field to be the complex numbers.

If X is a Banach space, define $d(x, y) = ||x - y||$ to be the *distance between x and y*. Then,

(M_1) $d(x, y) \geq 0$; $d(x, y) = 0$ if and only if $x = y$,

(M_2) $d(x, y) = d(y, x)$,

(M_3) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

Any set that has a distance function $d(x, y)$ satisfying (M_1), (M_2), and (M_3) is called a *metric space* with *metric d*. Therefore, a Banach space is a metric space whose metric is the norm. Moreover, by (B_3), a Banach space X is a complete metric space; that is, if $d(x_k, x_m) \to 0$ as $k, m \to \infty$, then there is an $x \in X$ such that $d(x_k, x) \to 0$.

Theorem 8.14 *For $1 \le p \le \infty$, $L^p(E)$ is a Banach space with norm $||f|| = ||f||_{p,E}$.*

Proof. Parts (B_1) and (B_2) in the definition of a Banach space are clearly fulfilled by $L^p(E)$, parts (a) and (c) of (B_2) being Theorem 5.11 and Minkowski's inequality, respectively. (Regarding part (a), we do not distinguish between two L^p functions that are equal a.e.; thus, the zero element of $L^p(E)$ means any function equal to zero a.e. in E.)

To verify (B_3), suppose that $\{f_k\}$ is a Cauchy sequence in $L^p(E)$. If $p = \infty$, then $|f_k - f_m| \le ||f_k - f_m||_\infty$ except for a set $Z_{k,m}$ of measure zero. If $Z = \bigcup_{k,m} Z_{k,m}$, then Z has measure zero, and $|f_k - f_m| \le ||f_k - f_m||_\infty$ outside Z for all k and m. Hence, $\{f_k\}$ converges uniformly outside Z to a bounded limit f, and it follows that $||f_k - f||_\infty \to 0$. (Note that convergence in L^∞ is equivalent to uniform convergence outside a set of measure zero.)

In case $1 \le p < \infty$, Tchebyshev's inequality (5.49) implies that

$$\left|\{\mathbf{x} \in E : |f_k(\mathbf{x}) - f_m(\mathbf{x})| > \varepsilon\}\right| \le \varepsilon^{-p} \int_E |f_k - f_m|^p .$$

Hence, $\{f_k\}$ is a Cauchy sequence in measure. By Theorems 4.22 and 4.23, there is a subsequence $\{f_{k_j}\}$ and a function f such that $f_{k_j} \to f$ a.e. in E. Given $\varepsilon > 0$, there is a K such that

$$\left(\int_E \left|f_{k_j} - f_k\right|^p\right)^{1/p} = ||f_{k_j} - f_k||_p < \varepsilon \text{ if } k_j, k > K.$$

Letting $k_j \to \infty$, we obtain by Fatou's lemma that $||f - f_k||_p \le \varepsilon$ if $k > K$. Hence, $||f - f_k||_p \to 0$ as $k \to \infty$. Finally, since $||f||_p \le ||f - f_k||_p + ||f_k||_p < +\infty$, it follows that $f \in L^p(E)$, which completes the proof.

A metric space X is said to be *separable* if it has a countable dense subset; that is, X is separable if there exists a countable set $\{x_k\}$ in X with the property that for every $x \in X$ and every $\varepsilon > 0$, there is an x_k with $d(x, x_k) < \varepsilon$. In the next theorem, we will show that L^p is separable if $1 \le p < \infty$. Note that L^∞ is not separable: take $L^\infty(0,1)$, for example, and consider the functions $f_t(x) = \chi_{(0,t)}(x), 0 < t < 1$. There are an uncountable number of these, and $||f_t - f_{t'}||_\infty = 1$ if $t \ne t'$ (see also Exercise 10).

Theorem 8.15 *If $1 \le p < \infty$, $L^p(E)$ is separable.*

Proof. Suppose first that $E = \mathbf{R^n}$, and consider a grid of dyadic cubes in $\mathbf{R^n}$. Let D be the set of all (finite) linear combinations of characteristic functions

of these cubes, the coefficients being complex numbers with rational real and imaginary parts. Then D is a countable subset of $L^p(\mathbf{R^n})$. To see that D is dense, use the method of successively approximating more and more general functions: First, consider characteristic functions of open sets (every open set is the countable union of nonoverlapping dyadic cubes by Theorem 1.11), of G_δ sets, and of measurable sets with finite measure; then consider simple functions whose supports have finite measure, nonnegative functions in $L^p(\mathbf{R^n})$, and, finally, arbitrary functions in $L^p(\mathbf{R^n})$. The details are left to the reader (cf. Lemma 7.3). This proves the case $E = \mathbf{R^n}$.

For an arbitrary measurable E, let D' denote the restrictions to E of the functions in D. Then D' is dense in $L^p(E)$, $1 \le p < \infty$. In fact, given p and $f \in L^p(E)$, let $f_1 = f$ on E and $f_1 = 0$ off E. Then $f_1 \in L^p(\mathbf{R^n})$, so that given $\varepsilon > 0$, there exists $g \in D$ with $\left(\int_{\mathbf{R^n}} |f_1 - g|^p\right)^{1/p} < \varepsilon$. Therefore, $\left(\int_E |f - g|^p\right)^{1/p} < \varepsilon$. This shows that D' is dense in $L^p(E)$ and completes the proof.

As we have already noted, Minkowski's inequality fails when $0 < p < 1$. Therefore, $||\cdot||_{p,E}$ is not a norm for such p. However, we still have the following facts.

Theorem 8.16 *If $0 < p < 1$, $L^p(E)$ is a complete, separable metric space, with distance defined by*

$$d(f,g) = ||f - g||_{p,E}^p.$$

Proof. With $d(f,g)$ so defined, properties (M_1) and (M_2) of a metric space are clear. To verify (M_3), which is the triangle inequality, we first claim that

$$(a+b)^p \le a^p + b^p \text{ if } a, b \ge 0, 0 < p \le 1.$$

If both a and b are zero, this is obvious. If, say, $a \ne 0$, then dividing by a^p, we reduce the inequality to $(1+t)^p \le 1 + t^p, t > 0$ $(t = b/a)$. This is clear since both sides are equal when $t = 0$ and the derivative of the right side majorizes that of the left for $t > 0$.

It follows that $|f(\mathbf{x}) - g(\mathbf{x})|^p \le |f(\mathbf{x}) - h(\mathbf{x})|^p + |h(\mathbf{x}) - g(\mathbf{x})|^p$ if $0 < p \le 1$. Integrating, we obtain $||f - g||_p^p \le ||f - h||_p^p + ||h - g||_p^p$, which is just the triangle inequality. The proofs that L^p is complete and separable with respect to $||\cdot||_p$ are the same as in Theorems 8.14 and 8.15.

It is worth noting that in case $0 < p \le 1$, the triangle inequality is equivalent to the basic estimate

$$||f+g||_p^p \le ||f||_p^p + ||g||_p^p \qquad (0 < p \le 1). \tag{8.17}$$

See also Exercise 27(b).

The analogous results for series are listed in the next theorem.

Theorem 8.18

(i) *If $1 \le p \le \infty$, l^p is a Banach space with $\|a\| = \|a\|_p$. For $1 \le p < \infty$, l^p is separable; l^∞ is not separable.*

(ii) *If $0 < p < 1$, l^p is a complete, separable metric space, with distance $d(a,b) = \|a-b\|_p^p$.*

Proof. We will show that l^p is complete and separable when $1 \le p < \infty$ and that l^∞ is not separable. The rest of the proof of (i) and the proof of (ii) are left to the reader.

Suppose that $1 \le p < \infty$, $a^{(i)} = \left\{a_k^{(i)}\right\} \in l^p$ for $i = 1, 2, \ldots$, and $\left\|a^{(i)} - a^{(j)}\right\|_p \to 0$ as $i, j \to \infty$. Since $\left\|a^{(i)} - a^{(j)}\right\|_p \ge \left|a_k^{(i)} - a_k^{(j)}\right|$ for every k, it follows that $\left|a_k^{(i)} - a_k^{(j)}\right| \to 0$ for every k as $i, j \to \infty$. Let $a_k = \lim_{i\to\infty} a_k^{(i)}$ and $a = \{a_k\}$. We will show that $a \in l^p$ and $\left\|a^{(i)} - a\right\|_p \to 0$. Given $\varepsilon > 0$, there exists N such that

$$\left(\sum_k \left|a_k^{(i)} - a_k^{(j)}\right|^p\right)^{1/p} = \|a^{(i)} - a^{(j)}\|_p < \varepsilon \quad \text{if } i, j > N.$$

Restricting the summation to $k \le M$ and letting $j \to \infty$, we obtain

$$\left(\sum_{k=1}^{M} \left|a_k^{(i)} - a_k\right|^p\right)^{1/p} \le \varepsilon \quad \text{for any } M, \text{ if } i > N.$$

Letting $M \to \infty$, we get $\left\|a^{(i)} - a\right\|_p \le \varepsilon$ if $i > N$; that is, $\left\|a^{(i)} - a\right\|_p \to 0$. The fact that $\|a\|_p \le \left\|a - a^{(i)}\right\|_p + \left\|a^{(i)}\right\|_p$ shows that $a \in l^p$. Hence, l^p is complete.

To prove that l^p is separable when $p < \infty$, let D be the set of all sequences $\{d_k\}$ such that (a) the real and imaginary parts of d_k are rational, and (b) $d_k = 0$ for $k \ge N$ (N may vary from sequence to sequence). Then D is a countable subset of l^p. If $a = \{a_k\} \in l^p$ and $\varepsilon > 0$, choose N so that $\sum_{k=N+1}^{\infty} |a_k|^p < \varepsilon/2$. Choose $d_1, \ldots, d_N$ with rational real and imaginary parts such that $\sum_{k=1}^{N} |a_k - d_k|^p < \varepsilon/2$. Then $d = \{d_1, \ldots, d_N, 0, \ldots\}$ belongs to D and $\|a - d\|_p^p < \varepsilon$. It follows that D is dense in l^p and therefore that l^p is separable.

To see that l^∞ is not separable, consider the subclass of sequences $a = \{a_k\}$ for which each a_k is 0 or 1. The number of such sequences is uncountable,

and $\|a - a'\|_\infty = 1$ for any two different such sequences. Hence, l^∞ cannot be separable.

We know from Lusin's theorem that measurable functions have continuity properties. The next theorem gives a useful continuity property of functions in L^p.

Theorem 8.19 (Continuity in L^p) *If $f \in L^p(\mathbf{R^n})$, $1 \le p < \infty$, then*

$$\lim_{|\mathbf{h}|\to 0} \|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})\|_p = 0.$$

Proof. Let C_p denote the class of $f \in L^p$ such that $\|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})\|_p \to 0$ as $|\mathbf{h}| \to 0$. We claim that (a) a finite linear combination of functions in C_p is in C_p, and (b) if $f_k \in C_p$ and $\left\|f_k - f\right\|_p \to 0$, then $f \in C_p$. Both of these facts follow easily from Minkowski's inequality; for (b), for example, note that

$$\begin{aligned}
&\|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})\|_p \\
&\quad \le \left\|f(\mathbf{x}+\mathbf{h}) - f_k(\mathbf{x}+\mathbf{h})\right\|_p + \left\|f_k(\mathbf{x}+\mathbf{h}) - f_k(\mathbf{x})\right\|_p + \left\|f_k(\mathbf{x}) - f(\mathbf{x})\right\|_p \\
&\quad = \left\|f_k(\mathbf{x}+\mathbf{h}) - f_k(\mathbf{x})\right\|_p + 2\left\|f_k(\mathbf{x}) - f(\mathbf{x})\right\|_p .
\end{aligned}$$

Since $f_k \in C_p$, we have $\limsup_{|\mathbf{h}|\to 0} \|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})\|_p \le 2\|f_k(\mathbf{x}) - f(\mathbf{x})\|_p$, and (b) follows by letting $k \to \infty$.

Clearly, the characteristic function of a cube belongs to C_p. Hence, in view of the fact that linear combinations of characteristic functions of cubes are dense in $L^p(\mathbf{R^n})$ (see the proof of Theorem 8.15), it follows from (a) and (b) that $L^p(\mathbf{R^n})$ is contained in C_p, and the proof is complete.

We remark without proof that Theorem 8.19 is also true for $0 < p < 1$. (Use the same ideas for $\|\cdot\|_p^p$.) It fails, however, for $p = \infty$, as shown by the function $\chi = \chi_{(0,\infty)}(x)$ on $(-\infty, +\infty)$. In fact, $\chi \in L^\infty(-\infty, +\infty)$ but $\|\chi(x+h) - \chi(x)\|_\infty = 1$ for all $h \ne 0$.

8.5 The Space L^2 and Orthogonality

For complex-valued measurable f, $f = f_1 + if_2$ with f_1 and f_2 real-valued and measurable, we have $\int_E f = \int_E f_1 + i\int_E f_2$ (see p. 96 in Section 5.3). We will use the fact that $\left|\int_E f\right| \le \int_E |f|$ (see Exercise 1).

Among the L^p spaces, L^2 has the special property that the product of any two of its elements is integrable (Schwarz's inequality). This simple fact leads to some important extra structure in L^2, which we will now discuss.

Consider $L^2 = L^2(E)$, where E is a fixed subset of $\mathbf{R^n}$ of positive measure, and write $\|f\| = \|f\|_{2,E}$, $\int_E f = \int f$, etc. For $f, g \in L^2$, define the *inner product* of f and g by

$$\langle f, g\rangle = \int f\,\overline{g}, \tag{8.20}$$

where $\overline{g}$ denotes the complex conjugate of g. Note that by Schwarz's inequality,

$$|\langle f, g\rangle| \le \|f\|\,\|g\|.$$

Moreover, the inner product has the following properties:

(a) $\langle g, f\rangle = \overline{\langle f, g\rangle}$,
(b) $\langle f_1 + f_2, g\rangle = \langle f_1, g\rangle + \langle f_2, g\rangle$, $\langle f, g_1 + g_2\rangle = \langle f, g_1\rangle + \langle f, g_2\rangle$,
(c) $\langle \alpha f, g\rangle = \alpha\langle f, g\rangle$, $\langle f, \alpha g\rangle = \overline{\alpha}\langle f, g\rangle$, $\alpha \in \mathbf{C}$,
(d) $\langle f, f\rangle^{1/2} = \|f\|$.

If $\langle f, g\rangle = 0$, then f and g are said to be *orthogonal*. A set $\{\phi_\alpha\}_{\alpha\in A}$ is *orthogonal* if any two of its elements are orthogonal; $\{\phi_\alpha\}$ is *orthonormal* if it is orthogonal and $\|\phi_\alpha\| = 1$ for all α. Note that if $\{\phi_\alpha\}$ is orthogonal and $\|\phi_\alpha\| \ne 0$ for every α, then $\{\phi_\alpha / \|\phi_\alpha\|\}$ is orthonormal. Henceforth, we will assume that $\|\phi_\alpha\| \ne 0$ for all α for an orthogonal system $\{\phi_\alpha\}$. This implies that no element is zero. Furthermore, since for $\alpha \ne \beta$,

$$\|\phi_\alpha - \phi_\beta\|^2 = \int (\phi_\alpha - \phi_\beta)\left(\overline{\phi_\alpha} - \overline{\phi_\beta}\right) = \|\phi_\alpha\|^2 + \|\phi_\beta\|^2 \ne 0,$$

it implies that no two elements are equal.

A simple example of an infinite orthogonal system in $L^2(E)$ is $\{\chi_{E_j}\}$ where $\{E_j\}$ is an infinite collection of disjoint measurable subsets of E with $0 < |E_j| < \infty$ (cf. Exercise 33 of Chapter 3. See also Exercise 24 of this chapter).

Theorem 8.21 *Any orthogonal system $\{\phi_\alpha\}$ in L^2 is countable.*

Proof. We may assume that $\{\phi_\alpha\}$ is orthonormal. Then for $\alpha \ne \beta$, as above,

$$\|\phi_\alpha - \phi_\beta\|^2 = \int (\phi_\alpha - \phi_\beta)\left(\overline{\phi_\alpha} - \overline{\phi_\beta}\right) = \|\phi_\alpha\|^2 + \|\phi_\beta\|^2 = 2,$$

so that $\|\phi_\alpha - \phi_\beta\| = \sqrt{2}$. Since L^2 is separable, it follows that $\{\phi_\alpha\}$ must be countable.

A collection $\psi_1, \ldots, \psi_N$ is said to be *linearly independent* if $\sum_{k=1}^N a_k \psi_k(\mathbf{x}) = 0$ (a.e.) implies that every a_k is zero. Any collection of functions is called *linearly independent* if each finite subcollection is linearly independent. No function in a linearly independent set can be zero a.e.

Theorem 8.22 *If $\{\psi_k\}$ is orthogonal, it is linearly independent.*

Proof. Suppose that $a_1\psi_{k1} + \cdots + a_N\psi_{k_N} = 0$. Multiplying both sides by $\overline{\psi_{k1}}$ and integrating, we obtain by orthogonality that $a_1 = 0$. Similarly, $a_2 = \cdots = a_N = 0$.

The converse of Theorem 8.22 is not true. However, the next result shows that if $\{\psi_k\}$ is linearly independent, then the system formed from suitable linear combinations of its elements is orthogonal.

Theorem 8.23 (Gram–Schmidt Process) *If $\{\psi_k\}$ is linearly independent, then the system $\{\phi_k\}$ defined by*

$$\begin{aligned}
\phi_1 &= \psi_1 \\
\phi_2 &= a_{21}\psi_1 + \psi_2 \\
\vdots \quad & \quad \vdots \\
\phi_k &= a_{k1}\psi_1 + \cdots + a_{k,k-1}\psi_{k-1} + \psi_k \\
\vdots \quad & \quad \vdots
\end{aligned}$$

is orthogonal for proper selection of the a_{ij}.

Proof. Having $\phi_1 = \psi_1$, we proceed by induction, assuming that $\phi_1, \ldots, \phi_{k-1}$ have been chosen as required. We will determine constants $b_{k1}, \ldots, b_{k,k-1}$ so that the function ϕ_k defined by

$$\phi_k = b_{k1}\phi_1 + \cdots + b_{k,k-1}\phi_{k-1} + \psi_k$$

is orthogonal to $\phi_1, \ldots, \phi_{k-1}$. If $j < k$,

$$\langle \phi_k, \phi_j \rangle = b_{kj} \langle \phi_j, \phi_j \rangle + \langle \psi_k, \phi_j \rangle$$

by orthogonality. Since $\langle \phi_j, \phi_j \rangle \neq 0$, b_{kj} can be chosen so that $\langle \phi_k, \phi_j \rangle = 0$, $j < k$. Since each ϕ_j with $j < k$ is a linear combination of $\psi_1, \ldots, \psi_j$, the theorem follows.

When the ϕ_k are selected by the Gram–Schmidt process, we shall say that they are *generated* from the ψ_k. Note that the triangular character of the matrix in Theorem 8.23 means that each ψ_k can also be written as a linear combination of the ϕ_j, $j \leq k$.

We call an orthogonal system $\{\phi_k\}$ *complete* if the only function that is orthogonal to every ϕ_k is zero; that is, $\{\phi_k\}$ is complete if $\langle f, \phi_k \rangle = 0$ for all k implies that $f = 0$ a.e. Thus, a complete orthogonal system is one that is maximal in the sense that it is not properly contained in any larger orthogonal system.

The *span* of a set of functions $\{\psi_k\}$ is the collection of all finite linear combinations of the ψ_k. In speaking of the span of $\{\psi_k\}$, we may always assume that $\{\psi_k\}$ is orthogonal by discarding any dependent functions and applying the Gram–Schmidt process to the resulting linearly independent set.

A set $\{\psi_k\}$ is called a *basis* for L^2 if its span is dense in L^2; that is, $\{\psi_k\}$ is a basis if given $f \in L^2$ and $\varepsilon > 0$, there exist N and $\{a_k\}_{k=1}^N$ such that $\left\| f - \sum_{k=1}^N a_k \psi_k \right\| < \varepsilon$. The a_k can always be chosen with rational real and imaginary parts. Any countable dense set in L^2 is of course a basis. It follows that L^2 has an infinite orthogonal basis.

Theorem 8.24 *Any orthogonal basis in L^2 is complete. In particular, there exists a complete orthonormal basis for L^2.*

Proof. Let $\{\psi_k\}$ be an orthogonal basis for L^2. We may assume that $\{\psi_k\}$ is orthonormal. To show that it is complete, let $\langle f, \psi_k \rangle = 0$ for all k. Then $\langle f, f \rangle = \left\langle f, f - \sum_{k=1}^N a_k \psi_k \right\rangle$ for any finite sum $\sum_{k=1}^N a_k \psi_k$. By Schwarz's inequality, $|\langle f, f \rangle| \leq \|f\| \left\| f - \sum_{k=1}^N a_k \psi_k \right\|$, and so, since the term on the right can be chosen arbitrarily small, $\langle f, f \rangle = 0$. Therefore, $f = 0$ a.e., which completes the proof.

Let us show that every complete orthogonal system in $L^2(E)$ is infinite if (as always) $|E| > 0$. If not, there is a set E with $|E| > 0$ and a complete orthogonal system $\{\phi_k\}_{k=1}^N$ in L^2 with N finite. Assuming as we may that the system is orthonormal, its completeness implies that for every $f \in L^2(E)$, we have $f = \sum_{k=1}^N \langle f, \phi_k \rangle \phi_k$ a.e. in E, and consequently,

$$\langle f, g \rangle = \sum_{k=1}^N \langle f, \phi_k \rangle \overline{\langle g, \phi_k \rangle} \quad \text{if } f, g \in L^2(E).$$

(In the next section, similar facts are derived in the limit case $N = \infty$.) Therefore, if f and g are any two orthogonal functions in $L^2(E)$, then the sequences $\{\langle f, \phi_k\rangle\}_{k=1}^N$ and $\{\langle g, \phi_k\rangle\}_{k=1}^N$, considered as vectors, are orthogonal in the vector space $\mathbf{C}^N$ of N-tuples of complex numbers. We observed on p. 198 in Section 8.5 that there is an infinite orthogonal system $\{\psi_j\}$ in $L^2(E)$. Hence, there is an infinite collection $\left\{\{\langle \psi_j, \phi_k\rangle\}_{k=1}^N\right\}_j$ of orthogonal vectors in the finite dimensional space $\mathbf{C}^N$, which is impossible.

From now on, we will consider only orthogonal systems in L^2 that are (countably) infinite.

8.6 Fourier Series and Parseval's Formula

Let $\{\phi_k\}$ be any (infinite) orthonormal system in L^2. If $f \in L^2$, the numbers defined by

$$c_k = c_k(f) = \langle f, \phi_k\rangle = \int_E f\,\overline{\phi_k}$$

are called the *Fourier coefficients* of f with respect to $\{\phi_k\}$. The series $\sum_k c_k\phi_k$ is called the *Fourier series* of f with respect to $\{\phi_k\}$, and denoted $S[f] = \sum_k c_k\phi_k$. We also write

$$f \sim \sum_k c_k\phi_k.$$

The first question we ask is how well $S[f]$, or more precisely, the sequence of its partial sums, approximates f. Fix N and let $L = \sum_{k=1}^N \gamma_k\phi_k$ be a linear combination of $\phi_1, \ldots, \phi_N$. We wish to know what choice of $\gamma_1, \ldots, \gamma_N$ makes $\|f - L\|$ a minimum. Note that since $\{\phi_k\}$ is orthonormal, $\|L\|^2 = \langle L, L\rangle = \sum_{k=1}^N |\gamma_k|^2$. Hence,

$$\begin{aligned}\|f - L\|^2 &= \int \left(f - \sum_{k=1}^N \gamma_k\phi_k\right)\left(\overline{f} - \sum_{k=1}^N \overline{\gamma_k}\,\overline{\phi_k}\right)\\ &= \|f\|^2 - \sum_{k=1}^N (\overline{\gamma_k}\,c_k + \gamma_k\,\overline{c_k}) + \sum_{k=1}^N |\gamma_k|^2,\end{aligned}$$

where the c_k are the Fourier coefficients of f. Since

$$|c_k - \gamma_k|^2 = (c_k - \gamma_k)\,(\overline{c_k} - \overline{\gamma_k}) = |c_k|^2 - (\overline{\gamma_k}\,c_k + \gamma_k\,\overline{c_k}) + |\gamma_k|^2,$$

we obtain

$$\|f - L\|^2 = \|f\|^2 + \sum_{k=1}^{N} |c_k - \gamma_k|^2 - \sum_{k=1}^{N} |c_k|^2 .$$

Therefore,

$$\min_{\gamma_1,\ldots,\gamma_N} \|f - L\|^2 = \|f\|^2 - \sum_{k=1}^{N} |c_k|^2 ; \tag{8.25}$$

that is, the minimum is achieved when $\gamma_k = c_k$ for $k = 1, \ldots, N$, or equivalently, when L is the Nth partial sum of $S[f]$. Writing $s_N = s_N(f) = \sum_{k=1}^{N} c_k \phi_k$, we have from (8.25) that

$$\left\|f - s_N\right\|^2 = \|f\|^2 - \sum_{k=1}^{N} |c_k|^2 . \tag{8.26}$$

Theorem 8.27 *Let $\{\phi_k\}$ be an orthonormal system in L^2 and let $f \in L^2$.*

(i) *Of all linear combinations $\sum_1^N \gamma_k \phi_k$ with N fixed, the one that best approximates f in L^2 is given by the partial sum $s_N = \sum_1^N c_k \phi_k$ of the Fourier series of f.*

(ii) (*Bessel's inequality*) *The sequence $\{c_k\}$ of Fourier coefficients of f belongs to l^2 and*

$$\left(\sum_{k=1}^{\infty} |c_k|^2\right)^{1/2} \leq \|f\|.$$

Proof. Part (i) has been proved. Note that since $\left\|f - s_N\right\|^2 \geq 0$, Bessel's inequality follows from (8.26) by letting $N \to \infty$, which completes the proof.

If f is a function for which equality holds in Bessel's inequality, that is, if

$$\left(\sum_{k=1}^{\infty} |c_k|^2\right)^{1/2} = \|f\|, \tag{8.28}$$

then f is said to satisfy *Parseval's formula*. From (8.26), we immediately obtain the next result.

Theorem 8.29 *Parseval's formula holds for f if and only if $\|s_N - f\| \to 0$, that is, if and only if $S[f]$ converges to f in L^2 norm.*

The following theorem is of great importance.

Theorem 8.30 (Riesz–Fischer Theorem) *Let $\{\phi_k\}$ be any orthonormal system and let $\{c_k\}$ be any sequence in l^2. Then there is an $f \in L^2$ such that $S[f] = \sum c_k \phi_k$, that is, such that $\{c_k\}$ is the sequence of Fourier coefficients of f with respect to $\{\phi_k\}$. Moreover, f can be chosen to satisfy Parseval's formula.*

Proof. Let $t_N = \sum_{k=1}^{N} c_k \phi_k$. Then if $M < N$,

$$\|t_N - t_M\|^2 = \left\| \sum_{M+1}^{N} c_k \phi_k \right\|^2 = \sum_{M+1}^{N} |c_k|^2 .$$

The fact that $\{c_k\} \in l^2$ implies that $\{t_N\}$ is a Cauchy sequence in L^2. Since L^2 is complete, there is an $f \in L^2$ such that $\|f - t_N\| \to 0$. If $N \geq k$,

$$\int f \overline{\phi_k} = \int (f - t_N) \overline{\phi_k} + \int t_N \overline{\phi_k} = \int (f - t_N) \overline{\phi_k} + c_k.$$

Since the integral on the right is bounded in absolute value by $\|f - t_N\| \|\phi_k\| = \|f - t_N\|$, we obtain by letting $N \to \infty$ that $\int f \overline{\phi_k} = c_k$. Thus, $S[f] = \sum_k c_k \phi_k$, so that $t_N = s_N(f)$, and it follows from Theorem 8.29 that Parseval's formula holds for f. This completes the proof.

There is no guarantee that the Fourier coefficients of a function uniquely determine the function. However, if $\{\phi_k\}$ is complete, we can show that the correspondence between a function and its Fourier coefficients is unique; that is, if f and g have the same Fourier coefficients with respect to a complete system, then $f = g$ a.e. This is simple, since the vanishing of all the Fourier coefficients of $f - g$ implies that $f - g = 0$ a.e. An important related fact is the following.

Theorem 8.31 *Let $\{\phi_k\}$ be an orthonormal system. Then $\{\phi_k\}$ is complete if and only if Parseval's formula holds for every $f \in L^2$.*

Proof. Suppose that $\{\phi_k\}$ is complete. If $f \in L^2$, Bessel's inequality implies that its Fourier coefficients $\{c_k\}$ belong to l^2. Hence, by the Riesz–Fischer theorem,

there exists a g in L^2 with $S[g] = \sum c_k \phi_k$ and $\|g\| = \left(\sum |c_k|^2\right)^{1/2}$. Since f and g have the same Fourier coefficients and $\{\phi_k\}$ is complete, we see that $f = g$ a.e. Hence, $\|f\| = \|g\| = \left(\sum |c_k|^2\right)^{1/2}$, which is Parseval's formula.

Conversely, suppose that Parseval's formula holds with respect to $\{\phi_k\}$ for every $f \in L^2$. If $\langle f, \phi_k \rangle = 0$ for all k, then $\|f\| = \left(\sum \left|\langle f, \phi_k \rangle\right|^2\right)^{1/2} = 0$. Therefore, $f = 0$ a.e., so that $\{\phi_k\}$ is complete, which proves the result.

Suppose that $\{\phi_k\}$ is orthonormal and complete and that $f, g \in L^2$. Let $c_k = \langle f, \phi_k \rangle$, $d_k = \langle g, \phi_k \rangle$, $c = \{c_k\}$, $d = \{d_k\}$, and $(c, d) = \sum c_k \overline{d_k}$. We claim that

$$\langle f, g \rangle = (c, d). \tag{8.32}$$

To prove this, observe that by Parseval's formula, $\langle f + g, f + g \rangle = (c + d, c + d)$, or

$$\langle f, f \rangle + \langle g, g \rangle + 2\,\mathrm{Re}\,\langle f, g \rangle = (c, c) + (d, d) + 2\,\mathrm{Re}\,(c, d),$$

where $\mathrm{Re}\, z$ denotes the real part of z. Cancelling equal terms gives $\mathrm{Re}\,\langle f, g \rangle = \mathrm{Re}\,(c, d)$. Applying this to the function $if(\mathbf{x})$, we obtain $\mathrm{Re}\,\langle if, g \rangle = \mathrm{Re}\,(ic, d)$. But $\mathrm{Re}\,\langle if, g \rangle = \mathrm{Re}\,[i\langle f, g \rangle] = -\mathrm{Im}\,\langle f, g \rangle$. Similarly, $\mathrm{Re}\,(ic, d) = -\mathrm{Im}\,(c, d)$. Therefore, $\mathrm{Im}\,\langle f, g \rangle = \mathrm{Im}\,(c, d)$, and (8.32) is proved.

Another corollary of Theorem 8.31 is given in the next result. First, we make several definitions. Let X_1 and X_2 be metric spaces with metrics d_1 and d_2, respectively. Then X_1 and X_2 are said to be *isometric* if there is a mapping T of X_1 onto X_2 such that

$$d_1(f, g) = d_2(Tf, Tg)$$

for all $f, g \in X_1$. Such a T is called an *isometry*. Thus, an isometry is a mapping that preserves distances. An isometry is automatically one-to-one, and two isometric metric spaces may be regarded as the same space with a relabeling of the points. For example, two L^2 spaces, $L^2(E)$ and $L^2(E')$, are isometric if there is a mapping T of $L^2(E)$ onto $L^2(E')$ such that $\|f - g\|_{2,E} = \|Tf - Tg\|_{2,E'}$ for all $f, g \in L^2(E)$. The isometries we shall encounter will be *linear*, that is, will satisfy

$$T(\alpha f + \beta g) = \alpha Tf + \beta Tg \quad \text{for all scalars } \alpha, \beta.$$

If T is a linear map of $L^2(E)$ onto $L^2(E')$, then since $Tf - Tg = T(f - g)$, it follows that T is an isometry if and only if

$$\|f\|_{2,E} = \|Tf\|_{2,E'}$$

for all $f \in L^2(E)$. Similarly, a linear map T of $L^2(E)$ onto l^2 is an isometry if and only if $\|f\|_{2,E} = \|Tf\|_{l^2}$ for all $f \in L^2(E)$.

Theorem 8.33 *All spaces $L^2(E)$ are linearly isometric with l^2, and so with one another.*

Proof. For a given E, define a linear correspondence between $L^2(E)$ and l^2 by choosing a complete orthonormal system $\{\phi_k\}$ in $L^2(E)$ and mapping an $f \in L^2(E)$ onto the sequence $\{\langle f, \phi_k\rangle\}$ of its Fourier coefficients. This mapping is onto all of l^2 by the Riesz–Fischer theorem and is an isometry by Theorem 8.31.

8.7 Hilbert Spaces

A set H is called a *Hilbert space over the complex numbers* $\mathbf{C}$ if it satisfies the following three conditions:

(H_1) H is a vector space over $\mathbf{C}$; that is, if $f, g \in H$ and $\alpha \in \mathbf{C}$, then $f + g \in H$ and $\alpha f \in H$. The zero element of H will be denoted by 0.

(H_2) For every pair $f, g \in H$, there is a complex number (f, g), called the *inner product* of f and g, which satisfies

(a) $(g, f) = \overline{(f, g)}$,

(b) $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$,

(c) $(\alpha f, g) = \alpha(f, g)$ for $\alpha \in \mathbf{C}$,

(d) $(f, f) \geq 0$, and $(f, f) = 0$ if and only if $f = 0$.

Notice that (a), (b), and (c) imply that $(f, g_1 + g_2) = (f_1 g_1) + (f, g_2)$, $(f, \alpha g) = \overline{\alpha}(f, g)$, and $(0, f) = 0$. Define

$$\|f\| = (f, f)^{1/2}.$$

Before stating the third condition, we claim that

$$|(f, g)| \leq \|f\|\,\|g\| \quad (\textit{Schwarz's inequality}). \tag{8.34}$$

If $\|g\| = 0$, this is obvious. Otherwise, letting $\lambda = -(f, g)/\|g\|^2$, we obtain

$$0 \leq (f + \lambda g, f + \lambda g) = \|f\|^2 - 2\frac{|(f, g)|^2}{\|g\|^2} + \frac{|(f, g)|^2}{\|g\|^2} = \|f\|^2 - \frac{|(f, g)|^2}{\|g\|^2},$$

and Schwarz's inequality follows at once. This simple proof also shows that if equality holds in (8.34) and $||g|| \neq 0$, then f is a constant multiple of g, namely,

$$f = -\lambda g = \frac{(f,g)}{(g,g)}\, g.$$

We will show that $\|\cdot\|$ is a norm on H by proving the triangle inequality. In fact,

$$\|f+g\|^2 = (f+g, f+g) = \|f\|^2 + 2\,\mathrm{Re}\,(f,g) + \|g\|^2.$$

Since $|\mathrm{Re}\,(f,g)| \leq |(f,g)| \leq \|f\|\,\|g\|$, it follows that the right side is at most $(\|f\| + \|g\|)^2$. Taking square roots, we obtain $\|f+g\| \leq \|f\| + \|g\|$, as desired. Hence, H is a normed linear space.

We also require

(H_3) H is complete with respect to $\|\cdot\|$.

In particular, a Hilbert space is a Banach space.

As for L^2 spaces, a linear map T of a Hilbert space H onto a Hilbert space H' is an isometry if and only if $\|f\|_H = \|Tf\|_{H'}$ for all $f \in H$.

A Hilbert space is called *infinite dimensional* if it cannot be spanned by a finite number of elements; hence, an infinite dimensional Hilbert space has an infinite linearly independent subset. The space L^2 with inner product $(f,g) = \int f\,\overline{g}$ and the space l^2 with $(c,d) = \sum_k c_k\,\overline{d_k}$ are examples of separable infinite dimensional Hilbert spaces. In fact, there are essentially no other examples, as the following theorem shows.

Theorem 8.35 *All separable infinite dimensional Hilbert spaces are linearly isometric with l^2 and so with one another.*

Proof. The proof is a repetition of the ideas leading to Theorem 8.33, so we shall be brief. Let H be a separable infinite dimensional Hilbert space, and let $\{e'_k\}$ be a countable dense subset. Discarding those e'_k that are spanned by other e'_i, we obtain a linearly independent set $\{e_k\}$ with the same dense span as $\{e'_k\}$. Since H is infinite dimensional, $\{e_k\}$ is infinite. Using the Gram–Schmidt process, we may assume that $\{e_k\}$ is orthonormal: $(e_i, e_k) = 0$ for $i \neq k$ and $\|e_k\| = 1$ for all k. It follows that $\{e_k\}$ is complete; in fact, if $(f, e_k) = 0$ for all k, then

$$\left\| f - \sum_{k=1}^{N} a_k e_k \right\|^2 = \|f\|^2 + \sum_{k=1}^{N} |a_k|^2 \geq \|f\|^2$$

for $N = 1, 2, \ldots$. If f were not zero, the span of the e_k could not be dense. Hence, $f = 0$, which shows that H has a complete orthonormal system $\{e_k\}$.

Next, we will show that Bessel's inequality and an analogue of the Riesz–Fischer theorem hold for $\{e_k\}$. Let $f \in H$ and $c_k = (f, e_k)$. Then

$$0 \leq \left\| f - \sum_{k=1}^{N} c_k e_k \right\|^2 = \|f\|^2 - \sum_{k=1}^{N} |c_k|^2 .$$

Letting $N \to \infty$, we obtain Bessel's inequality $\left(\sum |c_k|^2\right)^{1/2} \leq \|f\|$. In particular, $\{c_k\}$ belongs to l^2.

To derive the Riesz–Fischer theorem, let $\{\gamma_k\}$ be a sequence in l^2 and set $t_N = \sum_{k=1}^{N} \gamma_k e_k$. Then

$$\|t_M - t_N\|^2 = \sum_{k=M+1}^{N} |\gamma_k|^2 \to 0 \quad \text{as } M, N \to \infty,\ M < N.$$

Since H is complete, there is a $g \in H$ such that $\|g - t_N\| \to 0$. We have

$$(g, e_k) = (g - t_N, e_k) + (t_N, e_k) = (g - t_N, e_k) + \gamma_k \quad (k \leq N).$$

Letting $N \to \infty$, it follows from Schwarz's inequality that $(g, e_k) = \gamma_k$. Hence, $t_N = \sum_{k=1}^{N} (g, e_k)\, e_k$ and $\|g - t_N\|^2 = \|g\|^2 - \sum_{k=1}^{N} |\gamma_k|^2$. Letting $N \to \infty$ in the last equation, we see that g satisfies Parseval's formula $\|g\| = \left(\sum |\gamma_k|^2\right)^{1/2}$. This gives the analogue of the Riesz–Fischer theorem.

Now, let $f \in H$ and set $c_k = (f, e_k)$. Choose $\{\gamma_k\} = \{c_k\}$ in the version of the Riesz–Fischer theorem just derived, and let $g \in H$ satisfy $(g, e_k) = c_k$ and $||g|| = \left(\sum |c_k|^2\right)^{1/2}$. We see by the completeness of $\{e_k\}$ that $g = f$, so that Parseval's formula holds: $\|f\| = \left(\sum |c_k|^2\right)^{1/2}$. The fact that H is linearly isometric with l^2 now follows as in the proof of Theorem 8.33.

Exercises

1. For complex-valued measurable f, $f = f_1 + if_2$ with f_1 and f_2 real-valued and measurable, we have $\int_E f = \int_E f_1 + i \int_E f_2$ by definition. Prove that $\int_E f$ is finite if and only if $\int_E |f|$ is finite, and $\left|\int_E f\right| \leq \int_E |f|$. (Note that

$\left|\int_E f\right| = \left[\left(\int_E f_1\right)^2 + \left(\int_E f_2\right)^2\right]^{1/2}$, and use the fact that $\left(a^2+b^2\right)^{1/2} = a\cos\alpha + b\sin\alpha$ for an appropriate α, while $\left(a^2+b^2\right)^{1/2} \geq |a\cos\alpha + b\sin\alpha|$ for all α.)

2. Prove the converse of Hölder's inequality for $p = 1$ and ∞. Show also that for $1 \leq p \leq \infty$, a real-valued measurable f belongs to $L^p(E)$ if $fg \in L^1(E)$ for every $g \in L^{p'}(E)$, $1/p + 1/p' = 1$. The negation is also of interest: if $f \notin L^p(E)$, then there exists $g \in L^{p'}(E)$ such that $fg \notin L^1(E)$. (To verify the negation, construct g of the form $\sum a_k g_k$ for appropriate a_k and g_k, with g_k satisfying $\int_E fg_k \to +\infty$.)

3. Prove Theorems 8.12 and 8.13. Show that Minkowski's inequality for series fails when $p < 1$.

4. Let f and g be real-valued and not identically 0 (i.e., neither function equals 0 a.e.), and let $1 < p < \infty$. Prove that equality holds in the inequality $\left|\int fg\right| \leq \|f\|_p \|g\|_{p'}$ if and only if fg has constant sign a.e. and $|f|^p$ is a multiple of $|g|^{p'}$ a.e.
If $\|f+g\|_p = \|f\|_p + \|g\|_p$ and $g \neq 0$ in Minkowski's inequality, show that f is a multiple of g.
Find analogues of these results for the spaces l^p.

5. For $0 < p \leq \infty$ and $0 < |E| < +\infty$, define

$$N_p[f] = \left(\frac{1}{|E|}\int_E |f|^p\right)^{1/p},$$

where $N_\infty[f]$ means $\|f\|_\infty$. Prove that if $p_1 < p_2$, then $N_{p_1}[f] \leq N_{p_2}[f]$. Prove also that if $1 \leq p \leq \infty$, then $N_p[f+g] \leq N_p[f] + N_p[g]$, $(1/|E|)\int_E |fg| \leq N_p[f]N_{p'}[g]$, $1/p + 1/p' = 1$, and that $\lim_{p\to\infty} N_p[f] = \|f\|_\infty$. Thus, N_p behaves like $\|\cdot\|_p$ but has the advantage of being monotone in p. Recall Exercise 28 of Chapter 5.

6. (a) Let $1 \leq p_i, r \leq \infty$ and $\sum_{i=1}^k \frac{1}{p_i} = \frac{1}{r}$. Prove the following generalization of Hölder's inequality:

$$\|f_1 \cdots f_k\|_r \leq \|f_1\|_{p_1} \cdots \|f_k\|_{p_k}.$$

(See also Exercise 12 of Chapter 7.)

(b) Let $1 \leq p < r < q \leq \infty$ and define $\theta \in (0,1)$ by $\frac{1}{r} = \frac{\theta}{p} + \frac{1-\theta}{q}$. Prove the interpolation estimate

$$\|f\|_r \leq \|f\|_p^\theta \|f\|_q^{1-\theta}.$$

In particular, if $A = \max\{\|f\|_p, \|f\|_q\}$, then $\|f\|_r \leq A$.

7. Show that when $0 < p < 1$, the neighborhoods $\{f : \|f\|_p < \varepsilon\}$ of zero in $L^p(0,1)$ are not convex. (Let $f = \chi_{(0,\varepsilon^p)}$, and $g = \chi_{(\varepsilon^p, 2\varepsilon^p)}$. Show that $\|f\|_p = \|g\|_p = \varepsilon$, but that $\|\frac{1}{2}f + \frac{1}{2}g\|_p > \varepsilon$.)

8. Prove the following *integral version of Minkowski's inequality* for $1 \le p < \infty$ and a measurable function $f(\mathbf{x}, \mathbf{y})$:

$$\left[\int \left[\int |f(\mathbf{x}, \mathbf{y})|\, d\mathbf{x}\right]^p d\mathbf{y}\right]^{1/p} \le \int \left[\int |f(\mathbf{x}, \mathbf{y})|^p\, d\mathbf{y}\right]^{1/p} d\mathbf{x}.$$

(For $1 < p < \infty$, note that the pth power of the left-hand side equals $\iint \left[\int \left|f(\mathbf{z}, \mathbf{y})\right| d\mathbf{z}\right]^{p-1} |f(\mathbf{x}, \mathbf{y})|\, d\mathbf{x}\, d\mathbf{y}$. Integrate first with respect to $\mathbf{y}$ and apply Hölder's inequality.)

9. If f is real-valued and measurable on E, $|E| > 0$, define its *essential infimum* on E by

$$\operatorname*{ess\,inf}_E f = \sup\{\alpha : |\{\mathbf{x} \in E : f(\mathbf{x}) < \alpha\}| = 0\}.$$

If $f \ge 0$, show that $\operatorname{ess}_E \inf f = (\operatorname{ess}_E \sup 1/f)^{-1}$.

10. Prove that $L^\infty(E)$ is not separable for any E with $|E| > 0$. (Construct a sequence of decreasing subsets of E whose measures strictly decrease. Consider the characteristic functions of the class of sets obtained by taking all possible unions of the differences of these subsets.)

11. If $f_k \to f$ in L^p, $1 \le p < \infty$, $g_k \to g$ pointwise, and $\|g_k\|_\infty \le M$ for all k, prove that $f_k g_k \to fg$ in L^p.

12. Let $f, \{f_k\} \in L^p, 0 < p \le \infty$. Show that if $\|f - f_k\|_p \to 0$, then $\|f_k\|_p \to \|f\|_p$. Conversely, if $f_k \to f$ a.e. and $\|f_k\|_p \to \|f\|_p$, $0 < p < \infty$, show that $\|f - f_k\|_p \to 0$. Show that the converse may fail for $p = \infty$. (For the converse when $0 < p < \infty$, note that $|f - f_k|^p \le c(|f|^p + |f_k|^p)$ with $c = \max\{2^{p-1}, 1\}$; then apply, for example, the sequential version of Lebesgue's dominated convergence theorem given in Exercise 23 of Chapter 5.)

13. Suppose that $f_k \to f$ a.e. and that $f_k, f \in L^p$, $1 < p \le \infty$. If $\|f_k\|_p \le M < +\infty$, show that $\int f_k g \to \int fg$ for all $g \in L^{p'}, 1/p + 1/p' = 1$. Show that the result is false if $p = 1$. (When $p > 1$, use Egorov's theorem in case the domain of integration has finite measure.)

14. Verify that the following systems are orthogonal:

(a) $\left\{\frac{1}{2}, \cos x, \sin x, \ldots, \cos kx, \sin kx, \ldots\right\}$ on any interval of length 2π.

(b) $\left\{e^{2\pi i k x/(b-a)};\ k = 0, \pm 1, \pm 2, \ldots\right\}$ on (a, b).

15. If $f \in L^2(0, 2\pi)$, show that

$$\lim_{k\to\infty} \int_0^{2\pi} f(x) \cos kx \, dx = \lim_{k\to\infty} \int_0^{2\pi} f(x) \sin kx \, dx = 0.$$

Prove that the same is true if $f \in L^1(0, 2\pi)$. (This last statement is the *Riemann–Lebesgue lemma*. To prove it, approximate f in L^1 norm by L^2 functions. See Theorem 12.21.)

16. A sequence $\{f_k\}$ in L^p is said to *converge weakly in* L^p to a function f (belonging to L^p) if $\int f_k g \to \int fg$ for all $g \in L^{p'}$. Prove that if $f_k \to f$ in L^p norm, $1 \le p \le \infty$, then $\{f_k\}$ converges weakly in L^p to f. Note by Exercise 15 that the converse is not true. See Exercise 28 of Chapter 10.

17. Suppose that $f_k, f \in L^2$ and that $\int f_k g \to \int fg$ for all $g \in L^2$ (i.e., $\{f_k\}$ converges weakly in L^2 to f). If $\|f_k\|_2 \to \|f\|_2$, show that $f_k \to f$ in L^2 norm. The same is true for L^p, $1 < p < \infty$, by a 1913 result of Radon.

18. Prove the *parallelogram law* for L^2:

$$\|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2.$$

Is this true for L^p when $p \neq 2$? The geometric interpretation is that the sum of the squares of the lengths of the diagonals of a parallelogram equals the sum of the squares of the edge lengths.

19. Prove that a finite dimensional Hilbert space is isometric with $\mathbf{R^n}$ for some n.

20. Construct a function in $L^1(-\infty, +\infty)$ that is not in $L^2(a, b)$ for any $a < b$. (Let $g(x) = x^{-1/2}$ on (0,1) and $g(x) = 0$ elsewhere, so that $\int_{-\infty}^{+\infty} g = 2$. Consider the function $f(x) = \sum a_k g(x - r_k)$, where $\{r_k\}$ is the rational numbers and $\{a_k\}$ satisfies $a_k > 0$, $\sum a_k < +\infty$.)

21. If $f \in L^p(\mathbf{R^n})$, $0 < p < \infty$, show that

$$\lim_{Q \searrow x} \frac{1}{|Q|} \int_Q |f(\mathbf{y}) - f(\mathbf{x})|^p \, d\mathbf{y} = 0 \quad \text{a.e.}$$

Note by Exercise 5 that if this condition holds for a given p, then it also holds for all smaller p.

22. Let $\{\phi_k\}$ be a complete orthonormal system in L^2 and let $m = \{m_k\}$ be a fixed bounded sequence of numbers. If $f \in L^2$, $f \sim \sum c_k \phi_k$, define Tf by $Tf \sim \sum m_k c_k \phi_k$. Such an operator is called a *Fourier multiplier operator*. Show that T is bounded on L^2, that is, that there is a constant c

independent of f such that $||Tf||_2 \leq c||f||_2$ for all $f \in L^2$. Show also that the smallest possible choice for c is $||m||_{l^\infty}$.

23. Show that every subset Λ of a separable metric space (M,d) is separable. (Let $D = \{f_k\}$ be a countable dense set in M, and for $j = 1,2,\ldots$, define $D_j = \{f \in D : \inf_{\lambda\in\Lambda} d(\lambda,f) < 1/j\}$. If $f_k \in D_j$, pick $\lambda_{k,j} \in \Lambda$ with $d(\lambda_{k,j},f_k) < 1/j$ and show that $\{\lambda_{k,j}\}$ is dense in Λ.)

24. Let E be a measurable set in $\mathbf{R^n}$ with $0 < |E| < \infty$. Construct an orthogonal system $\{\phi_j\}_{j=0}^{\infty}$ in $L^2(E)$ with $\phi_0 = 1$ everywhere in E. (Use Exercise 32 of Chapter 3 with $\theta = 1/2$, and choose ϕ_j for $j \geq 1$ to be appropriate simple functions with values ± 1.)

25. If f is a measurable function on $\mathbf{R^n}$, define $\langle f\rangle = \sup_{\alpha>0} \alpha|\{|f| > \alpha\}|$, and recall that f belongs to weak $L^1(\mathbf{R^n})$ if and only if $\langle f\rangle < \infty$. Show that weak L^1 $(\mathbf{R^n})$ has all the properties of a Banach space with respect to $\langle\cdot\rangle$ except the triangle inequality. Show however that there is a constant $\kappa > 1$ such that the quasi-triangle inequality $\langle f+g\rangle \leq \kappa(\langle f\rangle+\langle g\rangle)$ holds for all measurable f,g. (To show that κ cannot be 1, consider the case of one dimension and the functions $f = \chi_{[0,1/2]}+2\chi_{(1/2,1]}, g = 2\chi_{[0,1/2]}+\chi_{(1/2,1]}$.)

26. Show that $\liminf_{p\to\infty} ||f||_{L^p(E)} \geq ||f||_{L^\infty(E)}$ even if $|E| = \infty$.

27. (a) Prove Minkowski's inequality for infinite series:

$$\left\| \left(\sum_{k=1}^{\infty} |f_k| \right) \right\|_p \leq \sum_{k=1}^{\infty} ||f_k||_p, \quad 1 \leq p \leq \infty.$$

(b) Show that in part (a), the opposite inequality holds if $0 < p \leq 1$:

$$\sum_{k=1}^{\infty} ||f_k||_p \leq \left\| \left(\sum_{k=1}^{\infty} |f_k| \right) \right\|_p, \quad 0 < p \leq 1.$$

(For (b), assuming that $\sum |f_k|$ is positive and finite a.e., multiply and divide each $|f_k|^p$ in the summation on the left side of the inequality by $\left(\sum |f_k|\right)^{p(1-p)}$ and apply Hölder's inequality with exponents $q = 1/p$ and $q' = 1/(1-p)$.)

28. Let $\phi(t)$ be a continuous function on $[0,\infty)$ that is positive, increasing, and convex on $(0,\infty)$ and that satisfies $\phi(0) = 0, \lim_{t\to\infty} \phi(t) = \infty$, and $|\phi(2t)| \leq c|\phi(t)|$ for some constant c independent of t. For example, the function $\phi(t) = t\left(1+\log^+ t\right)$ has these properties (see Exercise 25 of Chapter 7). If E is a measurable set in $\mathbf{R^n}$, define the *Orlicz space*

$L_\phi(E)$ to be the collection of all measurable f on E that are finite a.e. in E and satisfy $\phi(|f|) \in L(E)$. Show that $L_\phi(E)$ is a Banach space with norm

$$||f||_{L_\phi(E)} = \inf\left\{\lambda > 0 : \int_E \phi\left(\frac{|f|}{\lambda}\right) d\mathbf{x} \le 1\right\}.$$

In case $\phi(t) = t\left(1 + \log^+ t\right)$, the class is often denoted $L\log L(E)$ (see Exercise 22 of Chapter 9 for a result about functions in $L\log L\,(\mathbf{R^n})$).

29. Let $1 \le p < \infty$ and $f \in L^p\,(\mathbf{R^n})$. Show that $||f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})||_p$ (where the norm is taken with respect to $\mathbf{x}$) is a uniformly continuous function of $\mathbf{h}$. Is the same true when $0 < p < 1$?

30. Let $1 \le p < \infty$ and E be a measurable set in $\mathbf{R^n}$.

(a) Prove that if f_1, f_2, g_1, g_2 are nonnegative and measurable on E, then

$$\left(\int_E \left[(f_1+f_2)^p + (g_1+g_2)^p\right] d\mathbf{x}\right)^{1/p}$$
$$\le \left[\int_E \left(f_1^p + g_1^p\right) d\mathbf{x}\right]^{1/p} + \left[\int_E \left(f_2^p + g_2^p\right) d\mathbf{x}\right]^{1/p}.$$

(b) If the right side in part (a) is replaced by

$$\left[\int_E \left(f_1^p + f_2^p\right) dx\right]^{1/p} + \left[\int_E \left(g_1^p + g_2^p\right) dx\right]^{1/p},$$

is the resulting inequality true?

(c) If $\{f_{i,j}\}_{i,j=1}^\infty$ are measurable functions on E, show that

$$\left[\int_E \sum_j \left(\sum_i |f_{i,j}(\mathbf{x})|\right)^p d\mathbf{x}\right]^{1/p} \le \sum_i \left[\int_E \left(\sum_j |f_{i,j}(\mathbf{x})|^p\right) d\mathbf{x}\right]^{1/p}.$$

$\Big($For $i = 1, \ldots, N$ (N finite), consider the sequences $F_i = \left\{|f_{i,j}|\right\}_j$ and note that $||\sum_{i=1}^N F_i||_{l^p} = \left(\sum_j \left(\sum_{i=1}^N |f_{i,j}|\right)^p\right)^{1/p}$.$\Big)$

31. Let $a = \{a_k\}$ be a sequence of real or complex numbers. Show that $||a||_p \le ||a||_1$ if $1 \le p \le \infty$ and more generally that $||a||_p \le ||a||_q$ if $0 < q \le p \le \infty$. (If $1 \le p < \infty$, the inequality $|a_1|^p + |a_2|^p \le (|a_1| + |a_2|)^p$ may be used together with an induction argument.)

32. For nonnegative measurable functions f and g on $(0, \infty)$, let

$$F(x) = \int_0^\infty f\left(\frac{x}{y}\right) g(y) \frac{dy}{y}, \quad x \in (0, \infty).$$

Also, if $1 \le p < \infty$, set

$$[f]_p = \left(\int_0^\infty f(x)^p \frac{dx}{x} \right)^{1/p},$$

and if $p = \infty$, define $[f]_\infty = \operatorname*{ess\,sup}_{(0,\infty)} f$. Prove that for $1 \le p \le \infty$, $[F]_p \le [f]_1 \, [g]_p$.

9

Approximations of the Identity and Maximal Functions

9.1 Convolutions

The convolution of two functions f and g that are measurable in $\mathbf{R^n}$ is defined by

$$(f * g)(\mathbf{x}) = \int_{\mathbf{R^n}} f(\mathbf{t})g(\mathbf{x} - \mathbf{t})\, d\mathbf{t}, \quad \mathbf{x} \in \mathbf{R^n},$$

provided the integral exists.

In Theorem 6.14, we saw that if $f, g \in L^1(\mathbf{R^n})$, then $f * g$ exists a.e. and is measurable in $\mathbf{R^n}$, and $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$. Moreover, according to Corollary 6.16, $\|f * g\|_1 = \|f\|_1 \|g\|_1$ if f and g are nonnegative and measurable. In this section, we will study some additional properties of convolutions, beginning with the following theorem.

Theorem 9.1 *Let $1 \leq p \leq \infty$, $f \in L^p(\mathbf{R^n})$ and $g \in L^1(\mathbf{R^n})$. Then $f * g \in L^p(\mathbf{R^n})$ and*

$$\|f * g\|_p \leq \|f\|_p \|g\|_1.$$

Proof. We may suppose that $1 < p \leq \infty$, since when $p = 1$, the result is just Theorem 6.14. Let us first prove the result in case f and g are nonnegative. Then $f * g$ is nonnegative and measurable on $\mathbf{R^n}$ by Corollary 6.16. If $p = \infty$,

$$(f * g)(\mathbf{x}) \leq \int_{\mathbf{R^n}} \|f\|_\infty g(\mathbf{x} - \mathbf{t})\, d\mathbf{t} = \|f\|_\infty \int_{\mathbf{R^n}} g(\mathbf{x} - \mathbf{t})\, d\mathbf{t} = \|f\|_\infty \|g\|_1.$$

Therefore, $\|f * g\|_\infty \leq \|f\|_\infty \|g\|_1$, as claimed. If $1 < p < \infty$, we write

$$(f * g)(\mathbf{x}) = \int_{\mathbf{R^n}} \left[f(\mathbf{t})g(\mathbf{x} - \mathbf{t})^{1/p} \right] g(\mathbf{x} - \mathbf{t})^{1/p'} d\mathbf{t}, \quad \frac{1}{p} + \frac{1}{p'} = 1.$$

By Hölder's inequality with exponents p and p',

$$\begin{aligned}(f * g)(\mathbf{x}) &\leq \left(\int_{\mathbf{R^n}} f(\mathbf{t})^p g(\mathbf{x}-\mathbf{t})\, d\mathbf{t}\right)^{1/p} \left(\int_{\mathbf{R^n}} g(\mathbf{x}-\mathbf{t})\, d\mathbf{t}\right)^{1/p'} \\ &= (f^p * g)(\mathbf{x})^{1/p} \|g\|_1^{1/p'}.\end{aligned}$$

Now raise the first and last terms in this inequality to the pth power and integrate the result. Since $\int_{\mathbf{R^n}} (f^p * g)\, d\mathbf{x} = \|f\|_p^p \|g\|_1$ by Corollary 6.16, we obtain

$$\|f * g\|_p^p \leq \|f\|_p^p \|g\|_1^{1+(p/p')} = \|f\|_p^p \|g\|_1^p.$$

The theorem follows for $f, g \geq 0$ by taking pth roots.

For general $f \in L^p(\mathbf{R^n})$ and $g \in L^1(\mathbf{R^n})$, let us first show that $f * g$ exists a.e. and is measurable. By the case already considered, we have $|f| * |g| \in L^p(\mathbf{R^n})$. Hence, $|f| * |g| < \infty$ a.e., so that $f(\mathbf{x}-\mathbf{t})g(\mathbf{t}) \in L^1(d\mathbf{t})$ for a.e. $\mathbf{x}$. Consequently, $f * g$ exists and is finite a.e. To show it is measurable, define

$$f_N = f\chi_{\{|\mathbf{x}|<N\}}, \quad N = 1, 2, \ldots,$$

and note that $f_N \in L^1(\mathbf{R^n})$ for each N since $f \in L^p(\mathbf{R^n})$. By Theorem 6.14, each $f_N * g$ is measurable since it is the convolution of functions in $L^1(\mathbf{R^n})$. Also,

$$\lim_{N\to\infty} f_N(\mathbf{x}-\mathbf{t})g(\mathbf{t}) = f(\mathbf{x}-\mathbf{t})g(\mathbf{t}) \text{ and } \left|f_N(\mathbf{x}-\mathbf{t})g(\mathbf{t})\right| \leq \left|f(\mathbf{x}-\mathbf{t})g(\mathbf{t})\right|.$$

Therefore, $\lim_{N\to\infty}(f_N * g) = f * g$ a.e. by the Lebesgue dominated convergence theorem. It follows that $f * g$ is measurable. The rest of the proof is now an immediate corollary of the inequality $|f * g| \leq |f| * |g|$ and the result for nonnegative functions.

See Exercise 21 for a useful sufficient condition that a convolution $f * g$ be measurable on $\mathbf{R^n}$, namely, that f, g are locally integrable on $\mathbf{R^n}$ and $|f| * |g|$ is finite a.e. in $\mathbf{R^n}$.

Theorem 9.1 is an important special case of the next result, whose proof is left to the reader. (See Exercise 2.)

Theorem 9.2 (Young's Convolution Theorem) *Let p and q satisfy $1 \leq p, q \leq \infty$ and $1/p + 1/q \geq 1$, and let r be defined by $1/r = 1/p + 1/q - 1$. If $f \in L^p(\mathbf{R^n})$ and $g \in L^q(\mathbf{R^n})$, then $f * g \in L^r(\mathbf{R^n})$ and*

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

Note that when $q = 1$, Young's theorem reduces to Theorem 9.1. See also Exercise 3.

A convolution $f * K$ with K fixed defines a transformation $T : f \to f * K$, which is called the *convolution operator with kernel K*. Theorem 9.1 states that a convolution operator with an integrable kernel maps functions in L^p into the same L^p. The next result shows an effect that convolution operators with smooth kernels have on L^p.

For a positive integer m, we denote by C^m the class of functions $f(\mathbf{x}), \mathbf{x} \in \mathbf{R^n}$, whose partial derivatives up to and including those of order m exist and are continuous. The subset of C^m of functions with compact support is denoted C_0^m. Similarly, C^∞ is the class of infinitely differentiable functions, and C_0^∞ is the corresponding subset of functions with compact support. (For the existence of such functions, see Exercise 4.) Finally, if $\alpha = (\alpha_1, \ldots, \alpha_n)$, where the α_k are nonnegative integers, then the αth partial derivative of f is denoted by

$$(D^\alpha f)(\mathbf{x}) = \left(\frac{\partial^\alpha f}{\partial \mathbf{x}^\alpha}\right)(\mathbf{x}) = \left(\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f\right)(\mathbf{x}).$$

Theorem 9.3 *If $1 \le p \le \infty$, $f \in L^p(\mathbf{R^n})$, and $K \in C_0^m$, then $f * K \in C^m$ with bounded partial derivatives of all orders at most m, and*

$$D^\alpha(f * K)(\mathbf{x}) = (f * D^\alpha K)(\mathbf{x})$$

if $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\alpha_1 + \cdots + \alpha_n \le m$.

Proof. We first claim that if K is any continuous kernel with compact support, then $f * K$ is bounded and continuous. In fact, if S denotes the support of K, then

$$\begin{aligned}
\left|\int_{\mathbf{R^n}} f(\mathbf{x}-\mathbf{t})K(\mathbf{t})\,d\mathbf{t}\right| &\le \left(\int_S |f(\mathbf{x}-\mathbf{t})|\,d\mathbf{t}\right)(\max|K|) \\
&\le \left(\int_S |f(\mathbf{x}-\mathbf{t})|^p d\mathbf{t}\right)^{1/p} |S|^{1/p'}(\max|K|) \\
&\le ||f||_p\,|S|^{1/p'}(\max|K|) < \infty,
\end{aligned}$$

$1/p + 1/p' = 1$, which shows that $f * K$ is bounded. Also,

$$\begin{aligned}
&|(f * K)(\mathbf{x}+\mathbf{h}) - (f * K)(\mathbf{x})| \\
&\quad = \left| \int_{\mathbf{R}^n} f(\mathbf{t}) K(\mathbf{x}+\mathbf{h}-\mathbf{t})\, d\mathbf{t} - \int_{\mathbf{R}^n} f(\mathbf{t}) K(\mathbf{x}-\mathbf{t})\, d\mathbf{t} \right| \\
&\quad = \left| \int_{\mathbf{R}^n} f(\mathbf{x}-\mathbf{t})[K(\mathbf{t}+\mathbf{h}) - K(\mathbf{t})]\, d\mathbf{t} \right| \\
&\quad \le \left(\int_{\mathbf{R}^n} |f(\mathbf{x}-\mathbf{t})|^p\, d\mathbf{t} \right)^{1/p} \left(\int_{\mathbf{R}^n} |K(\mathbf{t}+\mathbf{h}) - K(\mathbf{t})|^{p'} d\mathbf{t} \right)^{1/p'} \\
&\quad = \|f\|_p \|K(\mathbf{t}+\mathbf{h}) - K(\mathbf{t})\|_{p'}.
\end{aligned}$$

The last expression tends to zero as $|\mathbf{h}| \to 0$ since K is uniformly continuous and has compact support. (Note that in case $p' < \infty$, this also follows from Theorem 8.19 since $K \in L^{p'}$. See also Exercise 3.)

Next, let $K \in C_0^m, m \ge 1$. Fix $i = 1, \ldots, n$ and let $\mathbf{h} = (0, \ldots, 0, h, 0, \ldots, 0)$, where h is in the ith coordinate position. Note that

$$\begin{aligned}
\frac{(f * K)(\mathbf{x}+\mathbf{h}) - (f * K)(\mathbf{x})}{h} &= \int_{\mathbf{R}^n} f(\mathbf{t}) \left\{ \frac{K(\mathbf{x}-\mathbf{t}+\mathbf{h}) - K(\mathbf{x}-\mathbf{t})}{h} \right\} d\mathbf{t} \\
&= \int_{\mathbf{R}^n} f(\mathbf{t}) \frac{\partial K}{\partial x_i}(\mathbf{x}-\mathbf{t}+\mathbf{h}')\, d\mathbf{t},
\end{aligned}$$

by the mean-value theorem, where $\mathbf{h}' = (0, \ldots, 0, h', 0, \ldots, 0)$ for some h' depending on $\mathbf{x}$ and $\mathbf{t}$ which is between 0 and h. Hence, as $h \to 0$, $(\partial K/\partial x_i)(\mathbf{x}-\mathbf{t}+\mathbf{h}')$ converges to $(\partial K/\partial x_i)(\mathbf{x}-\mathbf{t})$ uniformly in $\mathbf{t}$. Since $\partial K/\partial x_i$ has compact support, it follows that the last integral converges to $[f * (\partial K/\partial x_i)](\mathbf{x})$. Therefore, $[\partial(f * K)/\partial x_i](\mathbf{x})$ exists and equals $[f * (\partial K/\partial x_i)](\mathbf{x})$, which is bounded and continuous by our earlier remarks. The proof of the theorem for $m = 1$ is now complete. The proof for $m = 2, 3, \ldots$ follows by repeated application of the case $m = 1$.

It follows from Theorem 9.3 that $f * K \in C^\infty$ if $f \in L^p$, $1 \le p \le \infty$, and $K \in C_0^\infty$. If, in addition, f has compact support, then so has $f * K$. In fact, if S_1 is the support of K and S_2 is the support of f, then the formula $(f * K)(\mathbf{x}) = \int_{S_2} f(\mathbf{t}) K(\mathbf{x}-\mathbf{t})\, d\mathbf{t}$ implies that $(f * K)(\mathbf{x}) = 0$ unless there are points $\mathbf{t} \in S_2$ for which $\mathbf{x} - \mathbf{t} \in S_1$. Hence, the support of $f * K$ is contained in $\{\mathbf{x} : \mathbf{x} = \mathbf{s}_1 + \mathbf{s}_2, \mathbf{s}_1 \in S_1, \mathbf{s}_2 \in S_2\}$ and so is bounded. An application of this fact is given in Exercise 5.

Theorem 9.4 *If $f \in L(\mathbf{R^n})$ and K is bounded and uniformly continuous on $\mathbf{R^n}$, then $f * K$ is bounded and uniformly continuous on $\mathbf{R^n}$.*

The proof is similar to the first part of the proof of Theorem 9.3 and is left as an exercise. See also Exercise 3.

9.2 Approximations of the Identity

Given $K(\mathbf{x})$ and $\varepsilon > 0$, let

$$K_\varepsilon(\mathbf{x}) = \varepsilon^{-n} K\left(\frac{\mathbf{x}}{\varepsilon}\right) = \varepsilon^{-n} K\left(\frac{x_1}{\varepsilon}, \ldots, \frac{x_n}{\varepsilon}\right).$$

For example, if $K(\mathbf{x}) = \chi_{\{|\mathbf{x}|<1\}}(\mathbf{x})$, then $K_\varepsilon(\mathbf{x}) = \varepsilon^{-n}\chi_{\{|\mathbf{x}|<\varepsilon\}}(\mathbf{x})$. In this case, taking successively smaller values of ε produces kernels with successively higher peaks and smaller supports. The effect on any positive K with compact support is roughly the same.

In general, K_ε has the following basic properties.

Lemma 9.5 *If $K \in L^1(\mathbf{R^n})$ and $\varepsilon > 0$, then*

(i) $\int_{\mathbf{R^n}} K_\varepsilon = \int_{\mathbf{R^n}} K$,
(ii) $\int_{|\mathbf{x}|>\delta} |K_\varepsilon| \to 0$ as $\varepsilon \to 0$, for any fixed $\delta > 0$.

Proof. Part (i) follows immediately from the change of variables $\mathbf{y} = \mathbf{x}/\varepsilon$ (see Exercise 20 of Chapter 5). For part (ii), fix $\delta > 0$, and let $\mathbf{y} = \mathbf{x}/\varepsilon$. Then

$$\int_{|\mathbf{x}|>\delta} |K_\varepsilon(\mathbf{x})|\, d\mathbf{x} = \varepsilon^{-n} \int_{|\mathbf{x}|>\delta} \left|K\left(\frac{\mathbf{x}}{\varepsilon}\right)\right| d\mathbf{x} = \int_{|\mathbf{y}|>\delta/\varepsilon} |K(\mathbf{y})|\, d\mathbf{y}.$$

Since $K \in L$ and $\delta/\varepsilon \to +\infty$ as $\varepsilon \to 0$, it follows that the last integral tends to zero as $\varepsilon \to 0$. This completes the proof.

Note that for $K \geq 0$, property (i) means that the areas under the graphs of K and K_ε are the same, while (ii) means that for small ε, the bulk of the area under the graph of K_ε is concentrated in the region above a small neighborhood of the origin.

For any $K \in L$, we can expect from (ii) that the effect of letting $\varepsilon \to 0$ in the formula $(f * K_\varepsilon)(\mathbf{x}) = \int f(\mathbf{t}) K_\varepsilon(\mathbf{x} - \mathbf{t}) d\mathbf{t}$ will be to emphasize the values of $f(\mathbf{t})$

when $\mathbf{t}$ is near $\mathbf{x}$. As a simple illustration, let $Q_\varepsilon(\mathbf{x})$ denote the cube in $\mathbf{R^n}$ of edge length ε centered at $\mathbf{x}$ and consider the kernel $k(\mathbf{t}) = \chi_{Q_1(\mathbf{0})}(\mathbf{t})$. Then

$$k_\varepsilon(\mathbf{x}-\mathbf{t}) = \varepsilon^{-n}k((\mathbf{x}-\mathbf{t})/\varepsilon) = \varepsilon^{-n}\chi_{Q_\varepsilon(\mathbf{x})}(\mathbf{t}),$$

and

$$\left(f * k_\varepsilon\right)(\mathbf{x}) = \frac{1}{\varepsilon^n}\int\limits_{Q_\varepsilon(\mathbf{x})} f(\mathbf{t})\,d\mathbf{t} = \frac{1}{|Q_\varepsilon(\mathbf{x})|}\int\limits_{Q_\varepsilon(\mathbf{x})} f(\mathbf{t})\,d\mathbf{t}.$$

In this case, the Lebesgue Differentiation Theorem 7.2 implies that $(f * k_\varepsilon)(\mathbf{x}) \to f(\mathbf{x})$ a.e as $\varepsilon \to 0$ if f is locally integrable.

The next four theorems show that $(f * K_\varepsilon)(\mathbf{x}) \to f(\mathbf{x})$ in various senses (e.g., in norm or pointwise) as $\varepsilon \to 0$ if K is suitably restricted. A family $\{K_\varepsilon : \varepsilon > 0\}$ of kernels for which $f * K_\varepsilon \to f$ in some sense is called an *approximation of the identity*. See also Exercise 23(a).

In what follows, we shall use the notation $f_\varepsilon(\mathbf{x})$ for the convolution $(f * K_\varepsilon)(\mathbf{x})$.

Theorem 9.6 *Let $f_\varepsilon = f * K_\varepsilon$, where $K \in L^1(\mathbf{R^n})$ and $\int_{\mathbf{R^n}} K = 1$. If $f \in L^p(\mathbf{R^n})$, $1 \le p < \infty$, then*

$$\|f_\varepsilon - f\|_p \to 0 \text{ as } \varepsilon \to 0.$$

Proof. By Lemma 9.5(i),

$$f(\mathbf{x}) = f(\mathbf{x})\int\limits_{\mathbf{R^n}} K_\varepsilon(\mathbf{t})\,d\mathbf{t} = \int\limits_{\mathbf{R^n}} f(\mathbf{x})K_\varepsilon(\mathbf{t})\,d\mathbf{t}.$$

Therefore,

$$\begin{aligned}|f_\varepsilon(\mathbf{x}) - f(\mathbf{x})| &= \left|\int\limits_{\mathbf{R^n}} [f(\mathbf{x}-\mathbf{t}) - f(\mathbf{x})]\,K_\varepsilon(\mathbf{t})\,d\mathbf{t}\right| \\ &\le \int\limits_{\mathbf{R^n}} |f(\mathbf{x}-\mathbf{t}) - f(\mathbf{x})||K_\varepsilon(\mathbf{t})|^{1/p}|K_\varepsilon(\mathbf{t})|^{1/p'}\,d\mathbf{t},\end{aligned}$$

where $1/p + 1/p' = 1$ ($1/p' = 0$ if $p = 1$). Applying Hölder's inequality with exponents p and p' and then raising both sides to the pth power and integrating with respect to $\mathbf{x}$, we obtain

$$\begin{aligned}
&\int_{\mathbf{R}^n} |f_\varepsilon(\mathbf{x}) - f(\mathbf{x})|^p \, d\mathbf{x} \\
&\quad \le \int_{\mathbf{R}^n} \left[\int_{\mathbf{R}^n} |f(\mathbf{x}-\mathbf{t}) - f(\mathbf{x})|^p |K_\varepsilon(\mathbf{t})| \, d\mathbf{t} \right] \left[\int_{\mathbf{R}^n} |K_\varepsilon(\mathbf{t})| \, d\mathbf{t} \right]^{p/p'} d\mathbf{x} \\
&\quad = \|K\|_1^{p/p'} \int_{\mathbf{R}^n} \left[\int_{\mathbf{R}^n} |f(\mathbf{x}-\mathbf{t}) - f(\mathbf{x})|^p |K_\varepsilon(\mathbf{t})| \, d\mathbf{t} \right] d\mathbf{x}.
\end{aligned}$$

Changing the order of integration in the last expression (which is justified since the integrand is nonnegative), we obtain

$$\|f_\varepsilon - f\|_p^p \le \|K\|_1^{p/p'} \int_{\mathbf{R}^n} |K_\varepsilon(\mathbf{t})| \phi(\mathbf{t}) \, d\mathbf{t},$$

where $\phi(\mathbf{t}) = \int_{\mathbf{R}^n} |f(\mathbf{x}-\mathbf{t}) - f(\mathbf{x})|^p \, d\mathbf{x} = \|f(\mathbf{x}-\mathbf{t}) - f(\mathbf{x})\|_p^p$. For $\delta > 0$, write

$$I_\varepsilon = \int_{\mathbf{R}^n} |K_\varepsilon(\mathbf{t})| \phi(\mathbf{t}) \, d\mathbf{t} = \int_{|\mathbf{t}|<\delta} + \int_{|\mathbf{t}|\ge\delta} = A_{\varepsilon,\delta} + B_{\varepsilon,\delta}.$$

Given $\eta > 0$, we can choose δ so small that $\phi(\mathbf{t}) < \eta$ if $|\mathbf{t}| < \delta$ (note that $\phi(\mathbf{t}) \to 0$ as $|\mathbf{t}| \to 0$ by Theorem 8.19). Then

$$A_{\varepsilon,\delta} \le \eta \int_{|t|<\delta} |K_\varepsilon(\mathbf{t})| \, d\mathbf{t} \le \eta \|K\|_1$$

for all ε. Moreover, ϕ is a bounded function by Minkowski's inequality (note that $\|\phi\|_\infty \le (2\|f\|_p)^p$), and therefore $B_{\varepsilon,\delta}$ is less than a constant multiple of $\int_{|\mathbf{t}|\ge\delta} |K_\varepsilon(\mathbf{t})| \, d\mathbf{t}$, which tends to zero with ε. This proves that $I_\varepsilon \to 0$ as $\varepsilon \to 0$, and the theorem follows.

We leave it as an exercise to show that Theorem 9.6 and the next corollary are false when $p = \infty$.

Corollary 9.7 *For $1 \le p < \infty$, C_0^∞ is dense in $L^p(\mathbf{R}^n)$.*

Proof. Let $f \in L^p$, $1 \le p < \infty$. Given $\eta > 0$, write $f = g + h$ where g has compact support and $||h||_p < \eta$. Choose a kernel $K \in C_0^\infty$ with $\int_{\mathbf{R}^n} K = 1$, and

let $g_\varepsilon = g * K_\varepsilon$. Then $g_\varepsilon \in C_0^\infty$ and, by Theorem 9.6, $\|g - g_\varepsilon\|_p \to 0$. By Minkowski's inequality, $\|f - g_\varepsilon\|_p \le \|g - g_\varepsilon\|_p + \|h\|_p < \|g - g_\varepsilon\|_p + \eta$. Choosing ε so that $\|g - g_\varepsilon\|_p < \eta$, we obtain $\|f - g_\varepsilon\|_p < 2\eta$, and the corollary follows.

The next result is a substitute for Theorem 9.6 in case $f \in L^\infty$.

Theorem 9.8 *Let $f_\varepsilon = f * K_\varepsilon$, where $K \in L^1(\mathbf{R^n})$ and $\int_{\mathbf{R^n}} K = 1$. If $f \in L^\infty(\mathbf{R^n})$, then $f_\varepsilon \to f$ as $\varepsilon \to 0$ at every point of continuity of f, and the convergence is uniform on any set where f is uniformly continuous.*

Proof. Note that for every $\varepsilon > 0$, $f_\varepsilon(\mathbf{x})$ converges absolutely for all $\mathbf{x}$ since $f \in L^\infty$ and $K \in L^1$. As before,

$$|f_\varepsilon(\mathbf{x}) - f(\mathbf{x})| \le \int_{\mathbf{R^n}} |f(\mathbf{x} - \mathbf{t}) - f(\mathbf{x})||K_\varepsilon(\mathbf{t})|\, d\mathbf{t}.$$

If f is continuous at $\mathbf{x}$, then given $\eta > 0$, there exists $\delta > 0$ such that $|f(\mathbf{x} - \mathbf{t}) - f(\mathbf{x})| < \eta$ if $|\mathbf{t}| < \delta$. Hence,

$$\int_{\mathbf{R^n}} |f(\mathbf{x} - \mathbf{t}) - f(\mathbf{x})||K_\varepsilon(\mathbf{t})|\, d\mathbf{t} \le \eta \int_{|\mathbf{t}|<\delta} |K_\varepsilon(\mathbf{t})|\, d\mathbf{t} + 2\|f\|_\infty \int_{|\mathbf{t}|\ge\delta} |K_\varepsilon(\mathbf{t})|\, d\mathbf{t}.$$

Since $\int_{|\mathbf{t}|<\delta} |K_\varepsilon(\mathbf{t})|\, d\mathbf{t} \le \|K\|_1$ and $\int_{|t|\ge\delta} |K_\varepsilon(\mathbf{t})|\, d\mathbf{t} \to 0$ as $\varepsilon \to 0$ for any fixed δ, it follows that $|f_\varepsilon(\mathbf{x}) - f(\mathbf{x})| \to 0$. Since δ may be chosen to be independent of $\mathbf{x}$ on any set where f is uniformly continuous, the same proof shows that f_ε converges uniformly to f on such a set.

Before stating the next result, we generalize some notation already used in special cases (see (7.45a) and Exercise 31 of Chapter 7.). If $\psi(\mathbf{x})$ and $\phi(\mathbf{x})$ are defined in a neighborhood of $\mathbf{x}_0$ and if $\phi > 0$ there, we say that

$$\psi(\mathbf{x}) = O(\phi(\mathbf{x})) \text{ as } \mathbf{x} \to \mathbf{x}_0$$

if there is a constant c such that $|\psi(\mathbf{x})/\phi(\mathbf{x})| \le c$ near $\mathbf{x}_0$. If, in addition, $\lim_{\mathbf{x}\to\mathbf{x}_0} \psi(\mathbf{x})/\phi(\mathbf{x}) = 0$, we say that

$$\psi(\mathbf{x}) = o(\phi(\mathbf{x})) \text{ as } \mathbf{x} \to \mathbf{x}_0.$$

The common terminology for these situations is that ψ is "big oh" or "little oh" of ϕ as $\mathbf{x} \to \mathbf{x}_0$. In particular, the expressions $\psi(\mathbf{x}) = O(1)$ and $\psi(\mathbf{x}) = o(1)$ as $\mathbf{x} \to \mathbf{x}_0$ mean, respectively, that ψ is bounded and that $\psi \to 0$ as $\mathbf{x} \to \mathbf{x}_0$. The notation most commonly occurs when $\mathbf{x}_0$ is zero or infinite. A similar

notation is used when $\mathbf{x}$ is a discontinuous variable, say a sequence of integers tending to $+\infty$. For example, $a_k = O(1)$ and $a_k = o(1)$ as $k \to +\infty$ mean, respectively, that $\{a_k\}$ is a bounded sequence and that $a_k \to 0$ as $k \to +\infty$.

The next theorem concerns the pointwise convergence of f_ε when $f \in L^1$ (see Exercise 12 for the case $f \in L^p$).

Theorem 9.9 *Let $f_\varepsilon = f * K_\varepsilon$, where $f \in L^1(\mathbf{R^n})$, $K \in L^1(\mathbf{R^n}) \cap L^\infty(\mathbf{R^n})$, $\int_{\mathbf{R^n}} K = 1$, and $K(\mathbf{x}) = o(|\mathbf{x}|^{-n})$ as $|\mathbf{x}| \to +\infty$. Then $f_\varepsilon \to f$ as $\varepsilon \to 0$ at each point of continuity of f.*

Proof. If f is continuous at $\mathbf{x}$, then given $\eta > 0$, choose $\delta > 0$ such that $|f(\mathbf{x} - \mathbf{t}) - f(\mathbf{x})| < \eta$ if $|\mathbf{t}| < \delta$. Note that $f_\varepsilon(\mathbf{x})$ converges absolutely for all $\mathbf{x}$ since $f \in L^1$ and $K \in L^\infty$. As usual,

$$\begin{aligned}|f_\varepsilon(\mathbf{x}) - f(\mathbf{x})| &\le \eta \int_{|\mathbf{t}|<\delta} |K_\varepsilon(\mathbf{t})|\, d\mathbf{t} + \int_{|\mathbf{t}|\ge\delta} |f(\mathbf{x}-\mathbf{t}) - f(\mathbf{x})|\, |K_\varepsilon(\mathbf{t})|\, d\mathbf{t} \\ &\le \eta \|K\|_1 + \int_{|\mathbf{t}|\ge\delta} |f(\mathbf{x}-\mathbf{t})|\, |K_\varepsilon(\mathbf{t})|\, d\mathbf{t} + |f(\mathbf{x})| \int_{|\mathbf{t}|\ge\delta} |K_\varepsilon(\mathbf{t})|\, d\mathbf{t}.\end{aligned}$$

The last term on the right tends to zero with ε by Lemma 9.5. It is enough, therefore, to show that the second term tends to zero with ε. Write $|K(\mathbf{x})| = \mu(\mathbf{x})|\mathbf{x}|^{-n}$, where $\mu(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to +\infty$. Then

$$\begin{aligned}\int_{|\mathbf{t}|\ge\delta} |f(\mathbf{x}-\mathbf{t})|\, |K_\varepsilon(\mathbf{t})|\, d\mathbf{t} &= \int_{|\mathbf{t}|\ge\delta} |f(\mathbf{x}-\mathbf{t})|\, \mu\left(\frac{\mathbf{t}}{\varepsilon}\right) |\mathbf{t}|^{-n}\, d\mathbf{t} \\ &\le \delta^{-n} \left\{ \sup_{|\mathbf{t}|\ge\delta} \mu\left(\frac{\mathbf{t}}{\varepsilon}\right) \right\} \int_{|\mathbf{t}|\ge\delta} |f(\mathbf{x}-\mathbf{t})|\, d\mathbf{t}.\end{aligned}$$

Note that $\sup_{|\mathbf{t}|\ge\delta} \mu(\mathbf{t}/\varepsilon) \to 0$ as $\varepsilon \to 0$. Hence, since $\int_{|\mathbf{t}|\ge\delta} |f(\mathbf{x}-\mathbf{t})\, d\mathbf{t} \le \|f\|_1$, the last expression tends to zero with ε, and the theorem follows.

There are many classical kernels that satisfy the restrictions we have imposed. Let us list three important examples for the case $n = 1$.

The Poisson kernel. Let

$$K(x) = P(x) = \frac{1}{\pi} \frac{1}{1+x^2} \quad [x \in (-\infty, +\infty)].$$

Then $P \in L^1(-\infty, +\infty) \cap L^\infty(-\infty, +\infty)$, $\int_{-\infty}^{+\infty} P = 1$, P is positive, and $P(x) = o(|x|^{-1})$ as $|x| \to +\infty$. In fact, $P(x) = O(|x|^{-2})$ as $|x| \to +\infty$. We have

$$P_\varepsilon(x) = \frac{1}{\varepsilon} P\left(\frac{x}{\varepsilon}\right) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + x^2}, \quad \varepsilon > 0. \tag{9.10}$$

P_ε is called the *Poisson kernel*, and the convolution

$$f_\varepsilon(x) = (f * P_\varepsilon)(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \frac{\varepsilon}{\varepsilon^2 + (x-t)^2}\, dt$$

is called the *Poisson integral of* f.

Setting $\varepsilon = y$ and letting $f(x, y) = f_y(x)$, we obtain a function $f(x, y)$ defined in the upper half plane $\{(x, y) : -\infty < x < +\infty, y > 0\}$. Notice that $y/(y^2 + x^2)$ is the imaginary part of $-1/z, z = x + iy$, and so is harmonic in the upper half-plane; that is, $P_y(x)$ satisfies *Laplace's equation*

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) P_y(x) = 0 \quad \text{if } y > 0.$$

We leave it as an exercise to show that if $f \in L^p$, $1 \le p \le \infty$, then

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) f(x, y) = \int_{-\infty}^{+\infty} f(t) \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) P_y(x - t)\, dt,$$

so that $(\partial^2/\partial x^2 + \partial^2/\partial y^2) f(x, y) = 0$ for $y > 0$. Hence, $f(x, y)$ is also harmonic in the upper half-plane.

If f is integrable on $(-\infty, +\infty)$, it follows from Theorem 9.9 that $f(x, y) \to f(x)$ as $y \to 0$ wherever f is continuous. Thus, $f(x, y)$ solves the *Dirichlet problem* for the upper half-plane; that is, if $f(x)$ is continuous and integrable on $(-\infty, +\infty)$, then $f(x, y)$ defines a function that is harmonic in the upper half-plane and that tends to $f(x)$ as $y \to 0$. See also Exercises 15, 16.

The Fejér kernel. Let

$$K(x) = \frac{1}{\pi} \left(\frac{\sin x}{x}\right)^2 \quad [x \in (-\infty, +\infty)].$$

Then K satisfies the same conditions as $P(x)$ in the previous example, and $K_\varepsilon(x) = (1/\pi)[\varepsilon \sin^2(x/\varepsilon)/x^2]$. Setting $w = 1/\varepsilon$, $w > 0$, we obtain the *Fejér kernel*

$$F(x, w) = \frac{1}{\pi} \frac{\sin^2 wx}{wx^2}. \tag{9.11}$$

If $f \varepsilon L(-\infty, +\infty)$ and if f is continuous at x, then by Theorem 9.9,

$$\lim_{w \to +\infty} \frac{1}{\pi} \int_{-\infty}^{+\infty} f(x-t) \frac{\sin^2 wt}{wt^2} dt = f(x).$$

The Gauss–Weierstrass kernel. The function

$$K(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} \quad [x \in (-\infty, +\infty)]$$

also satisfies all the required conditions (see Exercise 11 of Chapter 6). Here, $K_\varepsilon(x) = (1/\sqrt{\pi}\varepsilon)e^{-x^2/\varepsilon^2}$, and letting $\varepsilon = \sqrt{y}$, $y > 0$, we obtain the *Gauss–Weierstrass kernel*

$$W(x, y) = \frac{1}{\sqrt{\pi y}} e^{-x^2/y}. \tag{9.12}$$

The convolution

$$Wf(x, y) = \big(f * W(\cdot, y)\big)(x) = \frac{1}{\sqrt{\pi y}} \int_{-\infty}^{+\infty} f(x-t) e^{-t^2/y}\, dt$$

is called the *Gauss–Weierstrass integral of* f. If f is integrable on $(-\infty, +\infty)$ and continuous at x, then

$$\lim_{y \to 0+} Wf(x, y) = f(x).$$

Notice that $W(x, y)$ satisfies the *heat equation*

$$\frac{\partial^2}{\partial x^2} W = 4 \frac{\partial}{\partial y} W$$

in the upper half-plane, as does $Wf(x, y)$ if $f \in L^p(-\infty, \infty)$ for some p, $1 \leq p \leq \infty$.

Higher dimensional versions of the Gauss–Weierstrass and Poisson kernels are discussed in Chapter 13.

If we strengthen the condition $K(x) = o(|\mathbf{x}|^{-n})$, $|\mathbf{x}| \to +\infty$, used in Theorem 9.9, we can obtain the convergence of f_ε to f almost everywhere. The following result is fairly typical of theorems of this kind. Its hypotheses are met by any of the three examples just listed.

Theorem 9.13 *Suppose that $f \in L(\mathbf{R^n})$, K is bounded, $K(\mathbf{x}) = O(|\mathbf{x}|^{-n-\lambda})$ as $|\mathbf{x}| \to +\infty$ for some $\lambda > 0$, and $\int_{\mathbf{R^n}} K = 1$. If $f_\varepsilon = f * K_\varepsilon$, then $f_\varepsilon \to f$ at each point of the Lebesgue set of f.*

Proof. Let $\mathbf{x}_0$ be a point of the Lebesgue set of f (see (7.14)), so that $\rho^{-n} \int_{|\mathbf{x}|<\rho} |f(\mathbf{x}_0 + \mathbf{x}) - f(\mathbf{x}_0)|\, d\mathbf{x} \to 0$ as $\rho \to 0$. By considering the function $f(\mathbf{x}_0 + \mathbf{x})$, we may assume that $\mathbf{x}_0 = 0$. Since the hypothesis on K implies that $K(\mathbf{x}) = o(|\mathbf{x}|^{-n})$, the conclusion follows from Theorem 9.9 if f is continuous at 0. Hence, subtracting from f a continuous function with compact support, which equals $f(0)$ at 0, we may suppose that $f(0) = 0$.

The hypotheses $|K(\mathbf{x})| \le M$ and $K(\mathbf{x}) = O(|\mathbf{x}|^{-n-\lambda})$ can be combined into a single estimate:

$$|K(\mathbf{x})| \le \frac{M_1}{(1 + |\mathbf{x}|)^{n+\lambda}}.$$

Hence,

$$|K_\varepsilon(\mathbf{x})| \le M_1 \frac{\varepsilon^\lambda}{(\varepsilon + |\mathbf{x}|)^{n+\lambda}}.$$

Therefore,

$$|f_\varepsilon(0)| \le M_1 \int_{\mathbf{R^n}} |f(\mathbf{x})| \frac{\varepsilon^\lambda}{(\varepsilon + |\mathbf{x}|)^{n+\lambda}}\, d\mathbf{x},$$

and it remains to show that the integral tends to zero. We will use the following lemma, which is of some independent interest.

Lemma 9.14 *Suppose that $f(\mathbf{x})$ is integrable over a spherical shell $a \le |\mathbf{x}| \le b$ and that $\phi(\rho)$ is continuous for $a \le \rho \le b$, $0 \le a < b < +\infty$. Let $F(\rho) = \int_{a \le |\mathbf{x}| \le \rho} f(\mathbf{x})\, d\mathbf{x}$ for $a \le \rho \le b$. Then*

$$\int_{a \le |\mathbf{x}| \le b} f(\mathbf{x}) \phi(|\mathbf{x}|)\, d\mathbf{x} = \int_a^b \phi(\rho)\, dF(\rho),$$

the integral on the right being a Riemann–Stieltjes integral.

Proof. Note that this reduces to the formula in Theorem 7.32(i) in case $n = 1$. In any case, writing $f = f^+ - f^-$, we see that F is the difference of two bounded increasing functions. Hence, F is of bounded variation on $[a, b]$ and $\int_a^b \phi\, dF$ is

well-defined. We may assume that $f \geq 0$. Let $I = \int_{a \leq |\mathbf{x}| \leq b} f(\mathbf{x})\phi(|\mathbf{x}|)\, d\mathbf{x}$ and let $\{a = \rho_0 < \rho_1 < \cdots < \rho_k = b\}$ be a partition of $[a, b]$. Then

$$I = \sum_{i=1}^{k} \int\limits_{\rho_{i-1} \leq |\mathbf{x}| \leq \rho_i} f(\mathbf{x})\phi(|\mathbf{x}|)\, d\mathbf{x},$$

and since $f \geq 0$,

$$\sum_{i=1}^{k} m_i \int\limits_{\rho_{i-1} \leq |\mathbf{x}| \leq \rho_i} f(\mathbf{x})\, d\mathbf{x} \leq I \leq \sum_{i=1}^{k} M_i \int\limits_{\rho_{i-1} \leq |\mathbf{x}| \leq \rho_i} f(\mathbf{x})\, d\mathbf{x},$$

where m_i and M_i are, respectively, the minimum and maximum of ϕ in $[\rho_{i-1}, \rho_i]$. This can be rewritten

$$\sum_{i=1}^{k} m_i[F(\rho_i) - F(\rho_{i-1})] \leq I \leq \sum_{i=1}^{k} M_i[F(\rho_i) - F(\rho_{i-1})].$$

By Theorem 2.24, the extreme terms in this inequality converge to $\int_a^b \phi(\rho) dF(\rho)$ as the norm of the partition tends to zero, and the lemma follows.

Returning to the proof of Theorem 9.13, let $F(\rho) = \int_{|\mathbf{x}| \leq \rho} |f(\mathbf{x})|\, d\mathbf{x}$. The hypotheses that $\mathbf{x}_0 = 0$ is a Lebesgue point of f and that $f(0) = 0$ imply that given $\eta > 0$, there is a $\delta > 0$ such that $F(\rho) < \eta\rho^n$ if $\rho \leq \delta$. Write

$$\int\limits_{\mathbf{R}^n} |f(\mathbf{x})| \frac{\varepsilon^\lambda}{(\varepsilon + |\mathbf{x}|)^{n+\lambda}}\, d\mathbf{x} = \int\limits_{|\mathbf{x}| \leq \delta} + \int\limits_{|\mathbf{x}| > \delta} = A + B.$$

Taking $\phi(\rho) = \varepsilon^\lambda/(\varepsilon + \rho)^{n+\lambda}$ and $[a, b] = [0, \delta]$ in Lemma 9.14, we have

$$A = \int\limits_0^\delta \frac{\varepsilon^\lambda}{(\varepsilon + \rho)^{n+\lambda}}\, dF(\rho).$$

Integrating by parts and observing that $F(0) = 0$, we obtain

$$A = \frac{\varepsilon^\lambda}{(\varepsilon + \delta)^{n+\lambda}} F(\delta) + (n + \lambda) \int\limits_0^\delta F(\rho) \frac{\varepsilon^\lambda}{(\varepsilon + \rho)^{n+\lambda+1}}\, d\rho.$$

The first term on the right tends to zero as $\varepsilon \to 0$. The definition of δ and the change of variables $\rho = \varepsilon t$ show that the second term is at most

$$(n+\lambda)\eta \int_0^\delta \rho^n \frac{\varepsilon^\lambda}{(\varepsilon+\rho)^{n+\lambda+1}}\, d\rho = (n+\lambda)\eta \int_0^{\delta/\varepsilon} \frac{t^n}{(1+t)^{n+\lambda+1}}\, dt.$$

Hence,

$$\limsup_{\varepsilon\to 0} A \le (n+\lambda)\eta \int_0^\infty \frac{t^n}{(1+t)^{n+\lambda+1}} = c\,\eta,$$

where $c\ (=c_{\lambda,n})$ is finite since $\lambda > 0$.

Finally, to estimate B, note that if $|\mathbf{x}| > \delta$, then $\varepsilon + |\mathbf{x}| > \delta$, so that

$$B \le \frac{\varepsilon^\lambda}{\delta^{n+\lambda}} \int_{|\mathbf{x}|>\delta} |f(\mathbf{x})|\, d\mathbf{x} \le \frac{\varepsilon^\lambda}{\delta^{n+\lambda}} \|f\|_1.$$

Hence, $\lim_{\varepsilon\to 0} B = 0$. Combining these estimates, we obtain $\limsup_{\varepsilon\to 0}(A + B) \le c\,\eta$, and the theorem follows by letting $\eta \to 0$. See Exercise 12 for the case $f \in L^p$, $p > 1$.

The kernels $\{K_\varepsilon\}$ for K satisfying the kinds of conditions earlier are examples of approximations of the identity.

9.3 The Hardy–Littlewood Maximal Function

Let f^* denote the Hardy–Littlewood maximal function of f:

$$f^*(\mathbf{x}) = \sup \frac{1}{|Q|} \int_Q |f(\mathbf{y})|\, d\mathbf{y},$$

where the supremum is taken over all cubes Q with center $\mathbf{x}$ and edges parallel to the coordinate axes (see (7.5)). If $f \in L^p(\mathbf{R^n})$ for some $p \ge 1$, then f is locally integrable in $\mathbf{R^n}$, and consequently, $|f| \le f^*$ a.e. by Theorem 7.11.

We observed on p. 136 in Section 7.2 that f^* is not integrable over $\mathbf{R^n}$ (unless $f = 0$ a.e.), but does satisfy the weak-type condition

$$|\{\mathbf{x} \in \mathbf{R^n} : f^*(\mathbf{x}) > \alpha\}| \le \frac{c}{\alpha}\|f\|_1 \quad (\alpha > 0), \tag{9.15}$$

where c depends only on n (Lemma 7.9). The behavior of f^* on the other L^p spaces, $1 < p \leq \infty$, turns out to be better. For example, it is clear from the definition of f^* that $f^*(\mathbf{x}) \leq \|f\|_\infty$ for all $\mathbf{x}$. Thus, f^* is bounded if f is, and $\|f^*\|_\infty \leq \|f\|_\infty$. The following theorem describes the behavior of f^* when $f \in L^p, p > 1$.

Theorem 9.16 *Let $1 < p \leq \infty$ and $f \in L^p(\mathbf{R^n})$. Then $f^* \in L^p(\mathbf{R^n})$ and*

$$\|f^*\|_p \leq c\|f\|_p,$$

where c depends only on n and p.

Proof. Let $f \in L^p(\mathbf{R^n})$. We may assume that $1 < p < \infty$ since the result is obvious with constant $c = 1$ when $p = \infty$. The idea is to obtain information for L^p by interpolating between the known results for L^1 and L^∞. For $\alpha > 0$, let

$$\omega(\alpha) = \left|\left\{\mathbf{x} \in \mathbf{R^n} : f^*(\mathbf{x}) > \alpha\right\}\right|$$

denote the distribution function of f^*. Fix $\alpha > 0$ and define a function g by $g(\mathbf{x}) = f(\mathbf{x})$ when $|f(\mathbf{x})| \geq \alpha/2$ and $g(\mathbf{x}) = 0$ otherwise. Note that $g \in L^1(\mathbf{R^n})$ since

$$\begin{aligned}\|g\|_1 &= \int\limits_{\{\mathbf{x}\in\mathbf{R^n}:|f(\mathbf{x})|\geq\alpha/2\}} |f(\mathbf{x})|\,d\mathbf{x} \\ &\leq \int\limits_{\mathbf{R^n}} |f(\mathbf{x})| \left(\frac{|f(\mathbf{x})|}{\alpha/2}\right)^{p-1} d\mathbf{x} = \left(\frac{2}{\alpha}\right)^{p-1} \|f\|_p^p < \infty.\end{aligned}$$

Also, the difference $f - g \in L^\infty(\mathbf{R^n})$; in fact, $\|f - g\|_\infty \leq \alpha/2$. Then since $|f(\mathbf{x})| \leq |g(\mathbf{x})| + \alpha/2$,

$$f^*(\mathbf{x}) \leq \sup \frac{1}{|Q|} \int\limits_Q |g(\mathbf{y})|\,d\mathbf{y} + \frac{\alpha}{2} = g^*(\mathbf{x}) + \frac{\alpha}{2}.$$

In particular,

$$\{\mathbf{x} \in \mathbf{R^n} : f^*(\mathbf{x}) > \alpha\} \subset \left\{\mathbf{x} \in \mathbf{R^n} : g^*(\mathbf{x}) > \frac{\alpha}{2}\right\},$$

so that, by (9.15),

$$\begin{aligned}\omega(\alpha) &\le \left|\left\{\mathbf{x} \in \mathbf{R^n} : g^*(\mathbf{x}) > \frac{\alpha}{2}\right\}\right| \\ &\le \frac{2c}{\alpha}\|g\|_1 = \frac{2c}{\alpha} \int\limits_{\{\mathbf{x}\in\mathbf{R^n}:|f(\mathbf{x})|\ge\alpha/2\}} |f(\mathbf{x})|\, d\mathbf{x}.\end{aligned}$$

We have the formula $\int_{\mathbf{R^n}} f^{*p} d\mathbf{x} = p\int_0^\infty \alpha^{p-1}\omega(\alpha)\, d\alpha$, which was stated on two occasions: Exercise 16 of Chapter 5, and Exercise 5 of Chapter 6. Hence,

$$\int\limits_{\mathbf{R^n}} f^{*p}\, d\mathbf{x} \le p\int\limits_0^\infty \alpha^{p-1}\left[\frac{2c}{\alpha}\int\limits_{\{\mathbf{x}\in\mathbf{R^n}:|f(\mathbf{x})|\ge\alpha/2\}} |f(\mathbf{x})|\, d\mathbf{x}\right] d\alpha.$$

Interchanging the order of integration in the expression on the right (which is justified since the integrand is nonnegative), we obtain

$$\int\limits_{\mathbf{R^n}} f^{*p}\, d\mathbf{x} \le 2cp\int\limits_{\mathbf{R^n}} |f(\mathbf{x})|\left(\int\limits_0^{2|f(\mathbf{x})|} \alpha^{p-2} d\alpha\right) d\mathbf{x}.$$

Since $p-2 > -1$ (i.e., $p > 1$), the inner integral equals $(2|f(\mathbf{x})|)^{p-1}/(p-1)$ a.e. (wherever $f(\mathbf{x})$ is finite), so that

$$\int\limits_{\mathbf{R^n}} f^{*p}\, d\mathbf{x} \le \frac{2^p pc}{p-1}\int\limits_{\mathbf{R^n}} |f(\mathbf{x})|^p d\mathbf{x} = \frac{2^p pc}{p-1}\|f\|_p^p.$$

Taking pth roots, we see that $\|f^*\|_p \le C_p\|f\|_p$, where $C_p^p = 2^p pc/(p-1)$. This completes the proof. Note that the constant C_p tends to $+\infty$ as $p \to 1$ and is bounded as $p \to \infty$.

The Hardy–Littlewood maximal function plays an important role in many parts of analysis concerned with operator theory and differentiation. It arose naturally in Chapter 7 in connection with Lebesgue's differentiation theorem, and it will be used in the proof of Theorem 9.19 and frequently in later chapters. As another illustration of its usefulness, we have the following result.

Theorem 9.17 *Let $K(x)$ be nonnegative and integrable on $\mathbf{R^n}$ and suppose that $K(\mathbf{x})$ depends only on $|\mathbf{x}|$ and decreases as $|\mathbf{x}|$ increases (i.e., $K(\mathbf{x}) = \phi(|\mathbf{x}|)$, where $\phi(t)$, $t > 0$, is monotone decreasing). Then*

$$\sup_{\varepsilon>0} |(f * K_\varepsilon)(\mathbf{x})| \le cf^*(\mathbf{x}),$$

with c independent of f. In particular, for such kernels K,

$$||\sup_{\varepsilon>0} |(f * K_\varepsilon)|\,||_p \le c||f||_p, \quad \text{if } 1 < p \le \infty,$$

$$|\{\mathbf{x} : \sup_{\varepsilon>0} |(f * K_\varepsilon)(\mathbf{x})| > \alpha\}| \le \frac{c}{\alpha}||f||_1, \ \alpha > 0, \quad \text{if } p = 1,$$

with c independent of f and α.

