HALSEY L. ROYDEN | PATRICK M. FITZPATRICK

REAL ANALYSIS

5e

REAL ANALYSIS

Fifth Edition

Halsey L. Royden
Stanford University

Patrick M. Fitzpatrick
University of Maryland, College Park

Pearson

Content Management: *Jonathan Krebs, Jeff Weidenaar*
Content Production: *Manoj Lodhi*
Product Management: *Jessica Darczuk*
Rights and Permissions: *Anjali Singh*

Please contact https://support.pearson.com/getsupport/s/ with any queries on this content.

Library of Congress Cataloging-in-Publication Data

Names: Royden, H. L., author. | Fitzpatrick, Patrick, author.
Title: Real analysis / Halsey L. Royden, Stanford University, Patrick M. Fitzpatrick, University of Maryland, College Park.
Description: Fifth edition. | Hoboken, NJ : Pearson Education, Inc., [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022011423 | ISBN 9780137906529 | ISBN 0137906528
Subjects: LCSH: Functions of real variables. | Functional analysis. | Measure theory.
Classification: LCC QA331.5 .R6 2023 | DDC 515/.8—dc23/eng20220611
LC record available at https://lccn.loc.gov/2022011423

ScoutAutomatedPrintCode

Pearson

ISBN-10: 0-13-790652-8
ISBN-13: 978-0-13-790652-9

To the memory of my wife, Teresita Lega,

Patrick M. Fitzpatrick

To John Slavins

Halsey L. Royden

About the Authors

Halsey Royden was born in Pheonix, Arizona. He earned a BA from Stanford University at the age of 19, and one year later, an MA. After earning a PhD from Harvard University, he returned to Stanford to join the Department of Mathematics, where he remained for his professional career. He spent several sabbaticals at the Institute for Advanced Studies, Princeton. Between 1973 and 1981, he was dean of the School of Humanities and Sciences. During 1973–1974, he was a Guggenheim Fellow. The first edition of his *Real Analysis* was published in 1964. His research interests were in complex analysis and differential geometry.

Patrick Fitzpatrick was born in Youghal, Ireland. He studied at Rutgers University, where, at the age of 19 he earned an undergraduate degree and later a PhD. After post doctoral positions as a Courant Instructor at the Courant Institute, New York University, and an L. E. Dickson Instructor at the University of Chicago, he joined the Department of Mathematics at the University of Maryland, where he has remained for his professional career. He has spent several sabbaticals in Italy. During 1997–2007, he was chair of the Department of Mathematics. His book *Advanced Calculus* is published by the AMS. The first coauthored edition of *Real Analysis* was published in 2010. His research interests are in topological methods in nonlinear analysis.

Contents

Preface x

I Lebesgue Integration for Functions of a Single Real Variable 1

Preliminaries on Sets, Mappings, and Relations 3
Unions and Intersections of Sets . . . 3
Equivalence Relations, the Axiom of Choice, and Zorn's Lemma . . . 5

1 The Real Numbers: Sets, Sequences, and Functions 7
1.1 The Field, Positivity, and Completeness Axioms . . . 7
1.2 The Natural and Rational Numbers . . . 11
1.3 Countable and Uncountable Sets . . . 13
1.4 Open Sets, Closed Sets, and Borel Sets of Real Numbers . . . 16
1.5 Sequences of Real Numbers . . . 20
1.6 Continuous Real-Valued Functions of a Real Variable . . . 25

2 Lebesgue Measure 29
2.1 Introduction . . . 29
2.2 Outer-Measure . . . 31
2.3 The σ-Algebra of Lebesgue Measurable Sets . . . 34
2.4 Finer Properties of Measurable Sets . . . 39
2.5 Countable Additivity and Continuity of Measure, and the Borel-Cantelli Lemma . . . 42
2.6 Vitali's Example of a Non-measurable Set . . . 45
2.7 The Cantor Set and the Cantor-Lebesgue Function . . . 46

3 Lebesgue Measurable Functions 51
3.1 Sums, Products, and Compositions . . . 51
3.2 Sequential Pointwise Limits and Simple Approximation . . . 55
3.3 Littlewood's Three Principles, Egoroff's Theorem, and Lusin's Theorem . . . 59

4 Lebesgue Integration 62
4.1 Comments on the Riemann Integral . . . 62
4.2 The Integral of a Bounded, Finitely Supported, Measurable Function . . . 64
4.3 The Integral of a Non-Negative Measurable Function . . . 70
4.4 The General Lebesgue Integral . . . 74
4.5 Countable Additivity and Continuity of Integration . . . 78

5 Lebesgue Integration: Further Topics **81**
5.1 Uniform Integrability and Tightness: The Vitali Convergence Theorems . . 81
5.2 Convergence in the Mean and in Measure: A Theorem of Riesz 85
5.3 Characterizations of Riemann and Lebesgue Integrability 88

6 Differentiation and Integration **92**
6.1 Continuity of Monotone Functions 92
6.2 Differentiability of Monotone Functions: Lebesgue's Theorem 94
6.3 Functions of Bounded Variation: Jordan's Theorem 99
6.4 Absolutely Continuous Functions 103
6.5 Integrating Derivatives: Differentiating Indefinite Integrals 107
6.6 Measurability: Images of Sets, Compositions of Functions 114
6.7 Convex Functions 118

7 The L^p Spaces: Completeness and Approximation **123**
7.1 Normed Linear Spaces 123
7.2 The Inequalities of Young, Hölder, and Minkowski 125
7.3 L^p Is Complete: Rapidly Cauchy Sequences and the Riesz-Fischer Theorem 130
7.4 Approximation and Separability 134

8 The L^p Spaces: Duality, Weak Convergence, and Minimization **137**
8.1 Bounded Linear Functionals on a Normed Linear Space 137
8.2 The Riesz Representation Theorem for the Dual of $L^p, 1 \leq p < \infty$ 144
8.3 Weak Sequential Convergence in L^p 147
8.4 The Minimization of Convex Functionals 150

II Measure and Integration: General Theory 157

9 General Measure Spaces: Their Properties and Construction **159**
9.1 Measurable Sets and Measure Spaces 159
9.2 Measures Induced by an Outer-Measure 164
9.3 The Carathéodory-Hahn Theorem 168

10 Particular Measures **174**
10.1 Lebesgue Measure on Euclidean Space 174
10.2 Lebesgue Measurability and Measure of Images of Mappings 182
10.3 Borel Measures on $\mathbf{R}^n$ and Regularity 190
10.4 Carathéodory Outer-Measures and Hausdorff Measures 193
10.5 Signed Measures: The Hahn and Jordan Decompositions 197

11 Integration over General Measure Spaces **201**
11.1 Measurable Functions: The Egoroff and Lusin Theorems and Sequential Approximation 201
11.2 Integration of Non-Negative Measurable Functions: Fatou's Lemma, the Monotone Convergence Theorem, and Beppo Levi's Theorem 207

11.3 Integration of General Measurable Functions: The Dominated Convergence Theorem and the Vitali Convergence Theorem 213
11.4 The Radon-Nikodym Theorem 219
11.5 Product Measures: The Tonelli and Fubini Theorems 225
11.6 Products of Lebesgue Measure on Euclidean Spaces: Cavalieri's Principle 233

12 General L^p Spaces: Completeness, Convolution, and Duality 238
12.1 The Spaces $L^p(X, \mu), 1 \leq p \leq \infty$ 238
12.2 Convolution, Smooth Approximation, and a Smooth Urysohn's Lemma . . . 243
12.3 The Riesz Representation Theorem for the Dual of $L^p(X, \mu)$, $1 \leq p < \infty$ . . 252
12.4 Weak Sequential Compactness in $L^p(X, \mu)$, $1 < p < \infty$ 255
12.5 The Kantorovitch Representation Theorem for the Dual of $L^\infty(X, \mu)$ 257

III Abstract Spaces: Metric, Topological, Banach, and Hilbert Spaces 261

13 Metric Spaces: General Properties 263
13.1 Examples of Metric Spaces 263
13.2 Open Sets, Closed Sets, and Convergent Sequences 265
13.3 Continuous Mappings Between Metric Spaces 269
13.4 Complete Metric Spaces 271
13.5 Compact Metric Spaces 274
13.6 Separable Metric Spaces Ulam's Regularity Theorem 280

14 Metric Spaces: Three Fundamental Theorems and Applications 285
14.1 The Arzelà-Ascoli Theorem 285
14.2 The Banach Contraction Principle and Picard's Theorem 289
14.3 The Baire Category Theorem 295
14.4 The Nikodym Metric Space: The Vitali-Hahn-Saks Theorem and the Dunford-Pettis Theorem 299

15 Topological Spaces: General Properties 305
15.1 Open Sets, Closed Sets, Bases, and Subbases 305
15.2 The Separation Properties 309
15.3 Countability and Separability 311
15.4 Continuous Mappings, Weak Topologies, and Metrizability 313
15.5 Compact Topological Spaces 317
15.6 Connected Topological Spaces 320

16 Topological Spaces: Three Fundamental Theorems 323
16.1 Urysohn's Lemma and the Tietze Extension Theorem 323
16.2 The Tychonoff Product Theorem 327
16.3 The Stone-Weierstrass Theorem 331

17 Continuous Linear Operators Between Banach Spaces **337**
17.1 Normed Linear Spaces . . . 337
17.2 Linear Operators . . . 340
17.3 Compactness Lost: Infinite Dimensional Normed Linear Spaces . . . 343
17.4 The Open Mapping and Closed Graph Theorems . . . 347
17.5 The Uniform Boundedness Principle . . . 353

18 Duality for Normed Linear Spaces **355**
18.1 Linear Functionals, Bounded Linear Functionals, and Weak Topologies . . . 355
18.2 The Hahn-Banach Theorem . . . 361
18.3 Reflexive Banach Spaces and Weak Sequential Convergence . . . 367
18.4 Locally Convex Topological Vector Spaces . . . 370
18.5 The Separation of Convex Sets and Mazur's Theorem . . . 375
18.6 The Krein-Milman Theorem . . . 379

19 Compactness Regained: The Weak Topology **382**
19.1 Alaoglu's Extension of Helly's Theorem . . . 382
19.2 Reflexivity and Weak Compactness: Kakutani's Theorem . . . 384
19.3 Compactness and Weak Sequential Compactness: The Eberlein-Šmulian Theorem . . . 386
19.4 Metrizability of Weak Topologies . . . 389

20 Continuous Linear Operators on Hilbert Spaces **392**
20.1 The Inner-Product and Orthogonality . . . 393
20.2 Separability, Bessel's Inequality, and Orthonormal Bases . . . 397
20.3 The Dual Space and Weak Sequential Compactness . . . 399
20.4 Symmetric Operators . . . 404
20.5 Compact Operators . . . 409
20.6 The Hilbert-Schmidt Theorem . . . 411
20.7 The Riesz-Schauder Theorem: Characterization of Fredholm Operators . . . 414

IV Measure and Topology: Invariant Measures **421**

21 Measure and Topology **423**
21.1 Locally Compact Topological Spaces . . . 424
21.2 Separating Sets and Extending Functions . . . 428
21.3 The Construction of Radon Measures . . . 430
21.4 The Representation of Positive Linear Functionals on $C_c(X)$: The Riesz-Markov Theorem . . . 434
21.5 The Representation of the Dual of $C(X)$: The Riesz-Kakutani Theorem . . . 439
21.6 Regularity Properties of Baire Measures . . . 446

22 Invariant Measures **452**
22.1 Topological Groups: The General Linear Group 452
22.2 Kakutani's Fixed-Point Theorem 455
22.3 Invariant Borel Measures on Compact Groups: von Neumann's Theorem . . 460
22.4 Measure Preserving Transformations and Ergodicity 463

Bibliography **470**

Index **472**

Preface

The first three editions of Halsey L. Royden's *Real Analysis* contributed to the education of generations of mathematical analysis students. The fourth and this fifth edition of *Real Analysis* preserve the goal of its venerable predecessors—to present the measure theory, integration theory, and the elements of metric, topological, Hilbert, and Banach spaces that a modern analyst should know.

As in the preceding editions, in Part I, Lebesgue measure and integration for functions of a real variable are considered. In this fifth edition, the treatment of general measure and integration is placed in Part II rather than Part III. What was formerly in Part II is placed in Part III and a brief Part IV. In many courses based on this book, including my own, it has been found preferable to follow in the course the new ordering. This brings measure and integration on Euclidean space closer to their origin, the case of real variables. It also presents the opportunity to more strongly foreshadow, in the context of general measure and integration, concepts that later appear in general spaces.

First, a few remarks regarding specific topics in Part I.

- Somewhat simpler proofs of the Vitali Covering Lemma and Lebesgue's Theorem on the differentiability almost everywhere of a monotone function are provided.
- We prove von Neuman's Composition Theorem, according to which a strictly increasing, continuous function f: $[a, b] \to \mathbf{R}$ has an absolutely continuous inverse function if and only if the composition g: f is measurable whenever the function g: $\mathbf{R} \to \mathbf{R}$ is measurable.
- It is shown that a bounded function on a closed, bounded interval is Riemann integrable if and only if its set of discontinuities has measure zero. Alongside this, we present an ancestor of the Dominated Convergence Theorem for the Lebesgue integral, but for the Riemann integrable, called the Arzelá Convergence Theorem. The difficulty in proving this theorem without leaving the context of Riemann integration is remarkable.
- The concept of uniform integrability is prominently presented, and the Vitali Convergence Theorem is proven and made the centerpiece of the proof of the fundamental theorem of calculus for the Lebesgue integral. We prove that the divided difference functions for an absolutely continuous function are uniformly integrable, so that the fundamental theorem follows by directly taking the limit in its elementary, discrete formulation.
- Following Peter Lax, we consider rapidly Cauchy sequences in the $L^p(E)$ spaces: such sequences converge pointwise and in L^p to function in L^p. The identification of such sequences provides a more conceptual proof of the completeness of L^p spaces.
- An elegant proof of Lusin's Theorem, due to Peter Loeb and Eric Talvila, is given, and from this theorem it immediately follows that a measurable function is the pointwise limit almost everywhere of sequence of continuous functions. This is made the basis for proving, for $1 \leq p < \infty$, the separability of $L^p(\mathbf{R})$.
- The change of variables theorem for the Lebesgue integral for functions of a real variable is proven. This is one of many proofs in which the characterization of Lebesgue

measurable sets as being G_δ sets from which a set of measure zero has been excised is used. The proof brings to light delicate points regarding the measurability of compositions, which are informed by the just mentioned von Neumann Composition Theorem.

- A brief section on convergence in measure and convergence in the mean is included.

The treatment of Lebesgue measure and integration on $\mathbf{R}^n$ now includes the following.