Proof. We first remark that there is a constant c depending only on n such that

$$\sup_{\delta>0} \delta^{-n} \int_{|\mathbf{y}|<\delta} |f(\mathbf{x}-\mathbf{y})|\, d\mathbf{y} \le cf^*(\mathbf{x}).$$

This follows by enclosing the ball $|\mathbf{y}| < \delta$ in the cube with center 0 and edge 2δ and observing that ratios of the two volumes are bounded independent of δ.

To prove the result, we will use a method based on Tonelli's theorem. Fix ε and let $E = \{(\mathbf{y}, t) : \mathbf{y} \in \mathbf{R^n},\ t > 0,\ K_\varepsilon(\mathbf{y}) > t\}$. Then E is a measurable subset of $\mathbf{R^{n+1}}$ by Theorem 5.1, and

$$K_\varepsilon(\mathbf{y}) = \int_0^{K_\varepsilon(\mathbf{y})} dt = \int_0^\infty \chi_E(\mathbf{y}, t)\, dt.$$

Hence,

$$|(f * K_\varepsilon)(\mathbf{x})| = \left| \int_{\mathbf{R^n}} f(\mathbf{x}-\mathbf{y}) K_\varepsilon(\mathbf{y})\, d\mathbf{y} \right| \le \int_{\mathbf{R^n}} |f(\mathbf{x}-\mathbf{y})| \left[\int_0^\infty \chi_E(\mathbf{y}, t)\, dt \right] d\mathbf{y}.$$

Changing the order of integration in the last expression, we obtain

$$\begin{aligned} |(f * K_\varepsilon)(\mathbf{x})| &\le \int_0^\infty \left[\int_{\mathbf{R^n}} |f(\mathbf{x}-\mathbf{y})| \chi_E(\mathbf{y}, t)\, d\mathbf{y} \right] dt \\ &= \int_0^\infty \left[\int_{\{\mathbf{y}: K_\varepsilon(\mathbf{y})>t\}} |f(\mathbf{x}-\mathbf{y})|\, d\mathbf{y} \right] dt. \end{aligned}$$

Let $E_t = \{\mathbf{y} : K_\varepsilon(\mathbf{y}) > t\}$, $t > 0$. Since $K(\mathbf{y})$ depends only on $|\mathbf{y}|$ and decreases as $|\mathbf{y}|$ increases, E_t is a ball with center $\mathbf{0}$ unless it is empty or the single point $\mathbf{0}$.

In the t-integration in the last display, we may ignore any values of t such that $|E_t| = 0$ since the inner integral is then zero. For all other t, E_t is a ball, and by our earlier remark, the inner integral satisfies

$$\int_{E_t} |f(\mathbf{x}-\mathbf{y})|\,d\mathbf{y} = |E_t| \left[\frac{1}{|E_t|} \int_{E_t} |f(\mathbf{x}-\mathbf{y})|\,d\mathbf{y} \right] \le |E_t|\, cf^*(\mathbf{x}).$$

By combining estimates, we have

$$|(f * K_\varepsilon)(\mathbf{x})| \le \int_0^\infty |E_t|\, cf^*(\mathbf{x})\,dt = cf^*(\mathbf{x}) \int_0^\infty |E_t|\,dt.$$

Finally, note that $|E_t|$ is the distribution function of K_ε, so that

$$\int_0^\infty |E_t|\,dt = \|K_\varepsilon\|_1 = \|K\|_1.$$

Therefore, $|(f * K_\varepsilon)(\mathbf{x})| \le c||K||_1 f^*(\mathbf{x})$, and the first statement of the theorem follows by taking the sup over $\varepsilon > 0$. The second statement is then a corollary of Theorem 9.16 if $p > 1$, and of (9.15) if $p = 1$.

We leave it to the reader (Exercise 18) to show that Theorem 9.17 can also be derived from the formula in the conclusion of Lemma 9.14 (see, e.g., Exercise 17 and the proof of Theorem 12.61).

In particular, for the kernel $K(\mathbf{x}) = 1/(1 + |\mathbf{x}|^{n+\lambda})$, $\lambda > 0$, Theorem 9.17 gives

$$\sup_{\varepsilon>0} \left| \int_{\mathbf{R}^n} f(\mathbf{x}-\mathbf{y}) \frac{\varepsilon^\lambda}{\varepsilon^{n+\lambda} + |\mathbf{y}|^{n+\lambda}}\,d\mathbf{y} \right| \le cf^*(\mathbf{x}), \tag{9.18}$$

a fact that will be used in the next section.

Note that the conclusion of Theorem 9.17 is valid for any K that is majorized in absolute value by a kernel satisfying the hypothesis of Theorem 9.17. This includes any K satisfying the hypothesis of Theorem 9.13.

9.4 The Marcinkiewicz Integral

We recall from Theorem 6.17, that if F is a closed subset of a bounded open interval (a, b) in $\mathbf{R}^1$, and if $\delta(x)$ denotes the distance from x to F, then the Marcinkiewicz integral

$$M_\lambda(x) = \int_a^b \frac{\delta^\lambda(y)}{|x-y|^{1+\lambda}}\,dy \quad (\lambda > 0)$$

is integrable over F. More generally, in Exercise 7 of Chapter 6, we considered the expression

$$\int_{\mathbf{R^1}} \frac{\delta^\lambda(y)f(y)}{|x-y|^{1+\lambda}}\,dy \quad (\lambda > 0),$$

where f is nonnegative and integrable over the complement of F. If $f = \chi_{(a,b)}$, this reduces to $M_\lambda(x)$. In case $\lambda = 1$, M_λ plays a role in proving Theorem 12.67.

Now we consider L^p estimates for n-dimensional analogues, namely, for the integral

$$J_\lambda(f)(\mathbf{x}) = \int_{\mathbf{R^n}} \frac{\delta^\lambda(\mathbf{y})f(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{n+\lambda}}\,d\mathbf{y} \quad (\mathbf{x} \in \mathbf{R^n}),$$

and the modified form

$$H_\lambda(f)(\mathbf{x}) = \int_{\mathbf{R^n}} \frac{\delta^\lambda(\mathbf{y})f(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{n+\lambda} + \delta(\mathbf{x})^{n+\lambda}}\,d\mathbf{y}.$$

Here again, $\lambda > 0$, $\delta(\mathbf{x})$ denotes the distance from $\mathbf{x}$ to a closed set $F \subset \mathbf{R^n}$ and f is nonnegative and measurable on $\mathbf{R^n}$. Notice that $H_\lambda(f)$ and $J_\lambda(f)$ are equal in F since δ is zero there. For the same reason, $J_\lambda(f)$ and $H_\lambda(f)$ are independent of the values of f on F. Therefore, we may assume for simplicity that $f = 0$ on F.

We will prove in the next theorem that if $f \in L^p(\mathbf{R^n} - F)$, $1 \le p < \infty$, then $H_\lambda(f) \in L^p(\mathbf{R^n})$. This implies the basic fact that $J_\lambda(f) \in L^p(F)$. (In general, $J_\lambda(f)$ diverges outside F: see, e.g., Exercise 9 of Chapter 6.) For the proof, it will be convenient to consider one more modification of $J_\lambda(f)$, namely,

$$H'_\lambda(f)(\mathbf{x}) = \int_{\mathbf{R^n}} \frac{\delta^\lambda(\mathbf{y})f(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{n+\lambda} + \delta(\mathbf{y})^{n+\lambda}}\,d\mathbf{y}.$$

As we move from a point $\mathbf{x}$ to another point $\mathbf{y}$, the distance from F does not increase by more than $|\mathbf{x}-\mathbf{y}|$. Hence,

$$|\delta(\mathbf{x}) - \delta(\mathbf{y})| \le |\mathbf{x}-\mathbf{y}|.$$

It follows that $\delta(\mathbf{y}) < |\mathbf{x} - \mathbf{y}| + \delta(\mathbf{x})$, so that we have

$$\delta^{n+\lambda}(\mathbf{y}) \leq 2^{n+\lambda}[|\mathbf{x} - \mathbf{y}|^{n+\lambda} + \delta(\mathbf{x})^{n+\lambda}],$$

as well as a similar inequality with $\mathbf{x}$ and $\mathbf{y}$ interchanged. We immediately obtain that

$$2^{-n-\lambda-1} H'_\lambda(f)(\mathbf{x}) \leq H_\lambda(f)(\mathbf{x}) \leq 2^{n+\lambda+1} H'_\lambda(f)(\mathbf{x}).$$

Thus, inequalities for H'_λ lead to ones for H_λ, but H'_λ is easier to deal with.

Theorem 9.19 *If $f \in L^p(\mathbf{R^n})$, $1 \leq p < \infty$, and $\lambda > 0$, then $H_\lambda(f) \in L^p(\mathbf{R^n})$ and*

$$\|H_\lambda(f)\|_p \leq c\,\|f\|_p,$$

where c is independent of f. In particular, $\|J_\lambda(f)\|_{p,F} \leq c\,\|f\|_p$.

Proof. Fix p, $1 \leq p < \infty$, and let g be any nonnegative function with $\|g\|_{p'} \leq 1$, where $1/p + 1/p' = 1$. By interchanging the order of integration, we obtain

$$\int_{\mathbf{R^n}} H'_\lambda(f)(\mathbf{x}) g(\mathbf{x})\, d\mathbf{x} = \int_{\mathbf{R^n}} f(\mathbf{y}) \delta^\lambda(\mathbf{y}) \left[\int_{\mathbf{R^n}} \frac{g(\mathbf{x})}{|\mathbf{x} - \mathbf{y}|^{n+\lambda} + \delta(\mathbf{y})^{n+\lambda}}\, d\mathbf{x} \right] d\mathbf{y}.$$

The outer integration on the right can be restricted to $\mathbf{R^n} - F$ without changing the value of the integral. However, if $\mathbf{y} \in \mathbf{R^n} - F$, then $\delta(\mathbf{y}) > 0$, and it follows from (9.18) that the inner integral on the right is bounded by $c\delta(\mathbf{y})^{-\lambda} g^*(\mathbf{y})$. Combining this estimate with Holder's inequality and Theorem 9.16, we obtain

$$\begin{aligned}
\int_{\mathbf{R^n}} H'_\lambda(f)(\mathbf{x}) g(\mathbf{x})\, d\mathbf{x} &\leq c \int_{\mathbf{R^n}} f(\mathbf{y}) g^*(\mathbf{y})\, d\mathbf{y} \\
&\leq c\,\|f\|_p \|g^*\|_{p'} \\
&\leq c_1 \|f\|_p \|g\|_{p'} \leq c_1 \|f\|_p.
\end{aligned}$$

By (8.9), the supremum of the left side for all such g is $\|H'_\lambda(f)\|_p$, so that $\|H'_\lambda(f)\|_p \leq c_1 \|f\|_p$, and the theorem follows.

Exercises

1. Use Minkowski's integral inequality (see Exercise 8 of Chapter 8) to prove Theorem 9.1 for $1 < p < \infty$.

2. (a) Prove Young's Theorem 9.2. (For $f, g \geq 0$ and $p, q, r < \infty$, write

$$(f * g)(\mathbf{x}) = \int f(\mathbf{t})^{p/r} g(\mathbf{x} - \mathbf{t})^{q/r} \cdot f(\mathbf{t})^{p(1/p-1/r)} \cdot g(\mathbf{x} - \mathbf{t})^{q(1/q-1/r)}\, d\mathbf{t},$$

and apply Hölder's inequality for three functions (Exercise 6 of Chapter 8) with exponents r, p_1, and p_2, where $1/p_1 = 1/p - 1/r$, $1/p_2 = 1/q - 1/r$.)

(b) Suppose that the conclusion of Theorem 9.2 holds for three indices $p, q, r > 0$. Show that $1/r = 1/p + 1/q - 1$. (Apply the inequality in the conclusion to the dilated functions $f(\lambda\mathbf{x})$ and $g(\lambda\mathbf{x})$, $\lambda > 0$, and let λ vary. A similar method is used in Exercise 13 of Chapter 14.)

3. (a) Show that if $f \in L^p(\mathbf{R^n})$ and $K \in L^{p'}(\mathbf{R^n})$, $1 \leq p \leq \infty$, $1/p + 1/p' = 1$, then $f * K$ is bounded and continuous in $\mathbf{R^n}$.

(b) Sketch the (trapezoidal) graph of $\chi_I * \chi_J$ where I and J are one-dimensional intervals. Consider also the case when the two intervals are the same.

4. (a) Show that the function h defined by $h(x) = e^{-1/x^2}$ for $x > 0$ and $h(x) = 0$ for $x \leq 0$ is in C^∞.

(b) Show that the function $g(x) = h(x - a)h(b - x)$, $a < b$, is C^∞ with support $[a, b]$.

(c) Construct a function in $C_0^\infty(\mathbf{R^n})$ whose support is a ball or an interval.

5. Let G and G_1 be bounded open subsets of $\mathbf{R^n}$ such that $\overline{G}_1 \subset G$. Construct a function $h \in C_0^\infty$ such that $h = 1$ in G_1 and $h = 0$ outside G. (Choose an open G_2 such that $\overline{G}_1 \subset G_2$, $\overline{G}_2 \subset G$. Let $h = \chi_{G_2} * K$ for a $K \in C^\infty$ with suitably small support and $\int K = 1$.)

6. Prove Theorem 9.4.

7. Let $f \in L^p(-\infty, +\infty)$, $1 \leq p \leq \infty$. Show that the Poisson integral of f, $f(x, y)$, is harmonic in the upper half-plane $y > 0$. (Show that $((\partial^2/\partial x^2) + (\partial^2/\partial y^2))f(x, y) = \int_{-\infty}^{+\infty} f(t)((\partial^2/\partial x^2) + (\partial^2/\partial y^2))P_y(x - t)\, dt$.)

8. (*Schur's lemma*) For $s, t \geq 0$, let $K(s, t)$ satisfy $K \geq 0$ and $K(\lambda s, \lambda t) = \lambda^{-1}K(s, t)$ for all $\lambda > 0$, and suppose that $\int_0^\infty t^{-1/p}K(1, t)\, dt = \gamma < +\infty$ for some p, $1 \leq p \leq \infty$. For example, $K(s, t) = 1/(s + t)$ has these properties. Show that if

$$(Tf)(s) = \int_0^\infty f(t)K(s,t)\,dt \quad (f \geq 0),$$

then $\|Tf\|_p \leq \gamma\|f\|_p$. (Note that $K(s,t) = s^{-1}K(1,t/s)$, and therefore $(Tf)(s) = \int_0^\infty f(ts)K(1,t)\,dt$. Now apply Minkowski's integral inequality [see Exercise 8 of Chapter 8].)

9. (a) The maximal function is defined as $f^*(\mathbf{x}) = \sup |Q|^{-1} \int_Q |f|$, where the supremum is taken over cubes Q with center $\mathbf{x}$. Let $f^{**}(\mathbf{x})$ be defined similarly, but with the supremum taken over all cubes Q *containing* $\mathbf{x}$. Thus, $f^*(\mathbf{x}) \leq f^{**}(\mathbf{x})$. Show that there is a positive constant c depending only on the dimension such that $f^{**}(\mathbf{x}) \leq cf^*(\mathbf{x})$.

(b) If f^{**} were instead defined to be $\sup |B|^{-1} \int_B |f|$ where the supremum is taken over all balls B containing $\mathbf{x}$, show that there are positive constants c_1 and c_2 depending only on the dimension so that $c_1 f^*(\mathbf{x}) \leq f^{**}(\mathbf{x}) \leq c_2 f^*(\mathbf{x})$ for all $\mathbf{x}$.

10. Let $T : f \to Tf$ be a function transformation that is *sublinear*; that is, T has the property that if Tf_1 and Tf_2 are defined, then so is $T(f_1 + f_2)$, and

$$|T(f_1 + f_2)(\mathbf{x})| \leq |(Tf_1)(\mathbf{x})| + |(Tf_2)(\mathbf{x})|.$$

Suppose also that there are constants c_1 and c_2 such that T satisfies $\|Tf\|_\infty \leq c_1\|f\|_\infty$ and $|\{\mathbf{x} : |(Tf)(\mathbf{x})| > \alpha\}| \leq c_2\alpha^{-1}\|f\|_1$, $\alpha > 0$. Show that for $1 < p < \infty$, there is a constant c_3 such that $\|Tf\|_p \leq c_3\|f\|_p$. This is a special case of an interpolation result due to Marcinkiewicz. (An example of such a T is the maximal function operator $Tf = f^*$, and the proof in the general case is like that for f^*.)

11. Generalize Theorem 9.6 as follows: Let $f_\varepsilon = f * K_\varepsilon$, $K \in L^1(\mathbf{R^n})$ and $\int_{\mathbf{R^n}} K = \gamma$. If $f \in L^p(\mathbf{R^n})$, $1 \leq p < \infty$, show that $\|f_\varepsilon - \gamma f\|_p \to 0$. Derive analogous results for Theorems 9.8, 9.9, and 9.13. (The case $\gamma \neq 0$ follows from the case $\gamma = 1$ by considering $K(\mathbf{x})/\gamma$.)

12. Show that the conclusions of Theorems 9.9 and 9.13 remain true if the assumption that $f \in L^1$ is replaced by $f \in L^p$, $p > 1$.

13. Let $f \in L^p(0,1)$, $1 \leq p < \infty$, and for each $k = 1, 2, \ldots$, define a function f_k on (0,1) by letting $I_{k,j} = \{x : (j-1)2^{-k} \leq x < j2^{-k}\}$, $j = 1, \ldots, 2^k$, and setting $f_k(x)$ equal to $|I_{k,j}|^{-1} \int_{I_{k,j}} f$ for $x \in I_{k,j}$. Prove that $f_k \to f$ in $L^p(0,1)$ norm. (Exercise 17 of Chapter 7 may be helpful for the case $p = 1$.)

14. Show that Theorem 9.6 and Corollary 9.7 fail for $p = \infty$.

15. Regarding the Dirichlet problem for the upper half-space, more can be said about the behavior of the Poisson integral $f(x,y)$ of $f(x)$ near a point of continuity of $f(x)$. Prove that if $f \in L^p(-\infty,\infty)$, $1 \leq p \leq \infty$, and f

is continuous at a point x_0, then $f(x,y) \to f(x_0)$ as (x,y) approaches x_0 unrestrictedly, that is, as $x \to x_0$ and $y \to 0$, $y > 0$.

16. Let $1 \le p \le \infty$, $f \in L^p(-\infty,\infty)$, and x_0 be a Lebesgue point of f. Show that the Poisson integral $f(x,y)$ of f converges nontangentially to $f(x_0)$, that is, show that for any $\gamma > 0$, $f(x,y) \to f(x_0)$ as $x \to x_0$ and $y \to 0$ with $|x - x_0| \le \gamma y$. (Note that the Poisson kernel satisfies $P_y(t+z) \le C_\gamma P_y(t)$ if $|z| \le \gamma y$, with C_γ independent of y, t, z.) (See also Theorems 12.42 and 12.64.)

17. Prove that the conclusion of Lemma 9.14 holds without assuming ϕ is continuous provided $\int_a^b \phi(\rho)\, dF(\rho)$ exists. Show, for example, that the conclusion holds if ϕ is monotone and finite on $[a,b]$. (For the second part, recall from Theorem 2.21 that $\int_a^b \phi\, dF$ exists if $\int_a^b F\, d\phi$ does.)

18. Derive the first part of Theorem 9.17 by using the formula in the conclusion of Lemma 9.14 (even though the function ϕ in Theorem 9.17 is not assumed to be continuous). (Use Exercise 17. Note that if $K(\mathbf{x}) = \phi(|\mathbf{x}|)$ satisfies the hypothesis of Theorem 9.17, then $\phi(|\mathbf{x}|) = o(|\mathbf{x}|^{-n})$ as $|\mathbf{x}| \to 0$ and as $|\mathbf{x}| \to \infty$).

19. Let $K(\mathbf{x})$ be a nonnegative, decreasing, integrable radial function on $\mathbf{R^n}$ (i.e., K satisfies the hypothesis of Theorem 9.17), and let $\int K = 1$. Use Theorem 9.17 to show that if $1 \le p \le \infty$ and $f \in L^p(\mathbf{R^n})$, then $f * K_\varepsilon \to f$ a.e. as $\varepsilon \to 0$. (In case $1 \le p < \infty$, a proof reminiscent of the proof of Lebesgue's differentiation theorem can be constructed by applying Corollary 9.7.)

20. Show that the conclusion $f * K_\varepsilon \to f$ in Exercise 19 is valid at every Lebesgue point of f. (A proof based on integrating the formula in Lemma 9.14 by parts is possible; cf. Exercise 18.)

21. Let f and g be locally integrable functions on $\mathbf{R^n}$ and suppose that $|f| * |g|$ is finite a.e. in $\mathbf{R^n}$. Prove that $f * g$ is measurable on $\mathbf{R^n}$. (See the argument in the last part of the proof of Theorem 9.1 involving the truncated functions f_N and also truncate g.)

22. As we know from Chapter 7, the maximal function f^* of an $f \in L^1(\mathbf{R^n})$ may not be locally integrable. Show that if f satisfies the stronger condition $\int_{\mathbf{R^n}} |f|(1 + \log^+ |f|)\, d\mathbf{x} < \infty$, then $f^* \in L^1(E)$ for every measurable set E with $|E| < \infty$, and

$$\int_E f^*\, d\mathbf{x} \le C\Big(|E| + \int_{\mathbf{R^n}} |f| \log^+ |f|\, d\mathbf{x}\Big),$$

with C independent of f and E. (Let $\omega(\alpha)$ denote the distribution function of f^* relative to E. Write $\int_0^\infty \omega(\alpha)\, d\alpha = \int_0^\gamma + \int_\gamma^\infty$ for $\gamma > 0$. The first integral on the right is bounded by $\gamma|E|$. In the second integral, use the estimate $\omega(\alpha) \le C\alpha^{-1} \int_{\{|f| > \alpha/2\}} |f|$; cf. the proof of Theorem 9.16.)

23. (a) Show that there is no identity element for the convolution operation on $L^1(\mathbf{R^n})$, that is, there is no function $k \in L^1(\mathbf{R^n})$ such that $f * k = f$ a.e. for every $f \in L^1(\mathbf{R^n})$.

(b) Show that if $f, g \in L^2(\mathbf{R^n})$, then $f * g$ belongs to $L^\infty(\mathbf{R^n})$ but not necessarily to $L^2(\mathbf{R^n})$. See also Lemma 13.49.

10

Abstract Integration

In the preceding chapters, we developed a theory of integration based on a theory of measurable sets. The notion of the measure of a set was in turn based on the primitive and classical notion of the measure (or volume) of an interval in $\mathbf{R^n}$; this led almost automatically by the process of covering to the notion of measure for more general sets.

In this chapter, we follow an alternate approach. We will consider a family of sets and assume that they all have *measures*, that is, assume that with each member of the family, we can associate a nonnegative number satisfying elementary and natural requirements that justify calling it a measure. Starting with this assumption, we will develop a theory of integration that follows the pattern of Lebesgue integration. The advantage of this method is that it can be applied not only to $\mathbf{R^n}$ but also to general abstract spaces with much less geometric structure than $\mathbf{R^n}$. Thus, it is important for applications. There are new questions that arise in the abstract setting, but many of the theorems and proofs are practically the same as those for Lebesgue measure in $\mathbf{R^n}$. In such cases, we will usually refer to earlier chapters for proofs.

It is natural to ask how we can construct such measures. One possible approach is to start with the more elementary notion of an *outer measure* in an abstract space and, as in the case of $\mathbf{R^n}$ discussed in Chapter 3, select a subclass of sets on which the outer measure has additional properties, qualifying it as a measure. This idea will be developed in Chapter 11.

10.1 Additive Set Functions and Measures

Let $\mathscr{S}$ be a fixed set, and let Σ be a σ-algebra of subsets of $\mathscr{S}$; that is, let Σ satisfy the following:

(a) $\mathscr{S} \in \Sigma$.

(b) If $E \in \Sigma$, then its complement CE $(=\mathscr{S} - E) \in \Sigma$ (i.e., Σ is closed under complements).

(c) If $E_k \in \Sigma$ for $k = 1, 2, \ldots$, then $\bigcup E_k \in \Sigma$ (i.e., Σ is closed under countable unions).

It is easy to see that the definition is unchanged if condition (a) is replaced by the assumption that Σ be nonempty; see also p. 49 in Section 3.2. Another widely used term for a σ-algebra is a *countably additive family of sets*.

Immediate consequences of the definition are that the following sets belong to Σ:

(1) The empty set $\emptyset\,(= C\mathscr{S})$,

(2) $\bigcap E_k$ if $E_k \in \Sigma,\ k = 1, 2, \ldots,$

(3) $\limsup E_k\,(= \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} E_k)$ and $\liminf E_k\,(= \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} E_k)$ if each $E_k \in \Sigma$,

(4) $E_1 - E_2\,(= E_1 \cap CE_2)$ if $E_1, E_2 \in \Sigma$.

We recall the basic fact that the collection of Lebesgue measurable subsets of $\mathbf{R^n}$ is a σ-algebra: see Theorem 3.20. In general, the elements E of a σ-algebra Σ are called Σ-*measurable* sets, or simply *measurable sets* if it is clear from context what Σ is.

If Σ is a σ-algebra, then a real-valued function $\phi(E)$, $E \in \Sigma$, is called an *additive set function* on Σ if

(i) $\phi(E)$ is finite for every $E \in \Sigma$,

(ii) $\phi(\bigcup E_k) = \sum \phi(E_k)$ for every countable family $\{E_k\}$ of disjoint sets in Σ.

Since $\bigcup E_k$ is independent of the order of the E_k's, the series in (ii) converges absolutely.

We obtain a simple example of a set function by choosing Σ to be the σ-algebra of all subsets of $\mathscr{S}$ and defining $\phi(E) = \chi_E(x_0)$ for a fixed $x_0 \in \mathscr{S}$. As another example, let Σ be the collection of all Lebesgue measurable subsets of $\mathbf{R^n}$, and define $\phi(E) = \int_E f$, where $f \in L(\mathbf{R^n})$.

A function $\mu(E)$ defined for E in Σ is called a *measure* on Σ if

(i) $0 \le \mu(E) \le +\infty$,

(ii) $\mu(\bigcup E_k) = \sum \mu(E_k)$ for every countable family $\{E_k\}$ of disjoint sets in Σ.

The choices $\mu \equiv 0$ or $\mu \equiv +\infty$ are always possible, but of little interest.

If μ is a measure on Σ, then the triplet $(\mathscr{S}, \Sigma, \mu)$ is called a *measure space*. For example, Lebesgue measure together with the class of Lebesgue measurable subsets of $\mathbf{R^n}$ is a measure space. As another example, let $\mathscr{S}$ be any countable set, $\mathscr{S} = \{x_k\}$, and let $\{a_k\}$ be a sequence of nonnegative numbers. Let Σ be the family of all subsets of $\mathscr{S}$, and define $\delta(E) = \sum a_{k_j}$ if $E = \{x_{k_j}\}$. Then $(\mathscr{S}, \Sigma, \delta)$ is a measure space. Such a space is called a *discrete* measure space.

The distinction between a measure and an additive set function is that a measure is nonnegative, but may be infinite, while an additive set function may take both positive and negative values, but is finite. Any nonnegative

additive set function is a finite measure and vice versa. There are similarities between many of the properties of set functions and measures. If $E_1 \subset E_2$ and μ is a measure, then $\mu(E_2 - E_1) + \mu(E_1) = \mu(E_2)$, so that $\mu(E_2 - E_1) = \mu(E_2) - \mu(E_1)$ if $\mu(E_1)$ is finite. If ϕ is an additive set function, then the formula $\phi(E_2 - E_1) = \phi(E_2) - \phi(E_1)$, $E_1 \subset E_2$, always holds. Choosing $E_1 = E_2$, we see that $\phi(\emptyset) = 0$ for an additive set function, and also $\mu(\emptyset) = 0$ for a measure, unless $\mu(E) = +\infty$ for all E. Moreover, if $E_1 \subset E_2$, then $\mu(E_1) \leq \mu(E_2)$ even if $\mu(E_1) = +\infty$, and $\phi(E_1) \leq \phi(E_2)$ if $\phi \geq 0$.

The next few results concern limit properties and a basic decomposition for additive set functions. Both $\mathscr{S}$ and Σ are fixed.

Theorem 10.1 *If $\{E_k\}$ is a monotone sequence of sets in Σ (i.e., $E_k \nearrow E$ or $E_k \searrow E$) and ϕ is an additive set function, then $\phi(E) = \lim_{k\to\infty} \phi(E_k)$.*

Proof. If $E_k \nearrow E$, then $E = \bigcup E_k = E_1 \cup (E_2 - E_1) \cup (E_3 - E_2) \cup \cdots$. Hence, by disjointness,

$$\begin{aligned}\phi(E) &= \phi(E_1) + \sum_{k=2}^{\infty} \phi(E_k - E_{k-1}) \\ &= \phi(E_1) + \lim_{N\to\infty} \sum_{k=2}^{N} [\phi(E_k) - \phi(E_{k-1})] = \lim_{N\to\infty} \phi(E_N).\end{aligned}$$

On the other hand, if $E_k \searrow E$, then $\mathscr{S} - E_k \nearrow \mathscr{S} - E$. Therefore, by the case already considered, we have $\phi(\mathscr{S} - E_k) \to \phi(\mathscr{S} - E)$. Since $\phi(\mathscr{S} - E_k) = \phi(\mathscr{S}) - \phi(E_k)$ and $\phi(\mathscr{S} - E) = \phi(\mathscr{S}) - \phi(E)$, the result follows.

The next theorem is similar to Fatou's lemma (Theorem 5.17).

Theorem 10.2 *Let ϕ be a nonnegative additive set function, and let $\{E_k\}$ be any sequence of sets in Σ. Then*

$$\phi(\liminf E_k) \leq \liminf_{k\to\infty} \phi(E_k) \leq \limsup_{k\to\infty} \phi(E_k) \leq \phi(\limsup E_k).$$

Proof. The sets $H_m = \bigcap_{k=m}^{\infty} E_k$ increase to $\liminf E_k$. Therefore, by the preceding theorem, $\phi(\liminf E_k) = \lim \phi(H_m)$. Since $H_m \subset E_m$ and $\phi \geq 0$, we have $\phi(H_m) \leq \phi(E_m)$ and $\lim \phi(H_m) \leq \liminf \phi(E_m)$. Therefore, $\phi(\liminf E_k) \leq \liminf \phi(E_m)$, which proves the first inequality. The proof of the third one is similar, and the second is obvious.

If $E \in \Sigma$, the collection of sets $E \cap A$ as A ranges over Σ forms a σ-algebra Σ' of subsets of E. In fact, Σ' is just the collection of all Σ-measurable subsets of E. If ψ is an additive set function on Σ, then its restriction to Σ' is additive on Σ'. On the other hand, if ϕ is an additive set function on Σ', then the function defined by $\psi(A) = \phi(A \cap E)$ is additive on Σ.

Now, let ϕ be an additive set function on the measurable subsets of a set $E \in \Sigma$, and define

$$\overline{V}(E) = \overline{V}(E;\phi) = \sup_{\substack{A\subset E\\ A\in\Sigma}} \phi(A), \qquad \underline{V}(E) = \underline{V}(E;\phi) = -\inf_{\substack{A\subset E\\ A\in\Sigma}} \phi(A),$$

$$V(E) = V(E;\phi) = \overline{V}(E) + \underline{V}(E) \tag{10.3}$$

to be the *upper, lower*, and *total variation* of ϕ on E, respectively. Note that all three are nonnegative since $\phi(\emptyset) = 0$. Moreover, as is easy to see from the definitions,

$$-\underline{V}(E) \le \phi(A) \le \overline{V}(E) \quad \text{if } A \subset E \text{ and } A \in \Sigma,$$

and therefore, $\sup_{A\subset E, A\in\Sigma} |\phi(A)| \le V(E)$. In fact,

$$\sup_{\substack{A\subset E\\ A\in\Sigma}} |\phi(A)| \le V(E) \le 2 \sup_{\substack{A\subset E\\ A\in\Sigma}} |\phi(A)|.$$

Also, each variation is monotone increasing with E; that is, if $E_1 \subset E_2$, then $\overline{V}(E_1) \le \overline{V}(E_2)$, etc.

In case Σ is the collection of all Lebesgue measurable sets in $\mathbf{R^n}$ and $\phi(E) = \int_E f$ for some fixed $f \in L(\mathbf{R^n})$, it is easy to see that $\overline{V}(E) = \int_E f^+$, $\underline{V}(E) = \int_E f^-$, and $V(E) = \int_E |f|$; consequently, $\overline{V}$, $\underline{V}$, and V are also additive set functions. More generally, we will show that if ϕ is any additive set function on a σ-algebra Σ, then $\overline{V}$, $\underline{V}$, and V are also additive set functions on Σ. A simple corollary of the finiteness of V is that an additive set function ϕ is not only finite but also bounded. The first step in proving that the three variations are additive is the following lemma.

Lemma 10.4 *If ϕ is an additive set function on Σ, then each of its three variations is countably subadditive; that is, if $E_k \in \Sigma, k = 1, 2, \ldots$, then*

$$\overline{V}\left(\bigcup E_k\right) \le \sum \overline{V}(E_k),$$

with similar formulas for $\underline{V}$ and V.

Proof. Let $H_1 = E_1,\ H_2 = E_2 - E_1,\ H_3 = E_3 - E_2 - E_1, \ldots$. Then the H_k are disjoint and $\bigcup E_k = \bigcup H_k$. If $A \in \Sigma$ and $A \subset \bigcup E_k$, then $A = \bigcup (A \cap H_k)$ and $\phi(A) = \sum \phi(A \cap H_k)$. Therefore, since $A \cap H_k \subset E_k$, we have $\phi(A) \leq \sum \overline{V}(E_k)$. Hence,

$$\overline{V}\left(\bigcup E_k\right) = \sup_{A \subset \bigcup E_k, A \in \Sigma} \phi(A) \leq \sum \overline{V}(E_k),$$

which proves the result for $\overline{V}$. The proof for $\underline{V}$ is similar, and the result for V follows by adding.

Lemma 10.5 *If ϕ is an additive set function on Σ, then its variations $\overline{V}(E)$, $\underline{V}(E)$, and $V(E)$ are finite for every $E \in \Sigma$.*

Proof. It is enough to show the result for V. Suppose that $V(E) = +\infty$ for some E. We claim that there would then exist sets $E_k \in \Sigma, k = 1, 2, \ldots$, such that $E_k \searrow$ and both $V(E_k) = +\infty$ and $|\phi(E_k)| \geq k - 1$. To see this, we argue by induction. Let $E_1 = E$, and suppose that $E_1 \supset E_2 \supset \cdots \supset E_N$ have been constructed with $|\phi(E_k)| \geq k - 1$ and $V(E_k) = +\infty$ for $k = 1, \ldots, N$. Since $V(E_N) = +\infty$, there exists $A \in \Sigma$ such that $A \subset E_N$ and $|\phi(A)| \geq |\phi(E_N)| + N$. If $V(A) = +\infty$, let $E_{N+1} = A$, noting that $|\phi(A)| \geq N$. If $V(A) < +\infty$, let $E_{N+1} = E_N - A$. Then $V(E_{N+1}) = +\infty$ since by Lemma 10.4, we have $V(E_N) \leq V(E_{N+1}) + V(A)$. Furthermore,

$$|\phi(E_{N+1})| = |\phi(E_N) - \phi(A)| \geq |\phi(A)| - |\phi(E_N)| \geq N.$$

This establishes the existence of sets E_k with the desired properties. Thus, by Theorem 10.1, we obtain $|\phi(\bigcap E_k)| = \lim |\phi(E_k)| = +\infty$, contradicting the finiteness of ϕ and completing the proof of the lemma.

The final step in proving that the variations are additive set functions is given in the next lemma.

Lemma 10.6 *If ϕ is an additive set function on Σ and $\{E_k\}$ is a sequence of disjoint sets in Σ, then $\overline{V}(\bigcup E_k) = \sum \overline{V}(E_k)$. Similar formulas hold for $\underline{V}$ and V.*

Proof. By Lemma 10.4, we have $\overline{V}(\bigcup E_k) \leq \sum \overline{V}(E_k)$. To show the opposite inequality, given $\varepsilon > 0$, choose $A_k \subset E_k$ with $\overline{V}(E_k) \leq \phi(A_k) + \varepsilon 2^{-k}$. This is

possible since $\overline{V}(E_k)$ is finite by the previous lemma. Since the E_k are disjoint so are the A_k, and we obtain

$$\sum \overline{V}(E_k) \leq \left(\sum \phi(A_k)\right) + \varepsilon = \phi\left(\bigcup A_k\right) + \varepsilon \leq \overline{V}\left(\bigcup E_k\right) + \varepsilon.$$

Since ε is an arbitrary positive number, the result for $\overline{V}$ follows. The analogous formula for $\underline{V}$ is proved similarly, and the one for V follows by adding.

Combining Lemmas 10.5 and 10.6, we immediately obtain the next theorem.

Theorem 10.7 *If ϕ is an additive set function on Σ, then so are its variations $\overline{V}$, $\underline{V}$, and V.*

The result that follows is basic and gives a decomposition of an additive set function into the difference of two *nonnegative* additive set functions. It may be compared to Theorem 2.6.

Theorem 10.8 (Jordan Decomposition) *If ϕ is an additive set function on Σ, then*

$$\phi(E) = \overline{V}(E) - \underline{V}(E), \quad E \in \Sigma.$$

Proof. If $A \subset E$ and $A \in \Sigma$, then $\phi(E) = \phi(A) + \phi(E - A)$. Choose measurable sets $A_k \subset E$ with $\phi(A_k) \to \overline{V}(E)$ as $k \to \infty$. Then $\phi(E - A_k) \to \phi(E) - \overline{V}(E)$, and $-\underline{V}(E) \leq \phi(E) - \overline{V}(E)$ since $\phi(E - A_k) \geq -\underline{V}(E)$. If it were true that $-\underline{V}(E) < \phi(E) - \overline{V}(E)$, there would be a measurable set $B \subset E$ with $\phi(B) < \phi(E) - \overline{V}(E)$, and consequently $\phi(E - B) > \overline{V}(E)$, a contradiction. Hence, $-\underline{V}(E) = \phi(E) - \overline{V}(E)$ as claimed.

A sequence $\{E_k\}$ of sets is said to *converge* if $\limsup E_k = \liminf E_k$. Thus, $\{E_k\}$ converges if each point that belongs to infinitely many E_k belongs to all E_k from some k on. For example, if either $E_k \nearrow E$ or $E_k \searrow E$, then $\{E_k\}$ converges to E. If $\{E_k\}$ is any sequence that converges, it is said to *converge to* the set $E = \limsup E_k = \liminf E_k$.

As a simple corollary of the Jordan decomposition, we obtain the following result.

Corollary 10.9 *If a sequence of sets $\{E_k\}$ of Σ converges to E, and if ϕ is an additive set function on Σ, then $\lim_{k\to\infty} \phi(E_k) = \phi(E)$.*

Proof. If $\phi \geq 0$, we may apply Theorem 10.2. Since the extreme terms there both equal $\phi(E)$, it follows that all four equal $\phi(E)$. Hence, $\lim \phi(E_k)$ exists and equals $\phi(E)$. For arbitrary ϕ, the result therefore holds for $\overline{V}$ and $\underline{V}$, and so, by the Jordan decomposition, for ϕ itself.

Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space. We have already observed that μ satisfies $\mu(E_1) \leq \mu(E_2)$ if $E_1 \subset E_2, E_1, E_2 \in \Sigma$. Another basic property of μ is given in the next theorem.

Theorem 10.10 *Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space, and let $\{E_k : k = 1, 2, \ldots\}$ be any sequence of measurable sets. Then*

$$\mu\left(\bigcup E_k\right) \leq \sum \mu(E_k).$$

Proof. Write $\bigcup E_k$ as a disjoint union as follows:

$$\bigcup E_k = E_1 \cup (E_2 - E_1) \cup (E_3 - E_2 - E_1) \cup \cdots .$$

Then

$$\begin{aligned}\mu\left(\bigcup E_k\right) &= \mu(E_1) + \mu(E_2 - E_1) + \mu(E_3 - E_2 - E_1) + \cdots \\ &\leq \mu(E_1) + \mu(E_2) + \mu(E_3) + \cdots = \sum \mu(E_k),\end{aligned}$$

which completes the proof.

By definition, a measure is countably additive on disjoint measurable sets (cf. Theorem 3.23). The next result shows that it shares another basic property of Lebesgue measure (see Theorem 3.26).

Theorem 10.11 *Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space, and let $\{E_k\}$ be a sequence of measurable sets.*

(i) *If $E_k \nearrow E$, then $\lim_{k\to\infty} \mu(E_k) = \mu(E)$.*
(ii) *If $E_k \searrow E$ and $\mu(E_{k_0}) < +\infty$ for some k_0, then $\lim_{k\to\infty} \mu(E_k) = \mu(E)$.*

Proof. Suppose that $E_k \nearrow E$. If $\mu(E_k) < +\infty$ for all k, we may use the same argument used to prove the first part of Theorem 10.1. If $\mu(E_k) = +\infty$ for some k, then $\lim \mu(E_k) = \mu(E) = +\infty$. To prove the second part, we may assume that $k_0 = 1$ and use the argument for Theorem 10.1 with $\mathscr{S}$ replaced by E_1.

Corollary 10.12 *Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space and let $\{E_k : k = 1, 2, \ldots\}$ be a sequence of measurable sets. Then*

(i) $\mu(\liminf E_k) \leq \liminf_{k\to\infty} \mu(E_k)$.

(ii) *If $\mu(\bigcup_{k_0}^{\infty} E_k) < +\infty$ for some k_0, then $\mu(\limsup E_k) \geq \limsup_{k\to\infty} \mu(E_k)$.*

Proof. Part (ii) is an immediate corollary of Theorem 10.2. For part (i), let $A_m = \bigcap_{k=m}^{\infty} E_k, m = 1, 2, \ldots$. Then $A_m \nearrow \liminf E_k$, and by Theorem 10.11, $\mu(\liminf E_k) = \lim_{m\to\infty} \mu(A_m)$. Since $A_m \subset E_m$, we have $\mu(A_m) \leq \mu(E_m)$ and $\lim_{m\to\infty} \mu(A_m) \leq \liminf_{m\to\infty} \mu(E_m)$. The result follows by combining inequalities.

10.2 Measurable Functions and Integration

We will now develop the notions of measurable functions and integration in a measure space. These will be used later in the chapter to prove several important results for set functions.

Let Σ be a fixed σ-algebra of subsets of $\mathscr{S}$, and let $f(x)$ be a real-valued function defined for x in a measurable set E. (As usual, f may take the values $\pm\infty$.) Then f is said to be Σ-*measurable*, or simply *measurable*, if $\{x \in E : f(x) > a\}$ is measurable for $-\infty < a < +\infty$. We will state some familiar results whose proofs depend only on the fact that the class of measurable sets forms a σ-algebra. The proofs are therefore similar to those in Chapter 4 for Lebesgue measurable functions, and details are left to the reader. On the other hand, the proofs of some other results (such as the monotone convergence theorem) follow a pattern different from their analogues in Chapter 5 due to the lack of a geometric interpretation of the integral in the general setting.

Theorem 10.13

(i) *If f and g are measurable on a set $E \in \Sigma$, then so are $f + g$, cf for real c, $\phi(f)$ if ϕ is continuous on $\mathbf{R}^1$, f^+, f^-, $|f|^p$ for $p > 0$, fg, and $1/f$ if $f \neq 0$ in E.*

(ii) *If $\{f_k\}$ are measurable on a set $E \in \Sigma$, then so are $\sup_k f_k$, $\inf_k f_k$, $\limsup_{k\to\infty} f_k$, $\liminf_{k\to\infty} f_k$, and, if it exists, $\lim_{k\to\infty} f_k$.*

(iii) *If f is a simple function taking values $\{v_k\}_{k=1}^N$ on disjoint sets $\{E_k\}_{k=1}^N$, respectively, then f is measurable if and only if each E_k is measurable. In particular, χ_E is measurable if and only if E is.*

(iv) *If f is nonnegative and measurable on $E \in \Sigma$, then there exist nonnegative, simple measurable $f_k \nearrow f$ on E.*

If $(\mathscr{S}, \Sigma, \mu)$ is a measure space, a measurable set E is said to have μ-*measure zero*, or *measure zero*, if $\mu(E) = 0$. A property is said to hold *almost everywhere in E with respect to* μ, or *a.e.* (μ), if it holds in E except at most for a subset of measure zero.

We have the following analogue of Egorov's theorem.

Theorem 10.14 (Egorov's Theorem) *Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space, and let E be a measurable set with $\mu(E) < +\infty$. Let $\{f_k\}$ be a sequence of measurable functions on E such that each f_k is finite a.e. (μ) in E and $\{f_k\}$ converges a.e. (μ) in E to a finite limit. Then, given $\varepsilon > 0$, there is a measurable set $A \subset E$ with $\mu(E - A) < \varepsilon$ such that $\{f_k\}$ converges uniformly on A.*

In general, we cannot choose A to be closed in Theorem 10.14; in fact, $\mathscr{S}$ has very little structure, and the notion of a closed set may not even be defined. The proof is similar to that for Lebesgue measure and is left as an exercise.

Let f be nonnegative on a measurable set E. Define the *integral of f over E with respect to* μ by

$$\int_E f\, d\mu = \sup \sum_j [\inf_{x \in E_j} f(x)]\, \mu(E_j), \quad f \geq 0, \tag{10.15}$$

where the supremum is taken over all decompositions $E = \bigcup E_j$ of E into the union of a *finite* number of disjoint measurable sets E_j. We adopt the convention $0 \cdot \infty = \infty \cdot 0 = 0$ for the terms of the sum in (10.15). By Theorem 5.8, the definition reduces to the usual Lebesgue integral in case $\mathscr{S} = \mathbf{R^n}$, Σ is the class of Lebesgue measurable sets, μ is Lebesgue measure, and f is nonnegative and Lebesgue measurable. Although definition (10.15) does not require the measurability of f, many of the familiar properties of the integral are valid only for measurable functions. All functions considered in the rest of this section are assumed to be measurable.

Theorem 10.16 *Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space, and let f be a nonnegative, simple measurable function defined on a measurable set E. If f takes values $v_1, \ldots, v_N$ on disjoint $E_1, \ldots, E_N$, then*

$$\int_E f\, d\mu = \sum v_j\, \mu(E_j).$$

Proof. Since f is measurable, each E_j is measurable by Theorem 10.13(iii). Clearly, $\int_E f\, d\mu \geq \sum v_j \mu(E_j)$. On the other hand, consider any decomposition $E = \bigcup A_k$ of E into a finite number of disjoint measurable sets, and let $w_k = \inf_{A_k} f$. If $A_k \cap E_j$ is not empty, then $w_k \leq v_j$. Therefore, by the additivity of μ,

$$\begin{aligned}\sum w_k \mu(A_k) &= \sum_j \sum_k w_k \mu(A_k \cap E_j) \\ &\le \sum_j v_j \sum_k \mu(A_k \cap E_j) = \sum v_j \mu(E_j).\end{aligned}$$

Taking the supremum over all such decompositions gives $\int_E f\,d\mu \le \sum v_j \mu(E_j)$, which completes the proof.

Note that the previous theorem holds even if some of the v_j are $+\infty$.

Theorem 10.17 *Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space, and let f and g be measurable functions defined on a set $E \in \Sigma$.*

(i) *If $0 \le f \le g$ on E, then $\int_E f\,d\mu \le \int_E g\,d\mu$.*
(ii) *If $f \ge 0$ on E and $\mu(E) = 0$, then $\int_E f\,d\mu = 0$.*

Proof. Both parts follow immediately from the definition (10.15). For part (ii), note that $\mu(E_j) = 0$ whenever $E_j \subset E$ and $E_j \in \Sigma$. Hence, each term of the sum in (10.15) is zero.

In order to further investigate the properties of the integral, we need the next two lemmas. In these and the results that follow, the measure space $(\mathscr{S}, \Sigma, \mu)$ is fixed.

Lemma 10.18

(i) *If f and g are nonnegative, simple measurable functions on E, and if c is a nonnegative constant, then $\int_E (f+g)\,d\mu = \int_E f\,d\mu + \int_E g\,d\mu$ and $\int_E cf\,d\mu = c\int_E f\,d\mu$.*
(ii) *If f is a nonnegative, simple measurable function on E, and $E = E_1 \cup E_2$ is the union of two disjoint measurable sets, then $\int_E f\,d\mu = \int_{E_1} f\,d\mu + \int_{E_2} f\,d\mu$.*

The proof of the first part of (i) is like the first part of the proof of Theorem 5.14. The second part of (i) follows immediately from Theorem 10.16. Details and the proof of (ii) are left as an exercise.

Lemma 10.19 *Let f_k, $k = 1, 2, \ldots$, and g be nonnegative, simple measurable functions defined on a set $E \in \Sigma$. If $f_k \nearrow$ and $\lim_{k\to\infty} f_k \ge g$ on E, then*

$$\lim_{k\to\infty}\int_E f_k\,d\mu \geq \int_E g\,d\mu.$$

Moreover, the conclusion remains true if some values of g are $+\infty$.

Proof. Suppose g takes finite values $v_1,\ldots,v_m$ on disjoint sets $E_1,\ldots,E_m$. By Lemma 10.18(ii), it is enough to show that

$$\lim_{k\to\infty}\int_{E_j} f_k\,d\mu \geq \int_{E_j} g\,d\mu \quad \text{for each } j.$$

We thus reduce the proof when $g < +\infty$ to the case when g is constant on E, that is, $g = v \geq 0$ on E. If $v = 0$, the result is obvious. Suppose then that $0 < v < +\infty$, and let $0 < \varepsilon < v$ and $A_k = \{x \in E : f_k(x) \geq v - \varepsilon\}, k = 1, 2, \ldots$. Since $f_k \nearrow$, we have $A_k \nearrow E$, so that $\mu(A_k) \to \mu(E)$. Moreover,

$$\int_E f_k\,d\mu \geq \int_{A_k} f_k\,d\mu \geq (v-\varepsilon)\mu(A_k).$$

Therefore, $\lim_{k\to\infty}\int_E f_k\,d\mu \geq (v-\varepsilon)\mu(E)$. Letting $\varepsilon \to 0$ and observing that $v\mu(E) = \int_E g\,d\mu$, we obtain the desired result. We leave it to the reader to check the case when some values of g are $+\infty$.

The next theorem is helpful in deriving properties of $\int f d\mu$ for arbitrary nonnegative f from those for simple f.

Theorem 10.20 *Let $\{f_k\}$ be a sequence of nonnegative, simple measurable functions defined on $E \in \Sigma$. If $f_k \nearrow f$ on E, then $\int_E f_k\,d\mu \to \int_E f\,d\mu$.*

Proof. Clearly, $\lim_{k\to\infty}\int_E f_k\,d\mu \leq \int_E f\,d\mu$. To show the opposite inequality, consider a partition $E = \bigcup E_j$ of E into a finite number of disjoint measurable sets E_j, and let $v_j = \inf_{E_j} f$ and $\sigma = \sum v_j\mu(E_j)$. The function g defined by $g = \sum v_j\chi_{E_j}$ is nonnegative and measurable, and $\int_E g\,d\mu = \sigma$. Since $\lim_{k\to\infty} f_k \geq g$, we have $\lim_{k\to\infty}\int_E f_k\,d\mu \geq \sigma$ by Lemma 10.19. Taking the supremum of such σ over all partitions of E, we obtain the desired result.

As a corollary, we have the following theorem.

Theorem 10.21 *Let f and g be nonnegative measurable functions defined on $E \in \Sigma$, and let c be a nonnegative constant. Then*

(i) $\int_E (f+g)\,d\mu = \int_E f\,d\mu + \int_E g\,d\mu$ *and* $\int_E cf\,d\mu = c\int_E f\,d\mu$.

(ii) *If* $E = E_1 \cup E_2$, *where* E_1 *and* E_2 *are disjoint and measurable, then* $\int_E f\,d\mu = \int_{E_1} f\,d\mu + \int_{E_2} f\,d\mu$.

Proof. By Theorem 10.13(iv), choose simple measurable f_k and g_k such that $0 \le f_k \nearrow f$ and $0 \le g_k \nearrow g$. Then $f_k + g_k$ is simple and measurable, and $0 \le f_k + g_k \nearrow f + g$. Therefore, by Theorem 10.20 and Lemma 10.18,

$$\begin{aligned}\int_E (f+g)\,d\mu &= \lim_{k\to\infty} \int_E (f_k + g_k)\,d\mu = \lim_{k\to\infty}\left(\int_E f_k\,d\mu + \int_E g_k\,d\mu\right)\\ &= \int_E f\,d\mu + \int_E g\,d\mu.\end{aligned}$$

This proves the first part of (i); the other parts are proved similarly.

If f is any real-valued measurable function defined on a measurable set E, we define its *integral with respect to* μ by

$$\int_E f(x)\,d\mu(x) = \int_E f\,d\mu = \int_E f^+\,d\mu - \int_E f^-\,d\mu, \tag{10.22}$$

provided not both integrals on the right are $+\infty$.

We say that f is *integrable with respect to* μ, or μ-*integrable*, over E if $\int_E f\,d\mu$ exists and is finite. When this is the case, we write $f \in L(E; d\mu)$ or $f \in L(E;\mu)$. The abbreviations $L(d\mu)$ or $L(\mu)$ are also useful when it is clear from context what the set E is.

It is immediate from (10.22) and Theorem 10.17 that $\int_E f\,d\mu = 0$ if $\mu(E) = 0$ and that $\int_E f\,d\mu \le \int_E g\,d\mu$ if $f \le g$ on E and both integrals exist. The familiar properties of the Lebesgue integral are shared by $\int_E f\,d\mu$; some of them are listed in the following theorem.

Theorem 10.23

(i) $|\int_E f\,d\mu| \le \int_E |f|\,d\mu$; *furthermore,* $f \in L(E; d\mu)$ *if and only if* $|f| \in L(E; d\mu)$.

(ii) *If* $|f| \le |g|$ *a.e.* (μ) *in* E, *and if* $g \in L(E; d\mu)$, *then* $f \in L(E; d\mu)$, *and* $\int_E |f|\,d\mu \le \int_E |g|\,d\mu$.

(iii) *If* $f \in L(E; d\mu)$, *then* f *is finite a.e.* (μ) *in* E.

(iv) *If* $f = g$ *a.e.* (μ) *in* E *and if* $\int_E f\,d\mu$ *exists, then* $\int_E g\,d\mu$ *exists and* $\int_E g\,d\mu = \int_E f\,d\mu$.

(v) *If* $\int_E f\,d\mu$ *exists and* c *is a constant, then* $\int_E cf\,d\mu$ *exists and* $\int_E cf\,d\mu = c\int_E f\,d\mu$.

(vi) *If* $f, g \in L(E; d\mu)$, *then* $f+g \in L(E; d\mu)$ *and* $\int_E (f+g)\,d\mu = \int_E f\,d\mu + \int_E g\,d\mu$.

(vii) *If* $f \geq 0$ *and* $m \leq g \leq M$ *on* E, *then*

$$m\int_E f\,d\mu \leq \int_E fg\,d\mu \leq M\int_E f\,d\mu.$$

Proof. The proofs are similar to those for Lebesgue integrals. As examples, we will prove (iii) and (iv). For (iii), suppose that $f \in L(E; d\mu)$. Then, by (i), $|f| \in L(E; d\mu)$. Let $Z = \{x \in E : |f(x)| = +\infty\}$. Then for any positive integer k,

$$k\,\mu(Z) \leq \int_Z |f|\,d\mu \leq \int_E |f|\,d\mu.$$

Since $f \in L(E; d\mu)$, it follows that $\mu(Z) = 0$, which proves (iii).

For (iv), since both $f^+ = g^+$ and $f^- = g^-$ a.e. (μ) in E, we may assume that $f \geq 0$. Then both $\int_E f\,d\mu$ and $\int_E g\,d\mu$ clearly exist. Let $E_1 = \{x \in E : f(x) \neq g(x)\}$. Since $\mu(E_1) = 0$, Theorem 10.21(ii) implies that

$$\int_E f\,d\mu = \int_{E-E_1} f\,d\mu = \int_{E-E_1} g\,d\mu = \int_E g\,d\mu,$$

as asserted.

Theorem 10.24 *If* $\{f_k\}$ *is a sequence of nonnegative measurable functions on* E, *then*

$$\int_E \left(\sum f_k\right) d\mu = \sum \int_E f_k\,d\mu.$$

Proof. Let $f = \sum_{k=1}^{\infty} f_k$. Since $f \geq \sum_{k=1}^{m} f_k$, the integral of f over E majorizes $\sum_{k=1}^{m} \int_E f_k\,d\mu$ for any m. Hence, the left side in the preceding equation majorizes the right. To show the opposite inequality, let $\{f_k^{(j)}\}$ be a sequence of nonnegative, simple measurable functions increasing to f_k. Let $s_j = \sum_{k=1}^{j} f_k^{(j)}$. Then s_j is nonnegative and simple, and $s_j \nearrow$. We will show that $s_j \nearrow f$. Clearly, $\lim_{j\to\infty} s_j \leq f$. On the other hand, for any m,

$$\lim_{j\to\infty} s_j \geq \lim_{j\to\infty} \sum_{k=1}^{m} f_k^{(j)} = \sum_{k=1}^{m} f_k.$$

Therefore, $\lim_{j\to\infty} s_j \geq f$. It follows from Theorem 10.20 that $\int_E s_j\,d\mu \to \int_E f\,d\mu$. Since $\sum_{k=1}^{j} f_k \geq s_j$, we obtain

$$\sum_{k=1}^{\infty} \int_E f_k\,d\mu = \lim_{j\to\infty} \sum_{k=1}^{j} \int_E f_k\,d\mu \geq \lim_{j\to\infty} \int_E s_j\,d\mu = \int_E f\,d\mu.$$

This proves the desired inequality, and the theorem follows.

The next three results are essentially corollaries of Theorem 10.24.

Theorem 10.25 *If $\int_E f\,d\mu$ exists, and if $E = \bigcup E_k$ is a countable union of disjoint measurable sets E_k, then*

$$\int_E f\,d\mu = \sum \int_{E_k} f\,d\mu.$$

Proof. Suppose first that $f \geq 0$. Let $f_k = f\chi_{E_k}$ on E, so that f_k is measurable and nonnegative, and $f = \sum f_k$. By Theorem 10.24,

$$\int_E f\,d\mu = \sum \int_E f_k\,d\mu = \sum \int_{E_k} f\,d\mu.$$

For arbitrary measurable f, the existence of $\int_E f\,d\mu$ implies that of $\int_{E_k} f\,d\mu$; in fact, the integrals of f^+ and f^- over any E_k are majorized by those over E. Moreover, by the case already considered,

$$\int_E f^+\,d\mu = \sum \int_{E_k} f^+\,d\mu, \qquad \int_E f^-\,d\mu = \sum \int_{E_k} f^-\,d\mu.$$

Since at least one of these sums is finite, the conclusion follows by subtraction. (Compare Theorem 5.24.)

Theorem 10.26 *If f_k are measurable and $0 \leq f_k \nearrow f$ on E, then $\int_E f_k\,d\mu \to \int_E f\,d\mu$.*

Proof. If $\int_E f_k\,d\mu = +\infty$ for some k, the result is obvious. We may therefore assume that each $f_k \in L(d\mu)$. Write $f = f_1 + \sum_{k=2}^{\infty}(f_k - f_{k-1})$. Since each term on

the right is nonnegative, we obtain from Theorems 10.24 and 10.23(vi) that

$$\int_E f\,d\mu = \int_E f_1\,d\mu + \sum_{k=2}^{\infty}\left(\int_E f_k\,d\mu - \int_E f_{k-1}\,d\mu\right) = \lim_{k\to\infty}\int_E f_k\,d\mu.$$

Theorem 10.27 (Monotone Convergence Theorem) *Let $\{f_k\}$ and f be measurable functions on E:*

(i) *Suppose that $f_k \nearrow f$ a.e. (μ) on E. If there exists $\phi \in L(E;d\mu)$ such that $f_k \geq \phi$ on E for all k, then $\int_E f_k\,d\mu \to \int_E f\,d\mu$.*

(ii) *Suppose that $f_k \searrow f$ a.e. (μ) on E. If there exists $\phi \in L(E;d\mu)$ such that $f_k \leq \phi$ on E for all k, then $\int_E f_k\,d\mu \to \int_E f\,d\mu$.*

Proof. The proof of (i) follows by applying Theorem 10.26 to the functions $f_k - \phi$. The details are as in the proof of Theorem 5.32. Part (ii) follows by applying (i) to the functions $-f_k$.

Theorem 10.28 (Uniform Convergence Theorem) *If $f_k \in L(E;d\mu)$, $k = 1, 2, \ldots$, and $\{f_k\}$ converges uniformly to f on E, $\mu(E) < +\infty$, then $f \in L(E;d\mu)$ and $\int_E f_k\,d\mu \to \int_E f\,d\mu$.*

The proof is the same as for Lebesgue measure (see Theorem 5.33) and is omitted.

Fatou's lemma and the Lebesgue dominated convergence theorem are true for abstract measures. They are stated below without proof; the proofs are like those of Theorems 5.34 and 5.36.

Theorem 10.29 (Fatou's Lemma) *If $\{f_k\}$ is a sequence of measurable functions on E and there exists $\phi \in L(E;d\mu)$ such that $f_k \geq \phi$ a.e. (μ) on E for all k, then*

$$\int_E (\liminf_{k\to\infty} f_k)\,d\mu \leq \liminf_{k\to\infty}\int_E f_k\,d\mu.$$

The case $\phi = 0$ (i.e., $f_k \geq 0$) is of special importance. In this case, we obtain the following useful corollary.

Corollary 10.30 *Let $\{f_k\}$ and f be nonnegative measurable functions on E such that $f_k \to f$ a.e. (μ) in E. If $\int_E f_k\,d\mu \leq M$ for all k, then $\int_E f\,d\mu \leq M$.*

Theorem 10.31 (Lebesgue's Dominated Convergence Theorem) *Let ϕ, $\{f_k\}$, and f be measurable functions on E such that $|f_k| \le \phi$ a.e. (μ) on E and $\phi \in L(E; d\mu)$. Then*

(i) $\int_E (\liminf_{k\to\infty} f_k)\, d\mu \le \liminf_{k\to\infty} \int_E f_k\, d\mu \le \limsup_{k\to\infty} \int_E f_k\, d\mu$
$\le \int_E (\limsup_{k\to\infty} f_k)\, d\mu$.

(ii) *If $f_k \to f$ a.e. (μ) in E, then $\int_E f_k\, d\mu \to \int_E f\, d\mu$.*

Corollary 10.32 (Bounded Convergence Theorem) *Suppose that $\{f_k\}$ and f are measurable functions on E such that $f_k \to f$ a.e. (μ) in E. If $\mu(E) < +\infty$ and there is a constant M such that $|f_k| \le M$ a.e. (μ) in E, then $\int_E f_k\, d\mu \to \int_E f\, d\mu$.*

We conclude our brief study of integration with respect to abstract measures by defining $L^p(E; d\mu) = L^p(E, \Sigma, d\mu)$, $0 < p < \infty$, to be the collection of all measurable real or complex-valued f such that $\int_E |f|^p\, d\mu < +\infty$. We set

$$\|f\|_p = \|f\|_{p,E,d\mu} = \left(\int_E |f|^p\, d\mu \right)^{1/p}, \quad 0 < p < \infty.$$

When $p = \infty$, $L^\infty(E; d\mu)$ is defined to be the collection of all measurable f such that $\|f\|_\infty < +\infty$, where

$$\|f\|_\infty = \|f\|_{\infty,E,d\mu} = \operatorname*{ess\,sup}_E |f| = \inf\{\alpha : \mu(x \in E : |f(x)| > \alpha) = 0\}.$$

We leave it to the reader to check that $|f| \le \|f\|_\infty$ a.e. (μ) in E and that for every $\alpha < \|f\|_\infty$, there is a set $E_\alpha \subset E$ such that $\mu(E_\alpha) > 0$ and $|f| > \alpha$ on E_α.

Observe that l^p is $L^p(\mathscr{S}, \Sigma, d\mu)$ when $\mathscr{S}$ is the set of integers, Σ is the set of all subsets of $\mathscr{S}$, and $\mu(E)$ is the number of elements of E.

For $1 \le p \le \infty$, Hölder's and Minkowski's inequalities hold:

$$\|fg\|_1 \le \|f\|_p \|g\|_{p'}, \quad \|f+g\|_p \le \|f\|_p + \|g\|_p,$$

$1/p + 1/p' = 1$. Moreover, L^p is a Banach space with norm $\|\cdot\|_p$ if $1 \le p \le \infty$. In general, L^p is not separable (see Exercise 9). However, if L^2 is separable, we can define orthogonality, linear independence, completeness, Fourier coefficients, and Fourier series as usual, obtaining Bessel's inequality and Parseval's formula, as well as the usual result relating L^2 and l^2.

10.3 Absolutely Continuous and Singular Set Functions and Measures

We now turn our attention from the familiar results earlier to some new ones arising naturally in the context of abstract measure spaces.

Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space, and let ϕ be an additive set function on Σ. If $E \in \Sigma$, then ϕ is said to be *absolutely continuous on E with respect to* μ if $\phi(A) = 0$ for every measurable $A \subset E$ with $\mu(A) = 0$. Note that this definition has a somewhat different pattern from the one for Lebesgue measure (see p. 130 in Section 7.1). However, in Theorem 10.34, we shall obtain a reformulation of the present definition in terms of the old one.

On the other hand, ϕ is said to be *singular on E with respect to* μ if there is a measurable set $Z \subset E$ such that $\mu(Z) = 0$ and $\phi(A) = 0$ for every measurable $A \subset E - Z$. Thus, ϕ is singular if it is supported on a set of μ-measure zero, so that E splits into the union of two sets, Z and $E - Z$, one with μ-measure zero and the other with the property that ϕ is zero on each measurable subset of it.

As examples, note that if $f \in L(E; d\mu)$, then the function $\phi(A) = \int_A f\, d\mu$ is absolutely continuous on E with respect to μ. If Z is any measurable subset of E with $\mu(Z) = 0$ and ψ is any additive set function on the measurable subsets of E, then the function $\phi(A) = \psi(A \cap Z)$ is singular on E with respect to μ.

We list several simple properties of such set functions in the next theorem.

Theorem 10.33

(i) *If ϕ is both absolutely continuous and singular on E with respect to μ, then $\phi(A) = 0$ for every measurable $A \subset E$.*

(ii) *If both ψ and ϕ are absolutely continuous (singular) on E with respect to μ, then so are $\psi + \phi$ and $c\phi$, where c is any real constant.*

(iii) *ϕ is absolutely continuous (singular) on E with respect to μ if and only if its variations $\overline{V}$ and $\underline{V}$ are, or, equivalently, if and only if its total variation V is.*

(iv) *If $\{\phi_k\}$ is a sequence of additive set functions that are absolutely continuous (singular) on E with respect to μ, and if $\phi(A) = \lim_{k\to\infty} \phi_k(A)$ exists for every measurable $A \subset E$, then ϕ is absolutely continuous (singular) on E with respect to μ.*

Proof. For part (i), suppose that ϕ is absolutely continuous and singular on E. Let Z be a subset of E with μ-measure zero such that $\phi(H) = 0$ if H is measurable and $H \subset E - Z$. If A is any measurable subset of E, then

$\phi(A) = \phi(A \cap Z) + \phi(A - Z)$. Since ϕ is absolutely continuous and $\mu(A \cap Z) = 0$, we have $\phi(A \cap Z) = 0$. Moreover, since $A - Z \subset E - Z$ and ϕ is singular, we have $\phi(A - Z) = 0$. Hence, $\phi(A) = 0$, and part (i) is proved. The proofs of parts (ii)–(iv) are left as exercises.

The next two theorems give alternate characterizations of absolutely continuous and singular set functions.

Theorem 10.34 *An additive set function ϕ is absolutely continuous on E with respect to μ if and only if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|\phi(A)| < \varepsilon$ for any measurable $A \subset E$ with $\mu(A) < \delta$.*

Proof. The sufficiency of the condition is immediate since if $\mu(A) = 0$, then $\mu(A) < \delta$ for all $\delta > 0$, so that $|\phi(A)| < \varepsilon$ for all ε. Consequently, $\phi(A) = 0$. For the converse, suppose that ϕ is absolutely continuous, but that there is an $\varepsilon > 0$ for which no $\delta > 0$ gives the desired result. Then, taking $\delta = 2^{-k}$ for $k = 1, 2, \ldots$, there would exist measurable $A_k \subset E$ with $\mu(A_k) < 2^{-k}$ and $|\phi(A_k)| \geq \varepsilon$. Let $A = \limsup A_k$. Then, for any m,

$$\mu(A) \leq \mu\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \sum_{k=m}^{\infty} 2^{-k},$$

so that $\mu(A) = 0$. Therefore, $\phi(A) = 0$. Assuming for the moment that $\phi \geq 0$, we obtain from Theorem 10.2 that $\phi(A) = \phi(\limsup A_k) \geq \limsup \phi(A_k) \geq \varepsilon$. This contradiction establishes the result in case $\phi \geq 0$. For the general case, the variation V of an absolutely continuous ϕ is absolutely continuous (by Theorem 10.33(iii)) and nonnegative. Since $|\phi(A)| \leq V(A)$, the theorem follows.

Theorem 10.35 *An additive set function ϕ is singular on E with respect to μ if and only if given $\varepsilon > 0$, there is a measurable subset E_0 of E such that $\mu(E_0) < \varepsilon$ and $V(E - E_0; \phi) < \varepsilon$. (Recall from p. 240 in Section 10.1 that $V(E - E_0; \phi)$ is equivalent in size to $\sup_{A \subset E - E_0, A \in \Sigma} |\phi(A)|$).*

Proof. If ϕ is singular, there exists $Z \subset E$ with $\mu(Z) = 0$ such that $V(E - Z; \phi) = 0$. Taking $E_0 = Z$, we obtain the necessity of the condition. To prove its sufficiency, choose for each $k = 1, 2, \ldots$ a measurable $E_k \subset E$ with $\mu(E_k) < 2^{-k}$ and $V(E - E_k; \phi) < 2^{-k}$. Let $Z = \limsup E_k$. Since $Z \subset \bigcup_{k=m}^{\infty} E_k$ for every m, it follows as usual that $\mu(Z) = 0$. Moreover, by Theorem 10.2,

$$V(E - Z; \phi) = V(E - \limsup E_k; \phi) = V(\liminf(E - E_k); \phi)$$
$$\leq \liminf V(E - E_k; \phi) = 0.$$

Hence, ϕ is singular with respect to μ, which completes the proof.

For any measurable set E and any additive set function ϕ, the next theorem gives a useful decomposition of E in terms of the sign of ϕ. In order to motivate it, let us first consider the special case when ϕ is the indefinite integral of an $f \in L(E; d\mu)$: $\phi(A) = \int_A f\, d\mu$ for measurable $A \subset E$. Letting $P = \{x \in E : f(x) \geq 0\}$, we see that $\phi(A) \geq 0$ for any measurable $A \subset P$ and that $\phi(A) \leq 0$ for any measurable $A \subset E - P$. It follows that $\overline{V}(E; \phi) = \overline{V}(P; \phi) = \phi(P)$ and $\underline{V}(E; \phi) = \underline{V}(E - P; \phi) = -\phi(E - P)$. As a simple consequence, since $P = \{x \in E : f(x) = f^+(x)\}$, we have

$$\overline{V}(E; \phi) = \int_E f^+\, d\mu, \quad \underline{V}(E; \phi) = \int_E f^-\, d\mu.$$

The splitting $E = P \cup (E - P)$ is the sort of decomposition of E that we have in mind. For an arbitrary ϕ, there is the following basic result.

Theorem 10.36 (Hahn Decomposition) *Let E be a measurable set and let ϕ be an additive set function defined on the measurable subsets A of E. Then there is a measurable $P \subset E$ such that $\phi(A) \geq 0$ for $A \subset P$ and $\phi(A) \leq 0$ for $A \subset E - P$. Equivalently, $\underline{V}(P; \phi) = \overline{V}(E - P; \phi) = 0$. Hence,*

$$\overline{V}(E; \phi) = \overline{V}(P; \phi) = \phi(P),$$

$$\underline{V}(E; \phi) = \underline{V}(E - P; \phi) = -\phi(E - P).$$

Proof. Denote $\overline{V}(A) = \overline{V}(A; \phi)$ and $\underline{V}(A) = \underline{V}(A; \phi)$ for measurable sets $A \subset E$. For each positive integer k, choose a measurable $A_k \subset E$ such that $\phi(A_k) > \overline{V}(E) - 2^{-k}$. Then $\overline{V}(A_k) > \overline{V}(E) - 2^{-k}$. Since $\overline{V}$ is additive, $\overline{V}(E - A_k) = \overline{V}(E) - \overline{V}(A_k) < 2^{-k}$. Moreover, by the Jordan decomposition (Theorem 10.8), $\overline{V}(A_k) - \underline{V}(A_k) = \phi(A_k) > \overline{V}(E) - 2^{-k}$, and therefore $\underline{V}(A_k) < 2^{-k}$. Let $P = \liminf A_k$. Since $\underline{V}$ is nonnegative, Theorem 10.2 implies that $\underline{V}(P) \leq \liminf \underline{V}(A_k) = 0$. Also,

$$\overline{V}(E - P) = \overline{V}(E - \liminf A_k) = \overline{V}(\limsup(E - A_k)) \leq \overline{V}\Big(\bigcup_{k=m}^{\infty}(E - A_k)\Big)$$

for any m. Therefore, for any m, by using Lemma 10.4, we obtain

$$\overline{V}(E-P) \leq \sum_{k=m}^{\infty} \overline{V}(E-A_k) < \sum_{k=m}^{\infty} 2^{-k},$$

which gives $\overline{V}(E-P) = 0$ and completes the proof.

In the next theorem, we use the Hahn decomposition to split E into sets where ϕ is comparable to μ. In doing so, we assume that ϕ is nonnegative and μ is finite; thus, we are in fact dealing with two finite measures.

Theorem 10.37 *Let ϕ be a nonnegative additive set function defined on the measurable subsets of a measurable set E, and let μ be a measure with $\mu(E) < +\infty$. Then given $a > 0$, there is a decomposition $E = Z \cup \left(\bigcup_{k=1}^{\infty} E_k\right)$ of E into disjoint measurable sets such that*

(i) $\mu(Z) = 0$

(ii) $a(k-1)\mu(A) \leq \phi(A) \leq ak\mu(A)$ *for measurable* $A \subset E_k$, $k = 1, 2, \ldots$

Proof. We may assume that $a = 1$ by considering ϕ/a. For each positive integer k, let $\psi_k(A) = \phi(A) - k\mu(A)$ for measurable $A \subset E$. Since ϕ and μ are finite and additive, ψ_k is an additive set function. By the Hahn decomposition, there is a set $P_k \subset E$ such that $\psi_k(A) \geq 0$ if $A \subset P_k$ and $\psi_k(A) \leq 0$ if $A \subset E - P_k$. Thus, $\phi(A) \geq k\mu(A)$ if $A \subset P_k$ and $\phi(A) \leq k\mu(A)$ if $A \subset E - P_k$.

Now, let $Q_k = \bigcup_{m=k}^{\infty} P_m$ for $k = 1, 2, \ldots$, and observe that $P_k \subset Q_k$ and $Q_k \searrow$. We will show that $\phi(A) \geq k\mu(A)$ if $A \subset Q_k$, and $\phi(A) \leq k\mu(A)$ if $A \subset E - Q_k$. To see this, write

$$Q_k = P_k \cup (P_{k+1} - P_k) \cup (P_{k+2} - P_{k+1} - P_k) \cup \cdots$$

and note that the terms on the right side are disjoint. Hence, if $A \subset Q_k$, we may write $A = \bigcup_{m=k}^{\infty} A_m$, where the A_m are disjoint and $A_m \subset P_m$, by simply intersecting A with each such term of Q_k. Then

$$\phi(A) = \sum_{m=k}^{\infty} \phi(A_m) \geq \sum_{m=k}^{\infty} m\mu(A_m) \geq k \sum_{m=k}^{\infty} \mu(A_m) = k\mu(A),$$

so that $\phi(A) \geq k\mu(A)$ for $A \subset Q_k$, as claimed. On the other hand, if $A \subset E - Q_k$, then $A \subset E - P_k$, so that $\phi(A) \leq k\mu(A)$. This proves the assertion above.

We can now give the decomposition of E. Let $Z = \bigcap_{k=1}^{\infty} Q_k = \limsup P_k$, and write

$$\begin{aligned} E &= Z \cup (E - Q_1) \cup (Q_1 - Q_2) \cup (Q_2 - Q_3) \cup \cdots \\ &= Z \cup E_1 \cup E_2 \cup E_3 \cup \cdots . \end{aligned}$$

The terms in this decomposition are disjoint. If $A \subset E_1 (= E - Q_1)$, then $\phi(A) \leq \mu(A)$ by what was shown earlier, and $\phi(A) \geq 0$ by hypothesis. For $k \geq 2$, we have $E_k = Q_{k-1} - Q_k = Q_{k-1} \cap (E - Q_k)$. Hence, if $k \geq 2$ and $A \subset E_k$, then $\phi(A) \geq (k-1)\,\mu(A)$ due to the fact that $A \subset Q_{k-1}$; also, $\phi(A) \leq k\mu(A)$ due to $A \subset E - Q_k$. Finally, since $Z \subset Q_k$ for all k, we have $\phi(Z) \geq k\mu(Z)$ for all k. Since ϕ is finite, it follows that $\mu(Z) = 0$, which completes the proof.

To give some idea of the significance of the last result, write $A = (A \cap Z) \cup [\bigcup_k (A \cap E_k)]$ for measurable $A \subset E$. Then

$$\phi(A) = \phi(A \cap Z) + \sum \phi(A \cap E_k).$$

The set function $\sigma(A) = \phi\,(A \cap Z)$ is singular with respect to μ. By (ii) of the theorem, ϕ is absolutely continuous with respect to μ on each E_k. Hence, the set function α defined by

$$\alpha(A) = \phi(A) - \sigma(A) = \sum \phi(A \cap E_k)$$

is absolutely continuous with respect to μ since if $\mu(A) = 0$, then $\mu(A \cap E_k) = 0$ and $\phi(A \cap E_k) = 0$ for all k. Note also that (ii) can be written

$$a(k-1) \int_A d\mu \leq \phi(A) \leq ak \int_A d\mu$$

for measurable $A \subset E_k$.

We will now use these ideas to decompose any set function into the sum of an absolutely continuous part, which will be an indefinite integral, and a singular part. This decomposition, which is of major importance, is stated in the following theorem. We assume that the measure μ defined on the measurable subsets of E is σ-*finite*, that is, that E can be written as a countable union of measurable sets with finite μ-measure.

Theorem 10.38 (Lebesgue Decomposition) *Let ϕ be an additive set function on the measurable subsets of a measurable set E, and let μ be a σ-finite measure on E. Then there is a unique decomposition*

$$\phi(A) = \alpha(A) + \sigma(A) \textit{ for measurable } A \subset E,$$

where α and σ are additive set functions, α is absolutely continuous with respect to μ, and σ is singular with respect to μ. These functions are

$$\alpha(A) = \int_A f\,d\mu, \qquad \sigma(A) = \phi(A \cap Z)$$

for appropriate $f \in L(E; d\mu)$ and Z with $\mu(Z) = 0$. Moreover, if $\phi \geq 0$, then $f \geq 0$.

Proof. Assuming that such a decomposition exists, we will show it is unique. If $\phi = \alpha_1 + \sigma_1$ is another decomposition of ϕ into absolutely continuous and singular parts, then $\alpha - \alpha_1 = \sigma_1 - \sigma$, which (being both absolutely continuous and singular with respect to μ) must vanish identically. Hence $\alpha = \alpha_1$ and $\sigma = \sigma_1$.

To show that the decomposition exists, first assume that $\phi \geq 0$ and $\mu(E) < +\infty$. Taking $a = 2^{-m}$, $m = 1, 2, \ldots$, in Theorem 10.37, we may write E as a disjoint union $E = Z^{(m)} \cup \left(\bigcup_k E_k^{(m)}\right)$, where

$$\mu(Z^{(m)}) = 0 \text{ and } 2^{-m}(k-1)\mu(A) \leq \phi(A) \leq 2^{-m}k\mu(A) \text{ if } A \subset E_k^{(m)}.$$

Given m, k, m', and k', let $\beta = 2^{-m}(k-1)$, $\gamma = 2^{-m}k$, $\beta' = 2^{-m'}(k'-1)$, and $\gamma' = 2^{-m'}k'$. If the intervals $[\beta, \gamma]$ and $[\beta', \gamma']$ are disjoint, we will show that the set $A = E_k^{(m)} \cap E_{k'}^{(m')}$ has μ-measure zero. In fact, we have both

$$\beta\mu(A) \leq \phi(A) \leq \gamma\mu(A) \quad \text{and} \quad \beta'\mu(A) \leq \phi(A) \leq \gamma'\mu(A).$$

If, for example, $\gamma < \beta'$, the inequalities $\beta'\mu(A) \leq \phi(A) \leq \gamma\mu(A)$ imply that $\mu(A) = 0$. A similar argument applies if $\gamma' < \beta$. Fixing m and k, and setting $m' = m + 1$, we see that there are at most four values of k' such that $E_{k'}^{(m+1)}$ intersects $E_k^{(m)}$ in a set of positive μ-measure, namely, $k' = 2k-2$, $2k-1$, $2k$, and $2k+1$. Hence,

$$E_k^{(m)} \subset E_{2k-2}^{(m+1)} \cup E_{2k-1}^{(m+1)} \cup E_{2k}^{(m+1)} \cup E_{2k+1}^{(m+1)} \cup Y_k^{(m)},$$

where $\mu(Y_k^{(m)}) = 0$.

Let

$$Z = \left(\bigcup_m Z^{(m)}\right) \cup \left(\bigcup_{k,m} Y_k^{(m)}\right),$$

so that $\mu(Z) = 0$, and define functions $\{f_m\}_{m=1}^{\infty}$ on E by $f_m(x) = 2^{-m}(k-1)$ if $x \in E_k^{(m)} - Z$ and $f_m(x) = 0$ if $x \in Z$. Therefore, if $x \in E_k^{(m)} - Z$, then

$f_m(x) = 2^{-m}(k-1)$ and $f_{m+1}(x)$ takes one of the four values $2^{-m-1}j, j = 2k-3, 2k-2, 2k-1, 2k$. Hence, $|f_m(x) - f_{m+1}(x)| \le 2^{-m}$ if $x \in E_k^{(m)} - Z$, and so also if $x \in E$. It follows that $\{f_m\}$ converges uniformly on E to a limit f. Since $f_m \ge 0$, also $f \ge 0$.

Since E is the disjoint union $Z \cup \bigcup_k \left(E_k^{(m)} - Z\right)$ and ϕ is absolutely continuous on each $E_k^{(m)}$,

$$\begin{aligned}\phi(A) &= \phi(A \cap Z) + \sum_k \phi\left(A \cap \left(E_k^{(m)} - Z\right)\right)\\ &= \phi(A \cap Z) + \sum_k \phi\left(A \cap E_k^{(m)}\right)\end{aligned}$$

for measurable $A \subset E$. Therefore,

$$\begin{aligned}\phi(A \cap Z) + \sum_k 2^{-m}(k-1)\mu\left(A \cap E_k^{(m)}\right) &\le \phi(A)\\ \le \phi(A \cap Z) + \sum_k 2^{-m}k\mu\left(A \cap E_k^{(m)}\right),\end{aligned}$$

which can be rewritten

$$\phi(A \cap Z) + \int_A f_m \, d\mu \le \phi(A) \le \phi(A \cap Z) + \int_A f_m \, d\mu + 2^{-m}\mu(A).$$

Since $\mu(A)$ is finite, we obtain from the uniform convergence theorem that $\int_A f_m \, d\mu \to \int_A f \, d\mu$. Therefore,

$$\phi(A) = \phi(A \cap Z) + \int_A f \, d\mu,$$

which proves the theorem in case $\phi \ge 0$ and $\mu(E) < +\infty$.

If $\phi \ge 0$ and $\mu(E) = +\infty$, then E can still be written as a disjoint union $E = \bigcup E_j$ with $\mu(E_j) < +\infty$, since E is σ-finite. Hence, there exist $Z_j \subset E_j$, $\mu(Z_j) = 0$, and nonnegative f_j on E_j such that for all measurable $A \subset E$,

$$\phi(A \cap E_j) = \phi(A \cap Z_j) + \int_{A \cap E_j} f_j \, d\mu.$$

Letting $Z = \bigcup Z_j$ and $f = \sum f_j \chi_{E_j}$, we obtain $\mu(Z) = 0$, $f \geq 0$, and

$$\phi(A) = \sum \phi(A \cap E_j) = \sum \phi(A \cap Z_j) + \sum \int_{A \cap E_j} f_j \, d\mu = \phi(A \cap Z) + \int_A f \, d\mu.$$

Of course, f is integrable since ϕ is finite. The proof is now complete if $\phi \geq 0$.

For an arbitrary ϕ, apply the decomposition to each of $\overline{V}$ and $\underline{V}$, and subtract the results. By the Jordan decomposition, we obtain $\phi(A) = \int_A f \, d\mu + \sigma(A)$, where $f \in L(E; d\mu)$ and σ is singular with respect to μ. It remains to show that there is a set Z, $\mu(Z) = 0$, such that $\sigma(A) = \phi(A \cap Z)$. Let Z be the set of μ-measure zero corresponding to σ in the definition of a singular set function. Then $\sigma(A \cap Z) = \sigma(A)$ and $\int_{A \cap Z} f \, d\mu + \sigma(A) = 0$. Hence, replacing A by $A \cap Z$ in the formula $\phi(A) = \int_A f \, d\mu + \sigma(A)$, we obtain $\sigma(A) = \phi(A \cap Z)$. This completes the proof. For a result concerning the uniqueness of f, see Exercise 6(a).

We have already noted that the indefinite integral of an integrable function is absolutely continuous. The following fundamental result gives a converse: namely, in a σ-finite space, the only absolutely continuous set functions are indefinite integrals.

Theorem 10.39 (Radon–Nikodym) *Let ϕ be an additive set function on the measurable subsets of a measurable E, and let μ be a σ-finite measure on E. If ϕ is absolutely continuous with respect to μ, there exists a unique $f \in L(E; d\mu)$ such that*

$$\phi(A) = \int_A f \, d\mu$$

for every measurable $A \subset E$.

Here, the function f is called *the Radon–Nikodym derivative of ϕ with respect to μ.*

Proof. The result follows from the Lebesgue decomposition. In fact, $\phi(A) = \int_A f \, d\mu + \phi(A \cap Z)$ for appropriate $f \in L(E; d\mu)$ and Z with $\mu(Z) = 0$. Since ϕ is absolutely continuous, we have $\phi(A \cap Z) = 0$, so that $\phi(A) = \int_A f \, d\mu$. For the uniqueness of f, see Exercise 6(a).

If ν and μ are two measures defined on the same family of measurable sets, we say that ν is *absolutely continuous with respect to* μ on a measurable

set E if $\nu(A) = 0$ for every $A \subset E$ with $\mu(A) = 0$. If ν is finite, Theorem 10.34 implies that a necessary and sufficient condition for ν to be absolutely continuous with respect to μ is that given $\varepsilon > 0$, there exist $\delta > 0$ such that $\nu(A) < \varepsilon$ if $\mu(A) < \delta$. The necessity of this condition may fail if ν is not finite; see Exercise 12.

We say that two measures ν and μ are *mutually singular* on E if E can be written as a disjoint union, $E = E_1 \cup E_2$, of two measurable sets with $\nu(E_1) = \mu(E_2) = 0$. The reader can check that the following analogue of Theorem 10.35 is valid: two measures ν and μ are mutually singular on E if and only if given $\varepsilon > 0$, there are disjoint measurable E_1 and E_2 with $E = E_1 \cup E_2$ and $\nu(E_1) < \varepsilon, \mu(E_2) < \varepsilon$.

We also note that if ν and μ are mutually singular on E and if $g \in L(E; d\nu)$, then the set function $\int_A g\, d\nu$ is singular with respect to μ. To see this, write $E = E_1 \cup E_2$, where E_1 and E_2 are disjoint with $\nu(E_1) = \mu(E_2) = 0$. Setting $Z = E_2$, we have $\mu(Z) = 0$ and $\nu(A) = 0$ for every measurable $A \subset E - Z = E_1$. Hence, $\int_A g\, d\nu = 0$ for such A, which proves the assertion.

We have the following analogue of the Lebesgue decomposition.

Theorem 10.40 *Let ν and μ be two σ-finite measures defined on the measurable subsets of a measurable E. Then there is a unique, nonnegative measurable f on E and a unique measure σ on the measurable subsets of E such that σ and μ are mutually singular on E and $\nu(A) = \int_A f\, d\mu + \sigma(A)$ for every measurable $A \subset E$. Moreover,*

$$\int_A g\, d\nu = \int_A gf\, d\mu + \int_A g\, d\sigma$$

whenever $\int_A g\, d\nu$ exists.

Before giving the proof, we add several remarks based on the theorem. First, $\int_A f\, d\mu$ is an absolutely continuous measure with respect to μ since $f \geq 0$. Next, if $E = Z \cup (E - Z)$, where $\mu(Z) = \sigma(E - Z) = 0$, then σ has the form

$$\sigma(A) = \nu(A \cap Z) \quad \text{for measurable } A \subset E,$$

as can be seen by replacing A by $A \cap Z$ in the decomposition of ν and observing that $\sigma(A) = \sigma(A \cap Z) + \sigma(A - Z) = \sigma(A \cap Z)$. Note also that if $\nu(E)$ is finite, then $f \in L(E; d\mu)$. Finally, if $g \in L(E; d\nu)$, then $g \in L(E; d\sigma)$ since $\sigma(A) \leq \nu(A)$ for all measurable $A \subset E$. In this case, the second formula in the theorem implies that $gf \in L(E; d\mu)$ and expresses the Lebesgue decomposition of $\int_A g\, d\nu$ with respect to μ.

Proof. If $\nu(E) < +\infty$, the Lebesgue decomposition implies that $\nu(A) = \int_A f\, d\mu + \sigma(A)$ for measurable $A \subset E$, where $f \geq 0, f \in L(E; d\mu)$, and σ

and μ are mutually singular. If $\nu(E) = +\infty$, then since ν is σ-finite, we have $E = \bigcup E_j$ with E_j disjoint and $\nu(E_j) < +\infty$. Choose $Z_j \subset E_j$ and f_j on E_j such that $\mu(Z_j) = 0$, $f_j \geq 0$ and

$$\nu(A \cap E_j) = \int_{A \cap E_j} f_j \, d\mu + \nu(A \cap Z_j) \quad \text{for measurable } A \subset E.$$

Let $Z = \bigcup Z_j$ and $f = \sum f_j \chi_{E_j}$. Then $\mu(Z) = 0$, and adding over j, we have

$$\nu(A) = \int_A f \, d\mu + \nu(A \cap Z) = \int_A f \, d\mu + \sigma(A),$$

as claimed. The proof of the uniqueness of f and σ is left as an exercise.

If g is the characteristic function χ_B of a measurable set B, the formula in question, namely, $\int_A g \, d\nu = \int_A gf \, d\mu + \int_A g \, d\sigma$, reduces to $\nu(A \cap B) = \int_{A \cap B} f \, d\mu + \sigma(A \cap B)$, which we know to be valid. Hence, the formula is also valid for any simple measurable g and, therefore, by the monotone convergence theorem, for any measurable $g \geq 0$. Now let g be any measurable function for which $\int_A g \, d\nu$ exists. Then at least one of $\int_A g^+ \, d\nu$ and $\int_A g^- \, d\nu$ is finite, and the formula for g follows by subtracting those for g^+ and g^-.

Corollary 10.41 *Let ν and μ be two σ-finite measures defined on the measurable subsets of a measurable E.*

(i) *Then ν is absolutely continuous with respect to μ on E if and only if there is a nonnegative measurable f such that $\nu(A) = \int_A f \, d\mu$ for every measurable $A \subset E$. In this case,*

$$\int_A g \, d\nu = \int_A gf \, d\mu$$

for any measurable g and $A \subset E$ for which $\int_A g \, d\nu$ exists.

(ii) *Let $g \in L(E; d\nu)$. Then $\int_A g \, d\nu = \int_A gf \, d\mu$ for some nonnegative f and all measurable $A \subset E$ if and only if $\int_A g \, d\nu$ is absolutely continuous with respect to μ.*

Proof. Let $\nu(A) = \int_A f \, d\mu + \sigma(A)$ be the decomposition given by Theorem 10.40. Part (i) follows from Theorem 10.40 since $\sigma \equiv 0$ if and only if ν is absolutely continuous with respect to μ. Part (ii) follows from the fact that the formula $\int_A g \, d\nu = \int_A gf \, d\mu + \int_A g \, d\sigma$ is the Lebesgue decomposition of $\int_A g \, d\nu$.

10.4 The Dual Space of L^p

If B is a Banach space (or, more generally, a normed linear space) over the real numbers, a real-valued *linear functional* l on B is by definition a finite real-valued function $l(f)$, $f \in B$, which satisfies

$$l(f_1 + f_2) = l(f_1) + l(f_2), \quad l(\alpha f) = \alpha\, l(f), \ -\infty < \alpha < +\infty.$$

Note that $l(0) = 0$.

A linear functional l is said to be *bounded* if there is a constant c such that $|l(f)| \le c\,\|f\|$ for all $f \in B$. A bounded linear functional l is continuous with respect to the norm in B, by which we mean that if $\|f - f_k\| \to 0$ as $k \to \infty$, then $l(f_k) \to l(f)$, since $|l(f) - l(f_k)| = |l(f - f_k)| \le c\,\|f - f_k\| \to 0$.

The *norm* $\|l\|$ *of a bounded linear functional* l is defined as

$$\|l\| = \sup_{\|f\| \le 1} |l(f)|. \tag{10.42}$$

Since $f/\|f\|$ has norm 1 for any $f \neq 0$, and since l is linear, we have $\|l\| = \sup_{f \neq 0} |l(f)|/\|f\|$.

The collection of all bounded linear functionals on B is called the *dual space* B' of B. We shall consider the case when $B = L^p = L^p(E; d\mu)$ and for simplicity restrict our attention to real-valued functions. Our goal is to show that if $1 \le p < \infty$ and μ is σ-finite, then the dual space $(L^p)'$ of L^p can be identified in a natural way with $L^{p'}$, $1/p + 1/p' = 1$. The main tool in doing so is the Radon–Nikodym theorem. The first result is the following.

Theorem 10.43 *Let $1 \le p \le \infty$, $1/p + 1/p' = 1$. If $g \in L^{p'}(E; d\mu)$, then the formula*

$$l(f) = \int_E fg\, d\mu$$

defines a bounded linear functional $l \in [L^p(E; d\mu)]'$. Moreover, $\|l\| \le \|g\|_{p'}$.

Proof. This follows easily from Hölder's inequality and the linear properties of the integral: we have

$$|l(f)| = \left| \int_E fg\, d\mu \right| \le \|g\|_{p'} \|f\|_p,$$

and therefore $\|l\| \le \|g\|_{p'}$.

The theorem shows that with each $g \in L^{p'}$, we can associate a bounded linear functional, $l(f) = \int_E fg\,d\mu$, on L^p. If μ is σ-finite on E, the correspondence between g and l is unique (see Exercise 6) and defines an embedding of $L^{p'}$ in $(L^p)'$. We now give the characterization of $(L^p)'$, $1 \le p < \infty$.

Theorem 10.44 *Let $1 \le p < \infty$, $1/p + 1/p' = 1$, and let μ be σ-finite. If $l \in [L^p(E; d\mu)]'$, there is a unique $g \in L^{p'}(E; d\mu)$ such that*

$$l(f) = \int_E fg\,d\mu.$$

Moreover, $\|l\| = \|g\|_{p'}$, and therefore the correspondence between l and g defines an isometry between $(L^p)'$ and $L^{p'}$.

Proof. Suppose first that $\mu(E) < +\infty$. Let $l \in (L^p)'$ and write $\|l\| = c$. Define a set function ϕ on the measurable sets $A \subset E$ by

$$\phi(A) = l(\chi_A).$$

Note that ϕ is finite; in fact, $|\phi(A)| \le c\|\chi_A\|_p = c\mu(A)^{1/p}$. Clearly, ϕ is finitely additive. To show that it is countably additive, suppose that $A = \bigcup_{k=1}^{\infty} A_k$, A_k measurable and disjoint. Write $A = (\bigcup_{k=1}^{m} A_k) \cup (\bigcup_{k=m+1}^{\infty} A_k) = A' \cup A''$. Then

$$\phi(A) = \phi(A') + \phi(A'') = \sum_{k=1}^{m} \phi(A_k) + \phi(A'').$$

Since $|\phi(A'')| \le c\mu(A'')^{1/p}$ (with $p \ne \infty$), Theorem 10.11 implies that $\phi(A'')$ tends to zero as $m \to \infty$. Hence, $\phi(A) = \sum_{k=1}^{\infty} \phi(A_k)$, which shows that ϕ is countably additive. The fact that $|\phi(A)| \le c\mu(A)^{1/p}$ also implies that ϕ is absolutely continuous with respect to μ.

By the Radon–Nikodym theorem, there is a $g \in L^1(E; d\mu)$ such that $\phi(A) = \int_A g\,d\mu$ for measurable $A \subset E$. This means that $l(\chi_A) = \int_E \chi_A g\,d\mu$, so that $l(f) = \int_E fg\,d\mu$ for any simple measurable f. To show the same formula holds for any $f \in L^p$, we first claim that $g \in L^{p'}$ and $\|g\|_{p'} \le c$. If $p > 1$, choose simple functions h_k with $0 \le h_k \nearrow |g|^{p'}$. Let $\{g_k\}$ be the simple functions defined by

$$g_k = h_k^{1/p} \operatorname{sign} g.$$

Then $\|g_k\|_p = \|h_k\|_1^{1/p}$, and

$$\int_E g_k g \, d\mu = l(g_k) \le c \, \|g_k\|_p = c \, \|h_k\|_1^{1/p} .$$

Since $g_k g = h_k^{1/p}|g| \ge h_k^{1/p+1/p'} = h_k$, we obtain $\|h_k\|_1 \le c\|h_k\|_1^{1/p}$. We may assume that $\|h_k\|_1 \ne 0$ for large k. (Otherwise, g would be zero a.e. (μ), and our claim would be obviously true.) Hence, dividing both sides of the last inequality by $\|h_k\|_1^{1/p}$, we have $\|h_k\|_1^{1/p'} \le c$, so that $\|g\|_{p'} \le c$ by the monotone convergence theorem. This proves the claim when $p > 1$. The case $p = 1$ is left as an exercise.

To show that $l(f) = \int_E fg \, d\mu$ for any $f \in L^p$, choose simple f_k converging to f in L^p norm (see Exercise 8). Then $l(f_k) \to l(f)$, and $\int_E f_k g \, d\mu \to \int_E fg \, d\mu$ by Hölder's inequality:

$$\left| \int_E f_k g \, d\mu - \int_E fg \, d\mu \right| \le \int_E |f_k - f| \, |g| \, d\mu \le \|f_k - f\|_p \, \|g\|_{p'} .$$

The fact that the formula holds for f_k thus implies that it holds for f by passing to the limit.

To complete the proof for the case $\mu(E) < +\infty$, it remains to show that $\|g\|_{p'} = c$ and that the correspondence between l and g is unique. However, we already know that $\|g\|_{p'} \le c$, and the opposite inequality follows from Theorem 10.43. For the uniqueness of the correspondence, see Exercise 6(c).

If $\mu(E) = +\infty$, then since μ is σ-finite, there exist $E_j \nearrow E$ with $\mu(E_j) < +\infty$. Let $l \in [L^p(E)]'$. We may view functions in $L^p(E_j)$ as those functions in $L^p(E)$ that vanish outside E_j. Since the restriction of l to $L^p(E_j)$ is a bounded linear functional, there is a unique $g_j \in L^{p'}(E_j)$, $\|g_j\|_{p',E_j} \le \|l\|$, such that

$$l(f) = \int_{E_j} fg_j \, d\mu$$

for every f in L^p that vanishes outside E_j. For such f, the fact that $E_j \subset E_{j+1}$ also gives

$$l(f) = \int_{E_{j+1}} fg_{j+1} \, d\mu = \int_{E_j} fg_{j+1} \, d\mu.$$

Therefore, $g_{j+1} = g_j$ a.e. (μ) in E_j. We may assume that $g_{j+1} = g_j$ everywhere in E_j. Define $g(x) = g_j(x)$ if $x \in E_j$. Then g is measurable and it follows that $\|g\|_{p'} \leq \|l\|$. If $f \in L^p(E)$, then

$$l(f\chi_{E_j}) = \int_{E_j} fg_j \, d\mu = \int_{E_j} fg \, d\mu.$$

Since $f\chi_{E_j}$ converges in L^p to f and $\int_{E_j} fg \, d\mu \to \int_E fg \, d\mu$ (note that $fg \in L^1$ by Hölder's inequality), we obtain $l(f) = \int_E fg \, d\mu$ in the limit. Therefore, by Theorem 10.43, $\|l\| \leq \|g\|_{p'}$, so that $\|l\| = \|g\|_{p'}$ and the proof is complete.

We remark that the proof of Theorem 10.44 fails in several places if $p = \infty$, for example, at the place where we conclude that ϕ is absolutely continuous. In fact, the theorem itself is false when $p = \infty$, that is, not every bounded linear functional on L^∞ can be represented $l(f) = \int fg \, d\mu$ for some $g \in L^1$; an example is indicated in Exercise 18.

10.5 Relative Differentiation of Measures

Lebesgue's differentiation theorem, Theorem 7.11, states that if f is locally integrable in $\mathbf{R^n}$, then

$$\lim_{h \to 0} \frac{1}{|Q_{\mathbf{x}}(h)|} \int_{Q_{\mathbf{x}}(h)} f(\mathbf{y}) \, d\mathbf{y} = f(\mathbf{x}) \text{ a.e.},$$

where $Q_{\mathbf{x}}(h)$ is the cube with center $\mathbf{x}$ and edge length h. We will now study an analogue of this result for other measures on $\mathbf{R^n}$. Specifically, if μ and ν are two σ-finite measures on the Borel subsets of $\mathbf{R^n}$, we will study the existence of

$$\lim_{h \to 0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))}$$

and its relation to the Lebesgue decomposition of ν with respect to μ.

We will follow the method used to prove Lebesgue's differentiation theorem. To do this, we must find a replacement for Vitali's lemma: the simple form of Vitali's lemma (see Lemma 7.4) relies heavily on the fact that expanding a cube concentrically by a factor (say 5) only enlarges its Lebesgue

measure proportionally, whereas no such relation may hold for general measures. In order to bypass this difficulty, we shall present a covering lemma that is purely geometric in nature, that is, which makes no mention of measure.

We consider only cubes whose edges are parallel to the coordinate axes and write $Q = Q_{\mathbf{x}}$ for those with center $\mathbf{x}$. We say that a family K of cubes has *bounded overlaps* if there is a constant c such that every $\mathbf{x} \in \mathbf{R^n}$ belongs to at most c cubes from K. Thus, K has bounded overlaps if and only if

$$\sum_{Q \in K} \chi_Q(\mathbf{x}) \leq c \quad \text{for all } \mathbf{x} \in \mathbf{R^n}.$$

Theorem 10.45 (Besicovitch Covering Lemma) *Let E be a bounded subset of $\mathbf{R^n}$ and let K be a family of cubes covering E that contains a cube $Q_{\mathbf{x}}$ with center $\mathbf{x}$ for each $\mathbf{x} \in E$. Then there exist points $\{\mathbf{x}_k\}$ in E such that*

(i) $E \subset \bigcup Q_{\mathbf{x}_k}$
(ii) $\{Q_{\mathbf{x}_k}\}$ *has bounded overlaps*

Moreover, the constant c for which $\sum \chi_{Q_{\mathbf{x}_k}} \leq c$ can be chosen to depend only on n.

In order to prove this, it will be convenient to first prove the following lemma.

Lemma 10.46 *Let $\{Q_k\}_{k=1}^{\infty}$ be a sequence of cubes with centers $\{\mathbf{x}_k\}$ such that if $j < k$, then $\mathbf{x}_k \notin Q_j$ and $|Q_k| \leq 2|Q_j|$. Then $\{Q_k\}$ has bounded overlaps, and the constant c for which $\sum \chi_{Q_k} \leq c$ can be chosen to depend only on n.*

Proof. We will consider only $n = 2$; the case $n \neq 2$ is similar and left as an exercise. Let $Q_{k_m}, m = 1, 2, \ldots,$ be those Q_k that contain the origin and whose centers are in the first quadrant, and let h_m denote the edge length of Q_{k_m}. Then Q_{k_1} covers at least the region

$$A = \left\{(x, y) : 0 \leq x \leq \tfrac{1}{2}h_1, 0 \leq y \leq \tfrac{1}{2}h_1\right\}.$$

Hence, no Q_{k_m} can have its center outside the set $\{(x, y) : 0 \leq x \leq h_1, 0 \leq y \leq h_1\}$; otherwise, we would have $h_m > 2h_1$ for some m, so that $|Q_{k_m}| > 4|Q_{k_1}|$, a contradiction. Therefore, the center of each $Q_{k_m}, m \geq 2$, must lie in one of the regions $A, B, C,$ or D indicated in the following illustration:

The center cannot be in A since that would contradict the assumption that the center of Q_{k_m}, $m \geq 2$, does not lie in Q_{k_1}. If it lies in B, then since Q_{k_m} contains 0, it covers B, and so there is at most one Q_{k_m} with center in B. Similar arguments hold for C and D. Applying the same reasoning to each quadrant, we see that there are at most 16 cubes in $\{Q_k\}$ that contain the origin. By translation, the same holds at any point of the plane, and the lemma is proved.

Proof of Besicovitch's lemma. Let

$$\alpha_1 = \sup\{|Q_{\mathbf{x}}| : \mathbf{x} \in E\}.$$

If $\alpha_1 = +\infty$, there are arbitrarily large $Q_{\mathbf{x}}$, and since E is bounded, we simply choose one that contains E. If $\alpha_1 < +\infty$, write $E_1 = E$ and choose $\mathbf{x}_1 \in E_1$ with $|Q_{\mathbf{x}_1}| > \alpha_1/2$. Let

$$E_2 = E_1 - Q_{\mathbf{x}_1}, \quad \alpha_2 = \sup\{|Q_{\mathbf{x}}| : \mathbf{x} \in E_2\}.$$

The definition of α_2 assumes that $E_2 \neq \emptyset$. Then $\alpha_2 > 0$ and we choose $\mathbf{x}_2 \in E_2$ with $\left|Q_{\mathbf{x}_2}\right| > \alpha_2/2$. Proceed in this way, obtaining at the kth stage

$$E_k = E_{k-1} - Q_{\mathbf{x}_{k-1}} = E - \bigcup_{j=1}^{k-1} Q_{\mathbf{x}_j}, \quad \alpha_k = \sup\{|Q_{\mathbf{x}}| : \mathbf{x} \in E_k\},$$
$$\mathbf{x}_k \in E_k, \quad \left|Q_{\mathbf{x}_k}\right| > \alpha_k/2.$$

We continue the process as long as $E_k \neq \emptyset$. Note that $\alpha_k > 0$ if $E_k \neq \emptyset$.

Since $\mathbf{x}_k \in E_k$, we have $\mathbf{x}_k \in E_j$ for all $j \leq k$. Therefore, $\left|Q_{\mathbf{x}_k}\right| \leq \alpha_k \leq \alpha_j < 2\left|Q_{\mathbf{x}_j}\right|$ if $j \leq k$. It follows that $\left\{Q_{\mathbf{x}_k}\right\}$ satisfies the hypothesis of Lemma 10.46 and so has bounded overlaps. It remains only to show that $E \subset \bigcup Q_{\mathbf{x}_k}$.

If some E_{k_0} is empty, then E is contained in the union of the $Q_{\mathbf{x}_k}, k \leq k_0 - 1$. If no E_k is empty, then $Q_{\mathbf{x}_k}$ is defined for every $k = 1, 2, \ldots$. Since $\alpha_k \searrow$ and $\alpha_k/2 < \left|Q_{\mathbf{x}_k}\right| \leq \alpha_k$, it follows either that $\left|Q_{\mathbf{x}_k}\right| \to 0$ or that there exists $\delta > 0$

such that $|Q_{\mathbf{x}_k}| \geq \delta$ for all k. The second possibility cannot arise; otherwise, E would not be bounded since $\mathbf{x}_k \in E$ but $\mathbf{x}_k$ is not in any $Q_{\mathbf{x}_j}$ with $j < k$. Hence, $|Q_{\mathbf{x}_k}| \to 0$, or equivalently, $\alpha_k \to 0$. If $x \in E - \bigcup Q_{\mathbf{x}_k}$, then $\mathbf{x} \in E_k$ for all k. Therefore, $|Q_{\mathbf{x}}| \leq \alpha_k$ for all k, which means that $|Q_{\mathbf{x}}| = 0$. This shows that $E - \bigcup Q_{\mathbf{x}_k}$ is actually empty and completes the proof.

A Borel measure μ on $\mathbf{R^n}$ (i.e., a measure on the Borel subsets of $\mathbf{R^n}$) is called *regular* if

$$\mu(E) = \inf\{\mu(G) : G \supset E, G \text{ open}\}$$

for every Borel set E. If μ is regular and E is a Borel set with $\mu(E) < \infty$, then any open set G that satisfies $E \subset G$ and $\mu(G) < \mu(E) + \varepsilon$ for some $\varepsilon > 0$ also satisfies $\mu(G - E) < \varepsilon$ since $\mu(G - E) = \mu(G) - \mu(E)$ when $\mu(E) < \infty$.

Now suppose that μ is a σ-finite regular Borel measure. Let E be a Borel set and $\varepsilon > 0$. We will show that there is an open set G satisfying $E \subset G$ and $\mu(G - E) < \varepsilon$ whether $\mu(E)$ is finite or not. In fact, since μ is σ-finite, we can write $E = \bigcup_1^\infty E_k$ with $\mu(E_k) < \infty$ for each k, and then by choosing open sets G_k with $\mu(G_k - E_k) < \varepsilon 2^{-k}$, we obtain an open set $G = \bigcup_1^\infty G_k$ such that $E \subset G$ and $G - E \subset \bigcup_1^\infty (G_k - E_k)$, and consequently

$$\mu(G - E) \leq \sum_1^\infty \mu(G_k - E_k) < \sum_1^\infty \frac{\varepsilon}{2^k} = \varepsilon,$$

as desired. Moreover, since the complement CE of E is also a Borel set, it follows that there is an open set $\widetilde{G}$ with $CE \subset \widetilde{G}$ and $\mu(\widetilde{G} - CE) < \varepsilon$. Then the closed set F defined by $F = C\widetilde{G}$ satisfies $F \subset E$ and $\mu(E - F) < \varepsilon$ since $E - F = \widetilde{G} - CE$. In case $\mu(F) < \infty$, we obtain $\mu(E) - \mu(F) < \varepsilon$. In any case, either both $\mu(E)$ and $\mu(F)$ are finite or $\mu(E) = \mu(F) = \infty$. Therefore, for any Borel set E,

$$\mu(E) = \sup_{\substack{F \text{ closed} \\ F \subset E}} \mu(F)$$

if μ is a σ-finite regular Borel measure.

From now on, we will consider two regular Borel measures μ and ν on $\mathbf{R^n}$ that are finite on the bounded Borel sets. Let $Q_{\mathbf{x}}(h)$ denote the cube with center $\mathbf{x}$ and edge length h. We will also assume that

(i) $\mu(Q_{\mathbf{x}}(h)) > 0$ for $\mathbf{x} \in \mathbf{R^n}$, $h > 0$.

(ii) Sets of the form

$$\left\{\mathbf{x} : \sup_{h>0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} > \alpha\right\}, \left\{\mathbf{x} : \limsup_{h\to 0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} > \alpha\right\}, \text{ etc.,}$$

are (Borel) measurable.

Assumption (ii) is made for simplicity and is not necessary (see Exercise 17 of Chapter 11). Also, the assumption that μ and ν are regular is redundant (see Theorem 11.24 and the remarks following it). We assume that μ and ν are finite on the bounded Borel sets in order to ensure their finiteness on every $Q_{\mathbf{x}}(h)$. Thus μ and ν are σ-finite measures.

We will also use the fact that the class of continuous functions with compact support is dense in $L(d\mu)$; that is, given $f \in L(d\mu)$, there exist continuous g_k with compact support such that $\int |f - g_k|\, d\mu \to 0$. (See Exercise 27.)

Note that when μ is Lebesgue measure and $\nu(E) = \int_E |f|\, d\mathbf{x}$ for a locally integrable f, then $\sup_{h>0} \nu(Q_{\mathbf{x}}(h))/\mu(Q_{\mathbf{x}}(h))$ is the Hardy–Littlewood maximal function of f (see (7.5)). In the next lemma, we estimate the size of this expression if μ and ν are any measures with the properties listed above. The results we will prove about differentiation are corollaries of this estimate.

Lemma 10.47 *Let μ and ν satisfy the stated conditions. Then there is a constant c depending only on n such that*

(a) $\mu\{\mathbf{x} \in \mathbf{R^n} : \sup_{h>0} [\nu(Q_{\mathbf{x}}(h))/\mu(Q_{\mathbf{x}}(h))] > \alpha\} \le (c/\alpha)\nu(\mathbf{R^n})$, *and*
(b) $\mu\{\mathbf{x} \in E : \limsup_{h\to 0} [\nu(Q_{\mathbf{x}}(h))/\mu(Q_{\mathbf{x}}(h))] > \alpha\} \le (c/\alpha)\nu(E)$

for any Borel set $E \subset \mathbf{R^n}$ and any $\alpha > 0$

Proof. (a) Fix $\alpha > 0$, and let

$$S = \left\{\mathbf{x} \in \mathbf{R^n} : \sup_{h>0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} > \alpha\right\}.$$

If B is any bounded Borel set and $\mathbf{x} \in S \cap B$, there is a cube $Q_{\mathbf{x}}$ with center $\mathbf{x}$ such that $\nu(Q_{\mathbf{x}})/\mu(Q_{\mathbf{x}}) > \alpha$. Using Besicovitch's lemma, select $\{Q_{\mathbf{x}_k}\}$ and c such that $\nu(Q_{\mathbf{x}_k}) > \alpha\mu(Q_{\mathbf{x}_k})$, $S \cap B \subset \bigcup Q_{\mathbf{x}_k}$, and $\sum \chi_{Q_{\mathbf{x}_k}} \le c$. We then have

$$\mu(S \cap B) \le \mu\left(\bigcup Q_{\mathbf{x}_k}\right) \le \sum \mu(Q_{\mathbf{x}_k}) < \frac{1}{\alpha} \sum \nu(Q_{\mathbf{x}_k}),$$
$$\sum \nu(Q_{\mathbf{x}_k}) = \sum \int_{\bigcup Q_{\mathbf{x}_k}} \chi_{Q_{\mathbf{x}_k}} d\nu \le c \int_{\bigcup Q_{\mathbf{x}_k}} d\nu = c\nu\left(\bigcup Q_{\mathbf{x}_k}\right).$$

Therefore, $\mu(S \cap B) \le c\nu(\bigcup Q_{\mathbf{x}_k})/\alpha$, so that $\mu(S \cap B) \le c\nu(\mathbf{R^n})/\alpha$. Letting $B \nearrow \mathbf{R^n}$, we obtain $\mu(S) \le c\nu(\mathbf{R^n})/\alpha$, as desired.

(b) Fix $\alpha > 0$, and let

$$T = \left\{ \mathbf{x} \in E : \limsup_{h \to 0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} > \alpha \right\}.$$

If $\nu(E) = +\infty$, there is nothing to prove. Otherwise, choose an open set $G \supset E$ with $\nu(G) < \nu(E) + \varepsilon$, and let B be a bounded Borel set. If $\mathbf{x} \in T \cap B$, there is a cube $Q_{\mathbf{x}}$ such that $Q_{\mathbf{x}} \subset G$ and $\nu(Q_{\mathbf{x}})/\mu(Q_{\mathbf{x}}) > \alpha$. By again using Besicovitch's lemma, there exists $\{Q_{\mathbf{x}_k}\}$, $Q_{\mathbf{x}_k} \subset G$, such that $\mu(T \cap B) \leq c\,\nu(\bigcup Q_{\mathbf{x}_k})/\alpha$. Therefore, $\mu(T \cap B) \leq c\,\nu(G)/\alpha \leq c[\nu(E) + \varepsilon]/\alpha$. The result now follows by first letting $\varepsilon \to 0$ and then letting $B \nearrow \mathbf{R^n}$.

The first result about differentiation of measures is the following.

Theorem 10.48 *Let ν and μ satisfy the stated conditions. If ν and μ are mutually singular, then*

$$\lim_{h \to 0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} = 0 \quad \text{a.e. } (\mu).$$

Proof. Since ν and μ are mutually singular, there is a set Z with $\nu(\mathbf{R^n} - Z) = \mu(Z) = 0$. Let $E = \mathbf{R^n} - Z$, and consider the sets

$$T_\alpha = \left\{ \mathbf{x} \in E : \limsup_{h \to 0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} > \alpha \right\}, \quad \alpha > 0,$$

$$T = \left\{ \mathbf{x} \in E : \limsup_{h \to 0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} > 0 \right\}.$$

By Lemma 10.47(b), we have $\mu(T_\alpha) \leq c\nu(E)/\alpha = 0$. Since T is the union of the T_{α_k} for any sequence $\alpha_k \to 0$, it also has μ-measure zero, and the result follows.

Theorem 10.49 *Let μ satisfy the stated conditions, and let f be a Borel measurable function that is integrable $(d\mu)$ over every bounded Borel set in $\mathbf{R^n}$. Then*

$$\lim_{h \to 0} \frac{1}{\mu(Q_{\mathbf{x}}(h))} \int_{Q_{\mathbf{x}}(h)} f\, d\mu = f(\mathbf{x}) \quad \text{a.e. } (\mu).$$

Proof. Assume first that $f \in L(\mathbf{R^n}; d\mu)$. For any integrable g, we have

$$\left| \frac{1}{\mu(Q_{\mathbf{x}}(h))} \int_{Q_{\mathbf{x}}(h)} f\, d\mu - f(\mathbf{x}) \right| \leq \frac{1}{\mu(Q_{\mathbf{x}}(h))} \int_{Q_{\mathbf{x}}(h)} |f - g|\, d\mu$$
$$+ \left| \frac{1}{\mu(Q_{\mathbf{x}}(h))} \int_{Q_{\mathbf{x}}(h)} g\, d\mu - f(\mathbf{x}) \right|.$$

If g is also continuous, the last term on the right converges to $|g(\mathbf{x}) - f(\mathbf{x})|$ as $h \to 0$. Hence, letting $L(\mathbf{x})$ denote the $\limsup$ as $h \to 0$ of the term on the left, we obtain

$$L(\mathbf{x}) \leq \sup_{h>0} \frac{1}{\mu(Q_{\mathbf{x}}(h))} \int_{Q_{\mathbf{x}}(h)} |f - g|\, d\mu + |g(\mathbf{x}) - f(\mathbf{x})|.$$

Therefore, the set S_ε where $L(\mathbf{x}) > \varepsilon$, $\varepsilon > 0$, is contained in the union of the two sets where the corresponding terms on the right side of the last inequality exceed $\varepsilon/2$. From Lemma 10.47 and Tchebyshev's inequality, we obtain

$$\mu(S_\varepsilon) \leq c \left(\frac{\varepsilon}{2}\right)^{-1} \int_{\mathbf{R^n}} |f - g|\, d\mu + \left(\frac{\varepsilon}{2}\right)^{-1} \int_{\mathbf{R^n}} |f - g|\, d\mu.$$

As noted before the proof of Lemma 10.47, g can be chosen such that $\int_{\mathbf{R^n}} |f - g|\, d\mu$ is arbitrarily small. Hence, $\mu(S_\varepsilon) = 0$ for every $\varepsilon > 0$, and the result follows.

The case when $f \notin L(\mathbf{R^n}; d\mu)$ is left as an exercise (cf. Theorem 7.11).

Combining the last two theorems, we obtain the main result:

Corollary 10.50 *Let ν and μ satisfy the stated conditions. If $\nu(E) = \int_E f\, d\mu + \sigma(E)$ is the decomposition of ν into parts that are absolutely continuous and singular with respect to μ, then*

$$\lim_{h \to 0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} = f(\mathbf{x}) \quad \text{a.e. } (\mu).$$

Exercises

1. Prove Theorem 10.13.
2. A measure space $(\mathscr{S}, \Sigma, \mu)$ is said to be *complete* if Σ contains all subsets of sets with measure zero; that is, $(\mathscr{S}, \Sigma, \mu)$ is complete if $Y \in \Sigma$ whenever $Y \subset Z$, $Z \in \Sigma$, and $\mu(Z) = 0$. In this case, show that if f is measurable and $g = f$ a.e. (μ), then g is also measurable (cf. Theorem 4.5 and Chapter 3, Exercise 34). Is this true if $(\mathscr{S}, \Sigma, \mu)$ is not complete?

   Give an example of an incomplete measure space with a measure that is neither identically infinite nor identically zero.
3. Prove Egorov's Theorem 10.14.
4. If $(\mathscr{S}, \Sigma, \mu)$ is a measure space, and if f and $\{f_k\}$ are measurable and finite a.e. (μ) in a measurable set E, then $\{f_k\}$ is said to *converge in* μ*-measure* on E to limit f if

$$\lim_{k\to\infty} \mu\{x \in E : |f(x) - f_k(x)| > \varepsilon\} = 0 \text{ for all } \varepsilon > 0.$$

   Formulate and prove analogues of Theorems 4.21 through 4.23.
5. Complete the proof of Lemma 10.18.
6. (a) If $f_1, f_2 \in L(d\mu)$ and $\int_E f_1\, d\mu = \int_E f_2\, d\mu$ for all measurable E, show that $f_1 = f_2$ a.e. (μ).

   (b) Prove the uniqueness of f and σ in Theorem 10.40.

   (c) Let μ be σ-finite, and let $f_1, f_2 \in L^{p'}(d\mu)$, $1/p + 1/p' = 1$, $1 \le p \le \infty$. If $\int f_1 g\, d\mu = \int f_2 g\, d\mu$ for all $g \in L^p(d\mu)$, show that $f_1 = f_2$ a.e. (μ).
7. Prove the integral convergence results in Theorems 10.27 through 10.29 and 10.31.
8. Show that for $1 \le p < \infty$, the class of simple functions vanishing outside sets of finite measure is dense in $L^p(d\mu)$. See also Exercise 27.
9. The *symmetric difference* of two sets E_1 and E_2 is defined as

$$E_1 \,\Delta\, E_2 = (E_1 - E_2) \cup (E_2 - E_1).$$

   Let $(\mathscr{S}, \Sigma, \mu)$ be a measure space, and identify measurable sets E_1 and E_2 if $\mu(E_1 \Delta E_2) = 0$. Show that Σ is a metric space with distance $d(E_1, E_2) = \mu(E_1 \Delta E_2)$ and that if μ is finite, then $L^p(\mathscr{S}, \Sigma, \mu)$ is separable if and only if Σ is $1 \le p < \infty$. (For the sufficiency in the second part, Exercise 8 may be helpful; for the necessity, let $\{f_k\}$ be a countable dense set in $L^p(\mathscr{S}, \Sigma, \mu)$ and consider the sets $\{1/2 < f_k \le 3/2\}$.)

10. If ϕ is a set function whose Jordan decomposition is $\phi = \overline{V} - \underline{V}$, define

$$\int_E f\,d\phi = \int_E f\,d\overline{V} - \int_E f\,d\underline{V},$$

provided not both integrals on the right are infinite with the same sign. If V is the total variation of ϕ on E, and if $|f| \leq M$, prove that $|\int_E f\,d\phi| \leq MV$.

11. Prove parts (ii)–(iv) of Theorem 10.33.

12. Give an example of a pair of measures ν and μ such that ν is absolutely continuous with respect to μ, but given $\varepsilon > 0$, there is no $\delta > 0$ such that $\nu(A) < \varepsilon$ for every A with $\mu(A) < \delta$. (Thus, the analogue for measures of Theorem 10.34 may fail.)

Prove the analogue of Theorem 10.35 for mutually singular measures ν and μ.

13. Show that the set P of the Hahn decomposition is unique up to null sets. (By a null set for ϕ, we mean a set N such that $\phi(A) = 0$ for every measurable $A \subset N$.)

14. Complete the proof of Theorem 10.44 for $p = 1$.

15. (Converse of Hölder's inequality) Let μ be a σ-finite measure and $1 \leq p \leq \infty$.

(a) Show that

$$\|f\|_p = \sup \left| \int fg\,d\mu \right|,$$

where the supremum is taken over all bounded measurable functions g that vanish outside a set (depending on g) of finite measure, and for which $\|g\|_{p'} \leq 1$ and $\int fg\,d\mu$ exists. (If $1 < p \leq \infty$ and $\|f\|_p < \infty$, this can be deduced from Theorem 10.44.)

(b) Show that a real-valued measurable f belongs to L^p if $fg \in L^1$ for all $g \in L^{p'}$, $1/p + 1/p' = 1$.

16. Consider a convolution operator $Tf(\mathbf{x}) = \int_{\mathbf{R}^n} f(\mathbf{y})K(\mathbf{x}-\mathbf{y})\,d\mathbf{y}$ with $K \geq 0$. If $1 \leq p \leq \infty$ and $\|Tf\|_p \leq M\|f\|_p$ for all f, show that $\|Tf\|_{p'} \leq M\|f\|_{p'}$ for all f, $1/p + 1/p' = 1$. (Use Exercise 15 to write $\|Tf\|_{p'} = \sup_{\|g\|_p \leq 1} |\int_{\mathbf{R}^n} (Tf)g\,d\mathbf{x}|$, and note that

$$\int_{\mathbf{R}^n} (Tf)(\mathbf{x})g(\mathbf{x})\,d\mathbf{x} = \int_{\mathbf{R}^n} (T\widetilde{g})(-\mathbf{y})f(\mathbf{y})\,d\mathbf{y}$$

where $\widetilde{g}(\mathbf{x}) = g(-\mathbf{x})$.)

Find a generalization if the hypothesis is instead that $\|Tf\|_q \leq M\|f\|_p$ for all f, where q is a fixed index with $1 \leq q \leq \infty$ and $q \neq p$.

17. Let μ be σ-finite and define $\mathscr{L}^p(d\mu)$ to be the class of complex-valued f with $\int |f|^p d\mu < +\infty$. Let l be a complex-valued bounded linear functional on $\mathscr{L}^p(d\mu)$. If $1 \le p < \infty$, show that there is a function $g \in \mathscr{L}^{p'}(d\mu)$ such that $l(f) = \int fg\, d\mu$. (Here, as usual, we define $\int h\, d\mu = \int h_1\, d\mu + i \int h_2\, d\mu$ if $h = h_1 + ih_2$ with h_1 and h_2 real-valued.) (Hint: Reduce to the real case.)

18. Give an example to show that $(L^\infty)'$ cannot be identified with L^1 as in Theorem 10.44. (Consider $L^\infty[-1,1]$ with Lebesgue measure, and let $\mathscr{C}$ be the subspace of continuous functions on $[-1,1]$ with the sup norm. Define $l(f) = f(0)$ for $f \in \mathscr{C}$. Then l is a bounded linear functional on $\mathscr{C}$, so by the Hahn–Banach theorem*, l has an extension $l \in (L^\infty[-1,1])'$. If there were a function $g \in L^1[-1,1]$ such that $l(f) = \int_{-1}^1 fg\, dx$ for all $f \in L^\infty[-1,1]$, then we would have $f(0) = \int_{-1}^1 fg\, dx$ for all $f \in \mathscr{C}$. Show that this implies that $g = 0$ a.e., so that $l \equiv 0$. The functional l is called the Dirac δ-*function*.)

To show that $(L^\infty(E;dx))'$ and $L^1(E;dx)$ are not isometrically isomorphic, one can combine the following three facts: $L^1(E;dx)$ is separable; $L^\infty(E;dx)$ is not separable; and, a Banach space is separable if its dual space is separable. For the latter, see the references in the footnote below, Theorem 8.11 on p. 192 of the first, or Theorem 3.26, p. 73, of the second.

19. Complete the proof of Theorem 10.49.

20. Under the hypothesis of Theorem 10.49, prove that

$$\lim_{h\to 0} \frac{1}{\mu(Q_{\mathbf{x}}(h))} \int_{Q_{\mathbf{x}}(h)} |f(\mathbf{y}) - f(\mathbf{x})|\, d\mu(\mathbf{y}) = 0 \quad \text{a.e. } (\mu).$$

21. Derive an analogue of the Besicovitch Covering Lemma for the case of two dimensions (x,y) when the squares $Q_{(x,y)}$ are replaced by rectangles $R_{(x,y)}(h)$ centered at (x,y) whose x and y dimensions are h and h^2, respectively. Use this result to prove that under the hypothesis of Theorem 10.49,

$$\lim_{h\to 0} \frac{1}{\mu(R_{(x,y)}(h))} \int_{R_{(x,y)}(h)} f\, d\mu = f(x,y) \quad \text{a.e. } (\mu).$$

22. Let μ be a measure and A be a set with $0 < \mu(A) < \infty$. Let f be measurable and bounded on A, and let ϕ be convex in an interval containing the range of f. Prove that

* (see, e.g., Theorem 2.5, p. 33 of M. Schechter, *Principles of Functional Analysis*, 2nd edition, Graduate Studies in Mathematics, vol. 36 (2001), *American Mathematical Society*; or Theorem 1.1, p. 1, of H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, 2011).

$$\phi\left(\frac{\int_A f\,d\mu}{\int_A d\mu}\right) \le \frac{\int_A \phi(f)\,d\mu}{\int_A d\mu}.$$

(This is Jensen's inequality for measures. See Theorem 7.44.)

23. A sequence $\{\phi_k\}$ of set functions is said to be *uniformly absolutely continuous* with respect to a measure μ if given $\varepsilon > 0$, there exists $\delta > 0$ such that if E satisfies $\mu(E) < \delta$, then $|\phi_k(E)| < \varepsilon$ for all k. If $\{f_k\}$ is a sequence of integrable functions on a finite measure space $(\mathscr{S}, \Sigma, \mu)$ that converges pointwise a.e. (μ) to an integrable f, show that $f_k \to f$ in $L(d\mu)$ norm if and only if the indefinite integrals of the f_k are uniformly absolutely continuous with respect to μ. (Cf. Exercise 17 of Chapter 7.)

24. Let $(\mathscr{S}, \Sigma, \mu)$ be a σ-finite measure space, and let f be Σ-measurable and integrable over $\mathscr{S}$. Let Σ_0 be a σ-algebra satisfying $\Sigma_0 \subset \Sigma$. Of course, f may not be Σ_0-measurable. Show that there is a unique function f_0 that is Σ_0-measurable such that $\int fg\,d\mu = \int f_0 g\,d\mu$ for every Σ_0-measurable g for which the integrals are finite. The function f_0 is called the *conditional expectation* of f with respect to Σ_0, denoted $f_0 = E(f|\Sigma_0)$. (Apply the Radon–Nikodym theorem to the set function $\phi(E) = \int_E f\,d\mu$, $E \in \Sigma_0$.)

25. Using the notation of the preceding exercise, prove the following:

(a) $E(af + bg|\Sigma_0) = aE(f|\Sigma_0) + bE(g|\Sigma_0)$, a, b constants.

(b) $E(f|\Sigma_0) \ge 0$ if $f \ge 0$.

(c) $E(fg|\Sigma_0) = gE(f|\Sigma_0)$ if g is Σ_0-measurable.

(d) If $\Sigma_1 \subset \Sigma_0 \subset \Sigma$, then $E(f|\Sigma_1) = E(E(f|\Sigma_0)|\Sigma_1)$.

26. (*Hardy's inequality*) Let $f \ge 0$ on $(0,\infty)$, $1 \le p < \infty$, $d\mu(x) = x^\alpha\,dx$ and $d\nu(x) = x^{\alpha+p}\,dx$ on $(0,\infty)$. Prove there exists a constant c independent of f such that

(i) $\int_0^\infty (\int_0^x f(t)\,dt)^p\,d\mu(x) \le c\int_0^\infty f^p(x)\,d\nu(x)$, $\alpha < -1$,

(ii) $\int_0^\infty (\int_x^\infty f(t)\,dt)^p\,d\mu(x) \le c\int_0^\infty f^p(x)\,d\nu(x)$, $\alpha > -1$.

(For (i), $(\int_0^x f(t)\,dt)^p \le cx^{p-\eta-1}\int_0^x f(t)^p t^\eta\,dt$ by Hölder's inequality, provided $p - \eta - 1 > 0$. Multiply both sides by x^α, integrate over $(0,\infty)$, change the order of integration, and observe that an appropriate η exists since $\alpha < -1$.)

27. If μ is a σ-finite regular Borel measure on $\mathbf{R^n}$, show that the class of continuous functions with compact support is dense in $L^p(d\mu)$, $1 \le p < \infty$. (By Exercise 8, it is enough to approximate χ_E, where E is a Borel set with finite measure. Given $\varepsilon > 0$, as shown in Section 10.5 on p. 269, there exist open G and closed F with $F \subset E \subset G$ and $\mu(G - F) < \varepsilon$. Now use Urysohn's lemma: if F_1 and F_2 are disjoint closed sets in $\mathbf{R^n}$, there is a continuous f on $\mathbf{R^n}$ with $0 \le f \le 1$, $f = 1$ on F_1, $f = 0$ on F_2.)

28. Let $1 < p \le \infty$ and μ be a σ-finite measure for which $L^{p'}(E; d\mu)$ is separable, $1/p + 1/p' = 1$. Show that every bounded sequence in $L^p(E; d\mu)$ has a weakly convergent subsequence, that is, if $\sup_k \|f_k\|_p < \infty$, show that

there exists $\{f_{k_j}\}$ and $f \in L^p$ such that $\int_E f_{k_j} g \, d\mu \to \int_E fg \, d\mu$ for all $g \in L^{p'}$. (Use the Bolzano–Weierstrass theorem to show that for every $g \in L^{p'}$, there is a subsequence $\{f_{k_j}\}$ depending on g such that $\int_E f_{k_j} g \, d\mu$ converges. By using a diagonal argument, $\{f_{k_j}\}$ can be chosen to be independent of g for all g in any fixed countable subset S of $L^{p'}$ and consequently for all $g \in L^{p'}$ by choosing S to be dense in $L^{p'}$. Finally, apply Theorem 10.44 to the linear functional $l \in (L^{p'})'$ defined by $l(g) = \lim_{k_j \to \infty} \int_E f_{k_j} g \, d\mu$.)

29. Let l^p be defined as in Section 8.3. Explain how Theorem 10.44 can be used to describe both the action and the norm of a continuous linear functional on l^p, $1 \le p < \infty$.

30. Let Σ be the σ-algebra of Lebesgue measurable sets in $\mathbf{R^1}$. For every $E \in \Sigma$, let $\Lambda(E)$ denote the Lebesgue measure of E, and define measures R and Δ by $R(E) = \Lambda(E \cap [0,1])$ and $\Delta(E) = \chi_E(0)$:

(a) Show that $\int_E f \, d\Delta = f(0)\Delta(E)$.

(b) Is either R or Δ absolutely continuous or singular with respect to Λ?

(c) Identify the functions f and the sets Z in the Lebesgue decompositions of Δ with respect to Λ, of Λ with respect to Δ, of R with respect to Λ, and of Λ with respect to R.

31. Prove the Besicovitch Covering Lemma in case $n = 1$.

32. Let $w(\mathbf{x})$ be a nonnegative locally integrable function on $\mathbf{R^n}$ such that $\int_Q w > 0$ for every cube Q in $\mathbf{R^n}$ with edges parallel to the coordinate axes. Consider the *w-weighted maximal function* $M_w f$ defined by

$$M_w f(\mathbf{x}) = \sup \frac{1}{\int_Q w(\mathbf{y}) \, d\mathbf{y}} \int_Q |f(\mathbf{y})| \, w(\mathbf{y}) \, d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n},$$

where f is a measurable function and the supremum is taken over all cubes $Q \subset \mathbf{R^n}$ *centered* at $\mathbf{x}$ with edges parallel to the coordinate axes. Show that

$$\int_{\{\mathbf{x}: M_w f(\mathbf{x}) > \alpha\}} w(\mathbf{x}) \, d\mathbf{x} \le \frac{C}{\alpha} \int_{\mathbf{R^n}} |f(\mathbf{x})| \, w(\mathbf{x}) \, d\mathbf{x}, \quad \alpha > 0,$$

and that for $1 < p \le \infty$,

$$\left(\int_{\mathbf{R^n}} |M_w f(\mathbf{x})|^p w(\mathbf{x}) \, d\mathbf{x} \right)^{1/p} \le C \left(\int_{\mathbf{R^n}} |f(\mathbf{x})|^p w(\mathbf{x}) \, d\mathbf{x} \right)^{1/p},$$

where the constant C is independent of w, f, and α. (Compare Lemma 10.47 and Theorem 9.16.)

11

Outer Measure and Measure

11.1 Constructing Measures from Outer Measures

A function $\Gamma = \Gamma(A)$ that is defined for every subset A of a set $\mathscr{S}$ is called an *outer measure* if it satisfies the following:

(i) $\Gamma(A) \geq 0,\ \Gamma(\emptyset) = 0$.

(ii) $\Gamma(A_1) \leq \Gamma(A_2)$ if $A_1 \subset A_2$.

(iii) $\Gamma(\bigcup A_k) \leq \sum \Gamma(A_k)$ for any countable collection of sets $\{A_k\}$.

For example, ordinary Lebesgue outer measure is an outer measure on the subsets of $\mathbf{R^n}$. Some other concrete examples will be constructed later in the chapter.

As with Lebesgue outer measure, it is possible to use any outer measure to introduce a class of measurable sets and a corresponding measure. In doing so, we base the definition of measurability on Carathéodory's Theorem 3.30. Thus, given an outer measure Γ, we say that a subset E of $\mathscr{S}$ is Γ-*measurable*, or simply *measurable*, if

$$\Gamma(A) = \Gamma(A \cap E) + \Gamma(A - E) \tag{11.1}$$

for *every* $A \subset \mathscr{S}$. Equivalently, E is measurable if and only if

$$\Gamma(A_1 \cup A_2) = \Gamma(A_1) + \Gamma(A_2) \text{ whenever } A_1 \subset E,\ A_2 \subset \mathscr{S} - E.$$

It follows that a set E is measurable if and only if its complement $\mathscr{S} - E$ is measurable.

As a simple example, let us show that any set Z with $\Gamma(Z) = 0$ is measurable. In fact, for such Z and any $A \subset \mathscr{S}$, property (ii) gives

$$\Gamma(A \cap Z) + \Gamma(A - Z) \leq \Gamma(Z) + \Gamma(A) = \Gamma(A).$$

But by (iii), the opposite inequality $\Gamma(A) \leq \Gamma(A \cap Z) + \Gamma(A - Z)$ is always true, and the measurability of Z follows.

If E is a measurable set, then $\Gamma(E)$ is called its Γ-*measure*, or simply its *measure*. The terminology is justified by the following theorem.

Theorem 11.2 *Let Γ be an outer measure on the subsets of $\mathscr{S}$.*

(i) *The family of Γ-measurable subsets of $\mathscr{S}$ forms a σ-algebra.*

(ii) *Γ is countably additive on disjoint measurable sets, that is, if $\{E_k\}$ is a countable collection of disjoint Γ-measurable sets, then $\Gamma(\bigcup E_k) = \sum \Gamma(E_k)$. More generally, for any A, measurable or not,*

$$\Gamma\left(A \cap \bigcup E_k\right) = \sum \Gamma(A \cap E_k) \quad \text{and}$$
$$\Gamma(A) = \sum \Gamma(A \cap E_k) + \Gamma\left(A - \bigcup E_k\right).$$

Proof. Let $\{E_k\}$ be a collection of disjoint measurable sets, and let $H = \bigcup_{k=1}^{\infty} E_k$ and $H_j = \bigcup_{k=1}^{j} E_k$, $j = 1, 2, \ldots$. We first claim that for every A,

$$\Gamma(A) = \sum_{k=1}^{j} \Gamma(A \cap E_k) + \Gamma(A - H_j).$$

The proof will be by induction on j. If $j = 1$, the formula follows from the measurability of E_1. Assuming that the formula holds for $j - 1$, we have

$$\begin{aligned}\Gamma(A) &= \Gamma(A \cap E_j) + \Gamma(A - E_j)\\ &= \Gamma(A \cap E_j) + \sum_{k=1}^{j-1} \Gamma((A - E_j) \cap E_k) + \Gamma((A - E_j) - H_{j-1}).\end{aligned}$$

Since the E_k are disjoint, $(A - E_j) \cap E_k = A \cap E_k$ for $k \leq j - 1$. Hence, since $(A - E_j) - H_{j-1} = A - H_j$, we obtain $\Gamma(A) = \sum_{k=1}^{j} \Gamma(A \cap E_k) + \Gamma(A - H_j)$, as required. This proves the claim.

Since $H_j \subset H$, we have $\Gamma(A - H_j) \geq \Gamma(A - H)$. Using this fact in the previous formula and letting $j \to \infty$, it follows that

$$\Gamma(A) \geq \sum_{k=1}^{\infty} \Gamma(A \cap E_k) + \Gamma(A - H) \geq \Gamma(A \cap H) + \Gamma(A - H).$$

However, we also have $\Gamma(A) \leq \Gamma(A \cap H) + \Gamma(A - H)$. Therefore, H is measurable and $\Gamma(A) = \sum_{k=1}^{\infty} \Gamma(A \cap E_k) + \Gamma(A - H)$. Replacing A by $A \cap H$ in

this equation, we obtain $\Gamma(A \cap H) = \sum_{k=1}^{\infty} \Gamma(A \cap E_k)$, and the proof of (ii) is complete.

Note we have also shown that a countable union of disjoint measurable sets is measurable, and we know that the complement of a measurable set is measurable. To prove (i), it remains to show that a countable union of arbitrary measurable sets is measurable. We will use the next lemma.

Lemma 11.3 *If E_1 and E_2 are measurable, then so is $E_1 - E_2$.*

Proof. We will show that $\Gamma(A \cup B) = \Gamma(A) + \Gamma(B)$ whenever $A \subset E_1 - E_2$ and $B \subset C(E_1 - E_2)$. Since $B = (B \cap E_2) \cup (B - E_2)$, we have $A \cup B = [A \cup (B - E_2)] \cup [B \cap E_2]$. Hence, since $A \cup (B - E_2) \subset CE_2$ and $B \cap E_2 \subset E_2$, it follows from the measurability of E_2 that $\Gamma(A \cup B) = \Gamma(A \cup (B - E_2)) + \Gamma(B \cap E_2)$. However, $A \subset E_1$ and $B - E_2 \subset C(E_1 - E_2) - E_2 \subset CE_1$. Therefore, since E_1 is measurable, $\Gamma(A \cup (B - E_2)) = \Gamma(A) + \Gamma(B - E_2)$. Combining equalities and using the measurability of E_2, we obtain

$$\Gamma(A \cup B) = \Gamma(A) + \Gamma(B - E_2) + \Gamma(B \cap E_2) = \Gamma(A) + \Gamma(B),$$

which proves the lemma.

Returning now to the proof of part (i) of Theorem 11.2, recall that the complement of a measurable set is measurable. Since $E_1 \cup E_2 = C(CE_1 - E_2)$, it then follows from Lemma 11.3 that $E_1 \cup E_2$ is measurable if E_1 and E_2 are. Therefore, any finite union of measurable sets is measurable. Now, let $\{E_k\}$ be a countable collection of measurable sets. If $H_j = \bigcup_{k=1}^{j} E_k$, then

$$\bigcup_{k=1}^{\infty} E_k = H_1 \cup \left[\bigcup_{j=1}^{\infty} (H_{j+1} - H_j) \right],$$

and since the H_j are measurable and increasing, the terms on the right are measurable and disjoint. Thus, by the case already proved, it follows that $\bigcup_{k=1}^{\infty} E_k$ is measurable. This completes the proof of the theorem.

According to Theorem 11.2, an outer measure Γ is a measure on the σ-algebra of Γ-measurable sets and so enjoys the usual properties of measures. We also have the following result.

Corollary 11.4 *Let Γ be an outer measure on $\mathscr{S}$, let $\{E_k\}$ be a collection of measurable sets, and let A be any set.*

(i) *If* $E_k \nearrow$, *then* $\Gamma(A \cap \lim E_k) = \lim_{k\to\infty} \Gamma(A \cap E_k)$; *if* $E_k \searrow$ *and if* $\Gamma(A \cap E_{k_0})$ *is finite for some* k_0, *then* $\Gamma(A \cap \lim E_k) = \lim_{k\to\infty} \Gamma(A \cap E_k)$.

(ii) $\Gamma(A \cap \liminf E_k) \leq \liminf_{k\to\infty} \Gamma(A \cap E_k)$; *if* $\Gamma(A \cap \bigcup_{k=k_0}^{\infty} E_k)$ *is finite for some* k_0, *then* $\Gamma(A \cap \limsup E_k) \geq \limsup_{k\to\infty} \Gamma(A \cap E_k)$.

Proof. We will prove the first statements in (i) and (ii); the proofs of the second statements are left as exercises. Let E_k be measurable and $E_k \nearrow$. To prove the first part of (i), we may assume that $\Gamma(A \cap E_k)$ is finite for each k; otherwise, the result is clear. The sets $E_1, E_2 - E_1, \ldots, E_{k+1} - E_k, \ldots$ are disjoint and measurable. Since

$$\lim E_k = \bigcup_{k=1}^{\infty} E_k = E_1 \cup \left[\bigcup_{k=1}^{\infty} (E_{k+1} - E_k) \right],$$

it follows from Theorem 11.2 that

$$\Gamma(A \cap \lim E_k) = \Gamma(A \cap E_1) + \sum_{k=1}^{\infty} \Gamma(A \cap (E_{k+1} - E_k)).$$

Moreover, since E_k and $E_{k+1} - E_k$ are disjoint and measurable and E_k has finite measure, we have $\Gamma(A \cap (E_{k+1} - E_k)) = \Gamma(A \cap E_{k+1}) - \Gamma(A \cap E_k)$. Therefore,

$$\begin{aligned} \Gamma(A \cap \lim E_k) &= \Gamma(A \cap E_1) + \sum_{k=1}^{\infty} [\Gamma(A \cap E_{k+1}) - \Gamma(A \cap E_k)] \\ &= \lim_{k\to\infty} \Gamma(A \cap E_{k+1}), \end{aligned}$$

which proves the first part of (i).

For the first part of (ii), let $\{E_k\}$ be measurable and define sets $X_j = \bigcap_{k=j}^{\infty} E_k$, $j = 1, 2, \ldots$. Then $X_j \nearrow \liminf E_k$, so that by (i), $\Gamma(A \cap \liminf E_k) = \lim_{j\to\infty} \Gamma(A \cap X_j)$. But since $A \cap X_j \subset A \cap E_j$, we have $\lim_{j\to\infty} \Gamma(A \cap X_j) \leq \liminf_{j\to\infty} \Gamma(A \cap E_j)$, and the result follows.

11.2 Metric Outer Measures

Now let us introduce a new assumption concerning the underlying space $\mathscr{S}$: namely, that it is a metric space with metric d. The distance between two sets A_1 and A_2 is then defined by

$$d(A_1, A_2) = \inf\{d(x, y) : x \in A_1, y \in A_2\},$$

as in Euclidean space (see p. 5 in Section 1.3). An outer measure Γ on $\mathscr{S}$ is called a *metric outer measure*, or *an outer measure in the sense of Carathéodory*, if

$$\Gamma(A_1 \cup A_2) = \Gamma(A_1) + \Gamma(A_2) \text{ whenever } d(A_1, A_2) > 0.$$

For example, by Lemma 3.16, Lebesgue outer measure satisfies this condition.

An outer measure in a metric space may not be a metric outer measure and may lack properties (in addition to the defining property) of metric outer measures. Consider, for example, the case when $\mathscr{S}$ is the x, y-plane and d is the usual Euclidean metric in $\mathbf{R^2}$. Define

$$r(A) = \frac{1}{d(A, Y)}, \quad A \subset \mathbf{R^2}, \text{ where } Y \text{ is the } y\text{-axis},$$

with the conventions $1/0 = \infty$ and $r(\emptyset) = 0$. We leave it as an exercise to check that r is an outer measure on $\mathbf{R^2}$ but not a metric outer measure. Furthermore, Y and all its subsets are r-measurable (with infinite measure), and no set $B \subset \mathbf{R^2}$ with $d(B, Y) > 0$ is r-measurable. In particular, r does not have the property in the next result, Theorem 11.5.

Since $\mathscr{S}$ is a metric space, it has the topology induced by its metric. Thus, a set G in $\mathscr{S}$ is said to be open if for every $x \in G$, there is a $\delta > 0$ such that the metric ball $\{y : d(x, y) < \delta\}$ lies in G. A closed set is by definition the complement of an open set, and $\mathscr{B}$ denotes the σ-algebra of *Borel subsets* of $\mathscr{S}$; that is, $\mathscr{B}$ is the smallest σ-algebra containing all the open (closed) subsets of $\mathscr{S}$.

Theorem 11.5 *Let Γ be a metric outer measure on a metric space $\mathscr{S}$. Then every Borel subset of $\mathscr{S}$ is Γ-measurable.*

Since the collection of Γ-measurable sets is a σ-algebra, it is enough to prove that every closed set is Γ-measurable. To prove this, we will use the following fact.

Lemma 11.6 *Let Γ be a metric outer measure on a space $\mathscr{S}$ with metric d. Let A be any set contained in an open set G, and let $A_k = \{x \in A : d(x, CG) \geq 1/k\}$, $k = 1, 2, \ldots$. Then $\lim_{k\to\infty} \Gamma(A_k) = \Gamma(A)$.*

Proof. Since G is open, we have $A_k \nearrow A$. Clearly, $\lim_{k\to\infty} \Gamma(A_k) \leq \Gamma(A)$. To prove the opposite inequality, let $D_k = A_{k+1} - A_k, k = 1, 2, \ldots$ Then $d(D_{k+1}, A_k) \geq [(1/k) - (1/(k+1))] > 0$ since if $x \in A_k$ and $y \in D_{k+1}$, then

$$\frac{1}{k} \leq d(x, CG) \leq d(x, y) + d(y, CG) < d(x, y) + \frac{1}{k+1},$$

where the second inequality is true since d is a metric. We also have

$$A = A_k \cup D_k \cup D_{k+1} \cup \cdots, \quad \Gamma(A) \leq \Gamma(A_k) + \Gamma(D_k) + \Gamma(D_{k+1}) + \cdots.$$

If $\sum \Gamma(D_j) < +\infty$, then $\sum_{j \geq k} \Gamma(D_j)$ tends to zero as $k \to \infty$, and it follows that $\Gamma(A) \leq \lim_{k\to\infty} \Gamma(A_k)$, as desired.

If $\sum \Gamma(D_j) = +\infty$, then at least one of $\sum \Gamma(D_{2j})$ and $\sum \Gamma(D_{2j+1})$ is infinite. We can therefore choose N so that $\Gamma(D_N) + \Gamma(D_{N-2}) + \Gamma(D_{N-4}) + \cdots$ is arbitrarily large. However, when $k \geq 2$, the fact that $\bigcup_{j=1}^{k-1} D_j \subset A_k$ implies that the distance between D_{k+1} and $\bigcup_{j=1}^{k-1} D_j$ is positive. Therefore,

$$\Gamma(D_N \cup D_{N-2} \cup D_{N-4} \cup \cdots) = \Gamma(D_N) + \Gamma(D_{N-2}) + \Gamma(D_{N-4}) + \cdots.$$

Since A_{N+1} contains $D_N \cup D_{N-2} \cup D_{N-4} \cup \cdots$, it follows that $\lim \Gamma(A_k) = +\infty$, and the lemma is proved.

Proof of Theorem 11.5. Let F be any closed set. It is enough to show that $\Gamma(A \cup B) = \Gamma(A) + \Gamma(B)$ for $A \subset CF$, $B \subset F$. If $A_k = \{x \in A : d(x, F) \geq 1/k\}$, then $d(A_k, B) \geq 1/k$, so that $\Gamma(A_k \cup B) = \Gamma(A_k) + \Gamma(B)$. Therefore, $\Gamma(A \cup B) \geq \Gamma(A_k) + \Gamma(B)$. Letting $k \to \infty$, it follows from the lemma that $\Gamma(A \cup B) \geq \Gamma(A) + \Gamma(B)$. Since the opposite inequality is also true, the theorem is proved.

If $\mathscr{S}$ is a metric space, the notions of upper and lower semicontinuity of functions can be defined just as in $\mathbf{R^n}$. For example, a real-valued function f defined near a point x_0 is said to be upper semicontinuous at x_0 if

$$\limsup_{x \to x_0} f(x) \leq f(x_0).$$

Here, of course, the notation $x \to x_0$ means that $d(x, x_0) \to 0$. The results of Theorem 4.14 are valid for metric spaces; for example, f is usc at every point of $\mathscr{S}$ if and only if $\{f \geq a\}$ is closed for every a. We thus obtain the following fact.

Corollary 11.7 *Let Γ be a metric outer measure on $\mathscr{S}$. Then every semicontinuous function on $\mathscr{S}$ is Γ-measurable.*

Proof. Suppose, for example, that f is upper semicontinuous on $\mathscr{S}$. Then $\{x : f(x) \geq a\}$ is closed for every a, and so is Γ-measurable by Theorem 11.5. Hence, f is Γ-measurable. If f is lower semicontinuous, then $-f$ is upper semicontinuous, and the corollary follows.

11.3 Lebesgue–Stieltjes Measure

In this section and the next, we will consider two specific examples of outer measures in the sense of Carathéodory. The first is known as Lebesgue–Stieltjes outer measure. It elucidates the connection between measures and monotone functions. The situation is relatively simple for measures on $\mathbf{R^1}$ and monotone functions of a single variable, and we shall restrict our attention to this case. Extensions to higher dimensions are possible but more complicated.

To construct a typical Lebesgue–Stieltjes outer measure, consider any fixed function f that is finite and monotone increasing (i.e., nondecreasing) on $(-\infty, +\infty)$. For each half-open finite interval of the form $(a, b]$, let

$$\lambda(a, b] = \lambda_f((a, b]) = f(b) - f(a).$$

Note that $\lambda \geq 0$ since f is increasing. If A is a nonempty subset of $\mathbf{R^1}$, let

$$\Lambda^*(A) = \Lambda_f^*(A) = \inf \sum \lambda(a_k, b_k],$$

where the inf is taken over all countable collections $\{(a_k, b_k]\}$ such that $A \subset \bigcup (a_k, b_k]$. Further, define $\Lambda^*(\emptyset) = 0$.

Theorem 11.8 *Λ^* is a Carathéodory outer measure on $\mathbf{R^1}$.*

Proof. We have $\Lambda^* \geq 0$ and $\Lambda^*(\emptyset) = 0$. First, we will show that if $A_1 \subset A_2$, then $\Lambda^*(A_1) \leq \Lambda^*(A_2)$. This is obvious if either $A_1 = \emptyset$ or $\Lambda^*(A_2) = +\infty$. In any other case, choose $\{(a_k, b_k]\}$ such that $A_2 \subset \bigcup (a_k, b_k]$ and $\sum \lambda(a_k, b_k] < \Lambda^*(A_2) + \varepsilon$. Then $A_1 \subset \bigcup (a_k, b_k]$, so that $\Lambda^*(A_1) \leq \sum \lambda(a_k, b_k]$. Therefore, $\Lambda^*(A_1) < \Lambda^*(A_2) + \varepsilon$, and the result follows by letting $\varepsilon \to 0$.

To show that Λ^* is subadditive, let $\{A_j\}_{j=1}^{\infty}$ be a collection of nonempty subsets of $\mathbf{R^1}$ and let $A = \bigcup A_j$. We may assume that $\Lambda^*(A_j) < +\infty$ for each j. Choose $\left\{(a_k^j, b_k^j]\right\}$ such that

$$A_j \subset \bigcup_k (a_k^j, b_k^j] \quad \text{and} \quad \sum_k \lambda(a_k^j, b_k^j] < \Lambda^*(A_j) + \varepsilon 2^{-j}.$$

Since $A \subset \bigcup_{j,k} (a_k^j, b_k^j]$, we have

$$\Lambda^*(A) \leq \sum_{j,k} \lambda(a_k^j, b_k^j] < \sum_j \Lambda^*(A_j) + \varepsilon.$$

It follows that $\Lambda^*(A) \leq \sum \Lambda^*(A_j)$ and therefore that Λ^* is an outer measure.

To show that Λ^* is a Carathéodory outer measure, observe that if $a = a_0 < a_1 < \cdots < a_N = b$, then

$$\lambda(a,b] = f(b) - f(a) = \sum_{k=1}^{N}[f(a_k) - f(a_{k-1})] = \sum_{k=1}^{N}\lambda(a_{k-1}, a_k].$$

It follows that in defining Λ^*, we can always work with arbitrarily short intervals $(a_k, b_k]$. Hence, if A_1 and A_2 satisfy $d(A_1, A_2) > 0$, then given $\varepsilon > 0$, we can choose $\{(a_k, b_k]\}$ such that each $(a_k, b_k]$ has length less than $d(A_1, A_2)$ and

$$A_1 \cup A_2 \subset \bigcup (a_k, b_k], \quad \sum \lambda(a_k, b_k] \leq \Lambda^*(A_1 \cup A_2) + \varepsilon.$$

Thus, the collection $\{(a_k, b_k]\}$ splits into two coverings, one of A_1 and the other of A_2. Therefore, $\Lambda^*(A_1) + \Lambda^*(A_2) \leq \sum \lambda(a_k, b_k]$, so that since ε is arbitrary, $\Lambda^*(A_1) + \Lambda^*(A_2) \leq \Lambda^*(A_1 \cup A_2)$. But the opposite inequality is always true, which completes the proof.

Λ_f^* is called the *Lebesgue–Stieltjes outer measure corresponding to* f, and its restriction to those sets that are Λ_f^*-measurable is called the *Lebesgue–Stieltjes measure corresponding to* f and denoted Λ_f or simply Λ. Every Borel set in $(-\infty, \infty)$ is Λ_f^*-measurable by Theorems 11.5 and 11.8. In particular, since $(a, b]$ is a Borel set, $\Lambda_f^*((a, b]) = \Lambda_f((a, b])$ for every $(a, b]$.

We leave it as an exercise to show that the Lebesgue–Stieltjes outer measure Λ_x^* corresponding to $f(x) = x$ coincides with ordinary Lebesgue outer measure in $\mathbf{R^1}$. Hence, by Carathéodory's Theorem 3.30, a set is Λ_x^*-measurable if and only if it is Lebesgue measurable.

An outer measure Γ defined on the subsets of a set $\mathscr{S}$ is said to be *regular* if for every $A \subset \mathscr{S}$ there is a Γ-measurable set E such that $A \subset E$ and $\Gamma(A) = \Gamma(E)$. Ordinary Lebesgue outer measure in $\mathbf{R^n}$ is regular by Theorem 3.8. The next theorem shows that any Lebesgue–Stieltjes outer measure is regular; in fact, it shows that any set in $\mathbf{R^1}$ can be included in a Borel set with the same Lebesgue–Stieltjes outer measure. Of course, Borel sets are Λ^*-measurable by Theorem 11.5.

Theorem 11.9 *Let Λ^* be a Lebesgue–Stieltjes outer measure. If A is a subset of $\mathbf{R^1}$, there is a Borel set B containing A such that $\Lambda^*(A) = \Lambda(B)$.*

Proof. Given $j = 1, 2, \ldots$, choose $\{(a_k^j, b_k^j]\}$ such that

$$A \subset \bigcup_k (a_k^j, b_k^j], \quad \sum_k \lambda(a_k^j, b_k^j] \leq \Lambda^*(A) + \frac{1}{j}.$$

Let $B_j = \bigcup_k (a_k^j, b_k^j]$ and $B = \bigcap B_j$. Then $A \subset B$ and B is a Borel set. Moreover,

$$\Lambda(B_j) \le \sum_k \lambda(a_k^j, b_k^j] \le \Lambda^*(A) + \frac{1}{j}.$$

Since $B \subset B_j$, it follows that $\Lambda(B) \le \Lambda^*(A) + (1/j)$, so that $\Lambda(B) \le \Lambda^*(A)$. But the opposite inequality is also true since $A \subset B$, and the theorem follows.

If μ is a finite Borel measure on $\mathbf{R}^1$, define

$$f_\mu(x) = \mu((-\infty, x]), \quad -\infty < x < +\infty.$$

Note that f_μ is monotone increasing and that $\mu((a,b]) = f_\mu(b) - f_\mu(a)$. It is natural to ask if the Lebesgue–Stieltjes measure induced by f_μ agrees with μ as a Borel measure. An affirmative answer would mean that every finite Borel measure is a Lebesgue–Stieltjes measure. We shall see later (Corollary 11.22) that this is actually the case and that the continuity from the right of f_μ (see Exercise 2) plays a role. The next result is also useful.

Theorem 11.10 *If f is an increasing function that is continuous from the right, then its Lebesgue–Stieltjes measure Λ satisfies*

$$\Lambda((a,b]) = f(b) - f(a).$$

In particular, $\Lambda(\{a\}) = f(a) - f(a-)$.

Proof. Since $(a,b]$ covers itself, we always have $\Lambda((a,b]) = \Lambda^*((a,b]) \le f(b) - f(a)$. To show the opposite inequality, suppose that $(a,b] \subset \bigcup (a_k, b_k]$. Given $\varepsilon > 0$, use the right continuity of f to choose $\{b_k'\}$ with

$$b_k < b_k', \quad f(b_k) > f(b_k') - \varepsilon 2^{-k}.$$

If a' satisfies $a < a' < b$, then $[a', b]$ is covered by the (a_k, b_k'), and therefore, there is a finite N such that $[a', b] \subset \bigcup_{k=1}^N (a_k, b_k')$. By discarding any unnecessary (a_k, b_k') and reindexing the rest, we may assume that $a_{k+1} < b_k'$ for $k = 1, \ldots, N-1$. Also, $a_1 < a'$ and $b < b_N'$, so that $f(a_1) \le f(a')$ and $f(b) \le f(b_N')$. We have

$$\begin{aligned}\sum_k \lambda(a_k, b_k] &\ge \sum_{k=1}^N \lambda(a_k, b_k] = \sum_{k=1}^N [f(b_k) - f(a_k)] \\ &= f(b_N) - f(a_1) + \sum_{k=1}^{N-1} [f(b_k) - f(a_{k+1})].\end{aligned}$$

Now,

$$\begin{aligned} f(b_N) - f(a_1) &= [f(b_N) - f(b'_N)] + [f(b'_N) - f(a_1)] \\ &\geq -\varepsilon + [f(b) - f(a')]. \end{aligned}$$

Also, since $f(b'_k) - f(a_{k+1}) \geq 0$ for $k = 1, \ldots, N-1$,

$$\begin{aligned} \sum_{k=1}^{N-1} [f(b_k) - f(a_{k+1})] &= \sum_{k=1}^{N-1} [f(b_k) - f(b'_k)] + \sum_{k=1}^{N-1} [f(b'_k) - f(a_{k+1})] \\ &\geq \sum_{k=1}^{\infty} (-\varepsilon 2^{-k}) + 0 = -\varepsilon. \end{aligned}$$

Combining estimates, we obtain

$$\sum_k \lambda(a_k, b_k] \geq -2\varepsilon + [f(b) - f(a')].$$

Letting $\varepsilon \to 0$ and $a' \to a$, we have $\sum_k \lambda(a_k, b_k] \geq f(b) - f(a)$. Hence, $\Lambda((a, b]) \geq f(b) - f(a)$, and the first statement of the theorem follows. The second statement is proved by applying the first to the intervals $(a - (1/k), a], k = 1, 2, \ldots$, which decrease to $\{a\}$. This completes the proof.

Let g be a Borel measurable function defined on $\mathbf{R^1}$, and let Λ_f be a Lebesgue–Stieltjes measure. Then the integral $\int g\, d\Lambda_f$ is called *the Lebesgue–Stieltjes integral of* g *with respect to* Λ_f.* The next theorem gives a relation between $\int g\, d\Lambda_f$ and the usual Riemann–Stieltjes integral $\int g\, df$.

Theorem 11.11 *Let f be an increasing function that is right continuous on $[a, b]$, and let g be a bounded Borel measurable function on $[a, b]$. If the Riemann–Stieltjes integral $\int_a^b g\, df$ exists, then*

$$\int_{(a,b]} g\, d\Lambda_f = \int_a^b g\, df.$$

Proof. Let $\Gamma = \{x_j\}$ be a partition of $[a, b]$, and let m_j and M_j be the inf and sup of g in $[x_{j-1}, x_j]$ respectively. Let

* In some other texts, any integral $\int f\, d\mu$ of the kind considered in Chapter 10 is called a Lebesgue–Stieltjes integral. We shall use the terminology only when μ is a Lebesgue–Stieltjes measure.

$$L_\Gamma = \sum m_j[f(x_j) - f(x_{j-1})], \quad U_\Gamma = \sum M_j[f(x_j) - f(x_{j-1})]$$

denote the corresponding lower and upper Riemann–Stieltjes sums. Define functions g_1 and g_2 by setting $g_1 = m_j$ in $(x_{j-1}, x_j]$ and $g_2 = M_j$ in $(x_{j-1}, x_j]$. Since f is right continuous, it follows from Theorem 11.10 that

$$\int_{(a,b]} g_1 \, d\Lambda = L_\Gamma, \quad \int_{(a,b]} g_2 \, d\Lambda = U_\Gamma.$$

Therefore, since $g_1 \le g \le g_2$, we obtain $L_\Gamma \le \int_{(a,b]} g \, d\Lambda \le U_\Gamma$. However, as $|\Gamma| \to 0$, both L_Γ and U_Γ converge to $\int_a^b g \, df$ by Theorem 2.29. This completes the proof.

We remark in passing that a right continuous function f of bounded variation can be written $f = f_1 - f_2$, where f_1 and f_2 are right continuous, bounded, and increasing. If Λ_1 and Λ_2 are the Lebesgue–Stieltjes measures corresponding to f_1 and f_2, consider the Borel set function $\Phi = \Lambda_1 - \Lambda_2$, and define

$$\int g \, d\Phi = \int g \, d\Lambda_1 - \int g \, d\Lambda_2.$$

If $\int_a^b g \, df_1$ and $\int_a^b g \, df_2$ exist and are finite for a bounded Borel measurable function g, it then follows from Theorems 11.11 and 2.16 that

$$\int_{(a,b]} g \, d\Phi = \int_a^b g \, df.$$

11.4 Hausdorff Measure

Our second example of a Carathéodory outer measure is Hausdorff outer measure in $\mathbf{R^n}$. To define it, fix $\alpha > 0$, and let A be any subset of $\mathbf{R^n}$. Given $\varepsilon > 0$, let

$$H_\alpha^{(\varepsilon)}(A) = \inf \sum_k \delta(A_k)^\alpha,$$

where $\delta(A_k)$ denotes the diameter of A_k (see p. 5 in Section 1.3), and the inf is taken over all countable collections $\{A_k\}$ such that $A \subset \bigcup A_k$ and $\delta(A_k) < \varepsilon$ for all k. We may always assume that the A_k in a given covering are disjoint and that $A = \bigcup A_k$.

If $\varepsilon' < \varepsilon$, each covering of A by sets with diameters less than ε' is also such a cover for ε. Hence, as ε decreases, the collection of coverings decreases, and consequently $H_\alpha^{(\varepsilon)}(A)$ increases. Define

$$H_\alpha(A) = \lim_{\varepsilon \to 0} H_\alpha^{(\varepsilon)}(A).$$

Theorem 11.12 *For $\alpha > 0, H_\alpha$ is a Carathéodory outer measure on $\mathbf{R^n}$.*

Proof. Clearly, $H_\alpha \geq 0$ and $H_\alpha(\emptyset) = 0$. If $A_1 \subset A_2$, then any covering of A_2 is also one of A_1, so that $H_\alpha^{(\varepsilon)}(A_1) \leq H_\alpha^{(\varepsilon)}(A_2)$. Letting $\varepsilon \to 0$, we obtain $H_\alpha(A_1) \leq H_\alpha(A_2)$. To show that H_α is subadditive, let $A = \bigcup A_k$, and choose a cover of A_k for each k. The union of these is a cover of A, and it is easy to show that $H_\alpha^{(\varepsilon)}(A) \leq \sum H_\alpha^{(\varepsilon)}(A_k) \leq \sum H_\alpha(A_k)$. Letting $\varepsilon \to 0$, we get $H_\alpha(A) \leq \sum H_\alpha(A_k)$. The details of this argument and the proof that H_α is a Carathéodory outer measure are left as exercises.

H_α is called *Hausdorff outer measure of dimension* α on $\mathbf{R^n}$, and the corresponding measure is called *Hausdorff measure of dimension* α and also denoted H_α. It has the following basic property.

Theorem 11.13

(i) *If $H_\alpha(A) < +\infty$, then $H_\beta(A) = 0$ for $\beta > \alpha$.*
(ii) *If $H_\alpha(A) > 0$, then $H_\beta(A) = +\infty$ for $\beta < \alpha$.*

Proof. Statements (i) and (ii) are equivalent. To prove (i), let $A = \bigcup A_k$, $\delta(A_k) < \varepsilon$. If $\beta > \alpha$, then

$$H_\beta^{(\varepsilon)}(A) \leq \sum \delta(A_k)^\beta \leq \varepsilon^{\beta-\alpha} \sum \delta(A_k)^\alpha.$$

Therefore, $H_\beta^{(\varepsilon)}(A) \leq \varepsilon^{\beta-\alpha} H_\alpha^{(\varepsilon)}(A)$. Letting $\varepsilon \to 0$, we obtain $H_\beta(A) = 0$ if $H_\alpha(A) < +\infty$, completing the proof.

The next theorem shows that Hausdorff outer measure is regular, by showing that any set in $\mathbf{R^n}$ can be included in a Borel set with the same Hausdorff outer measure.

Theorem 11.14 *Given $A \subset \mathbf{R^n}$ and $\alpha > 0$, there is a set B of type G_δ containing A such that $H_\alpha(A) = H_\alpha(B)$.*

Proof. Given $\varepsilon > 0$, choose $\{A_k\}$ such that $A = \bigcup A_k$, $\delta(A_k) < \varepsilon/2$ and

$$\sum \delta(A_k)^\alpha \le H_\alpha^{(\varepsilon/2)}(A) + \varepsilon \le H_\alpha(A) + \varepsilon.$$

Enclose A_k in an open set G_k with $\delta(G_k) \le (1+\varepsilon)\delta(A_k)$; this can be done by letting $G_k = \{\mathbf{x} : d(\mathbf{x}, A_k) < \varepsilon\delta(A_k)/2\}$. Let $G = \bigcup G_k$. Then G is open and $A \subset G$. Since $\delta(G_k) < (1+\varepsilon)\varepsilon/2 < \varepsilon$ for $0 < \varepsilon < 1$, we have

$$\begin{aligned} H_\alpha^{(\varepsilon)}(G) &\le \sum \delta(G_k)^\alpha \le (1+\varepsilon)^\alpha \sum \delta(A_k)^\alpha \\ &\le (1+\varepsilon)^\alpha [H_\alpha(A) + \varepsilon]. \end{aligned}$$

Now let $\varepsilon \to 0$ through a sequence $\{\varepsilon_j\}$, and let $G(j)$ be the corresponding open sets G as above. If $B = \bigcap G(j)$, then B is of type G_δ and $A \subset B$. Also, since $B \subset G(j)$ for each j, we have

$$H_\alpha^{(\varepsilon)}(B) \le (1+\varepsilon)^\alpha [H_\alpha(A) + \varepsilon] \quad \text{for } \varepsilon = \varepsilon_j.$$

Letting $j \to \infty$, we obtain $H_\alpha(B) \le H_\alpha(A)$. Since the opposite inequality is clearly true, the result follows.

If A is a subset of $\mathbf{R^1}$ and $A = \bigcup A_k$ with $\delta(A_k) < \varepsilon$, then $\delta(A_k) = |I_k|$, where I_k is the smallest interval containing A_k. Hence, in the one-dimensional case,

$$H_\alpha^{(\varepsilon)}(A) = \inf \sum |I_k|^a \quad (n = 1),$$

where the I_k's are intervals of length less than ε such that $A \subset \bigcup I_k$. If $\alpha = 1$, it follows that $H_1(A)$ is the usual Lebesgue outer measure of A. In $\mathbf{R^n}$, $n > 1$, H_n is not the same as Lebesgue outer measure (see Exercise 10). Nevertheless, there is a simple relation between the two, which is a corollary of the next lemma.

Let

$$H_\alpha'^{(\varepsilon)}(A) = \inf \sum \delta(Q_k)^\alpha,$$

where $\{Q_k\}$ is any collection of *cubes* with edges parallel to the axes such that $A \subset \bigcup Q_k$ and $\delta(Q_k) < \varepsilon$. Also, let

$$H'_\alpha(A) = \lim_{\varepsilon \to 0} H_\alpha'^{(\varepsilon)}(A).$$

Thus, H'_α is defined in the same way as H_α, except that cubes are used instead of arbitrary sets.

Lemma 11.15 *There is a constant c depending only on n and α such that*

$$H_\alpha(A) \le H'_\alpha(A) \le cH_\alpha(A), \quad A \subset \mathbf{R^n}.$$

Proof. Since every covering of A by cubes is a covering of A, we obtain $H_\alpha(A) \le H'_\alpha(A)$. Any set with diameter δ, say, is contained in a cube with edge length 2δ and so with diameter $2\sqrt{n}\,\delta$. Now let $A = \bigcup A_k, \delta(A_k) < \varepsilon$. Select cubes $Q_k \supset A_k$ with $\delta(Q_k) = 2\sqrt{n}\,\delta(A_k)$. Then

$$\sum \delta(A_k)^\alpha = (2\sqrt{n})^{-\alpha} \sum \delta(Q_k)^\alpha \ge (2\sqrt{n})^{-\alpha} H'^{(2\sqrt{n}\,\varepsilon)}_\alpha(A).$$

Therefore, $H^{(\varepsilon)}_\alpha(A) \ge (2\sqrt{n})^{-\alpha} H'^{(2\sqrt{n}\varepsilon)}_\alpha(A)$. Letting $\varepsilon \to 0$, we obtain $H_\alpha(A) \ge (2\sqrt{n})^{-\alpha} H'_\alpha(A)$, which completes the proof.

Theorem 11.16

(i) *There are positive constants c_1 and c_2 depending only on the dimension n such that $c_1 H_n(A) \le |A|_e \le c_2 H_n(A)$ for $A \subset \mathbf{R^n}$.*

(ii) *If $\alpha > n$, then $H_\alpha(A) = 0$ for every $A \subset \mathbf{R^n}$.*

Proof. We first claim that for every set $A \subset \mathbf{R^n}$,

$$H'_n(A) = \inf_{\{Q_k\}} \sum \delta(Q_k)^n,$$

where the inf is taken over all collections $\{Q_k\}$ of cubes with edges parallel to the coordinate axes that cover A, *without* restriction on the size of $\delta(Q_k)$. Let I denote the inf on the right side. It follows easily that $H'_n(A) \ge I$. To show the opposite inequality, let $\{Q_k : k = 1, 2, \ldots\}$ be a collection of cubes with edges parallel to the coordinate axes such that $A \subset \bigcup Q_k$, and let $\varepsilon, \eta > 0$. Pick cubes $\{Q^*_k\}$ satisfying $Q_k \subset (Q^*_k)^\circ$ and $|Q^*_k - Q_k| < \eta/2^k$ for each k. Decompose $(Q^*_k)^\circ = \bigcup_j \widetilde{Q}_{k,j}$ into the union of nonoverlapping cubes $\widetilde{Q}_{k,j}$ with $\delta(\widetilde{Q}_{k,j}) < \varepsilon$ (and edges parallel to the coordinate axes); see Theorem 1.11 and the comment after its proof about the size of the initial net of cubes used in its proof. Then, for each k, we have $|Q^*_k| = \sum_j |\widetilde{Q}_{k,j}|$, and consequently $\delta(Q^*_k)^n = \sum_j \delta(\widetilde{Q}_{k,j})^n$ since for any cube Q, $\delta(Q)^n$ is proportional to the volume of Q: in fact, $\delta(Q)^n = n^{n/2}|Q|$. Then $A \subset \bigcup_{k,j} \widetilde{Q}_{k,j}$ and

$$\begin{aligned} H_n'^{(\varepsilon)}(A) &\le \sum_{k,j} \delta(\widetilde{Q}_{k,j})^n = n^{n/2} \sum_k |Q_k^*| \\ &= n^{n/2} \sum_{k\ge 1} \big(|Q_k| + |Q_k^* - Q_k|\big) \\ &< \sum_k \delta(Q_k)^n \ + n^{n/2}\eta. \end{aligned}$$

Letting η and ε tend to 0, we obtain $H_n'(A) \le \sum \delta(Q_k)^n$. Hence, $H_n'(A) \le I$, and the claim is proved.

Next note that the inf denoted earlier by I satisfies $I = \inf \sum n^{n/2}|Q_k| = n^{n/2}|A|_e$ (see Exercise 22(a) of Chapter 3). Consequently, $H_n'(A) = n^{n/2}|A|_e$ for every $A \subset \mathbf{R^n}$. Part (i) now follows from the fact that H_n' and H_n are comparable (Lemma 11.15).

For part (ii), if $H_n(A)$ is finite, then $H_\alpha(A) = 0$ for $\alpha > n$ by Theorem 11.13. If $H_n(A) = +\infty$, write $A = \bigcup(A \cap Q_j)$, where the Q_j are disjoint (partly open) cubes. Since $|A \cap Q_j|_e$ is finite, so is $H_n(A \cap Q_j)$. Hence, $H_\alpha(A \cap Q_j) = 0$ for $\alpha > n$. Therefore, $H_\alpha(A) \le \sum H_\alpha(A \cap Q_j) = 0$.

It is natural to ask if $H_\alpha(A)$ is comparable to the expression

$$\inf \sum \delta(A_k)^\alpha,$$

where the inf is taken over all coverings $\{A_k\}$ of A, without any requirement on the size of the diameters. It is not difficult to see that the answer in general is *no*. In fact, this expression is finite for any bounded A, as is easily seen from covering A by itself. On the other hand, it is clear from Theorems 11.13 and 11.16 that if $|A|_e > 0$, then $H_\alpha(A) = +\infty$ for $\alpha < n$.

However, in case $\alpha = n$, $H_\alpha(A)$ is comparable to the previous expression. To see this, it is enough by Lemma 11.15 to prove that $H_n'(A)$ is comparable to it. But since $H_n'(A) = \inf \sum \delta(Q_k)^n$, where $\{Q_k\}$ is any collection of cubes covering A whose edges are parallel to the axes, this follows from the fact that $\inf \sum \delta(Q_k)^n$ and $\inf \sum \delta(A_k)^n$ are comparable (cf. the proof of Lemma 11.15).

We remark in passing that Hausdorff measure is particularly useful in measuring sets with Lebesgue measure zero since these may have positive Hausdorff measure for some $\alpha < n$. For example, it can be shown that the Cantor set in [0,1] has Hausdorff measure of dimension log 2/log 3 equal to 1; see Exercise 19.

11.5 The Carathéodory–Hahn Extension Theorem

In this section, we will settle the question that arose in the discussion preceding Theorem 11.10. We recall the situation. Let μ be a finite Borel measure

on $\mathbf{R^1}$, and let Λ be the Lebesgue–Stieltjes measure induced by the function $f(x) = \mu((-\infty, x])$. Since f is continuous from the right, we have by Theorem 11.10 that

$$\Lambda((a,b]) = \mu((a,b]) \quad [= f(b) - f(a)].$$

The point in question is whether this implies that μ and Λ agree on every Borel set in $\mathbf{R^1}$. More generally, we may ask if two Borel measures can be finite and equal on every $(a,b]$ without being identical. It is worthwhile to consider this question in a still more general context, which we now present.

Let $\mathscr{S}$ be a fixed set. By an *algebra* $\mathscr{A}$ of subsets of $\mathscr{S}$, we mean a nonempty collection of subsets of $\mathscr{S}$ that is closed under the operations of taking complements and *finite* unions; that is, $\mathscr{A}$ is an algebra if it satisfies the following:

(i) If $A \in \mathscr{A}$, then $CA\ (= \mathscr{S} - A) \in \mathscr{A}$.
(ii) If $A_1, \ldots, A_N \in \mathscr{A}$, then $\bigcup_{k=1}^{N} A_k \in \mathscr{A}$.

What distinguishes an algebra from a σ-algebra is that an algebra is only closed under *finite* unions. It follows from the definition that an algebra is also closed under finite intersections and differences (relative complements) and that both the empty set $\emptyset$ and the whole space $\mathscr{S}$ belong to it.

The collection of finite intervals $(a,b]$ on the line is clearly not an algebra. However, we can generate an algebra from it by adjoining $\emptyset, \mathbf{R^1}$, and all intervals of the form $(-\infty, a]$ and $(b, +\infty)$, as well as all possible finite disjoint unions of these and the intervals $(a,b]$. This algebra will be called the *algebra generated by the intervals* $(a,b]$.

By a *measure* λ *on an algebra* $\mathscr{A}$, we mean a function λ which is defined on the elements of $\mathscr{A}$ and which satisfies

(i) $\lambda(A) \geq 0$ and $\lambda(\emptyset) = 0$,
(ii) $\lambda(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \lambda(A_k)$ whenever $\{A_k\}$ is a countable collection of disjoint sets in $\mathscr{A}$ whose union also belongs to $\mathscr{A}$.

It follows easily that such a λ is monotone: if $A_1 \subset A_2$ and $A_1, A_2 \in \mathscr{A}$, then $\lambda(A_1) \leq \lambda(A_2)$.

A measure λ on $\mathscr{A}$ is called σ-*finite* (with respect to $\mathscr{A}$) if $\mathscr{S}$ can be written $\mathscr{S} = \bigcup S_k$ with $S_k \in \mathscr{A}$ and $\lambda(S_k) < +\infty$. For example, any Lebesgue–Stieltjes measure Λ on the line is a σ-finite measure on the algebra generated by the intervals $(a,b]$.

Using the ideas behind the construction of a Lebesgue–Stieltjes outer measure, we can construct an outer measure λ^* from λ. Thus, let λ be a measure

on an algebra $\mathscr{A}$ of subsets of $\mathscr{S}$. For any subset A of $\mathscr{S}$, define

$$\lambda^*(A) = \inf \sum \lambda(A_k), \tag{11.17}$$

where the infimum is taken over all countable collections $\{A_k\}$ such that $A_k \in \mathscr{A}$ and $A \subset \bigcup A_k$. It is always possible to find such a covering of A since $\mathscr{S}$ itself belongs to $\mathscr{A}$. The facts that $\mathscr{A}$ is an algebra and λ is monotone allow us to assume without loss of generality that the sets A_k covering A are disjoint since $\bigcup_{k\geq 1} A_k = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2 - A_1) \cup \cdots$.

Theorem 11.18 *Let λ be a measure on an algebra $\mathscr{A}$, and let λ^* be defined by (11.17). Then λ^* is an outer measure.*

The proof is similar to the first part of the proof of Theorem 11.8 and is left as an exercise.

While λ is assumed to be defined only on $\mathscr{A}$, λ^* is defined on every subset of $\mathscr{S}$. The next result shows that λ^* equals λ when restricted to $\mathscr{A}$.

Theorem 11.19 *Let λ be a measure on an algebra $\mathscr{A}$, and let λ^* be the corresponding outer measure. If $A \in \mathscr{A}$, then $\lambda^*(A) = \lambda(A)$ and A is measurable with respect to λ^*.*

Proof. Let $A \in \mathscr{A}$. Clearly, $\lambda^*(A) \leq \lambda(A)$. On the other hand, given disjoint $A_k \in \mathscr{A}$ with $A \subset \cup A_k$, let $A'_k = A_k \cap A$. Then $A'_k \in \mathscr{A}$ and A is the disjoint union of the A'_k. Hence, $\lambda(A) = \sum \lambda(A'_k)$. Since $A'_k \subset A_k$, it follows that $\lambda(A) \leq \sum \lambda(A_k)$. Therefore, $\lambda(A) \leq \lambda^*(A)$, and the proof of the first part of the theorem is complete.

For the second part, let $A \in \mathscr{A}$. To show that A is measurable with respect to λ^*, we must show that

$$\lambda^*(E) = \lambda^*(E \cap A) + \lambda^*(E - A) \text{ for every } E \subset \mathscr{S}.$$

Since λ^* is subadditive, the right side majorizes the left. To show the opposite inequality, we may assume that $\lambda^*(E)$ is finite. Given $\varepsilon > 0$, choose $\{E_k\}$ such that $E_k \in \mathscr{A}, E \subset \bigcup E_k$ and $\sum \lambda(E_k) < \lambda^*(E) + \varepsilon$. Since E_k and A are in $\mathscr{A}$ and $(E_k \cap A) \cup (E_k - A) = E_k$, we have $\lambda(E_k \cap A) + \lambda(E_k - A) = \lambda(E_k)$. Hence,

$$\sum \lambda(E_k \cap A) + \sum \lambda(E_k - A) < \lambda^*(E) + \varepsilon.$$

Therefore, since $E \cap A \subset \bigcup(E_k \cap A)$ and $E - A \subset \bigcup(E_k - A)$, it follows from the definition of λ^* that

$$\lambda^*(E \cap A) + \lambda^*(E - A) < \lambda^*(E) + \varepsilon.$$

Letting $\varepsilon \to 0$, we obtain the desired inequality, which completes the proof.

Let λ be a measure on an algebra $\mathscr{A}$, and let μ be a measure on a σ-algebra Σ that contains $\mathscr{A}$. Then μ is said to be an *extension of* λ *to* Σ if $\mu(A) = \lambda(A)$ for every $A \in \mathscr{A}$. If λ^* is the outer measure generated by λ and $\mathscr{A}^*$ denotes the σ-algebra of λ^*-measurable sets, it follows from the last theorem that λ^* is an extension of λ to $\mathscr{A}^*$. This proves the first part of the following theorem, which is the main result of this section.

Theorem 11.20 (Carathéodory–Hahn Extension Theorem) *Let λ be a measure on an algebra $\mathscr{A}$, let λ^* be the corresponding outer measure, and let $\mathscr{A}^*$ be the σ-algebra of λ^*-measurable sets.*

(i) *The restriction of λ^* to $\mathscr{A}^*$ is an extension of λ.*

(ii) *If λ is σ-finite with respect to $\mathscr{A}$, and if Σ is any σ-algebra with $\mathscr{A} \subset \Sigma \subset \mathscr{A}^*$, then λ^* is the only measure on Σ that is an extension of λ.*

Proof. As we have already observed, (i) follows from Theorem 11.19. To prove (ii), which states the uniqueness of the extension, let μ be any measure on $\Sigma, \mathscr{A} \subset \Sigma \subset \mathscr{A}^*$, which agrees with λ on $\mathscr{A}$. Given a set $E \in \Sigma$, consider any countable collection $\{A_k\}$ such that $E \subset \bigcup A_k$ and each $A_k \in \mathscr{A}$. Then

$$\mu(E) \leq \mu\left(\bigcup A_k\right) \leq \sum \mu(A_k) = \sum \lambda(A_k).$$

Therefore, by the definition of λ^*, we have $\mu(E) \leq \lambda^*(E)$. To show that equality holds, first suppose that there exists a set $A \in \mathscr{A}$ with $E \subset A$ and $\lambda(A) < +\infty$. Applying what has just been proved to $A - E$ (which belongs to Σ), we obtain $\mu(A - E) \leq \lambda^*(A - E)$. However,

$$\mu(E) + \mu(A - E) = \mu(A) = \lambda(A) = \lambda^*(A) = \lambda^*(E) + \lambda^*(A - E).$$

Since all these terms are finite (due to the fact that $\lambda(A)$ is finite), it follows that $\mu(E) \geq \lambda^*(E)$, so that $\mu(E) = \lambda^*(E)$ in this case.

In the general case, since λ is σ-finite, there exists disjoint $S_k \in \mathscr{A}$ such that $\mathscr{S} = \bigcup S_k$ and $\lambda(S_k) < +\infty$. We may apply the previous result to each $E \cap S_k$ (which is a subset of S_k), obtaining $\mu(E \cap S_k) = \lambda^*(E \cap S_k)$. By adding over k, we obtain $\mu(E) = \lambda^*(E)$, which completes the proof.

As a corollary, we can answer the questions raised at the beginning of this section.

Corollary 11.21 *Let μ and ν be two Borel measures on $\mathbf{R}^1$ which are finite and equal on every half-open interval $(a, b]$, $-\infty < a < b < +\infty$. Then $\mu(B) = \nu(B)$ for every Borel set $B \subset \mathbf{R}^1$.*

Proof. Such μ and ν must agree on the algebra generated by the $(a, b]$ and are σ-finite with respect to this algebra. Since the smallest σ-algebra containing all $(a, b]$ is the Borel σ-algebra, it follows from Theorem 11.20 that μ and ν are identical.

Although we have confined ourselves to $n = 1$, we note that an analogue of Corollary 11.21 can be formulated in higher dimensions. See Exercise 18.

Now let μ be any finite Borel measure on $\mathbf{R}^1$, and define f_μ by $f_\mu(x) = \mu((-\infty, x])$. Clearly, $\mu((a, b]) = f_\mu(b) - f_\mu(a)$. Moreover, if Λ_{f_μ} denotes the Lebesgue–Stieltjes measure constructed from f_μ, then since f_μ is continuous from the right, it follows from Theorem 11.10 that $\Lambda_{f_\mu}((a, b]) = f_\mu(b) - f_\mu(a)$. Therefore, by the previous corollary, μ and Λ_{f_μ} are identical as Borel measures, and we easily obtain

Corollary 11.22 *The class of finite Borel measures on $\mathbf{R}^1$ is identical with the class of Lebesgue–Stieltjes measures induced by bounded increasing functions that are continuous from the right.*

See also Exercise 4.

Let μ be a Borel measure on $\mathbf{R}^1$ which is finite on every $(a, b]$, $-\infty < a < b < +\infty$ (equivalently, μ is finite on every bounded Borel set). Consider the restriction of μ to the algebra $\mathscr{A}$ generated by the $(a, b]$, and let μ^* be the corresponding outer measure. The smallest σ-algebra containing $\mathscr{A}$ is the Borel sets $\mathscr{B}$. Thus, $\mathscr{A} \subset \mathscr{B} \subset \mathscr{A}^*$. Since μ and μ^* are measures on $\mathscr{B}$ that agree on $\mathscr{A}$, it follows that $\mu = \mu^*$ on $\mathscr{B}$. In particular, if $B \in \mathscr{B}$, we see from the definition of μ^* (see (11.17)) that

$$\mu(B) = \inf\left\{\sum \mu(A_k) : B \subset \bigcup A_k, A_k \in \mathscr{A}\right\}.$$

Each $A_k \in \mathscr{A}$ is a countable union of disjoint $(a, b]$. Hence, we obtain the formula

$$\mu(B) = \inf\left\{\sum \mu(a_k, b_k] : B \subset \bigcup (a_k, b_k]\right\}, \quad B \in \mathscr{B}. \tag{11.23}$$

We recall (see p. 269 in Section 10.5) that a Borel measure μ is said to be regular if

$$\mu(B) = \inf\{\mu(G) : B \subset G, G \text{ open}\}, \quad B \in \mathscr{B}.$$

Theorem 11.24 *If μ is a Borel measure on $\mathbf{R}^1$ which is finite on every bounded Borel set, then μ is regular.*

Proof. This is a corollary of (11.23). Given a Borel set B and $\varepsilon > 0$, find a cover $\{(a_k, b_k]\}$ of B such that

$$\sum \mu(a_k, b_k] \leq \mu(B) + \varepsilon.$$

Since μ is finite on bounded intervals, it follows from Theorem 10.11 that $\mu(a, b] = \lim_{\varepsilon \to 0+} \mu(a, b + \varepsilon)$. Hence, by slightly enlarging each $(a_k, b_k]$, we see that there is an open set $G, G = \bigcup(a_k, b_k + \varepsilon_k)$ for sufficiently small ε_k, containing B such that

$$\mu(G) \leq \sum \mu(a_k, b_k + \varepsilon_k) \leq \mu(B) + 2\varepsilon.$$

This completes the proof.

In Theorem 11.24, as in Corollary 11.21, we have limited ourselves to $n = 1$. For $n > 1$, see Exercise 18. In particular, note that the assumption in Section 10.5 on p. 269 concerning the regularity of μ and ν is redundant.

Exercises

1. (a) Prove the second statements in both parts of Corollary 11.4.
   (b) Verify the statements made before Theorem 11.5 about the function $r(A)$ defined on sets $A \subset \mathbf{R}^2$. (One way to see that a set B with $d(Y, B) > 0$ is not r-measurable is to denote the mirror reflection of B in the y-axis by B^* and check that the equation $r(B \cup B^*) = r(B) + r(B^*)$ is false.)
2. Let μ be a finite Borel measure on $\mathbf{R}^1$, and define $f_\mu(x) = \mu((-\infty, x])$, $-\infty < x < +\infty$. Show that f_μ is monotone increasing, $\mu((a, b]) = f_\mu(b) - f_\mu(a)$, f_μ is continuous from the right, and $\lim_{x \to -\infty} f_\mu(x) = 0$.
3. Let f be monotone increasing on $\mathbf{R}^1$.
   (a) Show that $\Lambda_f(\mathbf{R}^1)$ is finite if and only if f is bounded.

(b) Let f be bounded and right continuous, let $\mu = \Lambda_f$, and let $\bar{f}$ denote the function f_μ defined in Exercise 2. Show that f and $\bar{f}$ differ by a constant.
Thus, if we make the additional assumption that $\lim_{x\to-\infty} f(x) = 0$, then $f = \bar{f}$.

4. If we identify two functions on $\mathbf{R^1}$ which differ by a constant, prove that there is a one-to-one correspondence between the class of finite Borel measures on $\mathbf{R^1}$ and the class of bounded increasing functions that are continuous from the right.

5. Let f be monotone increasing and right continuous on $\mathbf{R^1}$.

(a) Show that Λ_f is absolutely continuous with respect to Lebesgue measure if and only if f is absolutely continuous on $\mathbf{R^1}$. (By absolutely continuous on $\mathbf{R^1}$, we mean absolutely continuous on every compact interval.)

(b) If Λ_f is absolutely continuous with respect to Lebesgue measure, show that its Radon–Nikodym derivative equals df/dx.

6. Prove that the Lebesgue–Stieltjes outer measure constructed from $f(x) = x$ is the same as Lebesgue outer measure.

7. If f is monotone increasing and continuous from the right on $\mathbf{R^1}$, show that $\Lambda_f^*(A) = \Lambda_f^{\circ *}(A)$, where $\Lambda_f^{\circ *}$ is defined in the same way as Λ_f^* except that we use *open* intervals (a_k, b_k).

8. If f is monotone increasing and continuous from the right, derive formulas for $\Lambda_f([a,b])$ and $\Lambda_f((a,b))$.

9. Complete the proof of Theorem 11.12.

10. Show that in $\mathbf{R^n}$, $n > 1$, the Hausdorff outer measure H_n is not identical to Lebesgue outer measure. (For example, let $n = 2$, and write $A = \bigcup A_k$, $\delta(A_k) < \varepsilon$. Enclose A_k in a circle C_k with the same diameter, and show that $\sum \delta(A_k)^2 \geq (4/\pi)|A|_e$. Thus, $H_2^\varepsilon(A) \geq (4/\pi)|A|_e$.)

11. If A is a subset of $\mathbf{R^n}$, define the *Hausdorff dimension* of A as follows: If $H_\alpha(A) = 0$ for all $\alpha > 0$, let $\dim A = 0$; otherwise, let

$$\dim A = \sup\{\alpha : H_\alpha(A) = +\infty\}.$$

(a) Show that $H_\alpha(A) = 0$ if $\alpha > \dim A$ and that $H_\alpha(A) = +\infty$ if $\alpha < \dim A$. Show that in $\mathbf{R^n}$ we have $\dim A \leq n$. See Exercise 19 in order to determine the Hausdorff dimension of the Cantor set.

(b) If $\dim A_k = d$ for each A_k in a countable collection $\{A_k\}$, show that $\dim(\bigcup A_k) = d$. Hence, show that every countable set has Hausdorff dimension 0.

12. Let Γ be an outer measure on $\mathscr{S}$, and let Γ' denote Γ restricted to the Γ-measurable sets. Since Γ' is a measure on an algebra, it induces an outer measure Γ^*. Show that $\Gamma^*(A) \geq \Gamma(A)$ for $A \subset \mathscr{S}$ and that equality holds

for a given A if and only if there is a Γ-measurable set E such that $A \subset E$ and $\Gamma(E) = \Gamma(A)$. Thus, $\Gamma = \Gamma^*$ if Γ is regular.

13. Let λ be a measure on an algebra $\mathscr{A}$, and let λ^* be the corresponding outer measure. Given A, show that there is a set H of the form $\bigcap_k \bigcup_j A_{k,j}$ such that $A_{k,j} \in \mathscr{A}$, $A \subset H$ and $\lambda^*(A) = \lambda^*(H)$. Thus, every outer measure that is induced by a measure on an algebra is regular.

14. Prove Theorem 11.18.

15. (a) Show that the intersection of a family of algebras is an algebra.

(b) A collection $\mathscr{C}$ of subsets of $\mathscr{S}$ is called a *subalgebra* if it is closed under finite intersections and if the complement of any set in $\mathscr{C}$ is the union of a finite number of disjoint sets in $\mathscr{C}$. Give an example of a subalgebra. Show that a subalgebra $\mathscr{C}$ generates an algebra by adding $\emptyset$, $\mathscr{S}$, and all finite disjoint unions of sets of $\mathscr{C}$.

16. If μ is a finite Borel measure on $\mathbf{R^1}$, show that $\mu(B) = \sup \mu(F)$ for every Borel set B, where the sup is taken over all closed subsets F of B.

17. Show that the conclusions of Theorems 10.48 and 10.49, and therefore also the conclusion of Corollary 10.50, remain true without the assumption (ii) stated before Lemma 10.47. (Show that without this assumption, the conclusions of Lemma 10.47 are true with μ replaced by μ^*; for example,

$$\mu^* \left\{ \mathbf{x} \in E : \sup_{h>0} \frac{\nu(Q_{\mathbf{x}}(h))}{\mu(Q_{\mathbf{x}}(h))} > \alpha \right\} \leq c \, \frac{\nu(\mathbf{R^n})}{\alpha}. \Bigg)$$

18. Derive analogues of Corollary 11.21 and Theorem 11.24 in $\mathbf{R^n}$, $n > 1$. (Use partly open n-dimensional intervals in place of the intervals $(a, b]$.)

19. Show that the Cantor set C in $[0, 1]$ has the Hausdorff measure of order $\log 2/\log 3$ equal to 1. (In order to show the inequality $H_{\log 2/\log 3}(C) \leq 1$, consider $H^{(\varepsilon)}_{\log 2/\log 3}(C)$ when $\varepsilon = 3^{-k}$, $k = 1, 2, \ldots$, and show that the 2^k intervals $\{I_j\}$ of length 3^{-k} remaining at the kth stage C_k of construction of C satisfy $\sum |I_j|^\alpha = 2^k 3^{-k\alpha} = 1$ if $\alpha = \log 2/\log 3$. A proof of the opposite inequality is harder. It may be helpful to note that if I is a closed interval, containing at its two endpoints intervals from C_k, then $|I|^\alpha \geq n_k(I)/2^k$, where $n_k(I)$ is the number of intervals of C_k contained in I and $\alpha = \log 2/\log 3$.)

12

A Few Facts from Harmonic Analysis

12.1 Trigonometric Fourier Series

The Lebesgue measure and integration have been decisive in the development of many branches of analysis and are applied there in ever greater degree. But, conversely, some of the applications have had considerable impact on the theory of integration. In this chapter, we will consider one topic where this interdependence has been particularly fruitful: *harmonic analysis* (see p. 305 in Section 12.1).

One of the principal goals of harmonic analysis, which is a vast field, is to represent very general functions f in terms of a collection of simpler oscillatory ones. The fact that typical representations involve integration of f accounts for the interrelation of the two fields. Oscillatory behavior of the simpler functions has advantages: it can help make them independent of one another, and it can be exploited in order to find particular linear combinations of them that approximate general functions.

We begin by describing some elementary notions and facts; the concept of an orthogonal system, and in particular of the trigonometric system, is basic here.

The notion of an orthogonal system, defined generally on a subset E of positive measure in $\mathbf{R}^n$, was introduced in Chapter 8, and we refer the reader to that place for the definitions and properties of general orthogonal systems, restating only a few facts here.

A system of complex-valued functions $\{\phi_\alpha(\mathbf{x})\}$, all belonging to $L^2(E)$, is called *orthogonal* over E if

$$\langle \phi_\alpha, \phi_\beta \rangle = \int_E \phi_\alpha \overline{\phi_\beta} \begin{cases} = 0 & \alpha \neq \beta \\ > 0 & \alpha = \beta. \end{cases}$$

The second condition means that $\phi_\alpha \not\equiv 0$. If $\langle \phi_\alpha, \phi_\alpha \rangle = 1$ for all α, the orthogonal system is called *normal*, or *orthonormal*. If $\{\phi_\alpha\}$ is orthogonal, the system $\{\phi_\alpha / ||\phi_\alpha||_2\}$ is orthonormal. Thus, by merely multiplying the functions of an orthogonal system by suitable constants, we can *normalize* it, and formulas

for orthonormal systems can be easily and automatically extended to general, not necessarily normal, orthogonal systems. On the other hand, certain important orthogonal systems very naturally appear, often for historical reasons, in a nonnormalized form, and because of this, it may be desirable not to insist on the normality of the system under consideration. Let us, therefore, briefly restate the definitions in this somewhat more general setup.

Since orthogonal systems are countable (see Theorem 8.21), we may index them by integers. Let $\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \ldots$ be an orthogonal system on $E \subset \mathbf{R^n}$. Thus,

$$\int_E \phi_k \overline{\phi_l} = \begin{cases} 0 & k \neq l. \\ \lambda_k > 0 & k = l. \end{cases}$$

Given any (complex-valued) $f \in L^2(E)$, we call the numbers

$$c_k = \frac{1}{\lambda_k} \int_E f \overline{\phi_k}$$

the *Fourier coefficients* of f and the series $S[f] = \sum c_k \phi_k(\mathbf{x})$ the *Fourier series* of f, with respect to $\{\phi_k\}$. As before, we write

$$f \sim \sum c_k \phi_k(\mathbf{x}).$$

If we set

$$\psi_k = \lambda_k^{-1/2} \phi_k, \quad d_k = \lambda_k^{1/2} c_k = \int_E f \overline{\psi_k}, \tag{12.1}$$

then $\{\psi_k\}$ is orthonormal over E, and $\{d_k\}$ is the sequence of Fourier coefficients of f with respect to $\{\psi_k\}$. Clearly,

$$d_k \psi_k = c_k \phi_k. \tag{12.2}$$

This set of formulas enables us to rewrite relations for orthonormal systems in forms valid for general orthogonal systems. Thus, Bessel's inequality $\sum |d_k|^2 \leq \int_E |f|^2$ and Parseval's formula $\sum |d_k|^2 = \int_E |f|^2$ take the forms

$$\sum \lambda_k |c_k|^2 \leq \int_E |f|^2, \quad \sum \lambda_k |c_k|^2 = \int_E |f|^2. \tag{12.3}$$

The notion of completeness of an orthogonal system ("the vanishing of all the Fourier coefficients implies the vanishing of the function") remains unchanged in the general case, and as in the case of normalized systems, the

validity of the second formula in (12.3) is a necessary and sufficient condition for the completeness of $\{\phi_k\}$.

Let s_n denote the nth partial sum of $S[f]$. As a corollary of the corresponding result for orthonormal systems, we see that the equation $\sum \lambda_k |c_k|^2 = \int_E |f|^2$ is equivalent to

$$\int_E |f - s_n|^2 \to 0.$$

Thus, if an orthogonal system is complete, the Fourier series of every $f \in L^2(E)$ converges to f, convergence being understood in the metric L^2. Of course, this says nothing about the pointwise convergence of $\sum c_k \phi_k(\mathbf{x})$. On the other hand, it holds for any rearrangement of the terms of $\sum c_k \phi_k$ since the orthogonality and completeness of a system are not affected by a permutation of the functions within the system.

We shall now consider a special orthogonal system, the *trigonometric system*. This name is given to the system of functions

$$e^{ikx} = \cos kx + i \sin kx, \ x \in (-\infty, \infty) \quad (k = 0, \ \pm 1, \ \pm 2, \ldots).$$

These functions are all periodic, with period 2π, and it is immediate that they form an orthogonal system over any interval $Q = (a, a + 2\pi)$ of length 2π, since if k and m are distinct integers, then

$$\int_Q e^{ikx} \overline{e^{imx}} \, dx = \int_Q e^{i(k-m)x} \, dx = \left[\frac{e^{i(k-m)x}}{k-m} \right]_a^{a+2\pi} = 0.$$

The system is not orthonormal since $\lambda_k = \int_Q |e^{ikx}|^2 dx = 2\pi$ for all k. Thus, with any $f \in L(Q)$, we may associate its Fourier coefficients

$$c_k = \frac{1}{2\pi} \int_Q f(t) \, \overline{e^{ikt}} dt = \frac{1}{2\pi} \int_Q f(t) e^{-ikt} dt \quad (k = 0, \pm 1, \pm 2, \ldots), \tag{12.4}$$

and its Fourier series

$$f \sim \sum_{-\infty}^{+\infty} c_k e^{ikx}. \tag{12.5}$$

In what follows, this series will be designated by $S[f]$, and its coefficients by $c_k[f]$.

Observe that if two functions ϕ and ψ are orthogonal over a set E and if $\int_E |\phi|^2 = \int_E |\psi|^2$, then the pair $\phi \pm \psi$ is also orthogonal over E, as seen from the equation

$$\int_E (\phi + \psi)(\overline{\phi} - \overline{\psi}) = \int_E |\phi|^2 - \int_E |\psi|^2 = 0.$$

Applying this to the pairs $e^{\pm ikx}$ $(k = 1, 2, \ldots)$, we see that the functions

$$\frac{1}{2}, \ldots, \frac{e^{ikx} + e^{-ikx}}{2}, \frac{e^{ikx} - e^{-ikx}}{2i}, \ldots (k = 1, 2, \ldots) \tag{12.6}$$

or, what is the same thing, the functions

$$\frac{1}{2}, \cos x, \sin x, \ldots, \cos kx, \sin kx, \ldots \tag{12.7}$$

are orthogonal over any interval Q of length 2π. Using the form (12.6), we find that the numbers λ_k for (12.7) are

$$\frac{1}{2}\pi, \pi, \pi, \ldots.$$

Thus, any $f \in L(Q)$ can be developed into a new Fourier series

$$f \sim \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx), \tag{12.8}$$

where

$$\begin{aligned} a_0 &= \left(\frac{1}{2}\pi\right)^{-1} \int_Q \frac{1}{2} f = \frac{1}{\pi}\int_Q f, \\ a_k &= \frac{1}{\pi}\int_Q f(t)\cos kt\, dt, \qquad b_k = \frac{1}{\pi}\int_Q f(t)\sin kt\, dt. \end{aligned} \tag{12.9}$$

The numbers a_k and b_k are easily expressible in terms of the coefficients c_k of (12.4):

$$\frac{1}{2}a_0 = c_0, \quad a_k = c_k + c_{-k}, \quad b_k = i(c_k - c_{-k}). \tag{12.10}$$

Hence,

$$\sum_{k=-n}^{n} c_k e^{ikx} = c_0 + \sum_{k=1}^{n} c_k e^{ikx} + \sum_{k=1}^{n} c_{-k} e^{-ikx}$$
$$= \frac{1}{2}a_0 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx),$$

and the nth partial sum of the series in (12.8) turns out to be the nth *symmetric* partial sum of $\sum c_k e^{ikx}$. The numbers $a_k = a_k[f]$ and $b_k = b_k[f]$ are called the *Fourier cosine* and *sine coefficients* of f, respectively.

To sum up, we may consider the trigonometric system in two forms. One consists of the functions e^{ikx} $(k = 0, \pm 1, \pm 2, \ldots)$, and the Fourier series has the form $\sum_{-\infty}^{+\infty} c_k e^{ikx}$, where the c_k are given by (12.4). The other consists of the functions (12.7), the Fourier series is (12.8), and the coefficients a_k and b_k are given by (12.9). The partial sums of (12.8) are the symmetric partial sums of (12.5). In both cases, the terms of the Fourier series are *harmonic oscillations*, and for this reason, the study of $S[f]$ is called the *harmonic analysis* of f.

Each form of the trigonometric system has its advantages. For example, if f is real-valued, then the numbers a_k and b_k are real, while the c_k have the property $c_{-k} = \overline{c_k}$. Note also that if $Q = (-\pi, \pi)$ and f is an *even function*, that is, if $f(-x) = f(x)$, then

$$a_k = \frac{2}{\pi}\int_0^{\pi} f(t)\cos kt\, dt, \quad b_k = 0,$$

and if f is an *odd function*, that is, if $f(-x) = -f(x)$, then

$$a_k = 0, \quad b_k = \frac{2}{\pi}\int_0^{\pi} f(t)\sin kt\, dt.$$

Thus, if f is even, (12.8) reduces to the cosine series $\frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k \cos kx$, and if f is odd, to the sine series $\sum_{k=1}^{\infty} b_k \sin kx$.

Since the terms of a (trigonometric) Fourier series are periodic with period 2π, if we expect to represent a function f by its Fourier series, we may assume from the start that f is defined everywhere (or almost everywhere [a.e.]) on the real axis and is periodic with period 2π. This amounts to considering the function as defined on the circumference of the unit circle. We do not distinguish between points that are congruent mod 2π. By an integrable function, we shall mean a function integrable over a period. Similarly, the L^p norm of a function will mean its L^p norm over a period and the familiar definitions

of other classes of function such as functions of bounded variation and Lipschitz continuous functions will also be restricted to a period. In what follows, *periodic* will mean periodic of period 2π.

In the preceding chapters, we proved a number of theorems about functions in $L^p(\mathbf{R^n})$ and, in particular, in $L^p(\mathbf{R^1})$. Usually, these results have analogues for periodic functions, where integrals over $\mathbf{R^1}$ are replaced by integrals over a period. The proofs are usually in essence identical with those for $\mathbf{R^1}$ (or are merely corollaries of the results for $\mathbf{R^1}$) and may be left as exercises.