- Convolution of pairs of functions on $\mathbf{R}^n$ are considered. First, Young's Convolution Inequality and Minkowski's Integral Inequality are proven. Based on these, we prove that, for $1 \leq p < \infty$, the compactly supported, infinitely differentiable functions are dense in $L^p(\mathbf{R}^n, \mu_n)$, and also prove a smooth version of Urysohn's Lemma in $\mathbf{R}^n$.
- Sufficient conditions for a mapping on $\mathbf{R}^n$ to preserve Lebesgue measurability of sets are provided. Being Lipschitz is one such condition. We prove the Vitali Partition Theorem, according to which an open subset of R^n is, after the excision of a set of measure zero, the union of a disjoint, countable collection of open balls. Using this, it follows that a linear operator $L: R^n \to R^n$ preserves distance between points if and only is it preserves Lebesgue measure. This provides a simple geometric foundation for the proof that multiplication by the absolute value of the determinant gives the change in Lebesgue measure induced by a linear operator.
- We prove that any finite Borel measure on $\mathbf{R}^n$ is regular, in anticipation of the later proof of Ulam's Regularity Theorem, according to which a finite Borel measure on a separable, complete metric space is regular.
- Care has been taken to explicitly present the set-theoretic properties of measurable rectangles that are at the heart of the proof of the Fubini and Tonelli Theorems, and which, once presented, actually suggest the method of proof.

There is more likely to be agreement about what an analyst should know about measure and integration than there is about what should be known about general spaces. Historically, important special cases of theorems in general spaces were first revealed for spaces of integrable functions. We have commented on these generalizations as the special cases occur, with a view toward motivating them. Consider three examples of these. (i) An important consequence of the Hahn-Banach Theorem regarding the existence of bounded linear functionals on a normed linear space that separate points is explicitly established, and used, when considering, for $1 \leq p < \infty$, the dual of the L^p spaces. (ii) In these same spaces, weak sequential compactness of closed, bounded, convex subsets is proven, and used to establish minima for convex, continuous functions: this foreshadows weak sequential compactness in reflexive Banach space. (iii) The Uniform Boundedness Principle is directly proved for linear functionals rather than as a consequence of the Baire Category Theorem, using an elegant proof of Hahn. This is used to show that weakly convergent sequences are bounded, in these same L^p spaces.

Regarding the selection of general spaces in Part III, normed linear spaces need little motivation, since the L^p spaces have already been broadly considered in the first two parts. Then there are important concepts regarding subsets of normed linear spaces that are independent of the ambient linear structure. These properties are captured in the structure of a metric space, and for these the concepts of compactness and completeness play leading roles. The importance of completeness is brought to the forefront in the Baire Category

Theorem, with its quite elementary proof, and its remarkable diverse applications to individual operators (the Open Mapping and Closed Graph Theorem), to sequences of operators and functions, both linear and non-linear, and to properties of set-functions that are limits of sequences of measures. For complete metric spaces, we also prove, again with quite elementary proofs, the Banach Contraction Principle and a corollary, the Picard Existence Theorem. The importance of compactness goes back to the proof at the beginning of calculus of Rolle's Theorem. We show that a metric space is compact if and only if every continuous real-valued function on it has a minimum value.

Perhaps, for the young analyst, the motivation to extend the concept of metric space to topological space is not so evident. However, we prove a theorem of Riesz which asserts that the closed unit ball of a normed linear space is compact with respect to the metric induced by the norm if and only if the space has finite dimension. We also show, in Part I, that sequential weak compactness is sometimes an able substitute for the loss of compactness for the metric induced by the norm. It is natural to seek the appropriate metric with respect to which convergence is weak sequential compactness. However, we prove that for an infinite dimensional normed linear space, there is not a metric with respect to which sequential convergence is weak sequential convergence. Topological spaces provide a more flexible structure that is not dependent on sequential arguments or countable constructions. For topological spaces, we prove two fundamental theorems, which are strong extensions of their forebears in metric spaces. For metric spaces, the countable product of such spaces is directly defined, and it follows immediately that the countable product of compact metric spaces is compact. The Tychonoff Product Theorem asserts that the arbitrary product of compact topological spaces is compact. The proof of Urysohn's Lemma in a metric space is an immediate consequence of the continuity of the distance functions. The proof of Urysohn's Lemma for a normal topological space is more delicate. It has many interesting applications, among them being the Urysohn Metrization Theorem: if a topological space has a countable base, then the topology is induced by a metric if and only if it is normal.

Selected topics in linear operator theory and, in particular, for linear operators on Hilbert spaces, are presented. These include the consideration of Fredholm operators, which are widely useful in both applied mathematics and modern topology. More particular specific topics, for instance, von Neumann's Theorem on the existence of Haar measure on a compact group and the Stone-Weierstrass Theorem, are also presented. These are both important and elegant. Nevertheless, a different selections of topics could also be well justified. I welcome comments regarding the selections which have, or should have, been made. And, of course, any other comments are also very welcome. I can be contacted through pmf@math.umd.edu. A list of errata and remarks is on https://www.pearsonhighered.com/mathstatsresources.

ACKNOWLEDGMENTS

Regarding the fourth edition, it was a pleasure to acknowledge my indebtedness to teachers, colleagues, and students. My undergraduate teachers, Professors John Bender and Silvio Aurora, revealed to me the beauty of mathematics and to the possibility that I could become a mathematician. A penultimate draft of the entire manuscript was read by Diogo Arsénio, whom I warmly thank for his observations and suggestions, which materially improved that draft. Here in my mathematical home, the University of Maryland, I have written notes for various analysis courses, which have been incorporated into the present edition. A number of students in my graduate analysis classes worked through parts of drafts of this edition, and their comments and suggestions have been most helpful: I thank Avner Halevy, J. J. Lee, Kevin McGoff, and Himanshu Tiagi. I am pleased to acknowledge a particular debt to Brendan Berg who created the index, proofread the final manuscript, and kindly refined my Tex skills. I have benefited from conversations with many friends and colleagues; in particular, with Diogo Arsénio, Stu Antman, Craig Evans, Manos Grillakis, Tim Healey, Brian Hunt, Jacobo Pejsachowicz, and Michael Renardy.

The fourth edition was finely read by Jose Renato Ramos Barbosa and by Richard Hevener, each of whom provided me with errata sheets and many excellent suggestions. Sam Punshon-Smith was an invaluable assistant in preparing the reprintings of that edition. Regarding the preparation of this fifth edition, I am endebted to Stephane Gilles for his generous, expert help during a rather difficult period.

I warmly thank the Publisher, and the reviewers: Randall Swift, California State Polytechnic University; Jianfeng Lu, Duke University; Jeffrey Neugebauer, Eastern Kentucky University; Karl Liechty, DePaul University; Jason Crann, Carleton University; John Harding, New Mexico State University.

Patrick M. Fitzpatrick
La Jolla, CA
May, 2022

PART ONE

LEBESGUE INTEGRATION FOR FUNCTIONS OF A SINGLE REAL VARIABLE

This page intentionally left blank

Preliminaries on Sets, Mappings, and Relations

Contents

Unions and Intersections of Sets 3
Equivalence Relations, the Axiom of Choice, and Zorn's Lemma 5

In these preliminaries, we describe some notions regarding sets, mappings, and relations that will be used throughout the book. Our purpose is descriptive and the arguments given are directed toward plausibility and understanding rather than rigorous proof based on an axiomatic basis for set theory. There is a system of axioms called the Zermelo-Fraenkel Axioms for Sets upon which it is possible to formally establish properties of sets and thereby properties of relations and functions.

UNIONS AND INTERSECTIONS OF SETS

For a set A,[1] the membership of the element x in A is denoted by $x \in A$ and the nonmembership of x in A is denoted by $x \notin A$. We often say a member of A belongs to A and call a member of A a *point* in A. Frequently, sets are denoted by braces, so that $\{x \mid \text{statement about } x\}$ is the set of all elements x for which the statement about x is true.

Two sets are the same provided they have the same members. Let A and B be sets. We call A a **subset** of B provided each member of A is a member of B; we denote this by $A \subseteq B$ and also say that A is contained in B or B contains A. A subset A of B is called a **proper subset** of B provided $A \neq B$. The **union** of A and B, denoted by $A \cup B$, is the set of all points that belong either to A or to B; that is, $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. The word *or* is used here in the nonexclusive sense, so that points which belong to both A and B belong to $A \cup B$. The **intersection** of A and B, denoted by $A \cap B$, is the set of all points that belong to both A and B; that is, $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. The **complement** of A in B, denoted by $B \sim A$, is the set of all points in B that are not in A; that is, $B \sim A = \{x \mid x \in B, x \notin A\}$. If, in a particular discussion, all of the sets are subsets of a reference set X, we often refer to $X \sim A$ simply as the complement of A.

The set that has no members is called the **empty-set** and denoted by $\emptyset$. A set that is not equal to the empty-set is called non-empty. We refer to a set that has a single member as a **singleton set**. Given a set X, the set of all subsets of X is denoted by $\mathcal{P}(X)$ or 2^X; it is called the **power set** of X.

In order to avoid the confusion that might arise when considering sets of sets, we often use the words "collection" and "family" as synonyms for the word "set." Let $\mathcal{F}$ be a collection of sets. We define the union of $\mathcal{F}$, denoted by $\bigcup_{F \in \mathcal{F}} F$, to be the set of points

[1] The *Oxford English Dictionary* devotes several hundred pages to the definition of the word "set."

that belong to at least one of the sets in $\mathcal{F}$. We define the intersection of $\mathcal{F}$, denoted by $\bigcap_{F \in \mathcal{F}} F$, to be the set of points that belong to every set in $\mathcal{F}$. The collection of sets $\mathcal{F}$ is said to be **disjoint** provided the intersection of any two distinct sets in $\mathcal{F}$ is empty. For a family $\mathcal{F}$ of sets, the following identities are established by checking set inclusions.

De Morgan's identities

$$X \sim \left[\bigcup_{F \in \mathcal{F}} F\right] = \bigcap_{F \in \mathcal{F}} [X \sim F] \quad \text{and} \quad X \sim \left[\bigcap_{F \in \mathcal{F}} F\right] = \bigcup_{F \in \mathcal{F}} [X \sim F],$$

that is, the complement of the union is the intersection of the complements, and the complement of the intersection is the union of the complements.

For a set Λ, suppose that for each $\lambda \in \Lambda$, there is defined a set E_λ. Let $\mathcal{F}$ be the collection of sets $\{E_\lambda \mid \lambda \in \Lambda\}$. We write $\mathcal{F} = \{E_\lambda\}_{\lambda \in \Lambda}$ and refer to this as an **indexing** (or **parametrization**) of $\mathcal{F}$ by the **index set** (or **parameter set**) Λ.

Mappings between sets

Given two sets A and B, by a **mapping** or **function** from A into B we mean a correspondence that assigns to each member of A a member of B. In the case B is the set of real numbers we always use the word "function." Frequently we denote such a mapping by $f\colon A \to B$, and for each member x of A, we denote by $f(x)$ the member of B to which x is assigned. For a subset A' of A, we define $f(A') = \{b \mid b = f(a) \text{ for some member } a \text{ of } A'\}$: $f(A')$ is called the image of A' under f. We call the set A the **domain** of the function f and $f(A)$ the **image** or **range** of f. If $f(A) = B$, the function f is said to be **onto**. If for each member b of $f(A)$ there is exactly one member a of A for which $b = f(a)$, the function f is said to be **one-to-one**. A mapping $f\colon A \to B$ that is both one-to-one and onto is said to be **invertible**; we say that this mapping establishes a **one-to-one correspondence** between the sets A and B. Given an invertible mapping $f\colon A \to B$, for each point b in B, there is exactly one member a of A for which $f(a) = b$ and it is denoted by $f^{-1}(b)$. This assignment defines the mapping $f^{-1}\colon B \to A$, which is called the **inverse** of f. Two sets A and B are said to be **equipotent** provided there is an invertible mapping from A onto B. Two sets which are equipotent are, from the set-theoretic point of view, indistinguishable.

Given two mappings $f\colon A \to B$ and $g\colon C \to D$ for which $f(A) \subseteq C$ then the composition $g \circ f\colon A \to D$ is defined by $[g \circ f](x) = g(f(x))$ for each $x \in A$. It is not difficult to see that the composition of invertible mappings is invertible. For a set D, define the identity mapping $id_D\colon D \to D$ by $id_D(x) = x$ for all $x \in D$. A mapping $f\colon A \to B$ is invertible if and only if there is a mapping $g\colon B \to A$ for which

$$g \circ f = id_A \text{ and } f \circ g = id_B.$$

Even if the mapping $f\colon A \to B$ is not invertible, for a set E, we define $f^{-1}(E)$ to be the set $\{a \in A \mid f(a) \in E\}$; it is called the **inverse image** of E under f. We have the following useful properties: for any two sets E_1 and E_2,

$$f^{-1}(E_1 \cup E_2) = f^{-1}(E_1) \cup f^{-1}(E_2),\ f^{-1}(E_1 \cap E_2) = f^{-1}(E_1) \cap f^{-1}(E_2)$$

and

$$f^{-1}(E_1 \sim E_2) = f^{-1}(E_1) \sim f^{-1}(E_2).$$

Finally, for a mapping $f\colon A \to B$ and a subset A' of its domain A, the **restriction** of f to A', denoted by $f|_{A'}$, is the mapping from A' to B which assigns $f(x)$ to each $x \in A'$.

EQUIVALENCE RELATIONS, THE AXIOM OF CHOICE, AND ZORN'S LEMMA

Given two non-empty sets A and B, the **Cartesian product** of A with B, denoted by $A \times B$, is defined to be the collection of all ordered pairs (a, b), where $a \in A$ and $b \in B$ and we consider $(a, b) = (a', b')$ if and only if $a = a'$ and $b = b'$.[2] For a non-empty set X, we call a subset R of $X \times X$ a **relation** on X and write $x \, R \, x'$ provided (x, x') belongs to R. The relation R is said to be **reflexive** provided $x \, R \, x$, for all $x \in X$; the relation R is said to be **symmetric** provided $x \, R \, x'$ if $x' \, R \, x$; the relation R is said to be **transitive** provided whenever $x \, R \, x'$ and $x' \, R \, x''$, then $x \, R \, x''$.

Definition *A relation R on a set X is called an* **equivalence relation** *provided it is reflexive, symmetric, and transitive.*

Given an equivalence relation R on a set X, for each $x \in X$, the set $R_x = \{x' \mid x' \in X, \, x \, R \, x'\}$ is called the **equivalence class** of x (with respect to R). The collection of equivalence classes is denoted by X/R. For example, given a set X, the relation of equipotence is an equivalence relation on the collection 2^X of all subsets of X. The equivalence class of a set with respect to the relation equipotence is called the **cardinality** of the set.