We would like to stress one point. The definition of a general orthogonal system presupposes that the functions in the system are of class L^2. This makes it possible to define Fourier coefficients for any $f \in L^2$. If f is not in L^2, it may be impossible to define its Fourier coefficients with respect to certain orthogonal systems. The situation is different for special orthogonal systems. For example, in the case of the functions $\{e^{ikx}\}$, which are bounded, the coefficients c_k are defined for any f that is merely integrable over Q and, in particular, for any $f \in L^p(Q), 1 \le p \le \infty$. Thus, the trigonometric system is richer in properties than general orthogonal systems.

Some simple developments are important for the general theory of Fourier series. We consider two here and refer the reader to Exercise 5 for others.

Example (a). Let f be periodic and equal to $\frac{1}{2}(\pi - x)$ for $0 < x < 2\pi$, with $f(0) = f(2\pi) = 0$. Since f is odd, its Fourier series is a sine series, and integration by parts gives

$$b_k = \frac{2}{\pi}\int_0^\pi \frac{1}{2}(\pi - x)\sin kx\,dx = \frac{1}{\pi}\left(\frac{\pi}{k} - \frac{1}{k}\int_0^\pi \cos kx\,dx\right) = \frac{1}{k}.$$

Thus,

$$f \sim \sum_{k=1}^{\infty} \frac{\sin kx}{k} = \frac{1}{2}\sum_{-\infty}^{+\infty}{}' \frac{e^{ikx}}{ik},$$

where $\sum'$ denotes $\sum_{k\neq 0}$.

Example (b). Let f be periodic and equal in $(-\pi, \pi)$ to the characteristic function of the interval $(-h, h), 0 < h < \pi$. Then f is even, and if $k \neq 0$, its cosine coefficient is

$$a_k = \frac{2}{\pi}\int_0^h \cos kx\,dx = \frac{2}{\pi}\frac{\sin kh}{k}.$$

Since $a_0 = 2h/\pi$, we obtain

$$f \sim \frac{2h}{\pi}\left\{\frac{1}{2} + \sum_{k=1}^{\infty} \frac{\sin kh}{kh} \cos kx\right\}$$
$$= \frac{h}{\pi}\left\{1 + \sum_{-\infty}^{+\infty}{}' \frac{\sin kh}{kh} e^{ikx}\right\}.$$

Series of the form

$$\sum_{-\infty}^{+\infty} c_k e^{ikx}, \qquad \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx),$$

whether they are Fourier series or not, are called *trigonometric* series. In defining the convergence of $\sum_{-\infty}^{+\infty} c_k e^{ikx}$, we usually consider the limit, ordinary or generalized, of the symmetric partial sums $\sum_{-n}^{+n}$, and once again

$$\sum_{-n}^{n} c_k e^{ikx} = \frac{1}{2}a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx),$$

where

$$\frac{1}{2}a_0 = c_0, \quad a_k = c_k + c_{-k}, \quad b_k = i\left(c_k - c_{-k}\right).$$

A finite sum $T = \sum_{-n}^{n} c_k e^{ikx}$ is called a *trigonometric polynomial of order n*, and if $|c_{-n}| + |c_n| \neq 0$, T is *strictly of order n*. If T is of order n and vanishes at more than $2n$ distinct points (i.e., distinct mod 2π), then it vanishes identically, that is, all the c_k are 0. In fact, $Te^{inx} = \sum_{-n}^{n} c_k e^{i(k+n)x}$ is an algebraic (power) polynomial in $z = e^{ix}$ of degree $\leq 2n$, and if it has more than $2n$ zeros, then it vanishes identically.

If the numbers a_k and b_k are real, the trigonometric series

$$S = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

is the real part of the power series

$$\frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k - ib_k)\, z^k$$

on the unit circle $z = e^{ix}$. The imaginary part is then the series

$$\sum_{k=1}^{\infty} (a_k \sin kx - b_k \cos kx) \tag{12.11}$$

(with vanishing constant term). If S is written in the complex form $\sum_{-\infty}^{+\infty} c_k e^{ikx}$, it is easy to see that (12.11) is

$$\sum_{-\infty}^{+\infty} (-i \operatorname{sign} k) c_k e^{ikx} \tag{12.12}$$

(where, by convention, sign 0 = 0). The series (12.11), or (12.12), is said to be *conjugate* to S. A series conjugate to a trigonometric series S is denoted by $\widetilde{S}$. If S has constant term 0, then

$$\widetilde{\widetilde{S}} = -S.$$

It is natural to study the properties of $\widetilde{S}[f]$ simultaneously with those of $S[f]$.

One more remark. Properties of functions in $L^p(\mathbf{R^n})$, and in particular in $L^p(\mathbf{R^1})$, are important for the theory of *Fourier integrals*, which for non-periodic functions play the same role as Fourier series in the periodic case. The two theories run largely parallel. In this chapter, we shall limit ourselves to Fourier series since our primary aim is to show the role that Lebesgue integration plays in some problems of representability of functions, and both the results and techniques of Fourier series are sufficiently indicative of the situation. In Chapter 13, we will study the main aspects of Fourier integrals in $\mathbf{R^n}$, $n \geq 1$.

12.2 Theorems about Fourier Coefficients

Theorem 12.13 *If a periodic f is the indefinite integral of its derivative f' (i.e., if f is periodic and absolutely continuous), and if $f \sim \sum c_k e^{ikx}$, then*

$$f' \sim \sum_{-\infty}^{+\infty} c_k (ik) e^{ikx}.$$

In symbols,

$$S[f'] = S'[f],$$

where $S'[f]$ denotes the result of the termwise differentiation of $S[f]$.

Proof. It is clear that f' is also periodic and that its constant term equals

$$(2\pi)^{-1}\int_0^{2\pi} f'(x)\,dx = (2\pi)^{-1}[f(2\pi) - f(0)] = 0.$$

If $k \neq 0$, integrating by parts and observing that the integrated term is zero, we have

$$(2\pi)^{-1}\int_0^{2\pi} f'(x)e^{-ikx}\,dx = (2\pi)^{-1}ik\int_0^{2\pi} f(x)e^{-ikx}dx = ikc_k,$$

which proves the theorem.

By repeated application of this result, we see that if a periodic f is the mth indefinite integral of an integrable function $f^{(m)}$, then

$$S\left[f^{(m)}\right] = S^{(m)}[f] = \sum (ik)^m c_k e^{ikx}.$$

Theorem 12.14 *If f is periodic, $f \sim \sum c_k e^{ikx}$, and if F is the indefinite integral of f, then $F(x) - c_0 x$ is periodic and*

$$F(x) - c_0 x \sim C_0 + \sum{}' \frac{c_k}{ik} e^{ikx},$$

where C_0 is a suitable constant (depending on the choice of the arbitrary constant of integration in F) and $\sum'$ again denotes $\sum_{k\neq 0}$.

Proof. Let $G(x) = F(x) - c_0 x$. The periodicity of G follows from the equation

$$G(x + 2\pi) - G(x) = \int_x^{x+2\pi} f\,dt - c_0(2\pi) = 2\pi c_0 - 2\pi c_0 = 0.$$

Since G is also absolutely continuous, Theorem 12.13 gives

$$S'[G] = S[G'] = S[f - c_0] = \sum_{k\neq 0} c_k e^{ikx}.$$

Hence, $S[G]$ is obtained by termwise integration of $\sum' c_k e^{ikx}$, which leads to the result, C_0 being the constant term of $S[F - c_0 x]$.

For the trigonometric system, we have *Bessel's inequality* (see (12.3))

$$\sum_{-\infty}^{+\infty} |c_k|^2 \leq \frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 \, dx,$$

but actually, as we shall see, we also have *Parseval's formula*

$$\sum_{-\infty}^{+\infty} |c_k|^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 \, dx \tag{12.15}$$

for every $f \in L^2$. We know by Theorem 8.31 that this is a corollary of the next result.

Theorem 12.16 *The trigonometric system is complete. More precisely, if all the Fourier coefficients of an integrable f are zero, then $f = 0$ a.e.*

Proof. Assume first that f is continuous and real-valued, with all $c_k = 0$. If $f \not\equiv 0$, then $|f|$ attains a nonzero maximum M at some point x_0. Suppose, for example, that $f(x_0) = M > 0$. Let $\delta > 0$ be so small that $f(x) > \frac{1}{2}M$ in the interval $I = (x_0 - \delta, x_0 + \delta)$. Consider the trigonometric polynomial

$$t(x) = 1 + \cos(x - x_0) - \cos\delta.$$

It is strictly greater than 1 inside I and does not exceed 1 in absolute value elsewhere. The hypothesis that all the Fourier coefficients of f are 0 implies that $\int_{-\pi}^{\pi} f\, T\, dx = 0$ for any trigonometric polynomial T, and in particular,

$$\int_{-\pi}^{\pi} f\, t^N dx = 0 \quad (N = 1, 2, \ldots).$$

We claim that this is impossible for N large enough. The absolute value of the part of the last integral extended over the complement of I is $\leq 2\pi \cdot M \cdot 1^N = 2\pi M$. If I' is the middle half of I, then $t(x) \geq \theta > 1$ in I', so that

$$\int_I f\, t^N \, dx \geq \int_{I'} f\, t^N \, dx \geq \frac{1}{2} M \cdot \theta^N |I'| \to +\infty.$$

Collecting results, we see that $\int_{-\pi}^{\pi} f\, t^N dx \to +\infty$; this contradiction shows that $f \equiv 0$.

If f is continuous but not real-valued, the hypothesis $\int_0^{2\pi} f(x) e^{-ikx} dx = 0$ for all k implies that $\int_0^{2\pi} \overline{f(x)}\, e^{-ikx} dx = 0$ for all k. By adding and subtracting

the last two equations, we see that both the real and imaginary parts of f have all their Fourier coefficients equal to 0 and so vanish identically.

Finally, if f is merely integrable, the hypothesis $c_0 = 0$ implies that the function $F(x) = \int_0^x f\,dt$ is periodic, and by Theorem 12.14, for a suitable C_0, the Fourier coefficients of the continuous function $F - C_0$ are all 0. Hence, $F - C_0 \equiv 0$, F is a constant, and $f = F' = 0$ a.e. This completes the proof of the theorem. Another proof is given on p. 340 in Section 12.6.

An immediate corollary is the following result.

Theorem 12.17 *Parseval's formula (12.15) holds for any $f \in L^2$.*

Parseval's formula can be written in more general forms, which are, however, corollaries of (12.15). Thus, besides $f \sim \sum c_k e^{ikx} \in L^2$, consider another function $g \sim \sum d_k e^{ikx} \in L^2$. Then, by an argument like that in Section 8.6 for (8.32),

$$\frac{1}{2\pi}\int_0^{2\pi} f\,\overline{g}\,dx = \sum_{-\infty}^{+\infty} c_k\,\overline{d_k}. \tag{12.18}$$

This reduces to (12.15) in the special case $f = g$.

The completeness of the trigonometric system also gives the next two theorems.

Theorem 12.19 *If the Fourier series of a continuous f converges uniformly, then the sum of the series is f.*

Proof. Let g be the sum of the uniformly convergent series $S[f]$. The Fourier coefficients of g can be obtained by multiplying $S[f]$ by e^{-ikx} and integrating the result termwise. Thus, $c_k[g] = c_k[f]$ for all k, so that $f \equiv g$.

Theorem 12.20 *If a periodic f is the integral of a function in L^2, then $S[f]$ converges absolutely and uniformly. In particular, the Fourier series of a continuously differentiable function converges uniformly to the function.*

Proof. Let f be the integral of $g \in L^2$, $g \sim \sum c_k e^{ikx}$ ($c_0 = 0$). Then $S[f] = C_0 + \sum' C_k e^{ikx}$, $C_k = c_k/ik, k \neq 0$. We have $\sum |c_k|^2 < +\infty$ by Bessel's inequality, so that $\sum |C_k| < +\infty$ by Schwarz's inequality. This completes the proof.

The theorem that follows is of basic importance.

Theorem 12.21 (Riemann–Lebesgue) *The Fourier coefficients c_k of any integrable f tend to 0 as $k \to \pm\infty$. Hence, also $a_k, b_k \to 0$ as $k \to +\infty$.*

Proof. First, we note the obvious but important inequality

$$|c_k[f]| \le \frac{1}{2\pi} \int_0^{2\pi} |f|\, dx.$$

We will give two proofs of the theorem.

(a) (See also Exercise 15 of Chapter 8.) If $f \in L^2$, then $c_k \to 0$ as a corollary of Bessel's inequality (p. 310, Section 12.2). If $f \in L$ and $\varepsilon > 0$, write $f = g + h$, where $g \in L^2$ and $\int_0^{2\pi} |h| < \varepsilon$. (This decomposition can be made in various ways: we may, for example, take M large enough and define h to be f wherever $|f| \ge M$ and 0 elsewhere; clearly, $|g| \le M$, and so $g \in L^2$.) Then

$$c_k[f] = c_k[g] + c_k[h].$$

Since $c_k[g] \to 0$ and $|c_k[h]| \le (2\pi)^{-1} \int_0^{2\pi} |h| < \varepsilon/2\pi$ for all k, the relation $c_k[f] \to 0$ follows.

(b) Observe that

$$c_k[f] = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx}\, dx = \frac{1}{2\pi} \int_{-(\pi/k)}^{2\pi-(\pi/k)} f\left(x + \frac{\pi}{k}\right) e^{-ik(x+(\pi/k))} dx$$

$$= -\frac{1}{2\pi} \int_0^{2\pi} f\left(x + \frac{\pi}{k}\right) e^{-ikx}\, dx.$$

Taking the semi-sum of the first and third integrals, we obtain

$$c_k[f] = \frac{1}{4\pi} \int_0^{2\pi} \left[f(x) - f\left(x + \frac{\pi}{k}\right)\right] e^{-ikx}\, dx,$$

$$|c_k[f]| \le \frac{1}{4\pi} \int_0^{2\pi} \left|f(x) - f\left(x + \frac{\pi}{k}\right)\right| dx.$$

However, we know that the last integral tends to 0 as $k \to \pm\infty$ (cf. Theorem 8.19; the analogous result for periodic functions is left to the reader). This completes the proof.

Given any finite periodic f, the expression

$$\sup_{x,h;|h|\le\delta} |f(x+h)-f(x)| \qquad (\delta > 0)$$

is called the *modulus of continuity* of f and denoted by $\omega(\delta)$ or $\omega(\delta,f)$ (cf. Exercise 17 of Chapter 1). If f is in $L^p, 1 \le p < \infty$, the expression

$$\sup_{|h|\le\delta} \left[\frac{1}{2\pi} \int_0^{2\pi} |f(x+h)-f(x)|^p \, dx \right]^{1/p}$$

is called the L^p *modulus of continuity* of f and denoted $\omega_p(\delta,f)$ or simply $\omega_p(\delta)$. Clearly, $\omega_p(\delta) \le \omega(\delta)$ and, as is easily seen from Hölder's inequality,

$$\omega_p(\delta) \le \omega_q(\delta) \quad \text{if } p \le q.$$

We know that if $f \in L^p$, then $\omega_p(\delta,f) \to 0$ with δ (Theorem 8.19). The last inequality in proof (b) of Theorem 12.21 gives

$$|c_k[f]| \le \frac{1}{2}\omega\left(\frac{\pi}{|k|},f\right), \qquad |c_k[f]| \le \frac{1}{2}\omega_1\left(\frac{\pi}{|k|},f\right). \tag{12.22}$$

These two inequalities contain the Riemann–Lebesgue theorem in a sharp form since they *quantitatively* estimate the magnitude of the Fourier coefficients of f in terms of various moduli of continuity.

The estimates (12.22) are also useful for families of functions. The following special case deserves a separate mention. A continuous periodic f is said to satisfy a *Lipschitz condition of order* α, $0 < \alpha \le 1$, if $\omega(\delta,f) = O(\delta^\alpha)$ or, equivalently, if there is a finite constant M independent of x, h such that

$$|f(x+h)-f(x)| \le M|h|^\alpha.$$

Theorem 12.23 *If f satisfies a Lipschitz condition of order $\alpha, 0 < \alpha < 1$, then*

$$|c_k[f]| = O(|k|^{-\alpha}).$$

If $\alpha = 1$, the stronger estimate

$$|c_k[f]| = o\left(\frac{1}{|k|}\right)$$

is valid.

Proof. The first part follows from the first inequality (12.22). If $\alpha = 1$, then f is absolutely continuous (see p. 150 in Section 7.5) and so equals the indefinite integral of its derivative f'. Since f' is bounded (and so is in L^2), its Fourier coefficients tend to zero. Hence, the coefficients of f are $o(1/|k|)$ by Theorem 12.13.

Theorem 12.24 *If a periodic f is of bounded variation over a period, then $|c_k[f]| = O(1/|k|)$. More precisely,*

$$|c_k| \leq \frac{V}{2\pi|k|},$$

where V is the total variation of f over a closed interval of length 2π.

Proof. Integrating by parts and taking account of the periodicity of f, we have

$$2\pi c_k[f] = \int_{-\pi}^{\pi} e^{-ikx} f(x)\, dx = \frac{1}{ik} \int_{-\pi}^{\pi} e^{-ikx}\, df(x),$$

where the last integral is a Riemann-Stieltjes integral. Hence,

$$2\pi|c_k[f]| \leq |k|^{-1} \int_{-\pi}^{\pi} |df(x)| = |k|^{-1} V.$$

It must be stressed that V is the total variation over a *closed* interval of periodicity.

12.3 Convergence of $S[f]$ and $\widetilde{S}[f]$

We shall now briefly discuss the problem of pointwise convergence of $S[f]$, treating side-by-side the parallel problem for $\widetilde{S}[f]$. Among many existing results, we will consider only the simplest. Without loss of generality, we may restrict our attention to real-valued f.

We begin by computing the partial sums of $S[f]$ and $\widetilde{S}[f]$. If a_k and b_k denote the cosine and sine coefficients of f and $n = 1, 2, \ldots$, then the nth partial sum of $S[f]$ is

$$
\begin{aligned}
s_n(x) &= \frac{1}{2}a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx) \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\, dt + \sum_{k=1}^{n} \left\{ \cos kx \cdot \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos kt\, dt + \sin kx \right. \\
&\quad \left. \cdot \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin kt\, dt \right\} \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^{n} \cos k(x-t) \right] dt = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) D_n(x-t)\, dt,
\end{aligned}
$$

where

$$
D_n(t) = \frac{1}{2} + \sum_{k=1}^{n} \cos kt.
$$

In case $n = 0$, we denote $s_0(x) = \frac{1}{2}a_0$ and $D_0(t) = \frac{1}{2}$. The trigonometric polynomial D_n is called the nth *Dirichlet kernel*. Similarly, if $n = 1, 2, \ldots$, the nth partial sum of $\widetilde{S}[f]$ is

$$
\begin{aligned}
\widetilde{s}_n(x) &= \sum_{k=1}^{n} (a_k \sin kx - b_k \cos kx) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left\{ \sum_{k=1}^{n} \sin k(x-t) \right\} dt \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \widetilde{D}_n(x-t)\, dt,
\end{aligned}
$$

where

$$
\widetilde{D}_n(t) = \sum_{k=1}^{n} \sin kt
$$

is the nth *conjugate Dirichlet kernel*. It will be convenient to define $\widetilde{s}_0(x) = 0$ and $\widetilde{D}_0(t) = 0$. Notice that D_n and $\widetilde{D}_n$ are even and odd functions of t, respectively, and that

$$
\frac{1}{\pi} \int_{-\pi}^{\pi} D_n(t)\, dt = 1, \quad \frac{1}{\pi} \int_{-\pi}^{\pi} \widetilde{D}_n(t)\, dt = 0. \tag{12.25}
$$

Moreover, D_n and $\widetilde{D}_n$ are, respectively, the real and imaginary parts of

$$\frac{1}{2}+\sum_{k=1}^{n} e^{ikt}=\frac{1}{2}+\sum_{k=1}^{n} z^k=\frac{1}{2}+\frac{z^{n+1}-z}{z-1} \qquad \left(z=e^{it},\, n\geq 1\right),$$

this expression being interpreted as $\frac{1}{2}+n$ when $t=0$. Then an elementary computation gives (even if $n=0$)

$$D_n(t)=\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{1}{2}t}, \quad \widetilde{D}_n(t)=\frac{\cos\frac{1}{2}t-\cos\left(n+\frac{1}{2}\right)t}{2\sin\frac{1}{2}t}, \tag{12.26}$$

with $D_n(0)=n+\frac{1}{2}$ and $\widetilde{D}_n(0)=0$.

A quicker, though somewhat artificial, method of obtaining the first formula in (12.26) is to multiply $D_n(t)$ termwise by $2\sin\frac{1}{2}t$, replace the products $2\sin\frac{1}{2}t\cos kt$ by $\sin\left(k+\frac{1}{2}\right)t-\sin\left(k-\frac{1}{2}\right)t$, and make use of cancellation of terms. The formula for $\widetilde{D}_n(t)$ can be derived similarly.

Given a function f and a fixed point x, let us consider the expressions

$$\phi_x(t)=\frac{1}{2}[f(x+t)+f(x-t)], \quad \psi_x(t)=\frac{1}{2}[f(x+t)-f(x-t)]$$

as functions of t. They are called the *even* and *odd parts of* f at the point x, respectively. Clearly,

$$f(x+t)=\phi_x(t)+\psi_x(t).$$

It turns out that the behaviors of $\phi_x(t)$ and $\psi_x(t)$ near $t=0$ are decisive for the behaviors of $S[f]$ and $\widetilde{S}[f]$, as the case may be, at the point x.

Returning to the formula for $s_n(x)$ and making use of the even character of $D_n(t)$, we can write

$$\begin{aligned} s_n(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)D_n(x-t)\,dt=\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)D_n(t-x)\,dt \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)D_n(t)\,dt=\frac{2}{\pi}\int_{0}^{\pi}\frac{1}{2}[f(x+t)+f(x-t)]D_n(t)\,dt \\ &= \frac{2}{\pi}\int_{0}^{\pi}\phi_x(t)D_n(t)\,dt. \end{aligned}$$

The first formula (12.25) immediately gives

$$\begin{aligned} s_n(x) - f(x) &= \frac{2}{\pi} \int_0^\pi [\phi_x(t) - f(x)]\, D_n(t)\, dt \\ &= \frac{2}{\pi} \int_0^\pi [\phi_x(t) - f(x)] \frac{\sin\left(n + \frac{1}{2}\right) t}{2 \sin \frac{1}{2} t}\, dt. \end{aligned}$$

It will be convenient to modify this formula slightly by replacing n by $n-1$ and taking the semi-sum of the two formulas. When $n \geq 1$, writing

$$s_n^{\#}(x) = \frac{1}{2}\left[s_n(x) + s_{n-1}(x)\right] = s_n(x) - \frac{1}{2}\left(a_n \cos nx + b_n \sin nx\right), \tag{12.27}$$

we obtain, after observing that

$$\frac{1}{2}\left(a_n \cos nx + b_n \sin nx\right) = \frac{2}{\pi} \int_0^\pi [\phi_x(t) - f(x)] \frac{\cos nt}{2}\, dt,$$

the formula

$$s_n^{\#}(x) - f(x) = \frac{2}{\pi} \int_0^\pi [\phi_x(t) - f(x)] \frac{\sin nt}{2 \tan \frac{1}{2} t}\, dt.$$

The right side here is the nth Fourier sine coefficient of the odd function

$$[\phi_x(t) - f(x)] \frac{1}{2} \cot \frac{1}{2} t,$$

and if this function happens to be integrable near $t = 0$, the Riemann–Lebesgue theorem immediately gives $s_n^{\#}(x) - f(x) \to 0$. Hence, making use of the fact that $a_n, b_n \to 0$, we obtain from (12.27) the following basic result.

Theorem 12.28 (Dini's Test) *Let f be periodic and integrable. If the integral*

$$\int_0^\pi \left|\phi_x(t) - f(x)\right| \frac{1}{2} \cot \frac{1}{2} t\, dt$$

is finite, then $S[f]$ converges at the point x to the value $f(x)$.

Since only small values of t matter here, and since for small t we have $\frac{1}{2}\cot\frac{1}{2}t \simeq t^{-1}$, Dini's condition can be restated in the form

$$\int_0^\pi \frac{|\phi_x(t) - f(x)|}{t}\,dt < +\infty,$$

or what is the same thing

$$\int_0^\pi \frac{|f(x+t) + f(x-t) - 2f(x)|}{t}\,dt < +\infty. \tag{12.29}$$

The following special case is useful. Suppose that f has a jump discontinuity at x, so that the one-sided limits $f(x+), f(x-)$ exist and are finite. Since changing f at a single point does not affect $S[f]$, we may assume that

$$f(x) = \frac{1}{2}[f(x+) + f(x-)],$$

in which case we say that f has a *regular discontinuity* at x. Condition (12.29) is then certainly satisfied if both

$$\int_0^\pi \frac{|f(x+t) - f(x+)|}{t}\,dt < +\infty, \quad \int_0^\pi \frac{|f(x-t) - f(x-)|}{t}\,dt < +\infty.$$

Thus, a corollary of Dini's test is that *if both $f(x+)$ and $f(x-)$ exist and are finite, and if both of the last two integrals are finite, then $S[f]$ converges at the point x to the value*

$$\frac{f(x+) + f(x-)}{2}.$$

There is a result analogous to Theorem 12.28 for $\widetilde{S}[f]$, and we will be brief here. Using the formula for $\widetilde{s}_n$ and the odd character of $\widetilde{D}_n$, we have if $n \geq 1$ that

$$\widetilde{s}_n(x) = -\frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\widetilde{D}_n(t)\,dt = -\frac{2}{\pi}\int_0^\pi \psi_x(t)\frac{\cos\frac{1}{2}t - \cos\left(n+\frac{1}{2}\right)t}{2\sin\frac{1}{2}t}\,dt,$$

$$\widetilde{s}_n^{\#}(x) = \frac{\widetilde{s}_n(x) + \widetilde{s}_{n-1}(x)}{2}$$
$$= -\frac{2}{\pi}\int_0^{\pi} \psi_x(t)\frac{1}{2}\cot\frac{1}{2}t\,dt + \frac{2}{\pi}\int_0^{\pi}\psi_x(t)\frac{\cos\left(n+\frac{1}{2}\right)t}{2\sin\frac{1}{2}t}\,dt,$$

provided that

$$\int_0^{\pi} |\psi_x(t)|\,\frac{dt}{t} < +\infty. \tag{12.30}$$

Under this hypothesis, the last term in the preceding equation tends to zero by the Riemann–Lebesgue theorem, and we obtain (see Exercise 22)

Theorem 12.31 *Under the hypothesis (12.30), the series $\widetilde{S}[f]$ converges at the point x to the sum*

$$-\frac{2}{\pi}\int_0^{\pi}\psi_x(t)\frac{1}{2}\cot\frac{1}{2}t\,dt = -\frac{1}{\pi}\int_0^{\pi}\frac{f(x+t)-f(x-t)}{2\tan\frac{1}{2}t}\,dt.$$

We denote the last integral, which converges absolutely when (12.30) holds, by $\widetilde{f}(x)$:

$$\widetilde{f}(x) = -\frac{1}{\pi}\int_0^{\pi}\frac{f(x+t)-f(x-t)}{2\tan\frac{1}{2}t}\,dt. \tag{12.32}$$

This function is called the *conjugate function* of f and is intimately connected with the behavior of $\widetilde{S}[f]$. We will study the existence and properties of $\widetilde{f}$ in detail later.

Observe that condition (12.30) is of a nature completely different from (12.29); (12.30) precludes the possibility that f may have a jump at x. See Exercise 15.

The proofs of Theorems 12.28 and 12.31 are based on the Riemann–Lebesgue theorem and give convergence results only at individual points. They cannot give uniform convergence in an interval without additional and rather strong assumptions. We consider one such assumption that, though very restrictive, leads to an important result.

Theorem 12.33 *If $f = 0$ in an interval (a, b), then $S[f]$ and $\widetilde{S}[f]$ converge uniformly in every smaller interval $(a+\varepsilon, b-\varepsilon)$. The sum of $S[f]$ is 0.*

Proof. The pointwise convergence in (a, b) is a corollary of Theorems 12.28 and 12.31, and it is only the question of uniformity that requires additional comment. We will consider only $S[f]$; the argument for $\widetilde{S}[f]$ is similar. Fix $\varepsilon > 0$. From (12.27), we deduce that

$$s_n^{\#}(x_0) = \frac{s_n(x_0) + s_{n-1}(x_0)}{2} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x_0 + t) \frac{\sin nt}{2 \tan \frac{1}{2} t} \, dt$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x_0 + t) \chi(t) \sin nt \, dt, \quad x_0 \in (a + \varepsilon, b - \varepsilon),$$

where $\chi(t)$ is periodic, equals $\frac{1}{2} \cot \frac{1}{2} t$ for $\varepsilon \le |t| \le \pi$, and is arbitrary for $|t| < \varepsilon$. Suppose χ is defined so that it is continuous everywhere. Write $f(x_0 + t)\chi(t) = g_{x_0}(t)$, treating t as the variable and x_0 as a parameter, and consider the modulus of continuity of $g_{x_0}(t)$ in the metric L^1. If we show that $\omega_1(\delta, g_{x_0})$ tends to 0 with δ uniformly for $x_0 \in (a + \varepsilon, b - \varepsilon)$, then the Fourier coefficients of $g_{x_0}(t)$ will also tend to 0 uniformly for such x_0 (see the second formula (12.22)), and the theorem will follow. Now, for $h > 0$,

$$\int_0^{2\pi} \left| g_{x_0}(t + h) - g_{x_0}(t) \right| dt = \int_0^{2\pi} \left| f(x_0 + t + h) \chi(t + h) - f(x_0 + t) \chi(t) \right| dt$$

$$\le \int_0^{2\pi} \left| f(x_0 + t + h) - f(x_0 + t) \right| |\chi(t + h)| \, dt$$

$$+ \int_0^{2\pi} \left| f(x_0 + t) \right| |\chi(t + h) - \chi(t)| \, dt.$$

The last integral clearly tends to 0 with h, uniformly in x_0, since $\max |\chi(t + h) - \chi(t)| \to 0$ as $h \to 0$. If $M = \max |\chi|$, the preceding integral is majorized by

$$M \int_0^{2\pi} \left| f(x_0 + t + h) - f(x_0 + t) \right| dt = M \int_0^{2\pi} |f(t + h) - f(t)| \, dt,$$

a quantity independent of x_0 and tending to 0. This completes the proof.

Two trigonometric series T_1 and T_2 are said to be *equiconvergent* at a point x_0 if their difference $T_1 - T_2$ converges to 0 at x_0. If $T_1 - T_2$ merely converges, but not necessarily to 0, then T_1 and T_2 are said to be *equiconvergent in the wider sense* at x_0. Each of two equiconvergent series may be individually divergent, but the character of divergence is so similar that divergence cancels out in $T_1 - T_2$.

Theorem 12.34 *Let f_1 and f_2 be two periodic functions that are equal in an interval (a, b). Then $S[f_1]$ and $S[f_2]$ are uniformly equiconvergent in every subinterval $(a + \varepsilon, b - \varepsilon)$; $\widetilde{S}[f_1]$ and $\widetilde{S}[f_2]$ are uniformly equiconvergent in the wider sense in every $(a + \varepsilon, b - \varepsilon)$.*

This is a corollary of Theorem 12.33, since, for example, $S[f_1] - S[f_2] = S[f]$ where $f\ (= f_1 - f_2)$ vanishes in (a, b).

Thus, if we change the values of f in an *arbitrary way* outside an interval (a, b), we do not affect the behavior of $S[f]$ in $(a + \varepsilon, b - \varepsilon)$. Likewise for $\widetilde{S}[f]$, although in this case, if the series converges, the value of the sum may change. Therefore, the convergence or divergence of $S[f]$ and $\widetilde{S}[f]$ at a point x_0 is a *local* property, that is, it depends only on the behavior of f near x_0.

12.4 Divergence of Fourier Series

Theorem 12.35 *There exists a continuous periodic f such that $S[f]$ diverges (more specifically, the partial sums of $S[f]$ are unbounded) at some point.*

Proof. Let $1 \le m < n$ and consider the polynomials

$$Q_{m,n}(x) = \frac{\cos mx}{n} + \frac{\cos(m+1)x}{n-1} + \cdots + \frac{\cos(m+n-1)x}{1}$$
$$- \frac{\cos(m+n+1)x}{1} - \frac{\cos(m+n+2)x}{2} - \cdots - \frac{\cos(m+2n)x}{n}.$$

We will show that all these polynomials are uniformly bounded, but that their partial sums are not. To prove the first statement, we need the fact that the partial sums of the series

$$\sum_{k=1}^{\infty} \frac{\sin kx}{k},$$

which we considered on p. 306 in Section 12.1, are uniformly bounded. This is an elementary fact that can be proved in many ways (see, e.g., Theorem 12.50(c)), but here we take it for granted. Thus, since

$$Q_{m,n}(x) = \sum_{k=1}^{n} \frac{\cos(m+n-k)x - \cos(m+n+k)x}{k}$$
$$= 2\sin(m+n)x \sum_{k=1}^{n} \frac{\sin kx}{k},$$

we obtain $|Q_{m,n}(x)| \leq C$, where C is independent of m and n. On the other hand, when $x = 0$, the partial sum

$$Q^{\#}_{m,n}(x) = \frac{\cos mx}{n} + \cdots + \frac{\cos(m+n-1)x}{1}$$

has the value $1 + (1/2) + \cdots + (1/n)$, which is of order $\log n$.

Now select integers m_k and n_k such that

$$m_k + 2n_k < m_{k+1} \quad (k = 1, 2, \ldots),$$

and choose a series of positive numbers α_k such that $\sum \alpha_k < +\infty$, $\alpha_k \log n_k \to +\infty$. (We will make the construction in a moment.) The series

$$\sum_{k=1}^{\infty} \alpha_k Q_{m_k,n_k}(x)$$

then converges uniformly to a continuous function f. In view of the previous inequality relating m_k and m_{k+1}, the polynomials Q_{m_k,n_k} do not overlap. Hence, the last series can be written as a single trigonometric series, whose coefficients (because of uniform convergence) are the Fourier coefficients of f. Thus, this series, unbracketed, is $S[f]$. But $S[f]$ has unbounded partial sums at $x = 0$ since a single block of terms, namely, $\alpha_k Q^{\#}_{m_k,n_k}(x)$, is of order $\alpha_k \log n_k$ at $x = 0$.

It is easy to verify that if we set

$$m_k = 5^{k^3}, \quad n_k = 2m_k = 2\left(5^{k^3}\right), \quad \alpha_k = 1/k^2,$$

then all the conditions required previously are fulfilled. This completes the proof.

We leave it to the reader to check that if we choose

$$m_k = 5^{k^2}, \quad n_k = 2m_k, \quad \alpha_k = 1/k^2$$

in the construction above, we get a continuous f whose partial sums are divergent but bounded at $x = 0$.

Theorem 12.35 asserts that the partial sums of $S[f]$ can be unbounded even if f is continuous. It is of interest to know *how unbounded* they can be. From the formula

$$s_n(x,f) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)D_n(t)\,dt,$$

we see that if $|f| \le 1$, then

$$\left|s_n(x,f)\right| \le \frac{1}{\pi}\int_{-\pi}^{\pi} |D_n(t)|\,dt = \frac{2}{\pi}\int_{0}^{\pi} |D_n(t)|\,dt,$$

uniformly in x. The right side here is called the nth *Lebesgue constant* and will be denoted by L_n. Note that L_n is actually the value of $s_n(0,f)$ for a specific f, namely, $f(t) = \operatorname{sign} D_n(t)$.

Theorem 12.36 *We have*

$$L_n = \frac{2}{\pi}\int_{0}^{\pi} |D_n(t)|\,dt = \frac{4}{\pi^2}\log n + O(1) \quad \text{as } n \to \infty.$$

Proof. Write

$$\begin{aligned} L_n &= \frac{2}{\pi}\int_{0}^{\pi} |D_n(t)|\;dt = \frac{2}{\pi}\int_{0}^{\pi} \left|\sin\left(n+\frac{1}{2}\right)t\right| \frac{1}{2\sin\frac{1}{2}t}\,dt \\ &= \frac{2}{\pi}\int_{0}^{\pi} \left|\sin\left(n+\frac{1}{2}\right)t\right| \frac{dt}{t} + \frac{2}{\pi}\int_{0}^{\pi} \left|\sin\left(n+\frac{1}{2}\right)t\right| \left(\frac{1}{2\sin\frac{1}{2}t} - \frac{1}{t}\right) dt. \end{aligned}$$

Since $\left(2\sin\frac{1}{2}t\right)^{-1} - t^{-1}$ is nonnegative and bounded for $0 < t \le \pi$, and since $|\sin\left(n+\frac{1}{2}\right)t| \le 1$, the last integral is nonnegative and majorized by an absolute constant. The change of variable $\left(n+\frac{1}{2}\right)t = u$ shows that the preceding term equals

$$\frac{2}{\pi}\int_{0}^{(n+(1/2))\pi} \frac{|\sin u|}{u}\,du.$$

We may disregard the parts of this integral extended over $(0, \pi)$ and $\left(n\pi, \left(n + \frac{1}{2}\right)\pi\right)$, since the integrand is bounded. In view of the periodicity of $\sin u$, what remains can be written as

$$\frac{2}{\pi} \int_{\pi}^{n\pi} \frac{|\sin u|}{u}\, du = \frac{2}{\pi} \int_{0}^{\pi} (\sin u) \left(\sum_{k=1}^{n-1} \frac{1}{u + k\pi} \right) du.$$

For $0 \le u \le \pi$, the sum in brackets is contained between $\pi^{-1} \sum_{k=2}^{n} (1/k)$ and $\pi^{-1} \sum_{k=1}^{n-1} (1/k)$ and so differs from $\pi^{-1} \log n$ by an amount that is bounded in n and u. If we now note that $\int_0^{\pi} \sin u\, du = 2$, and collect estimates, we obtain $L_n = (4/\pi^2) \log n + O(1)$.

Theorem 12.37 *If f is integrable, then at each point x_0 of continuity of f,*

$$s_n\left(x_0, f\right) = o(\log n).$$

The estimate is uniform over every closed interval of continuity of f.

Proof. We will prove only the first statement, leaving the second to the reader. Suppose, as we may, that $x_0 = 0$, $f(x_0) = 0$. Because of our results about localization (see Theorem 12.34), we may assume that f vanishes outside an arbitrarily small fixed interval $(-\delta, \delta)$. Then

$$|s_n(0)| = \left| \frac{1}{\pi} \int_{-\delta}^{\delta} f(t) D_n(t)\, dt \right| \le \sup_{|t| \le \delta} |f(t)| \cdot \int_{-\pi}^{\pi} |D_n(t)|\, dt.$$

Since the sup here is small with δ and the integral is of order $\log n$, the assertion follows.

12.5 Summability of Sequences and Series

Theorem 12.35 shows that even continuous functions, when developed into Fourier series, may not be representable by those series in terms of pointwise convergence. The situation can be remedied by considering *generalized* sums of the series. This topic is vast and basic for analysis, and we will study only a few facts important for the theory of Fourier series.

Consider a fixed doubly infinite matrix of numbers (real or complex):

$$\begin{array}{ccccc} \alpha_{00} & \alpha_{01} & \cdots & \alpha_{0n} & \cdots \\ \alpha_{10} & \alpha_{11} & \cdots & \alpha_{1n} & \cdots \\ & \vdots & & & \\ \alpha_{m0} & \alpha_{m1} & \cdots & \alpha_{mn} & \cdots \\ & \vdots & & & \end{array} \qquad (\mathscr{M})$$

Given an infinite sequence of numbers $s_0, s_1, \ldots, s_n, \ldots$, we transform it by using $(\mathscr{M})$ into a sequence $\sigma_0, \sigma_1, \ldots, \sigma_m, \ldots$ by means of the formulas

$$\sigma_m = \alpha_{m0}s_0 + \alpha_{m1}s_1 + \cdots + \alpha_{mn}s_n + \cdots \quad (m = 0, 1, 2, \ldots),$$

assuming that the series defining σ_m converges for each m. We may ask what conditions on $(\mathscr{M})$ will guarantee that whenever $\{s_n\}$ converges to a finite limit s, $\lim \sigma_m$ also exists and equals s. An answer is given by the following theorem.

Theorem 12.38 *Suppose that $(\mathscr{M})$ satisfies the following three conditions:*

(i) $\sum_n |\alpha_{mn}| \leq A$ *(for all m, with A independent of m),*
(ii) $\lim_{m\to\infty}(\sum_n \alpha_{mn}) = 1$,
(iii) $\lim_{m\to\infty} \alpha_{mn} = 0$ *for each n.*

Then for any sequence $\{s_n\}$ converging to a finite limit s, $\lim \sigma_m$ exists and equals s.

Theorem 12.38 is due to Toeplitz, and a matrix $(\mathscr{M})$ that satisfies (i)–(iii) is called a *Toeplitz matrix*.

Proof. First of all, since $\{s_n\}$ is bounded, (i) implies that σ_m exists for each m. Next, write $s_n = s + \varepsilon_n$, where $\varepsilon_n \to 0$. Then

$$\sigma_m = \sum_n \alpha_{mn}\,(s + \varepsilon_n) = s\sum_n \alpha_{mn} + \sum_n \alpha_{mn}\varepsilon_n.$$

We have $s\sum_n \alpha_{mn} \to s$ by (ii), and it remains only to show that the expression

$$\rho_m = \sum_n \alpha_{mn}\varepsilon_n$$

tends to 0 as $m \to \infty$. Given $\delta > 0$, split ρ_m into two sums,

$$\rho_m = \sum_{n \le n_0} \alpha_{mn} \varepsilon_n + \sum_{n > n_0} \alpha_{mn} \varepsilon_n = \rho'_m + \rho''_m,$$

say, where n_0 is so large that $|\varepsilon_n| \le \delta$ for $n > n_0$. By (i),

$$\left|\rho''_m\right| \le \sum_{n > n_0} |\alpha_{mn}| \, |\varepsilon_n| \le \sum_{n > n_0} |\alpha_{mn}| \, \delta \le A\delta.$$

On the other hand, ρ'_m consists of a fixed number of terms each of which, by (iii), tends to 0 as $m \to \infty$. Hence, $\left|\rho'_m\right| < A\delta$ for m large enough. Combining estimates, we see that $\rho_m \to 0$, which completes the proof.

It is useful to note that if $s = 0$, then condition (ii) is not required in the proof (and so in the statement of the theorem) above. It is also immediate from the proof that if $\{s_n\}$ depends on a parameter, and if $\{s_n\}$ tends uniformly to a bounded limit s, then $\{\sigma_m\}$ tends uniformly to s too.

If $\sigma_m \to s$, we shall say that the sequence $\{s_n\}$ (or the series whose partial sums are the s_n) is *summable to limit* (*sum*) s *by means of the matrix* $(\mathscr{M})$, or simply is *summable* $(\mathscr{M})$ *to* s.

The matrix $(\mathscr{M})$ is called *positive* if $\alpha_{mn} \ge 0$ for all m, n. Condition (i) is then a corollary of (ii). For positive $(\mathscr{M})$, Theorem 12.38 also holds if $s = \pm\infty$; we leave the proof to the reader.

Two methods of summability are of special significance for Fourier series.

(a) *The method of the arithmetic mean*. Given $s_0, s_1, \ldots, s_n, \ldots$, consider the *arithmetic means* $\sigma_0, \sigma_1, \ldots, \sigma_m, \ldots$ defined by

$$\sigma_m = \frac{s_0 + s_1 + \cdots + s_m}{m+1} \quad (m = 0, 1, 2, \ldots).$$

If $s_n \to s$ $(-\infty \le s \le +\infty)$, then $\sigma_m \to s$. This is clearly a special case of Theorem 12.38; the matrix is positive.

It is useful (see, e.g., the comments following the proof of Theorem 12.44) to note that if the s_n are the partial sums of a series $\sum_{k=0}^{\infty} u_k$, then

$$\begin{aligned}\sigma_m &= \frac{s_0 + s_1 + \cdots + s_m}{m+1} = \frac{u_0 + (u_0 + u_1) + \cdots + (u_0 + u_1 + \cdots + u_m)}{m+1} \\ &= \frac{1}{m+1} \sum_{k=0}^{m} (m + 1 - k)\, u_k.\end{aligned}$$

Thus,

$$\sigma_m = \sum_{k=0}^{m} \left(1 - \frac{k}{m+1}\right) u_k, \quad s_m - \sigma_m = \frac{1}{m+1} \sum_{k=0}^{m} k u_k. \tag{12.39}$$

(b) *The method of Abel*. Given a series $u_0 + u_1 + \cdots + u_n + \cdots$, consider the power series

$$f(r) = \sum_{n=0}^{\infty} u_n r^n, \quad 0 \le r < 1,$$

assuming that it converges for $0 \le r < 1$. If $f(r) \to s$ as $r \to 1$, we say that $\sum u_n$ is *Abel summable* (or *A-summable*) to sum s. The method can also be applied to sequences since any sequence $\{s_n\}$ can be written as the partial sums of the series $s_0 + (s_1 - s_0) + (s_2 - s_1) + \cdots$.

Let us now see the relation of Abel summability to the general scheme. We claim that for $0 \le r < 1$, the formula

$$\sum_{n=0}^{\infty} u_n r^n = (1-r) \sum_{n=0}^{\infty} s_n r^n \quad (s_n = u_0 + \cdots + u_n) \tag{12.40}$$

is valid assuming only that one of the two series that appear is convergent. If the right side converges, it equals

$$\begin{aligned} \sum_{n=0}^{\infty} s_n r^n - \sum_{n=0}^{\infty} s_n r^{n+1} &= \sum_{n=0}^{\infty} s_n r^n - \sum_{n=1}^{\infty} s_{n-1} r^n \\ &= s_0 + \sum_{n=1}^{\infty} (s_n - s_{n-1})\, r^n = \sum_{n=0}^{\infty} u_n r^n. \end{aligned}$$

Conversely, if $\sum_{n=0}^{\infty} u_n r^n$ converges for some r, $0 < r < 1$, its Cauchy product with the absolutely convergent series $\sum_{n=0}^{\infty} r^n = (1-r)^{-1}$ converges to sum

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} u_k r^k \cdot r^{n-k} \right) = \sum_{n=0}^{\infty} (u_0 + u_1 + \cdots + u_n)\, r^n = \sum_{n=0}^{\infty} s_n r^n.$$

This proves (12.40). Now, if $\{r_m\}$ is any sequence tending to 1, $0 < r_m < 1$, then the positive numbers

$$\alpha_{mn} = (1 - r_m)\, r_m^n$$

satisfy conditions (i), (ii), (iii) of Theorem 12.38. We leave the verification to the reader.

Theorem 12.41 (Abel) *If $\sum_{n=0}^{\infty} u_n$ converges to sum s, $-\infty \le s \le +\infty$, then it is A-summable to s.*

Proof. Suppose first that s is finite. Applying (12.40), we have to show that $(1-r)\sum_{n=0}^{\infty} s_n r^n \to s$ as $r \to 1$. It is enough to prove that this relation holds for any sequence $r = r_m, m = 0, 1, \ldots$, where $0 < r_m < 1$, $r_m \to 1$. This is a corollary of Theorem 12.38 since the numbers $\alpha_{mn} = (1-r_m)r_m^n$ satisfy (i), (ii), (iii). The matrix α_{mn} is positive, and so the proof holds for $s = \pm\infty$, the only prerequisite being that the series $\sum u_n r^n$ converges for $0 \le r < 1$.

We may also consider the power series

$$f(z) = \sum_{n=0}^{\infty} u_n z^n,$$

where z is a *complex* variable lying in the unit disc: $z = re^{ix}$, $0 \le r < 1$. If $f(z)$ tends to a limit s as z tends *nontangentially* to 1, that is, as $z \to 1$ in such a way that

$$\frac{|1-z|}{1-|z|} \le C < +\infty \quad (|z| < 1),$$

then $\sum_{n=0}^{\infty} u_n$ is said to be *nontangentially Abel summable* to sum s. The last inequality means that, in approaching 1, z remains between two chords of the unit circle through $z = 1$. In fact, if $z = x + iy$ is a point that satisfies $0 < x < 1$ and $|1-z| \le C(1-|z|)$, then $|y| < C(1-x)$ since $|y| < \sqrt{y^2 + (1-x)^2} = |1-z|$ and $C(1-|z|) \le C(1-x)$. Conversely (see Exercise 23(a)), given a constant $\gamma > 0$, there are constants C and δ with $C > 0$ and $0 < \delta < 1$ such that if $z = x+iy$ with $|z| < 1, 1-\delta < x < 1$ and $|y| < \gamma(1-x)$, then $|1-z| \le C(1-|z|)$.

See (12.65) for another characterization of the notion of nontangential approach of z to 1.

Theorem 12.42 (Abel–Stolz) *If $\sum_{n=0}^{\infty} u_n$ converges to a finite sum s, then it is nontangentially Abel summable to s.*

Proof. The proof is identical to that of Abel's theorem, except that now we use the formula $\sum u_n z^n = (1-z)\sum s_n z^n$ and consider any sequence $\{z_m\}$ tending

to 1 from the interior of the unit disc. The matrix α_{mn} is now $(1 - z_m)z_m^n$, conditions (ii) and (iii) of Theorem 12.38 are satisfied as before, and (i) takes the form

$$\frac{|1 - z_m|}{1 - |z_m|} \leq C.$$

Theorem 12.43 *If $\sum_{n=0}^{\infty} u_n$ is summable by the method of the arithmetic mean to sum s, then it is A-summable to s. If in addition s is finite, then $\sum_{n=0}^{\infty} u_n$ is nontangentially A-summable to s.*

Proof. Suppose that s is finite. By hypothesis,

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_n}{n+1} \to s.$$

Write $s_0 + s_1 + \cdots + s_n = t_n$. Applying formula (12.40) twice, we have

$$\sum_{n=0}^{\infty} u_n r^n = (1-r)\sum_{n=0}^{\infty} s_n r^n = (1-r)^2 \sum_{n=0}^{\infty} t_n r^n = (1-r)^2 \sum_{n=0}^{\infty} (n+1)\sigma_n r^n.$$

Again, it is enough to consider any sequence $r_m \to 1$, $0 < r_m < 1$. We then have to apply Theorem 12.38 with matrix

$$\alpha_{mn} = (1 - r_m)^2 (n+1) r_m^n,$$

and we easily verify that (α_{mn}) satisfies conditions (i), (ii), (iii) of Theorem 12.38. The rest of the proof of the theorem is left to the reader.

While convergence of a series implies summability A, the converse is generally false: for example, $\sum_{n=0}^{\infty}(-1)^n$ diverges but is A-summable to sum $\frac{1}{2}$ since $\sum_{n=0}^{\infty}(-r)^n = 1/(1+r) \to \frac{1}{2}$ as $r \to 1-$. If one makes additional assumptions on the terms of the series, however, the converse will hold. The following result is both elementary and useful.

Theorem 12.44 (Tauber) *If $\sum u_n$ is A-summable to sum s, $-\infty \leq s \leq +\infty$, and if $u_n = o(1/n)$ as $n \to \infty$, then $\sum u_n$ converges to sum s.*

Proof. Write $u_n = \varepsilon_n/n$, $n > 1$, where $\varepsilon_n \to 0$. Let r_m be a sequence tending to 1, which we shall determine in a moment. Then $s_m - f(r_m)$ is a transformation of the sequence ε_n:

$$s_m - f(r_m) = \sum_{n=1}^{m} \frac{\varepsilon_n}{n} - \sum_{n=1}^{\infty} \frac{\varepsilon_n}{n} r_m^n = \sum_{n=1}^{\infty} \alpha_{mn} \varepsilon_n,$$

where

$$\alpha_{mn} = \frac{1}{n}\left(1 - r_m^n\right) \quad \text{if } n \le m, \quad \alpha_{mn} = -\frac{1}{n} r_m^n \quad \text{if } n > m.$$

If we verify conditions (i) and (iii) of Theorem 12.38, then the fact that $\varepsilon_n \to 0$ will give $s_m - f(r_m) \to 0$, and so also $s_m \to s$. Condition (iii) is obvious for any $\{r_m\} \to 1$. As for (i), observing that

$$1 - r^n = (1 - r)\left(1 + \cdots + r^{n-1}\right) \le (1 - r)n,$$

we have

$$\begin{aligned}
\sum_n |\alpha_{mn}| &\le \sum_{n=1}^{m} \frac{1}{n}(1 - r_m)\, n + \sum_{n=m+1}^{\infty} \frac{1}{n} r_m^n \\
&\le m(1 - r_m) + \frac{1}{m+1} \sum_{n=0}^{\infty} r_m^n \\
&= m(1 - r_m) + \frac{1}{m+1} \frac{1}{1 - r_m}.
\end{aligned}$$

Hence, if we choose $r_m = 1 - (1/m)$, then $\sum_n |\alpha_{mn}| \le 2$. Thus, condition (i) holds, and the theorem follows.

If $\sum u_n$ is summable by the method of the arithmetic mean and $nu_n \to 0$, then $\sum u_n$ converges. Of course, this is a corollary of Theorems 12.43 and 12.44, but a direct proof is on the surface: By (12.39),

$$s_m - \sigma_m = \frac{1}{m+1} \sum_{n=0}^{m} n u_n,$$

and the assumption $nu_n \to 0$ clearly implies $s_m - \sigma_m \to 0$. Thus, if $\sigma_m \to s$, then also $s_m \to s$. Actually, this argument shows that if $nu_n \to 0$, then whether $\{\sigma_m\}$ converges or not, the difference $s_m - \sigma_m$ tends to 0, that is, the behavior of $\{s_m\}$ imitates that of $\{\sigma_m\}$. The same argument shows that *if $\{\sigma_m\}$ is a bounded*

sequence, and $|u_n| \le A/n$ *for* $n = 1, 2, \ldots$, *then the sequence* $\{s_m\}$ *is bounded*. The result that follows lies deeper.

Theorem 12.45 (Hardy) *If* $\sum u_n$ *is summable by the method of the arithmetic mean to a finite sum* s *and if*

$$|u_n| \le \frac{A}{n} \quad (n = 1, 2, \ldots),$$

then $\sum u_n$ *converges to* s.

Proof. Consider the expressions (which we shall call the *delayed arithmetic means*)

$$\sigma_{n,k} = \frac{s_{n+1} + s_{n+2} + \cdots + s_{n+k}}{k}.$$

They are easily expressible in terms of the σ_n:

$$\begin{aligned}\sigma_{n,k} &= \frac{\big(s_0 + \cdots + s_{n+k}\big) - (s_0 + \cdots + s_n)}{k} = \frac{n+k+1}{k}\sigma_{n+k} - \frac{n+1}{k}\sigma_n \\ &= \frac{n+1}{k}\big(\sigma_{n+k} - \sigma_n\big) + \sigma_{n+k}.\end{aligned}$$

It is clear that if k_n is any sequence of integers such that n/k_n is bounded as $n \to \infty$, then $\sigma_n \to s$ implies that $\sigma_{n,k_n} \to s$. Using the definition of $\sigma_{n,k}$, we also deduce that

$$\begin{aligned}\sigma_{n,k} &= s_n + \frac{(s_{n+1} - s_n) + \cdots + \big(s_{n+k} - s_n\big)}{k} \\ &= s_n + \frac{1}{k}\sum_{j=1}^{k}(k - j + 1)u_{n+j}.\end{aligned}$$

Hence, assuming as we may that $A = 1$,

$$\big|\sigma_{n,k} - s_n\big| \le \sum_{j=1}^{k}\big|u_{n+j}\big| \le \frac{k}{n+1}.$$

Let $k = k_n = [\varepsilon n]$, where $\varepsilon > 0$ is arbitrarily small and fixed, and $[x]$ designates the integral part of x. Then n/k_n is bounded, and so $\sigma_{n,k_n} \to s$. But by taking $k_n = [\varepsilon n]$ in the last estimate, we obtain

$$\limsup_{n\to\infty} \left|\sigma_{n,k_n} - s_n\right| \le \varepsilon.$$

Hence, $s_n \to s$, and the proof is complete.

We remark in passing that the conclusion of Hardy's theorem is true if the assumption of summability by the method of the arithmetic mean is replaced by A-summability (theorem of Littlewood)*.

12.6 Summability of $S[f]$ and $\widetilde{S}[f]$ by the Method of the Arithmetic Mean

Given a periodic f, we denote by $s_n(x) = s_n(x,f)$ the partial sums of $S[f]$ and by $\sigma_n(x) = \sigma_n(x,f)$ their arithmetic means. Thus (see p. 315 in Section 12.3 for the definition of the Dirichlet kernel D_n),

$$s_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)D_n(x-t)\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)D_n(t)\,dt,$$

$$\sigma_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)K_n(x-t)\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)K_n(t)\,dt,$$

where, by using (12.26),

$$K_n(t) = \frac{1}{n+1}\sum_{j=0}^{n} D_j(t) = \frac{1}{(n+1)2\sin\frac{1}{2}t}\sum_{j=0}^{n}\sin\left(j+\frac{1}{2}\right)t.$$

Note that $\sigma_0(x) = a_0/2$ and $K_0(t) = 1/2$. Multiplying and dividing the last sum termwise by $2\sin\frac{1}{2}t$ and using the equation $2\sin\left(j+\frac{1}{2}\right)t\,\sin\frac{1}{2}t = \cos jt - \cos(j+1)t$, we get

$$K_n(t) = \frac{1-\cos(n+1)t}{(n+1)\left(2\sin\frac{1}{2}t\right)^2} = \frac{2}{n+1}\left(\frac{\sin[(n+1)t/2]}{2\sin\frac{1}{2}t}\right)^2. \tag{12.46}$$

* See A. Zygmund, *Trigonometric Series*, vol. 1, 2nd edn., Cambridge University Press, Cambridge, U.K., 1968, p. 81.

The trigonometric polynomial $K_n(t)$ is called the nth *Fejér kernel*. An analogous kernel was considered in (9.11) for nonperiodic functions. The formula

$$\sigma_n(x,f) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)K_n(x-t)\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x-t)K_n(t)\,dt$$

is a periodic version of the notion of convolving a function and a kernel. Some of the facts we will prove below are similar to ones we have already had in Chapter 9, but rather than connecting the present case with those results, we shall give brief direct proofs of the theorems we need.

Using the formula $D_j(t) = \frac{1}{2} + \sum_{m=1}^{j} \cos mt$, $j \geq 1$, the Fejér kernel can be written (see (12.39)):

$$K_n(t) = \frac{1}{2} + \sum_{m=1}^{n}\left(1 - \frac{m}{n+1}\right)\cos mt = \frac{1}{2}\sum_{m=-n}^{n}\left(1 - \frac{|m|}{n+1}\right)e^{imt},$$

which should be interpreted as $1/2$ in case $n = 0$. K_n has the following properties:

(a) $K_n(t) \geq 0$; $K_n(-t) = K_n(t)$.
(b) $(1/\pi)\int_{-\pi}^{\pi} K_n(t)\,dt = 1$.
(c) $K_n(t) \leq (n+1)/2$; $K_n(t) \leq A/[(n+1)t^2]$ $(0 < |t| \leq \pi$; A an absolute constant).

Here, (a) and (b) are obvious from the various previously mentioned formulas for K_n. The first part of (c) follows from the formula $K_n(t) = (n+1)^{-1}\sum_{j=0}^{n} D_j(t)$ together with the obvious estimate $|D_j| \leq j + \frac{1}{2}$ and the identity $\sum_{j=0}^{n} j = n(n+1)/2$. The second part follows from (12.46) if we note that $|\sin u| \geq (2/\pi)|u|$ for $0 \leq |u| \leq \pi/2$.

From the second inequality in (c), we immediately deduce

(c′) $\int_{\delta \leq |t| \leq \pi} K_n(t)\,dt \to 0$ as $n \to \infty$ for any fixed δ, $0 < \delta \leq \pi$.

These properties of K_n lead to the next result, which is basic and related to Theorem 9.9.

Theorem 12.47 (Fejér) *Let f be integrable and periodic. Then*

$$\sigma_n(x) \to f(x)$$

at each point of continuity of f, and the convergence is uniform over every closed interval of continuity. In particular, $\sigma_n(x)$ tends to $f(x)$ uniformly everywhere if f is continuous everywhere. If f has a jump discontinuity at x_0, then

$$\sigma_n(x_0) \to \frac{1}{2}\left[f(x_0+) + f(x_0-)\right].$$

Proof. Suppose f is continuous on a closed interval $I = [a, b]$ (which may reduce to a point). Given $\varepsilon > 0$, we can find $\delta > 0$ so that $|f(x+t) - f(x)| < \varepsilon$ for $x \in I$, $|t| < \delta$. Using (b), we can write (assuming as we may that $\delta \le \pi$)

$$\begin{aligned}\sigma_n(x) - f(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x+t) - f(x)]K_n(t)\,dt \\ &= \frac{1}{\pi}\int_{|t|<\delta} + \frac{1}{\pi}\int_{\delta\le|t|\le\pi} = \alpha_n + \beta_n.\end{aligned}$$

Clearly, if $x \in I$, then

$$|\alpha_n| \le \frac{1}{\pi}\int_{|t|<\delta} \varepsilon K_n(t)\,dt \le \frac{\varepsilon}{\pi}\int_{-\pi}^{\pi} K_n(t)\,dt = \varepsilon.$$

Let $M = \max|f|$ in I. Then, for $x \in I$,

$$\begin{aligned}|\beta_n| &\le \frac{1}{\pi}\int_{\delta\le|t|\le\pi} (|f(x+t)| + M)K_n(t)\,dt \\ &\le \frac{1}{\pi}\left[\max_{\delta\le|t|\le\pi} K_n(t)\right]\int_{\delta\le|t|\le\pi} [|f(x+t)| + M]\,dt.\end{aligned}$$

The last integral is majorized by $\int_{-\pi}^{\pi}|f(x+t)|\,dt + 2\pi M = \int_{-\pi}^{\pi}|f(t)|\,dt + 2\pi M$, and by (c), the factor preceding it tends to 0 as $n \to \infty$. Hence, $|\beta_n| \to 0$ uniformly for $x \in I$, and $|\alpha_n| + |\beta_n| < 2\varepsilon$ for n large enough and $x \in I$. This proves the first part of Theorem 12.47.

The proof of the second part is similar. We may assume that $f(x_0) = \frac{1}{2}\left[f(x_0+) + f(x_0-)\right]$ since $\sigma_n(x_0)$ is unaffected by changing f at a single point. Then since K_n is even,

$$\sigma_n(x_0) - f(x_0) = \frac{1}{\pi}\int_0^\pi [f(x_0+t) + f(x_0-t) - f(x_0+) - f(x_0-)]\, K_n(t)\, dt,$$

$$\begin{aligned}\left|\sigma_n(x_0) - f(x_0)\right| &\le \frac{1}{\pi}\int_0^\pi \left|f(x_0+t) - f(x_0+)\right| K_n(t)\, dt \\ &\quad + \frac{1}{\pi}\int_0^\pi \left|f(x_0-t) - f(x_0-)\right| K_n(t)\, dt = a_n + b_n.\end{aligned}$$

To show, for example, that $a_n \to 0$, write $\int_0^\pi = \int_0^\delta + \int_\delta^\pi$ and use the fact that in $(0, \delta)$ the difference $\left|f(x_0+t) - f(x_0+)\right|$ is small, while in (δ, π) we have $\max K_n(t)$ tending to zero. The argument for b_n is similar, and the proof is complete.

The following result, although it is simple, deserves a statement.

Theorem 12.48

(a) *Let f be periodic and integrable. If $f(x) \le B$ for all x, then also $\sigma_n(x) \le B$. If $f(x) \ge A$, then $\sigma_n(x) \ge A$. If $|f(x)| \le M$, then $|\sigma_n(x)| \le M$.*

(b) *If $f(x) \to \pm\infty$ as $x \to x_0$, then $\sigma_n(x_0) \to \pm\infty$ as $n \to \infty$.*

We leave the proofs to the reader.

The next two results are corollaries of Fejér's theorem.

Theorem 12.49 *Let f be periodic and integrable, $f \sim \sum c_k e^{ikx}$, and let F be the indefinite integral of f. Then the series in the formula*

$$F(x) - c_0 x \sim C_0 + \sum{}' \frac{c_k}{ik} e^{ikx}$$

(see Theorem 12.14) converges uniformly to $F(x) - c_0 x$.

Proof. Let $S(x)$ denote the series on the right side of the formula. Then $S(x)$ is the Fourier series of a continuous function, and therefore its arithmetic means converge uniformly by Fejér's theorem. Since the terms of $S(x)$ are bounded uniformly in x and are also of order $o(1/|k|)$ uniformly in x, the difference between the partial sums and the arithmetic means of $S(x)$ tends uniformly to 0 by using an argument like the one in the discussion following the proof of Theorem 12.44. Also, $S(x) = F(x) - c_0 x$ by Theorem 12.19.

Theorem 12.50 (Dirichlet–Jordan) *If f is periodic and of bounded variation, then*

(a) *$S[f]$ converges to $f(x)$ at each point of continuity of f and to $\frac{1}{2}[f(x+) + f(x-)]$ at each point of discontinuity.*

(b) *The convergence of $S[f]$ is uniform over every closed interval of continuity of f.*

(c) *The partial sums of $S[f]$ are uniformly bounded.*

Parts (a) and (b) follow immediately from Fejér's theorem if one uses Theorem 12.45 and the fact that the Fourier coefficients of a function of bounded variation are $O(1/|k|)$ (see Theorem 12.24). For (c), use Theorem 12.48 and the remark before Theorem 12.45.

Perhaps it is of interest to observe here that the classical theorem of Weierstrass about the uniform approximability of functions that are continuous in finite closed intervals by *power* polynomials can be easily deduced from Fejér's theorem. Suppose that $f(x)$ is continuous for $a \le x \le b$. The formula $x = \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t$ establishes a one-to-one mapping between the intervals $a \le x \le b$ and $-1 \le t \le +1$, and every $f(x)$ continuous in $[a, b]$ becomes a $g(t)$ continuous in $[-1, +1]$. If we approximate $g(t)$ by polynomials in t, we at the same time approximate $f(x)$ by polynomials in x. Hence, we may assume from the start that $f(x)$ is defined and continuous in $[-1, +1]$. Write $x = \cos\theta$. The function $h(\theta) = f(\cos\theta)$ is then defined and continuous in $[0, \pi]$, and if we extend it to $[-\pi, \pi]$ by the condition of evenness, and after that to $(-\infty, +\infty)$ by periodicity, then $h(\theta)$ can be approximated arbitrarily closely and uniformly on $[0, \pi]$ by *cosine* polynomials

$$T(\theta) = \sum_{k=0}^{n} \alpha_k \cos k\theta,$$

for example, the arithmetic means of $S[h]$, these being cosine polynomials since h is even. It is easy to see that $\cos k\theta$ is a power polynomial of degree k in $x = \cos\theta$: for $k = 0, 1$, this is obvious, and for general k, it follows by induction from the formula $\cos k\theta + \cos(k-2)\theta = 2\cos\theta\cos(k-1)\theta$. Thus, the polynomials $T(\theta)$ above are power polynomials $P(\cos\theta)$ in $\cos\theta$, and the approximability of $h(\theta)$ by $T(\theta)$ is the same thing as the approximability of $f(x)$ by $P(x)$, which verifies Weierstrass's approximation theorem.

We shall now consider the arithmetic means of $S[f]$ when f is merely integrable. In Chapter 7, we introduced the notion of a Lebesgue point of a function in $\mathbf{R^n}$, but here we are only interested in the case $n = 1$. We recall the definition. A point x_0 is a *Lebesgue point* for a locally integrable f if

$$\frac{1}{2h}\int_{-h}^{h}\left|f(x_0+t)-f(x_0)\right|dt \to 0 \quad (h\to 0),$$

and we proved that almost all points have this property.

Simultaneously with $S[f]$, we shall also consider $\widetilde{S}[f]$, for f merely integrable. For $0<\varepsilon\le\pi$, we write

$$\begin{aligned}\widetilde{f}_\varepsilon(x) &= -\frac{1}{\pi}\int_{\varepsilon\le|t|\le\pi} f(x+t)\,\frac{dt}{2\tan\frac{1}{2}t}\\ &= -\frac{1}{\pi}\int_{\varepsilon}^{\pi}\frac{f(x+t)-f(x-t)}{2\tan\frac{1}{2}t}\,dt\end{aligned}$$

and call $\widetilde{f}_\varepsilon(x)$ the *truncated conjugate function* of f. If $\lim_{\varepsilon\to 0}\widetilde{f}_\varepsilon(x)$ exists, we will denote it $\widetilde{f}(x)$ and call it the *conjugate function* of f:

$$\begin{aligned}\widetilde{f}(x) &= \lim_{\varepsilon\to 0}\left(-\frac{1}{\pi}\int_{\varepsilon}^{\pi}\frac{f(x+t)-f(x-t)}{2\tan\frac{1}{2}t}\,dt\right)\\ &= \lim_{\varepsilon\to 0}\left(-\frac{1}{\pi}\int_{\varepsilon\le|t|\le\pi}\frac{f(x+t)}{2\tan\frac{1}{2}t}\,dt\right).\end{aligned}$$

We came across this function in Theorem 12.31 in connection with Dini's criterion. Occasionally, one also uses the notation

$$\widetilde{f}(x) = -\frac{1}{\pi}\,\text{p.v.}\int_{-\pi}^{\pi}\frac{f(x+t)}{2\tan\frac{1}{2}t}\,dt,$$

where p.v. stands for *principal value*, indicating that the integral, which as a Lebesgue integral is generally divergent at $t=0$, is given a new meaning by first removing a *symmetric* neighborhood around $t=0$ and then making that neighborhood shrink to 0. Formally, $\widetilde{f}$ is the convolution of f and $\frac{1}{2}\cot\frac{1}{2}t$, although the latter is not an integrable function. We will study the existence of $\widetilde{f}$ later.

The arithmetic means of $\widetilde{S}[f]$ will be denoted by $\widetilde{\sigma}_n(x)=\widetilde{\sigma}_n(x,f)$. From the formula on p. 315 in Section 12.3 for $\widetilde{s}_n$ and the oddness of the conjugate Dirichlet kernels $\widetilde{D}_n(t)$, $n\ge 1$, we obtain

$$\widetilde{\sigma}_n(x) = -\frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\widetilde{K}_n(t)\,dt,$$

where

$$\widetilde{K}_n(t) = \frac{1}{n+1} \sum_{j=0}^{n} \widetilde{D}_j(t).$$

Of course, $\widetilde{D}_0 = \widetilde{K}_0 = \widetilde{\sigma}_0 = 0$.

Theorem 12.51 (Lebesgue) *Suppose that f is periodic and integrable. Then at every Lebesgue point x_0 of f (in particular, for almost every x_0),*

(i) $\sigma_n(x_0) \to f(x_0)$ *as* $n \to \infty$,
(ii) $\widetilde{\sigma}_n(x_0) - \widetilde{f}_{1/n}(x_0) \to 0$ *as* $n \to \infty$.

Proof. Note that part (ii) does not assert that either $\widetilde{\sigma}_n(x_0)$ or $\widetilde{f}_{1/n}(x_0)$ has a limit, but only that their difference tends to 0. We will use the estimates

$$K_n(t) \le n, \quad K_n(t) \le \frac{A}{nt^2} \quad (n \ge 1,\ 0 < t \le \pi), \tag{12.52}$$

which are just variants of (c) on p. 333, Section 12.6. If we have to use both estimates, then clearly the first is preferable for $t \le 1/n$ and the second for $t \ge 1/n$. The proof that follows is basically a repetition of the argument for Theorem 9.13.

Let x_0 be a Lebesgue point of f. Assuming as we may that $f(x_0) = 0$, and letting

$$\phi(t) = \left|f(x_0+t)\right| + \left|f(x_0-t)\right|, \quad \psi(t) = \int_0^t \phi(u)\,du,$$

the condition that x_0 is a Lebesgue point takes the form $\psi(h)/h \to 0$ as $h \to 0$. The formula

$$\sigma_n(x_0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x_0+t)\,K_n(t)\,dt = \frac{1}{\pi}\int_0^{\pi} [f(x_0+t) + f(x_0-t)]\,K_n(t)\,dt$$

gives

$$|\sigma_n(x_0)| \le \frac{1}{\pi}\int_0^{\pi} \phi(t)K_n(t)\,dt = \frac{1}{\pi}\int_0^{1/n} + \frac{1}{\pi}\int_{1/n}^{\pi} = \alpha_n + \beta_n.$$

Clearly,

$$0 \le \alpha_n \le \int_0^{1/n} \phi(t) n \, dt = \frac{\psi(1/n)}{1/n} \to 0.$$

Next, using the second estimate for K_n and integrating by parts, we have

$$0 \le \beta_n \le \frac{A}{n} \int_{1/n}^{\pi} \frac{\phi(t)}{t^2} \, dt = \frac{A}{n} \left[\frac{\psi(t)}{t^2} \right]_{1/n}^{\pi} + \frac{2A}{n} \int_{1/n}^{\pi} \frac{\psi(t)}{t^3} \, dt.$$

The integrated term tends to 0 as $n \to \infty$. As for the last term, we will show that it also tends to 0. Given any $\varepsilon > 0$, take δ so small that $\psi(t)/t < \varepsilon$ if $0 < t < \delta$. Then

$$\frac{1}{n} \int_{1/n}^{\pi} \frac{\psi(t)}{t^3} \, dt \le \frac{1}{n} \int_{1/n}^{\delta} \frac{\varepsilon t}{t^3} \, dt + \frac{1}{n} \int_{\delta}^{\pi} \frac{\psi(t)}{t^3} \, dt.$$

The first term on the right is majorized by $(\varepsilon/n) \int_{1/n}^{\infty} t^{-2} \, dt = \varepsilon$, while the last term clearly tends to 0 as $n \to \infty$ since ψ is bounded. Collecting results, we conclude that $\sigma_n(x_0) \to 0$. This proves (i).

To prove (ii), we need estimates for $\widetilde{K}_n = [1/(n+1)] \sum_{j=0}^{n} \widetilde{D}_j$. The obvious inequality $\left|\widetilde{D}_j\right| \le j$ (recall that $\widetilde{D}_0(t) = 0$ and $\widetilde{D}_j(t) = \sum_{k=1}^{j} \sin kt$ if $j = 1, 2, \ldots$) shows that

$$\left|\widetilde{K}_n(t)\right| \le n. \tag{12.53}$$

On the other hand, from the formula

$$\widetilde{D}_j(t) = \frac{1}{2} \cot \frac{1}{2} t - \frac{\cos\left(j + \frac{1}{2}\right) t}{2 \sin \frac{1}{2} t}$$

(see (12.26)), we find that for $0 < |t| \le \pi$,

$$\widetilde{K}_n(t) = \frac{1}{n+1} \left[\frac{n}{2} \cot \frac{1}{2} t - \frac{1}{2 \sin \frac{1}{2} t} \sum_{j=1}^{n} \cos\left(j + \frac{1}{2}\right) t \right],$$

$$\widetilde{K}_n(t) - \frac{1}{2} \cot \frac{1}{2} t = -\frac{1}{n+1} \frac{1}{2 \sin \frac{1}{2} t} \sum_{j=0}^{n} \cos\left(j + \frac{1}{2}\right) t.$$

Then, using the identity $2\sin\frac{t}{2}\cos\left(j+\frac{1}{2}\right)t = \sin(j+1)t - \sin jt$, we obtain

$$\widetilde{K}_n(t) - \frac{1}{2}\cot\frac{1}{2}t = \frac{-\sin(n+1)t}{(n+1)\left(2\sin\frac{1}{2}t\right)^2},$$

which shows that

$$\left|\widetilde{K}_n(t) - \frac{1}{2}\cot\frac{1}{2}t\right| \le \frac{A}{nt^2} \quad (|t| \le \pi,\ n = 1, 2, \ldots). \tag{12.54}$$

The estimates (12.53) and (12.54) are analogues of (12.52), and they easily lead to (ii). We write, for $n = 1, 2, \ldots$,

$$\begin{aligned}
&\widetilde{\sigma}_n(x_0) - \widetilde{f}_{1/n}(x_0) \\
&\quad = -\frac{1}{\pi}\int_{-\pi}^{\pi} f(x_0+t)\,\widetilde{K}_n(t)\,dt + \frac{1}{\pi}\int_{1/n\le|t|\le\pi} f(x_0+t)\,\frac{1}{2}\cot\frac{1}{2}t\,dt \\
&\quad = -\frac{1}{\pi}\int_{|t|<1/n} f(x_0+t)\,\widetilde{K}_n(t)\,dt \\
&\qquad + \frac{1}{\pi}\int_{1/n\le|t|\le\pi} f(x_0+t)\left[\frac{1}{2}\cot\frac{1}{2}t - \widetilde{K}_n(t)\right]dt
\end{aligned}$$

and use the estimates (12.53) and (12.54) in the last two integrals, respectively. An argument identical to that for α_n and β_n in the preceding proof shows that these integrals tend to 0. This completes the proof.

We remark that part (i) of Lebesgue's theorem leads to a new proof of the completeness of the trigonometric system (see Theorem 12.16). For if all the Fourier coefficients of f are 0, then $\sigma_n(x,f)$ vanishes identically and consequently $f = 0$ a.e. by part (i).

Part (ii) of Lebesgue's theorem shows that $\lim \widetilde{\sigma}_n(x_0)$ exists at every Lebesgue point of f at which the conjugate function

$$\widetilde{f}(x_0) = \lim_{\varepsilon\to 0} \widetilde{f}_\varepsilon(x_0)$$

exists. The converse is also true, though it requires an additional argument. Let $1/(n+1) \le \varepsilon \le 1/n$. Then

$$\begin{aligned}\left|\widetilde{f}_\varepsilon(x_0) - \widetilde{f}_{1/n}(x_0)\right| &\le \frac{1}{\pi}\int_{1/(n+1)}^{1/n} \left|f(x_0+t) - f(x_0-t)\right| \frac{1}{2}\cot\frac{1}{2}t\,dt \\ &\le \frac{1}{\pi}\int_{1/(n+1)}^{1/n} |f(x_0+t) - f(x_0-t)|\,\frac{dt}{t} \qquad (12.55)\\ &\le \frac{n+1}{\pi}\int_0^{1/n} |f(x_0+t) - f(x_0-t)|\,dt \to 0\end{aligned}$$

in view of the Lebesgue point condition. Hence, we obtain

Theorem 12.56 *At every Lebesgue point x_0 of an integrable f, the existence of $\widetilde{f}(x_0)$ is equivalent to the summability of $\widetilde{S}[f]$ by the method of the arithmetic mean, and $\widetilde{f}(x_0) = \lim \widetilde{\sigma}_n(x_0, f)$.*

Suppose now that f is not only integrable but also in L^2. If $f \sim \sum c_k e^{ikx}$, this means that $\sum |c_k|^2 < +\infty$. Observing that

$$\widetilde{S}[f] = \sum c_k \varepsilon_k e^{ikx}, \quad \varepsilon_k = -i\,\mathrm{sign}\,k$$

(see (12.12); recall that $\varepsilon_0 = 0$), we see by the Riesz–Fischer Theorem 8.30 that there is a function $g \in L^2$ such that $\widetilde{S}[f] = S[g]$ and

$$\frac{1}{2\pi}\int_0^{2\pi} |g|^2 = \sum |c_k \varepsilon_k|^2 .$$

Therefore,

$$\frac{1}{2\pi}\int_0^{2\pi} |g|^2 \le \sum |c_k|^2 = \frac{1}{2\pi}\int_0^{2\pi} |f|^2,$$

that is, $\|g\|_2 \le \|f\|_2$. Since $\widetilde{\sigma}_n(x,f) = \sigma_n(x,g)$ and $\lim \sigma_n(x,g)$ exists and equals g a.e. by Theorem 12.51, $\widetilde{f}$ exists and equals g a.e. by Theorem 12.56, and we have proved the following result.

Theorem 12.57 *If f is periodic and in L^2, then the conjugate function*

$$\widetilde{f}(x) = -\frac{1}{\pi}\int_0^\pi \frac{f(x+t)-f(x-t)}{2\tan\frac{1}{2}t}\,dt = -\frac{1}{\pi}\lim_{\varepsilon\to 0}\int_\varepsilon^\pi$$

exists a.e. and is in L^2. Moreover, $\|\widetilde{f}\|_2 \le \|f\|_2$ (more precisely, $\|\widetilde{f}\|_2^2 = \|f\|_2^2 - 2\pi|c_0|^2$) and $\widetilde{S}[f] = S[\widetilde{f}]$.

The existence a.e. of $\widetilde{f}$ is a remarkable result, which shows that the odd part of f,

$$\psi_x(t) = \frac{1}{2}[f(x+t)-f(x-t)],$$

has special properties that are not immediate consequences of the theory of integration. Observing that $\frac{1}{2}\cot\frac{1}{2}t-(1/t)$ is bounded for $0<t<\pi$, we deduce from Theorem 12.57 that if $f\in L^2$, then the integral

$$\int_0^\pi \frac{f(x+t)-f(x-t)}{t}\,dt = \lim_{\varepsilon\to 0}\int_\varepsilon^\pi$$

exists a.e., a result that is not obvious even for continuous f.

That $\widetilde{f}$ exists a.e. for f merely integrable will be proved later (see Theorem 12.67).

In the theorems that follow, we will use the notation

$$\|f\|_p = \left(\int_{-\pi}^{\pi}|f(x)|^p dx\right)^{1/p},\ 1\le p<\infty;\quad \|f\|_\infty = \operatorname*{ess\,sup}_{|x|\le\pi}|f(x)|$$

(although sometimes it may be convenient to modify the definition of $\|f\|_p$ by inserting a numerical factor; e.g., by writing $\|f\|_p = [(1/2\pi)\int_{-\pi}^{\pi}|f|^p dx]^{1/p}$).

Theorem 12.58 *If $f\in L^p$, then*

(i) $\|\sigma_n\|_p \le \|f\|_p$, $1\le p\le\infty$,
(ii) $\|f-\sigma_n\|_p \to 0$, $1\le p<\infty$.

Proof. The theorem and its proof are repetitions of Theorems 9.1 and 9.6. If $p=\infty$, (i) is a corollary of Theorem 12.48(a). If $1<p<\infty$ and p' is the exponent conjugate to p, we have

$$|\sigma_n(x)| \le \frac{1}{\pi}\int_{-\pi}^{\pi} |f(t)|K_n(x-t)^{1/p}K_n(x-t)^{1/p'}\,dt$$
$$\le \left[\frac{1}{\pi}\int_{-\pi}^{\pi} |f(t)|^p K_n(x-t)\,dt\right]^{1/p}\left[\frac{1}{\pi}\int_{-\pi}^{\pi} K_n(x-t)\,dt\right]^{1/p'},$$

by Hölder's inequality, and so by property (b) on p. 333 in Section 12.6,

$$|\sigma_n(x)|^p \le \frac{1}{\pi}\int_{-\pi}^{\pi} |f(t)|^p K_n(x-t)\,dt,$$

an inequality that clearly also holds for $p = 1$. Integrating both sides over $-\pi \le x \le \pi$, and interchanging the order of integration on the right, we obtain (i).

Part (ii) is proved similarly. We write

$$\left|\sigma_n(x) - f(x)\right| \le \frac{1}{\pi}\int_{-\pi}^{\pi} |f(x+t) - f(x)|K_n(t)\,dt$$
$$\le \left[\frac{1}{\pi}\int_{-\pi}^{\pi} |f(x+t) - f(x)|^p K_n(t)\,dt\right]^{1/p}\left[\frac{1}{\pi}\int_{-\pi}^{\pi} K_n(t)\,dt\right]^{1/p'},$$
$$\left|\sigma_n(x) - f(x)\right|^p \le \frac{1}{\pi}\int_{-\pi}^{\pi} \left|f(x+t) - f(x)\right|^p K_n(t)\,dt.$$

Integrating both sides over $-\pi \le x \le \pi$ and interchanging the order of integration on the right, we obtain

$$\|\sigma_n - f\|_p^p \le \frac{1}{\pi}\int_{-\pi}^{\pi} \phi(t)K_n(t)\,dt$$

where

$$\phi(t) = \int_{-\pi}^{\pi} |f(x+t) - f(x)|^p\,dx.$$

Clearly, ϕ is a bounded function, and we know by Theorem 8.19 that it tends to 0 with t. Hence, by Fejér's Theorem 12.47,

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \phi(t)K_n(t)\,dt = \sigma_n(0,\phi) \to 0,$$

and (ii) follows.

We conclude this section by considering the maximal arithmetic means defined by

$$\sigma^*(x) = \sigma^*(x,f) = \sup_{n\geq 0}|\sigma_n(x,f)|.$$

In view of Theorem 12.51(i), $\sigma^*(x)$ is finite a.e. It has properties not unlike those of the Hardy–Littlewood maximal function f^* considered in Chapters 7 and 9 and that are easily deducible from those of f^*. (Using similar symbols, σ^* and f^*, for different notions should not cause confusion.) First, we consider an adaptation of the definition of f^* to the case of periodic functions. For periodic f, it is natural to set

$$f^*(x) = \sup_{0<h\leq\pi}\frac{1}{2h}\int_{-h}^{h}|f(x+t)|\,dt. \tag{12.59}$$

Clearly, f^* is also periodic.

Theorem 12.60 *Let f be periodic and integrable. Then*

$$\|f^*\|_p \leq C_p\|f\|_p, \quad 1 < p \leq \infty,$$
$$|\{x : 0 \leq x \leq 2\pi;\ f^*(x) > \alpha\}| \leq \frac{c}{\alpha}\|f\|_1, \quad \alpha > 0.$$

These inequalities are analogues (actually, corollaries) of Theorem 9.16 and Lemma 7.9. Let $g(x)$ be defined as equal to $f(x)$ in $(-\pi, 3\pi)$ and to 0 elsewhere. Then, in $(0, 2\pi)$, the maximal function f^* just defined is majorized by the Hardy–Littlewood maximal function of g, and the norms of g in $(-\infty, +\infty)$ are majorized by multiples of the corresponding norms of f in $(0, 2\pi)$.

The first part of the next result is an analogue of Theorem 9.17.

Theorem 12.61 *Let f be periodic and integrable. Then there is an absolute constant c such that*

(i) $\sigma^*(x,f) \leq cf^*(x)$,

(ii) $\sup_{n\geq 1}|\widetilde{\sigma}_n(x,f) - \widetilde{f}_{1/n}(x)| \leq cf^*(x)$.

Proof. The proof can be based on either Tonelli's theorem (see the proof of Theorem 9.17) or on the formula for integration by parts. We choose the second approach since it follows the same line as the proof of Theorem 12.51, but

is actually easier since we do not have to consider Lebesgue points. Using the notation and proof of Theorem 12.51, we have

$$\left|\sigma_n(x,f)\right| \le \frac{1}{\pi}\int_0^\pi (|f(x+t)| + |f(x-t|)K_n(t)\,dt$$

$$= \frac{1}{\pi}\int_0^{1/n} + \frac{1}{\pi}\int_{1/n}^{\pi} = \alpha_n + \beta_n,$$

where

$$\alpha_n \le \frac{1}{\pi} n \int_0^{1/n} \phi(t)\,dt, \quad \beta_n \le \frac{A}{\pi n}\int_{1/n}^{\pi} \frac{\phi(t)}{t^2}\,dt$$

and $\phi(t) = |f(x+t)| + |f(x-t)|$. Let $\psi(t) = \int_0^t \phi(u)\,du$. The inequality $(\psi(t)/2t) \le f^*(x)$ shows that $\alpha_n \le f^*(x)$. If we integrate the integral majorizing β_n by parts so as to introduce $\psi(t)$ and again use the inequality $(\psi(t)/2t) \le f^*(x)$, we obtain $\beta_n \le Af^*(x)$, and (i) follows. The proof of (ii) is left to the reader.

The following result is a corollary of Theorem 12.61 and complements Theorem 12.57. It will be useful later.

Theorem 12.62 *If f is periodic and in L^2, then the maximal conjugate function defined by*

$$\widetilde{f}_*(x) = \sup_{0<\varepsilon\le\pi} \left|\widetilde{f}_\varepsilon(x)\right|$$

is also in L^2 and

$$\|\widetilde{f}_*\| \le A\|f\|_2 \quad (A \text{ independent of } f).$$

We put the asterisk as a subscript here to avoid confusion with $\left(\widetilde{f}\right)^*$, the Hardy–Littlewood maximal function of $\widetilde{f}$, which also appears in the proof of the theorem.

Proof. First suppose that $0 < \varepsilon \le 1/2$. Pick $n = 2, 3, \ldots$ such that $1/(n+1) \le \varepsilon \le 1/n$. The inequalities in (12.55) give

$$\left|\widetilde{f}_\varepsilon(x) - \widetilde{f}_{1/n}(x)\right| \le \frac{n+1}{\pi} \int_0^{1/n} \{|f(x+t)| + |f(x-t)|\}\, dt \le f^*(x),$$

where the final inequality is true since $n \ge 2$. Combining this with Theorems 12.61(ii) and 12.57, we find that (with different A's at different places)

$$\begin{aligned}\left|\widetilde{f}_\varepsilon(x)\right| &\le \sup_n |\widetilde{\sigma}_n(x)| + Af^*(x) = \sup_n \left|\sigma_n(x, \widetilde{f})\right| + Af^*(x) \\ &\le A\left\{\left(\widetilde{f}\right)^*(x) + f^*(x)\right\} \quad \text{(by Theorem 12.61(i))}.\end{aligned}$$

If instead $1/2 < \varepsilon \le \pi$, then

$$\begin{aligned}\left|\widetilde{f}_\varepsilon(x)\right| &= \left|\frac{1}{\pi} \int_{\varepsilon \le |t| \le \pi} f(x+t)\, \frac{1}{2\tan\frac{t}{2}}\, dt\right| \\ &\le \frac{1}{\pi} \int_{\frac{1}{2} \le |t| \le \pi} |f(x+t)|\, \frac{1}{\left|2\tan\frac{t}{2}\right|}\, dt \\ &\le A \int_{|t| \le \pi} |f(x+t)|\, dt \le Af^*(x).\end{aligned}$$

Collecting estimates, we obtain

$$\begin{aligned}\widetilde{f}_* &\le A\{(\widetilde{f})^*(x) + f^*(x)\}, \\ \|\widetilde{f}_*\|_2 &\le A\{\|\widetilde{f}\|_2 + \|f\|_2\}, \\ \|\widetilde{f}_*\|_2 &\le A\|f\|_2.\end{aligned}$$

12.7 Summability of $S[f]$ by Abel Means

Given a periodic and integrable f,

$$f \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

let

$$f(r,x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)r^n, \quad 0 \le r < 1,$$

denote its Abel means. Since summability by arithmetic means implies Abel summability, the results of the preceding section immediately lead to results about A-summability of $S[f]$. For example, we have the relation

$$f(r,x_0) \to f(x_0) \quad (r \to 1) \tag{12.63}$$

at every Lebesgue point of f, and so a.e. In particular, the last relation holds at each point of continuity of f and uniformly over every closed interval of continuity.

However, an independent discussion of Abel summability has some merits, if only for the following two reasons: (a) the relation (12.63) holds at points that need not be Lebesgue points (Theorem 12.64); (b) instead of (12.63), we may consider the more general relation

$$f(r,x) \to f(x_0)$$

as (r,x) tends to $(1,x_0)$ (i.e., as re^{ix} tends to e^{ix_0}) not only radially but also along more general curves, for example, nontangentially (see p. 328 in Section 12.5).

We now derive a representation for $f(r,x)$ as an integral operator. Using the formulas for a_n and b_n, we have

$$\begin{aligned} f(r,x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\,dt + \sum_{n=1}^{\infty} r^n \left[\cos nx \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt \right. \\ &\qquad \left. + \sin nx \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt\right] \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\left[\frac{1}{2} + \sum_{1}^{\infty} r^n \cos n(x-t)\right]dt \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)P(r,x-t)\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)P(r,t)\,dt, \end{aligned}$$

where

$$P(r,t) = \frac{1}{2} + \sum_{n=1}^{\infty} r^n \cos nt$$

is called the periodic *Poisson kernel*. The function $f(r,x)$ is called the *Poisson integral of* f. All the formal operations performed above (like the interchange of the order of summation and integration) are easily justifiable since $0 \le r < 1$.

We can write $P(r,t)$ in a finite form by observing that $\frac{1}{2} + \sum_{n=1}^{\infty} r^n \cos nt$ is the real part of the series

$$\frac{1}{2} + z + z^2 + \cdots = \frac{1}{2}\frac{1+z}{1-z} \quad \left(z = re^{it},\ 0 \le r < 1\right),$$

and a simple computation shows that

$$P(r,t) = \frac{1}{2}\frac{1-r^2}{1-2r\cos t + r^2} = \frac{1}{2}\frac{1-r^2}{(1-r)^2 + 4r\sin^2 \frac{1}{2}t}.$$

This may be compared to the nonperiodic version of the Poisson kernel given in (9.10). The Poisson kernel has all the properties of the Fejér kernel but is also much smoother. We list the following properties:

(a) $P(r,t) \ge 0$; $P(r,-t) = P(r,t)$.

(b) $(1/\pi)\int_{-\pi}^{\pi} P(r,t)\,dt = 1$.

(c) $P(r,t) \le 1/(1-r)$; $P(r,t) \le A(1-r)/t^2$ ($\frac{1}{2} \le r < 1$, $|t| \le \pi$, A an absolute constant).

Properties (a) and (b) are obvious. The first part of (c) is a corollary of

$$P(r,t) \le \frac{1}{2} + r + r^2 + \cdots = \frac{1}{2}\frac{1+r}{1-r} \le \frac{1}{1-r},$$

and the second part follows from

$$P(r,t) = \frac{1}{2}\frac{(1-r)(1+r)}{(1-r)^2 + 4r\sin^2 \frac{1}{2}t} \le \frac{1-r}{4r\sin^2 \frac{1}{2}t}.$$

The estimates (c) are analogues of the corresponding estimates for the Fejér kernel $K_n(t)$ if we identify $(1-r)$ and $1/n$. Thus, results for the Fejér means have analogues for Abel means, and the proofs are basically the same. We shall, however, not dwell on this point and shall limit ourselves to several results of a somewhat different nature.

Given a periodic and integrable f, we will systematically denote by F its indefinite integral; F need not be periodic. Besides the ordinary derivative of F at x,

$$F'(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h},$$

we will also consider the *symmetric derivative*

$$F_s'(x) = \lim_{h\to 0+} \frac{F(x+h) - F(x-h)}{2h}.$$

Clearly, the existence of $F'(x)$ implies that of $F_s'(x)$ and $F_s'(x) = F'(x)$. The converse is not true, however, as shown by the simple example $F(x) = |x|$ at $x = 0$. Using the notions of the even and odd parts of a function introduced on p. 316 in Section 12.3, we see that F_s' is the ordinary derivative at 0 of the odd part of F at the point x (x is fixed, differentiation is with respect to h, at $h = 0$). Also, since

$$F_s'(x) = \lim_{h\to 0+} \frac{1}{h} \int_0^h \frac{f(x+t) + f(x-t)}{2}\, dt,$$

$F_s'(x)$ is the ordinary derivative at 0 of the integral of the even part of f at the point x.

Theorem 12.64 *Let f be periodic and integrable, and let F be the integral of f.*

(i) *At any point x_0 where $F_s'(x_0)$ exists, finite or infinite, $S[f]$ is Abel summable to the value $F_s'(x_0)$.*

(ii) *At any point x_0 where F has an ordinary and finite derivative $F'(x_0)$, the Poisson integral $f(r,x)$ tends to $F'(x_0)$ as (r,x) tends nontangentially to $(1,x_0)$.*

Proof. (i) Suppose, as we may, that $x_0 = 0$. Write

$$\phi(t) = \frac{1}{2}[f(t) + f(-t)], \quad \psi(t) = \int_0^t \phi(u)\, du,$$

and note that $\lim_{t\to 0+}[\psi(t)/t] = F_s'(0)$. We have

$$f(r,0) = \frac{2}{\pi}\int_0^\pi \phi(t)P(r,t)\, dt = \frac{2}{\pi}\int_0^\delta \phi(t)P(r,t)\, dt + o(1)$$

for any fixed δ, $0 < \delta \le \pi$, in view of the second part of property (c) for $P(r,t)$. Integration by parts shows that the last integral equals

$$-\frac{2}{\pi}\int_0^\delta \psi(t)P'(r,t)\, dt + o(1),$$

where

$$P'(r,t) = \frac{d}{dt}P(r,t) = -\frac{\left(1-r^2\right) r \sin t}{\left(1-2r\cos t + r^2\right)^2}.$$

Since $-P' \geq 0$ in $(0,\pi)$, if $\psi(t)/t$ is contained between m and M in $(0,\delta)$, then the last integral is contained between m and M multiplied by

$$\begin{aligned} -\frac{2}{\pi}\int_0^\delta tP'(r,t)\,dt &= \frac{2}{\pi}\int_0^\delta P(r,t)\,dt + o(1) \\ &= \frac{2}{\pi}\int_0^\pi P(r,t)\,dt + o(1) = 1 + o(1). \end{aligned}$$

Collecting results, we see that the $\limsup$ and $\liminf$ of $f(r,0)$ as $r \to 1$ are contained between m and M. This gives (i) when $F_s'(0)$ is either finite or infinite.

(ii) Assume again that $x_0 = 0$. We may also assume that $F(0) = 0$. Let us show that it suffices to prove the result in case $F'(0) = 0$. Indeed, denote $F'(0) = \alpha$, set $g(x) = f(x) - \alpha$, and let G be the integral of g. Then $G'(0) = F'(0) - \alpha = 0$, and if we can prove that the Poisson integral of g, which equals $f(r,x) - \alpha$, converges to 0 as (r,x) tends nontangentially to $(1,0)$, then the proof will be complete. Thus, assume that $F'(0) = 0$.

Suppose that $(r,x) \to (1,0)$ nontangentially, that is, see Exercise 23(b), suppose that $r \to 1$ and $x \to 0$ in such a way that

$$\frac{|x|}{1-r} \leq C \tag{12.65}$$

for some positive constant C. Given $\varepsilon > 0$, choose δ so small that $|F(u)/u| \leq \varepsilon$ for $|u| \leq 2\delta$. Write

$$\begin{aligned} f(r,x) &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)P(r,t)\,dt = \frac{1}{\pi}\int_{-\delta}^{\delta} f(x+t)P(r,t)\,dt + o(1) \\ &= -\frac{1}{\pi}\int_{-\delta}^{\delta} F(x+t)P'(r,t)\,dt + o(1), \end{aligned}$$

using integration by parts and property (c) of P on p. 348 in Section 12.7. If $|x| \leq \delta$, then $|x+t| \leq 2\delta$ in the last integral, and the integral itself is majorized in absolute value by

$$\varepsilon \int_{-\delta}^{\delta} (|x| + |t|)\, |P'(r,t)|\, dt = 2\varepsilon |x| \int_{0}^{\delta} |P'(r,t)|\, dt + 2\varepsilon \int_{0}^{\delta} t|P'(r,t)|\, dt. \tag{12.66}$$

Since

$$\int_{0}^{\delta} |P'(r,t)|\, dt = P(r,0) - P(r,\delta) < P(r,0) \leq \frac{1}{1-r}$$

and

$$\begin{aligned}\int_{0}^{\delta} t|P'(r,t)|\, dt &= -\int_{0}^{\delta} tP'(r,t)\, dt = -[tP(r,t)]_0^{\delta} + \int_{0}^{\delta} P(r,t)\, dt \\ &\leq \int_{0}^{\pi} P(r,t)\, dt = \frac{1}{2}\pi,\end{aligned}$$

condition (12.65) implies that the right side of (12.66) is less than a fixed multiple of ε. Hence, $f(r,x)$ tends to 0 under the hypothesis (12.65). This completes the proof of (ii).

12.8 Existence of $\widetilde{f}$

In this section, we prove the following basic result.

Theorem 12.67 *If f is periodic and integrable, then the conjugate function*

$$\widetilde{f}(x) = \lim_{\varepsilon \to 0} \widetilde{f}_{\varepsilon}(x) = \lim_{\varepsilon \to 0} \left\{ -\frac{1}{\pi} \int_{\varepsilon \leq |t| \leq \pi} \frac{f(x+t)}{2\tan \frac{1}{2}t}\, dt \right\}$$

exists a.e. and is in weak L^1 : for $\alpha > 0$,

$$|\{x : |x| \leq \pi, |\widetilde{f}(x)| > \alpha\}| \leq \frac{c}{\alpha} \|f\|_1,$$

where c is independent of f and α.

Remark: If f is integrable, $\widetilde{f}$ need not be, as the following example shows. Let f be any periodic integrable function, nonnegative in $\left(0, \frac{1}{2}\pi\right)$ and zero elsewhere in $(-\pi, \pi)$. Then for $-\frac{1}{2}\pi < x < 0$,

$$\begin{aligned}
\widetilde{f}(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\frac{dt}{2\tan\frac{1}{2}(x-t)} = \frac{1}{\pi}\int_{0}^{\frac{1}{2}\pi} f(t)\frac{dt}{2\tan\frac{1}{2}(x-t)} \\
&\le \frac{1}{\pi}\int_{0}^{|x|} f(t)\frac{dt}{2\tan\frac{1}{2}(x-t)} = -\frac{1}{\pi}\int_{0}^{|x|} f(t)\frac{dt}{2\tan\frac{1}{2}(t-x)}, \\
|\widetilde{f}(x)| &\ge \frac{1}{\pi}\int_{0}^{|x|} f(t)\frac{dt}{2\tan\frac{1}{2}(t+|x|)} \ge \frac{1}{\pi 2\tan|x|}\int_{0}^{|x|} f(t)\,dt.
\end{aligned}$$

Now choosing $f(t) = \left(t\log^2 t\right)^{-1}$ $\left(=(d/dt)\left[\log(1/t)\right]^{-1}\right)$ for $0 < t < \frac{1}{2}$ and $f(t) = 0$ for $\frac{1}{2} \le t < \pi$, we obtain $\left|\widetilde{f}(x)\right| \ge c\,\left[|x|\log(1/|x|)\right]^{-1}$ for some constant $c > 0$ and all $x \in (-\pi/2, 0)$. Clearly, $f \in L$ but $\widetilde{f} \notin L$.

The lemma that follows is essential for the proof of Theorem 12.67.*

Lemma 12.68 (The Decomposition Lemma) *Let Q be a finite interval in $\mathbf{R}^1$ and suppose that $f \in L(Q)$, $f \ge 0$. Then for any α satisfying*

$$\alpha \ge \frac{1}{|Q|}\int_Q f, \tag{12.69}$$

there is a sequence of nonoverlapping intervals $Q_1, Q_2, \ldots$ contained in Q such that

(i) $\alpha < \frac{1}{|Q_k|}\int_{Q_k} f \le 2\alpha \quad (k = 1, 2, \ldots)$,
(ii) $f(x) \le \alpha$ *a.e. in* $P = Q - \bigcup Q_k$,
(iii) $|\bigcup Q_k| \le \frac{1}{\alpha}\int_{\bigcup Q_k} f \le \frac{1}{\alpha}\int_Q f$.

Proof. We split Q in half, obtaining two subintervals Q' of equal length. For each Q', there are only two possibilities: $|Q'|^{-1}\int_{Q'} f \le \alpha$ or $|Q'|^{-1}\int_{Q'} f > \alpha$. Since $|Q'| = \frac{1}{2}|Q|$, the hypothesis (12.69) implies that $|Q'|^{-1}\int_{Q'} f \le 2\alpha$. Thus, for each Q', we have either

* Lemma 12.68 is due to A.P. Calderón and A. Zygmund; see the remarks after its proof.

$$|Q'|^{-1}\int_{Q'} f \le \alpha \quad \text{or} \quad \alpha < |Q'|^{-1}\int_{Q'} f \le 2\alpha.$$

If Q' satisfies the first condition, we call it an interval of the first kind—otherwise, of the second kind.

We save any Q' of the second kind. If Q' is of the first kind, we may repeat the previous argument by splitting Q' into 2 equal parts Q''. For each Q'', we again have either

$$|Q''|^{-1}\int_{Q''} f \le \alpha \quad \text{or} \quad \alpha < |Q''|^{-1}\int_{Q''} f \le 2\alpha.$$

Saving those of the second kind, we repeat the procedure for each Q'' of the first kind, and so on.

Let $Q_1, Q_2, \ldots, Q_k, \ldots$ be the sequence of all the intervals of the second kind in the previous procedure. Clearly, the Q_k are nonoverlapping and satisfy condition (i). Also, each $x \in P = Q - \bigcup Q_k$ belongs to a sequence of intervals $\{\bar{Q}\}$ with $|\bar{Q}|$ tending to 0 such that $|\bar{Q}|^{-1}\int_{\bar{Q}} f \le \alpha$. Since the ratio $|\bar{Q}|^{-1}\int_{\bar{Q}} f$ tends to $f(x)$ a.e. in P, (ii) follows. Finally, writing the first inequality (i) in the form $\alpha|Q_k| \le \int_{Q_k} f$, and summing over k, we deduce (iii).

Remarks

(1) Lemma 12.68 holds for periodic functions of period 2π considered on the circumference of the unit circle, and $\alpha \ge (2\pi)^{-1}\int_{-\pi}^{\pi} f$. The intervals Q_k may be thought of as nonoverlapping arcs on the circumference with lengths $|Q_k| \le \pi$, and we make no distinction between an arc and its periodic translates. The proof is identical with that above.

(2) Lemma 12.68 is valid for $Q = \mathbf{R^1}$, $f \in L^1(\mathbf{R^1})$, and any $\alpha > 0$. Moreover, it has an analogue in $\mathbf{R^n}$, $n > 1$, where Q and the Q_k are taken to be n-dimensional cubes with edges parallel to the coordinate axes. The proof in case $Q = \mathbf{R^1}$, $f \in L^1(\mathbf{R^1})$, and α is any positive number is left to the reader. For the analogue in $\mathbf{R^n}$, $n > 1$, see Lemma 14.55. See also Exercise 23 in Chapter 14.

Proof of Theorem 12.67 Assume first that the periodic function $f \in L$ is nonnegative. Fix any $\alpha \ge (2\pi)^{-1}\int_{-\pi}^{\pi} f$ and apply remark (1). We then obtain a sequence of nonoverlapping arcs $Q_1, Q_2, \ldots$ on the circumference Q of the unit circle such that

$$\alpha < \frac{1}{|Q_k|}\int_{Q_k} f \le 2\alpha, \qquad f \le \alpha \text{ a.e. in } P = Q - \bigcup Q_k. \tag{12.70}$$

Make a decomposition

$$f = g + h,$$

where g is defined as equal to f in P and as $|Q_k|^{-1} \int_{Q_k} f$ on each arc Q_k. Hence, h equals 0 in P and $f - |Q_k|^{-1} \int_{Q_k} f$ on each Q_k. Using (12.70), we have

$$\begin{aligned} &\text{(a)}\ \ 0 \le g \le \alpha \text{ a.e. in } P, \quad 0 \le g \le 2\alpha \text{ in each } Q_k, \\ &\text{(b)}\ \ h = 0 \text{ in } P, \quad \int_{Q_k} h = 0, \\ &\text{(c)}\ \ |h| \le f + |Q_k|^{-1} \int_{Q_k} f \le f + 2\alpha \text{ in each } Q_k. \end{aligned} \tag{12.71}$$

In particular, since g is bounded a.e., and so is in L^2, $\widetilde{g}(x)$ exists a.e. and $\|\widetilde{g}\|_2 \le \|g\|_2$ by Theorem 12.57.

To study $\widetilde{h}$, we will use the Marcinkiewicz Theorem 6.17. That theorem was stated for functions defined on intervals of $\mathbf{R^1}$, while here we need it for functions defined on the circumference of the unit circle. Clearly, these two situations are identical. We restate the theorem.

Theorem 12.72 *Let F be a closed subset of the unit circumference Q, and let $G = Q - F$. Let $\delta(x)$ be the (circular) distance of the point $x \in Q$ from F. Then, for each $\lambda > 0$, the integral*

$$M_\lambda(x) = \int_Q \frac{\delta^\lambda(y)}{|x-y|^{1+\lambda}}\, dy = \int_G \frac{\delta^\lambda(y)}{|x-y|^{1+\lambda}}\, dy$$

is finite a.e. in F. Moreover,

$$\int_F M_\lambda(x)\, dx \le 2\lambda^{-1}|G|.$$

We will need the result only in case $\lambda = 1$.

We will now study the existence of $\widetilde{h} = \lim \widetilde{h}_\varepsilon$, using properties (12.71)(b), (c) of h. Let Q_k^* denote the interior of the arc Q_k expanded concentrically twice, and let t_k denote the center of Q_k and d_k its length. Let $Q^* = \bigcup Q_k^*$ and let P^* be the complement of Q^* in Q; P^* is closed. We will consider only points $x \in P^*$ until near the end of the proof.

Fix $x \in P^*$, and consider the equation

$$\widetilde{h}_\varepsilon(x) = \frac{1}{\pi} \int\limits_{\varepsilon < |x-t| \le \pi} h(t) \frac{1}{2} \cot \frac{1}{2}(x-t)\, dt.$$

Since $h = 0$ outside $\bigcup Q_k$, the last integral is a sum

(i) of integrals extended over those Q_k that are totally outside $(x-\varepsilon, x+\varepsilon)$, and

(ii) of at most two integrals extended over portions of those Q_k that contain the points $x \pm \varepsilon$.

We will investigate these two cases separately, beginning with (ii). The interval Q_k that, say, contains $x+\varepsilon$ is distant from x by $\le \varepsilon$ and at the same time by $\ge \frac{1}{2}d_k$ (since $x \in P^*$). Hence, $\frac{1}{2}d_k \le \varepsilon$, and the integral under consideration is majorized in absolute value by

$$\frac{1}{\pi} \int\limits_{x+\varepsilon}^{x+\varepsilon+d_k} \frac{|h(t)|}{|x-t|}\, dt \le \frac{1}{\pi\varepsilon} \int\limits_{x+\varepsilon}^{x+3\varepsilon} |h(t)|\, dt \le \frac{1}{3\varepsilon} \int\limits_0^{3\varepsilon} |h(x+t)|\, dt.$$

We have just presupposed implicitly that $\varepsilon < \frac{1}{3}\pi$ in order to guarantee that $|x-t| \le |2\tan\frac{1}{2}(x-t)|$ whenever t satisfies $|x-t| \le 3\varepsilon$ (since $|\theta| \le |\tan\theta|$ when $|\theta| < \pi/2$), but since we are primarily interested in small ε, this is an unimportant restriction. A similar estimate is valid for the interval Q_k containing $x-\varepsilon$. Using definition (12.59), we see that the contribution of (ii) is majorized by $2h^*(x)$.

We also notice that $h = 0$ in P^* so that by Lebesgue's theorem on the differentiability of integrals, the contribution of (ii) tends to 0 with ε at almost every $x \in P^*$.

Consider now any integral from (i). Since $\int_{Q_k} h = 0$ (see (12.71)(b)), it can be written in the form

$$\frac{1}{\pi} \int\limits_{Q_k} h(t) \frac{1}{2} \left[\cot \frac{1}{2}(x-t) - \cot \frac{1}{2}(x-t_k) \right] dt$$

(t_k is the midpoint of Q_k). Let $\mathscr{I}_k(x)$ denote the last integral with the integrand replaced by its absolute value. Denoting absolute constants by A, we have

$$\begin{aligned}
\mathscr{I}_k(x) &= \frac{1}{\pi}\int_{Q_k} |h(t)|\frac{1}{2}\frac{\left|\sin\frac{1}{2}(t-t_k)\right|}{\left|\sin\frac{1}{2}(x-t)\sin\frac{1}{2}(x-t_k)\right|}\,dt \\
&\le Ad_k\int_{Q_k} |h(t)|\,\frac{dt}{|x-t||x-t_k|} \qquad (12.73)\\
&\le \frac{Ad_k}{(x-t_k)^2}\int_{Q_k} |h(t)|\,dt,
\end{aligned}$$

where to arrive at the last term, we have used the fact that $|t-t_k| \le \frac{1}{2}d_k$ and $|x-t_k| \ge d_k$ (recall that $x \notin Q^*$ since $x \in P^*$), so that

$$\frac{1}{2} \le \frac{|x-t|}{|x-t_k|} \le \frac{3}{2} \quad (x \in P^*,\ t \in Q_k). \qquad (12.74)$$

Using the first inequality in (12.71)(c) and (12.70), we also have

$$\int_{Q_k} |h| \le 2\int_{Q_k} f \le 4\alpha|Q_k|. \qquad (12.75)$$

If $\delta(t)$ denotes the distance from t to P^*, then $\delta(t) \ge \frac{1}{2}d_k$ for $t \in Q_k$. Collecting results and using (12.73) through (12.75), we get the final estimates

$$\begin{aligned}
\mathscr{I}_k(x) &\le \frac{Ad_k}{(x-t_k)^2}4\alpha|Q_k| \le A\alpha\int_{Q_k}\frac{\delta(t)}{(x-t)^2}\,dt,\\
\sum_k \mathscr{I}_k(x) &\le A\alpha\sum_k\int_{Q_k}\frac{\delta(t)}{(x-t)^2}\,dt
\end{aligned}$$

for $x \in P^*$ and those intervals Q_k that are entirely contained in the complement of $(x-\varepsilon, x+\varepsilon)$.

The last sum is majorized by

$$A\alpha\int_{Q}\frac{\delta(t)}{(x-t)^2}\,dt,$$

a quantity that is finite a.e. in P^* by Theorem 12.72 with $\lambda = 1$. Hence, in view of our observation that the integrals in (ii) tend to 0 a.e. in P^*, $\widetilde{h} = \lim \widetilde{h}_\varepsilon$ exists a.e. in P^* and

$$|\widetilde{h}(x)| \le A\alpha \int_Q \frac{\delta(t)}{(x-t)^2}\, dt \quad \text{(a.e. in } P^*). \tag{12.76}$$

Since $\widetilde{f} = \widetilde{g} + \widetilde{h}$, and $\widetilde{g}$ exists a.e., $\widetilde{f}$ exists a.e. in P^*, that is, everywhere in Q with the exception of a set of measure at most

$$|Q^*| \le 2 \sum |Q_k| \le \frac{2}{\alpha} \int_Q f. \tag{12.77}$$

Taking α arbitrarily large, we see that $\widetilde{f}$ exists a.e.

The assumption $f \ge 0$ can be dropped by considering the decomposition $f = f^+ - f^-$.

We still have to prove the weak integrability of $\widetilde{f}$. This can be deduced from the estimates above, and we shall be brief. We may again assume that $f \ge 0$. Fix any $\alpha \ge (2\pi)^{-1} \int_Q f$, and consider the decomposition $f = g + h$ corresponding to this α. Then

$$\left\{x : \left|\widetilde{f}\right| > \alpha\right\} \subset \left\{x : |\widetilde{g}| > \frac{1}{2}\alpha\right\} \cup \left\{x : \left|\widetilde{h}\right| > \frac{1}{2}\alpha\right\} = S \cup T.$$

Recall that $0 \le g \le 2\alpha$; we also have $\int_Q g = \int_Q f$ since $\int_Q h = 0$ by (12.71)(b). Hence, by Tchebyshev's inequality,

$$\begin{aligned} |S| &\le \left(\frac{1}{2}\alpha\right)^{-2} \int_Q \widetilde{g}^2 \\ &\le \frac{4}{\alpha^2} \int_Q g^2 \le \frac{8}{\alpha} \int_Q g = \frac{8}{\alpha} \int_Q f. \end{aligned}$$

To estimate $|T|$, let $T_1 = T \cap P^*, T_2 = T \cap Q^*$. Thus, $|T| = |T_1| + |T_2|$, and by (12.77),

$$|T_2| \le |Q^*| \le \frac{2}{\alpha} \int_Q f.$$

On T_1, $|\widetilde{h}(x)|$ is majorized a.e. by (see (12.76))

$$A\alpha \int_Q \frac{\delta(t)}{(x-t)^2}\,dt = A\alpha \int_{Q^*} \frac{\delta(t)}{(x-t)^2}\,dt = A\alpha M(x),$$

and if $|\widetilde{h}(x)|$ is to be $\geq \frac{1}{2}\alpha$ there, then necessarily $M(x) \geq 1/(2A)$. But $M(x)$ is the integral $M_\lambda(x)$ of Theorem 12.72 corresponding to $\lambda = 1$, $G = Q^*$, $F = P^*$. Therefore, by the estimate given in Theorem 12.72 and (12.77),

$$\int_{P^*} M \leq 2|Q^*| \leq \frac{4}{\alpha}\int_Q f,$$

and by Tchebyshev's inequality, the subset of P^* where $M(x) \geq 1/(2A)$ has measure at most $2A\int_{P^*} M \leq 8A\alpha^{-1}\int_Q f$. Thus, $|T_1| \leq 8A\alpha^{-1}\int_Q f$, and collecting estimates, we have

$$\left|\left\{x : \left|\widetilde{f}(x)\right| > \alpha\right\}\right| \leq \frac{c}{\alpha}\|f\|_1 \quad (c = 10 + 8A).$$

This was proved for $\alpha \geq (2\pi)^{-1}\int_Q f$ but is trivially true for smaller α since our set certainly has measure $\leq 2\pi$, while $2\pi \leq \alpha^{-1}\int_Q f$ for such α.

This completes the proof of Theorem 12.67. The theorem is strengthened by the following result.

Theorem 12.78 *If $f \in L$, then the maximal conjugate function*

$$\widetilde{f}_*(x) = \sup_{0<\varepsilon\leq\pi} \left|\widetilde{f}_\varepsilon(x)\right|$$

is in weak L^1, that is,

$$\left|\left\{x : \widetilde{f}_*(x) > \alpha\right\}\right| \leq \frac{c}{\alpha}\|f\|_1 \quad (\alpha > 0),$$

where c is independent of f and α.

Proof. The proof is practically the same as that of Theorem 12.67. Leaving the details for the reader to fill, we argue as follows. We fix $f \geq 0$ and $\alpha \geq (2\pi)^{-1}\int_Q f$ and make the previous decomposition $f = g + h$, so that $\widetilde{f}_* \leq \widetilde{g}_* + \widetilde{h}_*$. By Theorem 12.62, $\|\widetilde{g}_*\|_2 \leq A\|g\|_2$, and the estimates we had for $\widetilde{h}$ also hold for $\widetilde{h}_*$. The restriction $\varepsilon < \frac{1}{3}\pi$ that was imposed in the argument is unimportant since if $\varepsilon \geq \frac{1}{3}\pi$, then $|\widetilde{h}_\varepsilon(x)| \leq c\int_Q f$ for all x, so

that $\sup_{\varepsilon \geq \pi/3} |\widetilde{h}_\varepsilon(x)| \leq c \int_Q f$. Therefore, the set where $\sup_{\varepsilon \geq \pi/3} |\widetilde{h}_\varepsilon(x)| > \alpha$ is empty if $\alpha \geq c \int_Q f$, while if $\alpha < c \int_Q f$, its measure is $\leq 2\pi < 2\pi c \alpha^{-1} \int_Q f$.

12.9 Properties of $\widetilde{f}$ for $f \in L^p, 1 < p < \infty$

Theorem 12.79 *If $f \in L^p, 1 < p < \infty$, then $\widetilde{f} \in L^p$ and $\widetilde{S}[f] = S[\widetilde{f}]$. Moreover,*

$$\text{(a)} \quad ||\widetilde{f}||_p \leq A_p ||f||_p \quad \textit{and} \quad \text{(b)} \quad ||\widetilde{f}_*||_p \leq A_p ||f||_p. \tag{12.80}$$

The constant A_p depends only on p and is bounded for p away from 1 and ∞.

This is the central theorem of the section. Its proof is long and has to be split into several parts. Of course, (b) implies (a). The theorem is false for $p = 1$ (see the remark after the statement of Theorem 12.67), and Theorem 12.67 is a substitute for this case. See also Exercise 21. The theorem is also false for $p = \infty$: see Exercises 19 and 20.

Lemma 12.81 *If $f \in L^p, 1 < p \leq 2$, then $\widetilde{f}_* \in L^p$ and (12.80) holds.*

Proof. For $p = 2$, this is Theorem 12.62. For $p = 1$, $\widetilde{f}_*$ is in weak L^1 by Theorem 12.67. The lemma will be obtained from these extreme cases by an interpolation argument similar to the one we used in the proof of the Hardy–Littlewood maximal function result in Theorem 9.16. We will use the same letter A to denote different absolute positive constants.

Let $f \in L^p, 1 < p < 2$. For each fixed $\alpha > 0$, we make the decomposition $f = g + h$ with

$$g(x) = f(x) \text{ wherever} |f(x)| \leq \alpha, \quad g(x) = 0 \text{ otherwise},$$
$$h(x) = f(x) \text{ wherever} |f(x)| > \alpha, \quad h(x) = 0 \text{ otherwise}.$$

Clearly, $g \in L^2, h \in L^1, \widetilde{f}_* \leq \widetilde{g}_* + \widetilde{h}_*$. Let $\omega(\alpha) = |\{\widetilde{f}_* > \alpha\}|$ be the distribution function of $\widetilde{f}_*$. We have

$$\omega(\alpha) \leq \left|\left\{\widetilde{g}_* > \alpha/2\right\}\right| + \left|\left\{\widetilde{h}_* > \alpha/2\right\}\right|.$$

By Theorem 12.67, with $Q = (-\pi, \pi)$,

$$\left|\left\{\widetilde{h}_* > \alpha/2\right\}\right| \leq A\,(\alpha/2)^{-1} \int_Q |h| = A\alpha^{-1} \int_{\{|f|>\alpha\}} |f|,$$

and by Tchebyshev's inequality and Theorem 12.62,

$$\begin{aligned}\left|\{\widetilde{g}_* > \alpha/2\}\right| &\leq \left(\alpha/2\right)^{-2} \int_Q \widetilde{g}_*^2 \\ &\leq A\alpha^{-2} \int_Q g^2 = A\alpha^{-2} \int_{\{|f|\leq\alpha\}} f^2.\end{aligned}$$

We have (see Theorem 5.51 and Exercises 16 of Chapter 5 and 5 of Chapter 6)

$$||\widetilde{f}_*||_p^p = -\int_0^\infty \alpha^p \, d\omega(\alpha) = p \int_0^\infty \alpha^{p-1} \, \omega(\alpha) \, d\alpha.$$

Using the estimates above, the last integral is majorized by

$$Ap \int_0^\infty \alpha^{p-1} \left(\alpha^{-1} \int_{\{|f|>\alpha\}} |f| \, dx \right) d\alpha + Ap \int_0^\infty \alpha^{p-1} \left(\alpha^{-2} \int_{\{|f|\leq\alpha\}} f^2 dx \right) d\alpha.$$

Interchanging the order of integration, we can write this as

$$\begin{aligned}&Ap \int_Q |f(x)| \left(\int_0^{|f(x)|} \alpha^{p-2} \, d\alpha \right) dx + Ap \int_Q f^2(x) \left(\int_{|f(x)|}^\infty \alpha^{p-3} \, d\alpha \right) dx \\ &= Ap \left\{ \frac{1}{p-1} \int_Q |f(x)|^p dx + \frac{1}{2-p} \int_Q |f(x)|^p dx \right\},\end{aligned}$$

due to the fact that $1 < p < 2$. It follows that (12.80)(b), and so also (12.80)(a), holds with

$$A_p^p = Ap \left\{ \frac{1}{p-1} + \frac{1}{2-p} \right\} \quad (1 < p < 2), \tag{12.82}$$

which proves the lemma.

It is not surprising that A_p becomes infinite as $p \to 1$ since as we have observed, the integrability of f does not imply that of $\widetilde{f}$. On the other hand,

in view of the validity of (12.80)(b) for $p = 2$, it is natural to expect that there is a better estimate for A_p near $p = 2$ than (12.82) (which becomes infinite as $p \to 2$). This we shall see below.

Meanwhile, note that the exponent 2 plays a rather accidental role in Lemma 12.81. In fact, if we knew (12.80)(b) for any $p_0 > 1$, then the previous argument would give it for $1 < p \le p_0$, with A_p bounded for p away from $p = 1$ and $p = p_0$. The only change necessary in the proof is replacing the exponent 2 in the argument for g by p_0. Moreover, if instead of (12.80)(b) we only knew (12.80)(a) for some p_0, then by applying the same argument to $\widetilde{f}$ instead of $\widetilde{f}_*$, we would obtain (12.80)(a) for $1 < p \le p_0$, with A_p bounded for p away from 1 and p_0. We will use this idea below.

Inequality (12.80)(b) for any p implies

$$||\widetilde{f}_\varepsilon||_p \le A_p||f||_p \tag{12.83}$$

for the same p and all $\varepsilon > 0$. From the fact that $||\widetilde{f}_\varepsilon||_p$ equals $\sup_g |\int_Q \widetilde{f}_\varepsilon g|$ for all g with $||g||_{p'} \le 1$, $1/p + 1/p' = 1$, and the easily verifiable formula

$$\int_Q \widetilde{f}_\varepsilon\, g = -\int_Q f\, \widetilde{g}_\varepsilon$$

(apply Fubini's theorem), it follows that if (12.83) holds for any $p, 1 < p < \infty$, then it also holds for the conjugate exponent p', and $A_{p'} = A_p$ (see also Exercise 16 of Chapter 10). But we proved (12.83) for $1 < p \le 2$. Hence, it also holds for $2 \le p < \infty$, and an observation analogous to the one made earlier just before (12.83), but this time for $\widetilde{f}_\varepsilon$ rather than $\widetilde{f}$, shows that the constant A_p in (12.83) remains bounded for p away from 1 and ∞.

Using (12.82), we thus see that the A_p in (12.83) satisfies the inequality

$$A_p \le \frac{A}{p-1} \quad (1 < p \le 2), \tag{12.84}$$

and so also (since $A_p = A_{p'}$)

$$A_p \le Ap \quad \text{for } p \ge 2. \tag{12.85}$$

Since $\widetilde{f}_\varepsilon(x) \to \widetilde{f}(x)$ a.e. as $\varepsilon \to 0$ for any integrable f, (12.83) leads to the basic inequality $||\widetilde{f}||_p \le A_p||f||_p, 1 < p < \infty$, where A_p satisfies the two previous estimates. This proves (12.80)(a); we still have to prove (12.80)(b) for $2 < p < \infty$ and the formula $\widetilde{S}[f] = S[\widetilde{f}]$ for $f \in L^p$, $p > 1$.

Let us show that if $f \in L^p$, $p > 1$, then $\widetilde{S}[f] = S[\widetilde{f}]$. We may assume that $p < 2$. In view of Theorem 12.56, $\lim \widetilde{\sigma}_n(x)$ exists and equals $\widetilde{f}(x)$ a.e. for f merely integrable. Next, by Theorem 12.61(ii), $|\widetilde{\sigma}_n(x)|$ is majorized by $\widetilde{f}_*(x) + cf^*(x)$,

an integrable function in our case since $f \in L^p, 1 < p < 2$. Hence, by the dominated convergence theorem, each Fourier coefficient of $\widetilde{\sigma}_n$ tends to the corresponding Fourier coefficient of $\widetilde{f}$. But it also tends to the corresponding coefficient of $\widetilde{S}[f]$; in fact by (12.39), if $f \sim \sum c_k e^{ikx}$, then the kth Fourier coefficient of $\widetilde{\sigma}_n$ equals $[1 - |k|/(n+1)](-i \operatorname{sign} k)c_k$ if $|k| \le n$ and equals 0 if $|k| > n$. Thus, $\widetilde{S}[f] = S[\widetilde{f}]$.

To complete the proof of Theorem 12.79, it remains to show that (12.80)(b) holds for $1 < p < \infty$ (we have shown it only for $1 < p \le 2$), with A_p bounded for p away from 1 and ∞. It is immediate (see the proof of Theorem 12.62) that if $1/(n+1) \le \varepsilon \le 1/n$, then $|\widetilde{f}_\varepsilon(x) - \widetilde{f}_{1/n}(x)|$ is majorized by $f^*(x)$. Hence, by Theorem 12.61(ii),

$$\widetilde{f}_*(x) \le cf^*(x) + \sup_n |\widetilde{\sigma}_n(x)| .$$

By Theorem 12.61(i) applied to $\widetilde{S}[f] = S[\widetilde{f}]$,

$$\sup_n \left|\widetilde{\sigma}_n(x,f)\right| = \sigma^* \left(x,\widetilde{f}\right) \le c \left(\widetilde{f}\right)^* (x).$$

Hence,

$$\begin{aligned} \widetilde{f}_* &\le c \left\{f^* + \left(\widetilde{f}\right)^*\right\}, \\ ||\widetilde{f}_*||_p &\le c \left\{||f^*||_p + \left|\left|\left(\widetilde{f}\right)^*\right|\right|_p\right\} \\ &\le c\, C_p\{||f||_p + ||\widetilde{f}||_p\}, \end{aligned}$$

where C_p is the constant of the Hardy–Littlewood maximal Theorem 12.60. In view of (12.80)(a), the inequality $||\widetilde{f}_*||_p \le B_p||f||_p$ is thus established for all $p, 1 < p < \infty$, with $B_p = cC_p[1 + A_p]$, A_p being the constant in (12.80)(a). Combining the estimates (12.84) and (12.85) for A_p with the estimate for C_p on p. 228 in Section 9.3, it follows that B_p is bounded for p away from 1 and ∞. In particular, since C_p remains bounded as $p \to \infty$, it follows that B_p is $O(p)$ as $p \to \infty$. To estimate B_p as $p \to 1$, it is best to use the estimate derived in the proof of Lemma 12.81 (see (12.82)); thus, $B_p = O(1/(p-1))$ as $p \to 1$, so that B_p satisfies the same sorts of estimates as A_p. This completes the proof of Theorem 12.79.

We have the following important corollary.

Corollary 12.86 *Let $f \in L^p$, $1 < p < \infty$. Then $\widetilde{f}_\varepsilon$ converges in L^p norm to $\widetilde{f}$.*

This is an immediate consequence of the dominated convergence theorem since $\widetilde{f}_\varepsilon$ converges pointwise a.e. to $\widetilde{f}$ and $\left|\widetilde{f}_\varepsilon\right| \le \widetilde{f}_* \in L^p$.

12.10 Application of Conjugate Functions to Partial Sums of S[f]

The behavior of the partial sums $s_n(x) = s_n(x,f)$ is a much more delicate topic than the behavior of the arithmetic means. We will consider only the question of the convergence of s_n to f in the metric L^p. The main tool here is a connection between the partial sums and the conjugate function.

Instead of s_n, it will be convenient to consider the expressions (see (12.27))

$$s_n^{\#}(x) = \frac{s_n(x) + s_{n-1}(x)}{2} = s_n(x) - \frac{1}{2}(a_n \cos nx + b_n \sin nx) \tag{12.87}$$

when $n \geq 1$. We have

$$\begin{aligned} s_n^{\#}(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\frac{D_n(x-t) + D_{n-1}(x-t)}{2}\,dt \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\frac{\sin n(x-t)}{2\tan\frac{1}{2}(x-t)}\,dt \\ &= \sin nx \cdot \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{f(t)\cos nt}{2\tan\frac{1}{2}(x-t)}\,dt - \cos nx \cdot \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{f(t)\sin nt}{2\tan\frac{1}{2}(x-t)}\,dt. \end{aligned}$$

The last decomposition was purely formal, but from Section 12.8, we know that the cofactors of $\sin nx$ and $\cos nx$ exist a.e. in the principal value sense (and represent the conjugate functions of $f(t)\cos nt$ and $f(t)\sin nt$, respectively).

Theorem 12.88 (M. Riesz) *If $f \in L^p, 1 < p < \infty$, then*

(i) $||s_n||_p \leq c||f||_p, \quad ||f - s_n||_p \to 0,$

(ii) $||\widetilde{s}_n||_p \leq c||f||_p, \quad ||\widetilde{f} - \widetilde{s}_n||_p \to 0,$

where c depends only on p.

Proof. It is enough to prove (i), which implies (ii) since $\widetilde{s}_n(x,f) = s_n(x,\widetilde{f})$ and $||\widetilde{f}||_p \leq A_p||f||_p$, $1 < p < \infty$. The first part of (i) with $s_n^{\#}$ instead of s_n is an immediate corollary of the last formula for $s_n^{\#}$ and the inequality

$\|\widetilde{f}\|_p \le A_p\|f\|_p$: in fact, letting $g_n(t) = f(t)\cos nt$, $h_n(t) = f(t)\sin nt$ in the formula, we obtain

$$s_n^{\#}(x) = \sin nx\,\widetilde{g}_n(x) - \cos nx\,\widetilde{h}_n(x) \quad \text{(a.e.)},$$

$$||s_n^{\#}||_p \le ||\widetilde{g}_n||_p + ||\widetilde{h}_n||_p \le A_p(||g_n||_p + ||h_n||_p) \le A_p||f||_p$$

for $1 < p < \infty$. Since, by (12.87), $\|s_n^{\#} - s_n\|_p \le A\|f\|_1 \le A\|f\|_p$, we obtain $\|s_n\|_p \le A_p\|f\|_p$.

The relation $||f - s_n||_p \to 0$ is obvious for functions that have a continuous derivative (see Theorem 12.20), and since such functions are dense in L^p, it also follows in the general case.

Exercises

1. Prove the following versions of Theorems 12.13 and 12.14.
   (a) If $f \sim \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx)$, and f is the indefinite integral of f', then $f' \sim \sum_{k=1}^{\infty} k(b_k \cos kx - a_k \sin kx)$.
   (b) If $f \sim \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx)$, and F is the indefinite integral of f, then

$$F(x) - \frac{1}{2}a_0 x \sim \frac{1}{2}A_0 + \sum_{k=1}^{\infty}\frac{1}{k}(a_k \sin kx - b_k \cos kx).$$

2. Prove that if $f(x) \sim \sum c_k e^{ikx}$, then $f(x+\alpha) \sim \sum c_k e^{ik\alpha}e^{ikx}$.
3. If f is real or complex-valued and $f \sim \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx)$, show that

$$\frac{1}{\pi}\int_0^{2\pi}|f|^2 = \frac{1}{2}|a_0|^2 + \sum_{k=1}^{\infty}\left(|a_k|^2 + |b_k|^2\right).$$

4. Deduce from (12.18) that if $f, g \in L^2, f \sim \sum c_k e^{ikx}, g \sim \sum d_k e^{ikx}$, then

$$\frac{1}{2\pi}\int_0^{2\pi} f(t)g(x-t)\,dt = \sum c_k d_k e^{ikx}.$$

Thus, a Fourier coefficient of the convolution of f and g equals the product of the corresponding coefficients of f and g. (For nonperiodic versions of this result, see Theorems 13.30 and 13.59.)

5. Prove the following.

(a) If $f(x)$ is periodic and equal to sign x in $(-\pi, \pi)$, then

$$f(x) \sim \frac{4}{\pi}\left\{\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right\}.$$

(b) Let $0 < h < \frac{1}{2}\pi$, and let f be the triangular function defined as follows: f is periodic, even, continuous, $f(0) = 1, f(x) = 0$ for $2h \le x \le \pi$, and f is linear in $(0, 2h)$. Then

$$f \sim \frac{2h}{\pi}\left[\frac{1}{2} + \sum_{k=1}^{\infty}\left(\frac{\sin kh}{kh}\right)^2 \cos kx\right] = \frac{h}{\pi}\left[1 + \sum_{-\infty}^{+\infty}{}' \left(\frac{\sin kh}{kh}\right)^2 e^{ikx}\right].$$

(c) Let g be periodic and equal to $\frac{1}{2}\log\left(1/\left|2\sin\frac{1}{2}x\right|\right)$ in $(-\pi, \pi)$. Then

$$g \sim \sum_{k=1}^{\infty} \frac{\cos kx}{k}.$$

(For (b), the coefficients can be computed directly, or by using Exercise 4 and Example (b), Section 12.1, p. 306. Observe that the convolution of the characteristic function of an interval with itself is a triangular function. For (c), one may either integrate by parts in the formula for the cosine coefficients of g or consider the real part of the series

$$\sum_{k=1}^{\infty} \frac{z^k}{k} = \log\frac{1}{1-z}, \quad z = re^{ix},$$

for $r < 1$, and then let $r \to 1$.)

6. Using the formula for the Fourier series of $\frac{1}{2}(\pi - x)$ given in Example (a), Section 12.1, p. 306, prove the formulas

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}, \quad \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90},$$

$$\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = \pi^{2k} B_k \text{ for some rational } B_k \quad (k = 1, 2, \ldots).$$

7. Prove that each of the systems
(a) $\frac{1}{2}, \cos x, \cos 2x, \ldots, \cos kx, \ldots$
(b) $\sin x, \sin 2x, \ldots, \sin kx, \ldots$
is orthogonal and complete over $(0, \pi)$.

8. Let $\{\phi_j(\mathbf{x})\}$ and $\{\psi_k(\mathbf{y})\}$ be two orthogonal systems: the first over a set $A \subset \mathbf{R}^m$ and the second over a set $B \subset \mathbf{R}^\mathbf{n}$. Then the (double) system

$$\omega_{j,k}(\mathbf{x}, \mathbf{y}) = \phi_j(\mathbf{x})\psi_k(\mathbf{y})$$

is orthogonal over the Cartesian product $C = A \times B \subset \mathbf{R}^{m+n}$. If both $\{\phi_j\}$ and $\{\psi_k\}$ are complete, so is $\{\omega_{j,k}\}$.

Generalize this to the case of more than two orthogonal systems.

9. Let $\{(k_1, k_1, \ldots, k_n)\}$ be all lattice points in the space $\mathbf{R}^\mathbf{n}$ (i.e., all distinct points with integral coordinates). Then the system

$$\exp\{i\,(k_1x_1 + k_2x_2 + \cdots + k_nx_n)\}$$

is orthogonal and complete over any cube in $\mathbf{R}^\mathbf{n}$ with edge length 2π.

10. If $f \sim \sum c_k e^{ikx} \in L^2$ and $g \sim \sum d_k e^{ikx} \in L^2$, then $h = fg$ is integrable, and if $h \sim \sum C_n e^{inx}$, then

$$C_n = \sum_{k=-\infty}^{+\infty} c_k d_{n-k},$$

where the series on the right converges absolutely. (For $n = 0$, this means $(1/2\pi)\int_0^{2\pi} fg = \sum c_k d_{-k}$, which is a variant of (12.18).) See also Exercises 17 and 18.

11. Let $f \sim \sum c_k e^{ikx} \in L^2$. For each n, let

$$\gamma_n = \sum_{k \neq n} c_k \frac{1}{n-k} = \sum_{k \neq 0} c_{n-k} \frac{1}{k}.$$

Show that $\{\gamma_n\} \in l^2$ and, more precisely, that $\sum |\gamma_n|^2 \leq \pi^2 \sum |c_k|^2$. The numbers γ_n are discrete analogues of the formal Hilbert integral $\int [f(t)/(x-t)]\,dt$; see Section 3 of Chapter 13. (The numbers γ_n are the Fourier coefficients of the function fg, where $g \sim \sum_{k \neq 0} e^{ikx}/k = i(\pi - x)$ for $0 < x < 2\pi$ (see Example (a), Section 12.1, p. 306), and we have

$$\int_0^{2\pi} |fg|^2 \leq \pi^2 \int_0^{2\pi} |f|^2.)$$

12. Let f and g be periodic, $f \in L^p, g \in L^{p'}, 1 \le p \le \infty, 1/p + 1/p' = 1$. Consider the product $h_x(t) = f(x+t)g(t)$ as a function of t. Show that the nth Fourier coefficient of $h_x(t)$ tends to 0 as $n \to \infty$, uniformly in x. (Show that the L^1-modulus of continuity of h tends to 0 uniformly in x; apply (12.22).)

13. Let f be periodic and integrable. Show that for the partial sums s_n and $\widetilde{s}_n$, we have the following formulas:

(a) $s_n(x) = (1/\pi) \int_{-\pi}^{\pi} f(x+t)\,[(\sin nt)/t]\,dt + \varepsilon_n(x)$, where $\varepsilon_n(x)$ tends to 0 uniformly in x as $n \to \infty$.

(b) $\widetilde{s}_n(x) = (1/\pi) \int_{-\pi}^{\pi} f(x+t)\,[(1-\cos nt)/t]\,dt + \eta_n(x)$, where $\eta_n(x)$ tends to 0 uniformly in x.

(For (a), except for an error which is $o(1)$ uniformly in x, we have

$$s_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+t) \frac{\sin nt}{2 \tan \frac{1}{2} t}\, dt,$$

and since $\frac{1}{2} \cot \frac{1}{2} t - \frac{1}{t}$ is bounded in $(-\pi, \pi)$, the result follows easily from Exercise 12 with $p = 1, p' = \infty$.)

14. Let $\widetilde{L}_n = 2\pi^{-1} \int_0^{\pi} |\widetilde{D}_n(t)|\,dt$, $\widetilde{D}_n$ denoting the conjugate Dirichlet kernel. Prove the following analogue of Theorem 12.36:

$$\widetilde{L}_n = \frac{2}{\pi} \log n + o(\log n) \quad \text{as } n \to \infty.$$

Show also that $\widetilde{L}_n = |\widetilde{s}_n(0, \operatorname{sign} \widetilde{D}_n)| = \max\{|\widetilde{s}_n(0,f)| : f \text{ with } |f| \le 1\}$.

15. Show that if x is a point of continuity of an integrable f, then $\widetilde{s}_n(x) = o(\log n)$. If f has a jump discontinuity at x and the value of the jump is $d = f(x+) - f(x-)$, show that $\widetilde{s}_n(x) = -(d/\pi) \log n + o(\log n)$. (The second statement follows from the first by considering [if, e.g., $x = 0$] the function $h(t) = \frac{1}{2}(\pi - t)$ of Example (a), Section 12.1, p. 306, whose conjugate function $\widetilde{h}$ satisfies $\widetilde{h} = -g$, where g is the function of Exercise 5(c).)

16. Prove the following more general form of Theorem 12.38. Suppose that $\liminf s_n = \underline{s}$ and $\limsup s_n = \bar{s}$ are finite. Then both $\liminf \sigma_n$ and $\limsup \sigma_n$ are contained between the numbers

$$\frac{1}{2}(\underline{s} + \bar{s}) \pm A\, \frac{1}{2}(\bar{s} - \underline{s}),$$

where A is the same as in condition (i) of Theorem 12.38.

17. Prove the following extension of Exercise 10. Let $f \in L^p$ and $g \in L^{p'}$, $1 < p < \infty$, $1/p + 1/p' = 1$ (thus also $1 < p' < \infty$). If $f \sim \sum c_k e^{ikx}$ and

$g \sim \sum d_k e^{ikx}$, then the Fourier coefficients C_n of the (integrable) function $h = fg$ are given by

$$C_n = \sum_{k=-\infty}^{\infty} c_k d_{n-k} = \lim_{M\to+\infty} \sum_{k=-M}^{M} .$$

Thus, the C_n are the same as in Exercise 10, but the series representing the C_n are no longer claimed to be absolutely convergent, and we must consider the limits of their *symmetric* partial sums. (The proof is parallel to that of Exercise 10. We write

$$C_n = \frac{1}{2\pi}\int_0^{2\pi} fge^{-inx} = \frac{1}{2\pi}\int_0^{2\pi} (f - s_M)ge^{-inx} + \frac{1}{2\pi}\int_0^{2\pi} s_M ge^{inx},$$

where $s_M = s_M(x,f)$, observe that the first integral on the right is majorized by $\int_0^{2\pi} |f - s_M|\,|g|$, apply Hölder's inequality, and use the fact that $||f - s_M||_p \to 0$ by Theorem 12.88.)

18. The result of Exercise 17 is valid when $p = 1, p' = \infty$ (or when $p = \infty, p' = 1$), but the series defining the C_n must be taken in the sense of the (symmetric) first arithmetic means:

$$C_n = \lim_{M\to+\infty} \sum_{k=-M}^{M} c_k \left(1 - \frac{|k|}{M+1}\right) d_{n-k}.$$

19. We know that Theorem 12.79 is false for $p = 1$. Show that it is also false for $p = \infty$, that is, that the conjugate function of a bounded function need not be bounded. (Consider, e.g., the two series $\sum_{k=1}^{\infty}(\sin kx)/k \sim \frac{1}{2}(\pi - x), 0 < x < 2\pi$, and $\sum_{k=1}^{\infty}(\cos kx)/k \sim \frac{1}{2}\log(1/|2\sin\frac{1}{2}x|), -\pi < x < \pi$ [see Exercise 5].)

20. There is a substitute result for Theorem 12.79 in case $p = \infty$. Let f be a periodic function with $|f| \le 1$. Then there are absolute constants $\lambda, \mu > 0$ such that

$$\int_{-\pi}^{\pi} e^{\lambda|\widetilde{f}|} \le \mu.$$

See also Exercise 27 in Chapter 14. (Write

$$e^{\lambda|\widetilde{f}|} - \lambda|\widetilde{f}| - 1 = \sum_{n=2}^{\infty} \frac{\lambda^n |\widetilde{f}|^n}{n!}$$

and integrate termwise. Use (12.80)(a), (12.85), and the fact that $n^n \leq c^n n!$.)

21. Suppose that f is a periodic function for which

$$\int_{-\pi}^{\pi} |f| \log^+ |f| < +\infty,$$

where $\log^+$ stands for the positive part of log. This clearly implies that $f \in L^1$. Show that $\widetilde{f} \in L^1$ and that there are absolute constants A and B such that

$$\int_{-\pi}^{\pi} |\widetilde{f}| \leq A \int_{-\pi}^{\pi} |f| \log^+ |f| \, dx + B.$$

(Write $\omega(\alpha) = |\{x : |x| < \pi, |\widetilde{f}(x)| > \alpha\}|$. Then

$$\int_{-\pi}^{\pi} \left|\widetilde{f}\right| = \int_0^{\infty} \omega(\alpha) \, d\alpha = \int_0^2 + \int_2^{\infty}.$$

For the first integral on the right, use the fact that $\omega(\alpha) \leq 2\pi$, and for the second, use an argument like that in the proof of Lemma 12.81.)

22. The discussion that precedes Theorem 12.31 shows only that the averages $(\widetilde{s}_n(x) + \widetilde{s}_{n-1}(x))/2$ converge. Give the remaining details of the proof of Theorem 12.31. (Consider $\widetilde{s}_n(x) - \widetilde{s}_{n-1}(x)$.)

23. (a) Prove the statement about nontangential approach made before Theorem 12.42, namely, that given $\gamma > 0$, there exist $C, \delta > 0$ with $\delta < 1$ such that $|1 - z| \leq C(1 - |z|)$ if $z = x + iy$, $|z| < 1$, $1 - \delta < x < 1$ and $|y| < \gamma(1 - x)$. (It may be helpful to use the asymptotic estimates $|\sin\theta| \simeq |\theta|$ and $1 - \cos\theta \simeq \theta^2/2$ as $\theta \to 0$.)

(b) Verify that (12.65) characterizes the nontangential approach of $z = re^{ix}$ to 1, $|z| < 1$.

13

The Fourier Transform

In this chapter, we will study properties of the Fourier transform $\widehat{f}(\mathbf{x})$ of a function f on $\mathbf{R^n}$, $n \geq 1$, defined formally (for the moment) as

$$\widehat{f}(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} f(\mathbf{y})\, e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n}. \tag{13.1}$$

Here $\mathbf{x} \cdot \mathbf{y} = \sum_1^n x_k y_k$ is the usual dot product of $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, and i is the complex number $i = \sqrt{-1} = e^{i\pi/2}$. Both f and $\widehat{f}$ may be complex-valued.

Different normalizations of $\widehat{f}$ are common in the literature, such as

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbf{R^n}} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}, \quad \int_{\mathbf{R^n}} f(\mathbf{y}) e^{2\pi i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}, \ldots,$$

but the important properties of $\widehat{f}$ are unaffected by normalization, and passing from one normalization to another is easy by scaling.

We will often abuse notation by denoting

$$\widehat{f}(\mathbf{x}) = \widehat{f(\mathbf{x})} \quad \text{or} \quad \widehat{f}(\mathbf{x}) = \big(f(\mathbf{x})\big)\widehat{\ }$$

instead of the more cumbersome notations $\widehat{f(\mathbf{y})}(\mathbf{x})$, $(f(\cdot))\widehat{\ }(\mathbf{x})$, etc. For example, we will do this in Theorem 13.8 when computing the Fourier transform of $e^{-|\mathbf{x}|^2}$ since the notations $\widehat{e^{-|\mathbf{x}|^2}}$ and $(e^{-|\mathbf{x}|^2})\widehat{\ }$ are somewhat simpler than $(e^{-|\cdot|^2})\widehat{\ }(\mathbf{x})$.

Note that $\widehat{f}(\mathbf{x})$ is a formal analogue of the sequence $\{c_j\}_{-\infty}^{\infty}$ of trigonometric Fourier coefficients of a periodic function on the line, with the continuous variable $\mathbf{x}$ now playing the role of j.

One of our main goals is to derive an analogue of Parseval's formula (12.15), that is, to prove that the mapping $f \to \widehat{f}$ is essentially an isometry on $L^2(\mathbf{R^n})$. Of course, an important requirement for achieving this is to find an interpretation of $\widehat{f}$ in case $f \in L^2(\mathbf{R^n})$. Unlike the formulas for Fourier coefficients in the one-dimensional periodic case, the integral in (13.1) may not converge absolutely for every $f \in L^2(\mathbf{R^n})$. However, as is easy to see, (13.1) does converge absolutely if $f \in L^1(\mathbf{R^n})$. Properties of $\widehat{f}$ when $f \in L^1(\mathbf{R^n})$ are simpler to derive precisely because of this absolute convergence, and we will

begin with that case. Furthermore, properties of $\widehat{f}$ when $f \in L^1(\mathbf{R^n})$ will be useful later for studying $\widehat{f}$ for other classes of functions f.

13.1 The Fourier Transform on L^1

In this section, we list some properties of the Fourier transform of functions in $L^1(\mathbf{R^n})$.

(1) Let $f \in L^1(\mathbf{R^n})$, $n \geq 1$. Define $\widehat{f}(\mathbf{x})$ by (13.1) and note that (cf. Exercise 1 of Chapter 8)

$$\begin{aligned} |\widehat{f}(\mathbf{x})| &= \left| \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y} \right| \\ &\leq \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} |f(\mathbf{y})|\, d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n}. \end{aligned}$$

Thus, the integral in (13.1) converges absolutely (i.e., exists in the usual Lebesgue sense) for every $f \in L^1(\mathbf{R^n})$ and every $\mathbf{x} \in \mathbf{R^n}$, and the mapping $f \to \widehat{f}$ sends the space $L^1(\mathbf{R^n})$ into $L^\infty(\mathbf{R^n})$ with

$$\sup_{\mathbf{x}\in\mathbf{R^n}} |\widehat{f}(\mathbf{x})| \leq (2\pi)^{-n} \|f\|_1. \tag{13.2}$$

The verification is immediate.
(2) The mapping $f \to \widehat{f}$ is linear on $L^1(\mathbf{R^n})$, that is, if $f_1, f_2 \in L^1(\mathbf{R^n})$ and $c_1, c_2 \in \mathbf{C}$ (the class of complex numbers), then

$$\widehat{\left(c_1 f_1 + c_2 f_2\right)}(\mathbf{x}) = c_1 \widehat{f_1}(\mathbf{x}) + c_2 \widehat{f_2}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n}. \tag{13.3}$$

We leave the simple proof to the reader.
(3) Next, let us show that $\widehat{f}(\mathbf{x})$ is a uniformly continuous function of $\mathbf{x}$ on $\mathbf{R^n}$ if $f \in L^1(\mathbf{R^n})$. For any $\mathbf{x}, \mathbf{h} \in \mathbf{R^n}$,

$$\begin{aligned} (2\pi)^n \left|\widehat{f}(\mathbf{x}+\mathbf{h}) - \widehat{f}(\mathbf{x})\right| &= \left| \int_{\mathbf{R^n}} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}} \left\{ e^{-i\mathbf{h}\cdot\mathbf{y}} - 1 \right\} d\mathbf{y} \right| \\ &\leq \int_{\mathbf{R^n}} |f(\mathbf{y})|\, \min\{|\mathbf{h}||\mathbf{y}|\,,\, 2\}\, d\mathbf{y} \\ &\leq \int_{|\mathbf{y}|<N} |f(\mathbf{y})|\, |\mathbf{h}||\mathbf{y}|\, d\mathbf{y} + \int_{|\mathbf{y}|\geq N} |f(\mathbf{y})|\, 2\, d\mathbf{y} \\ &= I + II, \end{aligned}$$

say, where $N > 0$ will be chosen momentarily. Let $\varepsilon > 0$. Note that II depends only on N and f, not on $\mathbf{x}$ or $\mathbf{h}$, and $II \to 0$ as $N \to \infty$ since $f \in L^1(\mathbf{R^n})$. Fix N so large that $II < \varepsilon/2$. For I, we have

$$I \leq \int_{|\mathbf{y}| \leq N} |f(\mathbf{y})|\, |\mathbf{h}| N\, d\mathbf{y} \leq \|f\|_1 N |\mathbf{h}| < \frac{\varepsilon}{2}$$

if $|\mathbf{h}| < \varepsilon/(2\|f\|_1 N)$. Here, we have assumed that $\|f\|_1 \neq 0$, but otherwise $\widehat{f} \equiv 0$ and the result is trivial. Hence, $I + II < \varepsilon$ uniformly in $\mathbf{x}$ and $\mathbf{h}$ provided $|\mathbf{h}|$ is small, which proves the result.

(4) There is a version for the Fourier transform of the Riemann–Lebesgue Theorem 12.21 for Fourier coefficients, namely,

Theorem 13.4 (Riemann–Lebesgue) *If* $f \in L^1(\mathbf{R^n})$, *then* $\lim_{|\mathbf{x}| \to \infty} \widehat{f}(\mathbf{x}) = 0$.

Proof. We will give two proofs. First, for any $\mathbf{x} \in \mathbf{R^n}$, $|\mathbf{x}| \neq 0$, we write

$$\begin{aligned}(2\pi)^n \widehat{f}(\mathbf{x}) &= \int_{\mathbf{R^n}} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y} \\ &= \int_{\mathbf{R^n}} f\Big(\mathbf{y} + \frac{\pi\mathbf{x}}{|\mathbf{x}|^2}\Big) e^{-i\mathbf{x}\cdot\mathbf{y}} e^{-i\pi\frac{|\mathbf{x}|^2}{|\mathbf{x}|^2}}\, d\mathbf{y} = -\int_{\mathbf{R^n}} f\Big(\mathbf{y} + \frac{\pi\mathbf{x}}{|\mathbf{x}|^2}\Big) e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y}.\end{aligned}$$

Adding the first and last formulas gives

$$2\,(2\pi)^n \widehat{f}(\mathbf{x}) = \int_{\mathbf{R^n}} \left[f(\mathbf{y}) - f\Big(\mathbf{y} + \frac{\pi\mathbf{x}}{|\mathbf{x}|^2}\Big) \right] e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y}$$

and

$$2\,(2\pi)^n |\widehat{f}(\mathbf{x})| \leq \int_{\mathbf{R^n}} \left| f(\mathbf{y}) - f\Big(\mathbf{y} + \frac{\pi\mathbf{x}}{|\mathbf{x}|^2}\Big) \right| d\mathbf{y}.$$

The last integral tends to 0 as $|\mathbf{x}| \to \infty$ by continuity in L^1, Theorem 8.19, and the first proof is complete.

We may also proceed by computing the Fourier transform of step functions and using a density argument. Let $I = \prod_{k=1}^n [a_k, b_k]$ be any interval in $\mathbf{R^n}$ with positive measure $|I|$. For any $\mathbf{x} = (x_1, \ldots, x_n)$, by Fubini's theorem,

$$(2\pi)^n \widehat{\chi_I}(\mathbf{x}) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} e^{-ix_1y_1} \cdots e^{-ix_ny_n} dy_1 \cdots dy_n$$
$$= \prod_{k=1}^{n} \int_{a_k}^{b_k} e^{-ix_kt} dt = \prod_{k=1}^{n} F_{[a_k,b_k]}(x_k),$$

where for any one-dimensional interval $[a, b]$ and any $s \in \mathbf{R^1}$, we denote

$$F_{[a,b]}(s) = \int_a^b e^{-ist} dt = \begin{cases} b - a \text{ if } s = 0 \\ \dfrac{e^{-isb} - e^{-isa}}{-is} \text{ if } s \neq 0. \end{cases}$$

The simple inequality $|e^{i\phi} - e^{i\psi}| \leq |\phi - \psi|$, $\phi, \psi \in \mathbf{R^1}$ shows that $\left|F_{[a,b]}(s)\right| \leq b - a$. Also, $\left|F_{[a,b]}(s)\right| \leq 2/|s|$ if $s \neq 0$. Since $|\mathbf{x}| \leq |x_1| + \cdots + |x_n|$, there exists $k_0 = 1, \ldots, n$ depending on $\mathbf{x}$ such that $|\mathbf{x}| \leq n|x_{k_0}|$. Combining estimates, if $|\mathbf{x}| \neq 0$, we obtain

$$(2\pi)^n \left|\widehat{\chi_I}(\mathbf{x})\right| \leq \prod_{k=1}^{n} \left|F_{[a_k,b_k]}(x_k)\right| \leq \frac{2}{|x_{k_0}|} \prod_{\substack{k=1 \\ k \neq k_0}}^{n} (b_k - a_k)$$
$$\leq \frac{2n}{|\mathbf{x}|} \frac{|I|}{\ell}, \quad \text{where } \ell = \min_k (b_k - a_k).$$

Hence, $\widehat{\chi_I}(\mathbf{x}) = O(|\mathbf{x}|^{-1})$ as $|\mathbf{x}| \to \infty$, and in particular, $\widehat{\chi_I}(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$. If $|I| = 0$, then $\widehat{\chi_I} \equiv 0$.

If f is a linear combination of characteristic functions of intervals in $\mathbf{R^n}$ (i.e., if f is a step function), it then follows from (13.3) that $\widehat{f}(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$. Finally, given any $f \in L^1(\mathbf{R^n})$ and $\varepsilon > 0$, choose a step function g such that $\|f - g\|_1 < \varepsilon$ (see the comment at the end of the proof of Lemma 7.3 for the existence of such a g). Then for any $\mathbf{x} \in \mathbf{R^n}$, by (13.2) and (13.3),

$$|\widehat{f}(\mathbf{x})| \leq |\widehat{f}(\mathbf{x}) - \widehat{g}(\mathbf{x})| + |\widehat{g}(\mathbf{x})| = |\widehat{(f-g)}(\mathbf{x})| + |\widehat{g}(\mathbf{x})|$$
$$< \frac{\varepsilon}{(2\pi)^n} + |\widehat{g}(\mathbf{x})|.$$

Since $\widehat{g}$ tends to 0 at infinity, so does $\widehat{f}$, and Theorem 13.4 is again proved.

The next three properties show how the Fourier transform interacts with translations, dilations, and rotations in $\mathbf{R^n}$. The proofs of the first two are left as exercises.

(5) (Translation) Let $f \in L^1(\mathbf{R^n})$ and $\mathbf{h} \in \mathbf{R^n}$. Define the *translation* $\tau_{\mathbf{h}} f$ *of* f *by* $\mathbf{h}$ as $(\tau_{\mathbf{h}} f)(\mathbf{x}) = f(\mathbf{x} + \mathbf{h})$. Then

$$\widehat{\tau_{\mathbf{h}} f}(\mathbf{x}) = e^{i\mathbf{x}\cdot\mathbf{h}} \widehat{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n}. \tag{13.5}$$

Also,

$$\left(\tau_{\mathbf{h}} \widehat{f}\right)(\mathbf{x}) = \widehat{f}(\mathbf{x} + \mathbf{h}) = \left(f(\mathbf{x}) e^{-i\mathbf{x}\cdot\mathbf{h}}\right)\widehat{\ }, \quad \mathbf{x} \in \mathbf{R^n},$$

that is, if $E_{\mathbf{h}} f$ is the function defined by $\left(E_{\mathbf{h}} f\right)(\mathbf{x}) = f(\mathbf{x}) e^{-i\mathbf{x}\cdot\mathbf{h}}$, then $\tau_{\mathbf{h}} \widehat{f} = \widehat{E_{\mathbf{h}} f}$.

(6) (Dilation) Let $f \in L^1(\mathbf{R^n})$ and $\lambda \in \mathbf{R^1} - \{0\}$. Define the *dilation* $\delta_\lambda f$ *of* f *by* λ to be the function $\left(\delta_\lambda f\right)(\mathbf{x}) = f(\lambda \mathbf{x})$. Then

$$\widehat{\delta_\lambda f}(\mathbf{x}) = \frac{1}{|\lambda|^n} \widehat{f}\left(\frac{\mathbf{x}}{\lambda}\right) \quad \left(= \frac{1}{|\lambda|^n} \left(\delta_{\frac{1}{\lambda}} \widehat{f}\right)(\mathbf{x})\right), \quad \mathbf{x} \in \mathbf{R^n}. \tag{13.6}$$

(7) (Rotation) Let $\mathcal{O}$ be an orthogonal linear transformation of $\mathbf{R^n}$ and set $(\mathcal{O} f)(\mathbf{x}) = f(\mathcal{O}\mathbf{x})$, $\mathbf{x} \in \mathbf{R^n}$. If $f \in L^1(\mathbf{R^n})$, then

$$\widehat{\mathcal{O} f}(\mathbf{x}) = (\mathcal{O} \widehat{f})(\mathbf{x}) \ \left(= \widehat{f}(\mathcal{O}\mathbf{x})\right), \quad \mathbf{x} \in \mathbf{R^n}. \tag{13.7}$$

In fact,

$$\begin{aligned}
\widehat{\mathcal{O} f}(\mathbf{x}) &= \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} f(\mathcal{O}\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y} \\
&= \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathcal{O}^{-1}\mathbf{y}} \frac{d\mathbf{y}}{|\det \mathcal{O}|} \\
&= \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} f(\mathbf{y}) e^{-i(\mathcal{O}\mathbf{x}\cdot\mathbf{y})} d\mathbf{y}
\end{aligned}$$

since $|\det \mathcal{O}| = 1$ and $\mathbf{x} \cdot \mathcal{O}^{-1}\mathbf{y} = \mathcal{O}\mathbf{x} \cdot \mathcal{O}\mathcal{O}^{-1}\mathbf{y} = \mathcal{O}\mathbf{x} \cdot \mathbf{y}$ by orthogonality. The last expression is $\widehat{f}(\mathcal{O}\mathbf{x})$ by definition, and (13.7) is proved.

See Exercise 5 for an analogue of (13.7) for general nonsingular linear transformations of $\mathbf{R^n}$.

By definition, a radial function of $\mathbf{x}$ is one that depends only on $|\mathbf{x}|$. Thus, f is a radial function on $\mathbf{R^n}$ if there is a function $g(t)$, $t \geq 0$, such that $f(\mathbf{x}) = g(|\mathbf{x}|)$ for all $\mathbf{x} \in \mathbf{R^n}$. Formula (13.7) says that rotations about the origin commute with the Fourier transform. As a consequence, we immediately obtain the next property.

(8) The Fourier transform of an integrable radial function is a radial function.

It is possible to explicitly compute the Fourier transforms of $e^{-|\mathbf{x}|^2}$ and $e^{-|\mathbf{x}|}$, and the formulas obtained have important applications. The computation of $\left(e^{-|\mathbf{x}|^2}\right)^{\widehat{}}$ is the simpler of the two, and the result will be used later in order to find $\left(e^{-|\mathbf{x}|}\right)^{\widehat{}}$.

(9) In shorthand notation, we have $\left(e^{-|\mathbf{x}|^2}\right)^{\widehat{}} = (2\sqrt{\pi})^{-n} e^{-|\mathbf{x}|^2/4}$, that is,

Theorem 13.8 *For all* $\mathbf{x} \in \mathbf{R^n}$,

$$\widehat{e^{-|\mathbf{x}|^2}} = \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} e^{-|\mathbf{y}|^2} e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y} = \frac{1}{(2\sqrt{\pi})^n} e^{-|\mathbf{x}|^2/4}.$$

Proof. By Fubini's theorem,

$$\begin{aligned}
&\frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} e^{-|\mathbf{y}|^2} e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y} \\
&= \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-y_1^2} \cdots e^{-y_n^2} e^{-ix_1y_1} \cdots e^{-ix_ny_n} dy_1 \cdots dy_n \\
&= \frac{1}{(2\pi)^n} \prod_{k=1}^{n} \int_{-\infty}^{\infty} e^{-y_k^2 - ix_ky_k} dy_k = \frac{1}{(2\pi)^n} \prod_{k=1}^{n} \int_{-\infty}^{\infty} e^{-(t^2+itx_k)} dt.
\end{aligned}$$

Writing $t^2 + itx_k = \left(t + i\frac{x_k}{2}\right)^2 + \frac{x_k^2}{4}$, we see that the last expression equals

$$\frac{1}{(2\pi)^n} \prod_{k=1}^{n} \int_{-\infty}^{\infty} e^{-(t+ix_k/2)^2} dt \cdot \prod_{k=1}^{n} e^{-x_k^2/4}.$$

Next, we use, without proof, the identities

$$\int_{-\infty}^{\infty} e^{-(t+is)^2} dt = \int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi} \quad \text{if } -\infty < s < \infty.$$

The first of these is a corollary of Cauchy's contour integration theorem; the second one is classical and has been observed in Exercise 11 of Chapter 6. Hence,

$$\widehat{e^{-|\mathbf{x}|^2}} = \frac{1}{(2\pi)^n} \cdot (\sqrt{\pi})^n \cdot e^{-|\mathbf{x}|^2/4} = \frac{1}{(2\sqrt{\pi})^n} e^{-|\mathbf{x}|^2/4},$$

and the proof is complete.

By rescaling $e^{-|\mathbf{x}|^2}$ and using the dilation property (13.6), we easily find a function that is equal to a multiple of its own Fourier transform:

$$\left(e^{-|\mathbf{x}|^2/2}\right)^{\widehat{}} = \frac{1}{(2\pi)^{n/2}} e^{-|\mathbf{x}|^2/2}, \quad \mathbf{x} \in \mathbf{R^n}. \tag{13.9}$$

In order to find an analogue in higher dimensions of the one-dimensional Gauss–Weierstrass kernel defined in (9.12), we first consider the kernel

$$K(\mathbf{x}) = \pi^{-n/2} e^{-|\mathbf{x}|^2}, \quad \mathbf{x} \in \mathbf{R^n},$$

and the corresponding approximation of the identity:

$$K_\varepsilon(\mathbf{x}) = \varepsilon^{-n} K(\mathbf{x}/\varepsilon) = \left(\sqrt{\pi}\varepsilon\right)^{-n} e^{-|\mathbf{x}|^2/\varepsilon^2}, \quad \varepsilon > 0.$$

Note that $\int_{\mathbf{R^n}} K(\mathbf{x})\, d\mathbf{x} = 1$, again by Exercise 11 of Chapter 6. Setting $\varepsilon = \sqrt{t}$, $t > 0$, we obtain the *n-dimensional Gauss–Weierstrass kernel* defined by

$$W(\mathbf{x}, t) = (\sqrt{\pi t})^{-n} e^{-|\mathbf{x}|^2/t}, \quad \mathbf{x} \in \mathbf{R^n},\ t > 0. \tag{13.10}$$

Then

$$\int_{\mathbf{R^n}} W(\mathbf{x}, t)\, d\mathbf{x} = 1, \quad t > 0.$$

For $\mathbf{x} \in \mathbf{R^n}$ and $t > 0$, the convolution $Wf(\mathbf{x}, t)$ defined by

$$Wf(\mathbf{x}, t) = [f * W(\cdot, t)](\mathbf{x}) = \int_{\mathbf{R^n}} f(\mathbf{x} - \mathbf{y}) W(\mathbf{y}, t)\, d\mathbf{y} \tag{13.11}$$

is called the *Gauss–Weierstrass integral of* f. The kernel $W(\mathbf{x}, t)$ satisfies the *heat equation*

$$\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\right) W(\mathbf{x}, t) = 4 \frac{\partial}{\partial t} W(\mathbf{x}, t) \tag{13.12}$$

in the upper half-space $\mathbf{R}_+^{\mathbf{n+1}} = \{(\mathbf{x}, t) : \mathbf{x} \in \mathbf{R^n}, t > 0\}$ (see Exercise 6(a)).

Due to the dilation property (13.6) and Theorem 13.8, the Gauss–Weierstrass kernel satisfies

$$W(\mathbf{x}, t) = \left(e^{-t|\mathbf{x}|^2/4}\right)^{\widehat{}}, \quad W(\mathbf{x}, t)^{\widehat{}} = (2\pi)^{-n} e^{-t|\mathbf{x}|^2/4} \tag{13.13}$$

if $(\mathbf{x}, t) \in \mathbf{R}^{\mathbf{n}+\mathbf{1}}_{+}$, where the Fourier transforms are taken in the $\mathbf{x}$ variable. We will use the first of these equations to derive an *inversion formula* for the Fourier transform, that is, a way to recover f from $\widehat{f}$. We need the following basic result in order to accomplish this.

(10) (Shifting hats) If $f, g \in L^1(\mathbf{R}^{\mathbf{n}})$, then

$$\int_{\mathbf{R}^{\mathbf{n}}} \widehat{f}(\mathbf{x})\, g(\mathbf{x})\, d\mathbf{x} = \int_{\mathbf{R}^{\mathbf{n}}} f(\mathbf{x})\, \widehat{g}(\mathbf{x})\, d\mathbf{x}. \tag{13.14}$$

Proof. Note that both integrals in (13.14) are finite; for example, the one on the left side is finite since $g \in L^1(\mathbf{R}^{\mathbf{n}})$ and $\widehat{f} \in L^\infty(\mathbf{R}^{\mathbf{n}})$ by (13.2). The formula itself is a corollary of Fubini's theorem since

$$\begin{aligned}
\int_{\mathbf{R}^{\mathbf{n}}} \widehat{f}(\mathbf{x})\, g(\mathbf{x})\, d\mathbf{x} &= \int_{\mathbf{R}^{\mathbf{n}}} \left(\frac{1}{(2\pi)^n} \int_{\mathbf{R}^{\mathbf{n}}} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y} \right) g(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbf{R}^{\mathbf{n}}} f(\mathbf{y}) \left(\frac{1}{(2\pi)^n} \int_{\mathbf{R}^{\mathbf{n}}} g(\mathbf{x}) e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{x} \right) d\mathbf{y} \\
&= \int_{\mathbf{R}^{\mathbf{n}}} f(\mathbf{y})\, \widehat{g}(\mathbf{y})\, d\mathbf{y},
\end{aligned}$$

where the change in the order of integration is justified because $f(\mathbf{y})g(\mathbf{x}) \in L^1(\mathbf{R}^{\mathbf{2n}}, d\mathbf{y}d\mathbf{x})$.

(11) We have the following inversion result (see also (13.28)).

Theorem 13.15 (Inversion of the Fourier transform on L^1) *If $f \in L^1(\mathbf{R}^{\mathbf{n}})$, then at every point $\mathbf{x}$ of the Lebesgue set of f,*

$$f(\mathbf{x}) = \lim_{t \to 0+} \int_{\mathbf{R}^{\mathbf{n}}} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} e^{-t|\mathbf{y}|^2}\, d\mathbf{y}. \tag{13.16}$$

If in addition $\widehat{f} \in L^1(\mathbf{R}^{\mathbf{n}})$, then at every Lebesgue point $\mathbf{x}$ of f,

$$f(\mathbf{x}) = \int_{\mathbf{R}^{\mathbf{n}}} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}. \tag{13.17}$$

In particular, (13.16) and (13.17) hold a.e. in $\mathbf{R}^{\mathbf{n}}$ and at every point of continuity of f.

Note that (13.17) can be rewritten as

$$f(\mathbf{x}) = (2\pi)^n \widehat{\widehat{f}}(-\mathbf{x}) = (2\pi)^n \left(\widehat{f}(-\mathbf{x})\right)^{\widehat{}}. \tag{13.18}$$

Before proving Theorem 13.15, we mention that the periodic analogue of (13.17) is that if the sum of the Fourier coefficients of a periodic $f \in L^1(-\pi, \pi)$ converges absolutely, then $S[f]$, the Fourier series of f, converges to f at every Lebesgue point of f. This is a corollary of the first part of Lebesgue's Theorem 12.51 (together with Theorem 12.38) since $S[f]$ converges everywhere if the sum of the Fourier coefficients of f converges absolutely.

Proof. To prove Theorem 13.15, let $f \in L^1(\mathbf{R^n})$ and write

$$Wf(\mathbf{x}, t) = \int_{\mathbf{R^n}} f(\mathbf{x} - \mathbf{y}) W(\mathbf{y}, t)\, d\mathbf{y} = \int_{\mathbf{R^n}} f(\mathbf{x} + \mathbf{y}) W(\mathbf{y}, t)\, d\mathbf{y},$$

where the last equality is true since $W(\mathbf{y}, t)$ is an even function of $\mathbf{y}$. By (13.13) and (13.14), the last integral equals

$$\begin{aligned}\int_{\mathbf{R^n}} (\tau_{\mathbf{x}} f)(\mathbf{y}) \left(e^{-t|\mathbf{y}|^2/4}\right)^{\widehat{}} d\mathbf{y} &= \int_{\mathbf{R^n}} \widehat{\tau_{\mathbf{x}} f}(\mathbf{y})\, e^{-t|\mathbf{y}|^2/4} d\mathbf{y} \\ &= \int_{\mathbf{R^n}} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} e^{-t|\mathbf{y}|^2/4} d\mathbf{y} \quad \text{by (13.5).}\end{aligned}$$

In summary, we have

$$Wf(\mathbf{x}, t) = \int_{\mathbf{R^n}} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} e^{-t|\mathbf{y}|^2/4} d\mathbf{y} \quad \text{for all } (\mathbf{x}, t) \in \mathbf{R}^{\mathbf{n}+\mathbf{1}}_+. \tag{13.19}$$

Now let $t \to 0+$. Due to the convolution representation of $Wf(\mathbf{x}, t)$ and Theorem 9.13, $Wf(\mathbf{x}, t)$ converges to $f(\mathbf{x})$ at every Lebesgue point $\mathbf{x}$ of f. The same is true if t is replaced by $4t$, and consequently, (13.16) follows from (13.19). Assuming in addition that $\widehat{f} \in L^1(\mathbf{R^n})$, we easily obtain from Lebesgue's dominated convergence theorem that, as $t \to 0$, the integral on the right in (13.19) tends to $\int_{\mathbf{R^n}} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}$. Formula (13.17) now follows at every Lebesgue point of f, and the proof is complete.

The integrals in (13.16), namely,

$$\int_{\mathbf{R^n}} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} e^{-t|\mathbf{y}|^2} d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n}, t > 0,$$

are referred to as the *Gauss–Weierstrass means* of the Fourier inversion integral $\int_{\mathbf{R^n}} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}$, although the latter integral may not exist in the Lebesgue sense if $f \in L^1(\mathbf{R^n})$. In general, the integrals

$$\int_{\mathbf{R^n}} g(\mathbf{y}) e^{-t|\mathbf{y}|^2} d\mathbf{y}$$

are called the Gauss–Weierstrass means of $\int_{\mathbf{R^n}} g$ and may exist and be finite for every $t > 0$ while $\int_{\mathbf{R^n}} g$ may not. The analogous expressions

$$\int_{\mathbf{R^n}} g(\mathbf{y}) e^{-t|\mathbf{y}|} d\mathbf{y}, \quad t > 0,$$

are called the *Abel means* of $\int_{\mathbf{R^n}} g$ (see (13.28)).

(12) An immediate corollary of (13.16) in Theorem 13.15 is

Corollary 13.20 (Uniqueness) *If $f, g \in L^1(\mathbf{R^n})$ and $\widehat{f}(\mathbf{x}) = \widehat{g}(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{R^n}$, then $f = g$ a.e. in $\mathbf{R^n}$. In particular, if $f \in L^1(\mathbf{R^n})$ and $\widehat{f} = 0$ everywhere in $\mathbf{R^n}$, then $f = 0$ a.e. in $\mathbf{R^n}$.*

(13) Another corollary of Theorem 13.15 is the following simple sufficient condition for integrability of $\widehat{f}$.

Corollary 13.21 *Suppose that $f \in L^1(\mathbf{R^n})$, f is continuous at $\mathbf{x} = \mathbf{0}$ and $\widehat{f} \geq 0$ everywhere. Then $\widehat{f} \in L^1(\mathbf{R^n})$ and $\|\widehat{f}\|_1 = f(\mathbf{0})$.*

Proof. Under the hypotheses, we may set $\mathbf{x} = \mathbf{0}$ in (13.16) to obtain

$$f(\mathbf{0}) = \lim_{t \to 0+} \int_{\mathbf{R^n}} \widehat{f}(\mathbf{y}) e^{-t|\mathbf{y}|^2} d\mathbf{y}.$$

In particular, the integral $\int_{\mathbf{R^n}} \widehat{f}(\mathbf{y}) e^{-t|\mathbf{y}|^2} d\mathbf{y}$ is bounded in t for $t > 0$. Its integrand is nonnegative since $\widehat{f} \geq 0$ by assumption. Letting $t \to 0$, Fatou's lemma then implies that $\widehat{f}$ is integrable. Finally, (13.17) with $\mathbf{x} = \mathbf{0}$ gives

$$f(\mathbf{0}) = \int_{\mathbf{R^n}} \widehat{f}(\mathbf{y})\, d\mathbf{y} = \|\widehat{f}\|_1,$$

completing the proof.

(14) Next, we compute the Fourier transform of $e^{-|\mathbf{x}|}$ in $\mathbf{R^n}$, $n \geq 1$. Let

$$\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} e^{-s} ds, \quad \alpha > 0,$$

denote the classical *Gamma function*.

Theorem 13.22 *For all* $\mathbf{x} \in \mathbf{R^n}$ *and* $\varepsilon > 0$,

$$\widehat{e^{-\varepsilon|\mathbf{x}|}} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}} \frac{\varepsilon}{\left(\varepsilon^2 + |\mathbf{x}|^2\right)^{\frac{n+1}{2}}}.$$

Proof. Because of the dilation property (13.6), it is enough to prove the result in case $\varepsilon = 1$. We consider first the one-dimensional case, $n = 1$, where the computation is simplest. Let $x \in (-\infty, \infty)$. Then

$$\begin{aligned}
2\pi\widehat{e^{-|x|}} &= \int_{-\infty}^{\infty} e^{-|t|} e^{-ixt}\, dt \\
&= \int_0^\infty e^{-t-ixt}\, dt + \int_{-\infty}^0 e^{t-ixt}\, dt \\
&= \int_0^\infty e^{-(1+ix)t}\, dt + \int_0^\infty e^{-(1-ix)t}\, dt \\
&= \left[-\frac{e^{-(1+ix)t}}{1+ix} - \frac{e^{-(1-ix)t}}{1-ix} \right]_{t=0}^{\infty} \\
&= \frac{1}{1+ix} + \frac{1}{1-ix} = \frac{2}{1+x^2}.
\end{aligned}$$

Hence,

$$\widehat{e^{-|x|}} = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in (-\infty, \infty), \tag{13.23}$$

and we are done since when $n = 1$, then $\Gamma((n+1)/2) = \Gamma(1) = 1$.

Combining (13.23) with the inversion formula (13.17) in case $n=1$, we obtain

$$e^{-|x|} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+t^2} e^{ixt}\, dt = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+t^2} e^{-ixt}\, dt, \quad x \in (-\infty, \infty).$$

Next, with the purpose of introducing an exponential factor e^{-t^2} in the computation, we write

$$\frac{1}{1+t^2} = \int_0^{\infty} e^{-(1+t^2)s}\, ds.$$

Consequently,

$$e^{-|x|} = \frac{1}{\pi}\int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\left(1+t^2\right)s} ds\right) e^{-ixt} dt, \quad x \in (-\infty, \infty). \tag{13.24}$$

Now consider the higher dimensional case:

$$(2\pi)^n \widehat{e^{-|\mathbf{x}|}} = \int_{\mathbf{R^n}} e^{-|\mathbf{y}|} e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n}.$$

On the right side, express the factor $e^{-|\mathbf{y}|}$ by using (13.24) with x there chosen to be $|\mathbf{y}|$, obtaining

$$\begin{aligned} e^{-|\mathbf{y}|} &= \frac{1}{\pi}\int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\left(1+t^2\right)s} ds\right) e^{-i|\mathbf{y}|t}\, dt \\ &= \frac{1}{\pi}\int_0^{\infty} e^{-s} \left(\int_{-\infty}^{\infty} e^{-t^2 s} e^{-i|\mathbf{y}|t} dt\right) ds \\ &= \frac{1}{\pi}\int_0^{\infty} e^{-s} \sqrt{\pi}\, e^{-|\mathbf{y}|^2/(4s)} \frac{1}{\sqrt{s}}\, ds, \end{aligned}$$

where the last equality follows by dilation from the one-dimensional version of Theorem 13.8. By substitution, we obtain that for any $\mathbf{x} \in \mathbf{R^n}$,

$$
\begin{aligned}
(2\pi)^n \widehat{e^{-|\mathbf{x}|}} &= \int_{\mathbf{R^n}} \left(\frac{1}{\pi} \int_0^\infty e^{-s} \sqrt{\pi}\, e^{-|\mathbf{y}|^2/(4s)} \frac{ds}{\sqrt{s}} \right) e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y} \\
&= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-s}}{\sqrt{s}} \left(\int_{\mathbf{R^n}} e^{-|\mathbf{y}|^2/(4s)} e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y} \right) ds \\
&= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-s}}{\sqrt{s}} \left(\sqrt{\pi}\right)^n \left(2\sqrt{s}\right)^n e^{-s|\mathbf{x}|^2}\, ds,
\end{aligned}
$$

by another application of Theorem 13.8, this time in the n-dimensional form. Regrouping terms gives

$$
\begin{aligned}
(2\pi)^n \widehat{e^{-|\mathbf{x}|}} &= 2^n \left(\sqrt{\pi}\right)^{n-1} \int_0^\infty e^{-\left(1+|\mathbf{x}|^2\right)s} s^{\frac{n-1}{2}}\, ds \\
&= \frac{2^n \left(\sqrt{\pi}\right)^{n-1}}{(1+|\mathbf{x}|^2)^{\frac{n+1}{2}}} \int_0^\infty e^{-s} s^{\frac{n-1}{2}}\, ds.
\end{aligned}
$$

The last integral equals $\Gamma\left(\frac{n+1}{2}\right)$, and therefore

$$
\widehat{e^{-|\mathbf{x}|}} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}} \frac{1}{(1+|\mathbf{x}|^2)^{\frac{n+1}{2}}}, \quad \mathbf{x} \in \mathbf{R^n},
$$

as desired. This completes the proof of Theorem 13.22.

(15) (The Poisson Integral in $\mathbf{R^{n+1}_+}$) For $(\mathbf{x}, \varepsilon) \in \mathbf{R^{n+1}_+} = \{(\mathbf{x}, \varepsilon) : \mathbf{x} \in \mathbf{R^n}, \varepsilon > 0\}$, we denote

$$
P(\mathbf{x}, \varepsilon) = \widehat{e^{-\varepsilon|\mathbf{x}|}} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}} \frac{\varepsilon}{\left(\varepsilon^2 + |\mathbf{x}|^2\right)^{\frac{n+1}{2}}}, \tag{13.25}
$$

and call $P(\mathbf{x}, \varepsilon)$ the *Poisson kernel for the half-space* $\mathbf{R^{n+1}_+}$. The case $n = 1$ is considered in (9.10), and the periodic version is defined in Section 12.7 on p. 348.

As a function of $\mathbf{x}$, $P(\mathbf{x}, \varepsilon)$ is clearly positive and of class $L^1(\mathbf{R^n}) \cap L^\infty(\mathbf{R^n})$ for each $\varepsilon > 0$. Also, as a function of $(\mathbf{x}, \varepsilon)$, it is infinitely differentiable in $\mathbf{R^{n+1}_+}$. By direct computation, it satisfies *Laplace's equation*

$$
\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2} + \frac{\partial^2}{\partial \varepsilon^2} \right) P(\mathbf{x}, \varepsilon) = 0 \quad \text{in } \mathbf{R^{n+1}_+}. \tag{13.26}
$$

Moreover, $P(\mathbf{x}, \varepsilon)$ is an approximation of the identity since if $P(\mathbf{x})$ is defined by $P(\mathbf{x}) = P(\mathbf{x}, 1)$, then $P(\mathbf{x}, \varepsilon) = \varepsilon^{-n} P(\mathbf{x}/\varepsilon)$, and by (13.17),

$$\int_{\mathbf{R}^n} P(\mathbf{x}, 1)\, d\mathbf{x} = \int_{\mathbf{R}^n} \widehat{e^{-|\mathbf{x}|}}\, d\mathbf{x} = e^{-|\mathbf{x}|}\Big|_{\mathbf{x}=\mathbf{0}} = 1.$$

Hence, since $P(\mathbf{x}, 1) = O(1/|\mathbf{x}|^{n+1})$ as $|\mathbf{x}| \to \infty$, Theorem 9.13 implies that the *Poisson integral* of f, defined to be the convolution

$$Pf(\mathbf{x}, \varepsilon) = \int_{\mathbf{R}^n} f(\mathbf{x} - \mathbf{y}) P(\mathbf{y}, \varepsilon)\, d\mathbf{y}, \tag{13.27}$$

converges to $f(\mathbf{x})$ as $\varepsilon \to 0$ at each point $\mathbf{x}$ of the Lebesgue set of f if $f \in L^1(\mathbf{R}^n)$. If $f \in L^p(\mathbf{R}^n)$, $p > 1$, the same is true (cf. Exercise 12 of Chapter 9). Note that

$$\begin{aligned} Pf(\mathbf{x}, \varepsilon) &= \int_{\mathbf{R}^n} f(\mathbf{x} - \mathbf{y}) \widehat{e^{-\varepsilon|\mathbf{y}|}}\, d\mathbf{y} = \int_{\mathbf{R}^n} f(\mathbf{x} + \mathbf{y}) \widehat{e^{-\varepsilon|\mathbf{y}|}}\, d\mathbf{y} \\ &= \int_{\mathbf{R}^n} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} e^{-\varepsilon|\mathbf{y}|}\, d\mathbf{y} \quad \text{if } f \in L^1(\mathbf{R}^n). \end{aligned}$$

Hence, if $f \in L^1(\mathbf{R}^n)$ and $\mathbf{x}$ is a Lebesgue point of f, then

$$f(\mathbf{x}) = \lim_{\varepsilon \to 0+} \int_{\mathbf{R}^n} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} e^{-\varepsilon|\mathbf{y}|} d\mathbf{y}. \tag{13.28}$$

Recall that the integrals

$$\int_{\mathbf{R}^n} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} e^{-\varepsilon|\mathbf{y}|} d\mathbf{y}, \quad \varepsilon > 0,$$

are called the Abel means of the (formal) integral $\int_{\mathbf{R}^n} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}$.

We also have by inversion that

$$\widehat{P(\mathbf{x}, \varepsilon)} = e^{-\varepsilon|\mathbf{x}|}, \quad \mathbf{x} \in \mathbf{R}^n,\ \varepsilon > 0, \tag{13.29}$$

where the Fourier transform is taken in the $\mathbf{x}$ variable.

(16) (The Convolution Property) By Young's theorem, the convolution of any two integrable functions is also integrable and therefore has a well-defined Fourier transform.

Theorem 13.30 *If $f, g \in L^1(\mathbf{R^n})$, then*

$$\widehat{f * g}(\mathbf{x}) = (2\pi)^n \widehat{f}(\mathbf{x})\,\widehat{g}(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbf{R^n}.$$

This follows easily from Fubini's theorem; we omit the details. In case $n = 1$, the result is listed in Exercise 6 Chapter 6. See also Theorem 13.59 for an important related result.

(17) (Differentiation Properties) We continue our list of properties of the Fourier transform with two important differentiation formulas.

Theorem 13.31

(a) *Let x_k, $k = 1, \ldots, n$, denote the kth coordinate of $\mathbf{x}$, $\mathbf{x} \in \mathbf{R^n}$. If f and $x_k f(\mathbf{x})$ belong to $L^1(\mathbf{R^n})$, then $\widehat{f}$ has a partial derivative with respect to x_k everywhere in $\mathbf{R^n}$, and*

$$\frac{\partial \widehat{f}}{\partial x_k}(\mathbf{x}) = \big(-ix_k f(\mathbf{x})\big)\widehat{\ } \quad \left[= \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} \big(-iy_k f(\mathbf{y})\big) e^{-i\mathbf{x}\cdot\mathbf{y}}\, d\mathbf{y}\right]. \tag{13.32}$$

(b) *Fix $k = 1, \ldots, n$ and let $\mathbf{h} = (0, \ldots, 0, h_k, 0, \ldots, 0) \neq \mathbf{0}$ lie on the kth coordinate axis. Suppose that $f \in L^1(\mathbf{R^n})$ and that there is a function g such that*

$$\lim_{h_k \to 0} \int_{\mathbf{R^n}} \left| \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})}{h_k} - g(\mathbf{x}) \right| d\mathbf{x} = 0. \tag{13.33}$$

Then $g \in L^1(\mathbf{R^n})$ and

$$\widehat{g}(\mathbf{x}) = ix_k \widehat{f}(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbf{R^n}. \tag{13.34}$$

A few remarks about condition (13.33) are listed after the proof of the theorem.

Proof. (a) Denoting nonzero points on the kth coordinate axis by $\mathbf{h} = (0, \ldots, 0, h_k, 0, \ldots, 0)$, we have

$$\frac{\widehat{f}(\mathbf{x} + \mathbf{h}) - \widehat{f}(\mathbf{x})}{h_k} = \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} f(\mathbf{y}) \left\{ \frac{e^{-ih_k y_k} - 1}{h_k} \right\} e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}.$$

Since the expression in curly brackets converges to $-iy_k$ as $h_k \to 0$ and is bounded in absolute value by $|y_k|$, formula (13.32) follows from the Lebesgue dominated convergence theorem and the hypothesis that $f(\mathbf{y})y_k$ is integrable.

(b) If g satisfies (13.33), then g is clearly integrable since f is. With $\mathbf{h} = (0, \ldots, 0, h_k, 0, \ldots, 0)$ as usual, (13.33) together with (13.2) implies the pointwise equality

$$\lim_{h_k \to 0} \left(\frac{f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})}{h_k} \right)^{\widehat{}} = \widehat{g}(\mathbf{x}), \text{ or}$$

$$\lim_{h_k \to 0} \frac{1}{h_k} \left(\tau_{\mathbf{h}} f - f\right)^{\widehat{}}(\mathbf{x}) = \widehat{g}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n}.$$

Equivalently, by (13.3) and (13.5), for all $\mathbf{x}$,

$$\lim_{h_k \to 0} \widehat{f}(\mathbf{x}) \left\{ \frac{e^{ix_k h_k} - 1}{h_k} \right\} = \widehat{g}(\mathbf{x}),$$

and (13.34) follows.

We now make some remarks related to (13.33). Given f, a function g that satisfies (13.33) is clearly unique up to a set of Lebesgue measure 0. If such a g exists, it is called the *partial derivative of f with respect to x_k in the L^1 sense*, the idea being that the difference quotient of f in the kth coordinate converges in the L^1 norm.

A simple case when (13.33) holds with g equal to the ordinary partial derivative $\partial f/\partial x_k$ of f is when $f \in C_0^1(\mathbf{R^n})$, since then both f and $\partial f/\partial x_k$ have compact support and the difference quotient of f in the variable x_k converges uniformly to $\partial f/\partial x_k$. Hence, by (13.34), for every $k = 1, \ldots, n$, we have

$$\left(\frac{\partial f}{\partial x_k} \right)^{\widehat{}}(\mathbf{x}) = ix_k \widehat{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n}, \quad \text{if } f \in C_0^1(\mathbf{R^n}). \tag{13.35}$$

As we will see in Theorem 13.41, the restriction in (13.35) that $f \in C_0^1(\mathbf{R^n})$ can be considerably weakened.

By iteration, (13.35) leads easily to analogous formulas for higher-order derivatives. For example, if $f \in C_0^2(\mathbf{R^n})$ and $j, k = 1, \ldots, n$, then

$$\left(\frac{\partial^2 f}{\partial x_j \partial x_k} \right)^{\widehat{}}(\mathbf{x}) = ix_j \left(\frac{\partial f}{\partial x_k} \right)^{\widehat{}}(\mathbf{x}) = (ix_j)\,(ix_k)\widehat{f}(\mathbf{x}).$$

More generally, if $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index of nonnegative integers and D^α is the differential operator defined by

$$D^\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}},$$

then for any $N = 1, 2, \ldots$ and any α with $\alpha_1 + \cdots + \alpha_n \leq N$,

$$\widehat{D^\alpha f}(\mathbf{x}) = (ix_1)^{\alpha_1} \cdots (ix_n)^{\alpha_n} \widehat{f}(\mathbf{x}) \quad \text{if } f \in C_0^N(\mathbf{R^n}). \tag{13.36}$$

A higher-order analogue of part (a) of Theorem 13.31 is that if $N = 1, 2, \ldots$ and f has the property that $p(\mathbf{x})f(\mathbf{x}) \in L^1(\mathbf{R^n})$ for every polynomial $p(\mathbf{x})$ of degree N (or less), then $\widehat{f} \in C^N(\mathbf{R^n})$ and

$$\left(D^\alpha \widehat{f}\right)(\mathbf{x}) = \left[(-ix_1)^{\alpha_1} \cdots (-ix_n)^{\alpha_n} f(\mathbf{x})\right]^\wedge \text{ if } \alpha_1 + \cdots + \alpha_n \leq N. \tag{13.37}$$

Verification is left to the reader.

In particular, (13.36) and (13.37) hold for derivatives of any order if $f \in C_0^\infty(\mathbf{R^n})$.

Note that if $f \in C_0^N(\mathbf{R^n})$ for some N, then for every multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_1 + \cdots + \alpha_n \leq N$, the fact that $D^\alpha f \in L^1(\mathbf{R^n})$ implies that $\widehat{D^\alpha f}$ is a bounded function on $\mathbf{R^n}$. Hence, by (13.36), the function $(ix_1)^{\alpha_1} \cdots (ix_n)^{\alpha_n} \widehat{f}(\mathbf{x})$ is bounded in $\mathbf{x}$ if $\alpha_1 + \cdots + \alpha_n \leq N$. Consequently,

$$|\widehat{f}(\mathbf{x})| \leq C \frac{1}{(1 + |\mathbf{x}|)^N} \quad \text{if } f \in C_0^N(\mathbf{R^n}).$$

Also, since $\widehat{D^\alpha f}$ tends to 0 at infinity, $\hat{f}(\mathbf{x}) = o(|\mathbf{x}|^{-n})$ as $|\mathbf{x}| \to \infty$. Roughly speaking, this means that the smoother a function with compact support is, the faster its Fourier transform tends to 0 at infinity. In particular, by choosing $N = n + 1$, we obtain a simple sufficient condition for $\widehat{f}$ to be integrable, namely,

Corollary 13.38 *If $f \in C_0^{n+1}(\mathbf{R^n})$, then $\widehat{f} \in L^1(\mathbf{R^n})$. In particular, the inversion formula (13.17) holds everywhere in $\mathbf{R^n}$ for such f.*

The class $\mathscr{S} = \mathscr{S}(\mathbf{R^n})$ of *Schwartz functions*, or *rapidly decreasing functions*, on $\mathbf{R^n}$ is defined to be the collection of all $f \in C^\infty(\mathbf{R^n})$ such that $p(\mathbf{x})D^\alpha f(\mathbf{x})$ is bounded on $\mathbf{R^n}$ for every polynomial $p(\mathbf{x})$ and every multi-index α. The bound may vary with α and the polynomial p. Thus, if $\mathbf{x}^\beta$ denotes the

monomial $\mathbf{x}^\beta = x_1^{\beta_1} \cdots x_n^{\beta_n}$, where $\beta = (\beta_1, \ldots, \beta_n)$ and each β_k is a nonnegative integer, then

$$\mathscr{S} = \left\{ f \in C^\infty(\mathbf{R^n}) : \sup_{\mathbf{x} \in \mathbf{R^n}} \left| \mathbf{x}^\beta D^\alpha f(\mathbf{x}) \right| < \infty \text{ for all } \alpha, \beta \right\}. \tag{13.39}$$

Clearly, $C_0^\infty(\mathbf{R^n}) \subset \mathscr{S}$, but the containment is proper since $e^{-|\mathbf{x}|^2}$ belongs to $\mathscr{S}$ but does not have compact support. The function $e^{-|\mathbf{x}|}$ is not differentiable at the origin, and so it does not belong to $\mathscr{S}$. The functions 1, $|\mathbf{x}|^2$, and $e^{-1/|\mathbf{x}|^2}$ (defined to be 0 at the origin) are simple examples of infinitely differentiable functions that do not belong to $\mathscr{S}$. Note that $\mathscr{S} \subset L^p(\mathbf{R^n})$ for every p, $0 < p \le \infty$. Also, if $f \in \mathscr{S}$, then $p(\mathbf{x}) D^\alpha f(\mathbf{x}) \in \mathscr{S}$ for every polynomial $p(\mathbf{x})$ in $\mathbf{R^n}$ and every multi-index α.

We leave the proof of the next result to the reader (see Exercise 8).

Theorem 13.40 *If $f \in \mathscr{S}$, then $\widehat{f} \in \mathscr{S}$. Also, if $f \in \mathscr{S}$, then the formulas in (13.36) and (13.37) hold for all α, that is,*

$$\widehat{D^\alpha f}(\mathbf{x}) = (i\mathbf{x})^\alpha \widehat{f}(\mathbf{x}) \text{ and } \left(D^\alpha \widehat{f}\right)(\mathbf{x}) = \left[(-i\mathbf{x})^\alpha f(\mathbf{x})\right]^\wedge$$

for all $\mathbf{x} \in \mathbf{R^n}$ and all α.

In particular, if $f \in C_0^\infty(\mathbf{R^n})$, then $\widehat{f} \in \mathscr{S}(\mathbf{R^n})$. On the other hand, the only function f in $C_0^\infty(\mathbf{R^n})$ such that $\widehat{f}$ has compact support is $f \equiv 0$ (see Exercise 28).

Theorem 13.40 shows that the formula $(\partial f/\partial x_k)^\wedge(\mathbf{x}) = i x_k \widehat{f}(\mathbf{x})$ is valid if $f \in \mathscr{S}$, and we observed in (13.35) that it is also true if $f \in C_0^1(\mathbf{R^n})$. Let us now show that it holds for a much larger class of functions. Recall that (by Theorem 7.29) an absolutely continuous function $F(t)$, $t \in [a, b] \subset \mathbf{R^1}$, has a first derivative F' a.e. in $[a, b]$ and $F(b) - F(a) = \int_a^b F'(t)\, dt$.

Theorem 13.41 *Let $k = 1, \ldots, n$ and $f \in L^1(\mathbf{R^n})$. Suppose that f is locally absolutely continuous on every line parallel to the kth coordinate axis, that is, suppose that $f(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n)$, when considered as a function of x_k, is absolutely continuous on every interval $-\infty < a \le x_k \le b < \infty$ for all $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n$. If $\partial f/\partial x_k \in L^1(\mathbf{R^n})$, then*

$$\left(\frac{\partial f}{\partial x_k}\right)^\wedge(\mathbf{x}) = i x_k \widehat{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n}.$$

In particular, the formula holds for every $k = 1, \ldots, n$ and every f such that $f \in L^1(\mathbf{R^n}) \cap Lip_{\text{loc}}(\mathbf{R^n})$ and $|\nabla f| \in L^1(\mathbf{R^n})$.

Proof. The second statement of the theorem (concerning the assumption that f is locally Lipschitz continuous) follows from the first statement since if f is locally Lipschitz continuous, then it is locally absolutely continuous on every line.

To prove the first statement, let f satisfy its hypothesis. By Theorem 13.31(b), it is enough to verify (13.33) with g equal to $\partial f/\partial x_k$, that is, to show that $\partial f/\partial x_k$ is the partial derivative of f with respect to x_k in the L^1 sense. We will show this only when $n \geq 2$ and $k = 1$, leaving the remaining cases for the reader to check. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, denote

$$\mathbf{x} = (x_1, \hat{\mathbf{x}}) \text{ with } \hat{\mathbf{x}} = (x_2, \ldots, x_n) \in \mathbf{R^{n-1}}.$$

By hypothesis, for every $\hat{\mathbf{x}}, f(x_1, \hat{\mathbf{x}})$ is an absolutely continuous function of x_1 on every compact interval in $\mathbf{R^1}$. Hence, by Theorem 7.29, for every $\mathbf{x} \in \mathbf{R^n}$ and every $\mathbf{h} = (h_1, 0, \ldots, 0) \in \mathbf{R^n}$, we have

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = f\left(x_1 + h_1, \hat{\mathbf{x}}\right) - f(x_1, \hat{\mathbf{x}}) = \int_0^{h_1} \frac{\partial f}{\partial x_1}\left(x_1 + t, \hat{\mathbf{x}}\right) dt.$$

Thus, if $h_1 \neq 0$,

$$\frac{f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})}{h_1} - \frac{\partial f}{\partial x_1}(\mathbf{x}) = \frac{1}{h_1}\int_0^{h_1}\left[\frac{\partial f}{\partial x_1}\left(x_1 + t, \hat{\mathbf{x}}\right) - \frac{\partial f}{\partial x_1}\left(x_1, \hat{\mathbf{x}}\right)\right] dt.$$

Taking the L^1 norm in $\mathbf{x}$, and denoting $\mathbf{t} = (t, 0, \ldots, 0)$, we obtain

$$\int_{\mathbf{R^n}} \left|\frac{f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})}{h_1} - \frac{\partial f}{\partial x_1}(\mathbf{x})\right| d\mathbf{x} \leq \frac{1}{h_1}\int_0^{h_1}\left\|\frac{\partial f}{\partial x_1}(\mathbf{x}+\mathbf{t}) - \frac{\partial f}{\partial x_1}(\mathbf{x})\right\|_1 dt.$$

Now let $h_1 \to 0$. By continuity in L^1 (Theorem 8.19), the integrand in the last integral tends to 0 as $t \to 0$, that is,

$$\left\|\frac{\partial f}{\partial x_1}(\mathbf{x}+\mathbf{t}) - \frac{\partial f}{\partial x_1}(\mathbf{x})\right\|_1 \to 0 \text{ as } t \to 0,$$

and the result follows.

(18) (The Principal Value Fourier Transform of x^{-1}) We close this section with the computation in one dimension of a different notion of the Fourier transform, which we call the *principal value Fourier transform*, of the function $1/x$. Since $1/x$ fails to be integrable in $\mathbf{R^1}$, both near the origin and near infinity,

its Fourier transform cannot be defined by using the one-dimensional version of (13.1). Furthermore, since $1/x$ does not belong to L^2 near the origin, the method developed in the next section for extending the definition of the Fourier transform to L^2 functions will not apply to $1/x$.

However, as we will see, the symmetry and oddness of $1/x$ can be exploited in order to define its Fourier transform as a limit, denoted p.v. $(x^{-1})\widehat{\ }$ and defined by

$$\text{p.v.}\left(\frac{1}{x}\right)\widehat{\ } = \lim_{\substack{\varepsilon\to 0+\\ \omega\to\infty}} \frac{1}{2\pi} \int_{\varepsilon<|y|<\omega} \frac{1}{y} e^{-ixy}\, dy, \quad x \in (-\infty, \infty). \tag{13.42}$$

We call this expression the *principal value Fourier transform of* x^{-1}. Its existence and evaluation are treated in the next theorem and will be used in Section 13.3 in order to define the Hilbert transform of a general L^2 function.

We may rewrite (13.42) in the equivalent form

$$\text{p.v.}\left(\frac{1}{x}\right)\widehat{\ } = \lim_{\substack{\varepsilon\to 0+\\ \omega\to\infty}} \widehat{K_{\varepsilon,\omega}}(x), \quad x \in \mathbf{R^1}, \tag{13.43}$$

where $K_{\varepsilon,\omega}$ is the doubly truncated function

$$K_{\varepsilon,\omega}(x) = \frac{\chi_{\{\varepsilon<|x|<\omega\}}(x)}{x}, \quad 0 < \varepsilon < \omega < \infty. \tag{13.44}$$

Note that each $K_{\varepsilon,\omega} \in L^1(\mathbf{R^1})$, and therefore $\widehat{K_{\varepsilon,\omega}}$ is well-defined and satisfies $\|\widehat{K_{\varepsilon,\omega}}\|_\infty \le \|K_{\varepsilon,\omega}\|_1/(2\pi)$. However, $\|K_{\varepsilon,\omega}\|_1$ is unbounded in ε, ω; in fact,

$$\|K_{\varepsilon,\omega}\|_1 = \int_{\varepsilon<|x|<\omega} \frac{dx}{|x|} = 2\log\frac{\omega}{\varepsilon}.$$

Theorem 13.45 *The principal value Fourier transform* $\widehat{x^{-1}}$ *defined in (13.42) exists and is finite everywhere in* $\mathbf{R^1}$, *and*

$$\text{p.v.}\left(\frac{1}{x}\right)\widehat{\ } = -\frac{i}{2}\operatorname{sign} x, \quad -\infty < x < \infty. \tag{13.46}$$

Here, the right side of (13.46) should be interpreted as 0 *when* $x = 0$, *that is,* p.v. $\widehat{x^{-1}}$ *equals* 0 *when* $x = 0$.

Furthermore, $\widehat{K_{\varepsilon,\omega}}$ is bounded uniformly in x, ε, and ω:

$$\sup_{\substack{\varepsilon,\omega \\ 0<\varepsilon<\omega<\infty}} \left\|\widehat{K_{\varepsilon,\omega}}\right\|_\infty \le \frac{4}{\pi}. \tag{13.47}$$

Proof. The proof of (13.46) is simple and based on the classical formula

$$\lim_{\substack{\varepsilon\to 0+ \\ \omega\to\infty}} \int_\varepsilon^\omega \frac{\sin t}{t}\,dt = \frac{\pi}{2},^* \tag{13.48}$$

which can be verified by contour integration. We will take (13.48) for granted. If $x \in \mathbf{R}^1$ and $0 < \varepsilon < \omega < \infty$, then since $1/y$ is an odd function of y,

$$\frac{1}{2\pi}\int_{\varepsilon<|y|<\omega} \frac{1}{y}\,e^{-ixy}dy = \frac{1}{2\pi}\int_{\varepsilon<|y|<\omega} \frac{-i\sin xy}{y}\,dy = \frac{-i}{\pi}\int_\varepsilon^\omega \frac{\sin xt}{t}\,dt.$$

The last expression is 0 if $x = 0$, while if $x \neq 0$, it is

$$\frac{-i}{\pi}\,(\operatorname{sign} x)\int_\varepsilon^\omega \frac{\sin|x|t}{t}\,dt = \frac{-i}{\pi}\,(\operatorname{sign} x)\int_{\varepsilon|x|}^{\omega|x|} \frac{\sin t}{t}\,dt.$$

Then (13.46) follows immediately from (13.48) by letting $\varepsilon \to 0$, $\omega \to \infty$.

To prove (13.47), note as earlier that if $0 < \varepsilon < \omega < \infty$ and $x \neq 0$, then

$$\widehat{K_{\varepsilon,\omega}}(x) = \frac{1}{2\pi}\int_{\varepsilon<|y|<\omega} \frac{e^{-ixy}}{y}\,dy = -\frac{i}{\pi}(\operatorname{sign} x)\int_{\varepsilon|x|}^{\omega|x|} \frac{\sin t}{t}\,dt.$$

Hence, it suffices to show that

$$\sup_{0<\lambda<\Lambda<\infty} \left|\int_\lambda^\Lambda \frac{\sin t}{t}\,dt\right| \le 4.$$

We will verify this when $0 < \lambda \le 1 \le \Lambda < \infty$, leaving the cases when λ and Λ do not straddle 1 to the reader. If $0 < \lambda \le 1 \le \Lambda < \infty$, integration by parts gives

* The improper integral $\int_0^\infty \frac{\sin t}{t}\,dt$ converges absolutely at $t = 0$ and so is improper only at ∞.

$$\int_{\lambda}^{\Lambda} \frac{\sin t}{t}\,dt = \int_{\lambda}^{1} \frac{\sin t}{t}\,dt + \int_{1}^{\Lambda} \frac{\sin t}{t}\,dt = \int_{\lambda}^{1} \frac{\sin t}{t}\,dt - \frac{\cos t}{t}\Big|_{1}^{\Lambda} + \int_{1}^{\Lambda} \frac{\cos t}{t^2}\,dt.$$

Therefore,

$$\left|\int_{\lambda}^{\Lambda} \frac{\sin t}{t}\,dt\right| \le \int_{0}^{1} 1\,dt + 2 + \int_{1}^{\infty} \frac{1}{t^2}\,dt = 4,$$

as desired.

13.2 The Fourier Transform on L^2

We will now define the Fourier transform of a general $f \in L^2(\mathbf{R^n})$ and study its main properties. The same notation, namely $\widehat{f}$, will be used for the Fourier transform no matter whether f belongs to $L^1(\mathbf{R^n})$ or to $L^2(\mathbf{R^n})$ since the two definitions will turn out to agree a.e. when $f \in L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$. Note that $L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ is a dense subset of both $L^1(\mathbf{R^n})$ and $L^2(\mathbf{R^n})$.

We will see that a striking and fundamental difference between the maps $f \to \widehat{f}$ when $f \in L^1(\mathbf{R^n})$ and when $f \in L^2(\mathbf{R^n})$ is that the mapping for $L^2(\mathbf{R^n})$ turns out to be essentially an isometry of $L^2(\mathbf{R^n})$ onto itself. Proving this is the main goal of the section.

Since the functions f now under consideration are generally complex-valued, it will be useful to recall that

$$|f|^2 = f\bar{f}, \quad \int_E |f|^2 = \int_E f\bar{f}, \quad \overline{\int_E f} = \int_E \bar{f}, \quad \text{etc.},$$

where $\bar{z}$ denotes the complex conjugate of z, $z \in \mathbf{C}$. Also, if $f \in L^1(\mathbf{R^n})$, then $\overline{\widehat{f}(\mathbf{x})} = \widehat{\bar{f}}(-\mathbf{x})$, $\mathbf{x} \in \mathbf{R^n}$, since

$$\overline{\frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}}\,d\mathbf{y}} = \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} \overline{f(\mathbf{y})} e^{i\mathbf{x}\cdot\mathbf{y}}\,d\mathbf{y}.$$

The next lemma gives a useful fact about the convolution $f * g$ of two L^2 functions. See Exercise 13 for an analogue when $f \in L^p(\mathbf{R^n})$ and $g \in L^{p'}(\mathbf{R^n})$, $1/p + 1/p' = 1$, $1 \le p \le \infty$.