Let R be an equivalence relation on a set X. Since R is symmetric and transitive, $R_x = R_{x'}$ if and only if $x \, R \, x'$ and therefore the collection of equivalence classes is disjoint. Since the relation R is reflexive, X is the union of the equivalence classes. Therefore, X/R is a disjoint collection of non-empty subsets of X whose union is X. Conversely, given a disjoint collection $\mathcal{F}$ of non-empty subsets of X whose union is X, the relation of belonging to the same set in $\mathcal{F}$ is an equivalence relation R on X for which $\mathcal{F} = X/R$.

Given an equivalence relation on a set X, it is often necessary to choose a subset C of X which consists of exactly one member from each equivalence class. Is it obvious that there is such a set? Ernst Zermelo called attention to this question regarding the choice of elements from collections of sets. Suppose, for instance, we define two real numbers to be rationally equivalent provided their difference is a rational number. It is easy to check that this is an equivalence relation on the set of real numbers. But it is not easy to identify a set of real numbers that consists of exactly one member from each rational equivalence class.

Definition *Let $\mathcal{F}$ be a non-empty family of non-empty sets. A* **choice function** *f on $\mathcal{F}$ is a function f from $\mathcal{F}$ to $\cup_{F \in \mathcal{F}} F$ with the property that for each set F in $\mathcal{F}$, $f(F)$ is a member of F.*

Zermelo's Axiom of Choice *Let $\mathcal{F}$ be a non-empty collection of non-empty sets. Then there is a choice function on $\mathcal{F}$.*

Very roughly speaking, a choice function on a family of non-empty sets "chooses" a member from each set in the family. We have adopted an informal, descriptive approach to set theory and accordingly we will freely employ, without further ado, the Axiom of Choice.

[2] In a formal treatment of set theory based on the Zermelo-Fraenkel Axioms, an ordered pair (a, b) is defined to be the set $\{\{a\}, \{a, b\}\}$ and a function with domain in A and image in B is defined to be a non-empty collection of ordered pairs in $A \times B$ with the property that if the ordered pairs (a, b) and (a, b') belong to the function, then $b = b'$.

Definition *A relation R on a set non-empty X is called a* **partial ordering** *provided it is reflexive, transitive, and, for x, x' in X,*

$$\text{if } x\,R\,x' \text{ and } x'\,R\,x, \text{ then } x = x'.$$

A subset E of X is said to be **totally ordered** *provided for x, x' in E, either $x\,R\,x'$ or $x'\,R\,x$. A member x of X is said to be an* **upper bound** *for a subset E of X provided $x'Rx$ for all $x' \in E$, and said to be* **maximal** *provided the only member x' of X for which $x\,R\,x'$ is $x' = x$.*

For a family $\mathcal{F}$ of sets and $A, B \in \mathcal{F}$, define $A\,R\,B$ provided $A \subseteq B$. This relation of **set inclusion** is a partial ordering of $\mathcal{F}$. Observe that a set F in $\mathcal{F}$ is an upper bound for a subfamily $\mathcal{F}'$ of $\mathcal{F}$ provided every set in $\mathcal{F}'$ is a subset of F and a set F in $\mathcal{F}$ is maximal provided it is not a proper subset of any set in $\mathcal{F}$. Similarly, given a family $\mathcal{F}$ of sets and $A, B \in \mathcal{F}$ define $A\,R\,B$ provided $B \subseteq A$. This relation of **set containment** is a partial ordering of $\mathcal{F}$. Observe that a set F in $\mathcal{F}$ is an upper bound for a subfamily $\mathcal{F}'$ of $\mathcal{F}$ provided every set in $\mathcal{F}'$ contains F and a set F in $\mathcal{F}$ is maximal provided it does not properly contain any set in $\mathcal{F}$.

Zorn's Lemma *Let X be a partially ordered set for which every totally ordered subset has an upper bound. Then X has a maximal member.*

We will use Zorn's Lemma to prove some of our most important results, including the Hahn-Banach Theorem, the Tychonoff Product Theorem, and the Krein-Milman Theorem. Zorn's Lemma is equivalent to Zermelo's Axiom of Choice. In the book Functional Analysis by Theo Bühler and Deitmar Salamon, there is a discussion and concise proof of the equivalence of the Axiom of Choice and Zorn's Lemma.

We have defined the Cartesian product of two sets. It is useful to define the Cartesian product of a general parametrized collection of sets. For a collection of sets $\{E_\lambda\}_{\lambda \in \Lambda}$ parametrized by the set Λ, the Cartesian product of $\{E_\lambda\}_{\lambda \in \Lambda}$, which we denote by $\Pi_{\lambda \in \Lambda} E_\lambda$, is defined to be the set of functions f from Λ to $\bigcup_{\lambda \in \Lambda} E_\lambda$ such that for each $\lambda \in \Lambda$, $f(\lambda)$ belongs to E_λ. It is clear that the Axiom of Choice is equivalent to the assertion that the Cartesian product of a non-empty family of non-empty sets is non-empty. Note that the Cartesian product is defined for a parametrized family of sets and that two different parametrizations of the same family will have different Cartesian products. This general definition of Cartesian product is consistent with the definition given for two sets. Indeed, consider two non-empty sets A and B. Define $\Lambda = \{\lambda_1, \lambda_2\}$ where $\lambda_1 \neq \lambda_2$ and then define $E_{\lambda_1} = A$ and $E_{\lambda_2} = B$. The mapping that assigns to the function $f \in \Pi_{\lambda \in \Lambda} E_\lambda$ the ordered pair $(f(\lambda_1), f(\lambda_2))$ is an invertible mapping of the Cartesian product $\Pi_{\lambda \in \Lambda} E_\lambda$ onto the collection of ordered pairs $A \times B$ and therefore these two sets are equipotent. For two sets E and Λ, define $E_\lambda = E$ for all $\lambda \in \Lambda$. Then the Cartesian product $\Pi_{\lambda \in \Lambda} E_\lambda$ is equal to the set of all mappings from Λ to E and is denoted by E^Λ.

C H A P T E R 1

The Real Numbers: Sets, Sequences, and Functions

Contents

1.1 The Field, Positivity, and Completeness Axioms 7
1.2 The Natural and Rational Numbers 11
1.3 Countable and Uncountable Sets 13
1.4 Open Sets, Closed Sets, and Borel Sets of Real Numbers 16
1.5 Sequences of Real Numbers 20
1.6 Continuous Real-Valued Functions of a Real Variable 25

We assume the reader has a familiarity with the properties of real numbers, sets of real numbers, sequences of real numbers, and real-valued functions of a real variable, which are usually treated in an undergraduate course in analysis. This familiarity will enable the reader to assimilate the present chapter, which is devoted to rapidly but thoroughly establishing those results which will be needed and referred to later. We assume that the set of real numbers, which is denoted by $\mathbf{R}$, satisfies three types of axioms. We state these axioms and derive from them properties on the natural numbers, rational numbers, and countable sets. With this as background, we establish properties of open and closed sets of real numbers; convergent, monotone, and Cauchy sequences of real numbers; and continuous real-valued functions of a real variable.

1.1 THE FIELD, POSITIVITY, AND COMPLETENESS AXIOMS

We assume as given the set $\mathbf{R}$ of real numbers such that for each pair of real numbers a and b, there are defined real numbers $a+b$ and ab called the sum and product, respectively, of a and b for which the following Field Axioms, Positivity Axioms, and Completeness Axiom are satisfied.

The field axioms

Commutativity of Addition: For all real numbers a and b,

$$a+b=b+a.$$

Associativity of Addition: For all real numbers a, b, and c,

$$(a+b)+c=a+(b+c).$$

The Additive Identity: There is a real number, denoted by 0, such that

$$0+a=a+0=a \qquad \text{for all real numbers } a.$$

The Additive Inverse: For each real number a, there is a real number b such that

$$a + b = 0.$$

Commutativity of Multiplication: For all real numbers a and b,

$$ab = ba.$$

Associativity of Multiplication: For all real numbers a, b, and c,

$$(ab)c = a(bc).$$

The Multiplicative Identity: There is a real number, denoted by 1, such that

$$1a = a1 = a \qquad \text{for all real numbers } a.$$

The Multiplicative Inverse: For each real number $a \neq 0$, there is a real number b such that

$$ab = 1.$$

The Distributive Property: For all real numbers a, b, and c,

$$a(b + c) = ab + ac.$$

The Nontriviality Assumption:

$$1 \neq 0.$$

Any set that satisfies these axioms is called a **field**. It follows from the commutativity of addition that the additive identity, 0, is unique, and we infer from the commutativity of multiplication that the multiplicative unit, 1, also is unique. The additive inverse and multiplicative inverse also are unique. We denote the additive inverse of a by $-a$ and, if $a \neq 0$, its multiplicative inverse by a^{-1} or $1/a$. If we have a field, we can perform all the operations of elementary algebra, including the solution of simultaneous linear equations. We use the various consequences of these axioms without explicit mention[1].

The positivity axioms

In the real numbers there is a natural notion of order: greater than, less than, and so on. A convenient way to codify these properties is by specifying axioms satisfied by the set of positive numbers. There is a set of real numbers, denoted by $\mathcal{P}$, called the set of **positive numbers.** It has the following two properties:

P1 If a and b are positive, then ab and $a + b$ are also positive.
P2 For a real number a, exactly one of the following three alternatives is true:

$$a \text{ is positive}, \qquad -a \text{ is positive}, \qquad a = 0.$$

[1]A systematic development of the consequences of the Field Axioms may be found in the first chapter of the classic book *A Survey of Modern Algebra* by Garrett Birkhoff and Saunders MacLane [BM97].

The Positivity Axioms lead in a natural way to an ordering of the real numbers: for real numbers a and b, we define $a > b$ to mean that $a - b$ is positive, and $a \geq b$ to mean that $a > b$ or $a = b$. We then define $a < b$ to mean that $b > a$, and $a \leq b$ to mean that $b \geq a$.

Using the Field Axioms and the Positivity Axioms, it is possible to formally establish the familiar properties of inequalities (see Problem 2). Given real numbers a and b for which $a < b$, we define $(a, b) = \{x \mid a < x < b\}$, and say a point in (a, b) lies between a and b. We call a non-empty set I of real numbers an **interval** provided for any two points in I, and all the points that lie between these points also belong to I. Of course, the set (a, b) is an interval, as are the following sets:

$$[a, b] = \{x \mid \ a \leq x \leq b\}\,; [a, b) = \{x \mid \ a \leq x < b\}\,; (a, b] = \{x \mid \ a < x \leq b\}\,. \tag{1}$$

The completeness axiom

A non-empty set E of real numbers is said to be **bounded above** provided there is a real number b such that $x \leq b$ for all $x \in E$: the number b is called an **upper bound** for E. Similarly, we define what it means for a set to be **bounded below** and for a number to be a **lower bound** for a set. A set that is bounded above need not have a largest member. But the next axiom asserts that it does have a smallest upper bound.

The Completeness Axiom *Let E be a non-empty set of real numbers that is bounded above. Then among the set of upper bounds for E there is a smallest, or least, upper bound.*

For a non-empty set E of real numbers that is bounded above, the **least upper bound** of E, the existence of which is asserted by the Completeness Axiom, will be denoted by l.u.b. E. The least upper bound of E is usually called the **supremum** of E and denoted by $\sup E$. It follows from the Completeness Axiom that every non-empty set E of real numbers that is bounded below has a **greatest lower bound**; it is denoted by g.l.b. E and usually called the **infimum** of E and denoted by $\inf E$. A non-empty set of real numbers is said to be **bounded** provided it is both bounded below and bounded above.

The triangle inequality

We define the **absolute value** of a real number x, $|x|$, to be x if $x \geq 0$ and to be $-x$ if $x < 0$. The following inequality, called the **Triangle Inequality**, is fundamental in mathematical analysis: for any pair of real numbers a and b,

$$|a + b| \leq |a| + |b|.$$

The extended real numbers

It is convenient to introduce the symbols ∞ and $-\infty$ and write $-\infty < x < \infty$ for all real numbers x. We call the set $\mathbf{R} \cup \pm\infty$ the **extended real numbers**. If a non-empty set E of real numbers is not bounded above we define its supremum to be ∞. It is also convenient to define $-\infty$ to be the supremum of the empty-set. Therefore, every set of real numbers has a supremum that belongs to the extended real numbers. Similarly, we can extend the concept of infimum so every set of real numbers has an infimum that belongs to the extended real numbers. We define $(-\infty, \infty) = \mathbf{R}$. For $a, b \in \mathbf{R}$, we define

$$(a, \infty) = \{x \in \mathbf{R} \mid a < x\}, \ (-\infty, b) = \{x \in \mathbf{R} \mid x < b\}$$

and

$$[a, \infty) = \{x \in \mathbf{R} \mid a \leq x\}, \ (-\infty, b] = \{x \in \mathbf{R} \mid x \leq b\}.$$

PROBLEMS

1. For $a \neq 0$ and $b \neq 0$, show that $(ab)^{-1} = a^{-1}b^{-1}$.
2. Verify the following:
   (i) For each real number $a \neq 0$, $a^2 > 0$. In particular, $1 > 0$ since $1 \neq 0$ and $1 = 1^2$.
   (ii) For each positive number a, its multiplicative inverse a^{-1} also is positive.
   (iii) If $a > b$, then
   $$ac > bc \text{ if } c > 0 \text{ and } ac < bc \text{ if } c < 0.$$
3. For a non-empty set of real numbers E, show that $\inf E = \sup E$ if and only if E consists of a single point.
4. Let a and b be real numbers.
   (i) Show that if $ab = 0$, then $a = 0$ or $b = 0$.
   (ii) Verify that $a^2 - b^2 = (a - b)(a + b)$ and conclude from part (i) that if $a^2 = b^2$, then $a = b$ or $a = -b$.
   (iii) Let c be a positive real number. Define $E = \{x \in \mathbf{R} \mid x^2 < c\}$. Verify that E is non-empty and bounded above. Define $x_0 = \sup E$. Show that $x_0^2 = c$. Use part (ii) to show that there is a unique $x > 0$ for which $x^2 = c$. It is denoted by $\sqrt{c}$.
5. Let a, b, and c be real numbers such that $a \neq 0$ and consider the quadratic equation
   $$ax^2 + bx + c = 0, \ x \in \mathbf{R}.$$
   (i) Suppose $b^2 - 4ac > 0$. Use the Field Axioms and the preceding problem to complete the square and thereby show that this equation has exactly two solutions given by
   $$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$
6. Use the Completeness Axiom to show that every non-empty set of real numbers that is bounded below has an infimum and that
   $$\inf E = -\sup\{-x \mid x \in E\}.$$
7. For real numbers a and b, verify the following:
   (i) $|ab| = |a||b|$.
   (ii) $|a + b| \leq |a| + |b|$.
   (iii) For $\epsilon > 0$,
   $$|x - a| < \epsilon \text{ if and only if } a - \epsilon < x < a + \epsilon.$$

1.2 THE NATURAL AND RATIONAL NUMBERS

It is tempting to define the natural numbers to be the numbers $1, 2, 3, \ldots$ and so on. However, it is necessary to be more precise. A convenient way to do this is to first introduce the concept of an *inductive set.*

Definition *A set E of real numbers is said to be* **inductive** *provided it contains* 1 *and if the number x belongs to E, the number $x+1$ also belongs to E.*

The whole set of real numbers $\mathbf{R}$ is inductive. From the inequality $1 > 0$, we infer that the sets $\{x \in \mathbf{R} \mid x \geq 0\}$ and $\{x \in \mathbf{R} \mid x \geq 1\}$ are inductive. The set of **natural numbers**, denoted by $\mathbf{N}$, is defined to be *the intersection of all inductive subsets* of $\mathbf{R}$. The set $\mathbf{N}$ is inductive. To see this, observe that the number 1 belongs to $\mathbf{N}$ since 1 belongs to every inductive set. Furthermore, if the number k belongs to $\mathbf{N}$, then k belongs to every inductive set. Thus, $k+1$ belongs to every inductive set and therefore $k+1$ belongs to $\mathbf{N}$.