Lemma 13.49 *Let $f, g \in L^2(\mathbf{R^n})$. Then $f * g$ is uniformly continuous and bounded on $\mathbf{R^n}$, and $\|f * g\|_\infty \le \|f\|_2 \|g\|_2$.*

Proof. We will use Schwarz's inequality. Let $f, g \in L^2(\mathbf{R^n})$. For any $\mathbf{x} \in \mathbf{R^n}$,

$$\begin{aligned}|(f * g)(\mathbf{x})| &\le \int_{\mathbf{R^n}} |f(\mathbf{x}-\mathbf{y})g(\mathbf{y})|\, d\mathbf{y} \\ &\le \left(\int_{\mathbf{R^n}} |f(\mathbf{x}-\mathbf{y})|^2 d\mathbf{y}\right)^{1/2} \left(\int_{\mathbf{R^n}} |g(\mathbf{y})|^2 d\mathbf{y}\right)^{1/2} = \|f\|_2 \|g\|_2.\end{aligned}$$

In particular, $f * g$ is finite everywhere and $\|f * g\|_\infty \le \|f\|_2 \|g\|_2$. Also, if $\mathbf{x}, \mathbf{h} \in \mathbf{R^n}$, then

$$\begin{aligned}|(f * g)(\mathbf{x}+\mathbf{h}) - (f * g)(\mathbf{x})| &= \left|\int_{\mathbf{R^n}} [f(\mathbf{x}+\mathbf{h}-\mathbf{y}) - f(\mathbf{x}-\mathbf{y})] g(\mathbf{y})\, d\mathbf{y}\right| \\ &\le \|f(\cdot + \mathbf{h}) - f(\cdot)\|_2 \|g\|_2 \to 0 \text{ as } |\mathbf{h}| \to 0\end{aligned}$$

by continuity in L^2, and the proof is complete.

We can now derive the key fact needed in order to extend the definition of the Fourier transform to $L^2(\mathbf{R^n})$.

Lemma 13.50 *If $f \in L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$, then $\widehat{f} \in L^2(\mathbf{R^n})$ and*

$$\|\widehat{f}\|_2 = (2\pi)^{-n/2} \|f\|_2.$$

Proof. Let f satisfy the hypothesis, and set $g(\mathbf{x}) = \overline{f(-\mathbf{x})}$. Then $g \in L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$. By Lemma 13.49, $f * g$ is continuous in $\mathbf{R^n}$, and

$$\begin{aligned}(f * g)(\mathbf{x}) &= \int_{\mathbf{R^n}} f(\mathbf{y}) g(\mathbf{x}-\mathbf{y})\, d\mathbf{y} = \int_{\mathbf{R^n}} f(\mathbf{y}) \overline{f(\mathbf{y}-\mathbf{x})}\, d\mathbf{y}, \\ (f * g)(\mathbf{0}) &= \int_{\mathbf{R^n}} f(\mathbf{y}) \overline{f(\mathbf{y})}\, d\mathbf{y} = \|f\|_2^2.\end{aligned}$$

Using Theorem 13.30, and since $\widehat{g}(\mathbf{x}) = \overline{\widehat{f}(\mathbf{x})}$, we also have

$$\widehat{f * g}(\mathbf{x}) = (2\pi)^n \widehat{f}(\mathbf{x}) \widehat{g}(\mathbf{x}) = (2\pi)^n |\widehat{f}(\mathbf{x})|^2 \ge 0.$$

Corollary 13.21 applied to $f * g$ then implies that $\widehat{f * g} \in L^1(\mathbf{R^n})$ with L^1 norm

$$\|\widehat{f * g}\|_1 = (f * g)(\mathbf{0}),$$

or equivalently,

$$(2\pi)^n \int_{\mathbf{R^n}} |\widehat{f}(\mathbf{x})|^2 d\mathbf{x} = \|f\|_2^2,$$

and the proof is complete.

A simple method often referred to as *extension by continuity* and based on the fact that $L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ is dense in $L^2(\mathbf{R^n})$ (in fact, the subset $C_0^\infty(\mathbf{R^n})$ of $L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ is dense in $L^2(\mathbf{R^n})$ by Corollary 9.7) allows us to use Lemma 13.50 to define the Fourier transform on $L^2(\mathbf{R^n})$. We proceed as follows.

Given an $f \in L^2(\mathbf{R^n})$, choose a sequence $\{f_j\}_{j=1}^\infty \subset L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ such that $\|f_j - f\|_2 \to 0$ as $j \to \infty$. By Lemma 13.50, for all j and ℓ,

$$\|\widehat{f_j} - \widehat{f_\ell}\|_2 = (2\pi)^{-n/2} \|f_j - f_\ell\|_2.$$

Since $\{f_j\}$ is a Cauchy sequence in $L^2(\mathbf{R^n})$, then so is $\{\widehat{f_j}\}$, and hence there is a function $\mathscr{F}f \in L^2(\mathbf{R^n})$ such that $\widehat{f_j} \to \mathscr{F}f$ in $L^2(\mathbf{R^n})$. Note that $\mathscr{F}f$ is independent of the particular sequence $\{f_j\}$; in fact, if $\{g_j\} \subset L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ is another sequence that satisfies $\|g_j - f\|_2 \to 0$, then $\widehat{g_j} \to \mathscr{F}f$ in $L^2(\mathbf{R^n})$ since

$$\begin{aligned}\|\widehat{g_j} - \mathscr{F}f\|_2 &\le \|\widehat{g_j} - \widehat{f_j}\|_2 + \|\widehat{f_j} - \mathscr{F}f\|_2 \\ &= (2\pi)^{-n/2}\|g_j - f_j\|_2 + \|\widehat{f_j} - \mathscr{F}f\|_2 \to 0 \text{ as } j \to \infty.\end{aligned}$$

As usual, L^2 functions that are equal a.e. are considered to be the same function.

Moreover, in case $f \in L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$, we see that $\mathscr{F}f = \widehat{f}$ a.e. by choosing $\{f_j\}$ earlier to be the sequence with $f_j = f$ for every j.

By definition, the *Fourier transform of an L^2 function f* is $\mathscr{F}f$. We adopt the convention of denoting $\mathscr{F}f$ by $\widehat{f}$ if $f \in L^2(\mathbf{R^n})$.

Many of the properties of $\widehat{f}$ that were proved in Section 13.1 for integrable f extend to analogous properties when $f \in L^2(\mathbf{R^n})$ simply by using the definition of $\widehat{f}$ for $f \in L^2(\mathbf{R^n})$. For example, if $f \in L^2(\mathbf{R^n})$, then $\overline{\widehat{f}(\mathbf{x})} = \widehat{\overline{f}}(-\mathbf{x})$ and

$$\widehat{\tau_{\mathbf{h}} f}(\mathbf{x}) = e^{i\mathbf{x}\cdot\mathbf{h}}\widehat{f}(\mathbf{x}) \quad \Big[\text{where } \big(\tau_{\mathbf{h}} f\big)(\mathbf{x}) = f(\mathbf{x} + \mathbf{h}),\ \mathbf{h} \in \mathbf{R^n}\Big],$$

$$\widehat{\delta_\lambda f}(\mathbf{x}) = \frac{1}{|\lambda|^n}\widehat{f}\left(\frac{\mathbf{x}}{\lambda}\right) \quad \Big[\text{where } \big(\delta_\lambda f\big)(\mathbf{x}) = f(\lambda\mathbf{x}),\ \lambda \in \mathbf{R^1} - \{\mathbf{0}\}\Big].$$

For functions $f \in L^2(\mathbf{R^n})$, such formulas must be interpreted as holding for a.e. $\mathbf{x}$ instead of all $\mathbf{x}$ as is true for integrable functions. In order to verify the formula for $\widehat{\tau_{\mathbf{h}} f}$, choose a sequence $\{f_j\} \subset L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ with $\|f_j - f\|_2 \to 0$. Then $\tau_{\mathbf{h}} f_j$ converges in $L^2(\mathbf{R^n})$ to $\tau_{\mathbf{h}} f$, and consequently, $\widehat{\tau_{\mathbf{h}} f_j}$ converges to $\widehat{\tau_{\mathbf{h}} f}$ in $L^2(\mathbf{R^n})$. But by (13.5), the formula holds for every f_j, that is, $\widehat{\tau_{\mathbf{h}} f_j}(\mathbf{x}) = e^{i\mathbf{x}\cdot\mathbf{h}} \widehat{f_j}(\mathbf{x})$ for all $\mathbf{x}$. On the other hand, since $\|\widehat{f_j} - \widehat{f}\|_2 \to 0$, then $e^{i\mathbf{x}\cdot\mathbf{h}} \widehat{f_j}(\mathbf{x})$ converges in $L^2(\mathbf{R^n})$ to $e^{i\mathbf{x}\cdot\mathbf{h}} \widehat{f}(\mathbf{x})$. The desired formula for $\widehat{\tau_h f}$ now follows immediately. The formulas for $\widehat{\overline{f}}(\mathbf{x})$ and for $\widehat{\delta_\lambda f}$ can be proved similarly. Some more properties of $\widehat{f}$ that extend easily from the case when $f \in L^1(\mathbf{R^n})$ to the case when $f \in L^2(\mathbf{R^n})$ will be derived below.

Two notable exceptions to the parallel properties of the Fourier transform of L^1 and L^2 functions are that when $f \in L^2(\mathbf{R^n})$, $\widehat{f}$ may not be continuous or bounded, and the conclusion of the Riemann–Lebesgue theorem that $\widehat{f}(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$ may fail. In fact, the next result, which is basic, shows that every L^2 function is the Fourier transform of an L^2 function.

Theorem 13.51 (Plancherel) *The mapping $\mathscr{F} : f \to \mathscr{F} f = \widehat{f}$ is a one-to-one linear transformation of $L^2(\mathbf{R^n})$ onto itself, and*

$$\|\widehat{f}\|_2 = (2\pi)^{-n/2} \|f\|_2, \quad f \in L^2(\mathbf{R^n}). \tag{13.52}$$

Formula (13.52) is called *Plancherel's formula*.

Proof. Let $f \in L^2(\mathbf{R^n})$ and choose $\{f_j\} \subset L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ such that $\|f_j - f\|_2 \to 0$. Then $\|\widehat{f_j} - \widehat{f}\|_2 \to 0$. Hence $\|f_j\|_2 \to \|f\|_2$ and $\|\widehat{f_j}\|_2 \to \|\widehat{f}\|_2$. Plancherel's formula (13.52) for f then follows immediately since $\|\widehat{f_j}\|_2 = (2\pi)^{-n/2} \|f_j\|_2$ for every j by Lemma 13.50.

Next, let us show that linearity of $\mathscr{F}$ on $L^2(\mathbf{R^n})$ follows in a standard way from the corresponding property (13.3) for integrable functions. To see this, let $f, g \in L^2(\mathbf{R^n})$ and choose $\{f_j\}, \{g_j\} \in L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ such that $f_j \to f$ and $g_j \to g$ in $L^2(\mathbf{R^n})$. For any constants c_1, c_2, we have $\widehat{\big(c_1 f_j + c_2 g_j\big)} = c_1 \widehat{f_j} + c_2 \widehat{g_j}$ by (13.3). However, since $c_1 f_j + c_2 g_j \to c_1 f + c_2 g$ in $L^2(\mathbf{R^n})$,

$$\widehat{\big(c_1 f_j + c_2 g_j\big)} \to \widehat{\big(c_1 f + c_2 g\big)} \text{ in } L^2(\mathbf{R^n}),$$

and since $\widehat{f_j} \to \widehat{f}$, $\widehat{g_j} \to \widehat{g}$ in $L^2(\mathbf{R^n})$,

$$c_1 \widehat{f_j} + c_2 \widehat{g_j} \to c_1 \widehat{f} + c_2 \widehat{g} \text{ in } L^2(\mathbf{R^n}).$$

Hence, $\widehat{\big(c_1 f + c_2 g\big)} = c_1 \widehat{f} + c_2 \widehat{g}$, as desired.

The fact that $\mathscr{F}$ is a one-to-one map is a corollary of its linearity and Plancherel's formula since if $f, g \in L^2(\mathbf{R^n})$ and $\widehat{f} = \widehat{g}$, then

$$\|f - g\|_2 = (2\pi)^{n/2}\|(f-g)\widehat{\ }\|_2 = (2\pi)^{n/2}\|\widehat{f} - \widehat{g}\|_2 = 0,$$

and therefore $f = g$ a.e.

To complete the proof of Theorem 13.51, it remains to show that the range of $\mathscr{F}$ is all of $L^2(\mathbf{R^n})$. We will prove this by using Corollary 13.38 and the following extension to L^2 of property (13.14) about shifting hats:

$$\int_{\mathbf{R^n}} \widehat{f}\, g\, d\mathbf{x} = \int_{\mathbf{R^n}} f\, \widehat{g}\, d\mathbf{x} \quad \text{if } f, g \in L^2\left(\mathbf{R^n}\right). \tag{13.53}$$

To prove (13.53), let $f, g \in L^2(\mathbf{R^n})$ and choose $\{f_j\}$, $\{g_j\}$ in $L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$ such that $f_j \to f$, $g_j \to g$ in $L^2(\mathbf{R^n})$. By (13.14), for every j,

$$\int_{\mathbf{R^n}} \widehat{f_j} g_j\, \mathbf{x} = \int_{\mathbf{R^n}} f_j \widehat{g_j}\, d\mathbf{x}.$$

Then (13.53) follows by passing to the limit since $\widehat{f_j} \to \widehat{f}$ and $\widehat{g_j} \to \widehat{g}$ in $L^2(\mathbf{R^n})$.

Finally, let $f \in L^2(\mathbf{R^n})$ and define F by

$$F(\mathbf{x}) = (2\pi)^n \widehat{f}(-\mathbf{x}).$$

Then $F \in L^2(\mathbf{R^n})$, and we claim that $f = \widehat{F}$. For any $\varphi \in C_0^\infty(\mathbf{R^n})$, Corollary 13.38 implies that

$$\varphi(\mathbf{x}) = (2\pi)^n \widehat{\widehat{\varphi}(-\mathbf{x})},$$

and therefore

$$\begin{aligned}
\int_{\mathbf{R^n}} f(\mathbf{x})\varphi(\mathbf{x})\, d\mathbf{x} &= \int_{\mathbf{R^n}} f(\mathbf{x})\, (2\pi)^n \widehat{\widehat{\varphi}(-\mathbf{x})}\, d\mathbf{x} \\
&= \int_{\mathbf{R^n}} \widehat{f}(\mathbf{x})(2\pi)^n \widehat{\varphi}(-\mathbf{x})\, d\mathbf{x} \quad \text{by (13.53)} \\
&= \int_{\mathbf{R^n}} \widehat{f}(-\mathbf{x})\, (2\pi)^n \widehat{\varphi}(\mathbf{x})\, d\mathbf{x} = \int_{\mathbf{R^n}} F(\mathbf{x})\widehat{\varphi}(\mathbf{x})\, d\mathbf{x} \\
&= \int_{\mathbf{R^n}} \widehat{F}(\mathbf{x})\varphi(\mathbf{x})\, d\mathbf{x} \quad \text{by (13.53) again.}
\end{aligned}$$

Hence,

$$\int_{\mathbf{R^n}} f\varphi\, d\mathbf{x} = \int_{\mathbf{R^n}} \widehat{F}\varphi\, d\mathbf{x} \text{ for all } \varphi \in C_0^\infty(\mathbf{R^n}).$$

Therefore $f = \widehat{F}$ a.e., which verifies our claim and completes the proof of Theorem 13.51.

An alternate proof (still using (13.53)) that $\mathscr{F}$ is surjective from L^2 to L^2 is indicated in Exercise 14.

The next theorem records some facts about the inverse map $\mathscr{F}^{-1}$ of $\mathscr{F}$ on $L^2(\mathbf{R^n})$, including analogues of (13.17) and (13.18). Its proof follows easily from the statement and the proof of Plancherel's theorem and is omitted.

Theorem 13.54 *The inverse map $\mathscr{F}^{-1}$ of $\mathscr{F}$ is a one-to-one linear mapping of $L^2(\mathbf{R^n})$ onto itself, and*

$$\|\mathscr{F}^{-1}(f)\|_2 = (2\pi)^{n/2}\|f\|_2.$$

Also, for every $f \in L^2(\mathbf{R^n})$ and a.e. $\mathbf{x} \in \mathbf{R^n}$,

$$(\mathscr{F}^{-1}f)(\mathbf{x}) = (2\pi)^n\widehat{f}(-\mathbf{x})$$

and

$$f(\mathbf{x}) = (2\pi)^n\widehat{\widehat{f}}(-\mathbf{x}) = (2\pi)^n\left(\widehat{f}(-\mathbf{x})\right)^{\widehat{}}.$$

$\mathscr{F}$ fails to be isometry on $L^2(\mathbf{R^n})$, that is, $\|f\|_2$ and $\|\widehat{f}\|_2$ are not identical, only because of the factor $(2\pi)^{-n/2}$ in (13.52). However, simple renormalizations of $\mathscr{F}$ lead to isometries. We leave it to the reader to check that both of the maps $\mathscr{F}_1$ and $\mathscr{F}_2$ defined by

$$\mathscr{F}_1 f(\mathbf{x}) = (2\pi)^{n/2}\,\widehat{f}(\mathbf{x}) \quad \text{and} \quad \mathscr{F}_2 f(\mathbf{x}) = (2\pi)^n\,\widehat{f}(2\pi\mathbf{x})$$

are isometries of $L^2(\mathbf{R^n})$ onto itself.

Plancherel's formula also gives analogues of (8.32) and (12.18), namely,

Theorem 13.55 *If $f, g \in L^2(\mathbf{R^n})$, then*

$$\int_{\mathbf{R^n}} f(\mathbf{x})\,\overline{g(\mathbf{x})}\, d\mathbf{x} = (2\pi)^n \int_{\mathbf{R^n}} \widehat{f}(\mathbf{x})\,\overline{\widehat{g}(\mathbf{x})}\, d\mathbf{x}.$$

The proof is left as an exercise.

Next, we turn to an analogue of the fact (see p. 303 in Section 12.1 and Theorem 8.29) that the trigonometric Fourier series $S[f]$ of an $f \in L^2(-\pi, \pi)$ converges to f in $L^2(-\pi, \pi)$ norm.

Theorem 13.56 *Let $f \in L^2(\mathbf{R^n})$. Then as $k \to \infty$,*

$$\int_{|\mathbf{y}|<k} \widehat{f}(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y} \to f(\mathbf{x}) \text{ in } L^2(\mathbf{R^n}) \text{ norm}, \tag{13.57}$$

and

$$\frac{1}{(2\pi)^n} \int_{|\mathbf{y}|<k} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y} \to \widehat{f}(\mathbf{x}) \text{ in } L^2(\mathbf{R^n}) \text{ norm}. \tag{13.58}$$

Proof. It is enough to prove (13.58) since it is equivalent to (13.57) by Theorem 13.54. If $f \in L^2(\mathbf{R^n})$ and $k = 1, 2, \ldots$, the functions f_k defined by $f_k(\mathbf{y}) = f(\mathbf{y})\chi_{\{|\mathbf{y}|<k\}}(\mathbf{y})$ belong to $L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$, and, by the Lebesgue dominated convergence theorem, $\|f_k - f\|_2 \to 0$. Hence $\|\widehat{f_k} - \widehat{f}\|_2 \to 0$, which is the same as (13.58) in case k is restricted to the positive integers. Note that the proof works equally well if the sequence $k = 1, 2, \ldots$ is replaced by any sequence $\{t_k\}$ of real numbers with $t_k \to \infty$. Consequently, k can be a continuous real variable tending to ∞, and the proof is complete.

We close this section with two results about the Fourier transform of the convolution $f * g$ of an L^1 function and an L^2 function. Recall that $f * g$ then belongs to L^2 by Young's convolution theorem, and therefore $\widehat{f * g}$ is well-defined as an L^2 function.

Theorem 13.59 *If $f \in L^1(\mathbf{R^n})$ and $g \in L^2(\mathbf{R^n})$, then*

$$\widehat{f * g}(\mathbf{x}) = (2\pi)^n \widehat{f}(\mathbf{x})\, \widehat{g}(\mathbf{x}) \text{ a.e. in } \mathbf{R^n}.$$

Proof. Fix f and g satisfying the hypothesis, and choose $g_k \in L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$, $k = 1, 2, \ldots$, such that $g_k \to g$ in $L^2(\mathbf{R^n})$. By Young's Theorem 9.2, it follows that $f * g_k \to f * g$ in $L^2(\mathbf{R^n})$, and hence

$$\widehat{f * g_k} \to \widehat{f * g} \quad \text{in } L^2(\mathbf{R^n}).$$

On the other hand, we have $\widehat{f * g_k} = (2\pi)^n \widehat{f}\,\widehat{g_k}$ by Theorem 13.30, and therefore

$$(2\pi)^n \widehat{f}\,\widehat{g_k} \to \widehat{f * g} \text{ in } L^2(\mathbf{R^n}).$$

But also $(2\pi)^n \widehat{f}\,\widehat{g_k} \to (2\pi)^n \widehat{f}\,\widehat{g}$ in $L^2(\mathbf{R^n})$ since $\widehat{f}$ is bounded, and the result follows.

Corollary 13.60 *Let $K \in L^1(\mathbf{R^n})$ and $f \in L^2(\mathbf{R^n})$. Then*

$$\|f * K\|_2 \le (2\pi)^n \|\widehat{K}\|_\infty \|f\|_2.$$

Thus, the function operator T defined by $T : f \to f * K$, when considered as a transformation from $L^2(\mathbf{R^n})$ into itself, has operator norm at most $(2\pi)^n \|\widehat{K}\|_\infty$.

Proof. Let $f \in L^2(\mathbf{R^n})$ and $K \in L^1(\mathbf{R^n})$. By Plancherel's Theorem (13.51) and Theorem 13.59,

$$\begin{aligned}\|f * K\|_2 &= (2\pi)^{n/2} \|\widehat{f * K}\|_2 = (2\pi)^{n/2} \|(2\pi)^n\, \widehat{f}\, \widehat{K}\|_2 \\ &\le (2\pi)^n \|\widehat{K}\|_\infty (2\pi)^{n/2} \|\hat{f}\|_2 = (2\pi)^n \|\widehat{K}\|_\infty \|f\|_2,\end{aligned}$$

and the proof is complete.

13.3 The Hilbert Transform on L^2

In this section, we will use Plancherel's theorem to define the Hilbert transform $Hf(x)$ of a general $f \in L^2(-\infty, \infty)$ and to prove the important classical result that the mapping $H : f \to Hf$ is bounded on $L^2(-\infty, \infty)$.

Given an $f \in L^2(-\infty, \infty)$, its *Hilbert transform* is *formally* defined for $-\infty < x < \infty$ to be the principal value integral

$$Hf(x) = \text{p.v.}\, \frac{1}{\pi} \int_{-\infty}^{\infty} f(x-y) \frac{1}{y}\, dy = \lim_{\substack{\varepsilon \to 0+ \\ \omega \to \infty}} \frac{1}{\pi} \int_{\varepsilon < |y| < \omega} f(x-y) \frac{1}{y}\, dy. \tag{13.61}$$

We include the constant factor $1/\pi$ in the definition of Hf for historical and natural reasons (see Exercise 17). Since $1/x$ is real-valued, it is natural to also assume that f is real-valued by considering the real and imaginary parts of f separately. Hf is generally regarded as the nonperiodic analogue of the

conjugate function $\widetilde{f}$ studied in Chapter 12 (see Exercise 17 and the comments at the end of this section). Sometimes, we also write simply

$$Hf = \frac{1}{\pi}\,\text{p.v.}\,(f * K), \text{ where } K(x) = \frac{1}{x}.$$

The standard convolution $f * (1/x)$ may diverge everywhere in $\mathbf{R^1}$ even when f is a smooth L^2 function since $1/x$ is not integrable near $x = 0$. Our goal is to give a sense in which the principal value definition (13.61) is well-defined for every $f \in L^2(\mathbf{R^1})$.

We will use the truncated kernels $K_{\varepsilon,\omega}$ in (13.44):

$$K_{\varepsilon,\omega}(x) = \frac{\chi_{\{\varepsilon<|x|<\omega\}}(x)}{x}, \quad 0 < \varepsilon < \omega < \infty.$$

Since each $K_{\varepsilon,\omega} \in L^1(\mathbf{R^1})$, Young's convolution theorem implies that the doubly *truncated Hilbert transform* $H_{\epsilon,\omega}f$ defined by

$$\big(H_{\epsilon,\omega}f\big)(x) = \frac{1}{\pi}\big(f * K_{\epsilon,\omega}\big)(x) = \frac{1}{\pi}\int\limits_{\epsilon<|y|<\omega} f(x-y)\frac{1}{y}\,dy \tag{13.62}$$

belongs to $L^2(\mathbf{R^1})$ for all $0 < \varepsilon < \omega < \infty$ if $f \in L^2(\mathbf{R^1})$.

However, as we have already noted, the norms $\|K_{\varepsilon,\omega}\|_1 = 2\log(\omega/\varepsilon)$ are unbounded in ε and ω. Consequently, the simple estimate obtained by applying Young's theorem, namely,

$$\|H_{\varepsilon,\omega}f\|_2 \le \|f\|_2\|K_{\varepsilon,\omega}\|_1,$$

does not guarantee boundedness of $\|H_{\varepsilon,\omega}f\|_2$ in ε, ω if $f \in L^2(\mathbf{R^n})$.

Instead of Young's theorem, the next result uses Theorem 13.45, (13.47), and Corollary 13.60 to interpret the convergence of (13.62) when $f \in L^2(\mathbf{R^1})$ (see also Exercise 16 regarding the convergence of the singly truncated Hilbert transform $H_\varepsilon f(x) = (1/\pi)\int_{|y|>\varepsilon} f(x-y)/y\,dy$ as $\varepsilon \to 0$).

Theorem 13.63 *If $f \in L^2(\mathbf{R^1})$, then the truncated Hilbert transform $H_{\varepsilon,\omega}f$ defined in (13.62) converges in $L^2(\mathbf{R^1})$, that is, there is a function $Hf \in L^2(\mathbf{R^1})$ such that*

$$\|H_{\varepsilon,\omega}f - Hf\|_2 \to 0 \quad \textit{as } \varepsilon \to 0 \textit{ and } \omega \to \infty. \tag{13.64}$$

Furthermore,

$$\|Hf\|_2 = \|f\|_2, \tag{13.65}$$

$$\widehat{Hf}(x) = (-i \operatorname{sign} x)\widehat{f}(x) \text{ a.e.}, \tag{13.66}$$

and there is a constant c, independent of f, ε, and ω, such that

$$\|H_{\varepsilon,\omega} f\|_2 \le c\,\|f\|_2. \tag{13.67}$$

Proof. Let $f \in L^2(\mathbf{R^1})$. Corollary 13.60 gives

$$\begin{aligned}\|H_{\varepsilon,\omega} f\|_2 &= \frac{1}{\pi}\|f * K_{\varepsilon,\omega}\|_2 \le \frac{1}{\pi}\cdot 2\pi\,\|\widehat{K_{\varepsilon,\omega}}\|_\infty \|f\|_2 \\ &\le 2\cdot\frac{4}{\pi}\,\|f\|_2 \text{ by (13.47)},\end{aligned}$$

which proves (13.67) with $c = 8/\pi$.

To show that $H_{\varepsilon,\omega} f$ converges in $L^2(\mathbf{R^1})$, let $0 < \varepsilon < \omega < \infty$ and $0 < \varepsilon' < \omega' < \infty$. By Plancherel's theorem and Theorem 13.59,

$$\begin{aligned}\|H_{\varepsilon,\omega} f - H_{\varepsilon',\omega'} f\|_2 &= (2\pi)^{1/2}\|\widehat{H_{\varepsilon,\omega} f} - \widehat{H_{\varepsilon',\omega'} f}\|_2 \\ &= (2\pi)^{1/2}\frac{1}{\pi}\,2\pi\,\|\left(\widehat{K_{\varepsilon,\omega}} - \widehat{K_{\varepsilon',\omega'}}\right)\widehat{f}\|_2.\end{aligned}$$

The last expression tends to 0 as $\varepsilon, \varepsilon' \to 0$ and $\omega, \omega' \to \infty$ by Lebesgue's dominated convergence theorem since $\widehat{f} \in L^2(\mathbf{R^1})$ and $\widehat{K_{\varepsilon,\omega}}(x) - \widehat{K_{\varepsilon',\omega'}}(x)$ is uniformly bounded in $x, \varepsilon, \omega, \varepsilon', \omega'$ by (13.47) and tends pointwise to 0 for all x as $\varepsilon, \varepsilon' \to 0$ and $\omega, \omega' \to \infty$ by Theorem 13.45. Since $L^2(\mathbf{R^1})$ is complete, it follows that $H_{\varepsilon,\omega} f$ converges in $L^2(\mathbf{R^1})$. Let Hf denote its limit in $L^2(\mathbf{R^1})$. This proves (13.64).

Then, by Plancherel's theorem, as $\varepsilon, \varepsilon' \to 0$ and $\omega, \omega' \to \infty$,

$$\widehat{H_{\varepsilon,\omega} f} \to \widehat{Hf} \text{ in } L^2(\mathbf{R^1}). \tag{13.68}$$

On the other hand, by Theorem 13.59 again,

$$\widehat{H_{\varepsilon,\omega} f} = \frac{1}{\pi}\widehat{\left(f * K_{\varepsilon,\omega}\right)} = \frac{1}{\pi}\cdot 2\pi\widehat{f}\,\widehat{K_{\varepsilon,\omega}}.$$

But $\widehat{f}(x)\,\widehat{K_{\varepsilon,\omega}}(x)$ converges in $L^2(\mathbf{R^1})$ to $\widehat{f}(x)\ \{(-i/2)\operatorname{sign} x\}$, again by dominated convergence and the pointwise convergence of $\widehat{K_{\varepsilon,\omega}}(x)$ to $(-i/2)\operatorname{sign} x$ proved in Theorem 13.45. Hence,

$$\widehat{H_{\varepsilon,\omega}f}(x) \to \frac{1}{\pi}\cdot 2\pi\widehat{f}(x)\left\{\frac{-i}{2}\operatorname{sign} x\right\} = (-i\operatorname{sign} x)\widehat{f}(x) \text{ in } L^2(\mathbf{R^1}). \tag{13.69}$$

Combining (13.69) and (13.68) proves (13.66).

Finally, we have

$$\begin{aligned}\|Hf\|_2 &= (2\pi)^{1/2}\|\widehat{Hf}\|_2 = (2\pi)^{1/2}\|(-i\operatorname{sign} x)\widehat{f}\|_2\\ &= (2\pi)^{1/2}\|\widehat{f}\|_2 = \|f\|_2.\end{aligned}$$

This verifies (13.65), and the proof is complete.

Corollary 13.70 *If $f \in L^2(\mathbf{R^1})$, then $H(Hf) = -f$ a.e. In particular, the map $H : f \to Hf$ is an isometry of $L^2(\mathbf{R^1})$ onto itself.*

Proof. If $f \in L^2(\mathbf{R^1})$, then by (13.66),

$$\widehat{H(Hf)}(x) = (-i\operatorname{sign} x)\,\widehat{Hf}(x) = (-i\operatorname{sign} x)^2\widehat{f}(x) = -\widehat{f}(x) \text{ a.e.}$$

Hence, $H(Hf) = -f$ a.e. Since H is clearly a linear operator on $L^2(\mathbf{R^1})$, we obtain that $H(-Hf) = f$ a.e. Therefore, H maps $L^2(\mathbf{R^1})$ onto itself. The fact that H is an isometry is due to (13.65).

In passing, we mention without proof that Hf can be defined by (13.61) if f belongs to $L^p(\mathbf{R^1})$ for any p, $1 \le p < \infty$, and that the operator $H : f \to Hf$ shares the main properties of the conjugate function map $C : f \to \widetilde{f}$ defined in Chapter 12. The theory of the Hilbert transform serves as a gateway to the very broad field of *singular integral operators* in $\mathbf{R^n}$, $n = 1, 2, \ldots$. A fairly typical form of a singular integral is

$$Tf(\mathbf{x}) = \int_{\mathbf{R^n}} f(\mathbf{y})K(\mathbf{x},\mathbf{y})\,d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n},$$

where the kernel K satisfies $|K(\mathbf{x},\mathbf{y})| \le A|\mathbf{x}-\mathbf{y}|^{-n}$ for some fixed constant A and has additional cancellation properties that ensure a rich theory. Many related topics can be found in *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, New Jersey, 1970, and

Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press, Princeton, New Jersey, 1993, both authored by E. M. Stein, as well as in the references listed there.

13.4 The Fourier Transform on L^p, $1 < p < 2$

In this section, the properties of the Fourier transform of L^1 and L^2 functions will be used to define the Fourier transform $\widehat{f}$ when f belongs to L^p for some p with $1 < p < 2$ and also to show that the map $\mathscr{F} : f \to \widehat{f}$ is bounded from L^p into $L^{p'}$, $1/p + 1/p' = 1$.

Let us first show that the set $L^1(\mathbf{R^n}) + L^2(\mathbf{R^n})$ defined by

$$L^1(\mathbf{R^n}) + L^2(\mathbf{R^n}) = \left\{f : f = f_1 + f_2, f_1 \in L^1(\mathbf{R^n}), f_2 \in L^2(\mathbf{R^n})\right\}$$

contains $L^p(\mathbf{R^n})$, $1 \le p \le 2$. This is obvious if p is either 1 or 2. If $1 < p < 2$ and $f \in L^p(\mathbf{R^n})$, the functions f_1 and f_2 in the simple decomposition

$$f = f\chi_{\{|f|\ge 1\}} + f\chi_{\{|f|<1\}} = f_1 + f_2$$

satisfy

$$\int_{\mathbf{R^n}} |f_1|\, d\mathbf{x} \le \int_{\mathbf{R^n}} |f_1|^p d\mathbf{x} = \int_{\{|f|\ge 1\}} |f|^p d\mathbf{x} \le \|f\|_p^p < \infty$$

and

$$\int_{\mathbf{R^n}} |f_2|^2 d\mathbf{x} \le \int_{\mathbf{R^n}} |f_2|^p d\mathbf{x} = \int_{\{|f|<1\}} |f|^p d\mathbf{x} \le \|f\|_p^p < \infty,$$

and therefore $f \in L^1(\mathbf{R^n}) + L^2(\mathbf{R^n})$.

Now, given any $f \in L^1(\mathbf{R^n}) + L^2(\mathbf{R^n})$ and a decomposition $f = f_1 + f_2$ with $f_1 \in L^1(\mathbf{R^n})$ and $f_2 \in L^2(\mathbf{R^n})$, we may define $\widehat{f}$ by

$$\widehat{f} = \widehat{f_1} + \widehat{f_2}$$

provided the right side is independent of any particular such decomposition of f. To show that this is the case, suppose that $f = f_1 + f_2 = g_1 + g_2$ with $f_j, g_j \in L^j(\mathbf{R^n})$, $j = 1, 2$. Then $f_1 - g_1 = g_2 - f_2 \in L^1(\mathbf{R^n}) \cap L^2(\mathbf{R^n})$, and consequently, $\widehat{f_1} - \widehat{g_1} = (f_1 - g_1)\widehat{\ } = (g_2 - f_2)\widehat{\ } = \widehat{g_2} - \widehat{f_2}$ a.e., or equivalently $\widehat{f_1} + \widehat{f_2} = \widehat{g_1} + \widehat{g_2}$ a.e. as desired.

In particular, $\widehat{f}$ is well-defined in this way if $f \in L^p(\mathbf{R^n})$ and $1 \le p \le 2$. Also, if f belongs to either $L^1(\mathbf{R^n})$ or $L^2(\mathbf{R^n})$, then the present definition of $\widehat{f}$ agrees with the appropriate earlier one by considering the trivial decomposition of f as the sum of f and 0.

The next result is the main one of the section.

Theorem 13.71 (Titchmarsh, Hausdorff–Young) *If $1 \le p \le 2$, then there is a constant C depending only on p such that*

$$\|\widehat{f}\|_{p'} \le C\,\|f\|_p, \quad \frac{1}{p} + \frac{1}{p'} = 1, \tag{13.72}$$

for all $f \in L^p(\mathbf{R^n})$.

Proof. It suffices to prove the theorem for $1 < p < 2$ since the endpoint cases $p = 1$ and $p = 2$ have already been established. We will prove the result for $1 < p < 2$ by adapting the interpolation method of Marcinkiewicz to the present situation. Recall that, in a different context, this method of interpolation is used in the proof of Theorem 9.16 (see also Exercise 10 of Chapter 9). The Marcinkiewicz method will not give the best value of the constant C in (13.72) (see the comments after the proof).

Fix f and p with $f \in L^p(\mathbf{R^n})$ and $1 < p < 2$. Since the Fourier transform is a linear operator, we may assume without loss of generality that $\|f\|_p = 1$. Let $\omega(\alpha)$ denote the distribution function of $|\widehat{f}|$:

$$\omega(\alpha) = |\{\mathbf{x} \in \mathbf{R^n} : |\widehat{f}(\mathbf{x})| > \alpha\}|, \quad \alpha > 0.$$

It suffices to show that there is a constant C depending only on p such that

$$p' \int_0^\infty \alpha^{p'-1}\omega(\alpha)\,d\alpha \le C$$

since the expression on the left side equals $\|\widehat{f}\|_{p'}^{p'}$.

For each $\alpha > 0$, define functions g and h by

$$g = g_\alpha = f\,\chi_{\{|f|^{p-1} \ge 1/\alpha\}} \quad \text{and} \quad h = h_\alpha = f\,\chi_{\{|f|^{p-1} < 1/\alpha\}}.$$

Note that $f = g + h$. Also, $g \in L^1(\mathbf{R^n})$ and $h \in L^2(\mathbf{R^n})$ with norms satisfying

$$\|g\|_1 = \int_{\{|f|^{p-1} \ge \alpha^{-1}\}} |f|\,d\mathbf{x} \le \alpha \int_{\{|f|^{p-1} \ge \alpha^{-1}\}} |f|^p d\mathbf{x} \le \alpha\|f\|_p^p = \alpha, \tag{13.73}$$

and

$$\|h\|_2^2 = \int\limits_{\{0<|f|<\alpha^{-1/(p-1)}\}} |f|^2 d\mathbf{x} \tag{13.74}$$
$$\leq \alpha^{-\frac{2-p}{p-1}} \int\limits_{\mathbf{R^n}} |f|^p d\mathbf{x} < \infty.$$

Then $\widehat{f} = \widehat{g} + \widehat{h}$ and, by (13.73),

$$|\widehat{f}| \leq |\widehat{g}| + |\widehat{h}| \leq \frac{\alpha}{(2\pi)^n} + |\widehat{h}| < \frac{\alpha}{2} + |\widehat{h}|.$$

Hence,

$$\left\{|\widehat{f}| > \alpha\right\} \subset \left\{|\widehat{h}| > \frac{\alpha}{2}\right\} \quad \text{and}$$
$$\omega(\alpha) = \left|\left\{|\widehat{f}| > \alpha\right\}\right| \leq \left|\left\{|\widehat{h}| > \frac{\alpha}{2}\right\}\right|$$
$$\leq \frac{1}{(\alpha/2)^2} \int\limits_{\mathbf{R^n}} |\widehat{h}|^2 d\mathbf{x} = \frac{4}{\alpha^2} \frac{1}{(2\pi)^n} \|h\|_2^2$$
$$\leq \frac{1}{\alpha^2} \int\limits_{\{0<|f|<\alpha^{-1/(p-1)}\}} |f|^2 d\mathbf{x},$$

where we used (13.74) to obtain the last estimate. Therefore,

$$\int\limits_0^\infty \alpha^{p'-1}\omega(\alpha)\, d\alpha \leq \int\limits_0^\infty \alpha^{p'-3} \left(\int\limits_{\{0<|f|<\alpha^{-1/(p-1)}\}} |f|^2 d\mathbf{x} \right) d\alpha$$
$$= \int\limits_{\{0<|f|<\infty\}} |f|^2 \left(\int\limits_0^{|f|^{1-p}} \alpha^{p'-3} d\alpha \right) d\mathbf{x}$$
$$= \frac{1}{p'-2} \int\limits_{\{0<|f|<\infty\}} |f|^2 |f|^{(1-p)(p'-2)} d\mathbf{x} \quad \text{since } p' > 2$$
$$= \frac{1}{p'-2} \int\limits_{\mathbf{R^n}} |f|^p d\mathbf{x} = \frac{1}{p'-2}.$$

Collecting estimates, we obtain

$$\|\widehat{f}\|_{p'} \leq \left(\frac{p'}{p'-2}\right)^{\frac{1}{p'}} = \left(\frac{p}{2-p}\right)^{\frac{1}{p'}},$$

which proves (13.72) with $C = \left(\frac{p}{2-p}\right)^{1/p'}$ and $1 < p < 2$.

The value obtained here for the constant C is unbounded as p tends to 2. In fact, when $p = 2$, the proof fails to produce a finite right-hand side in (13.72) even though we know by Plancherel's theorem that $\|\widehat{f}\|_2 = (2\pi)^{-n/2}\|f\|_2$. This nonoptimality in the proof is due to the use of estimates of distribution functions. For example, while Plancherel's theorem is a sharp estimate, its application in the proof of Theorem 13.71 is preceded by using Tchebyshev's inequality in order to bound the measure of $\{|\widehat{h}| > \alpha/2\}$. In general, the Marcinkiewicz method is better suited to situations when only weak type distribution function estimates are known at the endpoint p values. A different interpolation technique, not treated in this text, known as the Riesz convexity theorem or the Riesz–Thorin theorem, and based on strong type norm estimates at the endpoints and complex-variable methods, produces a smaller value of the constant C in (13.72). This technique yields a value of C that is bounded for $1 < p < 2$. In fact, the best possible value of C, which depends on n and p, is determined in results due to K. I. Babenko and to W. Beckner.*

Exercises

1. Let $f \in L^1(\mathbf{R^n})$. Show that if f is real-valued and even, then $\widehat{f}$ is also real-valued and even. If f is real-valued and odd, show that $\widehat{f}$ is odd and purely imaginary.
Find a complex-valued f such that $\widehat{f}(\mathbf{x})$ is real-valued and finite for all $\mathbf{x}$.

2. Let f be an even integrable function on $(-\infty, \infty)$. Show that

$$\widehat{f}(x) = \frac{1}{\pi}\int_0^\infty f(y)\cos(|x|y)\,dy.$$

* See Inequalities in Fourier analysis, *Annals of Mathematics* 102 (1972), 159–182, by W. Beckner.

3. Let $f(\mathbf{x}) \in L^1(\mathbf{R^n})$, $n \geq 2$, and suppose that f is a radial function: $f(\mathbf{x}) = g(|\mathbf{x}|)$ for all $\mathbf{x}$. Assume that the polar coordinate representation

$$\int_{\mathbf{R^n}} f(\mathbf{x})\, d\mathbf{x} = \int_0^\infty \left(\int_{|\mathbf{x}'|=1} f(r\mathbf{x}')\, d\mathbf{x}' \right) r^{n-1} dr$$

is valid, where $d\mathbf{x}'$ denotes the differential element of surface area on the unit sphere $\{\mathbf{x}' \in \mathbf{R^n} : |\mathbf{x}'| = 1\}$, and $\mathbf{x} = r\mathbf{x}'$, $r = |\mathbf{x}|$ if $|\mathbf{x}| \neq 0$. Show that

$$\widehat{f}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\mathbf{x}|^{-\frac{n-2}{2}} \int_0^\infty g(r)\, r^{\frac{n}{2}} J_{\frac{n-2}{2}}(r|\mathbf{x}|)\, dr,$$

where $J_{(n-2)/2}(t)$, $t > 0$, is the *Bessel function* of order $(n-2)/2$, which may be defined as follows. For $t \geq 0$ and $\mathbf{x}'$ with $|\mathbf{x}'| = 1$, show that the function of $(t, \mathbf{x}')$ defined by $\int_{|\mathbf{y}'|=1} e^{-it\mathbf{x}'\cdot\mathbf{y}'} d\mathbf{y}'$ is independent of $\mathbf{x}'$, and then set

$$\int_{|\mathbf{y}'|=1} e^{-it\mathbf{x}'\cdot\mathbf{y}'} d\mathbf{y}' = (2\pi)^{\frac{n}{2}} t^{-\frac{n-2}{2}} J_{\frac{n-2}{2}}(t), \quad t > 0.$$

Note in particular that the formula above for $\widehat{f}$ shows that $\widehat{f}$ is a radial function if f is radial.

4. Verify properties (13.5) and (13.6) of the Fourier transform.

5. Let L be a nonsingular linear transformation of $\mathbf{R^n}$ and define $(Lf)(\mathbf{x}) = f(L\mathbf{x})$. If $f \in L^1(\mathbf{R^n})$, show that

$$\widehat{Lf}(\mathbf{x}) = |\det L|^{-1} \widehat{f}\big((L^{-1})^*\mathbf{x}\big), \quad \mathbf{x} \in \mathbf{R^n},$$

where $(L^{-1})^*$ $(= (L^*)^{-1})$ is the adjoint (or transpose) of L^{-1}. Equivalently, $(Lf)\widehat{\ } = |\det L|^{-1} \widetilde{L}\, \widehat{f}$ where $\widetilde{L} = (L^{-1})^*$.

6. (a) Verify (13.12). Also, given a function $f \in L^1(\mathbf{R^n})$, find a function $f(\mathbf{x}, t)$ defined on $\mathbf{R^{n+1}_+}$ such that

$$\sum_{k=1}^n \frac{\partial^2}{\partial x_k^2} f(\mathbf{x}, t) = \frac{\partial}{\partial t} f(\mathbf{x}, t) \quad \text{in } \mathbf{R^{n+1}_+}$$

and

$$\lim_{t \to 0} f(\mathbf{x}, t) = f(\mathbf{x}) \quad \text{at every Lebesgue point } \mathbf{x} \text{ of } f.$$

Such a function $f(\mathbf{x}, t)$ is said to solve the *Dirichlet problem for the heat equation* in $\mathbf{R}^{\mathbf{n}+\mathbf{1}}_+$. (Consider dilations of the Gauss–Weierstrass integral $Wf(\mathbf{x}, t)$.)

(b) Let $Pf(\mathbf{x}, \varepsilon)$ be the Poisson integral of f defined by (13.27) for $(\mathbf{x}, \varepsilon) \in \mathbf{R}^{\mathbf{n}+\mathbf{1}}_+$. If $f \in L^1(\mathbf{R}^{\mathbf{n}})$, show that Pf is harmonic in $\mathbf{R}^{\mathbf{n}+\mathbf{1}}_+$, that is, Pf satisfies Laplace's equation

$$\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2} + \frac{\partial^2}{\partial \varepsilon^2}\right) Pf(\mathbf{x}, \varepsilon) = 0 \text{ if } \mathbf{x} \in \mathbf{R}^{\mathbf{n}+\mathbf{1}}_+,\ \varepsilon > 0.$$

Hence, since $\lim_{\varepsilon \to 0} Pf(\mathbf{x}, \varepsilon) = f(\mathbf{x})$ at every Lebesgue point $\mathbf{x}$ of f, Pf solves the Dirichlet problem for Laplace's equation in the half-space $\mathbf{R}^{\mathbf{n}+\mathbf{1}}_+$.

7. Let $f \in L^1(\mathbf{R}^{\mathbf{n}})$, $P(\mathbf{x}, \varepsilon) = P_\varepsilon(\mathbf{x})$ be the Poisson kernel for $\mathbf{R}^{\mathbf{n}+\mathbf{1}}_+$, and $Pf(\mathbf{x}, \varepsilon)$ be the Poisson integral of f. Show that for all $\mathbf{x} \in \mathbf{R}^{\mathbf{n}}$ and all $\varepsilon, \delta > 0$,

$$\big(P_\varepsilon * P_\delta\big)(\mathbf{x}) = P_{\varepsilon+\delta}(\mathbf{x}) \text{ and } \big(Pf(\cdot, \varepsilon) * P_\delta\big)(\mathbf{x}) = Pf(\mathbf{x}, \varepsilon + \delta).$$

Prove analogues for $W(\mathbf{x}, t)$ and $Wf(\mathbf{x}, t)$.

8. Prove Theorem 13.40. Note in particular that if $\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$ denotes the *Laplace differential operator* (or *Laplacian*) in $\mathbf{R}^{\mathbf{n}}$, then

$$\widehat{\Delta f}(\mathbf{x}) = -|\mathbf{x}|^2 \widehat{f}(\mathbf{x}) \text{ if } f \in \mathscr{S}.$$

(Use Theorem 13.31 and the fact that if $f \in \mathscr{S}$ then $p(\mathbf{x})f(\mathbf{x}) \in \mathscr{S}$ for every polynomial $p(\mathbf{x})$.)

9. Let $f \in L^1(\mathbf{R}^{\mathbf{n}})$. If $\widehat{f} \in \mathscr{S}$, show that there is a function $F \in \mathscr{S}$ such that $f = F$ in the Lebesgue set of f.

10. If $f, g \in \mathscr{S}$, prove that $f * g \in \mathscr{S}$.

11. Suppose that $f \in L^1(\mathbf{R}^{\mathbf{n}})$ and f has a partial derivative in the L^1 sense with respect to x_k for every $k = 1, \ldots, n$. Prove that $\widehat{f}(\mathbf{x}) = o(|\mathbf{x}|^{-1})$ as $|\mathbf{x}| \to \infty$. Compare the estimate preceding Corollary 13.38 with $N = 1$ there.

12. (a) Show that the principal value Fourier transform in $\mathbf{R}^{\mathbf{1}}$ of $1/|x|$ is identically infinite.

(b) Show that the principal value Fourier transform in $\mathbf{R}^{\mathbf{1}}$ of $1/|x|^{1-\alpha}$, $0 < \alpha < 1$, equals $c_\alpha |x|^{-\alpha}$ if $x \neq 0$. (We will see in Chapter 14 that the function $1/|x|^{1-\alpha}$, $0 < \alpha < 1$, is the kernel of the one-dimensional fractional integral operator of order α.)

13. Let $1 \le p \le \infty$, $1/p + 1/p' = 1$, $f \in L^p(\mathbf{R}^{\mathbf{n}})$, and $g \in L^{p'}(\mathbf{R}^{\mathbf{n}})$. Show that $f * g$ is uniformly continuous and bounded on $\mathbf{R}^{\mathbf{n}}$.

14. Part of Plancherel's theorem states that the range of the map $\mathscr{F} : f \to \widehat{f}$, $f \in L^2(\mathbf{R^n})$, is all of $L^2(\mathbf{R^n})$, that is, that $\mathscr{F}$ is surjective. Show that this can also be proved as follows without using Corollary 13.38. The range space $\mathscr{F}(L^2(\mathbf{R^n}))$ is a closed subset of $L^2(\mathbf{R^n})$. If it were a proper subset, there would exist (by using an orthogonal basis in $L^2(\mathbf{R^n})$) a nontrivial $g \in L^2(\mathbf{R^n})$ such that $\int_{\mathbf{R^n}} g\widehat{f}\, d\mathbf{x} = 0$ for all $f \in L^2(\mathbf{R^n})$. Now use (13.53) to deduce a contradiction.

15. Verify Theorem 13.55.

16. Let $f \in L^2(-\infty, \infty)$. Show that the singly truncated Hilbert transform

$$H_\varepsilon f(x) = \frac{1}{\pi} \int\limits_{|y|>\varepsilon} f(x-y)\,\frac{1}{y}\,dy, \quad \varepsilon > 0,\ x \in (-\infty, \infty),$$

is finite and continuous at every x and that $\|H_\varepsilon f - Hf\|_2 \to 0$ as $\varepsilon \to 0$. Show also that there is a constant c independent of f and ε such that $\|H_\varepsilon f\|_2 \le c\|f\|_2$.

17. Given a real-valued function $f \in L^2(\mathbf{R^1})$ (or more generally in $L^p(\mathbf{R^1})$ for some p, $1 \le p < \infty$), consider the Cauchy integral $Cf(z)$, $z = x + iy$, defined in the upper half-space by

$$Cf(z) = \frac{1}{i\pi} \int\limits_{-\infty}^{\infty} f(t)\,\frac{1}{t-z}\,dt, \quad z = x + iy,\ y > 0.$$

(a) Show that for every such z, $Cf(z)$ converges absolutely, has real part equal to the Poisson integral of f, that is, equal to

$$f(x,y) = (f * P_y)(x) = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} f(t)\,\frac{y}{(x-t)^2 + y^2}\,dt,$$

and has imaginary part

$$\widetilde{f}(x,y) = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} f(t)\,\frac{x-t}{(x-t)^2 + y^2}\,dt.$$

The latter expression $\widetilde{f}(x,y)$ is called the *conjugate Poisson integral of f in the upper half-space* R^2_+. Recall that the numerical factor $1/\pi$ is needed in order to guarantee that $\int_{-\infty}^{\infty} P_y(x)\,dx = 1$. In this way, the factor $1/\pi$ is also natural in the definition of $\widetilde{f}(x,y)$ even though its kernel $\widetilde{P}(x) = (1/\pi)x/(x^2+1)$ is not integrable at infinity.

(b) Show that at every Lebesgue point x of f, the conjugate Poisson integral $\widetilde{f}(x,y)$ is equiconvergent as $y \to 0+$ with the singly truncated

Hilbert transform $H_y f(x)$ defined in Exercise 16, that is, $\widetilde{f}(x,y) - H_y f(x) \to 0$ as $y \to 0+$ if x is a Lebesgue point of f. Compare Theorem 12.51 (ii). (Define

$$k(x) = \widetilde{P}(x) - \frac{1}{\pi}\frac{\chi_{\{|x|>1\}}(x)}{x},$$

where $\widetilde{P}(x)$ is as in part (a), and note that $k(x)$ is bounded with $k(x) = O(|x|^{-3})$ as $|x| \to \infty$. Also, $\int_{-\infty}^{\infty} k(x)\,dx = 0$. Now compute the dilated kernels $k_y(x) = y^{-1}k(x/y)$, $y > 0$, and use the ideas of Theorem 9.13 and Exercise 11 of Chapter 9.)

18. Let $f \in \mathscr{S}(-\infty,\infty)$ and let Hf denote the Hilbert transform of f. If $\int_{-\infty}^{\infty} f(x)\,dx \neq 0$, show that there is no function $g \in \mathscr{S}(-\infty,\infty)$ such that $Hf = g$ a.e.

19. If $f, g \in L^2(-\infty,\infty)$ and Hf, Hg denote their Hilbert transforms, show that

$$\int_{-\infty}^{\infty} (Hf)\,(Hg)\,dx = \int_{-\infty}^{\infty} f\,g\,dx \text{ and } \int_{-\infty}^{\infty} f\,(Hg)\,dx = -\int_{-\infty}^{\infty} (Hf)\,g\,dx.$$

20. Let $-\infty < a < b < \infty$ and $\chi_{(a,b)}(x)$ denote the characteristic function of the interval (a,b). Prove that the truncated Hilbert transform $H_{\varepsilon,\omega}\chi_{(a,b)}$ satisfies

$$\lim_{\substack{\varepsilon\to 0+\\ \omega\to\infty}} \big(H_{\varepsilon,\omega}\chi_{(a,b)}\big)(x) = \frac{1}{\pi}\log\left|\frac{a-x}{b-x}\right|, \quad x \neq a, b,$$

where the convergence is in the pointwise sense. Deduce that the Hilbert transform of a function that is bounded and has compact support may not be bounded or have compact support.

21. Let $-\infty < a < b < \infty$ and $\chi_{(a,b)}(x)$ denote the characteristic function of the interval (a,b). Set $c = (a+b)/2$ and $\ell = (b-a)/2$. Show that

$$\widehat{\chi_{(a,b)}}(x) = \frac{1}{\pi}\frac{\sin \ell x}{x}e^{-icx}, \quad x \in (-\infty,\infty),$$

where the right side should be interpreted as ℓ/π when $x = 0$. Note that $\widehat{\chi_{(a,b)}}$ is not integrable (at infinity). Also, given $h > 0$, sketch the graph of the convolution $\frac{1}{2h}\chi_{(-h,h)} * \chi_{(-h,h)}$ and show that the Fourier transform of this function is the integrable function $(\sin^2 hx)/(\pi h x^2)$ (cf. Exercise 5(b) of Chapter 12).

These examples further illustrate the maxim first observed on p. 387 in Section 13.1 that the smoother a compactly supported function is, the smaller its Fourier transform is at infinity.

22. Let $1 < p < 2$ and $f \in L^p(\mathbf{R^n})$. Show that the functions

$$\widehat{f_k}(\mathbf{x}) = \frac{1}{(2\pi)^n} \int\limits_{|\mathbf{y}|<k} f(\mathbf{y}) e^{-i\mathbf{x}\cdot\mathbf{y}} d\mathbf{y}, \quad k > 0,$$

converge in $L^{p'}(\mathbf{R^n})$ to $\widehat{f}$ as $k \to \infty$, $1/p + 1/p' = 1$.

23. Let $1 < p < 2, f \in L^1(\mathbf{R^n})$ and $g \in L^p(\mathbf{R^n})$. Show that $\widehat{f * g} = (2\pi)^n \widehat{f}\,\widehat{g}$ a.e.

24. Let $f \in L^p(0, 2\pi)$, $1 < p < 2$, be periodic with period 2π, and let $c_k = (1/2\pi) \int_0^{2\pi} f(t) e^{-ikt} dt$, $k = 0, \pm 1, \ldots$, be its Fourier coefficients. Prove that there is a constant C independent of f such that

$$\left(\sum_k |c_k|^{p'} \right)^{1/p'} \le C \left(\int\limits_0^{2\pi} |f|^p dt \right)^{1/p}, \quad \frac{1}{p} + \frac{1}{p'} = 1.$$

This is the periodic Hausdorff–Young theorem. (Write

$$\sum_{k=-\infty}^{\infty} |c_k|^{p'} = \sum_{j=-\infty}^{\infty} \sum_{\substack{k \\ 2^{j-1} < |c_k| \le 2^j}} |c_k|^{p'} \approx \sum_{j=-\infty}^{\infty} 2^{jp'} \sum_{\substack{k \\ 2^{j-1} < |c_k| \le 2^j}} 1$$

in analogy with the *first* equality in Theorem 5.51. After summing by parts, argue similarly to the proof of Theorem 13.71 with 2^{j-1} playing the role of α. This method will not produce the optimal value of the constant C.)

25. Let $f \in L^p(-\infty, \infty)$, $1 \le p \le 2$. Derive the following pointwise formula for $\widehat{f}$:

$$\widehat{f}(x) = \frac{d}{dx} \left(\frac{1}{2\pi} \int\limits_{-\infty}^{\infty} f(y) \frac{e^{-ixy} - 1}{-iy} dy \right) \text{ a.e.}$$

$\Big($Define $F(x) = \int_0^x \widehat{f}$ and $F_N(x) = \int_0^x \widehat{f_N}$ where $f_N = f\chi_{(-N,N)}$, $N > 0$. By Hölder's inequality, $|F(x) - F_N(x)| \le |x|^{1/p} ||\widehat{f} - \widehat{f_N}||_{p'} \to 0$ as $N \to \infty$. Then $\widehat{f}(x) = F'(x) = (d/dx) \lim_{N\to\infty} F_N(x)$ a.e. Also,

$$F_N(x) = \frac{1}{2\pi} \int\limits_{-N}^{N} f(y) \frac{e^{-ixy} - 1}{-iy} dy.\Big)$$

26. Let μ be a finite Borel measure on $\mathbf{R^n}$. Define the Fourier transform of μ by

$$\widehat{\mu}(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbf{R^n}} e^{-i\mathbf{x}\cdot\mathbf{y}} d\mu(\mathbf{y}), \quad \mathbf{x} \in \mathbf{R^n}.$$

(a) Show that $\widehat{\mu}$ is bounded and continuous on $\mathbf{R^n}$ and satisfies $\overline{\widehat{\mu}(-\mathbf{x})} = \widehat{\mu}(\mathbf{x})$ for all $\mathbf{x}$.

(b) If $\mathbf{x}_0$ is a point in $\mathbf{R^n}$ and $\{E\}$ is the class of Borel sets in $\mathbf{R^n}$, define the *delta measure* $\delta_{\mathbf{x}_0}$ at $\mathbf{x}_0$ by $\delta_{\mathbf{x}_0}(E) = 1$ if $\mathbf{x}_0 \in E$ and $\delta_{\mathbf{x}_0}(E) = 0$ if $\mathbf{x}_0 \notin E$. Show that

$$\widehat{\delta_{\mathbf{x}_0}}(\mathbf{x}) = \frac{1}{(2\pi)^n} e^{-i\mathbf{x}_0\cdot\mathbf{x}}.$$

Note that $\widehat{\delta_{\mathbf{0}}}(\mathbf{x})$ is identically equal to the constant $1/(2\pi)^n$.

(c) Let $m = 1, 2, \ldots$, $\{\mathbf{x}_j\}_{j=1}^m \subset \mathbf{R^n}$, and $\{z_j\}_{j=1}^m$ be complex numbers. Show that

$$\sum_{j=1}^m \sum_{k=1}^m \widehat{\mu}(\mathbf{x}_j - \mathbf{x}_k) z_j \overline{z_k} = (2\pi)^n \int_{\mathbf{R^n}} \left| \sum_{j=1}^m z_j \widehat{\delta_{\mathbf{x}_j}}(\mathbf{y}) \right|^2 d\mu(\mathbf{y}).$$

In particular, the sum on the left side is nonnegative. (Conversely, it is known that any bounded continuous complex-valued function g on $\mathbf{R^n}$ that satisfies $\overline{g(-\mathbf{x})} = g(\mathbf{x})$ for all $\mathbf{x}$ is the Fourier transform of a finite Borel measure provided g is nonnegative definite in the following sense:

$$\sum_{j=1}^m \sum_{k=1}^m g(\mathbf{x}_j - \mathbf{x}_k) z_j \overline{z_k} \geq 0$$

for all such m, $\{\mathbf{x}_j\}_{j=1}^m$ and $\{z_j\}_{j=1}^m$. See S. Bochner, *Lectures on Fourier Integrals*, Annals of Math. Studies 42, Princeton Univ. Press, 1959.)

27. (a) If $f \in L^1((0,\infty); dt/t)$, that is, if $\int_0^\infty |f(t)|\, dt/t < \infty$, define the *Mellin transform* $m(f)$ of f by

$$m(f)(x) = \frac{1}{2\pi} \int_0^\infty f(t) t^{-ix} \frac{dt}{t}, \quad -\infty < x < \infty.$$

If $f, g \in L^1((0,\infty); dt/t)$, define an analogue $f\#g$ of ordinary convolution by

$$(f\#g)(s) = \int_0^\infty f(st^{-1})g(t)\,\frac{dt}{t}, \quad 0 < s < \infty.$$

Show that $f\#g \in L^1((0,\infty); dt/t)$ and

$$m(f\#g)(x) = m(f)(x)\,m(g)(x), \quad x \in \mathbf{R^1}.$$

(b) For any f defined on $(0,\infty)$, let $(Ef)(x) = f(e^x)$, $x \in \mathbf{R^1}$. If $f \in L^1((0,\infty); dt/t)$, show that $Ef \in L^1(\mathbf{R^1})$ and $m(f)(x) = \widehat{Ef}(x)$. (Exercise 11 of Chapter 7 may be helpful.) If $f \in L^2((0,\infty); dt/t)$, explain why $Ef \in L^2(\mathbf{R^1})$, and define $m(f)$ by $m(f)(x) = \widehat{Ef}(x)$, $x \in \mathbf{R^1}$. Show that

$$\frac{1}{2\pi}\int_{e^{-k}}^{e^k} f(t)t^{-ix}\frac{dt}{t} \to m(f)(x) \text{ in } L^2(\mathbf{R^1}) \text{ norm as } k \to \infty, \text{ and}$$

$$\int_{-\infty}^{\infty} |m(f)(x)|^2 dx = \frac{1}{2\pi}\int_0^\infty |f(t)|^2\frac{dt}{t}.$$

28. Show that $f \equiv 0$ if $f, \widehat{f} \in C_0^\infty(\mathbf{R^n})$. (In case $n = 1$ and f is supported in the open interval $(-\pi,\pi)$, the Fourier series $S[f]$ of f in $(-\pi,\pi)$ converges everywhere in $(-\pi,\pi)$ to f. If $\widehat{f}$ has compact support, $S[f]$ reduces to a trigonometric polynomial and therefore has at most a finite number of roots (mod 2π); see Section 12.1, p. 307. An alternate proof based on the identity theorem for analytic functions is possible.)

29. There is an analogue for smooth truncations of the limit inversion fact in (13.57). For $k = 1, 2, \ldots$, let $\chi_k(\mathbf{x}) \in C_0^\infty(\mathbf{R^n})$ with $\chi_k(\mathbf{x}) = 1$ if $|\mathbf{x}| \le k$, $\chi_k(\mathbf{x}) = 0$ if $|\mathbf{x}| \ge 2k$, and $0 \le \chi_k \le 1$. The operators

$$\Psi_k f(\mathbf{x}) = \int_{\mathbf{R^n}} \chi_k(\mathbf{y})\widehat{f}(\mathbf{y})e^{i\mathbf{x}\cdot\mathbf{y}}\,d\mathbf{y}, \quad f \in L^2(\mathbf{R^n}),$$

are often called *pseudodifferential cutoff operators*. Show that $\|\Psi_k f\|_2 \le \|f\|_2$ and $\|\Psi_k f - f\|_2 \to 0$ as $k \to \infty$. Show also that $\Psi_k f = \psi_k * f$, where $(2\pi)^n\widehat{\psi_k} = \chi_k$. Deduce that $\Psi_k f$ is infinitely differentiable if $f \in L^2(\mathbf{R^n})$, and that if $f \in \mathscr{S}$, then $D^\alpha\Psi_k f \to D^\alpha f$ in $L^2(\mathbf{R^n})$ for every multi-index α.

14

Fractional Integration

In this chapter, we will study an important class of convolution operators known as fractional integral operators. The behavior of these operators on functions in the L^p spaces, including L^p for various "endpoint" values of p, is of particular interest. In addition, a number of closely related topics dealing with how much a function differs from its integral average are treated. Results of this second type are generally called mean oscillation estimates. The classes of Hölder continuous functions as well as the class of functions of bounded mean oscillation arise naturally in this context.

In the next chapter, the norm estimates for fractional integrals that are derived in this chapter will be used in order to obtain Poincaré–Sobolev inequalities.

Many of the results and methods in Chapters 14 and 15 can be adapted to geometric settings that are more varied than the usual Euclidean one, as well as to measures more general than Lebesgue measure.

14.1 Subrepresentation Formulas and Fractional Integrals

Let f be a real-valued measurable function on $\mathbf{R^n}$, $n \geq 1$, and let $0 < \alpha < n$. The *fractional integral* or *Riesz potential of f of order α* is defined by

$$I_\alpha f(\mathbf{x}) = \int_{\mathbf{R^n}} f(\mathbf{y}) \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}}\, d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n}, \tag{14.1}$$

provided the integral exists. By allowing f to vary, the mapping defined by $I_\alpha : f \to I_\alpha f$, that is, the convolution operator with kernel $|\mathbf{x}|^{\alpha-n}$, is called the *fractional integral operator of order α*. The main mapping properties of I_α, including answers to the questions of existence, finiteness, and measurability of $I_\alpha f$ for various classes of measurable f, will be studied in Section 14.3. For now, we simply note that if f is nonnegative and measurable on $\mathbf{R^n}$, then since $|\mathbf{x}|^{\alpha-n}$ is also nonnegative and measurable, Corollary 6.16 guarantees existence and measurability (but not finiteness) of $I_\alpha f$ on $\mathbf{R^n}$.

The case $\alpha = 1$ will play a special role, although the theory for general α, $0 < \alpha < n$, will be developed. As a motivation for studying fractional integrals, we begin by deriving a basic subrepresentation formula for any

sufficiently smooth function f in terms of the Riesz potential of order $\alpha = 1$ of the first partial derivatives of f. The formula may be regarded as a weak substitute in $\mathbf{R^n}$, $n > 1$, for the fundamental theorem of calculus in $\mathbf{R^1}$, that is, as an n-dimensional version of the formula in Theorem 7.29 showing that an absolutely continuous function f defined on an interval $[a, b] \subset (-\infty, \infty)$ satisfies

$$f(x) - f(y) = \int_y^x f'(t)\,dt, \quad x, y \in [a, b].$$

In particular, after taking absolute values and integrating in y, this formula yields the estimate

$$\frac{1}{b-a}\int_a^b |f(x) - f(y)|\,dy \le \int_a^b |f'|, \quad x \in [a, b].$$

Moreover, since

$$f(x) - \frac{1}{b-a}\int_a^b f(y)\,dy = \frac{1}{b-a}\int_a^b [f(x) - f(y)]\,dy,$$

we also obtain the pointwise inequality

$$\left| f(x) - \frac{1}{b-a}\int_a^b f(y)\,dy \right| \le \int_a^b |f'|, \quad x \in [a, b].$$

In order to derive analogues of these inequalities in higher dimensions, we will initially assume that f is a function defined in an open ball $B \subset \mathbf{R^n}$ and that f belongs to the class $C^1(B)$ of functions with continuous first partial derivatives in B. The C^1 restriction will be considerably relaxed in Theorem 15.16. The gradient vector of such an f will be denoted

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right);$$

its magnitude is

$$|\nabla f| = \left(\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \right)^{1/2}.$$

Note that since B is open, neither f nor $|\nabla f|$ may belong to $L^1(B)$ if $f \in C^1(B)$.

Theorem 14.2 (Subrepresentation Formula) *Let B be an open ball in $\mathbf{R^n}$ and $f \in C^1(B)$. Then*

$$\frac{1}{|B|}\int_B |f(\mathbf{x}) - f(\mathbf{y})|\,d\mathbf{y} \le c_n \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}}d\mathbf{y}, \quad \mathbf{x} \in B, \tag{14.3}$$

where c_n is a constant that depends only on n. If in addition $f \in L^1(B)$, then

$$\left|f(\mathbf{x}) - f_B\right| \le c_n \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}}d\mathbf{y}, \quad \mathbf{x} \in B, \tag{14.4}$$

where $f_B = \frac{1}{|B|}\int_B f(\mathbf{y})\,d\mathbf{y}$ is the integral average of f on B.

Before proving the theorem, we remark that the integrals on the right sides of (14.3) and (14.4) are essentially $I_1(|\nabla f|)(x)$, except that their domain of integration is restricted to B. In fact, f is assumed to be defined only on B. If g is any function defined on B, but not necessarily outside B, we can extend g to all of $\mathbf{R^n}$ by defining it to be 0 outside B, and we will then denote the extension by $g\chi_B$:

$$(g\chi_B)(\mathbf{x}) = \begin{cases} g(\mathbf{x}) \text{ if } \mathbf{x} \in B \\ 0 \text{ if } \mathbf{x} \in \mathbf{R^n} - B. \end{cases}$$

Sometimes, we will also write $g(\mathbf{x})\chi_B(\mathbf{x})$ instead of $(g\chi_B)(\mathbf{x})$. With this notation, the integrals on the right sides of (14.3) and (14.4) can be written simply as $I_1(|\nabla f|\chi_B)(\mathbf{x})$.

Proof. We will prove Theorem 14.2 only in case $n \ge 2$, leaving it to the reader to adapt our earlier comments about an absolutely continuous function on a closed interval $[a,b] \subset \mathbf{R^1}$ to a continuously differentiable f on an open interval (a,b), $b-a < \infty$.

Fix an open ball B and an $f \in C^1(B)$. It suffices to prove (14.3) since (14.4) follows immediately from (14.3) and the simple estimate

$$|f(\mathbf{x}) - f_B| = \left|\frac{1}{|B|}\int_B [f(\mathbf{x}) - f(\mathbf{y})]\,d\mathbf{y}\right| \le \frac{1}{|B|}\int_B |f(\mathbf{x}) - f(\mathbf{y})|\,d\mathbf{y},$$

assuming of course that $f \in L^1(B)$ in order to guarantee that f_B is well-defined. To prove (14.3), fix $\mathbf{x} \in B$, and for any $\mathbf{y} \in B$, write

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \frac{d}{dt}\Big[f(t\mathbf{y} + (1-t)\mathbf{x})\Big]\,dt$$
$$= \int_0^1 (\nabla f)(t\mathbf{y} + (1-t)\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})\,dt,$$

where use of the chain rule is justified since $f \in C^1(B)$ and the line segment $\{t\mathbf{y} + (1-t)\mathbf{x} : 0 \le t \le 1\}$ lies in B. Hence, by Tonelli's theorem,

$$\int_B |f(\mathbf{y}) - f(\mathbf{x})|\,d\mathbf{y} \le \int_0^1 \left(\int_B |(\nabla f)(t\mathbf{y} + (1-t)\mathbf{x})|\,|\mathbf{y} - \mathbf{x}|\,d\mathbf{y}\right) dt.$$

In the inner integral on the right side, we make the affine change of variables $\mathbf{w} = t\mathbf{y} + (1-t)\mathbf{x} = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$ in $\mathbf{y}$, with t and $\mathbf{x}$ fixed, and note that $\mathbf{w} \in B$. Also, $|\mathbf{w} - \mathbf{x}| = |\mathbf{y} - \mathbf{x}|\,t$, and consequently $|\mathbf{w} - \mathbf{x}| \le t\,\mathrm{diam}(B)$. By using (6.13) and Exercise 20 of Chapter 5, it follows that the right side of the preceding equation is at most

$$\int_0^1 \int_B \chi_{\{\mathbf{w}:|\mathbf{w}-\mathbf{x}|\le t\,\mathrm{diam}(B)\}}(\mathbf{w})\,|\nabla f(\mathbf{w})|\,\frac{|\mathbf{w} - \mathbf{x}|}{t}\,\frac{d\mathbf{w}}{t^n}dt$$
$$\le \int_B |\nabla f(\mathbf{w})|\,|\mathbf{w} - \mathbf{x}| \left(\int_{|\mathbf{w}-\mathbf{x}|/\mathrm{diam}(B)}^{\infty} \frac{1}{t^{n+1}}\,dt\right) d\mathbf{w}$$
$$= \frac{\mathrm{diam}(B)^n}{n} \int_B |\nabla f(\mathbf{w})|\,\frac{1}{|\mathbf{w} - \mathbf{x}|^{n-1}}d\mathbf{w}.$$

However, $\mathrm{diam}(B)^n$ is a fixed multiple of $|B|$; in fact, by using polar coordinates $\mathbf{y} = r\mathbf{y}'$ with $r = |\mathbf{y}|$ and $|\mathbf{y}'| = 1$, and denoting the radius of B by $r(B)$ and the surface area of the unit ball in $\mathbf{R^n}$ by w_n, we have

$$|B| = \int_{|\mathbf{y}|<r(B)} d\mathbf{y} = \int_0^{r(B)} r^{n-1}dr \int_{|\mathbf{y}'|=1} d\mathbf{y}' = \frac{r(B)^n}{n} w_n = \frac{w_n}{n}\left[\frac{\mathrm{diam}(B)}{2}\right]^n.$$

Combining estimates gives

$$\int_B |f(\mathbf{y}) - f(\mathbf{x})|\,d\mathbf{y} \le \frac{2^n}{w_n}\,|B| \int_B |\nabla f(\mathbf{w})|\,\frac{1}{|\mathbf{w} - \mathbf{x}|^{n-1}}d\mathbf{w}.$$

This proves (14.3) with $c_n = 2^n w_n^{-1}$, and the proof of Theorem 14.2 is complete.

If f satisfies the stronger condition that $f \in C^1(\overline{B})$, where $\overline{B}$ is the closure of B, then (14.3) and (14.4) hold for all $\mathbf{x} \in \overline{B}$ (see Exercise 1). Also, the integral on the right side of (14.3) is then finite for $\mathbf{x} \in \overline{B}$; in fact, it is bounded by

$$\left(\max_{\overline{B}} |\nabla f|\right) \int\limits_{|\mathbf{x}-\mathbf{y}| \le \mathrm{diam}(B)} \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y} = \left(\max_{\overline{B}} |\nabla f|\right) w_n \,\mathrm{diam}(B).$$

Here, we have used the formula

$$\int\limits_{|\mathbf{x}-\mathbf{y}| \le \mathrm{diam}(B)} \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y} = \int\limits_{|\mathbf{y}| \le \mathrm{diam}(B)} \frac{d\mathbf{y}}{|\mathbf{y}|^{n-1}} = w_n \,\mathrm{diam}(B).$$

Note that when $f \in C^1(\overline{B})$, we clearly also have that $f \in L^1(B)$.

The role played by balls in Theorem 14.2 can be played by some other types of sets, with similar proofs. See Exercise 4 for the case of the Cartesian product $B_1 \times B_2$ of two balls, and see Exercise 5 for general intervals in $\mathbf{R^n}$.

From now on, all balls B are assumed to be open. As in Chapter 1, we use the notation

$$B = B(\mathbf{x}; r) = \{\mathbf{y} : |\mathbf{x}-\mathbf{y}| < r\} \tag{14.5}$$

for the (open) ball with center $\mathbf{x}$ and radius r. The radius of a ball B will often be denoted by $r(B)$. As shown in the proof of Theorem 14.2, there is a constant c_n depending only on n such that $|B| = c_n r(B)^n$ for every ball B.

Next, we list a corollary of Theorem 14.2 that gives analogues of the subrepresentation formula (14.4) *without* the average f_B on the left side if f vanishes on appropriate subsets of B.

Corollary 14.6 *Suppose that B is an open ball in $\mathbf{R^n}$ and $f \in C^1(B)$.*

(i) *If $f = 0$ in a measurable set $E \subset B$ satisfying $|E| \ge \gamma\, |B|$ for some constant $\gamma > 0$, then*

$$|f(\mathbf{x})| \le \frac{c_n}{\gamma} \int\limits_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y}, \quad \mathbf{x} \in B,$$

where c_n is a constant that depends only on n.

(ii) *If f has compact support in B, then*

$$|f(\mathbf{x})| \le c_n \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n},$$

where c_n depends only on n.

Proof. Let B, f, and E satisfy the hypothesis of part (i). If $\mathbf{x} \in B$, then since $f = 0$ in E, we have

$$\begin{aligned} |f(\mathbf{x})| &= \frac{1}{|E|} \int_E |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y} \\ &\le \frac{1}{\gamma |B|} \int_B |f(\mathbf{y}) - f(\mathbf{x})|\, d\mathbf{y} \le \frac{c_n}{\gamma} \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y} \end{aligned}$$

by (14.3), with the same constant c_n as in (14.3), and the proof of (i) is complete.

Next, suppose that B is an open ball, $f \in C^1(B)$, and f has compact support in B. Extend f to $\mathbf{R^n}$ by setting $f = 0$ outside B. Then $f \in C^1(\mathbf{R^n})$. Let B^* be an open ball concentric with B such that $r(B^*) > r(B)$. By part (i) applied to B^*, and with E chosen to be $E = B^* - B$ and $\gamma = |B^* - B|/|B^*|$, we obtain

$$|f(\mathbf{x})| \le \frac{c_n}{\gamma} \int_{B^*} \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y} = \frac{c_n}{\gamma} I_1(|\nabla f| \chi_B)(\mathbf{x})$$

for all $\mathbf{x} \in B^*$, and so for all $\mathbf{x} \in \mathbf{R^n}$ since f is supported in B. The corollary follows by choosing B^* with $r(B^*) = 2r(B)$ since γ then depends only on n. In fact, by instead letting $r(B^*) \to \infty$ and observing that γ then tends to 1, we obtain $|f(\mathbf{x})| \le c_n I_1(|\nabla f| \chi_B)(\mathbf{x})$ with the same constant c_n as in Theorem 14.2.

Corollary 14.7 *If $f \in C^1(\mathbf{R^n})$ and there is a sequence $\{B_i\}_{i=1}^{\infty}$ of balls increasing to $\mathbf{R^n}$ such that $f_{B_i} \to 0$, then*

$$|f(\mathbf{x})| \le c_n \int_{\mathbf{R^n}} \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n},$$

where c_n depends only on n.

Note that the expression on the right side of the conclusion is $c_n I_1(|\nabla f|)(\mathbf{x})$.

Proof. Fix $f \in C^1(\mathbf{R^n})$ and a sequence of balls $B_i \nearrow \mathbf{R^n}$ with $f_{B_i} \to 0$. The conclusion of the corollary follows immediately by applying (14.4) to the balls B_i and using Fatou's lemma.

The assumption in Corollary 14.7 that there are balls $B_i \nearrow \mathbf{R^n}$ such that $f_{B_i} \to 0$ is satisfied by any sequence of balls increasing to $\mathbf{R^n}$ if either

(1) $f \in L^r(\mathbf{R^n})$ for some r with $1 \le r < \infty$ or
(2) $f \in L^r_{loc}(\mathbf{R^n})$ for some r with $1 \le r \le \infty$ and $\lim_{|\mathbf{x}|\to\infty} f(\mathbf{x}) = 0$.

For example, in case (1), by Hölder's inequality,

$$|f_B| \le \frac{1}{|B|}\int_B |f|\,d\mathbf{y} \le \left(\frac{1}{|B|}\int_B |f|^r\,d\mathbf{y}\right)^{1/r} \le |B|^{-1/r}\|f\|_{L^r(\mathbf{R^n})},$$

which tends to 0 as $B \nearrow \mathbf{R^n}$ since r is finite. The verification in case (2) is left to the reader.

Note that part (ii) of Corollary 14.6 is a special case of Corollary 14.7.

The significance of the subrepresentation formulas in Theorem 14.2 and Corollaries 14.6 and 14.7 will be more apparent after we study the behavior of L^q norms of $I_\alpha f$ when $f \in L^p$ in Section 14.3. For example, when $n > 1$, by combining (14.4) with norm estimates for $I_\alpha f$ in case $\alpha = 1$, we will be able to bound $L^q(B)$ norms of $f - f_B$ by $L^p(B)$ norms of $|\nabla f|$ for appropriate values of p and q. The inequalities obtained are called Poincaré–Sobolev estimates; they are derived in Chapter 15 under less restrictive smoothness assumptions on f than continuous differentiability. In case $n = 1$, Poincaré–Sobolev estimates can be derived directly from Theorem 7.29.

14.2 L^1, L^1 Poincaré Estimates, the Subrepresentation Formula, and Hölder Classes

In this section, we begin by considering the relationship between the inequality

$$\frac{1}{|B|}\int_B \left|f(\mathbf{x}) - f_B\right|\,d\mathbf{x} \le c\,r(B)\frac{1}{|B|}\int_B \left|\nabla f(\mathbf{x})\right|\,d\mathbf{x} \tag{14.8}$$

and the subrepresentation formula (14.4). Note that (14.4) is a pointwise estimate, while (14.8) is not. We call inequality (14.8) the L^1, L^1 *Poincaré estimate* for f and B. The "L^1, L^1" part of the terminology is due to the fact that the

exponents of both $\left|f(\mathbf{x}) - f_B\right|$ and $|\nabla f(\mathbf{x})|$ in (14.8) are 1. One of our main goals is to show that for a given f, the subrepresentation formula (14.4) is equivalent to the L^1, L^1 Poincaré estimate *provided* the ball B is allowed to vary. Another goal is to apply the ideas used to prove this equivalence in order to obtain pointwise estimates that characterize some other mean oscillation inequalities of the form

$$\frac{1}{|B|}\int_B \left|f(\mathbf{x}) - f_B\right| \, d\mathbf{x} \le a(B), \quad B \subset B_0, \tag{14.9}$$

where $a(B)$ is a nonnegative functional defined on balls.

Such functionals $a(B)$ may depend on f. For example, by choosing

$$a(B) = c\,\frac{r(B)}{|B|}\int_B |\nabla f| \, d\mathbf{x},$$

(14.9) becomes (14.8). In this section, we will consider only two types of functionals. The first one is

$$a(B) = c\,\frac{r(B)}{|B|}\int_B g \, d\mathbf{x}, \tag{14.10}$$

where g is fixed and nonnegative, for example, $g = |\nabla f|$. The other one is

$$a(B) = c\, r(B)^{\beta}, \quad 0 < \beta \le 1, \tag{14.11}$$

where β is fixed (independent of B). We will call (14.11) the *Hölder* β*-functional*, or the *Lipschitz* β*-functional*. Hölder functionals can depend on f only indirectly through the constants c and β. Theorems 14.12 and 14.25 give pointwise characterizations of those f that satisfy (14.9) for $a(B)$ as in (14.10) and (14.11), respectively. Similar ideas are used to prove both characterizations.

On the other hand, the important special case of (14.11) when $\beta = 0$ lies deeper and requires a different treatment. Then $a(B)$ is identically constant, and when B_0 is replaced by $\mathbf{R^n}$, condition (14.9) becomes

$$\frac{1}{|B|}\int_B \left|f - f_B\right| \, d\mathbf{x} \le C, \quad B \subset \mathbf{R^n}.$$

Such f are said to belong to the class $BMO(\mathbf{R^n})$ of functions of *bounded mean oscillation on* $\mathbf{R^n}$. They will be characterized in Section 14.5 in terms of the size of the distribution function of $\left|f - f_B\right|$ on B, rather than in terms of a pointwise condition.

We will continue to assume that all balls B are open and to use the notation $f_B = |B|^{-1}\int_B f \, d\mathbf{x}$ if $f \in L^1(\mathbf{B})$.

Theorem 14.12 *Let B_0 be a ball in $\mathbf{R}^n$, $f \in L^1(B_0)$, and g be a nonnegative measurable function on B_0. Then there is a constant C_1 such that*

$$\frac{1}{|B|}\int_B |f - f_B|\, d\mathbf{x} \le C_1 r(B)\frac{1}{|B|}\int_B g\, d\mathbf{x} \quad \textit{for all balls } B \subset B_0 \tag{14.13}$$

if and only if there is a constant C_2 such that

$$|f(\mathbf{x}) - f_B| \le C_2 \int_B \frac{g(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{n-1}} d\mathbf{y} \quad \textit{for a.e. } \mathbf{x} \in B \textit{ and all balls } B \subset B_0. \tag{14.14}$$

The constants C_1 and C_2 are equivalent in the sense that there is a positive constant c_n depending only on n such that $c_n^{-1}C_1 \le C_2 \le c_n C_1$.

Before proving the theorem, we note that (14.13) remains the same if f is changed arbitrarily in a set of measure 0. The possible exclusion of a set of measure 0 in (14.14) is then natural since (14.14) is a pointwise inequality for $|f - f_B|$. In Remark 14.19, we will see that if either (14.13) or (14.14) holds, then (14.14) holds for every point $\mathbf{x}$ of the Lebesgue set of f in B.

Proof of Theorem 14.12. To show that (14.14) implies (14.13), we first integrate (14.14) over B to obtain

$$\int_B |f(\mathbf{x}) - f_B|\, d\mathbf{x} \le C_2 \int_B g(\mathbf{y}) \left(\int_B \frac{d\mathbf{x}}{|\mathbf{x}-\mathbf{y}|^{n-1}} \right) d\mathbf{y}.$$

Since $|\mathbf{x} - \mathbf{y}| \le 2r(B)$ if $\mathbf{x}, \mathbf{y} \in B$, then

$$\int_B \frac{d\mathbf{x}}{|\mathbf{x}-\mathbf{y}|^{n-1}} \le \int_{|\mathbf{x}-\mathbf{y}|<2r(B)} \frac{d\mathbf{x}}{|\mathbf{x}-\mathbf{y}|^{n-1}} = w_n 2r(B).$$

Combining estimates, we immediately obtain (14.13) with $C_1 = 2w_n C_2$.

The proof of the converse is longer and based on adding (14.13) over an appropriate chain $\{B_k\}$ of balls B_k associated with each point $\mathbf{x} \in B_0$ that shrink regularly to $\mathbf{x}$ (in the sense of Section 7.2, p. 141). The chain associated with $\mathbf{x}$ is described in the next lemma. Its key properties are that if $\mathbf{x}_k$ and r_k denote the center and radius of B_k, then $r_k \approx |\mathbf{x} - \mathbf{x}_k| \to 0$ as $k \to \infty$, and every two successive balls B_k, B_{k+1} have substantial overlap uniformly in k, while the entire collection $\{B_k\}$ has bounded overlaps (see Section 10.5, p. 267) uniformly in $\mathbf{x}$. A more precise formulation is as follows.

Lemma 14.15 *Let B be an open ball in $\mathbf{R^n}$ and let $\mathbf{x} \in B$. Then there is a sequence $\{B_k\}_{k=1}^{\infty}$ of balls with the following properties:*

(i) $B_k \subset B$,

(ii) $B_k = B(\mathbf{x}_k; r_k)$ *with* $r_1 \geq \frac{1}{8} r(B)$ *and* $r_k = \frac{1}{2}|\mathbf{x} - \mathbf{x}_k| \to 0$ *as* $k \to \infty$,

(iii) $B_k \subset B(\mathbf{x}; 3r_k)$,

(iv) *If* $k < \ell$ *and* $B_k \cap B_\ell \neq \emptyset$, *then* $\ell = k+1$,

(v) $B_k \cap B_{k+1}$ *contains a ball* $\widetilde{B}_k$ *with* $|B_k|, \left|B_{k+1}\right| \leq c_n \left|\widetilde{B}_k\right|$.

Taking Lemma 14.15 temporarily for granted, let us finish the proof of Theorem 14.12. We must prove (14.14) assuming that (14.13) is true. Fix a ball B with $B \subset B_0$. For each $\mathbf{x} \in B$, let $\{B_k\}_{k=1}^{\infty}$ be a chain of subballs of B with the properties in Lemma 14.15. Then $B_k = B(\mathbf{x}_k; r_k) \subset B(\mathbf{x}; 3r_k)$ for each $k \geq 1$, and $\mathbf{x}_k \to \mathbf{x}$ and $r_k \to 0$ as $k \to \infty$. Since $|B(\mathbf{x}; 3r_k)| = 3^n |B_k|$, the balls B_k shrink regularly to $\mathbf{x}$, and Theorem 7.16 implies that

$$f(\mathbf{x}) = \lim_{k\to\infty} f_{B_k} \quad \text{for a.e. } \mathbf{x} \in B.$$

In fact, this equality holds at every Lebesgue point of f in B. Fix such an $\mathbf{x}$. Then,

$$f_B - f(\mathbf{x}) = \lim_{k\to\infty} \left(f_B - f_{B_k}\right) = \left(f_B - f_{B_1}\right) + \sum_{k=1}^{\infty} \left(f_{B_k} - f_{B_{k+1}}\right),$$

and therefore,

$$\left|f(\mathbf{x}) - f_B\right| \leq \left|f_B - f_{B_1}\right| + \sum_{k=1}^{\infty} \left|f_{B_k} - f_{B_{k+1}}\right| = I + II. \tag{14.16}$$

We have

$$\begin{aligned} I &= \left|\frac{1}{|B_1|} \int_{B_1} (f - f_B)\right| \leq \frac{1}{|B_1|} \int_{B_1} |f - f_B| \\ &\leq c_n \frac{1}{|B|} \int_B |f - f_B| \quad \text{since } B_1 \subset B \text{ and } r(B) \leq 8r_1 = 8r(B_1) \qquad (14.17) \\ &\leq c_n C_1 \frac{r(B)}{|B|} \int_B g(\mathbf{y})\, d\mathbf{y} \quad \text{by (14.13)} \\ &\leq c_n C_1 \int_B \frac{g(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|^{n-1}}\, d\mathbf{y} \end{aligned}$$

since for all $\mathbf{y} \in B$, we have $|\mathbf{x} - \mathbf{y}| < 2r(B)$, and consequently,

$$\frac{r(B)}{|B|} = c_n \frac{1}{r(B)^{n-1}} \le c_n \frac{1}{|\mathbf{x} - \mathbf{y}|^{n-1}}, \quad \mathbf{y} \in B.$$

Next, in order to estimate II, let $\{\widetilde{B}_k\}$ be balls as described in part (v) of Lemma 14.15. Then for $k = 1, 2, \ldots,$

$$\begin{aligned}
\left|f_{B_k} - f_{B_{k+1}}\right| &\le \left|f_{B_k} - f_{\widetilde{B}_k}\right| + \left|f_{\widetilde{B}_k} - f_{B_{k+1}}\right| \\
&= \left|\frac{1}{|\widetilde{B}_k|} \int_{\widetilde{B}_k} (f - f_{B_k})\right| + \left|\frac{1}{|\widetilde{B}_k|} \int_{\widetilde{B}_k} (f - f_{B_{k+1}})\right| \\
&\le \frac{c_n}{|B_k|} \int_{B_k} |f - f_{B_k}| + \frac{c_n}{|B_{k+1}|} \int_{B_{k+1}} |f - f_{B_{k+1}}|
\end{aligned}$$

since $\widetilde{B}_k \subset B_k \cap B_{k+1}$ and $|B_k|, |B_{k+1}| \le c_n |\widetilde{B}_k|$. Hence,

$$\begin{aligned}
II &\le 2 \sum_{k=1}^{\infty} \frac{c_n}{|B_k|} \int_{B_k} |f - f_{B_k}| \qquad (14.18) \\
&\le c_n C_1 \sum_{k=1}^{\infty} \frac{r_k}{|B_k|} \int_{B_k} g(\mathbf{y})\, d\mathbf{y} \quad \text{by (14.13)} \\
&= c_n C_1 \int_B g(\mathbf{y}) \left\{ \sum_{k=1}^{\infty} \chi_{B_k}(\mathbf{y}) \frac{r_k}{|B_k|} \right\} d\mathbf{y}
\end{aligned}$$

since $B_k \subset B$. By Lemma 14.15(iv), for each $\mathbf{y} \in B$, there are at most two nonzero terms in the sum in the last integrand. Also, if $\mathbf{y} \in B_k$, then since $B_k \subset B(\mathbf{x}; 3r_k)$, we have $|\mathbf{x} - \mathbf{y}| < 3r_k$ and therefore,

$$\frac{r_k}{|B_k|} = c_n \frac{1}{r_k^{n-1}} \le c_n \frac{1}{|\mathbf{x} - \mathbf{y}|^{n-1}}, \quad \mathbf{y} \in B_k.$$

Hence, the entire sum in curly brackets is also bounded by $c_n |\mathbf{x} - \mathbf{y}|^{-(n-1)}$. Combining estimates, we obtain

$$I + II \le c_n C_1 \int_B \frac{g(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|^{n-1}}\, d\mathbf{y},$$

which completes the proof of Theorem 14.12.

Remark 14.19 *As mentioned after the statement of Theorem 14.12, the set of points* $\mathbf{x}$ *for which (14.14) holds can be assumed to contain all points of the Lebesgue set of* f *in* B *for all balls* $B \subset B_0$*. Indeed, the proof of Theorem 14.12 shows that if (14.13) holds, then (14.14) is true if* $\mathbf{x}$ *is a Lebesgue point of* f *in* B*. On the other hand, the proof also shows that if (14.14) is true, then so is (14.13), and consequently, (14.14) holds for all* $\mathbf{x}$ *in the Lebesgue set of* f *in* B *for all balls* $B \subset B_0$*. A related fact will be used in the proof of Theorem 14.25.*

Proof of Lemma 14.15. We will construct a chain with the desired properties in case $B = B(\mathbf{0}; r)$ and $\mathbf{x} = (x, 0, \ldots, 0)$ with $0 \le x < r$. The construction in the general case is similar and left as an exercise.

Fix $\mathbf{x}$ and r as above and define real numbers $\{x_k\}_{k=1}^{\infty}$ by

$$x_1 = \frac{1}{2}\left(-\frac{r}{2} + x\right); \quad x_{k+1} = \frac{1}{2}(x_k + x) = x_k + \frac{1}{2}(x - x_k) \text{ if } k \ge 1. \tag{14.20}$$

Then $\{x_k\}$ is strictly increasing and $-r/4 \le x_k < x$ for all k. Denote

$$\mathbf{x}_k = (x_k, 0, \ldots, 0) \text{ and } r_k = \frac{1}{2}(x - x_k) \quad \text{if } k \ge 1.$$

We will show that the balls B_k defined by

$$B_k = B(\mathbf{x}_k; r_k), \quad k \ge 1, \tag{14.21}$$

have the desired properties. To prove property (i), let $\mathbf{y} \in B_k$ for some k and note that $r(B_k) = r_k = (x - x_k)/2$. Then

$$|\mathbf{y}| \le |\mathbf{y} - \mathbf{x}_k| + |\mathbf{x}_k| < r_k + |x_k| = \frac{1}{2}(x - x_k + 2|x_k|).$$

If $x_k \ge 0$, the last expression is $(x + x_k)/2 < x < r$, while if $x_k < 0$, it is

$$\frac{1}{2}(x - 3x_k) \le \frac{1}{2}(x - 3x_1) < \frac{1}{2}\left(r + \frac{3}{4}r\right) < r.$$

In either case, $\mathbf{y} \in B(\mathbf{0}; r) = B$, which proves (i).

For (ii), since $|\mathbf{x} - \mathbf{x}_k| = x - x_k$ for all k, we have $r_k = |\mathbf{x} - \mathbf{x}_k|/2$ by the definition of r_k. Also,

$$r_1 = \frac{1}{2}(x - x_1) = \frac{x}{4} + \frac{r}{8} \ge \frac{r}{8}.$$

Moreover, $\mathbf{x}_k \to \mathbf{x}$ since $x - x_{k+1} = \frac{1}{2}(x - x_k)$ if $k \geq 1$ by (14.20) and therefore by iteration, we have

$$|\mathbf{x} - \mathbf{x}_k| = x - x_k = \frac{x - x_1}{2^{k-1}} \quad \text{if } k \geq 1. \tag{14.22}$$

For (iii), note that the first equality in (14.22) combined with the definition of r_k gives $|\mathbf{x} - \mathbf{x}_k| = 2r_k$, and then for every $\mathbf{y} \in B_k$, we obtain

$$|\mathbf{y} - \mathbf{x}| \leq |\mathbf{y} - \mathbf{x}_k| + |\mathbf{x}_k - \mathbf{x}| < r_k + 2r_k = 3r_k.$$

This proves (iii). Note also that $r_k = (x - x_1)/2^k = x_{k+1} - x_k$, $k \geq 1$, and therefore,

$$r(B_k) = r_k \leq \frac{r(B)}{2^{k-1}}, \quad k \geq 1, \tag{14.23}$$

a fact that will be used in the proof of Theorem 14.25.

To prove (iv), fix k and suppose that $B_k \cap B_\ell \neq \emptyset$ for some $\ell > k$. We must show that $\ell = k + 1$. For any $\mathbf{y} \in B_k \cap B_\ell$,

$$\begin{aligned} x_\ell - x_k = |\mathbf{x}_\ell - \mathbf{x}_k| &\leq |\mathbf{x}_\ell - \mathbf{y}| + |\mathbf{y} - \mathbf{x}_k| \\ &< r_\ell + r_k = \frac{x - x_1}{2^\ell} + \frac{x - x_1}{2^k}. \end{aligned}$$

But by (14.22),

$$x_\ell - x_k = (x - x_k) - (x - x_\ell) = \frac{x - x_1}{2^{k-1}} - \frac{x - x_1}{2^{\ell-1}}.$$

Combining the previous two inequalities and dividing by $x - x_1$ give

$$\frac{2}{2^k} - \frac{2}{2^\ell} < \frac{1}{2^\ell} + \frac{1}{2^k}, \quad \text{or equivalently} \quad \frac{1}{2^k} < \frac{3}{2^\ell}.$$

This is possible only if $\ell = k + 1$ since we have assumed that $\ell > k$. Thus, (iv) is proved.

Finally, in order to verify (v), fix $k \geq 1$ and define

$$\widetilde{x_k} = x_k + \frac{3}{4} r_k, \quad \widetilde{\mathbf{x}_k} = (\widetilde{x_k}, 0, \ldots, 0), \quad \widetilde{B_k} = B\left(\widetilde{\mathbf{x}_k}; \frac{1}{4} r_k\right).$$

First note that

$$\left|\widetilde{B_k}\right| = c_n r\left(\widetilde{B_k}\right)^n = c_n \left(\frac{r_k}{4}\right)^n = \frac{1}{4^n} |B_k|,$$

and since $r_k = 2r_{k+1}$, also $|\widetilde{B}_k| = |B_{k+1}|/2^n$. Thus, (v) will be proved if we show that $\widetilde{B}_k \subset B_k \cap B_{k+1}$. Let $\mathbf{y} \in \widetilde{B}_k$. Then,

$$\begin{aligned} |\mathbf{y} - \mathbf{x}_k| &\le |\mathbf{y} - \widetilde{\mathbf{x}}_k| + |\widetilde{\mathbf{x}}_k - \mathbf{x}_k| \\ &< \frac{1}{4} r_k + \frac{3}{4} r_k = r_k, \end{aligned}$$

and so $\widetilde{B}_k \subset B_k$. Similarly, since $\widetilde{x}_k = x_{k+1} - \frac{1}{4} r_k$ (recall that $r_k = x_{k+1} - x_k$), we have

$$\begin{aligned} |\mathbf{y} - \mathbf{x}_{k+1}| &\le |\mathbf{y} - \widetilde{\mathbf{x}}_k| + |\widetilde{\mathbf{x}}_k - \mathbf{x}_{k+1}| \\ &< \frac{1}{4} r_k + \frac{1}{4} r_k = \frac{1}{2} r_k = r_{k+1}. \end{aligned}$$

It follows that $\widetilde{B}_k \subset B_{k+1}$, which completes the proof of Lemma 14.15.

An immediate corollary of Theorem 14.12 is the following equivalence between the truths of the L^1, L^1 Poincaré inequality and the subrepresentation inequality for any integrable f that has first partial derivatives. However, the corollary does not assert the truth of either of these inequalities for such f. See also Theorem 15.16 and Exercise 22 in Chapter 15.

Corollary 14.24 *Let B_0 be a ball in $\mathbf{R^n}$ and f be an integrable function on B_0, all of whose first partial derivatives exist a.e. in B_0. Then the following two conditions are equivalent:*

(a) *There is a constant C_1 such that*

$$\frac{1}{|B|} \int_B |f - f_B| \, d\mathbf{x} \le C_1 \, r(B) \frac{1}{|B|} \int_B |\nabla f| \, d\mathbf{x} \quad \textit{for all balls } B \subset B_0.$$

(b) *There is a constant C_2 such that*

$$|f(\mathbf{x}) - f_B| \le C_2 \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^{n-1}} d\mathbf{y} \quad \textit{for a.e. } \mathbf{x} \in B \textit{ and all balls } B \subset B_0.$$

The constants C_1 and C_2 are equivalent in the sense that there is a positive constant c_n depending only on n such that $c_n^{-1} C_1 \le C_2 \le c_n C_1$.

By choosing $g = 1$ in (14.13), we obtain the condition

$$\frac{1}{|B|} \int_B |f - f_B| \le C r(B) \quad \text{for all balls } B \subset B_0.$$

This amounts to (14.9) for the Hölder (Lipschitz) functional $a(B) = Cr(B)^\beta$ in the special case $\beta = 1$. On the other hand, when $g = 1$, condition (14.14) becomes

$$|f(\mathbf{x}) - f_B| \le Cr(B) \quad \text{a.e. in } B \text{ for all balls } B \subset B_0,$$

since

$$\int_B \frac{1}{|\mathbf{x} - \mathbf{y}|^{n-1}} d\mathbf{y} \approx r(B) \quad \text{if } \mathbf{x} \in B,$$

with constants of comparability that depend only on n. Theorem 14.12 guarantees the equivalence of these two conditions on f. The next result, which is a companion for Hölder β-functionals of Theorem 14.12, characterizes such an f as being Lipschitz continuous on B_0 after possible redefinition in a subset of B_0 of measure zero.

Theorem 14.25 (Campanato, Meyers) *Let $0 < \beta \le 1$, B_0 be a ball in $\mathbf{R}^n$, and f be a function defined on B_0 with $f \in L^1(B_0)$. Then the following three conditions are equivalent.*

(i) *There is a constant C_1 such that*

$$\frac{1}{|B|} \int_B |f - f_B| \le C_1 r(B)^\beta \quad \textit{for all balls } B \subset B_0. \tag{14.26}$$

(ii) *There is a constant C_2 such that*

$$|f(\mathbf{x}) - f_B| \le C_2 r(B)^\beta \quad \textit{a.e. in } B \textit{ for all balls } B \subset B_0. \tag{14.27}$$

(iii) *There is a constant C_3 such that*

$$|f(\mathbf{x}) - f(\mathbf{y})| \le C_3 |\mathbf{x} - \mathbf{y}|^\beta \quad \textit{for a.e. } \mathbf{x}, \mathbf{y} \in B_0, \tag{14.28}$$

and consequently, after redefinition of f in at most a subset of B_0 of measure zero,

$$|f(\mathbf{x}) - f(\mathbf{y})| \le C_3 |\mathbf{x} - \mathbf{y}|^\beta \quad \textit{for all } \mathbf{x}, \mathbf{y} \in B_0. \tag{14.29}$$

Furthermore, any two of the constants C_1, C_2, C_3 are equivalent in the same sense as in Theorem 14.12; for example, if (14.26) is true, then (14.28) holds with $C_3 = c_{n,\beta} C_1$ for some constant $c_{n,\beta}$ that depends at most on n and β.

A function f for which (14.29) holds is said to satisfy a *Hölder (Lipschitz) condition of order* β on B_0, or to be *Hölder (Lipschitz) continuous of order* β on B_0.

Proof. We will show that (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (i); some details in the proof that (ii) $\Longrightarrow$ (iii) will be left as exercises. Fix $0 < \beta \leq 1$ and a ball B_0. Let $f \in L^1(B_0)$ and suppose that (i) holds for f. To show that (ii) holds, fix a ball $B \subset B_0$ and follow the notation and proof of Theorem 14.12 through (14.17), but now estimate (14.17) by using (14.26) to obtain

$$I \leq c_n \frac{1}{|B|} \int_B |f - f_B| \leq c_n C_1 r(B)^\beta.$$

Next, in order to estimate term II in (14.16), combine (14.18) with (14.26) to obtain

$$II \leq 2 \sum_{k=1}^{\infty} \frac{c_n}{|B_k|} \int_{B_k} |f - f_{B_k}| \leq 2c_n C_1 \sum_{k=1}^{\infty} r(B_k)^\beta.$$

Now recall from (14.23) that $r(B_k) \leq r(B)/2^{k-1}$ for all $k \geq 1$. Hence, since $\beta > 0$, we obtain

$$II \leq 2c_n C_1 \sum_{k=1}^{\infty} \frac{1}{(2^{k-1})^\beta} r(B)^\beta = c_{n,\beta} C_1 r(B)^\beta.$$

Combining the estimates for I and II, we see that the inequality in (ii) holds if $\mathbf{x}$ is a Lebesgue point of f in B. This completes the proof that (i) $\Longrightarrow$ (ii).