Principle of Mathematical Induction *For each natural number n, let $S(n)$ be some mathematical assertion. Suppose $S(1)$ is true. Also suppose that whenever k is a natural number for which $S(k)$ is true, then $S(k+1)$ is also true. Then, $S(n)$ is true for every natural number n.*

Proof Define $A = \{k \in \mathbf{N} \mid S(k) \text{ is true}\}$. The assumptions mean precisely that A is an inductive set. Thus $\mathbf{N} \subseteq A$. Therefore, $S(n)$ is true for every natural number n. $\square$

Theorem 1 *Every non-empty set of natural numbers has a smallest member.*

Proof Let E be a non-empty set of natural numbers. Since the set $\{x \in \mathbf{R} \mid x \geq 1\}$ is inductive, the natural numbers are bounded below by 1. Therefore E is bounded below by 1. As a consequence of the Completeness Axiom, E has an infimum; define $c = \inf E$. Since $c+1$ is not a lower bound for E, there is an $m \in E$ for which $m < c+1$. We claim that m is the smallest member of E. Otherwise, there is an $n \in E$ for which $n < m$. Since $n \in E$, $c \leq n$. Thus $c \leq n < m < c+1$ and therefore $m - n < 1$. Therefore, the natural number m belongs to the interval $(n, n+1)$. An induction argument shows that for every natural number n, $(n, n+1) \cap \mathbf{N} = \emptyset$ (see Problem 8). This contradiction confirms that m is the smallest member of E. $\square$

Archimedean Property *For each pair of positive real numbers a and b, there is a natural number n for which $na > b$.*

Proof Define $c = b/a > 0$. We argue by contradiction. If the theorem is false, then c is an upper bound for the natural numbers. By the Completeness Axiom, the natural numbers have a supremum; define $c_0 = \sup \mathbf{N}$. Then $c_0 - 1$ is not an upper bound for the natural numbers. Choose a natural number n such that $n > c_0 - 1$. Therefore, $n+1 > c_0$. But the natural numbers are inductive so that $n+1$ is a natural number. Since $n+1 > c_0$, c_0 is not an upper bound for the natural numbers. This contradiction completes the proof. $\square$

We frequently use the Archimedean Property of $\mathbf{R}$ reformulated as follows; for each positive real number ϵ, there is a natural number n for which $1/n < \epsilon$[2].

[2] Archimedeas explicitly asserted that it was his fellow Greek, Eurathostenes, who identified the property that we have here attributed to Archimedeas.

We define the set of **integers**, denoted by **Z**, to be the set of numbers consisting of the natural numbers, their negatives, and the number 0. The set of **rational numbers**, denoted by **Q**, is defined to be the set of quotients of integers, that is, numbers x of the form $x = m/n$, where m and n are integers and $n \neq 0$. A real number is called **irrational** if it is not rational. As we argued in Problem 4, there is a unique positive number x for which $x^2 = 2$; it is denoted by $\sqrt{2}$. This number is not rational. Indeed, suppose p and q are natural numbers for which $(p/q)^2 = 2$. Then $p^2 = 2q^2$. The prime factorization theorem[3] tells us that 2 divides p^2 just twice as often as it divides p. Hence 2 divides p^2 an even number of times. Similarly, 2 divides $2q^2$ an odd number of times. Thus $p^2 \neq 2q^2$ and therefore $\sqrt{2}$ is irrational.

Definition *A set E of real numbers is said to be* **dense** *in* **R** *provided between any two real numbers there lies a member of E.*

Theorem 2 *The rational numbers are dense in* **R**.

Proof Let a and b be real numbers with $a < b$. First suppose that $a > 0$. By the Archimedean Property of **R**, there is a natural number q for which $(1/q) < b-a$. Again using the Archimedean Property of **R**, the set of natural numbers $S = \{n \in \mathbf{N} \mid n/q \geq b\}$ is non-empty. According to Theorem 1, S has a smallest member p. Observe that $1/q < b - a < b$ and hence $p > 1$. Therefore $p - 1$ is a natural number (see Problem 9) and so, by the minimality of the choice of p, $(p-1)/q < b$. We also have

$$a = b - (b - a) < (p/q) - (1/q) = (p-1)/q.$$

Therefore, the rational number $r = (p-1)/q$ lies between a and b. If $a < 0$, by the Archimedean property of **R**, there is a natural number n for which $n > -a$. We infer from the first case considered that there is a rational number r that lies between $n+a$ and $n+b$. Therefore the rational number $r - n$ lies between a and b. □

PROBLEMS

8. Use an induction argument to show that for each natural number n, the interval $(n, n+1)$ fails to contain any natural number.
9. Use an induction argument to show that if $n > 1$ is a natural number, then $n - 1$ also is a natural number. Then use another induction argument to show that if m and n are natural numbers with $n > m$, then $n - m$ is a natural number.
10. Show that for any real number r, there is exactly one integer in the interval $[r, r+1)$.
11. Show that any non-empty set of integers that is bounded above has a largest member.
12. Show that the irrational numbers are dense in **R**.
13. Show that each real number is the supremum of a set of rational numbers and also the supremum of a set of irrational numbers.
14. Show that if $r > 0$, then, for each natural number n, $(1+r)^n \geq 1 + n \cdot r$.

[3] This theorem asserts that each natural number may be uniquely expressed as the product of prime natural numbers; see [BM97].

15. Use induction arguments to prove that for every natural number n,

    (i) $\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$,

    (ii) $1^3 + 2^3 + \ldots + n^3 = (1 + 2 + \ldots + n)^2$,

    (iii) $1 + r + \ldots + r^n = \frac{1-r^{n+1}}{1-r}$ if $r \neq 1$.

1.3 COUNTABLE AND UNCOUNTABLE SETS

In the preliminaries, we called two sets A and B equipotent provided there is a one-to-one mapping f of A onto B. We refer to such an f as a one-to-one correspondence between the sets A and B. Equipotence defines an equivalence relation among sets, that is, it is reflexive, symmetric, and transitive (see Problem 20). It is convenient to denote the initial segment of natural numbers $\{k \in \mathbf{N} \,|\, 1 \leq k \leq n\}$ by $\{1, \ldots, n\}$. The first observation regarding equipotence is that for any natural numbers n and m, the set $\{1, \ldots, n+m\}$ is not equipotent to the set $\{1, \ldots, n\}$. This observation is often called **the pigeonhole principle** and may be proved by an induction argument with respect to n (see Problem 21).

Definition *A set E is said to be* **finite** *provided either it is empty or there is a natural number n for which E is equipotent to $\{1, \ldots, n\}$. We say that E is* **countably infinite** *provided E is equipotent to the set $\mathbf{N}$ of natural numbers. A set that is either finite or countably infinite is said to be* **countable**. *A set that is not countable is called* **uncountable**.

Observe that if a set is equipotent to a countable set, then it is countable. In the proof of the following theorem we will use the pigeonhole principle and Theorem 1, which tells us that every non-empty set of natural numbers has a smallest, or first, member.

Theorem 3 *A subset of a countable set is countable. In particular, every set of natural numbers is countable.*

Proof Let B be a countable set and A a non-empty subset of B. First consider the case that B is finite. Let f be a one-to-one correspondence between $\{1, \ldots, n\}$ and B. Define $g(1)$ to be the first natural number j, $1 \leq j \leq n$, for which $f(j)$ belongs to A. If $A = \{f(g(1))\}$ the proof is complete since $f \circ g$ is a one-to-one correspondence between $\{1\}$ and A. Otherwise, define $g(2)$ to be the first natural number j, $1 \leq j \leq n$, for which $f(j)$ belongs to $A \sim \{f(g(1))\}$. The pigeonhole principle tells us that this inductive selection process terminates after at most N selections, where $N \leq n$. Therefore, $f \circ g$ is a one-to-one correspondence between $\{1, \ldots, N\}$ and A. Thus, A is finite.

Now consider the case that B is countably infinite. Let f be a one-to-one correspondence between $\mathbf{N}$ and B. Define $g(1)$ to be the first natural number j for which $f(j)$ belongs to A. Arguing as in the first case, we see that if this selection process terminates, then A is finite. Otherwise, this selection process does not terminate and g is properly defined on all of N. It is clear that $f \circ g$ is a one-to-one mapping with domain $\mathbf{N}$ and image contained in A. An induction argument shows that $g(j) \geq j$ for all j. For each $x \in A$, there is some k for which $x = f(k)$. Hence x belongs to the set $\{f(g(1)), \ldots, f(g(k))\}$. Thus, the image of $f \circ g$ is A. Therefore A is countably infinite. $\square$

Corollary 4 *The following sets are countably infinite:*

(i) For each natural number n, the Cartesian product $\overbrace{\mathbf{N} \times \cdots \times \mathbf{N}}^{n \text{ times}}$.
(ii) The set of rational numbers $\mathbf{Q}$.

Proof We prove (i) for $n = 2$ and leave the general case as an exercise in induction. Define the mapping g from $\mathbf{N} \times \mathbf{N}$ to $\mathbf{N}$ by $g(m, n) = (m+n)^2 + n$. The mapping g is one-to-one. Indeed, if $g(m, n) = g(m', n')$, then $(m+n)^2 - (m'+n')^2 = n' - n$ and hence

$$|m + n + m' + n'| \cdot |m + n - m' - n'| = |n' - n|.$$

If $n \neq n'$, then the natural number $m + n + m' + n'$ both divides and is greater than the natural number $|n' - n|$, which is impossible. Thus $n = n'$, and hence $m = m'$. Therefore $\mathbf{N}\times\mathbf{N}$ is equipotent to $g(\mathbf{N}\times\mathbf{N})$, a subset of the countable set $\mathbf{N}$. We infer from the preceding theorem that $\mathbf{N} \times \mathbf{N}$ is countable. To verify the countability of $\mathbf{Q}$, we first infer from the prime factorization theorem that each positive rational number x may be written uniquely as $x = p/q$, where p and q are relatively prime natural numbers. Define the mapping g from $\mathbf{Q}$ to $\mathbf{N}$ by $g(x) = 2((p+q)^2 + q)$ for $x = p/q > 0$ with p and q relatively prime natural numbers, $g(0) = 1$, and $g(x) = g(-x) + 1$ for $x < 0$. We leave it as an exercise to show that g is one-to-one. Thus $\mathbf{Q}$ is equipotent to a subset of $\mathbf{N}$ and hence, by the preceding theorem, is countable. We leave it as an exercise to use the pigeonhole principle to show that neither $\mathbf{N} \times \mathbf{N}$ nor $\mathbf{Q}$ is finite. □

For a countably infinite set X, we say that $\{x_n \,|\, n \in \mathbf{N}\}$ is an **enumeration** of X provided

$$X = \{x_n \mid n \in \mathbf{N}\} \text{ and } x_n \neq x_m \text{ if } n \neq m.$$

Theorem 5 *A non-empty set is countable if and only if it is the image of a function whose domain is a non-empty countable set.*

Proof Let A be a non-empty countable set and f be mapping of A onto B. We suppose that A is countably infinite and leave the finite case as an exercise. By composing with a one-to-one correspondence between A and $\mathbf{N}$, we may suppose that $A = \mathbf{N}$. Define two points x, x' in A to be equivalent provided $f(x) = f(x')$. This is an equivalence relation, that is, it is reflexive, symmetric, and transitive. Let E be a subset of A consisting of one member of each equivalence class. Then the restriction of f to E is a one-to-one correspondence between E and B. But E is a subset of $\mathbf{N}$ and therefore, by Theorem 3, is countable. The set B is equipotent to E and therefore B is countable. The converse assertion is clear; if B is a non-empty countable set, then it is equipotent either to an initial segment of natural numbers or to the natural numbers. □

Corollary 6 *The union of a countable collection of countable sets is countable.*

Proof Let Λ be a countable set and for each $\lambda \in \Lambda$, let E_λ be a countable set. We will show that the union $E = \bigcup_{\lambda\in\Lambda} E_\lambda$ is countable. If E is empty, then it is countable. So we assume $E \neq \emptyset$. We consider the case that Λ is countably infinite and leave the finite case as an exercise. Let $\{\lambda_n \mid n \in \mathbf{N}\}$ be an enumeration of Λ. Fix $n \in \mathbf{N}$. If E_{λ_n} is finite and non-empty, choose a natural number $N(n)$ and a one-to-one mapping f_n of $\{1, \ldots, N(n)\}$ onto E_{λ_n}; if E_{λ_n} is countably infinite, choose a one-to-one mapping f_n of $\mathbf{N}$ onto E_{λ_n}. Define

$$E' = \big\{(n, k) \in \mathbf{N} \times \mathbf{N} | E_{\lambda_n} \text{is non-empty, and } 1 \leq k \leq N(n) \text{ if } E_{\lambda_n} \text{is also finite}\big\}.$$

Define the mapping f of E' to E by $f(n, k) = f_n(k)$. Then f is a mapping of E' onto E. However, E' is a subset of the countable set $\mathbf{N} \times \mathbf{N}$ and hence, by Theorem 3, is countable. Theorem 5 tells us that E also is countable. □

We call an interval of real numbers degenerate if it is empty or contains a single member.

Theorem 7 *A non-degenerate interval of real numbers is uncountable.*

Proof Let I be a non-degenerate interval of real numbers. Clearly I is not finite. We argue by contradiction to show that I is uncountable. Suppose I is countably infinite. Let $\{x_n \mid n \in \mathbf{N}\}$ be an enumeration of I. Let $[a_1, b_1]$ be a non-degenerate closed, bounded subinterval of I which fails to contain x_1. Then let $[a_2, b_2]$ is a non-degenerate closed, bounded subinterval of $[a_1, b_1]$, which fails to contain x_2. We inductively choose a countable collection $\{[a_n, b_n]\}_{n=1}^{\infty}$ of non-degenerate closed, bounded intervals, which is descending in the sense that, for each n, $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ and such that for each n, $x_n \notin [a_n, b_n]$. The non-empty set $E = \{a_n \mid n \in \mathbf{N}\}$ is bounded above by b_1. The Completeness Axiom tells us that E has a supremum. Define $x^* = \sup E$. Since x^* is an upper bound for E, $a_n \leq x^*$ for all n. On the other hand, since $\{[a_n, b_n]\}_{n=1}^{\infty}$ is descending, for each n, b_n is an upper bound for E. Hence, for each n, $x^* \leq b_n$. Therefore, x^* belongs to $[a_n, b_n]$ for each n. But x^* belongs to $[a_1, b_1] \subseteq I$ and therefore there is a natural number n_0 for which $x^* = x_{n_0}$. We have a contradiction since $x^* = x_{n_0}$ does not belong to $[a_{n_0}, b_{n_0}]$. Therefore I is uncountable. □

PROBLEMS

16. Show that the set $\mathbf{Z}$ of integers is countable.
17. Show that a set A is countable if and only if there is a one-to-one mapping of A to $\mathbf{N}$.
18. Use an induction argument to complete the proof of part (i) of Corollary 4.
19. Prove Corollary 6 in the case of a finite family of countable sets.
20. Let both $f: A \to B$ and $g: B \to C$ be one-to-one and onto. Show that the composition $g \circ f: A \to B$ and the inverse $f^{-1}: B \to A$ are also one-to-one and onto.
21. Use an induction argument to establish the pigeonhole principle.
22. Show that $2^{\mathbf{N}}$, the collection of all sets of natural numbers, is uncountable.
23. Show that the Cartesian product of a finite collection of countable sets is countable. Use the preceding problem to show that $\mathbf{N}^{\mathbf{N}}$, the collection of all mappings of $\mathbf{N}$ into $\mathbf{N}$, is not countable.
24. Show that a non-degenerate interval of real numbers fails to be finite.

25. Show that any two non-degenerate intervals of real numbers are equipotent.
26. Is the set $\mathbf{R} \times \mathbf{R}$ equipotent to $\mathbf{R}$?

1.4 OPEN SETS, CLOSED SETS, AND BOREL SETS OF REAL NUMBERS

Definition *A set $\mathcal{O}$ of real numbers is called* **open** *provided for each $x \in \mathcal{O}$, there is a $r > 0$ for which the interval $(x - r, x + r)$ is contained in $\mathcal{O}$.*

For $a < b$, the interval (a, b) is an open set. Indeed, let x belong to (a, b). Define $r = \min\{b - x, x - a\}$. Observe that $(x - r, x + r)$ is contained in (a, b). Thus (a, b) is an open bounded interval and each bounded open interval is of this form. For $a, b \in \mathbf{R}$, we defined

$$(a, \infty) = \{x \in \mathbf{R} \mid a < x\}, (-\infty, b) = \{x \in \mathbf{R} \mid x < b\} \text{ and } (-\infty, \infty) = \mathbf{R}.$$

Observe that each of these sets is an open interval. Moreover, it is not difficult to see that since each set of real numbers has an infimum and supremum in the set of extended real numbers, each unbounded open interval is of the above form.

Proposition 8 *The set of real numbers $\mathbf{R}$ and the empty-set $\emptyset$ are open; the intersection of any finite collection of open sets is open; and the union of any collection of open sets is open.*

Proof It is clear that $\mathbf{R}$ and $\emptyset$ are open and the union of any collection of open sets is open. Let $\{\mathcal{O}_k\}_{k=1}^n$ be a finite collection of open subsets of $\mathbf{R}$. If the intersection of this collection is empty, then the intersection is the empty-set and therefore is open. Otherwise, let x belong to $\cap_{k=1}^n \mathcal{O}_k$. For $1 \leq k \leq n$, choose $r_k > 0$ for which $(x - r_k, x + r_k) \subseteq \mathcal{O}_k$. Define $r = \min\{r_1, \ldots, r_n\}$. Then $r > 0$ and $(x - r, x + r) \subseteq \cap_{k=1}^n \mathcal{O}_k$. Therefore $\cap_{k=1}^n \mathcal{O}_k$. is open. $\square$

It is not true, however, that the intersection of any collection of open sets is open. For example, for each natural number n, let $\mathcal{O}_n$ be the open interval $(-1/n, 1/n)$. Then, by the Archimedean Property of $\mathbf{R}$, $\bigcap_{n=1}^\infty \mathcal{O}_n = \{0\}$, and $\{0\}$ is not an open set.

Proposition 9 *Every non-empty open set is the union of a countable, disjoint collection of open intervals.*

Proof Let $\mathcal{O}$ be a non-empty open subset of $\mathbf{R}$. Let x belong to $\mathcal{O}$. There is a $y > x$ for which $(x, y) \subseteq \mathcal{O}$ and a $z < x$ for which $(z, x) \subseteq \mathcal{O}$. Define the extended real numbers a_x and b_x by

$$a_x = \inf\{z \mid (z, x) \subseteq \mathcal{O}\} \text{ and } b_x = \sup\{y \mid (x, y) \subseteq \mathcal{O}\}.$$

Then $I_x = (a_x, b_x)$ is an open interval that contains x. We claim that

$$I_x \subseteq \mathcal{O} \text{ but } a_x \notin \mathcal{O}, b_x \notin \mathcal{O}. \tag{2}$$

Indeed, let w belong to I_x, say $x < w < b_x$. By the definition of b_x, there is a number $y > w$ such that $(x, y) \subseteq \mathcal{O}$, and so $w \in \mathcal{O}$. Moreover, $b_x \notin \mathcal{O}$, for if $b_x \in \mathcal{O}$, then for some $r > 0$ we have $(b_x - r, b_x + r) \subseteq \mathcal{O}$. Thus $(x, b_x + r) \subseteq \mathcal{O}$, contradicting the definition of b_x.

Similarly, $a_x \notin \mathcal{O}$. Consider the collection of open intervals $\{I_x\}_{x\in\mathcal{O}}$. Since each x in $\mathcal{O}$ is a member of I_x, and each I_x is contained in $\mathcal{O}$, we have $\mathcal{O} = \bigcup_{x\,\in\,\mathcal{O}} I_x$. We infer from (2) that $\{I_x\}_{x\in\mathcal{O}}$ is disjoint. Thus $\mathcal{O}$ is the union of a disjoint collection of open intervals. It remains to show that this collection is countable. By the density of the rationals, Theorem 2, each of these open intervals contains a rational number. This establishes a one-to-one correspondence between the collection of open intervals and a subset of the rational numbers. We infer from Theorem 3 and Corollary 4 that any set of rational numbers is countable. Therefore, $\mathcal{O}$ is the union of a countable disjoint collection of open intervals. □

Definition *For a set E of real numbers, a real number x is called a* **point of closure** *of E provided every open interval that contains x also contains a point in E. The collection of points of closure of E is called the* **closure** *of E and denoted by $\overline{E}$.*

It is clear that we always have $E \subseteq \overline{E}$. If E contains all of its points of closure, that is, $E = \overline{E}$, then the set E is said to be **closed.**

Proposition 10 *For a set of real numbers E, its closure $\overline{E}$ is closed. Moreover, $\overline{E}$ is the smallest closed set that contains E, in the sense that if F is closed and $E \subseteq F$, then $\overline{E} \subseteq F$.*

Proof The set $\overline{E}$ is closed provided it contains all its points of closure. Let x be a point of closure of $\overline{E}$. Consider an open interval I_x which contains x. There is a point $x' \in \overline{E} \cap I_x$. Since x' is a point of closure of E and the open interval I_x contains x', there is a point $x'' \in E \cap I_x$. Therefore, every open interval that contains x also contains a point of E and hence $x \in \overline{E}$. So the set $\overline{E}$ is closed. It is clear that if $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$, and hence if F is closed and contains E, then $\overline{E} \subseteq \overline{F} = F$. □

Proposition 11 *A set of real numbers is open if and only if its complement in* $\mathbf{R}$ *is closed.*

Proof First suppose E is an open subset of $\mathbf{R}$. Let x be a point of closure of $\mathbf{R} \sim E$. Then x cannot belong to E because otherwise there would be an open interval that contains x and is contained in E and thus is disjoint from $\mathbf{R} \sim E$. Therefore x belongs to $\mathbf{R} \sim E$ and hence $\mathbf{R} \sim E$ is closed. Now suppose $\mathbf{R} \sim E$ is closed. Let x belong to E. Then there must be an open interval that contains x that is contained in E, for otherwise every open interval that contains x contains points in $\mathbf{R} \sim E$ and therefore x is a point of closure of $\mathbf{R} \sim E$. Since $\mathbf{R} \sim E$ is closed, x also belongs to $\mathbf{R} \sim E$. This is a contradiction. □

Since $\mathbf{R} \sim [\mathbf{R} \sim E] = E$, it follows from the preceding proposition that *a set is closed if and only if its complement is open.* Therefore, by De Morgan's Identities, Proposition 8 may be reformulated in terms of closed sets as follows.

Proposition 12 *The empty-set $\emptyset$ and* $\mathbf{R}$ *are closed; the union of any finite collection of closed sets is closed; and the intersection of any collection of closed sets is closed.*

A collection of sets $\{E_\lambda\}_{\lambda\in\Lambda}$ is said to be a **cover** of a set E provided $E \subseteq \bigcup_{\lambda\in\Lambda} E_\lambda$. By a subcover of a cover of E, we mean a subcollection of the cover that itself also is a cover of E. If each set E_λ in a cover is open, we call $\{E_\lambda\}_{\lambda\in\Lambda}$ an **open cover** of F. If the cover

$\{E_\lambda\}_{\lambda \in \Lambda}$ contains only a finite number of sets, we call it a **finite cover**. This terminology is inconsistent: In "open cover," the adjective "open" refers to the sets in the cover; in "finite cover," the adjective "finite" refers to the collection and does not imply that the sets in the collection are finite sets. Thus the term "open cover" is an abuse of language and should properly be "cover by open sets." Unfortunately, the former terminology is well established in mathematics.

A set F of real numbers is said to be compact provided that every open cover of F has a finite subcover.

The Heine-Borel Theorem *A set of real numbers is compact if and only if it is closed and bounded.*

Proof We leave as an exercise the proofs, by contradiction, that a compact set is closed and is bounded. To prove the converse, let us first consider the case that F is the closed, bounded interval $[a, b]$. Let $\mathcal{F}$ be an open cover of $[a, b]$. Define E to be the set of numbers $x \in [a, b]$ with the property that the interval $[a, x]$ can be covered by a finite number of the sets of $\mathcal{F}$. Since $a \in E$, E is non-empty. Since E is bounded above by b, by the completeness of $\mathbf{R}$, E has a supremum; define $c = \sup E$. Since c belongs to $[a, b]$, there is an $\mathcal{O} \in \mathcal{F}$ that contains c. Since $\mathcal{O}$ is open, there is an $\epsilon > 0$, such that the interval $(c - \epsilon, c + \epsilon)$ is contained in $\mathcal{O}$. Now $c - \epsilon$ is not an upper bound for E, and so there must be an $x \in E$ with $x > c - \epsilon$. Since $x \in E$, there is a finite collection $\{\mathcal{O}_1, \ldots, \mathcal{O}_k\}$ of sets in $\mathcal{F}$ that covers $[a, x]$. Consequently, the finite collection $\{\mathcal{O}_1, \ldots, \mathcal{O}_k, \mathcal{O}\}$ covers the interval $[a, c+\epsilon)$. Thus $c = b$, for otherwise $c < b$ and c is not an upper bound for E. Thus, $[a, b]$ can be covered by a finite number of sets from $\mathcal{F}$, proving our special case.

Now let F be any closed and bounded set and $\mathcal{F}$ an open cover of F. Since F is bounded, it is contained in some closed, bounded interval $[a, b]$. The preceding proposition tells us that the set $\mathcal{O} = \mathbf{R} \sim F$ is open since F is closed. Let $\mathcal{F}^*$ be the collection of open sets obtained by adding $\mathcal{O}$ to $\mathcal{F}$, that is, $\mathcal{F}^* = \mathcal{F} \cup \mathcal{O}$. Since $\mathcal{F}$ covers F, $\mathcal{F}^*$ covers $[a, b]$. By the case just considered, there is a finite subcollection of $\mathcal{F}^*$ that covers $[a, b]$ and hence F. By removing $\mathcal{O}$ from this finite subcover of F if $\mathcal{O}$ belongs to the finite subcover, we have a finite collection of sets in $\mathcal{F}$ that covers F. $\square$

We say that a countable collection of sets $\{E_n\}_{n=1}^\infty$ is **descending** provided $E_{n+1} \subseteq E_n$ for every natural number n. It is said to be **ascending** provided each $E_n \subseteq E_{n+1}$.

The Nested Set Theorem *Let $\{F_n\}_{n=1}^\infty$ be a descending countable collection of non-empty closed sets of real numbers and F_1 be bounded. Then*

$$\bigcap_{n=1}^{\infty} F_n \neq \emptyset.$$

Proof We argue by contradiction. Suppose the intersection is empty. Then for each real number x, there is a natural number n for which $x \notin F_n$, that is, $x \in \mathcal{O}_n = \mathbf{R} \sim F_n$. Therefore $\bigcup_{n=1}^\infty \mathcal{O}_n = \mathbf{R}$. According to Proposition 11, since each F_n is closed, each $\mathcal{O}_n$ is open. Therefore $\{\mathcal{O}_n\}_{n=1}^\infty$ is an open cover of $\mathbf{R}$ and hence also of F_1. The Heine-Borel Theorem tells us that there is a natural number N for which $F_1 \subseteq \bigcup_{n=1}^N \mathcal{O}_n$. Since $\{F_n\}_{n=1}^\infty$ is descending, the collection of complements $\{\mathcal{O}_n\}_{n=1}^{\infty}$ is ascending. Therefore $\bigcup_{n=1}^{N} \mathcal{O}_n = \mathcal{O}_N = \mathbf{R} \sim F_N$. Hence $F_1 \subseteq \mathbf{R} \sim F_N$, which contradicts the assumption that F_N is a non-empty subset of F_1. $\square$

Definition *Given a set X, a collection $\mathcal{A}$ of subsets of X is called a σ-algebra (of subsets of X) provided (i) the empty-set, $\emptyset$, belongs to $\mathcal{A}$; (ii) the complement in X of a set in $\mathcal{A}$ also belongs to $\mathcal{A}$; (iii) the union of a countable collection of sets in $\mathcal{A}$ also belongs to $\mathcal{A}$.*

Given a set X, the collection $\{\emptyset, X\}$ is a σ-algebra which has two members and is contained in every σ-algebra of subsets of X. At the other extreme is the collection of sets 2^X which consists of all subsets of X and contains every σ-algebra of subsets of X. For any σ-algebra $\mathcal{A}$, we infer from De Morgan's Identities that $\mathcal{A}$ is closed with respect to the formation of intersections of countable collections of sets that belong to $\mathcal{A}$; moreover, since the empty-set belongs to $\mathcal{A}$, $\mathcal{A}$ is closed with respect to the formation of finite unions and finite intersections of sets that belong to $\mathcal{A}$. We also observe that a σ-algebra is closed with respect to relative complements since if A_1 and A_2 belong to $\mathcal{A}$, so does $A_1 \sim A_2 = A_1 \cap [X \sim A_2]$. The proof of the following proposition follows directly from the definition of σ-algebra.