Next, let us sketch the proof that (ii) $\Longrightarrow$ (iii), leaving some details to the reader. Suppose then that (ii) holds and let L_f denote the Lebesgue set of f in B_0. Since (ii) clearly implies (i) by integration, it follows from what was just proved that the inequality in (ii) is true for all $\mathbf{x} \in B \cap L_f$ for every ball $B \subset B_0$. Then, by the triangle inequality,

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq 2C_2 r(B)^\beta \quad \text{if } \mathbf{x}, \mathbf{y} \in B \cap L_f \text{ and } B \subset B_0. \tag{14.30}$$

We will prove (14.28) by using (14.30) to show that there is a constant C_3 such that $|f(\mathbf{x}) - f(\mathbf{y})| \leq C_3 |\mathbf{x} - \mathbf{y}|^\beta$ for all $\mathbf{x}, \mathbf{y} \in L_f$. By translation, we may assume that $B_0 = B(\mathbf{0}; r)$. If $\mathbf{x}, \mathbf{y} \in L_f$ and $|\mathbf{x} - \mathbf{y}| \geq r/4$, then (14.30) with B chosen to be B_0 gives

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq 2C_2 r^\beta \leq 2C_2 (4|\mathbf{x} - \mathbf{y}|)^\beta = 2^{2\beta+1} C_2 |\mathbf{x} - \mathbf{y}|^\beta.$$

It remains to consider points $\mathbf{x}, \mathbf{y} \in L_f$ with $0 < |\mathbf{x} - \mathbf{y}| < r/4$. Fix such $\mathbf{x}, \mathbf{y}$ and choose $\varepsilon, R > 0$ depending on $\mathbf{x}, \mathbf{y}$ with

$$|\mathbf{x}| + \varepsilon,\ |\mathbf{y}| + \varepsilon < r \quad \text{and} \quad \frac{R}{2} < |\mathbf{x} - \mathbf{y}| < R < |\mathbf{x} - \mathbf{y}| + \varepsilon. \tag{14.31}$$

Note that $R < r/2$ since $R < 2|\mathbf{x} - \mathbf{y}| < r/2$. Also, $\mathbf{y} \in B(\mathbf{x}; R)$ and $\mathbf{x} \in B(\mathbf{y}; R)$ since $|\mathbf{x} - \mathbf{y}| < R$. If either $B(\mathbf{x}; R)$ or $B(\mathbf{y}; R)$ lies in B_0, we are done since, if for example $B(\mathbf{x}; R) \subset B_0$, then by (14.30) applied to $B(\mathbf{x}; R)$ and the fact that $\mathbf{x}, \mathbf{y} \in B(\mathbf{x}; R) \cap L_f$, we obtain

$$|f(\mathbf{x}) - f(\mathbf{y})| \le 2C_2 R^\beta \le 2C_2(2|\mathbf{x} - \mathbf{y}|)^\beta = 2^{\beta+1} C_2 |\mathbf{x} - \mathbf{y}|^\beta,$$

which is the desired estimate. Thus, we may assume that neither $B(\mathbf{x}; R)$ nor $B(\mathbf{y}; R)$ is contained in $B_0 = B(\mathbf{0}; r)$. Then both $\mathbf{x}$ and $\mathbf{y}$ belong to an annulus near the boundary of B_0:

$$\frac{r}{2} < r - R < |\mathbf{x}|,\ |\mathbf{y}| < r, \tag{14.32}$$

and, by Exercise 8, there exist (open) balls $B^{\mathbf{x}}, B^{\mathbf{y}} \subset B_0$ such that $\mathbf{x} \in B^{\mathbf{x}}, \mathbf{y} \in B^{\mathbf{y}}$, $r(B^{\mathbf{x}}) = r(B^{\mathbf{y}}) = R$, and $B^{\mathbf{x}} \cap B^{\mathbf{y}} \neq \emptyset$. Since L_f is dense in B_0, there is then a point $\mathbf{z} \in L_f$ such that $\mathbf{z} \in B^{\mathbf{x}} \cap B^{\mathbf{y}}$. Then $\mathbf{x}, \mathbf{y}, \mathbf{z} \in L_f$, and by applying (14.30) to both $B^{\mathbf{x}}$ and $B^{\mathbf{y}}$, we have

$$|f(\mathbf{x}) - f(\mathbf{z})| \le 2C_2 R^\beta, \quad |f(\mathbf{y}) - f(\mathbf{z})| \le 2C_2 R^\beta.$$

Therefore,

$$\begin{aligned} |f(\mathbf{x}) - f(\mathbf{y})| &\le 2C_2 R^\beta + 2C_2 R^\beta \\ &\le 4C_2(2|\mathbf{x} - \mathbf{y}|)^\beta = 2^{\beta+2} C_2 |\mathbf{x} - \mathbf{y}|^\beta. \end{aligned}$$

Thus, (14.28) is now proved in all cases, with $C_3 = 2^{\beta+2} C_2 \le 8C_2$.

Statement (14.29) in part (iii), namely, the fact that f can be redefined in a set of measure 0 such that $|f(\mathbf{x}) - f(\mathbf{y})| \le C_3 |\mathbf{x} - \mathbf{y}|^\beta$ for all $\mathbf{x}, \mathbf{y} \in B_0$, follows from (14.28) with the same constant C_3 (see Exercise 9). Finally, the implication (iii) $\Longrightarrow$ (i) follows immediately by integrating either (14.28) or (14.29) with respect to $\mathbf{y}$ over B. This completes the proof of Theorem 14.25.

14.3 Norm Estimates for I_α

We will now determine values of p and q for which I_α is a bounded operator from $L^p(\mathbf{R^n})$ to $L^q(\mathbf{R^n})$ and study some closely related "endpoint" estimates. We always assume $0 < \alpha < n$. In case $\alpha = 1$, the results will be used in Chapter 15 to derive Poincaré–Sobolev estimates.

We use the notation $\|f\|_p$ for the $L^p(\mathbf{R^n})$ norm of f, $1 \le p \le \infty$.

It will turn out that the values of p and q for which the norm inequality

$$\|I_\alpha f\|_q \le c\,\|f\|_p \quad \text{for all } f \in L^p(\mathbf{R^n}) \tag{14.33}$$

is true, for some c independent of f, are limited to $1 < p < \frac{n}{\alpha}$ and $\frac{1}{q} = \frac{1}{p} - \frac{\alpha}{n}$. For the endpoint values $p = 1$ and $p = n/\alpha$, we will derive variants of (14.33).

Let us begin by listing three comments that explain why the restrictions on p and q just mentioned are necessary for (14.33).

If $p \ge \frac{n}{\alpha}$, there exists $f \in L^p(\mathbf{R^n})$ such that $I_\alpha f = \infty$ everywhere in $\mathbf{R^n}$. In particular, (14.33) cannot hold if $p \ge \frac{n}{\alpha}$. (14.34)

If $1 \le p < \frac{n}{\alpha}$, the only value of q for which (14.33) can possibly hold for all $f \in L^p(\mathbf{R^n})$ satisfies $\frac{1}{q} = \frac{1}{p} - \frac{\alpha}{n}$, that is, $q = pn/(n-\alpha)$. (14.35)

If $p = 1$ and $\frac{1}{q} = 1 - \frac{\alpha}{n}$, there exists $f \in L^1(\mathbf{R^n})$ such that $\|I_\alpha f\|_q = \infty$. Thus, (14.33) fails when $p = 1$ and $q = n/(n-\alpha)$. (14.36)

We will verify the first and third of these. The second one follows from a basic dilation property of I_α, namely, if $0 < \lambda < \infty$ and $\delta_\lambda f$ denotes the dilation of f defined by $(\delta_\lambda f)(\mathbf{x}) = f(\lambda \mathbf{x})$, then $\delta_\lambda(I_\alpha f) = \lambda^\alpha I_\alpha(\delta_\lambda f)$; see Exercise 13. The formula $1/q = 1/p - \alpha/n$ is often called the *dimensional balance formula for* I_α (or simply *the balance formula for* I_α).

To verify (14.34), let α, p satisfy $0 < \alpha < n$ and $n/\alpha \le p \le \infty$. Let ψ denote the characteristic function of $\{\mathbf{y} : |\mathbf{y}| > 2\}$, and let

$$f(\mathbf{y}) = \frac{\psi(\mathbf{y})}{|\mathbf{y}|^\alpha \log|\mathbf{y}|}.$$

Then f is clearly bounded and nonnegative on $\mathbf{R^n}$, and also $f \in L^p(\mathbf{R^n})$ for $n/\alpha \le p < \infty$ since

$$\int_{\mathbf{R^n}} f^p = \int_{|\mathbf{y}|>2} \frac{1}{|\mathbf{y}|^{\alpha p}(\log|\mathbf{y}|)^p}\,d\mathbf{y} < \infty,$$

where we have used the fact that $p > 1$ in case $p = n/\alpha$. However, $I_\alpha f(\mathbf{x}) = \infty$ for every $\mathbf{x}$ since

$$\begin{aligned} I_\alpha f(\mathbf{x}) &\geq \int\limits_{|\mathbf{y}|>2+|\mathbf{x}|} \frac{1}{|\mathbf{y}|^\alpha \log|\mathbf{y}|} \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}}\, d\mathbf{y} \\ &\geq \int\limits_{|\mathbf{y}|>2+|\mathbf{x}|} \frac{1}{|\mathbf{y}|^\alpha \log|\mathbf{y}|} \frac{1}{(2|\mathbf{y}|)^{n-\alpha}}\, d\mathbf{y} \\ &= 2^{\alpha-n} \int\limits_{|\mathbf{y}|>2+|\mathbf{x}|} \frac{1}{|\mathbf{y}|^n \log|\mathbf{y}|}\, d\mathbf{y} = \infty. \end{aligned}$$

We note that in case $p = \infty$, a much simpler function f can be chosen above; for example, the constant function $f \equiv 1$ has fractional integral equal to

$$\int\limits_{\mathbf{R^n}} \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}}\, d\mathbf{y} = \int\limits_{\mathbf{R^n}} \frac{1}{|\mathbf{y}|^{n-\alpha}}\, d\mathbf{y} = +\infty \quad \text{for all } \mathbf{x}.$$

Also, in case $p = n/\alpha$, there are functions $f \in L^{n/\alpha}(\mathbf{R^n})$ *with compact support* such that $I_\alpha f \notin L^\infty(\mathbf{R^n})$; see Exercise 14.

To verify (14.36), let $0 < \alpha < n$ and $p = 1$. If B denotes the unit ball $B(\mathbf{0}; 1)$, then $\chi_B \in L^1(\mathbf{R^n})$ and

$$\begin{aligned} I_\alpha \chi_B(\mathbf{x}) &= \int\limits_{|\mathbf{y}|<1} \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}}\, d\mathbf{y} \\ &\geq \int\limits_{|\mathbf{y}|<1} \frac{1}{(|\mathbf{x}|+1)^{n-\alpha}}\, d\mathbf{y} = c_n \frac{1}{(|\mathbf{x}|+1)^{n-\alpha}}, \quad \mathbf{x} \in \mathbf{R^n}. \end{aligned}$$

Hence, if $q = n/(n-\alpha)$, then

$$\int\limits_{\mathbf{R^n}} (I_\alpha \chi_B)^q d\mathbf{x} \geq c_n^q \int\limits_{\mathbf{R^n}} \frac{1}{(|\mathbf{x}|+1)^n}\, d\mathbf{x} = +\infty.$$

The failure of integrability of $(I_\alpha \chi_B)^{n/(n-\alpha)}$ is due to its size for large values of $|\mathbf{x}|$. However, there are functions $f \in L^1(\mathbf{R^n})$ such that $|I_\alpha f|^{n/(n-\alpha)}$ is not even locally integrable; see Exercise 15.

The next result gives basic norm estimates for the Riesz operators I_α.

Theorem 14.37 (Hardy–Littlewood, Sobolev*) *Let*

$$0 < \alpha < n, \quad 1 \le p < \frac{n}{\alpha}, \quad \textit{and} \quad \frac{1}{q} = \frac{1}{p} - \frac{\alpha}{n}.$$

Then for every $f \in L^p(\mathbf{R^n})$, $I_\alpha f$ exists a.e. and is measurable in $\mathbf{R^n}$. Moreover,

(a) *if $1 < p < \frac{n}{\alpha}$, then*

$$\|I_\alpha f\|_q \le c\,\|f\|_p$$

for a constant c that depends only on α, n, and p;

(b) *if $p = 1$, then*

$$\sup_{\lambda>0} \lambda \left|\{\mathbf{x} \in \mathbf{R^n} : |I_\alpha f(\mathbf{x})| > \lambda\}\right|^{\frac{1}{q}} \le c\|f\|_1 \qquad \left(q = \frac{n}{n-\alpha}\right),$$

with c depending only on α and n.

Proof. The theorem can be proved in several ways. We will use a method due to L. Hedberg and based on Hölder's inequality and norm estimates for the Hardy–Littlewood maximal function.

Let f be nonnegative and measurable on $\mathbf{R^n}$. Then, as noted at the beginning of this chapter, $I_\alpha f$ is measurable on $\mathbf{R^n}$ by Corollary 6.16. For $\delta > 0$ to be chosen and $\mathbf{x} \in \mathbf{R^n}$, we write

$$\begin{aligned} I_\alpha f(\mathbf{x}) &= \int_{|\mathbf{x}-\mathbf{y}|<\delta} \frac{f(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}}\,d\mathbf{y} + \int_{|\mathbf{x}-\mathbf{y}|\ge\delta} \frac{f(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}}\,d\mathbf{y} \\ &= J_1(\mathbf{x}) + J_2(\mathbf{x}), \text{ say.} \end{aligned}$$

By Hölder's inequality, if $1 < p < \frac{n}{\alpha}$ and $1/p + 1/p' = 1$, then

$$J_2(\mathbf{x}) \le \|f\|_p \left(\int_{|\mathbf{x}-\mathbf{y}|\ge\delta} \frac{1}{|\mathbf{x}-\mathbf{y}|^{(n-\alpha)p'}}\,d\mathbf{y} \right)^{1/p'} = c_{n,\alpha,p}\,\delta^{\alpha-\frac{n}{p}}\,\|f\|_p$$

since $(n-\alpha)p' > n$ due to the restriction $p < n/\alpha$. In case $p = 1$, so that $p' = \infty$, the similar estimate $J_2(\mathbf{x}) \le \delta^{\alpha-n}\|f\|_1$ follows immediately from the definition of $J_2(\mathbf{x})$.

* Hardy and Littlewood considered the case $n = 1$ and Sobolev the case $n > 1$. When $p > 1$, Thorin obtained estimates, and $p = 1$ was studied by Zygmund.

Next, we will show that $J_1(\mathbf{x}) \le c_{n,\alpha}\delta^\alpha f^*(\mathbf{x})$, where f^* denotes the Hardy–Littlewood maximal function of f. In fact, this follows from Theorem 9.17, but a direct proof is simple:

$$\begin{aligned} J_1(\mathbf{x}) &\le \sum_{k=1}^{\infty} \int\limits_{\delta 2^{-k} \le |\mathbf{x}-\mathbf{y}| < \delta 2^{-k+1}} \frac{f(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{y} \\ &\le \sum_{k=1}^{\infty} \frac{1}{\left(\delta 2^{-k}\right)^{n-\alpha}} \int\limits_{|\mathbf{x}-\mathbf{y}| < \delta 2^{-k+1}} f(\mathbf{y})\, d\mathbf{y} \\ &\le c_n \sum_{k=1}^{\infty} \frac{\left(\delta 2^{-k+1}\right)^n}{\left(\delta 2^{-k}\right)^{n-\alpha}} f^*(\mathbf{x}) = c_{n,\alpha}\, \delta^\alpha f^*(\mathbf{x}). \end{aligned}$$

The maximal function f^* used here can be formed by using either cubes or balls centered at $\mathbf{x}$ in its definition since the two resulting functions are pointwise equivalent in size, with constants of equivalence depending only on n (cf. the second part of Exercise 9 of Chapter 9).

By combining the estimates for J_1 and J_2, we obtain

$$I_\alpha f(\mathbf{x}) \le c\left[\delta^\alpha f^*(\mathbf{x}) + \delta^{\alpha - \frac{n}{p}} \|f\|_p\right], \tag{14.38}$$

where the constant c depends only on n, α, and p. Choosing δ (depending on f and $\mathbf{x}$) such that the two terms on the right side are the same, namely, $\delta = \left(\|f\|_p / f^*(\mathbf{x})\right)^{p/n}$, gives

$$I_\alpha f(\mathbf{x}) \le c\, \|f\|_p^{\frac{\alpha p}{n}} f^*(\mathbf{x})^{1-\frac{\alpha p}{n}}. \tag{14.39}$$

This choice of δ essentially amounts to minimizing the right side of (14.38) with respect to δ and makes sense unless $f^*(\mathbf{x})$ is 0 or ∞. However, if $f^*(\mathbf{x}) = 0$ for any $\mathbf{x}$, then $f = 0$ a.e. in $\mathbf{R^n}$ and conclusions (a) and (b) are trivially true. On the other hand, if $f^*(\mathbf{x}) = \infty$, then (14.39) is automatically true and no choice of δ is needed.

It follows from (14.39) that

$$\begin{aligned} \|I_\alpha f\|_q &\le c\, \|f\|_p^{\frac{\alpha p}{n}} \left(\int\limits_{\mathbf{R^n}} (f^*)^{q(1-\frac{\alpha p}{n})} \right)^{\frac{1}{q}} \\ &= c\, \|f\|_p^{\frac{\alpha p}{n}} \left(\int\limits_{\mathbf{R^n}} (f^*)^p \right)^{\frac{1}{q}} \quad \text{since } \frac{1}{q} = \frac{1}{p} - \frac{\alpha}{n} \\ &= c\, \|f\|_p^{\frac{\alpha p}{n}}\, \|f^*\|_p^{\frac{p}{q}} \le c\, \|f\|_p^{\frac{\alpha p}{n} + \frac{p}{q}} = c\, \|f\|_p, \end{aligned}$$

where in order to obtain the last inequality, we have assumed $p > 1$ and applied Theorem 9.16. Note that the constant c still depends only on n, α, and p. This completes the proof of part (a) when $f \geq 0$.

If $p = 1$, then (14.39) implies that for any $\lambda > 0$,

$$\left\{\mathbf{x} \in \mathbf{R^n} : I_\alpha f(\mathbf{x}) > \lambda\right\} \subset \left\{\mathbf{x} \in \mathbf{R^n} : f^*(\mathbf{x}) > \left(\frac{\lambda}{c\|f\|_1^{\alpha/n}}\right)^{\frac{n}{n-\alpha}}\right\},$$

assuming as we may that $\|f\|_1 \neq 0$. Applying the weak type estimate in Lemma 7.9 to the set on the right side yields

$$\begin{aligned}\left|\left\{\mathbf{x} \in \mathbf{R^n} : I_\alpha f(\mathbf{x}) > \lambda\right\}\right| &\leq c\left(\frac{\|f\|_1^{\alpha/n}}{\lambda}\right)^{\frac{n}{n-\alpha}} \|f\|_1 \\ &= c\left(\frac{\|f\|_1}{\lambda}\right)^{\frac{n}{n-\alpha}},\end{aligned}$$

with c depending only on n and α. This agrees with the inequality in part (b). The theorem is now proved for nonnegative f.

Next, consider a general $f \in L^p(\mathbf{R^n})$ for $1 \leq p < n/\alpha$ and $0 < \alpha < n$. We must show that $I_\alpha f$ exists a.e. and is measurable in $\mathbf{R^n}$ and that (a) and (b) hold. The results just derived for nonnegative measurable functions can be applied to $|f|$ to conclude that (a) and (b) hold with $I_\alpha(|f|)$ in place of $I_\alpha f$. Since $|I_\alpha f(\mathbf{x})| \leq I_\alpha(|f|)(\mathbf{x})$ at any point $\mathbf{x}$ where $I_\alpha f(\mathbf{x})$ exists, it then suffices to verify that $I_\alpha f(\mathbf{x})$ exists a.e. and is measurable in $\mathbf{R^n}$. This is a consequence of local integrability of both f and the Riesz kernel $1/|\mathbf{y}|^{n-\alpha}$, together with the fact that

$$\left(|f| * \frac{1}{|\mathbf{y}|^{n-\alpha}}\right)(\mathbf{x}) = I_\alpha(|f|)(\mathbf{x}) < \infty \quad \text{a.e. in } \mathbf{R^n};$$

see Exercise 21 of Chapter 9. Alternately, the general case can be concluded by writing $f = f^+ - f^-$ and applying the results for nonnegative functions to both f^+ and f^-. The proof of Theorem 14.37 is now complete.

14.4 Exponential Integrability of $I_\alpha f$

In Theorem 14.37, the endpoint value $p = n/\alpha$ corresponding to $q = \infty$ is excluded. In fact, statement (14.34) shows in a dramatic way that the range of I_α on $L^{n/\alpha}(\mathbf{R^n})$ is not contained in $L^\infty(\mathbf{R^n})$. In this section, we will derive variants of Theorem 14.37 for $p = n/\alpha$ either by restricting I_α to the subclass of compactly supported $f \in L^{n/\alpha}(\mathbf{R^n})$ or by appropriately modifying the

definition of I_α for general $f \in L^{n/\alpha}(\mathbf{R^n})$. The range functions then turn out to be locally exponentially integrable; in particular, they belong to $L^q_{loc}(\mathbf{R^n})$ for every $q < \infty$. Results in case $p = n/\alpha$ have been studied extensively and are often called estimates of Trudinger or Moser–Trudinger type. They help motivate the study in Section 14.5 of functions of bounded mean oscillation, which also exhibit local exponential integrability, although of a weaker kind than in the present section.

We begin with a result for compactly supported functions $f \in L^{n/\alpha}(\mathbf{R^n})$, $0 < \alpha < n$. Such f belongs to $L^1(\mathbf{R^n})$ by Hölder's inequality, and consequently $I_\alpha f$ is measurable and finite a.e. by Theorem 14.37.

We will often denote e^t by $\exp\{t\}$ for $t \geq 0$.

Theorem 14.40 *Let $0 < \alpha < n$. There are positive constants c_1 and c_2 depending only on α and n such that if $f \in L^{n/\alpha}(\mathbf{R^n})$ and $f = 0$ outside a ball B, then*

$$\frac{1}{|B|} \int_B \exp\left\{ c_1 \left(\frac{|I_\alpha f(\mathbf{x})|}{\|f\|_{n/\alpha}} \right)^{\frac{n}{n-\alpha}} \right\} d\mathbf{x} \leq c_2.$$

In the statement of the theorem, we have tacitly assumed that $\|f\|_{n/\alpha} \neq 0$ since otherwise $I_\alpha f = 0$ everywhere, while the integral in the conclusion is meaningless.

Proof. Fix α, f, and B satisfying the hypothesis. Replacing f by $|f|/\|f\|_{n/\alpha}$, we may assume that $f \geq 0$ and $\|f\|_{n/\alpha} = 1$. Denote $R = 2r(B)$. For $\mathbf{x} \in B$ and $\delta \in (0, R]$, split $I_\alpha f$ as in the proof of Theorem 14.37:

$$\begin{aligned} I_\alpha f(\mathbf{x}) &= \int_{|\mathbf{x}-\mathbf{y}|<\delta} f(\mathbf{y}) \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{y} + \int_{|\mathbf{x}-\mathbf{y}|\geq\delta} f(\mathbf{y}) \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{y} \\ &= J_1(\mathbf{x}) + J_2(\mathbf{x}). \end{aligned}$$

As before, $J_1(\mathbf{x}) \leq c_{n,\alpha} \delta^\alpha f^*(\mathbf{x})$. In order to estimate $J_2(\mathbf{x})$, note that since $\mathbf{x} \in B$ and $f(\mathbf{y}) = 0$ if $\mathbf{y} \notin B$, then

$$\begin{aligned} J_2(\mathbf{x}) &= \int_{\delta \leq |\mathbf{x}-\mathbf{y}| \leq R} f(\mathbf{y}) \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{y} \\ &\leq \|f\|_{n/\alpha} \left(\int_{\delta \leq |\mathbf{x}-\mathbf{y}| \leq R} \frac{1}{|\mathbf{x}-\mathbf{y}|^n} d\mathbf{y} \right)^{\frac{n-\alpha}{n}} \quad \text{by Hölder's inequality} \\ &= c_{n,\alpha} \left(\log \frac{R}{\delta} \right)^{\frac{n-\alpha}{n}} \quad \text{since } \|f\|_{n/\alpha} = 1. \end{aligned}$$

Hence,

$$I_\alpha f(\mathbf{x}) \le c_{n,\alpha}\left[\delta^\alpha f^*(\mathbf{x}) + \left(\log\frac{R}{\delta}\right)^{\frac{n-\alpha}{n}}\right] \quad \text{if } \mathbf{x}\in B \text{ and } 0<\delta\le R.$$

If $f^*(\mathbf{x}) < \infty$, choose δ satisfying

$$\delta^\alpha = \min\left\{\frac{1}{f^*(\mathbf{x})}, R^\alpha\right\},$$

noting again that if $f^*(\mathbf{x}) = 0$, then $f = 0$ a.e. in $\mathbf{R^n}$ and there is nothing to prove. It follows that

$$I_\alpha f(\mathbf{x}) \le c_{n,\alpha}\left[1 + \left\{\log^+\left(Rf^*(\mathbf{x})^{\frac{1}{\alpha}}\right)\right\}^{\frac{n-\alpha}{n}}\right], \quad \mathbf{x}\in B. \tag{14.41}$$

Let $S = \left\{\mathbf{x}\in B : I_\alpha f(\mathbf{x}) > 2c_{n,\alpha}\right\}$, with $c_{n,\alpha}$ as in the last inequality. If $\mathbf{x}\in S$, the inequality implies that $\log^+\left(Rf^*(\mathbf{x})^{1/\alpha}\right) > 1$ and therefore that $\log\left(Rf^*(\mathbf{x})^{1/\alpha}\right) > 1$. Hence,

$$I_\alpha f(\mathbf{x}) \le 2c_{n,\alpha}\left\{\log\left(Rf^*(\mathbf{x})^{\frac{1}{\alpha}}\right)\right\}^{\frac{n-\alpha}{n}} \quad \text{if } \mathbf{x}\in S.$$

Exponentiating, we obtain

$$\exp\left\{\left(\frac{I_\alpha f(\mathbf{x})}{2c_{n,\alpha}}\right)^{\frac{n}{n-\alpha}}\right\} \le Rf^*(\mathbf{x})^{\frac{1}{\alpha}}, \quad \mathbf{x}\in S.$$

On the other hand, if $\mathbf{x}\in B-S$, then $I_\alpha f(\mathbf{x}) \le 2c_{n,\alpha}$, so that

$$\exp\left\{\left(\frac{I_\alpha f(\mathbf{x})}{2c_{n,\alpha}}\right)^{\frac{n}{n-\alpha}}\right\} \le e \quad \text{if } \mathbf{x}\in B-S.$$

Combining the last two estimates yields

$$\begin{aligned}\frac{1}{|B|}\int_B \exp\left\{\left(\frac{\left|I_\alpha f(\mathbf{x})\right|}{2c_{n,\alpha}}\right)^{\frac{n}{n-\alpha}}\right\} d\mathbf{x} &\le \frac{1}{|B|}\int_S Rf^*(\mathbf{x})^{\frac{1}{\alpha}}\,d\mathbf{x} + \frac{1}{|B|}\int_{B-S} e\,d\mathbf{x}\\ &\le \frac{R}{|B|}\left(\int_B f^*(\mathbf{x})^{\frac{n}{\alpha}}d\mathbf{x}\right)^{\frac{1}{n}} |B|^{\frac{1}{n'}} + e\end{aligned}$$

by Hölder's inequality with exponents n and n', $1/n + 1/n' = 1$. By Theorem 9.16, since $n/\alpha > 1$,

$$\left(\int_B f^*(\mathbf{x})^{\frac{n}{\alpha}} d\mathbf{x} \right)^{\frac{1}{n}} \leq \|f^*\|_{n/\alpha}^{1/\alpha} \leq c\,\|f\|_{n/\alpha}^{1/\alpha} = c \cdot 1 = c,$$

where c depends only on n and α. Also, since $R = 2r(B)$ and $|B|$ is a multiple depending only on n of $r(B)^n$,

$$\frac{R}{|B|}\,|B|^{1/n'} = \frac{2r(B)}{|B|^{1/n}} = c_n.$$

The theorem now follows immediately from the estimates above.

Theorem 14.40 has a variant for arbitrary functions $f \in L^{n/\alpha}(\mathbf{R^n})$ and a modified version of $I_\alpha f$. The modification is defined by

$$\widetilde{I_\alpha} f(\mathbf{x}) = \int_{\mathbf{R^n}} f(\mathbf{y}) \left[\frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} - \frac{\chi(\mathbf{y})}{|\mathbf{y}|^{n-\alpha}} \right] d\mathbf{y}, \quad 0 < \alpha < n, \tag{14.42}$$

where $\chi(\mathbf{y})$ is the characteristic function of the complement of the closed unit ball, that is,

$$\chi(\mathbf{y}) = \chi_{\{|\mathbf{y}|>1\}}(\mathbf{y}) = \begin{cases} 1 & \text{if } |\mathbf{y}| > 1 \\ 0 & \text{if } |\mathbf{y}| \leq 1. \end{cases}$$

Before proceeding, let us show that if $f \in L^{n/\alpha}(\mathbf{R^n})$, then $\widetilde{I_\alpha} f$ exists as a Lebesgue integral and is finite a.e. in $\mathbf{R^n}$. In fact, we will prove the stronger property that

$$\int_{\mathbf{R^n}} |f(\mathbf{y})| \left| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} - \frac{\chi(\mathbf{y})}{|\mathbf{y}|^{n-\alpha}} \right| d\mathbf{y} \in L^1_{loc}(\mathbf{R^n}) \text{ if } f \in L^{n/\alpha}(\mathbf{R^n}). \tag{14.43}$$

The integral in (14.43) is a measurable function of $\mathbf{x}$ in $\mathbf{R^n}$ by Tonelli's theorem since its integrand is a measurable function of $(\mathbf{x}, \mathbf{y})$ in $\mathbf{R^{2n}}$ by Lemmas 5.2 and 6.15. To verify (14.43), we will show that if $f \in L^{n/\alpha}(\mathbf{R^n})$ and $N > 1$, then

$$\int_{|\mathbf{x}|<N} \left(\int_{\mathbf{R^n}} |f(\mathbf{y})| \left| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} - \frac{\chi(\mathbf{y})}{|\mathbf{y}|^{n-\alpha}} \right| d\mathbf{y} \right) d\mathbf{x} < \infty.$$

By Tonelli's theorem, this iterated integral equals

$$\begin{aligned}
&\int_{|\mathbf{y}|<2N} |f(\mathbf{y})| \left(\int_{|\mathbf{x}|<N} \left| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} - \frac{\chi(\mathbf{y})}{|\mathbf{y}|^{n-\alpha}} \right| d\mathbf{x} \right) d\mathbf{y} \\
&\quad + \int_{|\mathbf{y}|>2N} |f(\mathbf{y})| \left(\int_{|\mathbf{x}|<N} \left| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} - \frac{\chi(\mathbf{y})}{|\mathbf{y}|^{n-\alpha}} \right| d\mathbf{x} \right) d\mathbf{y} \\
&= J + K, \quad \text{say.}
\end{aligned}$$

Since $\chi(\mathbf{y})/|\mathbf{y}|^{n-\alpha} \leq 1$ for all $\mathbf{y}$,

$$\begin{aligned}
J &\leq \int_{|\mathbf{y}|<2N} |f(\mathbf{y})| \left(\int_{|\mathbf{x}-\mathbf{y}|<3N} \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{x} + \int_{|\mathbf{x}|<N} 1\, d\mathbf{x} \right) d\mathbf{y} \\
&\leq c_{n,\alpha} \left(\int_{|\mathbf{y}|<2N} |f(\mathbf{y})|\, d\mathbf{y} \right) \left(N^{\alpha} + N^{n} \right) < \infty
\end{aligned}$$

since f is locally integrable. Also, since $\chi(\mathbf{y}) = 1$ when $|\mathbf{y}| > 1$, we have

$$\begin{aligned}
K &= \int_{|\mathbf{y}|>2N} |f(\mathbf{y})| \left(\int_{|\mathbf{x}|<N} \left| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} - \frac{1}{|\mathbf{y}|^{n-\alpha}} \right| d\mathbf{x} \right) d\mathbf{y} \\
&\leq \int_{|\mathbf{y}|>2N} |f(\mathbf{y})| \left(\int_{|\mathbf{x}|<N} c_{n,\alpha} \frac{|\mathbf{x}|}{|\mathbf{y}|^{n-\alpha+1}}\, d\mathbf{x} \right) d\mathbf{y} \\
&\leq c_{n,\alpha} N^{n+1} \int_{|\mathbf{y}|>2N} |f(\mathbf{y})| \frac{1}{|\mathbf{y}|^{n-\alpha+1}}\, d\mathbf{y}.
\end{aligned}$$

By Hölder's inequality with exponents n/α and $(n/\alpha)' = n/(n-\alpha)$, the last expression is at most

$$c_{n,\alpha} N^{n+1} \|f\|_{n/\alpha} \left(\int_{|\mathbf{y}|>2N} |\mathbf{y}|^{-\frac{n-\alpha+1}{n-\alpha} n} d\mathbf{y} \right)^{\frac{n-\alpha}{n}},$$

which is finite since $\frac{n-\alpha+1}{n-\alpha}n > n$ and $\|f\|_{n/\alpha} < \infty$. In fact,

$$\left(\int\limits_{|\mathbf{y}|>2N} |\mathbf{y}|^{-\frac{n-\alpha+1}{n-\alpha}n} d\mathbf{y}\right)^{\frac{n-\alpha}{n}} = c_{n,\alpha}N^{-1}.$$

This completes the proof of (14.43).

We also note that if $f \in L^{n/\alpha}(\mathbf{R^n})$, then $\widetilde{I_\alpha} f$ is measurable but may not be locally essentially bounded (see Exercises 16 and 14).

We can now state and prove an analogue of Theorem 14.40 for $\widetilde{I_\alpha} f$.

Theorem 14.44 *Let $0 < \alpha < n$. There are positive constants c_1 and c_2 depending only on n and α such that for any $f \in L^{n/\alpha}(\mathbf{R^n})$ and any ball B,*

$$\frac{1}{|B|}\int\limits_B \exp\left\{c_1\left(\frac{\left|\widetilde{I_\alpha} f(\mathbf{x}) - \left[\widetilde{I_\alpha} f\right]_B\right|}{\|f\|_{n/\alpha}}\right)^{\frac{n}{n-\alpha}}\right\} d\mathbf{x} \le c_2.$$

As in Theorem 14.40, we assume that $\|f\|_{n/\alpha} \neq 0$. Note that the average $[\widetilde{I_\alpha} f]_B$ exists and is finite if $\|f\|_{n/\alpha} < \infty$ since $\widetilde{I_\alpha} f$ is then measurable by Exercise 16 and locally integrable by (14.43).

Proof. Part of the proof is like that of Theorem 14.40, and we will concentrate on the differences. Fix α, f, and B. We may assume that $\|f\|_{n/\alpha} = 1$ by linearity of $\widetilde{I_\alpha}$. If $\mathbf{x} \in B$, then

$$\begin{aligned}
\left|\widetilde{I_\alpha} f(\mathbf{x}) - \left[\widetilde{I_\alpha} f\right]_B\right| &\le \frac{1}{|B|}\int\limits_B \left|\widetilde{I_\alpha} f(\mathbf{x}) - \widetilde{I_\alpha} f(\mathbf{z})\right| d\mathbf{z} \\
&\le \frac{1}{|B|}\int\limits_B \left(\int\limits_{\mathbf{R^n}} |f(\mathbf{y})| \left|\frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} - \frac{1}{|\mathbf{z}-\mathbf{y}|^{n-\alpha}}\right| d\mathbf{y}\right) d\mathbf{z} \\
&= \frac{1}{|B|}\int\limits_B \left(\int\limits_{2B} \cdots + \int\limits_{\mathbf{R^n}-2B} \cdots\right) d\mathbf{z} = A_1 + A_2,
\end{aligned}$$

say, where $2B$ denotes the ball of radius $2r(B)$ concentric with B. If $\mathbf{y} \notin 2B$ and $\mathbf{z} \in B$, we have $|\mathbf{z}-\mathbf{y}| \ge r(B)$, $|\mathbf{x}-\mathbf{y}| \ge r(B)$, and $|\mathbf{x}-\mathbf{z}| \le 2r(B)$, and therefore,

$$A_2 \le \frac{1}{|B|}\int\limits_B \left(\int\limits_{|\mathbf{y}-\mathbf{x}|\ge r(B)} |f(\mathbf{y})|\, c\frac{r(B)}{|\mathbf{x}-\mathbf{y}|^{n-\alpha+1}} d\mathbf{y}\right) d\mathbf{z},$$

where c depends only on n and α. Since the inner integral is independent of $\mathbf{z}$, we obtain

$$\begin{aligned} A_2 &\leq cr(B) \int\limits_{|\mathbf{y}-\mathbf{x}|\geq r(B)} |f(\mathbf{y})| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha+1}} d\mathbf{y} \\ &\leq cr(B)\|f\|_{n/\alpha} \left(\int\limits_{|\mathbf{y}|\geq r(B)} |\mathbf{y}|^{-\frac{n-\alpha+1}{n-\alpha}n} d\mathbf{y} \right)^{\frac{n-\alpha}{n}} \\ &= cr(B) \cdot 1 \cdot r(B)^{-1} = c. \end{aligned}$$

Next,

$$A_1 \leq \frac{1}{|B|} \int\limits_B \left(\int\limits_{2B} |f(\mathbf{y})| \left\{ \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} + \frac{1}{|\mathbf{z}-\mathbf{y}|^{n-\alpha}} \right\} d\mathbf{y} \right) d\mathbf{z}.$$

Performing the integration in $\mathbf{z}$ gives

$$A_1 \leq \int\limits_{2B} |f(\mathbf{y})| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{y} + \frac{1}{|B|} \int\limits_{2B} |f(\mathbf{y})|\, c\, r(B)^{\alpha} d\mathbf{y} = A_{11} + A_{12},$$

where to obtain A_{12}, we have used the fact that for any $\mathbf{y} \in 2B$,

$$\int\limits_B \frac{1}{|\mathbf{z}-\mathbf{y}|^{n-\alpha}} d\mathbf{z} \leq \int\limits_{|\mathbf{z}-\mathbf{y}|\leq 3r(B)} \frac{1}{|\mathbf{z}-\mathbf{y}|^{n-\alpha}} d\mathbf{z} = c\, r(B)^{\alpha}$$

for a constant c that depends only on n and α. By Hölder's inequality,

$$\begin{aligned} A_{12} &\leq \frac{1}{|B|} \left(\int\limits_{2B} |f|^{\frac{n}{\alpha}} d\mathbf{y} \right)^{\frac{\alpha}{n}} |2B|^{1-\frac{\alpha}{n}} cr(B)^{\alpha} \\ &\leq c\, \|f\|_{n/\alpha} = c. \end{aligned}$$

It remains to consider A_{11}. Since $\mathbf{x} \in B$,

$$A_{11} \leq \int\limits_{|\mathbf{x}-\mathbf{y}|\leq 3r(B)} |f(\mathbf{y})| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{y}.$$

The integral on the right side can be estimated by the method used in the proof of Theorem 14.40, namely, by choosing $\delta \in (0, 3r(B)]$ such that

$$\int\limits_{|\mathbf{x}-\mathbf{y}|\le 3r(B)} |f(\mathbf{y})| \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-\alpha}} d\mathbf{y} = \int\limits_{|\mathbf{x}-\mathbf{y}|<\delta} \cdots + \int\limits_{\delta\le|\mathbf{x}-\mathbf{y}|\le 3r(B)} \cdots$$
$$\le c\left[1 + \left\{\log^+\left(3r(B)f^*(\mathbf{x})^{\frac{1}{\alpha}}\right)\right\}^{\frac{n-\alpha}{n}}\right]$$

(cf. (14.41)). Details are left as an exercise. Collecting estimates yields a similar inequality for $\widetilde{I_\alpha f}(\mathbf{x}) - \left[\widetilde{I_\alpha f}\right]_B$:

$$\left|\widetilde{I_\alpha f}(\mathbf{x}) - \left[\widetilde{I_\alpha f}\right]_B\right| \le c\left[1 + \left\{\log^+\left(3r(B)f^*(\mathbf{x})^{\frac{1}{\alpha}}\right)\right\}^{\frac{n-\alpha}{n}}\right], \quad \mathbf{x} \in B.$$

The remainder of the proof of Theorem 14.44 is essentially identical to the part of the proof of Theorem 14.40 after (14.41). Details are left to the reader. This completes the proof of Theorem 14.44.

We close this section by making some further comments about $I_\alpha f$ in case $f \in L^{n/\alpha}(\mathbf{R^n})$. As already noted, $I_\alpha f$ may then be identically infinite, while $\widetilde{I_\alpha f}$ is finite a.e. and locally exponentially integrable. On the other hand, if $I_\alpha f$ exists in the Lebesgue sense and is finite at any point $\mathbf{x}_0$ where $\widetilde{I_\alpha f}$ also exists and is finite, then the definitions (14.1) and (14.42) imply that

$$I_\alpha f(\mathbf{x}_0) - \widetilde{I_\alpha f}(\mathbf{x}_0) = \int\limits_{|\mathbf{y}|>1} f(\mathbf{y}) \frac{1}{|\mathbf{y}|^{n-\alpha}}\, d\mathbf{y}. \tag{14.45}$$

Moreover, the integral on the right side of (14.45) then exists and is finite. Conversely, if this integral exists and is finite, then since it is independent of $\mathbf{x}_0$ and since by (14.43) $\widetilde{I_\alpha f}$ is locally integrable when $\|f\|_{n/\alpha} < \infty$, $I_\alpha f$ must also be locally integrable, and the difference $I_\alpha f - \widetilde{I_\alpha f}$ must be identically constant a.e. In this case, for a.e. $\mathbf{x} \in \mathbf{R^n}$, we have $I_\alpha f(\mathbf{x}) - [I_\alpha f]_B = \widetilde{I_\alpha f}(\mathbf{x}) - \left[\widetilde{I_\alpha f}\right]_B$, and Theorem 14.44 immediately yields the next corollary.

Corollary 14.46 *Let $0 < \alpha < n$ and $f \in L^{n/\alpha}(\mathbf{R^n})$. If*

$$\int\limits_{|\mathbf{y}|>1} |f(\mathbf{y})| \frac{1}{|\mathbf{y}|^{n-\alpha}}\, d\mathbf{y} < \infty,$$

then $I_\alpha f$ exists and is finite a.e. in $\mathbf{R^n}$ and

$$\frac{1}{|B|}\int_B \exp\left\{c_1\left(\frac{\left|I_\alpha f(\mathbf{x}) - [I_\alpha f]_B\right|}{\|f\|_{n/\alpha}}\right)^{\frac{n}{n-\alpha}}\right\} d\mathbf{x} \le c_2$$

for every ball $B \subset \mathbf{R^n}$, with c_1 and c_2 independent of f and B.

See also Exercise 18.

14.5 Bounded Mean Oscillation

In this section, we will study functions that satisfy (14.9) with B_0 replaced by $\mathbf{R^n}$ for the constant functional $a(B) = c$; cf. (14.11) with exponent $\beta = 0$. Such functions have a remarkable local exponential integrability property discovered by F. John and L. Nirenberg.

We begin with some standard terminology. A locally integrable function f on $\mathbf{R^n}$ is said to be of *bounded mean oscillation on* $\mathbf{R^n}$ (or to be a *BMO function on* $\mathbf{R^n}$) if there is a constant $c \ge 0$ such that for every ball $B \subset \mathbf{R^n}$,

$$\frac{1}{|B|}\int_B \left|f(\mathbf{x}) - f_B\right| d\mathbf{x} \le c. \tag{14.47}$$

As usual, f_B denotes the average $|B|^{-1}\int_B f$. The collection of all such f is denoted BMO($\mathbf{R^n}$) and called *the class of functions of bounded mean oscillation on* $\mathbf{R^n}$. Equivalently, if we denote

$$\|f\|_* = \sup_B \frac{1}{|B|}\int_B \left|f(\mathbf{x}) - f_B\right| d\mathbf{x}, \tag{14.48}$$

where the supremum is taken over all balls $B \subset \mathbf{R^n}$, then $f \in$ BMO($\mathbf{R^n}$) means that f is locally integrable and $\|f\|_* < \infty$.

Note that if f and g are any two locally integrable functions, then $\|f+g\|_* \le \|f\|_* + \|g\|_*$, and $\|cf\|_* = |c|\,\|f\|_*$ for any constant c. However, $\|\cdot\|_*$ is not a norm in the usual sense since it vanishes on constant functions.

The next lemma shows that if f is measurable and (14.47) holds with the average f_B replaced by a different constant depending on B, then f belongs to BMO($\mathbf{R^n}$).

Lemma 14.49 *Let f be a measurable function on $\mathbf{R^n}$. If there is a constant C such that*

$$\frac{1}{|B|}\int_B |f(\mathbf{x}) - c(f,B)|\,d\mathbf{x} \le C$$

for every ball $B \subset \mathbf{R^n}$ and for some constant $c(f,B)$ depending on f and B, then $f \in BMO(\mathbf{R^n})$. Moreover, $\|f\|_* \le 2C$.

Proof. If f satisfies the hypothesis, then f is clearly locally integrable by the triangle inequality. Furthermore, for any ball B,

$$\begin{aligned}\frac{1}{|B|}\int_B \left|f(\mathbf{x}) - f_B\right|\,d\mathbf{x} &\le \frac{1}{|B|}\int_B |f(\mathbf{x}) - c(f,B)|\,d\mathbf{x} + \left|c(f,B) - f_B\right| \\ &\le C + \left|c(f,B) - f_B\right|,\end{aligned}$$

where $c(f,B)$ and C are as in the hypothesis. Also,

$$\begin{aligned}\left|c(f,B) - f_B\right| &= \left|\frac{1}{|B|}\int_B \big(f(\mathbf{x}) - c(f,B)\big)\,d\mathbf{x}\right| \\ &\le \frac{1}{|B|}\int_B |f(\mathbf{x}) - c(f,B)|\,d\mathbf{x} \le C.\end{aligned}$$

Therefore, $\|f\|_* \le C + C = 2C$, and the lemma is proved.

A simple corollary of Lemma 14.49 is that balls can be replaced by cubes in the definition of BMO($\mathbf{R^n}$). More precisely, suppose that f is locally integrable and satisfies the analogue of condition (14.47) for cubes, that is suppose that

$$\|f\|_{**} := \sup \frac{1}{|Q|}\int_Q \left|f(\mathbf{x}) - f_Q\right|\,d\mathbf{x} < \infty,$$

where the supremum is taken over all cubes Q with edges parallel to the coordinate axes, and $f_Q = |Q|^{-1}\int_Q f$. Then, given any ball B, by enclosing B in a cube Q with $|Q| \le c_n|B|$, we obtain

$$\int_B \left|f - f_Q\right|\,d\mathbf{x} \le \int_Q \left|f - f_Q\right|\,d\mathbf{x} \le \|f\|_{**}|Q| \le c_n\|f\|_{**}|B|.$$

It follows from Lemma 14.49 with $c(f,B)$ there chosen to be f_Q that $f \in$ BMO($\mathbf{R^n}$) and $\|f\|_* \le 2c_n\|f\|_{**}$. The converse is also true, that is, the definition of BMO($\mathbf{R^n}$) using balls implies the analogous definition using cubes,

and $\|f\|_{**} \le c\|f\|_*$ for some constant c that depends only on n; we leave the verification as an exercise.

Our main goal is to study the "size" of functions of bounded mean oscillation. To set the stage, we begin by listing a few examples. First, it is easy to see that $L^\infty(\mathbf{R^n}) \subset \mathrm{BMO}(\mathbf{R^n})$. In fact, if $f \in L^\infty(\mathbf{R^n})$, then f is clearly locally integrable, and for any ball B,

$$\begin{aligned}\frac{1}{|B|}\int_B |f(\mathbf{x}) - f_B|\,d\mathbf{x} &\le \|f - f_B\|_\infty \\ &\le \|f\|_\infty + |f_B| \le 2\|f\|_\infty.\end{aligned}$$

Hence, $f \in \mathrm{BMO}(\mathbf{R^n})$ and $\|f\|_* \le 2\|f\|_\infty$.

However, the containment $L^\infty(\mathbf{R^n}) \subset \mathrm{BMO}(\mathbf{R^n})$ is a proper containment. For example, the (essentially) unbounded function $\log|\mathbf{x}|$ is of bounded mean oscillation on $\mathbf{R^n}$; see Exercise 20. We also leave it as an exercise to check the following two facts: if $-\infty < \lambda < \infty$ and $\lambda \ne 0$, then $|\mathbf{x}|^\lambda \notin \mathrm{BMO}(\mathbf{R^n})$; and there are functions with compact support that belong to $L^p(\mathbf{R^n})$ for all p, $0 < p < \infty$, but do not belong to $\mathrm{BMO}(\mathbf{R^n})$. See Exercises 21 and 22.

Another subclass of $\mathrm{BMO}(\mathbf{R^n})$ is the collection of all functions $\widetilde{I_\alpha} f$ defined in (14.42) when $f \in L^{n/\alpha}(\mathbf{R^n})$, $0 < \alpha < n$. This follows immediately from the exponential integrability estimate in Theorem 14.44 and the simple inequality $t \le \exp(t^\gamma)$, $t \ge 0$, $\gamma \ge 1$. See also Exercise 27.

All the examples given so far of $\mathrm{BMO}(\mathbf{R^n})$ functions are locally exponentially integrable in the sense that there are positive constants c and γ such that

$$\int_B \exp\left\{c|f(\mathbf{x})|^\gamma\right\} d\mathbf{x} < \infty \text{ for every ball } B \subset \mathbf{R^n}.$$

A natural question is whether the same is true for every $f \in \mathrm{BMO}(\mathbf{R^n})$. If $\gamma > 1$, the answer is *no*, and an example is $f(\mathbf{x}) = \log|\mathbf{x}|$; for example, if $n = 1$ and $\gamma > 1$, then for any $c > 0$,

$$\int_0^{1/2} \exp\left\{c\left(\log\frac{1}{x}\right)^\gamma\right\} dx = \int_{\ln 2}^{\infty} e^{cu^\gamma} e^{-u}\,du = +\infty.$$

However, the answer is *yes* if $\gamma = 1$. This is a corollary of the following basic fact, which is the main result of the section.

Theorem 14.50 (John–Nirenberg) *There are positive constants c_1 and c_2 depending only on n such that if $f \in BMO(\mathbf{R^n})$, B is a ball in $\mathbf{R^n}$ and $\lambda > 0$, then*

$$\left|\left\{\mathbf{x} \in B : \left|f(\mathbf{x}) - f_B\right| > \lambda\right\}\right| \le c_1 \left(\exp\left\{-\frac{c_2\lambda}{\|f\|_*}\right\}\right) |B|. \tag{14.51}$$

Here, we have assumed that $\|f\|_* \neq 0$; otherwise, f is constant a.e. in $\mathbf{R^n}$ and the left side of (14.51) is zero for all B and all $\lambda > 0$.

Before giving a proof, let us deduce two corollaries of the theorem.

Corollary 14.52 *Let $f \in BMO(\mathbf{R^n})$ and $1 \le p < \infty$. There is a positive constant c depending only on n and p such that for every ball $B \subset \mathbf{R^n}$,*

$$\left(\frac{1}{|B|}\int_B \left|f - f_B\right|^p d\mathbf{x}\right)^{1/p} \le c\,\|f\|_*.$$

In particular, $f \in L^p_{\text{loc}}(\mathbf{R^n})$, and for every ball B,

$$\left(\frac{1}{|B|}\int_B |f|^p d\mathbf{x}\right)^{1/p} \le c\,\|f\|_* + |f_B|$$

Proof. To prove the corollary, we compute (cf. Exercise 16 of Chapter 5)

$$\begin{aligned}
\frac{1}{|B|}\int_B \left|f - f_B\right|^p d\mathbf{x} &= \frac{p}{|B|}\int_0^\infty \lambda^{p-1}\left|\left\{\mathbf{x} \in B : \left|f(\mathbf{x}) - f_B\right| > \lambda\right\}\right| d\lambda \\
&\le \frac{p}{|B|}\int_0^\infty \lambda^{p-1} c_1 \exp\left\{-\frac{c_2\lambda}{\|f\|_*}\right\} |B|\, d\lambda \quad \text{by (14.51)} \\
&= pc_1\left(\frac{\|f\|_*}{c_2}\right)^p \int_0^\infty \lambda^{p-1}e^{-\lambda}\, d\lambda.
\end{aligned}$$

The first part of the corollary now follows by taking pth roots, and the second part then follows from Minkowski's inequality. Note that the integral $\int_0^\infty \lambda^{p-1}e^{-\lambda}d\lambda$ is the classical gamma function $\Gamma(p)$.

The proof shows that the constant c can be chosen to satisfy

$$\begin{aligned}
c^p &= pc_1c_2^{-p}\int_0^\infty \lambda^{p-1}e^{-\lambda}\, d\lambda \\
&\ge pc_1c_2^{-p}\int_p^\infty p^{p-1}e^{-\lambda}\, d\lambda = pc_1c_2^{-p}p^{p-1}e^{-p},
\end{aligned}$$

whose pth root is of order $O(p^{1/p'}) \to \infty$ as $p \to \infty$. The constant c in the conclusion of the corollary must tend to ∞ as $p \to \infty$ since functions in $\mathrm{BMO}(\mathbf{R^n})$ may not be locally essentially bounded.

Corollary 14.53 *Let c_1 and c_2 be as in Theorem 14.50. If $f \in BMO(\mathbf{R^n})$ and c_0 is a positive constant such that $c_0\|f\|_* < c_2$, then*

$$\frac{1}{|B|}\int_B \exp\left\{c_0\left|f - f_B\right|\right\}\,d\mathbf{x} \le 1 + \frac{c_0c_1}{c_2 - c_0\|f\|_*}$$

for every ball $B \subset \mathbf{R^n}$. In particular, $\int_B \exp\{c_0|f|\}\,d\mathbf{x} < \infty$ for every ball $B \subset \mathbf{R^n}$.

Proof. The result can be deduced from (14.51) combined with the following formula (see Exercise 29 of Chapter 5):

$$\int_B \exp\left\{c_0|f - f_B|\right\}\,d\mathbf{x} = |B| + c_0\int_0^\infty e^{c_0\lambda}|\left\{\mathbf{x} \in B : |f(\mathbf{x}) - f_B| > \lambda\right\}|\,d\lambda.$$

Further details are left to the reader.

Theorem 14.50 will be proved in several steps. The main step is to derive an analogue of (14.51) for n-dimensional cubes Q with edges parallel to the coordinate axes, that is, to show that

$$\left|\left\{\mathbf{x} \in Q : \left|f(\mathbf{x}) - f_Q\right| > \lambda\right\}\right| \le c_1\left(\exp\left\{-\frac{c_2\lambda}{\|f\|_{**}}\right\}\right)|Q| \tag{14.54}$$

for all such Q and all $\lambda > 0$, and now with $c_1 = 4$ and $c_2 = 2^{-n-1}\ln 2$. As usual, we assume that $\|f\|_{**} \ne 0$ since otherwise there is nothing to prove.

The proof of (14.54) will be based on the following n-dimensional version of the Decomposition Lemma 12.68 (cf. the second remark on p. 353 in Section 12.8). All cubes considered below are assumed to have edges parallel to the coordinate axes.

Lemma 14.55 (Decomposition Lemma in R^n) *Let Q be a cube in $\mathbf{R^n}$ and suppose that $f \in L^1(Q)$ and $f \ge 0$. Then for any real number s satisfying*

$$s \ge \frac{1}{|Q|}\int_Q f,$$

there are nonoverlapping cubes $Q_1, Q_2, \ldots$ contained in Q such that

(i) $s < \frac{1}{|Q_k|}\int_{Q_k} f \le 2^n s$ *for all* k,

(ii) $f(\mathbf{x}) \le s$ *for a.e.* $\mathbf{x} \in Q - \bigcup Q_k$,

(iii) $\sum_k |Q_k| \le \frac{1}{s}\int_{\bigcup Q_k} f \le \frac{1}{s}\int_Q f$.

Proof. The proof is essentially identical to that of Lemma 12.68, but since the lemma is basic, we will repeat the main ideas.

Fix Q, f, and s as in the hypothesis, and subdivide Q into 2^n subcubes Q' of equal size by bisecting each edge of Q. For each Q', either $|Q'|^{-1}\int_{Q'} f \le s$ or $|Q'|^{-1}\int_{Q'} f > s$. Using the hypothesis that $|Q|^{-1}\int_Q f \le s$ together with the fact that $|Q| = 2^n|Q'|$, we have for each Q' that either

$$\frac{1}{|Q'|}\int_{Q'} f \le s \quad \text{or} \quad s < \frac{1}{|Q'|}\int_{Q'} f \le 2^n s.$$

If Q' satisfies the first condition, we call it a cube of the first kind; otherwise, we say it is of the second kind.

We save any cube of the second kind and repeat the process for each Q' of the first kind by subdividing it into 2^n cubes Q'' of equal size. For each Q'', we again have either

$$\frac{1}{|Q''|}\int_{Q''} f \le s \quad \text{or} \quad s < \frac{1}{|Q''|}\int_{Q''} f \le 2^n s.$$

Save those Q'' of the second kind, repeat the procedure for every Q'' of the first kind, and so on. Let $\{Q_k\}$ be all the cubes of the second kind in the construction above. The Q_k are clearly nonoverlapping and satisfy property (i). Each $\mathbf{x}$ in $Q - \bigcup Q_k$ belongs by construction to every $\overline{Q}$ in a sequence $\{\overline{Q}\}$ of cubes with $|\overline{Q}| \to 0$ and $|\overline{Q}|^{-1}\int_{\overline{Q}} f \le s$. Consequently, by the Lebesgue differentiation theorem, $f(\mathbf{x}) \le s$ for a.e. such $\mathbf{x}$, which proves property (ii). Finally, (iii) follows by adding over k the first inequality in (i), rewritten in the form $|Q_k| < s^{-1}\int_{Q_k} f$, $k = 1, 2, \ldots$. This completes the proof of Lemma 14.55.

Proof of Theorem 14.50. We begin by proving (14.54), which is the version of (14.51) for cubes instead of balls. We may consider only those $f \in \mathrm{BMO}(\mathbf{R^n})$ that are not identically constant a.e. in $\mathbf{R^n}$. Then $0 < \|f\|_{**} < \infty$, and by replacing f by $f/\|f\|_{**}$, we may consider only those f such that $\|f\|_{**} = 1$. Any such f satisfies

$$\frac{1}{|Q|}\int_Q |f - f_Q| \le 1 \text{ for every cube } Q \subset \mathbf{R^n}.$$

For $\lambda > 0$, define

$$F(\lambda) = \sup_{\substack{Q,f \\ \|f\|_{**}=1}} \frac{|\{\mathbf{x} \in Q : |f(\mathbf{x}) - f_Q| > \lambda\}|}{|Q|}.$$

Clearly, $F(\lambda) \le 1$ for all $\lambda > 0$, and our objective is to show that $F(\lambda) \le c_1 e^{-c_2\lambda}$ with $c_1 = 4$ and $c_2 = 2^{-n-1}\ln 2$.

Fix Q and f with $\|f\|_{**} = 1$ and apply Lemma 14.55 to $|f - f_Q|$ for a fixed value $s \ge 1$, noting that

$$\frac{1}{|Q|}\int_Q |f - f_Q| \le 1 \le s.$$

Then there are nonoverlapping subcubes Q_k, $k = 1, 2, \ldots$, of Q such that $|f - f_Q| \le s$ for almost all $\mathbf{x} \in Q - \bigcup_k Q_k$ and

$$s < \frac{1}{|Q_k|}\int_{Q_k} |f - f_Q| \le 2^n s \quad \text{for all } Q_k. \tag{14.56}$$

If $\lambda > s$, then, except possibly for a set of measure zero,

$$\{\mathbf{x} \in Q : |f(\mathbf{x}) - f_Q| > \lambda\} \subset \bigcup_k \{\mathbf{x} \in Q_k : |f(\mathbf{x}) - f_Q| > \lambda\}. \tag{14.57}$$

We have

$$|f_{Q_k} - f_Q| = \left|\frac{1}{|Q_k|}\int_{Q_k} (f - f_Q)\right| \le \frac{1}{|Q_k|}\int_{Q_k} |f - f_Q| \le 2^n s$$

by (14.56). Therefore, for any $\mathbf{x} \in Q$,

$$|f(\mathbf{x}) - f_Q| \le |f(\mathbf{x}) - f_{Q_k}| + |f_{Q_k} - f_Q| \le |f(\mathbf{x}) - f_{Q_k}| + 2^n s.$$

If the left side $|f(\mathbf{x}) - f_Q|$ exceeds λ, then by subtracting $2^n s$ from both sides, we obtain $|f(\mathbf{x}) - f_{Q_k}| > \lambda - 2^n s$. Hence, by (14.57), if $\lambda > s$, then

$$|\{\mathbf{x} \in Q : |f(\mathbf{x}) - f_Q| > \lambda\}| \le \sum_k |\{\mathbf{x} \in Q_k : |f(\mathbf{x}) - f_{Q_k}| > \lambda - 2^n s\}|.$$

If $\lambda > 2^n s$, it follows that

$$\frac{|\{\mathbf{x} \in Q : |f(\mathbf{x}) - f_Q| > \lambda\}|}{|Q|} \le \frac{1}{|Q|} \sum_k F(\lambda - 2^n s)\, |Q_k|$$
$$\le F(\lambda - 2^n s)\, \frac{1}{s}\, \frac{1}{|Q|} \int_Q |f - f_Q| \le \frac{F(\lambda - 2^n s)}{s},$$

where we have used property (iii) in Lemma 14.55 (applied to $|f - f_Q|$) in order to obtain the next-to-last estimate. Therefore,

$$F(\lambda) \le \frac{F(\lambda - 2^n s)}{s} \quad \text{if } s \ge 1 \text{ and } \lambda > 2^n s.$$

This inequality will now be iterated in order to prove the estimate $F(\lambda) \le 4 \exp\left\{-\left(2^{-n-1} \ln 2\right) \lambda\right\}$, $\lambda > 0$. Let

$$s = 2 \quad \text{and} \quad \gamma = 2^n s = 2^{n+1}.$$

Fix λ with $\lambda > \gamma$ and choose $m = 0, 1, 2, \ldots$ such that $(m+1)\gamma < \lambda \le (m+2)\gamma$. Then if $m \ne 0$,

$$\lambda > \lambda - \gamma > \lambda - 2\gamma > \cdots > \lambda - m\gamma > \gamma,$$

and we obtain $F(\lambda) \le 2^{-m} F(\lambda - m\gamma)$ after m iterations, even if $m = 0$. The trivial estimate $F(\lambda - m\gamma) \le 1$ together with $m \ge (\lambda/\gamma) - 2$ then gives

$$F(\lambda) \le 2^{-m} \le 2^{2 - \frac{\lambda}{\gamma}} = 4 e^{-\frac{\ln 2}{\gamma} \lambda} \quad \text{if } \lambda > \gamma.$$

Finally, if $0 < \lambda \le \gamma$, then it is also true that $F(\lambda) \le 4 e^{-\frac{\ln 2}{\gamma} \lambda}$ since $F(\lambda) \le 1$. This proves (14.54) for all $\lambda > 0$ and for all Q and f.

To complete the proof, it remains only to deduce inequality (14.51) for balls from its analogue (14.54) for cubes. The ideas needed are like those in the first paragraph after the proof of Lemma 14.49, and we will be brief. We will use the letters c, c_1, c_2 to denote various positive constants depending only on n, which may be different at each occurrence. Let B be a ball and $f \in \mathrm{BMO}(\mathbf{R^n})$, $\|f\|_* \ne 0$. Choose a cube satisfying $B \subset Q$ and $|Q| \le c\,|B|$. As usual,

$$\left|f_B - f_Q\right| \le \frac{1}{|B|} \int_B |f - f_Q| \le \frac{c}{|Q|} \int_Q |f - f_Q| \le c\, \|f\|_{**} \quad \text{and so}$$
$$\left|f(\mathbf{x}) - f_Q\right| \ge \left|f(\mathbf{x}) - f_B\right| - c\, \|f\|_{**}, \quad \mathbf{x} \in \mathbf{R^n}.$$

Then if λ is much larger than $\|f\|_{**}$, for example, if $\lambda/2 > c\,\|f\|_{**}$, we have

$$\begin{aligned}\left|\left\{\mathbf{x} \in B : \left|f(\mathbf{x}) - f_B\right| > \lambda\right\}\right| &\le \left|\left\{\mathbf{x} \in Q : \left|f(\mathbf{x}) - f_Q\right| > \lambda/2\right\}\right| \\ &\le c_1 \exp\left\{-c_2 \frac{\lambda}{2}\,\frac{1}{\|f\|_{**}}\right\} |Q| \quad \text{by (14.54).}\end{aligned}$$

Hence, since $|Q| \le c\,|B|$ and $c^{-1}\|f\|_{**} \le \|f\|_* \le c\,\|f\|_{**}$,

$$\left|\left\{\mathbf{x} \in B : |f(\mathbf{x}) - f_B| > \lambda\right\}\right| \le c_1 \exp\left\{-c_2 \frac{\lambda}{\|f\|_*}\right\} |B|, \quad \lambda > c\,\|f\|_*.$$

However, for the remaining values of λ, that is, when $0 < \lambda \le c\,\|f\|_*$, such an inequality is obvious, and Theorem 14.50 follows.

Exercises

1. Show that if $f \in C^1(\overline{B})$, where $\overline{B}$ denotes the closure of a ball B, then the inequalities in (14.3) and (14.4) are true for all $\mathbf{x} \in \overline{B}$.
2. Under the same assumptions as in Theorem 14.2, show that both of the estimates (14.3) and (14.4) can be improved by replacing the integral on their right-hand sides by

$$\int_B \frac{|\nabla f(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^n}\, d\mathbf{y} \le \sum_{i=1}^{n} \int_B \left|\frac{\partial f}{\partial y_i}(\mathbf{y})\right| \frac{|x_i - y_i|}{|\mathbf{x} - \mathbf{y}|^n}\, d\mathbf{y}.$$

3. Derive the L^1, L^1 Poincaré inequality (14.8) directly by the same method used to prove the subrepresentation Theorem 14.2, instead of obtaining it as a corollary of Theorem 14.2.
4. Let B_1 and B_2 be balls, with $B_1 \subset \mathbf{R}^{\mathbf{n_1}}$ and $B_2 \subset \mathbf{R}^{\mathbf{n_2}}$, $n_1, n_2 \ge 1$. If $f(\mathbf{x}_1, \mathbf{x}_2) \in C^1(B_1 \times B_2) \cap L^1(B_1 \times B_2)$, show that for every $(\mathbf{x}_1, \mathbf{x}_2) \in B_1 \times B_2$,

$$\begin{aligned}\left|f\left(\mathbf{x}_1, \mathbf{x}_2\right) - f_{B_1 \times B_2}\right| \le c &\int_{B_1 \times B_2} \left|\nabla_{\mathbf{y}_1} f\left(\mathbf{y}_1, \mathbf{y}_2\right)\right| k_1\left(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}_1, \mathbf{y}_2\right) d\mathbf{y}_1 d\mathbf{y_2} \\ &+ c \int_{B_1 \times B_2} \left|\nabla_{\mathbf{y}_2} f\left(\mathbf{y}_1, \mathbf{y}_2\right)\right| k_2\left(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}_1, \mathbf{y}_2\right) d\mathbf{y}_1 d\mathbf{y}_2,\end{aligned}$$

where

$$k_i\left(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}_1,\mathbf{y}_2\right)=\frac{1}{|B_1|\,|B_2|}\frac{|\mathbf{x}_i-\mathbf{y}_i|}{\left(\frac{|\mathbf{x}_1-\mathbf{y}_1|}{r(B_1)}+\frac{|\mathbf{x}_2-\mathbf{y}_2|}{r(B_2)}\right)^{n_1+n_2}},\quad i=1,2,$$

$f_{B_1\times B_2}=(|B_1|\,|B_2|)^{-1}\int_{B_1\times B_2}f$, and c is a constant that is independent of f, $\mathbf{x}_1$, $\mathbf{x}_2$, B_1, and B_2.

Note that each integral is at most a multiple of

$$\int\limits_{B_1\times B_2}\left(\left|\nabla_{\mathbf{y}_1}f(\mathbf{y}_1,\mathbf{y}_2)\right|\,r(B_1)+\left|\nabla_{\mathbf{y}_2}f(\mathbf{y}_1,\mathbf{y}_2)\right|\,r((B_2)\right)$$
$$\times\,K\left(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}_1,\mathbf{y}_2\right)d\mathbf{y}_1d\mathbf{y}_2$$

where

$$K\left(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}_1,\mathbf{y}_2\right)=\frac{1}{|B_1|\,|B_2|}\frac{1}{\left(\frac{|\mathbf{x}_1-\mathbf{y}_1|}{r(B_1)}+\frac{|\mathbf{x}_2-\mathbf{y}_2|}{r(B_2)}\right)^{n_1+n_2-1}}.$$

5. Let Q be a closed cube in $\mathbf{R^n}$, $n>1$. Show that if $f\in C^1(Q)$, then

$$\left|f(\mathbf{x})-f_Q\right|\le c_n\int\limits_Q\frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}}d\mathbf{y},\quad \mathbf{x}\in Q,$$

where $f_Q=|Q|^{-1}\int_Q f$ and c_n depends only on n. Also, by choosing Q to be the unit cube centered at the origin and making an affine change of variables, show that if $I=\prod_{i=1}^n[a_i,b_i]$ is an interval in $\mathbf{R^n}$ and $f\in C^1(I)$, then for all $\mathbf{x}\in I$,

$$\left|f(\mathbf{x})-f_I\right|\le c_n\frac{1}{|I|}\int\limits_I\left(\sum_{i=1}^n(b_i-a_i)\left|\frac{\partial f}{\partial y_i}(\mathbf{y})\right|\right)\frac{1}{\left(\sum_{i=1}^n\frac{|x_i-y_i|}{b_i-a_i}\right)^{n-1}}d\mathbf{y},$$

where $f_I=|I|^{-1}\int_I f$.

6. Let $I=\prod_{i=1}^n[a_i,b_i]$ be an interval in $\mathbf{R^n}$ and suppose that $f\in C^1(I)$. Show that

$$\int\limits_I|f(\mathbf{x})-f_I|\,d\mathbf{x}\le c_n\int\limits_I\sum_{i=1}^n(b_i-a_i)\left|\frac{\partial f}{\partial x_i}(\mathbf{x})\right|d\mathbf{x},$$

where $f_I=|I|^{-1}\int_I f$ and c_n is independent of f and I.

7. The proof of Lemma 14.15 was carried out for the ball $B = B(\mathbf{0}; r)$ and a point $\mathbf{x} = (x, 0, \ldots, 0)$ with $0 \le x < r$. Show that the proof is similar in the general case.
8. Verify the existence of balls $B^{\mathbf{x}}$ and $B^{\mathbf{y}}$ with the properties described after (14.32), that is, using the notation there and assuming that $\mathbf{x}$, $\mathbf{y}$, r, R, ε satisfy (14.31) and (14.32), show that there are open balls $B^{\mathbf{x}}, B^{\mathbf{y}} \subset B(\mathbf{0}; r)$ such that $\mathbf{x} \in B^{\mathbf{x}}$, $\mathbf{y} \in B^{\mathbf{y}}$, $r(B^{\mathbf{x}}) = r(B^{\mathbf{y}}) = R$, and $B^{\mathbf{x}} \cap B^{\mathbf{y}} \neq \emptyset$. (Denote $d = |\mathbf{x} - \mathbf{y}|$, $\mathbf{x}' = x/|\mathbf{x}|$, $\mathbf{y}' = y/|\mathbf{y}|$, and let $\mathbf{x}_d = (|\mathbf{x}| - d)\mathbf{x}'$ and $\mathbf{y}_d = (|\mathbf{y}| - d)\mathbf{y}'$. Then the balls $B^{\mathbf{x}} = B(\mathbf{x}_d; R)$ and $B^{\mathbf{y}} = B(\mathbf{y}_d; R)$ have the desired properties. The restriction on ε in (14.31) helps to show that $B^{\mathbf{x}}, B^{\mathbf{y}} \subset B(\mathbf{0}; r)$. Also, $(\mathbf{x}_d + \mathbf{y}_d)/2 \in B^{\mathbf{x}} \cap B^{\mathbf{y}}$. It may be helpful to note that $|\mathbf{x}_d - \mathbf{y}_d| < d$, for example, by the law of cosines.)
9. Show that a function that satisfies (14.28) can be redefined on a subset of B_0 of measure zero so that (14.29) holds with the same constant C_3 as in (14.28).
10. Prove that Theorems 14.12 and 14.25 remain true if B_0 is replaced by $\mathbf{R^n}$ and f is assumed to be locally integrable on $\mathbf{R^n}$.
11. Show that the function $|\mathbf{x}|^{1/2}$ is Hölder continuous of order β on $\mathbf{R^n}$ if and only if $\beta = 1/2$.
12. In the plane $\mathbf{R^2}$, consider the class of rectangles $I((x, y); h)$ with edges parallel to the coordinate axes, center (x, y), x-dimension h, and y-dimension h^2, $h > 0$, where (x, y) and h are allowed to vary. Suppose that $f(x, y)$ is a locally integrable function that satisfies

$$\frac{1}{h^3} \iint\limits_{I((x,y);h)} \left| f(u, v) - f_{I((x,y);h)} \right| du\, dv \le C h^{\beta}$$

for some C and β independent of (x, y) and h, with $0 < \beta \le 1$. Show that after possible redefinition on a set of measure zero, f satisfies

$$|f(x, y) - f(u, v)| \le c \left(|x - u| + |y - v|^{1/2} \right)^{\beta}$$

for all (x, y) and (u, v), where c depends only on C and β. (It may be helpful to consider $\mathbf{R^2}$ endowed with the metric $d((x, y), (u, v)) = |x - u| + |y - v|^{1/2}$ instead of the usual metric.)
13. Verify the dilation property $\delta_\lambda(I_\alpha f) = \lambda^\alpha I_\alpha(\delta_\lambda f)$, where $(\delta_\lambda f)(\mathbf{x}) = f(\lambda \mathbf{x})$, $\lambda > 0$, and use it to prove (14.35). Similarly, show that if $0 < \alpha < n$, $q > 0$, and there is a constant c such that

$$\sup_{t>0} t \left| \left\{ \mathbf{x} \in \mathbf{R^n} : \left| I_\alpha f(\mathbf{x}) \right| > t \right\} \right|^{1/q} \le c \, \|f\|_1$$

for all f, then $q = n/(n - \alpha)$.

14. Let B be the ball $B(\mathbf{0}; 1/2)$ in $\mathbf{R^n}$, and let $0 < \alpha < n$ and $0 \le \beta < 1 - (\alpha/n)$. Then the function

$$f(\mathbf{x}) = \chi_B(\mathbf{x}) \left\{ |\mathbf{x}|^\alpha \, |\log |\mathbf{x}||^{1-\beta} \right\}^{-1}$$

has compact support and belongs to $L^{n/\alpha}(\mathbf{R^n})$. Show that there is a positive constant c such that

$$I_\alpha f(\mathbf{x}) \ge c \begin{cases} \log\log \frac{1}{|\mathbf{x}|} & \text{if } \beta = 0 \\ \left(\log \frac{1}{|\mathbf{x}|}\right)^\beta & \text{if } 0 < \beta < 1 - \frac{\alpha}{n} \end{cases} \qquad \text{for all } \mathbf{x} \text{ near } \mathbf{0}.$$

In particular,

$$\lim_{\mathbf{x} \to \mathbf{0}} I_\alpha f(\mathbf{x}) = +\infty,$$

and $I_\alpha f$ is not essentially bounded in any neighborhood of the origin. How are $I_\alpha f$ and $\widetilde{I_\alpha} f$ related?

15. Let $0 < \alpha < n$, $\gamma = 2 - (\alpha/n)$, and $B = B(\mathbf{0}; 1/2)$. Show that the function

$$f(\mathbf{x}) = \chi_B(\mathbf{x}) \left\{ |\mathbf{x}|^n \, |\log |\mathbf{x}||^\gamma \right\}^{-1}$$

belongs to $L^1(\mathbf{R^n})$ but that $I_\alpha f \notin L^{n/(n-\alpha)}_{loc}(\mathbf{R^n})$. (Show that there is a constant $c > 0$ such that $I_\alpha f(\mathbf{x}) \ge c \, |\mathbf{x}|^{\alpha - n} \big| \log |\mathbf{x}| \big|^{1-\gamma}$ for all $\mathbf{x}$ near the origin.)

16. If $0 < \alpha < n$ and $f \in L^{n/\alpha}(\mathbf{R^n})$, show that $\widetilde{I_\alpha} f$ is a measurable function. (One way to proceed is to express $\widetilde{I_\alpha} f$ as the limit of a sequence of measurable functions: for example, by (14.43),

$$\widetilde{I_\alpha} f(\mathbf{x}) = \lim_{k \to \infty} \int_{|\mathbf{y}| < k} f(\mathbf{y}) \left[\frac{1}{|\mathbf{x} - \mathbf{y}|^{n-\alpha}} - \frac{\chi(\mathbf{y})}{|\mathbf{y}|^{n-\alpha}} \right] d\mathbf{y}$$

for almost every $\mathbf{x}$.)

17. Complete the proof of Theorem 14.44.

18. Let $0 < \alpha < n$ and $f \in L^{n/\alpha}(\mathbf{R^n})$. The discussion preceding Corollary 14.46 shows that the condition

$$\int_{|\mathbf{y}| > 1} |f(\mathbf{y})| \frac{1}{|\mathbf{y}|^{n-\alpha}} \, d\mathbf{y} < \infty$$

is necessary and sufficient for the existence and finiteness a.e. of $I_\alpha f$. Prove that the condition holds if there exists $p \in [1, n/\alpha)$ such that

$f \in L^{n/\alpha}(\mathbf{R^n}) \cap L^p(\mathbf{R^n})$. Consequently, the conclusion of Corollary 14.46 is valid for such f.

19. Let $0 < \beta \leq 1$ and B_0 be a ball in $\mathbf{R^n}$. Suppose that f is a measurable function on B_0 that satisfies

$$\sup_B \frac{1}{|B|^{1+(\beta/n)}} \int_B |f(\mathbf{x}) - c(f,B)|\, d\mathbf{x} < \infty,$$

where the supremum is taken over all balls $B \subset B_0$ and $c(f,B)$ is a constant depending on f and B. Show that $f \in L^1(B_0)$ and that (14.26) holds. (Argue as in the proof of Lemma 14.49.)

20. (a) Show that $\log|\mathbf{x}| \in \mathrm{BMO}(\mathbf{R^n})$.

(b) In case $n = 1$, show that the odd function $(\operatorname{sign} x)\log|x|$ does not belong to $\mathrm{BMO}(-\infty,\infty)$. Find an analogue of this fact in case $n > 1$.

(For part (a), it may help to consider two types of balls, those with center $\mathbf{x}_B \neq \mathbf{0}$ and radius $r(B) < |\mathbf{x}_B|/2$ and those with center $\mathbf{x}_B$ and radius $r(B) \geq |\mathbf{x}_B|/2$ [e.g., $\mathbf{x}_B = \mathbf{0}$], and apply Lemma 14.49. If B is of the first type, use the fact that $\log|\mathbf{x}|$ is continuously differentiable in B and choose $c(\log|\mathbf{x}|, B) = \log|\mathbf{x}_B|$ in the lemma. If B is of the second type, choose $c(\log|\mathbf{x}|, B) = \log r(B)$. For part (b), consider intervals centered at the origin.)

21. Show that $|\mathbf{x}|^\lambda \notin \mathrm{BMO}(\mathbf{R^n})$, $-\infty < \lambda < \infty$, $\lambda \neq 0$.
Show that $|\mathbf{x}|^\lambda \chi_{|\mathbf{x}|<1}(\mathbf{x}) \in \mathrm{BMO}(\mathbf{R^n})$ if $\lambda \geq 0$ but not if $\lambda < 0$.

22. Give an example of a compactly supported function that belongs to $L^p(\mathbf{R^n})$ for every p, $0 < p < \infty$, but does not belong to $\mathrm{BMO}(\mathbf{R^n})$.

23. Derive an analogue of the decomposition Lemma 14.55 for any given $s > 0$ assuming that $f \in L^1(\mathbf{R^n})$. (Given s and f, adjust the size of the cubes Q in the initial grid, which now covers all of $\mathbf{R^n}$, such that $|Q|^{-1}\int_Q f \leq s$ and all Q have the same edge length.)

24. Let $w(\mathbf{x})$ be measurable and positive a.e. in $\mathbf{R^n}$, and suppose that $\log w \in \mathrm{BMO}(\mathbf{R^n})$. Show that there are positive constants A_1 and A_2 depending on n, with A_1 also depending on $\|\log w\|_*$, such that

$$\left(\frac{1}{|B|}\int_B w^{A_1} d\mathbf{x}\right)\left(\frac{1}{|B|}\int_B w^{-A_1} d\mathbf{x}\right) \leq A_2 \quad \text{for all balls } B \subset \mathbf{R^n}.$$

(This can be deduced from Corollary 14.53 by writing

$$\int_B w^{A_1} d\mathbf{x} = \int_B \exp\left\{A_1\big(\log w - (\log w)_B\big)\right\} d\mathbf{x}\, \exp\left\{A_1(\log w)_B\right\},$$

together with a similar formula for $\int_B w^{-A_1} d\mathbf{x}$; note that the function $\log(1/w) = -\log w$ also belongs to $\mathrm{BMO}(\mathbf{R^n})$.)

25. Let $1 < p < \infty$. A nonnegative function $w(\mathbf{x})$ on $\mathbf{R^n}$ is said to satisfy *the A_p condition* if $w, w^{-1/(p-1)} \in L^1_{loc}(\mathbf{R^n})$ and there is a constant C such that for every ball $B \subset \mathbf{R^n}$,

$$\left(\frac{1}{|B|}\int_B w(\mathbf{x})\,d\mathbf{x}\right)\left(\frac{1}{|B|}\int_B w(\mathbf{x})^{\frac{1}{p-1}}\,d\mathbf{x}\right)^{p-1} \le C.$$

For such w, we will write $w \in A_p$. The opposite inequality with $C = 1$ is a consequence of Hölder's inequality. Note that $0 < w(\mathbf{x}) < \infty$ for a.e. $\mathbf{x}$ if $w \in A_p$.

(a) Show by direct estimation that $|\mathbf{x}|^\gamma \in A_p$ if $-n < \gamma < n(p-1)$. (Consider first the case when $r(B) < |\mathbf{x}_B|/2$, where $r(B)$ and $\mathbf{x}_B$ denote the radius and center of B.)

(b) Show that $\log w \in \mathrm{BMO}(\mathbf{R^n})$ if $w \in A_p$. (Consider separately the cases $p = 2$, $p < 2$ and $p > 2$. If $w \in A_2$, show that

$$\sup_B \frac{1}{|B|}\int_B \exp\left\{|\log w(\mathbf{x}) - \log w_B|\right\} d\mathbf{x} < \infty, \quad w_B = \frac{1}{|B|}\int_B w(\mathbf{x})\,d\mathbf{x}.$$

If $p < 2$ and $w \in A_p$, show that $w \in A_2$. In case $p > 2$, use the fact that if $w \in A_p$ then $w^{-1/(p-1)} \in A_{p'}$, $1/p + 1/p' = 1$.)

26. Let Q_0 be a cube in $\mathbf{R^n}$. We say that $f \in \mathrm{BMO}(Q_0)$ if $f \in L^1(Q_0)$ and

$$\|f\|_{**,Q_0} := \sup_{Q\subset Q_0} \frac{1}{|Q|}\int_Q |f - f_Q| < \infty,$$

where the supremum is taken over all cubes Q in Q_0 that have the same orientation as Q_0. Prove that if $f \in \mathrm{BMO}(Q_0)$, then there are positive constants c_1 and c_2 such that for all such Q and all $\lambda > 0$, (14.54) holds with $\|f\|_{**,Q_0}$ in place of $\|f\|_{**}$.

27. Let $\widetilde{f}$ be the periodic conjugate function of f as defined in Chapter 12:

$$\widetilde{f}(x) = -\frac{1}{\pi}\,\text{p.v.}\int_{|t|\le\pi} f(x-t)\frac{dt}{2\tan\frac{1}{2}t}.$$

(a) If $f \in L^\infty[-\pi,\pi]$, show that $\widetilde{f}$ belongs to the class $\mathrm{BMO}[-\pi,\pi]$ defined in Exercise 26 and that

$$\|\widetilde{f}\|_{**,[-\pi,\pi]} \le c\|f\|_{L^\infty[-\pi,\pi]}$$

with c independent of f.

(b) Deduce from part (a) the conclusion of Exercise 20 in Chapter 12.

Let $f \in L^\infty(\mathbf{R^1})$ and have compact support. Show that the Hilbert transform Hf of f belongs to $\mathrm{BMO}(\mathbf{R^1})$, and derive an analogue for $\|Hf\|_*$ of the estimate in part (a).

(For (a), given a small interval $I \subset [-\pi, \pi]$, let x_I denote the center of I and $2I$ denote the interval concentric with I with length $2|I|$. Decompose f on $(x_I - \pi, x_I + \pi)$ as $f = g + h$ where $g = f\chi_{2I}$. Use Hölder's inequality and Theorem 12.79, applied to the periodic extension of g, to show that $\int_I |\widetilde{g}| \le c|I|\,\|f\|_{L^\infty[-\pi,\pi]}$. Then estimate $\int_I |\widetilde{h}(x) - \widetilde{h}(x_I)|\,dx$ by using the mean value theorem.)

28. For a measurable function f on $\mathbf{R^n}$ and $0 < \alpha < n$, define the *fractional maximal function* $M_\alpha f$ of f by

$$M_\alpha f(\mathbf{x}) = \sup_{B:\mathbf{x}\in B} \frac{1}{r(B)^{n-\alpha}} \int_B |f(\mathbf{y})|\,d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n}.$$

Show that $M_\alpha f$ is a measurable function on $\mathbf{R^n}$ and that there is a constant c independent of f and $\mathbf{x}$ such that $M_\alpha f(\mathbf{x}) \le cI_\alpha(|f|)(\mathbf{x})$. As a consequence, the estimates in Theorem 14.37 for $I_\alpha f$ also hold for $M_\alpha f$. Show that the same is true if balls B are replaced by cubes Q in the definition of $M_\alpha f$, with $r(B)$ replaced by the edge length of Q.

29. Let $0 < \alpha < n$ and suppose that f satisfies

$$\int_{\mathbf{R^n}} |f|(1 + \log^+ |f|)\,d\mathbf{x} < \infty.$$

Prove that $I_\alpha f \in L^{n/(n-\alpha)}(E)$ for every measurable set $E \subset \mathbf{R^n}$ with $|E| < \infty$, and

$$\int_E |I_\alpha f|^{n/(n-\alpha)} d\mathbf{x} \le C\|f\|_{L^1(\mathbf{R^n})}^{\alpha/(n-\alpha)} \left(|E| + \int_{\mathbf{R^n}} |f| \log^+ |f|\,d\mathbf{x}\right),$$

where C is independent of f and E. (Combine (14.39) in case $p = 1$ with the estimate in Exercise 22 of Chapter 9.)

30. Let $k = 2, 3, \ldots$ and B be a ball in $\mathbf{R^n}$, $n > k$. Show that if $f \in C_0^k(B)$, then

$$|f(\mathbf{x})| \le c \int_B |\nabla^k f(\mathbf{y})| \frac{1}{|\mathbf{x} - \mathbf{y}|^{n-k}}\,d\mathbf{y}, \quad \mathbf{x} \in B,$$

where $|\nabla^k f(\mathbf{y})|$ denotes the sum $\sum_{|\alpha|=k} |D^\alpha f(\mathbf{y})|$, $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index of nonnegative integers, $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and c is a

constant independent of B and f. (In case $k=2$, apply part (ii) of Corollary 14.6 to f and each component of its gradient, and verify the estimate

$$\int_B \frac{1}{|\mathbf{x}-\mathbf{z}|^{n-1}} \frac{1}{|\mathbf{y}-\mathbf{z}|^{n-1}}\, d\mathbf{z} \le c_n \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-2}}, \quad \mathbf{x}, \mathbf{y} \in B,$$

when $n > 2$.)

15

Weak Derivatives and Poincaré–Sobolev Estimates

First-order Poincaré–Sobolev estimates in $\mathbf{R^n}$ are inequalities showing how L^p norms of the gradient of a function control the function itself. For a sufficiently smooth function f, the first-order partial derivatives $\partial f/\partial x_i$, $i=1,\ldots,n$, and the gradient ∇f of course have the usual meanings. When f is continuously differentiable and $n>1$, the first-order Poincaré–Sobolev estimates that we will derive are fairly simple consequences of the subrepresentation formulas and norm estimates for fractional integrals proved in Chapter 14. A notable exception to the simplicity of their derivation occurs when $p=1$, as we will see.

However, Poincaré–Sobolev estimates are also true for less smooth functions. Our first goal is to study a weaker notion of ∇f that may exist when the ordinary gradient does not. This will allow us to extend the subrepresentation formulas in Chapter 14 to more general functions than those of class C^1. Poincaré–Sobolev estimates for functions with weak derivatives can then be derived for $n>1$ by using the same pattern as for smooth functions, namely, by applying norm estimates for the fractional integral operator I_1. Results when $n=1$ will instead be obtained from the representation in Chapter 7 of an absolutely continuous function as the integral of its derivative.

15.1 Weak Derivatives

Let Ω be an open set in $\mathbf{R^n}$, $n\geq 1$. Let $L^1_{loc}(\Omega)$ denote the class of locally integrable real-valued functions f on Ω; as usual, we say f is locally integrable on Ω if it is integrable on every compact subset of Ω. The notion of the weak first-order partial derivatives of such an f is based on generalizing the standard formula for integration by parts by allowing functions that may be different from the ordinary partial derivatives of f to play their role in the formula.

The precise definition in case $n>1$ is as follows: if $f\in L^1_{loc}(\Omega)$ and $i=1,\ldots,n$, then $f(\mathbf{x})$ is said to have a *weak partial derivative in Ω with respect to* x_i, where $\mathbf{x}=(x_1,\ldots,x_n)$, if there is a function $g_i\in L^1_{loc}(\Omega)$ such that

$$\int_\Omega f(\mathbf{x})\,\frac{\partial\varphi}{\partial x_i}(\mathbf{x})\,d\mathbf{x} = -\int_\Omega g_i(\mathbf{x})\,\varphi(\mathbf{x})\,d\mathbf{x} \quad \text{for every } \varphi \in C_0^\infty(\Omega). \tag{15.1}$$

The justification for not including "integrated terms" on the right side of (15.1) is that each function φ has compact support in Ω.

An important part of the definition that f has a weak partial derivative g_i in Ω is that both f and g_i belong to $L^1_{loc}(\Omega)$, but neither f nor g_i is required to belong to $L^1(\Omega)$.

The functions φ in (15.1) are called *test functions* because they serve to test whether the same function g_i satisfies (15.1) as φ varies over a fairly large collection of functions with compact support. As we will see later, using $C_0^\infty(\Omega)$ as the class of test functions is sufficient to ensure that g_i is unique if it exists, and it will then make sense to refer to g_i as "the" weak partial derivative of f with respect to x_i.

On the other hand, if (15.1) holds for all $\varphi \in C_0^\infty(\Omega)$, then it also holds for all $\varphi \in Lip_0(\Omega)$, that is, for all φ that are Lipschitz continuous and compactly supported in Ω (see Theorem 15.7).

In case $n = 1$ and Ω is an open set in $(-\infty, \infty)$, the analogous definition is that a function $f \in L^1_{loc}(\Omega)$ has a *weak derivative in* Ω if there exists $g \in L^1_{loc}(\Omega)$ such that

$$\int_\Omega f(x)\varphi'(x)\,dx = -\int_\Omega g(x)\varphi(x)\,dx \quad \text{for every } \varphi \in C_0^\infty(\Omega). \tag{15.2}$$

There are simple, almost trivial, examples of functions that have weak partial derivatives but have ordinary partial derivatives nowhere. For example, in case $n = 1$, if $-\infty < s < t < \infty$, the function $\chi(x)$ that equals s for rational x and equals t for irrational x has this property since for every $\varphi \in C_0^\infty(\mathbf{R^1})$,

$$\int_{-\infty}^{\infty} \chi\,\varphi'\,dx = t\int_{-\infty}^{\infty} \varphi'\,dx = 0 = -\int_{-\infty}^{\infty} 0\,\varphi\,dx.$$

Therefore, χ has weak derivative 0 in $\mathbf{R^1}$ even though its ordinary derivative exists nowhere; in fact, χ is continuous nowhere. The averaging process that is inherent in integrating $\chi\,\varphi'$ compensates for the lack of ordinary differentiability of χ. We leave it as an exercise to construct a similar example in any dimension.

On the other hand, as we will see in Theorem 15.4, the weak first partial derivatives with respect to x_i, $i = 1, \ldots, n$, of a locally Lipschitz continuous function f agree with its corresponding ordinary partial derivatives $\partial f/\partial x_i$.

Let us begin by making several comments related to (15.1) and (15.2).

The integrals on both sides of (15.1) are well-defined and finite since if K denotes the support of φ, then

$$\int_\Omega \left| f \frac{\partial \varphi}{\partial x_i} \right| d\mathbf{x} = \int_K \left| f \frac{\partial \varphi}{\partial x_i} \right| d\mathbf{x} \le \left\| \frac{\partial \varphi}{\partial x_i} \right\|_{L^\infty(K)} \int_K |f| \, d\mathbf{x} < \infty,$$

where we have used the local integrability of f in Ω, and similarly,

$$\int_\Omega |g_i \varphi| \, d\mathbf{x} = \int_K |g_i \varphi| \, d\mathbf{x} \le \|\varphi\|_{L^\infty(K)} \int_K |g_i| \, d\mathbf{x} < \infty.$$

Next, let us consider the question of uniqueness of weak derivatives. Let $i = 1, \ldots, n$ and suppose that f has a weak partial derivative g_i with respect to x_i in Ω. Clearly, both f and g_i can be changed arbitrarily in any subset of Ω of measure zero without affecting (15.1). However, it turns out that g_i is uniquely determined a.e. in Ω by f. To prove this, it is enough to first note that if $\tilde{g}_i$ is another function in $L^1_{loc}(\Omega)$ that satisfies (15.1) for the same f, then

$$\int_\Omega (g_i - \tilde{g}_i) \, \varphi \, d\mathbf{x} = 0 \quad \text{for all } \varphi \in C_0^\infty(\Omega),$$

and then to apply part (i) of the next lemma with g chosen to be $g_i - \tilde{g}_i$.

Lemma 15.3 *Let Ω be an open set in $\mathbf{R^n}$ and let $g \in L^1_{loc}(\Omega)$.*

(i) *If*

$$\int_\Omega g \, \varphi \, d\mathbf{x} = 0 \quad \textit{for all } \varphi \in C_0^\infty(\Omega),$$

then $g = 0$ a.e. in Ω.

(ii) *If Ω is connected and*

$$\int_\Omega g \, \frac{\partial \varphi}{\partial x_i} \, d\mathbf{x} = 0 \quad \textit{for all } i = 1, \ldots, n \textit{ and all } \varphi \in C_0^\infty(\Omega),$$

then g is constant a.e. in Ω.

Part (ii) of Lemma 15.3 will be used in the proof of Theorem 15.6.

Proof. Part (i) can be proved in several ways. We will use a method based on a smooth approximation of the identity; the method is flexible and will

be used in other situations in this chapter. Let $k(\mathbf{x}) \in C_0^\infty(\{|\mathbf{x}| < 1\})$ satisfy $\int_{\mathbf{R}^n} k(\mathbf{x})\,d\mathbf{x} = 1$, and set $k_\varepsilon(\mathbf{x}) = \varepsilon^{-n}k(\mathbf{x}/\varepsilon)$, $\varepsilon > 0$. Each function $k_\varepsilon(\mathbf{x})$ is infinitely differentiable in $\mathbf{R}^n$ and supported in $\{|\mathbf{x}| < \varepsilon\}$. Let B be any fixed ball (or bounded open interval if $n = 1$) whose closure lies in Ω. Then there is a number $\varepsilon_0 > 0$ and a compact set K_0 with $B \subset K_0 \subset \Omega$ such that for all $\mathbf{y} \in B$ and all $\varepsilon \in (0, \varepsilon_0)$, the function $k_\varepsilon(\mathbf{y} - \mathbf{x})$ considered as a function of $\mathbf{x}$ has support in K_0. By hypothesis, we obtain

$$\int_\Omega g(\mathbf{x})\,k_\varepsilon(\mathbf{y} - \mathbf{x})\,d\mathbf{x} = 0 \quad \text{for all } \mathbf{y} \in B \text{ and all } \varepsilon < \varepsilon_0.$$

For such $\mathbf{y}$ and ε, the integral on the left equals

$$\int_{\mathbf{R}^n} (g\,\chi_{K_0})(\mathbf{x})k_\varepsilon(\mathbf{y} - \mathbf{x})\,d\mathbf{x},$$

which by Theorem 9.13 converges to $(g\chi_{K_0})(\mathbf{y})$ for a.e. $\mathbf{y} \in \mathbf{R}^n$ as $\varepsilon \to 0$, and so converges to $g(\mathbf{y})$ for a.e. $\mathbf{y} \in B$ as $\varepsilon \to 0$. Here, $g\,\chi_{K_0}$ is of course interpreted to be 0 outside K_0, that is,

$$g\,\chi_{K_0}(\mathbf{x}) = \begin{cases} g(\mathbf{x}) & \text{if } \mathbf{x} \in K_0 \\ 0 & \text{if } \mathbf{x} \in \mathbf{R}^n - K_0, \end{cases}$$

and we have used the fact that $g\,\chi_{K_0} \in L^1(\mathbf{R}^n)$. It follows that g is zero a.e in B and therefore also a.e. in Ω. This proves (i).

To prove (ii), let k_ε, B, ε_0, and K_0 be as above. If $\mathbf{y} \in B$ and $\varepsilon < \varepsilon_0$, then for every $i = 1, \ldots, n$, we have by hypothesis that

$$\begin{aligned} 0 &= \int_\Omega g(\mathbf{x})\frac{\partial}{\partial x_i}\Big[k_\varepsilon(\mathbf{y} - \mathbf{x})\Big]\,d\mathbf{x} \\ &= -\frac{\partial}{\partial y_i}\int_\Omega g(\mathbf{x})k_\varepsilon(\mathbf{y} - \mathbf{x})\,d\mathbf{x}. \end{aligned}$$

Hence, for all $\varepsilon < \varepsilon_0$, there is a constant $c_\varepsilon = c_{\varepsilon,B}$ such that

$$\int_\Omega g(\mathbf{x})k_\varepsilon(\mathbf{y} - \mathbf{x})\,d\mathbf{x} = c_\varepsilon \quad \text{for all } \mathbf{y} \in B.$$

As above, assuming that $\mathbf{y} \in B$ and $\varepsilon < \varepsilon_0$, g can be replaced by the integrable function $g\chi_{K_0}$ in this integration, and the domain Ω of integration can then be enlarged to $\mathbf{R}^n$. Now let $\varepsilon \to 0$. Then the integral converges to $g(\mathbf{y})$ a.e. in B, and consequently c_ε converges to some constant c, and $g(\mathbf{y}) = c$ a.e. in B.

Since Ω is connected by hypothesis, it follows that g is constant a.e. in Ω. This completes the proof of the lemma.

Since the function g_i in (15.1) is unique (if it exists), we may call it *the weak partial derivative of f with respect to x_i in Ω*. It is customary to use the familiar notation $\partial f/\partial x_i$ for g_i even though f may not have a partial derivative with respect to x_i in Ω in the ordinary sense. Similarly, when $n = 1$, the weak derivative of f will be denoted by f'. In situations when the notation might cause confusion, it can be clarified by also using the term "weak" or "ordinary" derivative as the case may be.

Let us now show that if $f \in Lip_{loc}(\Omega)$, then f has weak partial derivatives that are the same as its ordinary partial derivatives. Recall from the discussion preceding the Rademacher–Stepanov Theorem 7.53 that any $f \in Lip_{loc}(\Omega)$ has ordinary partial derivatives $\partial f/\partial x_i$, $i = 1, \ldots, n$, a.e. in Ω that are measurable and locally bounded in Ω. We note however that the full strength of the Rademacher–Stepanov theorem is not used here.

Theorem 15.4 *Let Ω be an open set in $\mathbf{R^n}$ and let $f \in Lip_{loc}(\Omega)$. Then for each $i = 1, \ldots, n$, f has a weak partial derivative g_i with respect to x_i in Ω given by the ordinary partial derivative $\partial f/\partial x_i$, that is,*

$$\int_\Omega f \frac{\partial \varphi}{\partial x_i}\, d\mathbf{x} = -\int_\Omega \frac{\partial f}{\partial x_i} \varphi\, d\mathbf{x} \quad \textit{if } \varphi \in C_0^\infty(\Omega) \textit{ and } i = 1, \ldots, n.$$

Proof. Let Ω, f, and φ satisfy the hypothesis, and let $\partial f/\partial x_i$ denote the ordinary partial derivative of f with respect to x_i. Note that both integrals in the conclusion are finite since φ and $\partial\varphi/\partial x_i$ are bounded and both f and $\partial f/\partial x_i$ are integrable (even bounded) on the support of φ.

We will prove the result in case $n > 1$ and $i = 1$; the general case is similar. Denote points $\mathbf{x} \in \mathbf{R^n}$ by

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) = (x_1, \hat{\mathbf{x}}) \quad \text{with } \hat{\mathbf{x}} = (x_2, \ldots, x_n) \in \mathbf{R^{n-1}},$$

and write

$$\Omega_{\hat{\mathbf{x}}} = \{x_1 : (x_1, \hat{\mathbf{x}}) \in \Omega\}$$
$$\text{and} \quad d\mathbf{x} = dx_1\, d\hat{\mathbf{x}} \quad \text{with} \quad d\hat{\mathbf{x}} = dx_2 \cdots dx_n.$$

By Theorem 6.8,

$$\int_\Omega f \frac{\partial \varphi}{\partial x_1}\, d\mathbf{x} = \int_{\mathbf{R^{n-1}}} \left[\int_{\Omega_{\hat{\mathbf{x}}}} f(x_1, \hat{\mathbf{x}}) \frac{\partial \varphi}{\partial x_1}(x_1, \hat{\mathbf{x}})\, dx_1 \right] d\hat{\mathbf{x}}.$$

Each set $\Omega_{\hat{\mathbf{x}}}$ is an open set in $(-\infty, \infty)$. If $\Omega_{\hat{\mathbf{x}}}$ is not empty, then Theorem 1.10 implies that it can be written as a countable union $\bigcup_j (\alpha_j, \beta_j)$ of disjoint open intervals (α_j, β_j), possibly of infinite length. The intervals (α_j, β_j) are the connected components of $\Omega_{\hat{\mathbf{x}}}$ and of course depend on Ω and $\hat{\mathbf{x}}$. When $f(x_1, \hat{\mathbf{x}})$ and $\varphi(x_1, \hat{\mathbf{x}})$ are considered as functions of x_1, $f(x_1, \hat{\mathbf{x}})$ is locally Lipschitz continuous on $\Omega_{\hat{\mathbf{x}}}$ and $\varphi(x_1, \hat{\mathbf{x}})$ is supported in the union $\bigcup_j (a_j, b_j)$ of open intervals (a_j, b_j) that satisfy $[a_j, b_j] \subset (\alpha_j, \beta_j)$ and $b_j - a_j < \infty$ for each j. Therefore, using Theorem 7.32(ii) and the absolute continuity on $[a_j, b_j]$ of $f(x_1, \hat{\mathbf{x}})$, we obtain

$$\begin{aligned}\int_{\alpha_j}^{\beta_j} f(x_1, \hat{\mathbf{x}}) \frac{\partial \varphi}{\partial x_1}(x_1, \hat{\mathbf{x}})\, dx_1 &= \int_{a_j}^{b_j} f(x_1, \hat{\mathbf{x}}) \frac{\partial \varphi}{\partial x_1}(x_1, \hat{\mathbf{x}})\, dx_1 \\ &= -\int_{a_j}^{b_j} \frac{\partial f}{\partial x_1}(x_1, \hat{\mathbf{x}}) \varphi(x_1, \hat{\mathbf{x}})\, dx_1 = -\int_{\alpha_j}^{\beta_j} \frac{\partial f}{\partial x_1}(x_1, \hat{\mathbf{x}}) \varphi(x_1, \hat{\mathbf{x}})\, dx_1.\end{aligned}$$

Adding over j gives

$$\int_{\Omega_{\hat{\mathbf{x}}}} f(x_1, \hat{\mathbf{x}}) \frac{\partial \varphi}{\partial x_1}(x_1, \hat{\mathbf{x}})\, dx_1 = -\int_{\Omega_{\hat{\mathbf{x}}}} \frac{\partial f}{\partial x_1}(x_1, \hat{\mathbf{x}}) \varphi(x_1, \hat{\mathbf{x}})\, dx_1.$$

Hence,

$$\begin{aligned}\int_{\Omega} f \frac{\partial \varphi}{\partial x_1}\, d\mathbf{x} &= \int_{\mathbf{R}^{n-1}} \left[-\int_{\Omega_{\hat{\mathbf{x}}}} \frac{\partial f}{\partial x_1}(x_1, \hat{\mathbf{x}}) \varphi(x_1, \hat{\mathbf{x}})\, dx_1 \right] d\hat{\mathbf{x}} \\ &= -\int_{\Omega} \frac{\partial f}{\partial x_1} \varphi\, d\mathbf{x},\end{aligned}$$

where the final equality is also due to Theorem 6.8, completing the proof.