Proposition 13 *Let $\mathcal{F}$ be a collection of subsets of a set X. Then the intersection $\mathcal{A}$ of all σ-algebras of subsets of X that contain $\mathcal{F}$ is a σ-algebra that contains $\mathcal{F}$. Moreover, it is the smallest σ-algebra of subsets of X that contains $\mathcal{F}$, in the sense that any σ-algebra that contains $\mathcal{F}$ also contains $\mathcal{A}$.*

Let $\{A_n\}_{n=1}^{\infty}$ be a countable collection of sets that belong to a σ-algebra $\mathcal{A}$. Since $\mathcal{A}$ is closed with respect to the formation of countable intersections and unions, the following two sets belong to $\mathcal{A}$:

$$\limsup\{A_n\}_{n=1}^{\infty} = \bigcap_{k=1}^{\infty}\left[\bigcup_{n=k}^{\infty} A_n\right] \text{ and } \liminf\{A_n\}_{n=1}^{\infty} = \bigcup_{k=1}^{\infty}\left[\bigcap_{n=k}^{\infty} A_n\right].$$

The set $\limsup\{A_n\}_{n=1}^{\infty}$ is the set of points that belong to A_n for countably infinitely many indices n while the set $\liminf\{A_n\}_{n=1}^{\infty}$ is the set of points that belong to A_n except for at most finitely many indices n.

Although the union of any collection of open sets is open and the intersection of any finite collection of open sets is open, as we have seen, the intersection of a *countable* collection of open sets need not be open. In our development of Lebesgue measure and integration on the real line, we will see that the smallest σ-algebra of sets of real numbers that contains the open sets is a natural object of study.

Definition *The collection $\mathcal{B}$ of Borel sets of real numbers is the smallest σ-algebra of sets of real numbers that contains all of the open sets of real numbers.*

Every open set is a Borel set and since a σ-algebra is closed with respect to the formation of complements, we infer from Proposition 11 that every closed set is a Borel set. Therefore, since each singleton set is closed, every countable set is a Borel set. A countable intersection of open sets is called a G_δ set. A countable union of closed sets is called an F_σ set.

Since a σ-algebra is closed with respect to the formation of countable unions and countable intersections, each G_δ set and each F_σ set is a Borel set. Moreover, both the lim inf and lim sup of a countable collection of sets of real numbers, each of which is either open or closed, are a Borel set.

PROBLEMS

27. Is the set of rational numbers open or closed?
28. What are the sets of real numbers that are both open and closed?
29. Find two sets A and B such that $A \cap B = \emptyset$ and $\overline{A} \cap \overline{B} \neq \emptyset$.
30. A point x is called an **accumulation point** of a set E provided it is a point of closure of $E \sim \{x\}$.
    (i) Show that the set E' of accumulation points of E is a closed set.
    (ii) Show that $\overline{E} = E \cup E'$.
31. A point x is called an **isolated point** of a set E provided there is an $r > 0$ for which $(x-r, x+r) \cap E = \{x\}$. Show that if a set E consists of isolated points, then it is countable.
32. A point x is called an **interior point** of a set E if there is an $r > 0$ such that the open interval $(x-r, x+r)$ is contained in E. The set of interior points of E is called the **interior** of E denoted by $\operatorname{int} E$. Show that
    (i) E is open if and only if $E = \operatorname{int} E$.
    (ii) E is dense if and only if $\operatorname{int}(\mathbf{R} \sim E) = \emptyset$.
33. Show that the Nested Set Theorem is false if F_1 is unbounded.
34. Show that the assertion of the Heine-Borel Theorem is equivalent to the Completeness Axiom for the real numbers. Show that the assertion of the Nested Set Theorem is equivalent to the Completeness Axiom for the real numbers.
35. Show that the collection of Borel sets is the smallest σ-algebra that contains the closed sets.
36. Show that the collection of Borel sets is the smallest σ-algebra that contains intervals of the form $[a, b)$, where $a < b$.
37. Show that each open set is an F_σ set.

1.5 SEQUENCES OF REAL NUMBERS

A **sequence** of real numbers is a real-valued function whose domain is the set of natural numbers. Rather than denoting a sequence with standard functional notation such as $f: \mathbf{N} \to \mathbf{R}$, it is customary to use subscripts, replace $f(n)$ with a_n, and denote a sequence by $\{a_n\}$. A natural number n is called an **index** for the sequence, and the number a_n corresponding to the index n is called the nth **term** of the sequence. Just as we say that a real-valued function is bounded provided its image is a bounded set of real numbers, we say a sequence $\{a_n\}$ is **bounded** provided there is some $c \geq 0$ such that $|a_n| \leq c$ for all n. A sequence is said to be **increasing** provided $a_n \leq a_{n+1}$ for all n, is said to be **decreasing** provided $\{-a_n\}$ is increasing, and is said to be **monotone** provided it is either increasing or decreasing.

Definition *A sequence* $\{a_n\}$ *is said to* **converge** *to the number* a*, provided for every* $\epsilon > 0$ *there is an index* N *for which*

$$\text{if } n \geq N, \text{ then } |a - a_n| < \epsilon. \tag{3}$$

We call a *the* **limit** *of the sequence and denote the convergence of* $\{a_n\}$ *by writing*

$$\{a_n\} \to a \text{ or } \lim_{n \to \infty} a_n = a.$$

We leave the proof of the following proposition as an exercise.

Proposition 14 *Let the sequence of real numbers* $\{a_n\}$ *converge to the real number* a*. Then the limit is unique, the sequence is bounded, and, for a real number* c*,*

$$\text{if } a_n \leq c \text{ for all } n, \text{ then } a \leq c.$$

Theorem 15 (**the Monotone Convergence Criterion for Real Sequences**) *A monotone sequence of real numbers converges if and only if it is bounded.*

Proof Let $\{a_n\}$ be an increasing sequence. If this sequence converges, then, by the preceding proposition, it is bounded. Now assume that $\{a_n\}$ is bounded. By the Completeness Axiom, the set $S = \{a_n \mid n \in N\}$ has a supremum: define $a = \sup S$. We claim that $\{a_n\} \to a$. Indeed, let $\epsilon > 0$. Since s is an upper bound for S, $a_n \leq a$ for all n. Since $a - \epsilon$ is not an upper bound for S, there is an index N for which $a_N > a - \epsilon$. Since the sequence is increasing, $a_n > a - \epsilon$ for all $n \geq N$. Thus if $n \geq N$, then $|a - a_n| < \epsilon$. Therefore $\{a_n\} \to a$. The proof for the case when the sequence is decreasing is the same. $\square$

For a sequence $\{a_n\}$ and a strictly increasing sequence of natural numbers $\{n_k\}$, we call the sequence $\{a_{n_k}\}$ whose kth term is a_{n_k} a **subsequence** of $\{a_n\}$.

Theorem 16 (**the Bolzano-Weierstrass Theorem**) *Every bounded sequence of real numbers has a convergent subsequence.*

Proof Let $\{a_n\}$ be a bounded sequence of real numbers. Choose $M \geq 0$ such that $|a_n| \leq M$ for all n. Let n be a natural number. Define $E_n = \overline{\{a_j \mid j \geq n\}}$. Then $E_n \subseteq [-M, M]$ and E_n is closed since it is the closure of a set. Therefore, $\{E_n\}$ is a descending sequence of nonempty closed bounded subsets of $\mathbf{R}$. The Nested Set Theorem tells us that $\bigcap_{n=1}^{\infty} E_n \neq \emptyset$; choose $a \in \bigcap_{n=1}^{\infty} E_n$. For each natural number k, a is a point of closure of $\{a_j \mid j \geq k\}$. Hence, for infinitely many indices $j \geq n$, a_j belongs to $(a - 1/k, a + 1/k)$. By induction, choose a strictly increasing sequence of natural numbers $\{n_k\}$ such that $|a - a_{n_k}| < 1/k$ for all k. We infer from the Archimedean Property of $\mathbf{R}$ that the subsequence $\{a_{n_k}\}$ converges to a. $\square$

Definition *A sequence of real numbers* $\{a_n\}$ *is said to be* **Cauchy** *provided for each* $\epsilon > 0$*, there is an index* N *for which*

$$\text{if } n, m \geq N, \text{ then } |a_m - a_n| < \epsilon. \tag{4}$$

Theorem 17 (**the Cauchy Convergence Criterion for Real Sequences**) *A sequence of real numbers converges if and only if it is Cauchy.*

Proof First suppose that $\{a_n\} \to a$. Observe that for all natural numbers n and m,

$$|a_n - a_m| = |(a_n - a) + (a - a_m)| \leq |a_n - a| + |a_m - a|. \tag{5}$$

Let $\epsilon > 0$. Since $\{a_n\} \to a$, we may choose a natural number N such that if $n \geq N$, then $|a_n - a| < \epsilon/2$. We infer from (5) that if $n, m \geq N$, then $|a_n - a_m| < \epsilon$. Therefore the sequence $\{a_n\}$ is Cauchy. To prove the converse, let $\{a_n\}$ be a Cauchy sequence. We claim that it is bounded. Indeed, for $\epsilon = 1$, choose N such that if $n, m \geq N$, then $|a_n - a_m| < 1$. Thus

$$|a_n| = |(a_n - a_N) + a_N| \leq |a_n - a_N| + |a_N| \leq 1 + |a_N| \text{ for all } n \geq N.$$

Define $M = 1 + \max\{|a_1|, \ldots, |a_N|\}$. Then $|a_n| \leq M$ for all n. Thus $\{a_n\}$ is bounded. The Bolzano-Weierstrass Theorem tells us there is a subsequence $\{a_{n_k}\}$ which converges to a real number a. We claim that the whole sequence converges to a. Indeed, let $\epsilon > 0$. Since $\{a_n\}$ is Cauchy we may choose a natural number N such that

$$\text{if } n, m \geq N, \text{ then } |a_n - a_m| < \epsilon/2.$$

On the other hand, since $\{a_{n_k}\} \to a$ we may choose a natural number n_k such that $|a - a_{n_k}| < \epsilon/2$ and $n_k \geq N$. Therefore

$$|a_n - a| = |(a_n - a_{n_k}) + (a_{n_k} - a)| \leq |a_n - a_{n_k}| + |a - a_{n_k}| < \epsilon \text{ for all } n \geq N. \qquad \square$$

Theorem 18 (**Linearity and Monotonicity of Convergence of Real Sequences**) *Let $\{a_n\}$ and $\{b_n\}$ be convergent sequences of real numbers. Then for each pair of real numbers α and β, the sequence $\{\alpha \cdot a_n + \beta \cdot b_n\}$ is convergent and*

$$\lim_{n \to \infty} [\alpha \cdot a_n + \beta \cdot b_n] = \alpha \cdot \lim_{n \to \infty} a_n + \beta \cdot \lim_{n \to \infty} b_n. \tag{6}$$

Moreover,

$$\text{if } a_n \leq b_n \text{ for all } n, \text{ then } \lim_{n \to \infty} a_n \leq \lim_{n \to \infty} b_n. \tag{7}$$

Proof Define

$$\lim_{n \to \infty} a_n = a \text{ and } \lim_{n \to \infty} b_n = b.$$

Observe that

$$\big|[\alpha \cdot a_n + \beta \cdot b_n] - [\alpha \cdot a + \beta \cdot b]\big| \leq |\alpha| \cdot |a_n - a| + |\beta| \cdot |b_n - b| \text{ for all } n. \tag{8}$$

Let $\epsilon > 0$. Choose a natural number N such that

$$|a_n - a| < \epsilon/[2 + 2|\alpha|] \text{ and } |b_n - b| < \epsilon/[2 + 2|\beta|] \text{ for all } n \geq N.$$

We infer from (8) that

$$|[\alpha \cdot a_n + \beta \cdot b_n] - [\alpha \cdot a + \beta \cdot b]| < \epsilon \text{ for all } n \geq N.$$

Therefore (6) holds. To verify (7), set $c_n = b_n - a_n$ for all n and $c = b - a$, Then $c_n \geq 0$ for all n and, by linearity of convergence, $\{c_n\} \to c$. We must show $c \geq 0$. Let $\epsilon > 0$. There is an N such that

$$-\epsilon < c - c_n < \epsilon \text{ for all } n \geq N.$$

In particular, $0 \leq c_N < c + \epsilon$. Since $c > -\epsilon$ for every positive number $\epsilon, c \geq 0$. $\square$

If a sequence $\{a_n\}$ has the property that for each real number c, there is an index N such that if $n \geq N$, then $a_n \geq c$, we say that $\{a_n\}$ **converges to infinity**, call ∞ the limit of $\{a_n\}$, and write $\lim_{n \to \infty} a_n = \infty$. Similar definitions are made at $-\infty$. With this extended concept of convergence we may assert that any monotone sequence $\{a_n\}$ of real numbers, bounded or unbounded, converges to an extended real number and therefore $\lim_{n \to \infty} a_n$ is properly defined.

The extended concept of supremum and infimum of a set and of convergence for any monotone sequence of real numbers allows us to make the following definition.

Definition *Let $\{a_n\}$ be a sequence of real numbers. The limit superior of $\{a_n\}$, denoted by $\limsup\{a_n\}$, is defined by*

$$\limsup\{a_n\} = \lim_{n \to \infty} [\sup\{a_k \mid k \geq n\}].$$

The limit inferior of $\{a_n\}$, denoted by $\liminf\{a_n\}$, is defined by

$$\liminf\{a_n\} = \lim_{n \to \infty} [\inf\{a_k \mid k \geq n\}].$$

We leave the proof of the following proposition as an exercise.

Proposition 19 *Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers.*

(i) $\limsup\{a_n\} = \ell \in \mathbf{R}$ *if and only if for each $\epsilon > 0$, there are infinitely many indices n for which $a_n > \ell - \epsilon$ and only finitely many indices n for which $a_n > \ell + \epsilon$.*

(ii) $\limsup\{a_n\} = \infty$ *if and only if $\{a_n\}$ is not bounded above.*

(iii) $\limsup\{a_n\} = -\liminf\{-a_n\}$.

(iv) A sequence of real numbers $\{a_n\}$ converges to an extended real number a if and only if

$$\liminf\{a_n\} = \limsup\{a_n\} = a.$$

(v) If $a_n \leq b_n$ for all n, then

$$\limsup\{a_n\} \leq \limsup\{b_n\}.$$

For each sequence $\{a_k\}$ of real numbers, there corresponds a sequence of **partial sums** $\{s_n\}$ defined by $s_n = \sum_{k=1}^{n} a_k$ for each index n. We say that the series $\sum_{k=1}^{\infty} a_k$ is **summable** to the real number s provided $\{s_n\} \to s$ and write $s = \sum_{k=1}^{\infty} a_k$.

We leave the proof of the following proposition as an exercise.

Proposition 20 *Let $\{a_n\}$ be a sequence of real numbers.*

(i) The series $\sum_{k=1}^{\infty} a_k$ is summable if and only if for each $\epsilon > 0$, there is an index N for which

$$\left| \sum_{k=n}^{n+m} a_k \right| < \epsilon \text{ for } n \geq N \text{ and any natural number } m.$$

(ii) If the series $\sum_{k=1}^{\infty} |a_k|$ is summable, then $\sum_{k=1}^{\infty} a_k$ also is summable.

(iii) If each term a_k is non-negative, then the series $\sum_{k=1}^{\infty} a_k$ is summable if and only if the sequence of partial sums is bounded.

Consider the series $\sum_{k=1}^{\infty} a_k$. It is said to be **absolutely convergent** provided that the series $\sum_{k=1}^{\infty} |a_k|$ converges. Given a permutation $\pi\colon \mathbf{N} \to \mathbf{N}$, the series $\sum_{k=1}^{\infty} a_{\pi(k)}$ is called a **rearrangement** of $\sum_{k=1}^{\infty} a_k$.

Theorem 21 (**the Riemann Rearrangement Theorem**) *If a series converges absolutely, then every rearrangement converges to the same sum. If a series converges, but does not converge absolutely, then for every real number s, there is a rearrangement that converges to s.*

A proof of this remarkable theorem may be found in Terence Tao's Analysis 1.

PROBLEMS

38. We call an extended real number a **cluster point** of a sequence $\{a_n\}$ if a subsequence converges to this extended real number. Show that $\liminf\{a_n\}$ is the smallest cluster point of $\{a_n\}$ and $\limsup\{a_n\}$ is the largest cluster point of $\{a_n\}$.
39. Prove Proposition 19.
40. Show that a sequence $\{a_n\}$ is convergent to an extended real number if and only if there is exactly one extended real number that is a cluster point of the sequence.
41. Show that $\liminf a_n \leq \limsup a_n$.
42. Prove that if, for all n, $a_n \geq 0$ and $b_n \geq 0$, then
$$\limsup\,[a_n \cdot b_n] \leq (\limsup\ a_n) \cdot (\limsup\ b_n),$$
provided the product on the right is not of the form $0 \cdot \infty$.
43. Show that every real sequence has a monotone subsequence. Use this to provide another proof of the Bolzano-Weierstrass Theorem.
44. Let p be a natural number greater than 1, and x a real number, $0 \leq x \leq 1$. Show that there is a sequence $\{a_n\}$ of integers with $0 \leq a_n < p$ for each n such that
$$x = \sum_{n=1}^{\infty} \frac{a_n}{p^n}$$
and that this sequence is unique except when x is of the form $q/p^n, 0 < q < p^n$, in which case there are exactly two such sequences. Show that, conversely, if $\{a_n\}$ is any sequence of integers with $0 \leq a_n < p$, the series
$$\sum_{n=1}^{\infty} \frac{a_n}{p^n}$$
converges to a real number x with $0 \leq x \leq 1$. If $p = 10$, this sequence is called the *decimal* expansion of x. For $p = 2$ it is called the *binary* expansion; and for $p = 3$, the *ternary* expansion.

45. Prove Proposition 20.
46. Show that the assertion of the Bolzano-Weierstrass Theorem is equivalent to the Completeness Axiom for the real numbers. Show that the assertion of the Monotone Convergence Theorem is equivalent to the Completeness Axiom for the real numbers.

1.6 CONTINUOUS REAL-VALUED FUNCTIONS OF A REAL VARIABLE

Let f be a real-valued function defined on a set E of real numbers. We say that f is **continuous at the point** x in E provided that for each $\epsilon > 0$, there is a $\delta > 0$ for which

$$\text{if } x' \in E \text{ and } |x' - x| < \delta, \text{ then } |f(x') - f(x)| < \epsilon.$$

The function f is said to be **continuous** (on E) provided it is continuous at each point in its domain E. The function f is said to be **Lipschitz** provided there is a $c \geq 0$ for which

$$|f(x') - f(x)| \leq c \cdot |x' - x| \text{ for all } x', x \in E.$$

It is clear that a Lipschitz function is continuous. Indeed, for a number $x \in E$ and any $\epsilon > 0$, $\delta = \epsilon / c$ responds to the ϵ challenge regarding the criterion for the continuity of f at x. Not all continuous functions are Lipschitz. For example, if $f(x) = \sqrt{x}$ for $0 \leq x \leq 1$, then f is continuous on $[0, 1]$ but is not Lipschitz.

We leave as an exercise the proof of the following characterization of continuity at a point in terms of sequential convergence.

Proposition 22 *A real-valued function f defined on a set E of real numbers is continuous at the point $x_* \in E$ if and only if whenever a sequence $\{x_n\}$ in E converges to x_*, its image sequence $\{f(x_n)\}$ converges to $f(x_*)$.*

We have the following characterization of continuity of a function on all of its domain.

Proposition 23 *Let f be a real-valued function defined on a set E of real numbers. Then f is continuous on E if and only if for each open set $\mathcal{O}$,*

$$f^{-1}(\mathcal{O}) = E \cap \mathcal{U} \text{ where } \mathcal{U} \text{ is an open set.} \tag{9}$$

Proof First assume the inverse image under f of any open set is the intersection of the domain with an open set. Let x belong to E. To show that f is continuous at x, let $\epsilon > 0$. The interval $I = (f(x) - \epsilon,\ f(x) + \epsilon)$ is an open set. Therefore, there is an open set $\mathcal{U}$ such that

$$f^{-1}(I) = \{x' \in E \mid f(x) - \epsilon < f(x') < f(x) + \epsilon\} = E \cap \mathcal{U}.$$

In particular, $f(E \cap \mathcal{U}) \subseteq I$ and x belongs to $E \cap \mathcal{U}$. Since $\mathcal{U}$ is open, there is a $\delta > 0$ such that $(x - \delta,\ x + \delta) \subseteq \mathcal{U}$. Thus if $x' \in E$ and $|x' - x| < \delta$, then $|f(x') - f(x)| < \epsilon$. Hence f is continuous at x.

Suppose now that f is continuous. Let $\mathcal{O}$ be an open set and x belong to $f^{-1}(\mathcal{O})$. Then $f(x)$ belongs to the open set $\mathcal{O}$ so that there is an $\epsilon > 0$, such that $(f(x) - \epsilon,\ f(x) + \epsilon) \subseteq \mathcal{O}$.

Since f is continuous at x, there is a $\delta > 0$ such that if x' belongs to E and $|x' - x| < \delta$, then $|f(x') - f(x)| < \epsilon$. Define $I_x = (x - \delta, x + \delta)$. Then $f(E \cap I_x) \subseteq \mathcal{O}$. Define

$$\mathcal{U} = \bigcup_{x \in f^{-1}(\mathcal{O})} I_x.$$

Since $\mathcal{U}$ is the union of open sets it is open. It has been constructed so that (9) holds. $\square$

The Extreme Value Theorem *A continuous real-valued function on a non-empty, closed, bounded set of real numbers takes a minimum and maximum value.*

Proof Let f be a continuous real-valued function on the non-empty closed bounded set E of real numbers. We first show that f is bounded on E, that is, there is a real number M such that

$$|f(x)| \leq M \text{ for all } x \in E. \tag{10}$$

Let x belong to E. Let $\delta > 0$ respond to the $\epsilon = 1$ challenge regarding the criterion for continuity of f at x. Define $I_x = (x - \delta,\, x + \delta)$. Therefore, if x' belongs to $E \cap I_x$, then $|f(x') - f(x)| < 1$ and so $|f(x')| \leq |f(x)| + 1$. The collection $\{I_x\}_{x \in E}$ is an open cover of E. The Heine-Borel Theorem tells us that there are a finite number of points $\{x_1, \ldots, x_n\}$ in E such that $\{I_{x_k}\}_{k=1}^n$ also covers E. Define $M = 1 + \max\{|f(x_1)|, \ldots, |f(x_n)|\}$. We claim that (10) holds for this choice of E. Indeed, let x belong to E. There is an index k such that x belongs to I_{x_k} and therefore $|f(x)| \leq 1 + |f(x_k)| \leq M$. To see that f takes a maximum value on E, define $m = \sup f(E)$. If f failed to take the value m on E, then the function $x \mapsto 1/(f(x) - m), x \in E$ is a continuous function on E which is unbounded. This contradicts what we have just proved. Therefore, f takes a maximum value on E. Since $-f$ is continuous, $-f$ takes a maximum value, that is, f takes a minimum value on E. $\square$

The Intermediate Value Theorem *Let f be a continuous real-valued function on the closed, bounded interval $[a, b]$ for which $f(a) < c < f(b)$. Then there is a point x_0 in (a, b) at which $f(x_0) = c$.*

Proof We will define by induction a descending countable collection $\{[a_n, b_n]\}_{n=1}^\infty$ of closed intervals whose intersection consists of a single point $x_0 \in (a, b)$ at which $f(x_0) = c$. Define $a_1 = a$ and $b_1 = b$. Consider the midpoint m_1 of $[a_1, b_1]$. If $c < f(m_1)$, define $a_2 = a_1$ and $b_2 = m_1$. If $f(m_1) \geq c$, define $a_2 = m_1$ and $b_2 = b_1$. Therefore, $f(a_2) \leq c \leq f(b_2)$ and $b_2 - a_2 = [b_1 - a_1]/2$. We inductively continue this bisection process to obtain a descending collection $\{[a_n, b_n]\}_{n=1}^\infty$ of closed intervals such that

$$f(a_n) \leq c \leq f(b_n) \text{ and } b_n - a_n = [b - a]/2^{n-1} \text{ for all } n. \tag{11}$$

According to the Nested Set Theorem, $\bigcap_{n=1}^\infty [a_n, b_n]$ is non-empty. Choose x_0 in $\bigcap_{n=1}^\infty [a_n, b_n]$. Observe that

$$|a_n - x_0| \leq b_n - a_n = [b - a]/2^{n-1} \text{ for all } n.$$

Therefore $\{a_n\} \to x_0$. By the continuity of f at x_0, $\{f(a_n)\} \to f(x_0)$. Since $f(a_n) \leq c$ for all n, and the set $(-\infty, c]$ is closed, $f(x_0) \leq c$. By a similar argument, $f(x_0) \geq c$. Hence $f(x_0) = c$. $\square$

Definition *A real-valued function f defined on a set E of real numbers is said to be* ***uniformly continuous*** *provided for each $\epsilon > 0$, there is a $\delta > 0$ such that for all x, x' in E,*

$$\text{if } |x - x'| < \delta, \text{ then } |f(x) - f(x')| < \epsilon.$$

Theorem 24 *A continuous real-valued function on a closed, bounded set of real numbers is uniformly continuous.*

Proof Let f be a continuous real-valued function on a closed bounded set E of real numbers. Let $\epsilon > 0$. For each $x \in E$, there is a $\delta_x > 0$ such that if $x' \in E$ and $|x' - x| < \delta_x$, then $|f(x') - f(x)| < \epsilon/2$. Define I_x to be the open interval $(x - \delta_x/2,\ x + \delta_x/2)$. Then $\{I_x\}_{x \in E}$ is an open cover of E. According to the Heine-Borel Theorem, there is a finite subcollection $\{I_{x_1}, \ldots, I_{x_n}\}$ which covers E. Define

$$\delta = \frac{1}{2}\min\{\delta_{x_1}, \ldots, \delta_{x_n}\}.$$

We claim that this $\delta > 0$ responds to the $\epsilon > 0$ challenge regarding the criterion for f to be uniformly continuous on E. Indeed, let x and x' belong to E with $|x - x'| < \delta$. Since $\{I_{x_1}, \ldots, I_{x_n}\}$ covers E, there is an index k for which $|x - x_k| < \delta_{x_k}/2$. Since $|x - x'| < \delta \leq \delta_{x_k}/2$,

$$|x' - x_k| \leq |x' - x| + |x - x_k| < \delta_{x_k}/2 + \delta_{x_k}/2 = \delta_{x_k}.$$

By the definition of δ_{x_k}, since $|x - x_k| < \delta_{x_k}$ and $|x' - x_k| < \delta_{x_k}$ we have $|f(x) - f(x_k)| < \epsilon/2$ and $|f(x') - f(x_k)| < \epsilon/2$. Therefore,

$$|f(x) - f(x')| \leq |f(x) - f(x_k)| + |f(x') - f(x_k)| < \epsilon/2 + \epsilon/2 = \epsilon. \qquad \square$$

Definition *A real-valued function f defined on a set E of real numbers is said to be* **increasing** *provided $f(x) \leq f(x')$ whenever x, x' belong to E and $x \leq x'$, and* **decreasing** *provided $-f$ is increasing. It is called* **monotone** *if it is either increasing or decreasing.*

Let f be a monotone real-valued function defined on an open interval I that contains the point x_0. We infer from Theorem 15 and its proof that if $\{x_n\}$ is a decreasing sequence in $I \cap (x_0, \infty)$ which converges to x_0, then the sequence $\{f(x_n)\}$ converges to a real number and the limit is independent of the choice of sequence $\{x_n\}$. We denote the limit by $f(x_0^+)$. Similarly, we define $f(x_0^-)$. Then clearly f is continuous at x_0 if and only if $f(x_0^-) = f(x_0) = f(x_0^+)$. If f fails to be continuous at x_0, then the only point of the image of f that lies strictly between $f(x_0^+)$ and $f(x_0^-)$ is $f(x_0)$ and f is said to have a **jump discontinuity** at x_0. Thus, by the Intermediate Value Theorem, a monotone function on an open interval is continuous if and only if its image is an interval (see Problem 55).

PROBLEMS

47. Let E be a closed set of real numbers and f a real-valued function that is defined and continuous on E. Show that there is a function g defined and continuous on all of $\mathbf{R}$ such that $f(x) = g(x)$ for each $x \in E$. (Hint: Take g to be linear on each of the intervals of which $\mathbf{R} \sim E$ is composed.)

48. Define the real-valued function f on $\mathbf{R}$ by setting

$$f(x) = \begin{cases} x & \text{if } x \text{ irrational} \\ p \sin \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ in lowest terms.} \end{cases}$$

At what points is f continuous?