Let f be a function defined on an open interval $(\alpha, \beta) \subset (-\infty, \infty)$, possibly of infinte length. We say that f is *absolutely continuous on* (α, β) if f is absolutely continuous on every compact subinterval $[a, b]$ of (α, β). By Theorem 7.29, this is equivalent to assuming that f has an ordinary derivative f' a.e. in (α, β) that is locally integrable in (α, β) and that

$$f(b) - f(a) = \int_a^b f'(x)\, dx \quad \text{if } \alpha < a \le b < \beta.$$

Remark 15.5 *The proof of Theorem 15.4 shows that the conclusion*

$$\int_\Omega f \frac{\partial \varphi}{\partial x_i}\, d\mathbf{x} = -\int_\Omega \frac{\partial f}{\partial x_i} \varphi\, d\mathbf{x} \quad \text{if } \varphi \in C_0^\infty(\mathbf{R^n})$$

is true for a particular i, $i = 1, \ldots, n$, if the assumption that $f \in Lip_{loc}(\Omega)$ is weakened by assuming both of the following:

(a) *f is absolutely continuous with respect to x_i on every connected component of the intersection of Ω with a.e. ($(n-1)$-dimensional measure) line parallel to the x_i axis.*

(b) *f and its ordinary partial derivative $\partial f/\partial x_i$ belong to $L^1_{loc}(\Omega)$.*

In case $n = 1$, condition (b) in Remark 15.5 is automatically true for any f that is absolutely continuous on every connected component of Ω. In fact, we have the next basic result.

Theorem 15.6 *Let Ω be an open set in $(-\infty, \infty)$ and $f \in L^1_{loc}(\Omega)$. Then f has a weak derivative in Ω if and only if f can be redefined in a subset of Ω of measure zero so that f is absolutely continuous on every connected component of Ω, that is, if and only if there is a function h on Ω such that $f = h$ a.e. in Ω and h is absolutely continuous on every connected component of Ω. Moreover, the weak derivative f' of f coincides with the ordinary derivative h'.*

Proof. The sufficiency of the condition follows from Remark 15.5, as we have already noted. To prove the necessity, suppose that f has weak derivative f' on Ω, and decompose Ω into the countable union of disjoint open intervals (α_j, β_j), possibly of infinite length: $\Omega = \bigcup_j (\alpha_j, \beta_j)$. These intervals are the connected components of Ω. Given an interval (α_j, β_j), choose a point $\gamma_j \in (\alpha_j, \beta_j)$ and define $\tilde{f}$ in (α_j, β_j) by

$$\tilde{f}(x) = \int_{\gamma_j}^{x} f'(t)\, dt, \quad x \in (\alpha_j, \beta_j).$$

In this way, $\tilde{f}$ is defined in all of Ω and is absolutely continuous on every (α_j, β_j) since $f' \in L^1_{loc}(\Omega)$. Clearly, f' is the weak derivative of f on each (α_j, β_j) since it is the weak derivative of f on Ω. Therefore, for every $\varphi \in C_0^\infty((\alpha_j, \beta_j))$, we have

$$\begin{aligned}\int_{\alpha_j}^{\beta_j} f\,\varphi'\,dx &= -\int_{\alpha_j}^{\beta_j} f'\,\varphi\,dx \\ &= -\int_{\alpha_j}^{\beta_j} (\tilde{f})'\,\varphi\,dx \quad \text{by Theorem 7.11} \\ &= \int_{\alpha_j}^{\beta_j} \tilde{f}\,\varphi'\,dx \quad \text{by Theorem 7.32(ii).}\end{aligned}$$

In the middle equality above, $(\tilde{f})'$ denotes the ordinary derivative of $\tilde{f}$. Hence, for all $\varphi \in C_0^\infty((\alpha_j, \beta_j))$,

$$\int_{\alpha_j}^{\beta_j} (f - \tilde{f})\,\varphi'\,dx = 0,$$

and it follows from the second part of Lemma 15.3 (in the one-dimensional case of an open interval) that there is a constant c_j such that

$$f(x) = \tilde{f}(x) + c_j \quad \text{for a.e. } x \in (\alpha_j, \beta_j).$$

Define $h(x)$ in Ω by $h(x) = \tilde{f}(x) + c_j$ if $x \in (\alpha_j, \beta_j)$, $j \geq 1$. Then h is absolutely continuous on every (α_j, β_j) and $f = h$ a.e. in Ω. Also, the ordinary derivative h' satisfies $h' = (\tilde{f})' = f'$ a.e. in Ω, and the proof is complete.

A simple application of Theorem 15.6 is that the step function

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ -1 & \text{if } -1 < x < 0 \end{cases}$$

does not have a weak derivative on $(-1, 1)$, even though it is infinitely differentiable on $(-1, 1)$ except at a single point. This follows immediately from Theorem 15.6 since f has a jump discontinuity in $(-1, 1)$, but a direct verification is not difficult. In fact, first note that if φ is infinitely differentiable and supported in $(-1, 1)$, then the left side of (15.2) equals

$$\begin{aligned}\int_{-1}^{1} f\varphi'\,dx &= \int_{0}^{1} \varphi'\,dx - \int_{-1}^{0} \varphi'\,dx \\ &= [\varphi(1) - \varphi(0)] - [\varphi(0) - \varphi(-1)] = -2\varphi(0)\end{aligned}$$

since $\varphi(1) = \varphi(-1) = 0$. If (15.2) were true, there would then exist $g \in L^1_{loc}(-1,1)$ such that

$$\int_{-1}^{1} g\varphi\, dx = 2\varphi(0) \quad \text{for all } \varphi \in C_0^\infty((-1,1)).$$

We leave it as an exercise to show that this is impossible; see Exercise 2. See also Exercise 3.

We conclude our basic comments related to (15.1) and (15.2) by showing that the class $C_0^\infty(\Omega)$ of test functions φ used in the definition of weak partial derivatives on Ω can be enlarged to the class $Lip_0(\Omega)$ of Lipschitz continuous functions with compact support in Ω. Thus, once it is verified by testing with the class $C_0^\infty(\Omega)$ that a function has weak derivatives, it is legitimate to use (15.1) (or (15.2)) for the larger class of functions $\varphi \in Lip_0(\Omega)$.

Theorem 15.7 *Let Ω be an open set in $\mathbf{R^n}$ and suppose that f has a weak partial derivative $\partial f/\partial x_i$ in Ω for some i, $i = 1, \ldots, n$. Then*

$$\int_\Omega f \frac{\partial \varphi}{\partial x_i}\, d\mathbf{x} = -\int_\Omega \frac{\partial f}{\partial x_i} \varphi\, d\mathbf{x} \quad \textit{for every } \varphi \in Lip_0(\Omega).$$

Proof. Consider again an approximation of the identity $k_\varepsilon(\mathbf{x}) = \varepsilon^{-n} k(\mathbf{x}/\varepsilon)$, $\varepsilon > 0$, $\mathbf{x} \in \mathbf{R^n}$, with $k \in C_0^\infty(\{|\mathbf{x}| < 1\})$ and $\int_{\mathbf{R^n}} k\, d\mathbf{x} = 1$. Let $\varphi \in Lip_0(\Omega)$. By extending φ to be zero outside Ω, we may think of φ as being Lipschitz continuous in all of $\mathbf{R^n}$. Since φ has compact support in Ω, there is a compact set $K_0 \subset \Omega$ and a number $\varepsilon_0 > 0$ such that K_0 contains the supports of φ and $\varphi * k_\varepsilon$ for all $\varepsilon < \varepsilon_0$. Moreover, we may assume that ε_0 is chosen so small that for all $\varepsilon < \varepsilon_0$ and $\mathbf{x} \in K_0$, $k_\varepsilon(\mathbf{x} - \mathbf{y})$ considered as a function of $\mathbf{y}$ is supported in Ω. By the definition of the weak derivative $\partial f/\partial x_i$,

$$\int_\Omega f(\mathbf{x}) \frac{\partial}{\partial x_i} (\varphi * k_\varepsilon)(\mathbf{x})\, d\mathbf{x} = -\int_\Omega \frac{\partial f}{\partial x_i}(\mathbf{x})\, (\varphi * k_\varepsilon)(\mathbf{x})\, d\mathbf{x}, \quad \varepsilon < \varepsilon_0.$$

Let us compute the limit as $\varepsilon \to 0$ of each side of this equation. The domains of integration on both sides may be replaced by K_0 because $\varepsilon < \varepsilon_0$. The expression on the right is then

$$-\int_{K_0} \frac{\partial f}{\partial x_i}\, (\varphi * k_\varepsilon)\, d\mathbf{x}.$$

Since $\partial f/\partial x_i \in L^1(K_0)$ and $(\varphi * k_\epsilon)(\mathbf{x})$ converges pointwise to $\varphi(\mathbf{x})$ and is bounded uniformly in ε and $\mathbf{x}$, this has limit equal to

$$-\int_{K_0} \frac{\partial f}{\partial x_i}\,\varphi\,d\mathbf{x} = -\int_{\Omega} \frac{\partial f}{\partial x_i}\,\varphi\,d\mathbf{x}.$$

On the other hand, for $\varepsilon < \varepsilon_0$, the integral on the left side earlier above is

$$\int_{K_0} f(\mathbf{x})\frac{\partial}{\partial x_i}(\varphi * k_\varepsilon)(\mathbf{x})\,d\mathbf{x}.$$

In order to compute its limit as $\varepsilon \to 0$, note that

$$\begin{aligned}\frac{\partial}{\partial x_i}(\varphi * k_\varepsilon)(\mathbf{x}) &= \int_{\Omega} \varphi(\mathbf{y})\frac{\partial}{\partial x_i}\big(k_\epsilon(\mathbf{x}-\mathbf{y})\big)\,d\mathbf{y}\\ &= \int_{\Omega} \varphi(\mathbf{y})\left[-\frac{\partial}{\partial y_i}\big(k_\epsilon(\mathbf{x}-\mathbf{y})\big)\right]d\mathbf{y} = \int_{\Omega} \frac{\partial\varphi}{\partial y_i}(\mathbf{y})k_\varepsilon(\mathbf{x}-\mathbf{y})\,d\mathbf{y},\end{aligned}$$

where to obtain the final equality, we have assumed that $\mathbf{x} \in K_0$ and $\varepsilon < \varepsilon_0$ and applied Theorem 15.4 to the Lipschitz function φ and the smooth test function $k_\varepsilon(\mathbf{x}-\mathbf{y})$ of $\mathbf{y}$. (See the related result in Exercise 4.) Since φ is Lipschitz continuous, the last integral is bounded uniformly in $\mathbf{x}$ and ε for $\mathbf{x} \in \mathbf{R^n}$ and $\varepsilon > 0$, and it converges a.e. to $(\partial\varphi/\partial x_i)(\mathbf{x})$ as $\varepsilon \to 0$. Since $f \in L^1(K_0)$, the Lebesgue dominated convergence theorem implies that

$$\begin{aligned}\lim_{\varepsilon\to 0}\int_{K_0} f(\mathbf{x})\frac{\partial}{\partial x_i}(\varphi * k_\varepsilon)(\mathbf{x})\,d\mathbf{x} &= \int_{K_0} f\frac{\partial\varphi}{\partial x_i}\,d\mathbf{x}\\ &= \int_{\Omega} f\frac{\partial\varphi}{\partial x_i}\,d\mathbf{x}.\end{aligned}$$

This completes the computation of the limits. It follows that

$$\int_{\Omega} f\frac{\partial\varphi}{\partial x_i}\,d\mathbf{x} = -\int_{\Omega} \frac{\partial f}{\partial x_i}\varphi\,d\mathbf{x},$$

and the theorem is proved.

If a locally integrable function f in Ω has weak partial derivatives $\partial f/\partial x_i$ in Ω for every $i = 1, \ldots, n$, $n > 1$, we write

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right) \quad \text{in } \Omega$$

and call this vector the *weak gradient of f in Ω*. It is customary to use the same notation ∇f for the weak gradient as for the ordinary gradient. By Theorem 15.4, the two notions are the same if $f \in Lip_{loc}(\Omega)$. Note that if f is any function with a weak gradient in Ω, then both f and $|\nabla f|$ are locally integrable in Ω by definition; here, the notation $|\nabla f|$ stands for

$$|\nabla f| = \left(\sum_{i=1}^{n} \left| \frac{\partial f}{\partial x_i} \right|^2 \right)^{1/2}.$$

15.2 Smooth Approximation and Sobolev Spaces

In Section 15.3, before deriving Poincaré–Sobolev estimates, we will show that the subrepresentation formulas for C^1 functions in Chapter 14 can be extended to functions with a weak gradient. The proofs of these extended versions of the formulas will be based on the known ones for C^1 functions and various results about approximation by smooth functions. Approximation theorems like Theorem 15.8 are generally attributed to K. Friedrichs, although their hypotheses may vary. Some variants are discussed farther below and in the exercises.

Theorem 15.8 *Let Ω be an open set in $\mathbf{R^n}$ and K be a compact subset of Ω. If f has a weak gradient ∇f in Ω, then there is a sequence $\{f_j\}_{j=1}^{\infty}$ of functions on Ω such that*

(i) *$f_j \in C_0^\infty(\Omega)$ for all j,*

(ii) *$f_j \to f$ a.e. in K and in $L^1(K)$ norm as $j \to \infty$,*

(iii) *$\nabla f_j \to \nabla f$ a.e. in K and in $L^1(K)$ norm as $j \to \infty$.*

In part (iii), the terminology $\nabla f_j \to \nabla f$ *in $L^1(K)$ norm* means that for every $i = 1, \ldots, n$, $\partial f_j/\partial x_i \to \partial f/\partial x_i$ in $L^1(K)$ norm as $j \to \infty$.

Before giving a proof, we note that more can be said about the supports of the approximating functions f_j: if G is any open set satisfying $K \subset G \subset \Omega$,

then conclusion (i) of the theorem can be replaced by $f_j \in C_0^\infty(G)$ for all j. Of course, the sequence $\{f_j\}$ then depends on G as well as on K. Indeed, to see why, we have only to apply the theorem with Ω replaced by G, after noting that if f has weak gradient ∇f in an open set Ω, then f also has weak gradient ∇f in any open set $G \subset \Omega$.

See Theorem 15.9 about the existence of a single subsequence $\{f_j\}$ that has the properties in Theorem 15.8 for every compact set $K \subset \Omega$.

Proof. We will use the standard method based on a smooth approximation of the identity, with a further refinement. As usual, $B(\mathbf{x}; r)$ denotes the (open) ball with center $\mathbf{x}$ and radius r.

Let f, Ω, and K satisfy the hypothesis, and let $k_\varepsilon(\mathbf{x})$ be as in the proof of Theorem 15.7. Choose an open set G and a number $\varepsilon_0 > 0$ such that $K \subset G$, G has compact closure in Ω, $B(\mathbf{x}; \varepsilon) \subset G$ for all $\mathbf{x} \in K$ and all $\varepsilon < \varepsilon_0$, and $B(\mathbf{y}; \varepsilon) \subset \Omega$ for all $\mathbf{y} \in G$ and all $\varepsilon < \varepsilon_0$. For example, G can be chosen as

$$G = \bigcup_{\mathbf{x} \in K} B(\mathbf{x}; \varepsilon_0),$$

where ε_0 is chosen less than half the distance from K to the complement of Ω (cf. Exercise 1(l) in Chapter 1). Note that $k_\varepsilon(\mathbf{x} - \mathbf{y})$ considered as a function of $\mathbf{y}$ has support in G for all $\mathbf{x} \in K$ and all $\varepsilon < \varepsilon_0$.

Let $g = f\chi_G$ denote the function on $\mathbf{R^n}$ obtained by extending f to be zero outside G. Then $g \in L^1(\mathbf{R^n})$ since $f \in L^1_{loc}(\Omega)$ by hypothesis. Now define

$$g_\varepsilon(\mathbf{x}) = (g * k_\varepsilon)(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n},\ \varepsilon > 0.$$

Then $g_\varepsilon \in C_0^\infty(\mathbf{R^n})$ by Theorem 9.3 and the comments after its proof. Moreover, since g vanishes outside G and k_ε has support in $B(\mathbf{0}; \varepsilon)$, it follows that g_ε has support in Ω if $\varepsilon < \varepsilon_0$.

By Theorems 9.6 and 9.13, $g_\varepsilon \to g$ in $L^1(\mathbf{R^n})$ norm and pointwise a.e. in $\mathbf{R^n}$ as $\varepsilon \to 0$. Hence, $g_\varepsilon \to f$ in $L^1(K)$ and pointwise a.e. in K.

Next, by Theorem 9.3, for all $\mathbf{x} \in \mathbf{R^n}$, all $\varepsilon > 0$ and every $i = 1, \ldots, n$, we have

$$\begin{aligned}
\frac{\partial g_\varepsilon}{\partial x_i}(\mathbf{x}) &= \int_{\mathbf{R^n}} g(\mathbf{y}) \frac{\partial}{\partial x_i} \big\{k_\varepsilon(\mathbf{x} - \mathbf{y})\big\}\, d\mathbf{y} \\
&= -\int_{\mathbf{R^n}} g(\mathbf{y}) \frac{\partial}{\partial y_i} \big\{k_\varepsilon(\mathbf{x} - \mathbf{y})\big\}\, d\mathbf{y} \\
&= -\int_G f(\mathbf{y}) \frac{\partial}{\partial y_i} \big\{k_\varepsilon(\mathbf{x} - \mathbf{y})\big\}\, d\mathbf{y}.
\end{aligned}$$

Considered as a function of $\mathbf{y}$, $k_\varepsilon(\mathbf{x}-\mathbf{y})$ has support in $B(\mathbf{x};\varepsilon)$ and therefore has support in G if $\mathbf{x} \in K$ and $\varepsilon < \varepsilon_0$. Since $\partial f/\partial x_i$ is the weak partial derivative of f with respect to x_i in Ω, it is also the weak partial derivative of f with respect to x_i in G. Consequently,

$$\frac{\partial g_\varepsilon}{\partial x_i}(\mathbf{x}) = \int_G \frac{\partial f}{\partial y_i}(\mathbf{y}) k_\varepsilon(\mathbf{x}-\mathbf{y})\, d\mathbf{y} \quad \text{if } \mathbf{x} \in K \text{ and } \varepsilon < \varepsilon_0.$$

The last integral equals

$$\left[\left(\frac{\partial f}{\partial x_i}\chi_G\right) * k_\varepsilon\right](\mathbf{x}),$$

which converges to $\left(\frac{\partial f}{\partial x_i}\chi_G\right)(\mathbf{x})$ in $L^1(\mathbf{R^n})$ and pointwise a.e. in $\mathbf{R^n}$ as $\varepsilon \to 0$. Restricting $\mathbf{x}$ to K, we obtain that $\partial g_\varepsilon/\partial x_i \to \partial f/\partial x_i$ in $L^1(K)$ and pointwise a.e. in K. The theorem now follows by choosing $f_j = g_{\varepsilon_j}$ for any sequence $\{\varepsilon_j\}_{j=1}^\infty \to 0$ such that $\varepsilon_j < \varepsilon_0$ for all j.

The proof of Theorem 15.8 can be modified to yield a single sequence $\{f_j\}$ that has the same properties as in Theorem 15.8 for every compact set $K \subset \Omega$. Indeed, we have the following result whose proof is left to the reader (see Exercise 5).

Theorem 15.9 *Under the same hypotheses on Ω and f as in Theorem 15.8, there is a sequence $\{f_j\}$ that satisfies the three properties in the conclusion of Theorem 15.8 for every compact set $K \subset \Omega$.*

It follows immediately that the sequence $\{f_j\}$ in Theorem 15.9 has the pointwise convergence properties $f_j \to f$ and $\nabla f_j \to \nabla f$ a.e. in Ω.

We arrive naturally at the definition of the *Sobolev space* $W^{1,p}(\Omega)$ by considering functions f that have a weak gradient in Ω such that both f and $|\nabla f|$ belong to $L^p(\Omega)$, that is, if Ω is an open set in $\mathbf{R^n}$ and $1 \le p \le \infty$, then $W^{1,p}(\Omega)$ is defined by

$$W^{1,p}(\Omega) = \left\{f \in L^p(\Omega) : f \text{ has a weak gradient in } \Omega \text{ satisfying } |\nabla f| \in L^p(\Omega)\right\}. \tag{15.10}$$

The purpose of the first superscript 1 in the notation $W^{1,p}$ is to indicate that only first-order derivatives enter the definition. We will not consider Sobolev spaces involving weak derivatives of order more than 1, although these spaces have important applications. In fact, we only consider a few

aspects of the spaces $W^{1,p}(\Omega)$, and then our primary concerns are the cases when Ω is either a ball or the entire space $\mathbf{R^n}$.

One fact about (15.10) deserves emphasis: in order that $f \in W^{1,p}(\Omega)$, the definition requires that f and $|\nabla f|$ belong to $L^p(\Omega)$ and not just to $L^p_{loc}(\Omega)$. On the other hand, in case Ω is a ball B, we will see in Section 15.3 that $f \in W^{1,p}(B)$ if $f \in L^1_{loc}(B)$ and f has a weak gradient in B satisfying $|\nabla f| \in L^p(B)$. In other words, in the case of a ball B, the requirement in (15.10) that $f \in L^p(B)$ can be replaced by assuming only that $f \in L^1_{loc}(B)$.

Some basic properties of Sobolev spaces are given in the exercises. For example (see Exercise 7), if Ω is any open set in $\mathbf{R^n}$ and $1 \le p \le \infty$, then $W^{1,p}(\Omega)$ is a Banach space with respect to the norm

$$\|f\|_{W^{1,p}(\Omega)} = \|f\|_{L^p(\Omega)} + \sum_{i=1}^{n} \left\| \frac{\partial f}{\partial x_i} \right\|_{L^p(\Omega)}. \tag{15.11}$$

An equivalent norm is

$$\|f\|_{L^p(\Omega)} + \left\| \left(\sum_{i=1}^{n} \left| \frac{\partial f}{\partial x_i} \right|^2 \right)^{1/2} \right\|_{L^p(\Omega)} = \|f\|_{L^p(\Omega)} + \|\nabla f\|_{L^p(\Omega)}.$$

Moreover, $W^{1,p}(\Omega)$ is separable if $1 \le p < \infty$.

In case $1 \le p < \infty$, another equivalent norm is

$$\left(\|f\|^p_{L^p(\Omega)} + \sum_{i=1}^{n} \left\| \frac{\partial f}{\partial x_i} \right\|^p_{L^p(\Omega)} \right)^{1/p},$$

and using this norm when $p = 2$ makes $W^{1,2}(\Omega)$ a Hilbert space with inner product

$$(f, g) = \int_\Omega fg \, d\mathbf{x} + \int_\Omega \nabla f \cdot \nabla g \, d\mathbf{x}. \tag{15.12}$$

See Exercises 11 and 12 for versions of the product rule and the chain rule in $W^{1,p}(\Omega)$.

Next, we state a way that functions in $W^{1,p}(\Omega)$ can be approximated by smooth functions when p is finite.

Theorem 15.13 *Let $1 \le p < \infty$ and Ω be an open set in $\mathbf{R^n}$. If $f \in W^{1,p}(\Omega)$, there is a sequence $\{f_j\}_{j=1}^{\infty}$ of functions on $\mathbf{R^n}$ such that*

(i) $f_j \in C_0^\infty(\mathbf{R^n})$ *for all j,*

(ii) $f_j \to f$ *a.e. in* Ω *and in* $L^p(\Omega)$ *norm as* $j \to \infty$,

(iii) $\nabla f_j \to \nabla f$ *a.e. in* Ω *and in* $L^p(K)$ *norm as* $j \to \infty$ *for every compact set* $K \subset \Omega$.

Moreover, in case $\Omega = \mathbf{R^n}$ *and* $f \in W^{1,p}(\mathbf{R^n})$ *for some* p *with* $1 \le p < \infty$, *the sequence* $\{f_j\}$ *can be chosen to converge to* f *in* $W^{1,p}(\mathbf{R^n})$ *norm, that is, so that* $f_j \to f$ *in* $L^p(\mathbf{R^n})$ *and* $\nabla f_j \to \nabla f$ *in* $L^p(\mathbf{R^n})$.

Here the terminology $\nabla f_j \to \nabla f$ *in* L^p *norm* means that $\partial f_j/\partial x_i \to \partial f/\partial x_i$ in L^p norm as $j \to \infty$ for every $i = 1, \dots, n$.

Proof. The proof is similar to that of Theorem 15.8, now using the L^p versions of Theorems 9.6 and 9.13. We will leave some details for the reader to check. Fix p, Ω, and f satisfying the hypotheses, and define $g = f\chi_\Omega$ by extending f to be zero outside Ω. Note that g belongs to $L^p(\mathbf{R^n})$ but may not have compact support and that $g = f$ everywhere if $\Omega = \mathbf{R^n}$. For $\varepsilon > 0$, define $g_\varepsilon = g * k_\varepsilon$ where k_ε is the same approximation of the identity used in the proofs of Theorems 15.7 and 15.8.

Then $g_\varepsilon \in C^\infty(\mathbf{R^n})$, although it may not have compact support, and $g_\varepsilon \to g$ in $L^p(\mathbf{R^n})$ and pointwise a.e. in $\mathbf{R^n}$. Hence, $g_\varepsilon \to f$ in $L^p(\Omega)$ and pointwise a.e. in Ω.

The proof that $\nabla g_\varepsilon \to \nabla f$ pointwise a.e. in Ω and also in $L^p(K)$ norm for every compact set K in Ω is left to the reader, as is the proof that if $\Omega = \mathbf{R^n}$, then the sequence can be chosen with $\nabla g_\varepsilon \to \nabla f$ in $L^p(\mathbf{R^n})$ norm.

Finally, whether $\Omega = \mathbf{R^n}$ or not, let $\eta(\mathbf{x})$ be a smooth cutoff function on $\mathbf{R^n}$ such that $\eta \in C_0^\infty(\mathbf{R^n})$ and

$$\eta(\mathbf{x}) = \begin{cases} 1 & \text{if } |\mathbf{x}| \le 1 \\ 0 & \text{if } |\mathbf{x}| \ge 2. \end{cases}$$

Then for any sequence $\varepsilon_j \to 0$, the functions $\{f_j\}$ defined on $\mathbf{R^n}$ by $f_j(\mathbf{x}) = \eta(\varepsilon_j \mathbf{x}) g_{\varepsilon_j}(\mathbf{x})$ have compact support in $\mathbf{R^n}$ and satisfy all the properties stated in the theorem. For example, let us show why $\nabla f_j \to \nabla f$ in $L^p(K)$ for every compact set $K \subset \Omega$. We compute

$$\nabla f_j(\mathbf{x}) = \varepsilon_j (\nabla\eta)(\varepsilon_j \mathbf{x}) g_{\varepsilon_j}(\mathbf{x}) + \eta(\varepsilon_j \mathbf{x}) \nabla g_{\varepsilon_j}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R^n}.$$

Therefore, if $\mathbf{x} \in \Omega$,

$$\begin{aligned} \left|\nabla f_j(\mathbf{x}) - \nabla f(\mathbf{x})\right| &\le \varepsilon_j \|\nabla\eta\|_{L^\infty(\mathbf{R^n})} |g_{\varepsilon_j}(\mathbf{x})| + \eta(\varepsilon_j \mathbf{x}) \left|\nabla g_{\varepsilon_j}(\mathbf{x}) - \nabla f(\mathbf{x})\right| \\ &\quad + \big(1 - \eta(\varepsilon_j \mathbf{x})\big) |\nabla f(\mathbf{x})| := \alpha_j(\mathbf{x}) + \beta_j(\mathbf{x}) + \gamma_j(\mathbf{x}). \end{aligned}$$

First consider α_j. Denote

$$M = \|\nabla\eta\|_{L^\infty(\mathbf{R^n})}\Big(\sup_j \|g_{\varepsilon_j}\|_{L^p(\mathbf{R^n})}\Big).$$

Then M is finite and

$$\|\alpha_j\|_{L^p(\Omega)} \le M\varepsilon_j \to 0 \quad \text{as } j \to \infty.$$

We estimate the norm of γ_j by using the facts that $\eta_j(\varepsilon_j\mathbf{x}) = 1$ if $|\varepsilon_j\mathbf{x}| \le 1$ and $|\nabla f| \in L^p(\Omega)$ with p finite:

$$\|\gamma_j\|_{L^p(\Omega)} \le \left(\int\limits_{\mathbf{x}\in\Omega;|\mathbf{x}|>\varepsilon_j^{-1}} |\nabla f|^p\, d\mathbf{x}\right)^{1/p} \to 0 \quad \text{as } j \to \infty.$$

Next, fix a compact set $K \subset \Omega$. Since $\nabla g_{\varepsilon_j} \to \nabla f$ in $L^p(K)$, then

$$\|\beta_j\|_{L^p(K)} \le \|\nabla g_{\varepsilon_j} - \nabla f\|_{L^p(K)} \to 0 \quad \text{as } j \to \infty.$$

Collecting estimates, it follows that $\nabla f_j \to \nabla f$ in $L^p(K)$ for every compact set $K \subset \Omega$.

Next, suppose that $\Omega = \mathbf{R^n}$ and recall that $\nabla g_{\varepsilon_j} \to \nabla f$ in $L^p(\mathbf{R^n})$ in this case. Hence,

$$\|\beta_j\|_{L^p(\mathbf{R^n})} \le \|\nabla g_{\varepsilon_j} - \nabla f\|_{L^p(\mathbf{R^n})} \to 0 \quad \text{as } j \to \infty.$$

Finally, the arguments proving that both $\|\alpha_j\|_{L^p(\mathbf{R^n})}$ and $\|\gamma_j\|_{L^p(\mathbf{R^n})}$ tend to 0 were already included above. Thus, in case $\Omega = \mathbf{R^n}$, we obtain that $\nabla f_j \to \nabla f$ in $L^p(\mathbf{R^n})$.

The remaining details of the proof are left to the reader.

In some situations, fewer properties of the approximating sequence $\{f_j\}$ are needed than the ones listed in Theorems 15.8 or 15.13. The following simple variant of Theorem 15.13 will be useful in the proof of Theorem 15.45.

Theorem 15.14 *Let f have a weak gradient in $\mathbf{R^n}$ that satisfies $|\nabla f| \in L^p(\mathbf{R^n})$ for some p, $1 \le p < \infty$. If either $f \in L^r(\mathbf{R^n})$ for some r with $1 \le r < \infty$ or $\lim_{|\mathbf{x}|\to\infty} f(\mathbf{x}) = 0$, then there is a sequence $\{f_j\}_{j=1}^\infty \subset C^\infty(\mathbf{R^n})$ with $\lim_{|\mathbf{x}|\to\infty} f_j(\mathbf{x}) = 0$ for every j such that $f_j \to f$ pointwise a.e. in $\mathbf{R^n}$ and $\nabla f_j \to \nabla f$ in $L^p(\mathbf{R^n})$ norm as $j \to \infty$.*

Proof. Let k_ε be as usual and let $f_j = f * k_{\varepsilon_j}$ for any sequence $\varepsilon_j \to 0$. Then $f_j \in C^\infty(\mathbf{R^n})$ for all j. Also, $f_j \to f$ a.e. in $\mathbf{R^n}$ since $f \in L^1_{loc}(\mathbf{R^n})$, and $\nabla f_j \to \nabla f$ in $L^p(\mathbf{R^n})$ as usual.

It remains to show that each $f_j(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$. First, suppose that $f(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$. Then, for every $\varepsilon > 0$,

$$\sup_{B(\mathbf{x};\varepsilon)} |f| \to 0 \quad \text{as} \quad |\mathbf{x}| \to \infty,$$

and we are done since

$$|(f * k_\varepsilon)(\mathbf{x})| \le \left(\sup_{B(\mathbf{x};\varepsilon)} |f| \right) \|k\|_{L^1(\mathbf{R^n})}.$$

Finally, assuming instead that $f \in L^r(\mathbf{R^n})$ for some r with $1 \le r < \infty$, we have

$$\begin{aligned} |(f * k_\varepsilon)(\mathbf{x})| &= \left| \int_{|\mathbf{x}-\mathbf{y}|<\varepsilon} f(\mathbf{y}) k_\varepsilon(\mathbf{x}-\mathbf{y})\, d\mathbf{y} \right| \\ &\le \left(\int_{|\mathbf{x}-\mathbf{y}|<\varepsilon} |f(\mathbf{y})|^r\, d\mathbf{y} \right)^{1/r} \|k_\varepsilon\|_{L^{r'}(\mathbf{R^n})}, \end{aligned}$$

$1/r + 1/r' = 1$. For each ε, the r'-norm of k_ε is independent of $\mathbf{x}$ and finite:

$$\|k_\varepsilon\|_{L^{r'}(\mathbf{R^n})} \le \|k_\varepsilon\|_{L^\infty(\mathbf{R^n})} |B(0;\varepsilon)|^{1/r'}.$$

Hence, since

$$\left(\int_{|\mathbf{x}-\mathbf{y}|<\varepsilon} |f(\mathbf{y})|^r\, d\mathbf{y} \right)^{1/r} \to 0 \quad \text{as } |\mathbf{x}| \to \infty,$$

where we use the fact that r is finite, we are again done, and the proof is complete.

In Theorem 15.14, either of the extra assumptions $\lim_{|\mathbf{x}|\to\infty} f(\mathbf{x}) = 0$ or $f \in L^r(\mathbf{R^n})$ for some r with $1 \le r < \infty$ can be replaced by the weaker condition that $\lim_{|\mathbf{x}|\to\infty} \int_{B(\mathbf{x};1)} |f|\, d\mathbf{y} = 0$ without affecting the conclusion of the theorem. Verification is left as an exercise.

In passing, we prove a result about extending functions with weak derivatives and compact support.

Theorem 15.15 *Suppose that f has compact support in an open set $\Omega \subset \mathbf{R^n}$ and weak partial derivative $\partial f/\partial x_i$ in Ω for some $i = 1, \ldots, n$. Extend f and $\partial f/\partial x_i$ to $\mathbf{R^n}$ by setting them equal to 0 outside Ω. Then f has a weak partial derivative $\partial f/\partial x_i$ in $\mathbf{R^n}$.*

Proof. Fix f and Ω satisfying the hypothesis. Choose a function $\psi \in C_0^\infty(\Omega)$ such that $\psi = 1$ on the support of f; cf. Exercise 5 of Chapter 9. Then for any test function $\varphi \in C_0^\infty(\mathbf{R^n})$, we have $\varphi\psi \in C_0^\infty(\Omega)$, and consequently,

$$\int_\Omega f \frac{\partial}{\partial x_i}(\varphi\psi)\, d\mathbf{x} = -\int_\Omega \frac{\partial f}{\partial x_i}\, \varphi\psi\, d\mathbf{x}.$$

Since f and $\partial f/\partial x_i$ are defined to be 0 outside Ω, we may think of both integrals as extended over $\mathbf{R^n}$. Furthermore, by Exercise 13, the set where $\partial f/\partial x_i \neq 0$ is contained in the union of the support of f and a subset of Ω of measure 0. Hence, since $\psi = 1$ on the support of f, the integrals are unchanged if ψ is replaced by 1 in their integrands, and we obtain

$$\int_{\mathbf{R^n}} f \frac{\partial \varphi}{\partial x_i}\, d\mathbf{x} = -\int_{\mathbf{R^n}} \frac{\partial f}{\partial x_i}\varphi\, d\mathbf{x}.$$

Of course, the extended functions f and $\frac{\partial f}{\partial x_i}$ belong to $L^1_{loc}(\mathbf{R^n})$, and the proof is complete.

15.3 Poincaré–Sobolev Estimates

In this section, we derive a variety of first-order Poincaré–Sobolev estimates for functions with a weak gradient. Results in case $n = 1$ (see Exercise 20) are easy consequences of Theorem 15.6 and the representation of an absolutely continuous function as the integral of its derivative, and so we generally assume that $n > 1$.

We begin by extending the subrepresentation inequalities in Chapter 14 to functions with a weak gradient. Poincaré–Sobolev estimates for $n > 1$ will follow by combining these inequalities with various results about fractional integrals.

Let us first extend Theorem 14.2. As usual, all balls $B \subset \mathbf{R^n}$ are open by definition, and if f has a weak gradient ∇f in a ball B, then f and $|\nabla f|$ must belong to $L^1_{loc}(B)$, but neither is assumed to belong to $L^1(B)$.

Theorem 15.16 *If f has weak gradient ∇f in a ball $B \subset \mathbf{R^n}$, then*

$$\frac{1}{|B|}\int_B |f(\mathbf{x}) - f(\mathbf{y})|\,d\mathbf{y} \le c_n \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}}d\mathbf{y} \quad \textit{for a.e. } \mathbf{x} \in B. \tag{15.17}$$

If in addition $f \in L^1(B)$, then

$$|f(\mathbf{x}) - f_B| \le c_n \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}}d\mathbf{y} \quad \textit{for a.e. } \mathbf{x} \in B, \tag{15.18}$$

where $f_B = |B|^{-1}\int_B f(\mathbf{y})\,d\mathbf{y}$. The constant c_n depends only on n.

Proof. Let f and B satisfy the hypothesis. Clearly, (15.18) follows from (15.17) if $f \in L^1(B)$. To prove (15.17), we will use Theorems 14.2 and 15.8. Choose balls $D \nearrow B$ concentric with B and with closures in B. It suffices to show that there is a constant c_n depending only on n such that

$$\frac{1}{|D|}\int_D |f(\mathbf{x}) - f(\mathbf{y})|\,d\mathbf{y} \le c_n \int_D \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^{n-1}}\,d\mathbf{y} \quad \text{for a.e. } \mathbf{x} \in D, \tag{15.19}$$

since (15.17) then follows by passing to the limit. Given such a ball D, choose $\{f_j\}$ as in Theorem 15.8 corresponding to the compact set $K = \overline{D}$. Since each f_j is infinitely differentiable in D, Theorem 14.2 gives

$$\frac{1}{|D|}\int_D |f_j(\mathbf{x}) - f_j(\mathbf{y})|\,d\mathbf{y} \le c_n I_1(|\nabla f_j|\chi_D)(\mathbf{x}), \quad \mathbf{x} \in D,$$

for every j, where I_1 denotes the fractional integral operator of order 1 and c_n depends only on n. Here, as usual, $|\nabla f_j|\chi_D$ denotes the extension of $|\nabla f_j|$ to $\mathbf{R^n}$ by setting it equal to 0 outside D. Since $f_j \to f$ a.e. in D, we obtain by applying Fatou's lemma to the left side that

$$\frac{1}{|D|}\int_D |f(\mathbf{x}) - f(\mathbf{y})|\,d\mathbf{y} \le c_n \liminf_{j\to\infty} I_1(|\nabla f_j|\chi_D)(\mathbf{x}) \quad \text{for a.e. } \mathbf{x} \in D.$$

To prove (15.19), it is then enough to show that

$$\liminf_{j\to\infty} I_1(|\nabla f_j|\chi_D) \le I_1(|\nabla f|\chi_D) \quad \text{a.e. in } D.$$

By Theorem 15.8, $\nabla f_j \to \nabla f$ in $L^1(D)$ norm, and therefore $|\nabla f_j|\chi_D \to |\nabla f|\chi_D$ in $L^1(\mathbf{R^n})$ norm. The weak type estimate in Theorem 14.37(b) then implies that

$I_1(|\nabla f_j|\chi_D)$ converges in measure on $\mathbf{R^n}$ to $I_1(|\nabla f|\chi_D)$. Consequently, there is a subsequence $\{j_\ell\} \to \infty$ such that $I_1(|\nabla f_{j_\ell}|\chi_D) \to I_1(|\nabla f|\chi_D)$ pointwise a.e. in $\mathbf{R^n}$. Hence, the desired inequality is true a.e. in $\mathbf{R^n}$ and therefore also a.e. in D. This proves (15.19) and Theorem 15.16.

Corollary 15.20 *If f has a weak gradient equal to zero in a ball B, then f is constant a.e. in B.*

Proof. This follows from (15.17) since if $|\nabla f| = 0$ in B, then the integral on the right side of (15.17) vanishes for all $\mathbf{x} \in B$, and therefore $|f(\mathbf{x}) - f(\mathbf{y})| = 0$ for a.e. $\mathbf{x}, \mathbf{y} \in B$.

The next result is an analogue of Corollary 14.6.

Corollary 15.21 *Let f have a weak gradient ∇f in a ball $B \subset \mathbf{R^n}$.*

(i) *If $f = 0$ in a measurable set $E \subset B$ satisfying $|E| \geq \gamma|B|$ for some constant $\gamma > 0$, then*

$$|f(\mathbf{x})| \leq \frac{c_n}{\gamma} \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^{n-1}}\, d\mathbf{y} \quad \textit{for a.e. } \mathbf{x} \in B.$$

(ii) *If f has compact support in B, then*

$$|f(\mathbf{x})| \leq c_n \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^{n-1}}\, d\mathbf{y} \quad \textit{for a.e. } \mathbf{x} \in B.$$

In both parts, the constant c_n depends only on n.

Proof. The proof of part (i) is omitted since it is essentially identical to the corresponding proof in Corollary 14.6.

To prove (ii), let f have compact support and weak gradient ∇f in a ball B. Extend f and ∇f to $\mathbf{R^n}$ by setting them equal to zero outside B. Let B^* be the ball of radius $2r(B)$ concentric with B. By Theorem 15.15, ∇f is the weak gradient of f in B^*. Also, by Exercise 13, $|\nabla f| = 0$ outside B. By part (i) of this corollary applied to B^* and its subset $E = B^* - B$, we then obtain

$$|f(\mathbf{x})| \leq c_n \int_{B^*} \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^{n-1}}\, d\mathbf{y} = c_n \int_B \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^{n-1}}\, d\mathbf{y}$$

for a.e. $\mathbf{x} \in B$, which completes the proof.

There is also an analogue of Corollary 14.7. In stating it, we assume f has a weak gradient in $\mathbf{R^n}$ and satisfies the extra condition that there is a sequence $\{B_i\}_{i=1}^{\infty}$ of balls such that $B_i \nearrow \mathbf{R^n}$ and $f_{B_i} \to 0$ as $i \to \infty$. As in Chapter 14, typical situations when this holds are (i) $f \in L^r(\mathbf{R^n})$ for some r with $1 \le r < \infty$, or (ii) $f \in L^r_{loc}(\mathbf{R^n})$ for some r with $1 \le r \le \infty$ and $\lim_{|\mathbf{x}|\to\infty} f(\mathbf{x}) = 0$.

Corollary 15.22 *Let f have weak gradient ∇f in $\mathbf{R^n}$ and suppose that $f_B \to 0$ for some sequence of balls $B \nearrow \mathbf{R^n}$. Then there is a constant c_n depending only on n such that*

$$|f(\mathbf{x})| \le c_n I_1(|\nabla f|)(\mathbf{x}) \quad \textit{for a.e. } \mathbf{x} \in \mathbf{R^n},$$

where I_1 is the fractional integral operator of order 1. In particular, the conclusion holds if $f \in W^{1,p}(\mathbf{R^n})$ for any p with $1 \le p < \infty$.

Proof. Fix a function f with a weak gradient in $\mathbf{R^n}$ and $f_B \to 0$ for some sequence of balls B increasing to $\mathbf{R^n}$. For any B in the sequence, since f is locally integrable in $\mathbf{R^n}$, we have that $f \in L^1(B)$, and Theorem 15.16 (see (15.18)) yields

$$|f(\mathbf{x}) - f_B| \le c_n I_1(|\nabla f|)(\mathbf{x}) \quad \text{for a.e. } \mathbf{x} \in B.$$

Now let $B \nearrow \mathbf{R^n}$. Since $f_B \to 0$, the result follows.

We now turn to the Poincaré–Sobolev inequalities themselves. Four separate ranges of p values will be considered: $1 < p < n$, $p = n$, $p > n$, and $p = 1$. The case $p = 1$ uses a more local subrepresentation formula than the ones derived so far. Our primary interest is again when $n > 1$. In fact, the range $1 < p < n$ does not exist if $n = 1$, and results when $p = n = 1$ and $p > n = 1$ are easy consequences of Theorem 15.6; see Exercise 20.

We begin with the range $1 < p < n$.

Theorem 15.23 *Let B be a ball in $\mathbf{R^n}$, $n > 1$, and let p and q satisfy $1 < p < n$ and $1/q = 1/p - 1/n$ (so that $q = pn/(n-p) > p$). If f is a locally integrable function on B with a weak gradient ∇f belonging to $L^p(B)$, then $f \in L^q(B)$ and*

$$\left(\int_B |f(\mathbf{x}) - f_B|^q d\mathbf{x}\right)^{1/q} \le c_{n,p} \left(\int_B |\nabla f(\mathbf{x})|^p d\mathbf{x}\right)^{1/p}, \tag{15.24}$$

where $f_B = |B|^{-1} \int_B f(\mathbf{x})\, d\mathbf{x}$ and the constant $c_{n,p}$ depends only on n and p.

If f also has compact support in B, then

$$\left(\int_B |f(\mathbf{x})|^q d\mathbf{x}\right)^{1/q} \le c_{n,p}\left(\int_B |\nabla f(\mathbf{x})|^p d\mathbf{x}\right)^{1/p}. \tag{15.25}$$

Some comments about Theorem 15.23 are given after its proof.

Proof. Fix B, p, and f satisfying the hypothesis of the first part of the theorem, and let $1/q = 1/p - 1/n$. Let us begin by showing that $f \in L^1(B)$. Theorem 14.37(a) in case $\alpha = 1$ gives

$$\|I_1(|\nabla f|\chi_B)\|_{L^q(\mathbf{R^n})} \le c_{n,p}\|(\nabla f)\chi_B\|_{L^p(\mathbf{R^n})} = c_{n,p}\|\nabla f\|_{L^p(B)} < \infty.$$

In particular, $I_1(|\nabla f|\chi_B)$ is finite a.e. in B. Since $f \in L^1_{loc}(B)$ by hypothesis, f is also finite a.e. in B. Furthermore, by (15.17),

$$\frac{1}{|B|}\int_B |f(\mathbf{x}) - f(\mathbf{y})|\, d\mathbf{y} \le c_n I_1(|\nabla f|\chi_B)(\mathbf{x}) \quad \text{a.e. in } B,$$

and therefore,

$$\frac{1}{|B|}\int_B |f(\mathbf{y})|\, d\mathbf{y} \le |f(\mathbf{x})| + c_n I_1(|\nabla f|\chi_B)(\mathbf{x}) \quad \text{a.e. in } B.$$

The fact that $f \in L^1(B)$ now follows by choosing a point $\mathbf{x} \in B$ such that both $f(\mathbf{x})$ and $I_1(|\nabla f|\chi_B)(\mathbf{x})$ are finite.

Hence, the average f_B is finite and (15.18) holds:

$$|f(\mathbf{x}) - f_B| \le c_n I_1(|\nabla f|\chi_B)(\mathbf{x}) \quad \text{a.e. in } B.$$

Taking L^q norms over B, we obtain

$$\begin{aligned}\|f - f_B\|_{L^q(B)} &\le c_n\|I_1(|\nabla f|\chi_B)\|_{L^q(B)}\\ &\le c_{n,p}\|\nabla f\|_{L^p(B)},\end{aligned}$$

as noted earlier. This proves (15.24). The fact that $f \in L^q(B)$ now follows immediately from Minkowski's inequality:

$$\begin{aligned}\|f\|_{L^q(B)} &\le \|f - f_B\|_{L^q(B)} + |f_B|\,|B|^{1/q}\\ &\le c_{n,p}\|\nabla f\|_{L^p(B)} + |f_B|\,|B|^{1/q} < \infty.\end{aligned}$$

This proves the first statement in Theorem 15.23.

If f also has compact support in B, a similar argument based on part (ii) of Corollary 15.21 yields (15.25). Theorem 15.23 is now proved.

We pause briefly in order to add several comments related to Theorem 15.23.

The proof of Theorem 15.23 fails if $p = 1$ since I_1 does not map $L^1(\mathbf{R^n})$ into $L^{n/(n-1)}(\mathbf{R^n})$; Theorem 14.37 provides only a weak type estimate for I_1 when $p = 1$. However, as will see in Theorem 15.37, Theorem 15.23 remains true if $p = 1$.

Inequality (15.24) is often called the *L^p, L^q Poincaré estimate for f and B*, and (15.25) for compactly supported f is called the *L^p, L^q Sobolev estimate for f*. The formula $1/q = 1/p - 1/n$ is called the *Poincaré–Sobolev dimensional balance formula* or simply the *balance formula*. Because of it and the fact that the measure of any ball B in $\mathbf{R^n}$ is a fixed multiple of $r(B)^n$, the L^p, L^q Poincaré estimate for B can be written in the equivalent normalized form

$$\left(\frac{1}{|B|}\int_B |f(\mathbf{x}) - f_B|^q d\mathbf{x}\right)^{1/q} \le c_{n,p} r(B) \left(\frac{1}{|B|}\int_B |\nabla f(\mathbf{x})|^p d\mathbf{x}\right)^{1/p}. \tag{15.26}$$

Similarly, the L^p, L^q Sobolev estimate can be written in the form

$$\left(\frac{1}{|B|}\int_B |f(\mathbf{x})|^q d\mathbf{x}\right)^{1/q} \le c_{n,p} r(B) \left(\frac{1}{|B|}\int_B |\nabla f(\mathbf{x})|^p d\mathbf{x}\right)^{1/p}. \tag{15.27}$$

A remarkable fact is that the hypothesis in the first part of Theorem 15.23 is nominally weaker than assuming f belongs to $W^{1,p}(B)$, because f is not assumed to belong to $L^p(B)$ or even to $L^1(B)$, but the conclusion is stronger since f turns out to belong to $L^q(B)$ for some $q > p$. In fact, as already shown in the proof of Theorem 15.23, under the same hypothesis as in the first part of Theorem 15.23, we have

$$\|f\|_{L^q(B)} \le c_{n,p}\|\nabla f\|_{L^p(B)} + |f_B|\,|B|^{1/q} < \infty. \tag{15.28}$$

From a heuristic point of view, the main content of Theorem 15.23 is that there is a gain in the order of integrability of f. An important corollary is that $W^{1,p}(B) \subset L^q(B)$ if $1 < p < n$ and $1/q = 1/p - 1/n$. See also Exercise 14.

Norm estimates over all of $\mathbf{R^n}$ for the range $1 < p < n$ can be derived easily from Corollary 15.22. The second part of the following result is usually referred to as the *first-order Sobolev embedding theorem* for $\mathbf{R^n}$ in case $1 < p < n$.

Theorem 15.29 *Let f have a weak gradient in $\mathbf{R^n}$ satisfying $|\nabla f| \in L^p(\mathbf{R^n})$ for some p with $1 < p < n$, and suppose that $f_B \to 0$ for some sequence of balls increasing*

to $\mathbf{R^n}$. *Let* $1/q = 1/p - 1/n$. *Then* $f \in L^q(\mathbf{R^n})$ *and*

$$\|f\|_{L^q(\mathbf{R^n})} \le c_{n,p}\|\nabla f\|_{L^p(\mathbf{R^n})}.$$

In particular, if $f \in W^{1,p}(\mathbf{R^n})$ *for some* p *with* $1 < p < n$, *then*

$$\|f\|_{L^q(\mathbf{R^n})} \le c_{n,p}\|\nabla f\|_{L^p(\mathbf{R^n})} \le c_{n,p}\|f\|_{W^{1,p}(\mathbf{R^n})},$$

$1/q = 1/p - 1/n$.

Proof. The first statement follows immediately by combining Corollary 15.22 and Theorem 14.37(a). Alternately, it can be derived by applying (15.24) to each ball in the sequence of balls in the hypothesis of the theorem.

The second statement is a corollary of the first one since $f_B \to 0$ for any sequence of balls $B \nearrow \mathbf{R^n}$ if $f \in L^p(\mathbf{R^n})$, $1 \le p < \infty$. See also Exercise 15.

Next, we consider the endpoint value $p = n$ that was omitted in Theorem 15.23. When $p = n$, the corresponding value of q in the balance formula $1/q = 1/p - 1/n$ is $q = \infty$. However, the function $\left|\log|\mathbf{x}|\right|^\beta$ belongs to $W^{1,n}(\{|\mathbf{x}| < 1/2\})$ if $n > 1$ and $0 < \beta < (n-1)/n$, but it is not bounded; see Exercise 16.

On the other hand, by using the estimates of Moser–Trudinger type in Chapter 14, we obtain the following result about local exponential integrability. The case $n = 2$ was studied by Pohozaev.

Theorem 15.30 *Let* B *be a ball in* $\mathbf{R^n}$, $n > 1$, *and* f *have a weak gradient in* B *satisfying* $|\nabla f| \in L^n(B)$. *There are positive constants* c_1 *and* c_2 *depending only on* n *such that*

$$\frac{1}{|B|}\int_B \exp\left\{c_1\left(\frac{|f(\mathbf{x}) - f_B|}{\|\nabla f\|_{L^n(B)}}\right)^{n/(n-1)}\right\} d\mathbf{x} \le c_2. \tag{15.31}$$

If f *is also supported in* B, *then (15.31) also holds with* f_B *replaced by* 0.

We have assumed here that $\|\nabla f\|_{L^n(B)} \ne 0$; otherwise, f is constant a.e. in B by Corollary 15.20.

Proof. Let B and f satisfy the first hypothesis. Since $|\nabla f| \in L^n(B)$, then $|\nabla f| \in L^p(B)$ for $1 \le p \le n$ by Hölder's inequality, and consequently $f \in L^1(B)$ by Theorem 15.23. The proof of (15.31) then follows immediately by combining (15.18) with the case $\alpha = 1$ of Theorem 14.40. If f also has compact support

in B, a similar argument based on the second part of Corollary 15.21 yields the second part of the theorem.

In case $n = 1$, a better result than (15.31) holds; see Exercise 20.

An immediate corollary of (15.31), under the same assumptions on f, is that for every r with $0 < r < \infty$,

$$\frac{1}{|B|} \int_B \left(\frac{|f(\mathbf{x}) - f_B|}{\|\nabla f\|_{L^n(B)}} \right)^r d\mathbf{x} \le c_{n,r},$$

or equivalently

$$\left(\frac{1}{|B|} \int_B |f(\mathbf{x}) - f_B|^r \, d\mathbf{x} \right)^{1/r} \le c_{n,r} \|\nabla f\|_{L^n(B)}, \quad 0 < r < \infty.$$

For a related global inequality, see Corollary 15.41.

Let us now consider the case $p > n$.

Theorem 15.32 (Morrey) *Let B be a ball in $\mathbf{R^n}$ and $n < p \le \infty$. If f has a weak gradient ∇f in B satisfying $|\nabla f| \in L^p(B)$, then after possible redefinition of f in a subset of B of measure zero, f is Hölder continuous of order $1-(n/p)$ on B. Moreover, there is a constant $c_{n,p}$ depending only on n and p such that*

$$|f(\mathbf{x}) - f(\mathbf{y})| \le c_{n,p} \|\nabla f\|_{L^p(B)} |\mathbf{x} - \mathbf{y}|^{1-\frac{n}{p}}, \quad \mathbf{x}, \mathbf{y} \in B.$$

The order $1 - (n/p)$ of Hölder continuity should be interpreted as 1 if $p = \infty$, that is, f is Lipschitz continuous on B if $p = \infty$.

Proof. Let B, p, and f satisfy the hypothesis. Then $|\nabla f| \in L^r(B)$ for $1 \le r \le p$, so $f \in L^1(B)$ by previous results. Therefore, using (15.18) and Hölder's inequality, for any ball $B_1 \subset B$ and a.e. $\mathbf{x} \in B_1$, we obtain

$$\begin{aligned} |f(\mathbf{x}) - f_{B_1}| &\le c_n \int_{B_1} \frac{|\nabla f(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^{n-1}} \, d\mathbf{y} \\ &\le c_n \, \|\nabla f\|_{L^p(B_1)} \left(\int_{B_1} \frac{1}{|\mathbf{x} - \mathbf{y}|^{(n-1)p'}} \, d\mathbf{y} \right)^{1/p'}, \end{aligned}$$

$1/p + 1/p' = 1$. Note that $(n-1)p' < n$ since $p > n$. Also, $B_1 \subset B(\mathbf{x}; 2r(B_1))$ for every $\mathbf{x} \in B_1$. Hence, if $\mathbf{x} \in B_1$, then

$$\begin{aligned}\left(\int_{B_1} \frac{1}{|\mathbf{x}-\mathbf{y}|^{(n-1)p'}}\, d\mathbf{y}\right)^{1/p'} &\le \left(\int_{|\mathbf{x}-\mathbf{y}|<2r(B_1)} \frac{1}{|\mathbf{x}-\mathbf{y}|^{(n-1)p'}}\, d\mathbf{y}\right)^{1/p'} \\ &= c_{n,p}\left(\int_0^{2r(B_1)} \frac{1}{r^{(n-1)p'}}\, r^{n-1}\, dr\right)^{1/p'} \\ &= c_{n,p} r(B_1)^{\frac{n-(n-1)p'}{p'}} = c_{n,p} r(B_1)^{1-\frac{n}{p}},\end{aligned}$$

where the exponent $1-\frac{n}{p}$ means 1 if $p=\infty$ ($p'=1$). Therefore, for a.e. $\mathbf{x} \in B_1$,

$$|f(\mathbf{x}) - f_{B_1}| \le c_{n,p}\|\nabla f\|_{L^p(B_1)} r(B_1)^{1-\frac{n}{p}} \le c_{n,p}\|\nabla f\|_{L^p(B)} r(B_1)^{1-\frac{n}{p}}.$$

The theorem now follows from Theorem 14.25.

Next, we record an immediate corollary of the case $p=\infty$ of Theorem 15.4 combined with Theorem 15.32. See also Exercise 10.

Corollary 15.33 *Let B be a ball in $\mathbf{R^n}$ and f be a locally integrable function in B. Then f has a weak gradient in $L^\infty(B)$ if and only if, after redefinition of f in at most a subset of B of measure zero, $f \in Lip(B)$. The Lipschitz constant of f and the norm $\|\nabla f\|_{L^\infty(B)}$ are equivalent in size, with constants of equivalence that are independent of f and B.*

Finally, we turn to the endpoint $p=1$. Results in this case are often referred to as Gagliardo, Nirenberg estimates. The main goal is to prove that the conclusion of Theorem 15.23 also holds when $p=1$, which is a surprising fact since the analogous strong type result for fractional integral operators is false. The local nature of ordinary differentiation helps lead to the following sharper version of the subrepresentation formula, which will play a key role when $p=1$.

Theorem 15.34 *Let B be a ball in $\mathbf{R^n}$ and let $f \in Lip_{loc}(B) \cap L^1(B)$. For $k=0,\pm1,\pm2,\ldots$, define sets $\Omega_k \subset B$ by*

$$\Omega_k = \left\{\mathbf{x} \in B : 2^k < |f(\mathbf{x}) - f_B| \le 2^{k+1}\right\}.$$

Then for all k,

$$|f(\mathbf{x}) - f_B| \le c_n I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) + \frac{4}{|B|}\int_B |f - f_B|\, d\mathbf{y}, \quad \mathbf{x} \in \Omega_k, \tag{15.35}$$

where I_1 is the fractional integral operator of order 1 *and c_n is a constant depending only on n.*

Before giving a proof, we note that the second term on the right side of (15.35) is bounded above by $8(|f|)_B$. It is also bounded above by

$$c_n \frac{r(B)}{|B|}\int_B |\nabla f|\, d\mathbf{y},$$

as can be seen by using the L^1, L^1 Poincaré estimate. Consequently, (15.35) can be restated with the second term on the right side replaced by either of these expressions. In our application, we will choose the second of the two.

To see why (15.35) is an improvement of (14.4) and (15.18), note that

$$\frac{r(B)}{|B|}\int_B |\nabla f|\, d\mathbf{y} \le c_n I_1(|\nabla f|\chi_B)(\mathbf{x}), \quad \mathbf{x} \in B,$$

due to the simple estimate $r(B)/|B| = c_n r(B)^{1-n} \le c_n|\mathbf{x} - \mathbf{y}|^{1-n}$ for all $\mathbf{x}, \mathbf{y} \in B$. Hence, the right side of (15.35) is bounded by a multiple of $I_1(|\nabla f|\chi_B)(\mathbf{x})$ if $\mathbf{x} \in B$.

Proof of Theorem 15.34. Fix B and f as in the hypothesis, and for each $k = 0, \pm 1, \pm 2, \ldots$, define a truncation g_k of $|f - f_B|$ on B by

$$g_k(\mathbf{x}) = \begin{cases} 2^{k-1} & \text{if } |f(\mathbf{x}) - f_B| \le 2^{k-1} \\ |f(\mathbf{x}) - f_B| & \text{if } 2^{k-1} < |f(\mathbf{x}) - f_B| \le 2^k, \text{ that is, if } \mathbf{x} \in \Omega_{k-1} \\ 2^k & \text{if } 2^k < |f(\mathbf{x}) - f_B|. \end{cases}$$

Then (see Exercise 17), each $g_k \in Lip_{loc}(B)$ and

$$|g_k(\mathbf{x}) - g_k(\mathbf{y})| \le |f(\mathbf{x}) - f(\mathbf{y})|, \quad \mathbf{x}, \mathbf{y} \in B.$$

Therefore, $|\nabla g_k| \le |\nabla f|$ a.e. in B, and we claim that

$$|\nabla g_k| \le |\nabla f|\,\chi_{\Omega_{k-1}} \quad \text{a.e. in } B. \tag{15.36}$$

To prove this, it is enough to show that $|\nabla g_k| = 0$ a.e. in $B - \Omega_{k-1}$. Let $\mathbf{x}_0 \in B - \Omega_{k-1}$. Then there are three possibilities: $|f(\mathbf{x}_0) - f_B| > 2^k$, $|f(\mathbf{x}_0) - f_B| < 2^{k-1}$, or $|f(\mathbf{x}_0) - f_B| = 2^{k-1}$. In either of the first two cases, since f is continuous, there is a neighborhood of $\mathbf{x}_0$ in which g_k is constant, namely, in which $g_k \equiv 2^k$ or $g_k \equiv 2^{k-1}$, respectively, and consequently $|\nabla g_k(\mathbf{x}_0)| = 0$ in either of the first two cases. On the other hand, if $|f(\mathbf{x}_0) - f_B| = 2^{k-1}$, then $g_k(\mathbf{x}_0) = 2^{k-1}$, and therefore, g_k has an absolute minimum at $\mathbf{x}_0$ (since $g_k \geq 2^{k-1}$ everywhere in B). Assuming as we may that $\nabla g_k(\mathbf{x}_0)$ exists, it follows that $|\nabla g_k(\mathbf{x}_0)| = 0$, and (15.36) is verified.

Note that g_k is integrable, even bounded, in B. For all $\mathbf{x} \in B$,

$$\begin{aligned} g_k(\mathbf{x}) &= [g_k(\mathbf{x}) - (g_k)_B] + (g_k)_B \\ &\leq c_n\, I_1(|\nabla g_k|\chi_B)(\mathbf{x}) + (g_k)_B \\ &\leq c_n\, I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) + (g_k)_B, \end{aligned}$$

where we have applied the subrepresentation formula to the locally Lipschitz function g_k and used (15.36). Also, $g_k \leq 2^{k-1} + |f - f_B|$ in B, and hence

$$(g_k)_B \leq 2^{k-1} + \frac{1}{|B|} \int_B |f - f_B|\, d\mathbf{y}.$$

Therefore, if $\mathbf{x} \in B$,

$$g_k(\mathbf{x}) \leq c_n I_1\left(|\nabla f|\chi_{\Omega_{k-1}}\right)(\mathbf{x}) + 2^{k-1} + \frac{1}{|B|} \int_B |f - f_B|\, d\mathbf{y}.$$

Restricting $\mathbf{x}$ to Ω_k, where we have $g_k(\mathbf{x}) = 2^k$, gives

$$2^k \leq c_n\, I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) + 2^{k-1} + \frac{1}{|B|} \int_B |f - f_B|\, d\mathbf{y},$$

and by then subtracting 2^{k-1} from both sides, it follows that

$$2^k \leq c_n\, I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) + \frac{2}{|B|} \int_B |f - f_B|\, d\mathbf{y}, \quad \mathbf{x} \in \Omega_k.$$

Combining this with the fact that $|f - f_B| \leq 2^{k+1}$ on Ω_k, we obtain the desired estimate

$$|f(\mathbf{x}) - f_B| \leq c_n\, I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) + \frac{4}{|B|} \int_B |f - f_B|\, d\mathbf{y}, \quad \mathbf{x} \in \Omega_k.$$

We can now extend Theorems 15.23 and 15.29 to $p = 1$.

Theorem 15.37

(i) *Let B be a ball in $\mathbf{R^n}$ and f be a function with a weak gradient ∇f in B that satisfies $|\nabla f| \in L^1(B)$. Then $f \in L^{n/(n-1)}(B)$ and*

$$\left(\int_B |f - f_B|^{n/(n-1)}\, d\mathbf{x}\right)^{(n-1)/n} \le c_n \int_B |\nabla f|\, d\mathbf{x}. \tag{15.38}$$

(ii) *Let f have a weak gradient ∇f in $\mathbf{R^n}$ satisfing $|\nabla f| \in L^1(\mathbf{R^n})$. If $f_B \to 0$ for some sequence of balls $B \nearrow \mathbf{R^n}$, then $f \in L^{n/(n-1)}(\mathbf{R^n})$ and*

$$\|f\|_{L^{n/(n-1)}(\mathbf{R^n})} \le c_n \|\nabla f\|_{L^1(\mathbf{R^n})}. \tag{15.39}$$

In particular, (15.39) holds if $f \in W^{1,1}(\mathbf{R^n})$. Thus, it holds if f has compact support and a weak gradient in $\mathbf{R^n}$ satisfing $|\nabla f| \in L^1(\mathbf{R^n})$.

The constants c_n depend only on n.

Once again, there are normalized versions: (15.38) can be rewritten as

$$\left(\frac{1}{|B|}\int_B |f - f_B|^{n/(n-1)}\, d\mathbf{x}\right)^{(n-1)/n} \le c_n \frac{r(B)}{|B|}\int_B |\nabla f|\, d\mathbf{x},$$

and if f has compact support in B, (15.39) is the same as

$$\left(\frac{1}{|B|}\int_B |f|^{n/(n-1)}\, d\mathbf{x}\right)^{(n-1)/n} \le c_n \frac{r(B)}{|B|}\int_B |\nabla f|\, d\mathbf{x}.$$

Proof. It is enough to prove part (i) since part (ii) follows from it by applying (15.38) to the sequence of balls in the hypothesis of part (ii) and using Fatou's lemma.

Fix a ball B and let f have weak gradient in B with $|\nabla f| \in L^1(B)$. Then $f \in L^1(B)$ by an argument like the one at the beginning of the proof of Theorem 15.23, except that the finiteness of $I_1(|\nabla f|\chi_B)$ a.e. in B now follows from the weak type estimate in Theorem 14.37(b) since $|\nabla f| \in L^1(B)$.

Assuming also that $f \in C^\infty(B)$, or even also just that $f \in Lip_{loc}(B)$, we can apply Theorem 15.34. Thus, let

$$\Omega_k = \left\{\mathbf{x} \in B : 2^k < |f(\mathbf{x}) - f_B| \le 2^{k+1}\right\}, \quad k = 0, \pm 1, \ldots.$$

Recall from the discussion following the statement of Theorem 15.34, that is, from the discussion about the size of the second term on the right side of (15.35), that

$$|f(\mathbf{x}) - f_B| \le c_n I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) + c_n \frac{r(B)}{|B|} \int_B |\nabla f|\, d\mathbf{y}, \quad \mathbf{x} \in \Omega_k. \tag{15.40}$$

Denote $q = n/(n-1)$. Since B is the disjoint union of the Ω_k, we have

$$\begin{aligned}\int_B |f - f_B|^q &= \sum_{k=-\infty}^{\infty} \int_{\Omega_k} |f - f_B|^q \\ &\le \sum_{k=-\infty}^{\infty} 2^{(k+1)q}|\Omega_k| = \sum_{k\le N} + \sum_{k\ge N+1} = S_1 + S_2,\end{aligned}$$

say, where N is chosen such that the second term on the right of (15.40) satisfies

$$2^{N-1} < c_n \frac{r(B)}{|B|} \int_B |\nabla f| \le 2^N.$$

Then

$$\begin{aligned}S_1 &= \sum_{k\le N} 2^{(k+1)q}|\Omega_k| \le 2^{(N+1)q}|B| \\ &= 2^{2q}\, 2^{(N-1)q}\, |B| \le 2^{2q} \left(c_n \frac{r(B)}{|B|} \int_B |\nabla f| \right)^q |B| \\ &= c_n \left(\int_B |\nabla f| \right)^q \quad \text{since } q = n/(n-1).\end{aligned}$$

In order to estimate S_2, note that for all $\mathbf{x} \in \Omega_k$, the choice of N and (15.40) imply that

$$2^k < |f(\mathbf{x}) - f_B| \le c_n I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) + 2^N.$$

If $k \ge N+1$, then by subtracting 2^N from both sides and noting that $2^{k-1} \le 2^k - 2^N$, we obtain

$$I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) > \frac{2^{k-1}}{c_n}, \quad \mathbf{x} \in \Omega_k,\ k \ge N+1.$$

Equivalently, if $k \geq N+1$, then

$$\Omega_k \subset \left\{\mathbf{x} : I_1(|\nabla f|\chi_{\Omega_{k-1}})(\mathbf{x}) > 2^{k-1}/c_n\right\}.$$

Therefore, by the weak type estimate in Theorem 14.37(b),

$$|\Omega_k| \leq c_n \left(\frac{1}{2^{k-1}} \int\limits_{\Omega_{k-1}} |\nabla f|\right)^q \quad \text{if } k \geq N+1.$$

Hence,

$$\begin{aligned} S_2 &= \sum_{k \geq N+1} 2^{(k+1)q} |\Omega_k| \\ &\leq c_n \sum_{k \geq N+1} \frac{2^{(k+1)q}}{2^{(k-1)q}} \left(\int\limits_{\Omega_{k-1}} |\nabla f|\right)^q \\ &= c_n \sum_{k \geq N+1} \left(\int\limits_{\Omega_{k-1}} |\nabla f|\right)^q \\ &\leq c_n \left(\sum_k \int\limits_{\Omega_{k-1}} |\nabla f|\right)^q = c_n \left(\int\limits_B |\nabla f|\right)^q. \end{aligned}$$

Here, we have used the estimate $\sum |a_k|^q \leq \left(\sum |a_k|\right)^q$ to obtain the last line (see Exercise 31 of Chapter 8). Combining the estimates for S_1 and S_2 proves (15.38) in case f is smooth.

Now consider a general f with a weak gradient satisfying $|\nabla f| \in L^1(B)$. Let D be a ball with $\overline{D} \subset B$, and choose an approximating sequence $\{f_j\}$ as in Theorem 15.8 with $K = \overline{D}$ there. By the case just proved, we have (with $q = n/(n-1)$)

$$\left(\int\limits_D |f_j - (f_j)_D|^q\right)^{1/q} \leq c_n \int\limits_D |\nabla f_j|$$

for every j. As $j \to \infty$, $f_j \to f$ a.e. in D, $(f_j)_D \to f_D$ since $f_j \to f$ in $L^1(D)$ norm, and $|\nabla f_j| \to |\nabla f|$ in $L^1(D)$. Therefore,

$$\left(\int\limits_D |f - f_D|^q\right)^{1/q} \leq c_n \int\limits_D |\nabla f|,$$

where we used Fatou's lemma for the left side of the inequality. Now enlarge the domain of integration on the right side to B and then let D increase to B. Since $f \in L^1(B)$, then $f_D \to f_B$, and another application of Fatou's lemma gives (15.38).

Finally, since (15.38) and the finiteness of f_B imply that $f \in L^q(B)$ by Minkowski's inequality, the proof is complete.

Part (ii) of Theorem 15.37 yields the next result about the space $W^{1,n}(\mathbf{R^n})$.

Corollary 15.41 *Let $f \in W^{1,n}(\mathbf{R^n})$ and $1 < n \le r < \infty$. Then $f \in L^r(\mathbf{R^n})$ and*

$$\|f\|_{L^r(\mathbf{R^n})} \le c_{n,r} \|f\|_{L^n(\mathbf{R^n})}^{\frac{n}{r}} \|\nabla f\|_{L^n(\mathbf{R^n})}^{1-\frac{n}{r}}, \tag{15.42}$$

where $c_{n,r}$ is a constant depending only on n and r. In particular,

$$\|f\|_{L^r(\mathbf{R^n})} \le c_{n,r} \|f\|_{W^{1,n}(\mathbf{R^n})}, \quad n \le r < \infty. \tag{15.43}$$

Proof. First suppose $f \in C_0^1(\mathbf{R^n})$ and note that the function $F = |f|^{\delta-1} f$ is then also of class $C_0^1(\mathbf{R^n})$ if $\delta \ge 1$. Moreover, $\partial F/\partial x_i = \delta |f|^{\delta-1} \partial f/\partial x_i$ for every $i = 1, \ldots, n$. Denote $n' = n/(n-1)$ and $\|f\|_{L^n(\mathbf{R^n})} = \|f\|_n$. Applying (15.39) to F, we obtain

$$\begin{aligned} \|f\|_{\delta n'}^{\delta} &\le c_n\, \delta \left\| |f|^{\delta-1} |\nabla f| \right\|_1 \\ &\le c_n\, \delta\, \|f\|_{(\delta-1)n'}^{\delta-1} \|\nabla f\|_n, \quad \delta \ge 1, \text{ by Hölder's inequality} \end{aligned} \tag{15.44}$$

with exponents n' and n.

We will successively choose $\delta = n, n+1, n+2, \ldots$ in (15.44). With $\delta = n$, since $(n-1)n' = n$, we have

$$\|f\|_{nn'}^{n} \le c_n\, n\, \|f\|_n^{n-1} \|\nabla f\|_n.$$

Next, choosing $\delta = n+1$,

$$\begin{aligned} \|f\|_{(n+1)n'}^{n+1} &\le c_n\, (n+1)\, \|f\|_{nn'}^{n} \|\nabla f\|_n \\ &\le c_n (n+1) n\, \|f\|_n^{n-1} \|\nabla f\|_n^2, \end{aligned}$$

where we have used the case $\delta = n$ to obtain the last inequality. Continuing inductively in this way gives

$$\|f\|_{(n+k)n'}^{n+k} \le c_n (n+k) \cdots (n+1) n\, \|f\|_n^{n-1} \|\nabla f\|_n^{k+1}, \quad k = 0, 1, \ldots.$$

Raising both sides to the power $1/(n+k)$, we obtain

$$\|f\|_r \le c_{n,r}\,\|f\|_n^{\frac{n}{r}}\,\|\nabla f\|_n^{1-\frac{n}{r}}, \quad r=(n+k)n',\ k=0,1,\ldots.$$

Also, the estimate is trivially true when $r=n$.

In order to obtain the same estimate for every $r\in[n,\infty)$, we use Hölder's inequality in the form (cf. Exercise 6 of Chapter 8)

$$\|f\|_r \le \|f\|_{r_1}^{\theta}\,\|f\|_{r_2}^{1-\theta}, \quad 1\le r_1\le r\le r_2\le\infty,\ \frac{1}{r}=\frac{\theta}{r_1}+\frac{1-\theta}{r_2}.$$

In fact, by choosing $r_1=n$ and $r_2=(n+k)n'$ for a fixed $k=0,1,\ldots$, it follows that for any $r\in[n,(n+k)n']$,

$$\begin{aligned}\|f\|_r &\le \|f\|_n^{\theta}\,\|f\|_{(n+k)n'}^{1-\theta}, \quad \frac{1}{r}=\frac{\theta}{n}+\frac{1-\theta}{(n+k)n'},\\ &\le \|f\|_n^{\theta}\Big(c_{n,k}\|f\|_n^{\frac{n-1}{n+k}}\,\|\nabla f\|_n^{\frac{k+1}{n+k}}\Big)^{1-\theta}\\ &= c_{n,r}\|f\|_n^{\theta+\frac{n-1}{n+k}(1-\theta)}\,\|\nabla f\|_n^{\frac{k+1}{n+k}(1-\theta)} = c_{n,r}\|f\|_n^{\frac{n}{r}}\,\|\nabla f\|_n^{1-\frac{n}{r}}.\end{aligned}$$

This proves the result when $f\in C_0^1(\mathbf{R^n})$. For general $f\in W^{1,n}(\mathbf{R^n})$, the same estimate then holds by applying the second statement in Theorem 15.13 with $p=n$. Details are left to the reader. This completes the proof.

Theorem 15.37(ii) has an elegant companion result relying heavily on the product structure of $\mathbf{R^n}$. There is an equally elegant analogue of Theorem 15.29 for functions defined in all of $\mathbf{R^n}$. The result corresponding to Theorem 15.37(ii) is given next; see Exercise 18 for the one corresponding to Theorem 15.29.

Theorem 15.45 *If $f\in W^{1,1}(\mathbf{R^n})$, then $f\in L^{n/(n-1)}(\mathbf{R^n})$ and*

$$\|f\|_{L^{n/(n-1)}(\mathbf{R^n})} \le \prod_{i=1}^{n}\left\|\frac{\partial f}{\partial x_i}\right\|_{L^1(\mathbf{R^n})}^{1/n}. \tag{15.46}$$

Moreover, the conclusion holds if the hypothesis that $f\in W^{1,1}(\mathbf{R^n})$ is replaced by assuming that f has a weak gradient in $\mathbf{R^n}$ that satisfies $|\nabla f|\in L^1(\mathbf{R^n})$ and either $f\in L^r(\mathbf{R^n})$ for some $r\in[1,\infty)$ or $\lim_{|\mathbf{x}|\to\infty} f(\mathbf{x})=0$.

In the proof, we will use the next lemma in case $n>1$.

Lemma 15.47 *Let $n = 2, 3, \ldots$ and $g_1(\mathbf{x}), \ldots, g_n(\mathbf{x})$ be n functions of $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{R^n}$ such that $g_i(\mathbf{x})$ is independent of x_i, $i = 1, \ldots, n$, that is, $g_i(\mathbf{x})$ depends only on $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \mathbf{R^{n-1}}$. If each $g_i \in L^{n-1}(\mathbf{R^{n-1}})$, then*

$$\left\| \prod_{i=1}^{n} g_i \right\|_{L^1(\mathbf{R^n})} \leq \prod_{i=1}^{n} \|g_i\|_{L^{n-1}(\mathbf{R^{n-1}})}.$$

Proof. The proof will be by induction on n. We may assume that every g_i is nonnegative.

If $n = 2$, the result is true since

$$\int_{\mathbf{R^2}} g_1(x_2) g_2(x_1)\, dx_1 dx_2 = \left(\int_{\mathbf{R^1}} g_1(x_2)\, dx_2 \right) \left(\int_{\mathbf{R^1}} g_2(x_1)\, dx_1 \right).$$

Suppose $n \geq 3$ and let $(n-1)'$ denote the conjugate index of $n-1$. Write $g_1 g_2 \cdots g_n = (g_1 g_2 \cdots g_{n-1})\, g_n$ and apply Hölder's inequality with indices $(n-1)'$ and $n-1$, obtaining

$$\begin{aligned} &\int_{\mathbf{R^{n-1}}} \left\{ \prod_{1}^{n} g_i \right\} dx_1 \cdots dx_{n-1} \\ &\quad \leq \left(\int_{\mathbf{R^{n-1}}} \left\{ \prod_{1}^{n-1} g_i^{(n-1)'} \right\} dx_1 \cdots dx_{n-1} \right)^{\frac{1}{(n-1)'}} \left(\int_{\mathbf{R^{n-1}}} g_n^{n-1}\, dx_1 \cdots dx_{n-1} \right)^{\frac{1}{n-1}}. \end{aligned} \tag{15.48}$$

The second factor on the right side of (15.48) is $\|g_n\|_{L^{n-1}(\mathbf{R^{n-1}})}$, which is independent of x_n since g_n is independent of x_n. We estimate the first factor on the right side by applying the inductive assumption as follows:

$$\begin{aligned} &\int_{\mathbf{R^{n-1}}} \left\{ \prod_{1}^{n-1} g_i^{(n-1)'} \right\} dx_1 \cdots dx_{n-1} \\ &\quad \leq \prod_{1}^{n-1} \left[\int_{\mathbf{R^{n-2}}} g_i^{(n-1)'(n-2)}\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_{n-1} \right]^{\frac{1}{n-2}}. \end{aligned}$$

Now raise both sides of this inequality to the power $1/(n-1)'$, note that $(n-1)'(n-2) = n-1$, and combine the result with (15.48) to obtain

$$\int_{\mathbf{R^{n-1}}} \left\{ \prod_1^n g_i \right\} dx_1 \cdots dx_{n-1}$$
$$\leq \left(\prod_1^{n-1} \left[\int_{\mathbf{R^{n-2}}} g_i^{n-1}\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_{n-1} \right]^{\frac{1}{n-1}} \right) \|g_n\|_{L^{n-1}(\mathbf{R^{n-1}})}.$$

On the right side of this inequality, each of the first $n-1$ factors in the product is a function only of x_n. Hence, by integrating both sides of the inequality with respect to x_n and then using Hölder's inequality with $n-1$ exponents all equal to $n-1$ (see Exercise 6 of Chapter 8), we obtain

$$\int_{\mathbf{R^n}} \left\{ \prod_1^n g_i \right\} d\mathbf{x} \leq \left(\prod_1^{n-1} \|g_i\|_{L^{n-1}(\mathbf{R^{n-1}})} \right) \|g_n\|_{L^{n-1}(\mathbf{R^{n-1}})}$$
$$= \prod_1^n \|g_i\|_{L^{n-1}(\mathbf{R^{n-1}})}.$$

This completes the proof of Lemma 15.47.

Proof of Theorem 15.45. First assume that $f \in C^1(\mathbf{R^n})$, $n > 1$, with $|\nabla f| \in L^1(\mathbf{R^n})$ and that $\lim_{|\mathbf{x}|\to\infty} f(\mathbf{x}) = 0$. Then for any $\mathbf{x} = (x_1, \ldots, x_n)$ and $i = 1, \ldots, n$, we have

$$f(\mathbf{x}) = \int_{-\infty}^{x_i} \frac{\partial f}{\partial x_i}(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n)\, dt$$

and

$$|f(\mathbf{x})| \leq \int_{-\infty}^{\infty} \left| \frac{\partial f}{\partial x_i}(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n) \right| dt \ := h_i(\mathbf{x}),$$

where h_i is defined by the last equality. Note that $h_i(\mathbf{x})$ is independent of x_i and belongs to $L^1(\mathbf{R^{n-1}})$. Raising both sides of the inequality to the power $1/(n-1)$, $n > 1$, and then taking the product over i of the result gives

$$|f(\mathbf{x})|^{\frac{n}{n-1}} \leq \prod_{i=1}^n h_i(\mathbf{x})^{\frac{1}{n-1}}, \quad \mathbf{x} \in \mathbf{R^n}.$$

Each $h_i^{1/(n-1)} \in L^{n-1}(\mathbf{R^{n-1}})$. Therefore, by Lemma 15.47 with g_i chosen to be $h_i^{1/(n-1)}$,

$$\int_{\mathbf{R^n}} |f(\mathbf{x})|^{\frac{n}{n-1}} d\mathbf{x} \leq \prod_{i=1}^{n} \left\| h_i^{\frac{1}{n-1}} \right\|_{L^{n-1}(\mathbf{R^{n-1}})}$$
$$= \prod_{i=1}^{n} \left\| \frac{\partial f}{\partial x_i} \right\|_{L^1(\mathbf{R^n})}^{\frac{1}{n-1}}.$$

Raising both sides to the power $(n-1)/n$ proves (15.46) when $n > 1$ and $f \in C^1(\mathbf{R^n})$ with limit 0 at ∞ and $|\nabla f| \in L^1(\mathbf{R^n})$. The general case when $n > 1$ follows by using the approximation result in Theorem 15.14 combined with Fatou's lemma; details as well as the proof in case $n = 1$ are left to the reader. This completes the proof.

Exercises

1. Construct an example of a function f that has weak first-order partial derivatives in $\mathbf{R^n}$, $n > 1$, but whose ordinary first-order partial derivatives exist nowhere in $\mathbf{R^n}$.
2. Let c be a real number different from 0. Show that there is no locally integrable function g on $(-1,1)$ such that $\int_{-1}^{1} g\varphi\, dx = c\varphi(0)$ for all $\varphi \in C_0^\infty((-1,1))$.
3. The fact that the Cantor–Lebesgue function f does not have a weak derivative on the interval $(0,1)$ is an immediate corollary of Theorem 15.6. Verify this fact instead directly from the definition of a weak derivative by choosing test functions that are adapted to the graph of f.
4. Consider the convolution $(\varphi * k)(\mathbf{x})$, $\mathbf{x} \in \mathbf{R^n}$, where $\varphi \in Lip_{loc}(\mathbf{R^n})$, $k \in L^1(\mathbf{R^n})$, and k has compact support. Show that for every $i = 1, \ldots, n$, the ordinary derivative $\partial(\varphi * k)/\partial x_i$ exists and is finite everywhere in $\mathbf{R^n}$, and
$$\frac{\partial}{\partial x_i}(\varphi * k)(\mathbf{x}) = \int_{\mathbf{R^n}} \frac{\partial \varphi}{\partial y_i}(\mathbf{y}) k(\mathbf{x}-\mathbf{y})\, d\mathbf{y}, \quad \mathbf{x} \in \mathbf{R^n}.$$
Show that the same is true without the assumption that k has compact support if φ satisfies the additional conditions $\varphi, \partial\varphi/\partial y_i \in L^\infty(\mathbf{R^n})$.
5. Verify Theorem 15.9. (Choose a sequence $\{K_\ell\}_{\ell=1}^\infty$ of compact subsets of Ω whose interiors increase to Ω. For each ℓ, use Theorem 15.8 to pick a function $f_\ell \in C_0^\infty(\Omega)$ such that $\|f_\ell - f\|_{L^1(K_\ell)} + \|\nabla f_\ell - \nabla f\|_{L^1(K_\ell)} < 1/\ell$.

Show that for every compact $K \subset \Omega$, $\{f_\ell\}$ and $\{\nabla f_\ell\}$ converge in $L^1(K)$ to f and ∇f, respectively. Then use a diagonal process to construct a subsequence $\{\ell_j\}$ of positive integers such that $\{f_{\ell_j}\}$ and $\{\nabla f_{\ell_j}\}$ also converge pointwise a.e. in Ω.)

6. Let Ω be an open set in $\mathbf{R^n}$ and let $f, g \in L^1_{loc}(\Omega)$. Show that for any $i = 1, \ldots, n$, f has weak partial derivative $\partial f/\partial x_i = g$ if and only if there are functions $\{f_j\}_{j=1}^{\infty} \subset C^\infty(\Omega)$ such that $f_j \to f$ in $L^1(K)$ and $\partial f_j/\partial x_i \to g$ in $L^1(K)$ as $j \to \infty$ for every compact set $K \subset \Omega$.

7. Prove that when $1 \le p \le \infty$, $W^{1,p}(\Omega)$ is a Banach space with respect to the norm (15.11) and that it is separable when $1 \le p < \infty$. Prove that $W^{1,2}(\Omega)$ is a Hilbert space with respect to the inner product (15.12). (To show separability, first note that the repeated Cartesian product of $L^p(\Omega)$ with itself is separable when $1 \le p < \infty$. Then apply the result in Exercise 23 of Chapter 8.)

8. Complete the proof of Theorem 15.13. Furthermore, given a number $\delta > 0$, show that the approximating functions $\{f_j\}$ can be chosen to have supports in the δ-neighborhood $\{\mathbf{x} : d(\mathbf{x}, \Omega) < \delta\}$ of Ω, where $d(\mathbf{x}, \Omega)$ denotes the distance from $\mathbf{x}$ to Ω. In particular, in case Ω is a ball B, if B^* is a larger ball concentric with B, the f_j can be chosen to have supports in B^*.

9. Let $1 \le p \le \infty$ and $f \in W^{1,p}(\mathbf{R^n})$. For $\mathbf{h}$ and $\mathbf{x}$ in $\mathbf{R^n}$, define the translated function $(\tau_{\mathbf{h}} f)(\mathbf{x}) = f(\mathbf{x} + \mathbf{h})$. Show that

$$\|\tau_{\mathbf{h}} f - f\|_{L^p(\mathbf{R^n})} \le |\mathbf{h}| \, \|\nabla f\|_{L^p(\mathbf{R^n})}.$$

(Consider first the case when $f \in C_0^\infty(\mathbf{R^n})$ and derive the estimate $\|\tau_{\mathbf{h}} f - f\|_{L^p(\{|\mathbf{x}|<k\})} \le |\mathbf{h}| \, \|\nabla f\|_{L^p(\{|\mathbf{x}|<k+|\mathbf{h}|\})}$ for $k = 1, 2, \ldots$. If p is finite, the general case follows from approximation and letting $k \to \infty$. If $p = \infty$, apply estimates for L^r when r is finite and let $r \nearrow \infty$.)

10. (a) If $f \in W^{1,\infty}(\mathbf{R^n})$, show that

$$|f(\mathbf{x}) - f(\mathbf{y})| \le |\mathbf{x} - \mathbf{y}| \, \|\nabla f\|_{L^\infty(\mathbf{R^n})} \quad \text{for a.e. } \mathbf{x}, \mathbf{y} \in \mathbf{R^n}.$$

In particular, a function f belongs to $W^{1,\infty}(\mathbf{R^n})$ if and only if there exists $g \in Lip(\mathbf{R^n})$ such that $f = g$ a.e. in $\mathbf{R^n}$.

(b) If $f \in L^\infty(\mathbf{R^n})$, show that $f \in W^{1,\infty}(\mathbf{R^n})$ if and only if there is a constant C such that $|f(\mathbf{x}) - f(\mathbf{y})| \le C\,|\mathbf{x} - \mathbf{y}|$ for a.e. $\mathbf{x}, \mathbf{y} \in \mathbf{R^n}$.

11. (Product Rule) Let $1 \le p \le \infty$ and Ω be an open set in $\mathbf{R^n}$. Prove that if $f, g \in W^{1,p}(\Omega) \cap L^\infty(\Omega)$, then $fg \in W^{1,p}(\Omega) \cap L^\infty(\Omega)$ and

$$\frac{\partial}{\partial x_i}(fg) = f\,\frac{\partial g}{\partial x_i} + \frac{\partial f}{\partial x_i}\,g \quad \text{in } \Omega,\ i = 1, \ldots, n.$$

(Note that both fg and the sum on the right side of the formula belong to $L^p(\Omega)$. It is then enough to check that the sum is the weak derivative $\partial(fg)/\partial x_i$. Use approximation by functions in $C_0^\infty(\mathbf{R^n})$ if p is finite. If $p = \infty$, use the result for finite p after noting that for every bounded open set $\Omega' \subset \Omega$, both f and g belong to $W^{1,p}(\Omega')$ for all finite p.)

12. (Chain Rule) Let $\phi \in C^1(-\infty, \infty)$ with ϕ' bounded and $\phi(0) = 0$. Let $1 \le p \le \infty$ and Ω be an open set in $\mathbf{R^n}$. If $f \in W^{1,p}(\Omega)$, prove that the composition $(\phi \circ f)(\mathbf{x}) = \phi(f(\mathbf{x}))$ satisfies $\phi \circ f \in W^{1,p}(\Omega)$ and

$$\frac{\partial}{\partial x_i}(\phi \circ f) = (\phi' \circ f)\,\frac{\partial f}{\partial x_i}, \quad i = 1, \ldots, n.$$

(The hypothesis $\phi(0) = 0$ guarantees that $|\phi \circ f| \le \|\phi'\|_\infty |f| \in L^p(\Omega)$. It is then enough to check that the expression on the right side of the formula is the weak partial derivative with respect to x_i of $\phi \circ f$. Approximate f by smooth functions if p is finite. If $p = \infty$, use the result for finite p and bounded open sets in Ω.)

13. Let f have compact support in an open set $\Omega \subset \mathbf{R^n}$ and weak derivative $\partial f/\partial x_i$ in Ω for some i. Show that the set where $\partial f/\partial x_i \neq 0$ is contained in the union of the support of f and a subset of Ω of measure 0. (The result in Exercise 2 of Chapter 7 may be helpful.)

14. Let B be a ball in $\mathbf{R^n}$, $1 < p < n$, and $1/q = 1/p - 1/n$. If f has weak gradient ∇f in B and $|\nabla f| \in L^p(B)$, show that

$$\begin{aligned}\|f\|_{L^q(B)} &\le c_{n,p}\|\nabla f\|_{L^p(B)} + c_n \frac{1}{r(B)}\|f\|_{L^p(B)} \\ &\le c_{n,p}\left(1 + \frac{1}{r(B)}\right)\|f\|_{W^{1,p}(B)},\end{aligned}$$

and if either $f_B = 0$ or f has compact support in B, then

$$\|f\|_{L^q(B)} \le c_{n,p}\|\nabla f\|_{L^p(B)} \le c_{n,p}\|f\|_{W^{1,p}(B)}.$$

15. Let f have weak gradient ∇f in $\mathbf{R^n}$ and suppose that $|\nabla f| \in L^p(\mathbf{R^n})$ for some p with $1 \le p < n$. Let $1/q = 1/p - 1/n$:

(a) Show that $f \in L^q(\mathbf{R^n})$ if and only if there is a sequence $\{B_k\}_{k=1}^\infty$ of balls increasing to $\mathbf{R^n}$ such that $|f_{B_k}|\,|B_k|^{1/q}$ is bounded in k.

(b) Suppose that $f_B \to 0$ for some sequence of balls $B \nearrow \mathbf{R^n}$. Show that

$$\sup\left(|f_B|\,|B|^{1/q}\right) < \infty,$$

where the supremum is taken over all balls $B \subset \mathbf{R^n}$.

16. Let $n > 1$ and $B = B(0; 1/2)$ be the ball in $\mathbf{R^n}$ of radius $1/2$ centered at the origin. If $0 < \beta < (n-1)/n$, show that $\left|\log|\mathbf{x}|\right|^{\beta}$ belongs to $W^{1,n}(B)$ but not to $L^{\infty}(B)$. Compare Exercise 14 in Chapter 14.

17. Let f and g be finite functions defined on a set $\Omega \subset \mathbf{R^n}$. Prove that the function $h(\mathbf{x}) = \max\{f(\mathbf{x}), g(\mathbf{x})\}$ (as well as its analogue with *max* replaced by *min*) satisfies

$$|h(\mathbf{x}) - h(\mathbf{y})| \le \max\{|f(\mathbf{x}) - f(\mathbf{y})|, |g(\mathbf{x}) - g(\mathbf{y})|\}, \quad \mathbf{x}, \mathbf{y} \in \Omega.$$

If Ω is open and $f, g \in Lip_{loc}(\Omega)$, deduce that $h \in Lip_{loc}(\Omega)$ and that

$$|\nabla h(\mathbf{x})| \le \max\{|\nabla f(\mathbf{x})|, |\nabla g(\mathbf{x})|\} \quad \text{for a.e. } \mathbf{x} \in \Omega.$$

In particular, if $-\infty < c < \infty$ and $f \in Lip_{loc}(\Omega)$, show that the truncated function

$$f_c(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } f(\mathbf{x}) \le c \\ c & \text{if } f(\mathbf{x}) > c \end{cases}$$

satisfies $|\nabla f_c| \le |\nabla f|$ a.e. in Ω. Show also that $|\nabla(|f|)| \le |\nabla f|$ a.e. in Ω.

18. Prove the following analogue of Theorem 15.45 when $1 < p < n$ and $1/q = 1/p - 1/n$. If $f \in W^{1,p}(\mathbf{R^n})$, then

$$\|f\|_{L^q(\mathbf{R^n})} \le c_{n,p} \prod_{i=1}^{n} \left\| \frac{\partial f}{\partial x_i} \right\|_{L^p(\mathbf{R^n})}^{1/n}.$$

(In case $f \in C_0^1(\mathbf{R^n})$, apply (15.46) to the function $F = |f|^{\delta-1} f$ with $\delta = q(n-1)/n$. Note that $1 < \delta < \infty$, and apply Hölder's inequality to the formula $\|\partial F/\partial x_i\|_{L^1(\mathbf{R^n})} = \|\delta\, |f|^{\delta-1} \partial f/\partial x_i\|_{L^1(\mathbf{R^n})}$ with exponents p' and p.)

19. (a) Show that there exists $f \in C^{\infty}(\mathbf{R^n})$ with $|\nabla f| \in L^p(\mathbf{R^n})$ for every p, $1 \le p \le \infty$, such that (15.39), (15.46), and the first inequality in the conclusion of Theorem 15.29 fail.

(b) Suppose that $1 \le p < n$, $f \in W^{1,p}(\mathbf{R^n})$, and $\partial f/\partial x_i = 0$ a.e. in $\mathbf{R^n}$ for some i, $i = 1, \ldots, n$. Prove that $f = 0$ a.e. in $\mathbf{R^n}$. (See Exercise 18 in case $1 < p < n$.)

20. Let $n = 1$, (a, b) be an open interval in $\mathbf{R^1}$, possibly of infinite length, and suppose that f has a weak derivative f' in (a, b). Verify the following analogues of Theorems 15.30 and 15.32:

(a) If $f' \in L^1(a, b)$, then $f \in L^{\infty}(a, b)$ and

$$\|f - f(c)\|_{L^{\infty}(a,b)} \le \|f'\|_{L^1(a,b)} \quad \text{for a.e. } c \in (a, b).$$

(b) If $f' \in L^p(a,b)$ for some p with $1 < p \leq \infty$, then

$$|f(x) - f(y)| \leq \|f'\|_{L^p(a,b)}|x-y|^{1/p'} \quad \text{for a.e. } x, y \in (a,b),$$

where $1/p + 1/p' + 1$.

21. Let $f \in W^{1,1}(\mathbf{R^n})$. Prove that

$$\|f\|_{L^r(\mathbf{R^n})} \leq \|f\|_{W^{1,1}(\mathbf{R^n})}, \quad 1 \leq r \leq \frac{n}{n-1}.$$

22. Verify the following remarkable fact. Let f and g be locally integrable functions on $\mathbf{R^n}$ that satisfy

$$\frac{1}{|B|}\int_B |f - f_B|\,d\mathbf{x} \leq Cr(B)\frac{1}{|B|}\int_B |g|\,d\mathbf{x}$$

for all balls $B \subset \mathbf{R^n}$, with C independent of B. If $1 < p < n$ and $1/q = 1/p - 1/n$, then for every ball B,

$$\left(\frac{1}{|B|}\int_B |f - f_B|^q\,d\mathbf{x}\right)^{1/q} \leq C'r(B)\left(\frac{1}{|B|}\int_B |g|^p\,d\mathbf{x}\right)^{1/p},$$

where C' depends only on n and C. (Recall Theorem 14.12.)
What can be said if $p \geq n$?

23. Let B be a ball in $\mathbf{R^n}$ with $r(B) = 1$ and $n \geq 2$. Let $k = 1, 2, \ldots$ and $f \in C_0^{k+1}(B)$. If $n = 2k$ or $n = 2k+1$, show that

$$\|f\|_{L^\infty(B)} \leq c_n\|\nabla^{k+1}f\|_{L^2(B)},$$

where the right-hand side denotes the $L^2(B)$ norm of the sum of the absolute value of every partial derivative of f of order $k+1$, and c_n is a constant independent of B and f. (One way to proceed is to first apply Corollary 14.6 to obtain $\|f\|_{L^\infty(B)} \leq c\|\nabla f\|_{L^r(B)}$ if $r > n$. Then show that $\|\nabla f\|_{L^r(B)} \leq c\|\nabla^2 f\|_{L^{rn/(r+n)}(B)}$, and continue considering higher order derivatives as necessary.)

Notations

$x + y$	addition
$\mathbf{x} \cdot \mathbf{y}$	dot product
$\lvert x \rvert$	absolute value
x^+	positive part
x^-	negative part
$f(x+)$	limit from the right
$f(x-)$	limit from the left
$\rightarrow$	converges to
$\overset{m}{\rightarrow}$	converges in measure to
$\nearrow$	increases to
$\searrow$	decreases to
$\{x_k\}$	sequence
$\{x : \ldots\}$	set of x satisfying ...
$x \in E$	x an element of E
$x \notin E$	x not an element of E
$\left.\begin{array}{l} E_1 \cup E_2 \\ \bigcup_k E_k \end{array}\right\}$	union
$\left.\begin{array}{l} E_1 \cap E_2 \\ \bigcup_k E_k \end{array}\right\}$	intersection
$E_1 - E_2$	difference, relative complement
$E_1 \subset E_2$	subset
CE	complement
$\bar{E}$	closure
$\mathring{E}$	interior
diam E	diameter
δ(E)	diameter
δ(x)	distance function
$\lvert E \rvert$	measure
$\lvert E \rvert_e$	outermeasure
$[a, b]$	closed interval
(a, b)	open interval
$\left.\begin{array}{l} (a, b] \\ [a, b) \end{array}\right\}$	partly open intervals
G_δ	countable intersection of open sets
F_σ	countable union of closed sets
$\emptyset$	empty set
sup	supremum, least upper bound
inf	infimum, greatest lower bound

lim sup	limit superior
lim inf	limit inferior
ess sup	essential supremum
ess inf	essential infimum
L, L^1 L^p Lip, Lip_{loc} L^∞ $W^{1,p}$ $\mathscr{S}$	classes of functions
l^p	classes of sequences
$\mathbf{R^1}$	real number system
$\mathbf{R^n}$	n-dimensional Euclidean space
$\mathbf{C}$	complex number system
$\int_E f$	Lebesgue integral in $\mathbf{R^n}$
$\int_E f d\mu$	Lebesgue integral in a measure space
$\int_a^b f d\phi$	Riemann–Stieltjes integral
$\|\cdot\|$	norm
$\|\cdot\|_p$	L^p norm
$\|\cdot\|_*, \|\cdot\|_{**}$	BMO constants
$f * g$	convolution
χ_E	characteristic function
a.e.	almost everywhere
(S, Σ, μ)	measure space
$o(\phi(x))$ $O(\phi(x))$	order of magnitude relations
$V(E)$ $\underline{V}(E), \bar{V}(E)$ $V[f; a, b]$	variations
$S[f]$	Fourier series
$\widetilde{S}[f]$	conjugate Fourier series
usc	upper semicontinuous
lsc	lower semicontinuous
R_Γ	Riemann–Stieltjes sum
S_Γ	sum of increments
$\omega(f; \delta)$ $\omega_p(f; \delta)$	moduli of continuity
$\omega_{f,E}(\alpha)$	distribution function
f^*	Hardy–Littlewood maximal function
$\widetilde{f}$	conjugate function
$\widehat{f}$	Fourier transform
$\mathscr{F}f$	Fourier transform
$I_\alpha f, \widetilde{I}_\alpha f$	fractional integrals

Hf	Hilbert transform
$H_\varepsilon f, H_{\varepsilon,\omega} f$	truncated Hilbert transform
$H_\alpha(A)$	Hausdorff measure
$\Lambda(A)$, $\Lambda^*(A)$	Lebesgue–Stieltjes measure, outer measure
$v(I)$	volume
$\frac{\partial f}{\partial x_i}$	partial derivative
∇f	gradient

Mathematics

Now considered a classic text on the topic, **Measure and Integral: An Introduction to Real Analysis** provides an introduction to real analysis by first developing the theory of measure and integration in the simple setting of Euclidean space, and then presenting a more general treatment based on abstract notions characterized by axioms and with less geometric content.

Published nearly forty years after the first edition, this long-awaited **Second Edition** also:

- Studies the Fourier transform of functions in the spaces L^1, L^2, and L^p, $1 < p < 2$
- Shows the Hilbert transform to be a bounded operator on L^2, as an application of the L^2 theory of the Fourier transform in the one-dimensional case
- Covers fractional integration and some topics related to mean oscillation properties of functions, such as the classes of Hölder continuous functions and the space of functions of bounded mean oscillation
- Derives a subrepresentation formula, which in higher dimensions plays a role roughly similar to the one played by the fundamental theorem of calculus in one dimension
- Extends the subrepresentation formula derived for smooth functions to functions with a weak gradient
- Applies the norm estimates derived for fractional integral operators to obtain local and global first-order Poincaré–Sobolev inequalities, including endpoint cases
- Proves the existence of a tangent plane to the graph of a Lipschitz function of several variables
- Includes many new exercises not present in the first edition

This widely used and highly respected text for upper-division undergraduate and first-year graduate students of mathematics, statistics, probability, or engineering is revised for a new generation of students and instructors. The book also serves as a handy reference for professional mathematicians.

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K24634
ISBN: 978-1-4987-0289-8
90000
9 781498 702898
www.crcpress.com