49. Let f and g be continuous real-valued functions with a common domain E.
   (i) Show that the sum, $f + g$, and product, fg, are also continuous functions.
   (ii) If h is a continuous function with image contained in E, show that the composition $f \circ h$ is continuous.
   (iii) Let $\max\{f, g\}$ be the function defined by $\max\{f, g\}(x) = \max\{f(x), g(x)\}$, for $x \in E$. Show that $\max\{f, g\}$ is continuous.
   (iv) Show that $|f|$ is continuous.

50. Show that a Lipschitz function is uniformly continuous but there are uniformly continuous functions that are not Lipschitz.

51. A continuous function φ on $[a, b]$ is called **piecewise linear** provided there is a partition $a = x_0 < x_1 < \cdots < x_n = b$ of $[a, b]$ for which φ is linear on each interval $[x_i, x_{i+1}]$. Let f be a continuous function on $[a, b]$ and ϵ a positive number. Show that there is a piecewise linear function φ on $[a, b]$ with $|f(x) - \varphi(x)| < \epsilon$ for all $x \in [a, b]$.

52. Show that a non-empty set E of real numbers is closed and bounded if and only if every continuous real-valued function on E takes a maximum value.

53. Show that a set E of real numbers is closed and bounded if and only if every open cover of E has a finite subcover.

54. Show that a non-empty set E of real numbers is an interval if and only if every continuous real-valued function on E has an interval as its image.

55. Show that a monotone function on an open interval is continuous if and only if its image is an interval.

56. Let f be a real-valued function defined on $\mathbf{R}$. Show that the set of points at which f is continuous is a G_δ set.

57. Let $\{f_n\}$ be a sequence of continuous functions defined on $\mathbf{R}$. Show that the set of points x at which the sequence $\{f_n(x)\}$ converges to a real number is the intersection of a countable collection of F_σ sets.

58. Let f be a continuous real-valued function on $\mathbf{R}$. Show that the inverse image with respect to f of an open set is open, of a closed set is closed, and of a Borel set is Borel.

59. A sequence $\{f_n\}$ of real-valued functions defined on a set E is said to converge uniformly on E to a function f if given $\epsilon > 0$, there is an N such that for all $x \in E$ and all $n \geq N$, we have $|f_n(x) - f(x)| < \epsilon$. Let $\{f_n\}$ be a sequence of continuous functions defined on a set E. Prove that if $\{f_n\}$ converges uniformly to f on E, then f is continuous on E.

CHAPTER 2

Lebesgue Measure

Contents

2.1 Introduction 29
2.2 Outer-Measure 31
2.3 The σ-Algebra of Lebesgue Measurable Sets 34
2.4 Finer Properties of Measurable Sets 39
2.5 Countable Additivity and Continuity of Measure, and the Borel-Cantelli Lemma 42
2.6 Vitali's Example of a Non-measurable Set 45
2.7 The Cantor Set and the Cantor-Lebesgue Function 46

2.1 INTRODUCTION

The Riemann integral of a bounded function over a closed, bounded interval is defined using approximations of the function by step-functions, which are constructed by partitioning its domain into subintervals. The generalization of the Riemann integral to the Lebesgue integral will be achieved by using approximations of the function by simple functions, which are constructed by partitioning into intervals the range of the function and considering preimages of these intervals. In this chapter, properties of individual measurable sets and of the collection of measurable sets are established.

Each interval is Lebesgue measurable. The richness of the collection of Lebesgue measurable sets provides better upper and lower approximations of a function, and therefore of its integral, than are possible by just employing step-functions. This leads to a larger class of functions that are Lebesgue integrable over very general domains and an integral that has better properties. For instance, under quite general circumstances, we prove that if a sequence of functions converges pointwise to a limiting function, then the integral of the limit function is the limit of the integrals of the approximating functions.

The length $\ell(I)$ of an interval I is defined to be the difference of the end-points of I if I is bounded, and ∞ if I is unbounded. Length is an example of a *set-function*, that is, a function that associates an extended real number to each set in a collection of sets. In the case of length, the domain is the collection of all intervals. In this chapter, the set-function length is extended to a large collection of sets of real numbers. For instance, the "length" of an open set is the sum of the lengths of the countable number of open intervals of which it is composed. However, the collection of sets consisting of intervals and open sets is still too limited for our purposes. We construct a collection of sets called **Lebesgue measurable sets**, and a set-function of this collection called **Lebesgue measure**, denoted by m. The collection of Lebesgue measurable sets is a σ-algebra[1] which contains all open sets and all closed sets. The set-function m possesses the following three properties.

[1] Recall that a collection of subsets of $\mathbf{R}$ is called a σ-algebra provided it contains $\mathbf{R}$ and is closed with respect to complements and countable unions; by De Morgan's Identities, such a collection is also closed with respect to countable intersections.

The measure of an interval is its length Each interval I is Lebesgue measurable and
$$m(I) = \ell(I).$$

Measure is translation invariant If E is Lebesgue measurable and y is any number, then the translate of E by y, $E + y = \{x + y \mid x \in E\}$, also is Lebesgue measurable and
$$m(E + y) = m(E).$$

Measure is countably additive over countable, disjoint unions of sets[2] If $\{E_k\}_{k=1}^{\infty}$ is a countable, disjoint collection of Lebesgue measurable sets, then
$$m\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} m(E_k).$$

It is not possible to construct a set-function that possesses the above three properties and is defined for all sets of real numbers. The response to this limitation is to construct a set-function on a very rich class of sets that does possess the above three properties. The construction has two stages.

We first construct a set-function called **outer-measure**, which we denote by m^*. It is defined for any set, and thus, in particular, for any interval. The outer-measure of an interval is its length. Outer-measure is translation invariant. However, outer-measure is not finitely additive, much less countably additive (see Corollary 24). But it is countably monotone in the sense that if $\{E_k\}_{k=1}^{\infty}$ is any countable collection of sets, disjoint or not, that covers a set E, then
$$m^*(E) \leq m^*\left(\bigcup_{k=1}^{\infty} E_k\right) \leq \sum_{k=1}^{\infty} m^*(E_k).$$
The second stage in the construction is to determine what it means for a set to be **Lebesgue measurable** and show that the collection of Lebesgue measurable sets is a σ-algebra that contains all open sets, and all sets of outer-measure zero. We then restrict the set-function m^* to the collection of Lebesgue measurable sets, denote it by m, and prove m is countably additive. We call m **Lebesgue measure**.

PROBLEMS

In the first three problems, let μ be a set-function defined for all sets in a σ-algebra $\mathcal{A}$ with values in $[0, \infty]$, and μ be countably additive over countable disjoint collections of sets in $\mathcal{A}$.

1. Prove that if A and B are two sets in $\mathcal{A}$ with $A \subseteq B$, then $\mu(A) \leq \mu(B)$. This property is called *monotonicity*.
2. Prove that if there is a set A in the collection $\mathcal{A}$ for which $\mu(A) < \infty$, then $\mu(\emptyset) = 0$.
3. Let $\{E_k\}_{k=1}^{\infty}$ be a countable collection of sets in $\mathcal{A}$. Prove that $\mu(\bigcup_{k=1}^{\infty} E_k) \leq \sum_{k=1}^{\infty} \mu(E_k)$.
4. A set-function c is defined on all subsets of $\mathbf{R}$ as follows: define $c(E)$ to be ∞ if E has infinitely many members and $c(E)$ to be the number of members in E if E is finite, and define $c(\emptyset) = 0$. Show that c is a countably additive and translation invariant set-function. This set-function is called the **counting measure**.

[2]For a collection of sets to be disjoint, we mean what is sometimes called pairwise disjoint, that is, that each pair of sets in the collection has empty intersection.

2.2 OUTER-MEASURE

Let I be an interval of real numbers. Define its length, $\ell(I)$, to be ∞ if I is unbounded and otherwise define its length to be the difference of its end-points. For a set A of real numbers, consider all countable collections $\{I_k\}_{k=1}^{\infty}$ of open, bounded intervals that cover A, in the sense that $A \subseteq \bigcup_{k=1}^{\infty} I_k$. For each such collection, consider the sum of the lengths of the intervals in the collection. The **outer-measure** of $A, m^*(A)$, is defined to be the infimum of all such sums, that is,

$$m^*(A) = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) \;\middle|\; A \subseteq \bigcup_{k=1}^{\infty} I_k \right\}.$$

It follows immediately from the definition of outer-measure that $m^*(\emptyset) = 0$. Moreover, since any cover of a set B is also a cover of any subset of B, outer-measure is **monotone** in the sense that

$$\text{if } A \subseteq B, \text{ then } m^*(A) \leq m^*(B).$$

Example A countable set C has outer-measure zero. Indeed, let C be enumerated as $C = \{c_k\}_{k=1}^{\infty}$. Let $\epsilon > 0$. For each k, define $I_k = (c_k - \epsilon/2^{k+1},\ c_k + \epsilon/2^{k+1})$. The countable collection of open intervals $\{I_k\}_{k=1}^{\infty}$ covers C. Therefore,

$$0 \leq m^*(C) \leq \sum_{k=1}^{\infty} \ell(I_k) = \sum_{k=1}^{\infty} \epsilon/2^k = \epsilon.$$

This inequality holds for each $\epsilon > 0$, and so $m^*(C) = 0$. According to Proposition 25, there is a non-countable set, the Cantor set, which has measure zero.

Lemma 1 **(von Neumann)** *For a bounded set E, define its* **integral count**, *$\mu^{int}(E)$, to be the number of integers in E. For each $\epsilon > 0$, define the ϵ-dilation T_ϵ: $\mathbf{R} \to \mathbf{R}$ by $T_\epsilon(x) = \epsilon \cdot x$. Then for each bounded interval I,*

$$\lim_{\epsilon \to \infty} \frac{\mu^{int}(T_\epsilon(I))}{\epsilon} = \ell(I). \tag{1}$$

Proof If I has end-points a and b, there is the following estimate for $\mu^{int}(I)$:

$$(b-a) - 1 \leq \mu^{int}(I) \leq (b-a) + 1.$$

For $\epsilon > 0$, replace the interval I by the dilated interval $T_\epsilon(I)$ to obtain the estimate

$$\epsilon \cdot (b-a) - 1 \leq \mu^{int}(T_\epsilon(I)) \leq \epsilon \cdot (b-a) + 1].$$

Divide this inequality by ϵ and take the limit as $\epsilon \to \infty$ to establish (1). $\square$

Proposition 2 *If the bounded interval I is covered by a finite collection $\{I_k\}_{k=1}^{n}$ of bounded intervals, then*

$$\ell(I) \leq \sum_{k=1}^{n} \ell(I_k).$$

Proof For each $\epsilon > 0$, the bounded interval $T_\epsilon(I)$ is covered by the collection of bounded intervals $\{T_\epsilon(I^k)\}_{k=1}^m$. It is clear that the integer count is finitely monotone, and so

$$\mu^{int}(T_\epsilon(I)) \leq \sum_{k=1}^{m} \mu^{int}(T_\epsilon(I^k)) \text{ for all } \epsilon.$$

Divide each side by ϵ, take the limit as $\epsilon \to \infty$ and, by the preceding lemma, conclude the proof. □

Proposition 3 *The outer-measure of an interval is its length.*

Proof We begin with the case of a closed, bounded interval $[a, b]$. Let $\epsilon > 0$. Since the open interval $(a - \epsilon, b + \epsilon)$ contains $[a, b]$, we have $m^*([a, b]) \leq \ell((a - \epsilon, b + \epsilon)) = b - a + 2\epsilon$. This holds for any $\epsilon > 0$, and so $m^*([a, b]) \leq b - a$. It remains to verify this inequality in the opposite direction, which this is equivalent to showing that if $\{I_k\}_{k=1}^\infty$ is any countable collection of open, bounded intervals covering $[a, b]$, then

$$b - a \leq \sum_{k=1}^{\infty} \ell(I_k).$$

By the Heine-Borel Theorem[3], any collection of open intervals covering $[a, b]$ has a finite subcollection that also covers $[a, b]$. Choose an index n for which $\{I_k\}_{k=1}^n$ covers $[a, b]$. To verify the above inequality, it suffices to show that

$$b - a \leq \sum_{k=1}^{n} \ell(I_k).$$

However, this follows from the preceding proposition. Now consider the case of a general bounded interval I. Given $\epsilon > 0$, there are two closed, bounded intervals J_1 and J_2 such that

$$J_1 \subseteq I \subseteq J_2 \text{ while } \ell(I) - \epsilon < \ell(J_1) \text{ and } \ell(J_2) < \ell(I) + \epsilon.$$

By the equality of outer-measure and length for closed bounded intervals and the monotonicity of outer-measure,

$$\ell(I) - \epsilon < \ell(J_1) = m^*(J_1) \leq m^*(I) \leq m^*(J_2) = \ell(J_2) < \ell(I) + \epsilon.$$

This holds for each $\epsilon > 0$. Therefore, $\ell(I) = m^*(I)$.

If I is an unbounded interval, then for each natural number n, there is an interval $J \subseteq I$ with $\ell(J) = n$. We have $m^*(I) \geq m^*(J) = \ell(J) = n$. This holds for each n, and therefore $m^*(I) = \infty$. □

Proposition 4 *Outer-measure is* **translation invariant**, *in the sense that for any set* E *and any* c, *if* $E + c = \{x + c | x \in E\}$,

$$m^*(E + c) = m^*(E).$$

---

[3]See page 18.

Proof Observe that if $\{I_k\}_{k=1}^{\infty}$ is any countable collection of sets, then $\{I_k\}_{k=1}^{\infty}$ covers E if and only if $\{I_k+c\}_{k=1}^{\infty}$ covers $E+c$. Moreover, if each I_k is an open interval, then each I_k+c is an open interval of the same length and so

$$\sum_{k=1}^{\infty} \ell(I_k) = \sum_{k=1}^{\infty} \ell(I_k+c).$$

It follows that $m^*(E+c)=m^*(E)$. □

Proposition 5 *Outer-measure is* **countably monotone**, *in the sense that if $\{E_k\}_{k=1}^{\infty}$ is any countable collection of sets, disjoint or not, that covers a set E, then*

$$m^*(E) \le m^*\left(\bigcup_{k=1}^{\infty} E_k\right) \le \sum_{k=1}^{\infty} m^*(E_k).$$

Proof The left-hand inequality follows from the monotonicity of outer-measure. If one of the E_k's has infinite outer-measure, the right-hand inequality holds trivially. We therefore assume that each of the E_k's has finite outer-measure. Let $\epsilon>0$. For each k, there is a countable collection $\{I_{k,i}\}_{i=1}^{\infty}$ of open, bounded intervals for which

$$E_k \subseteq \bigcup_{i=1}^{\infty} I_{k,i} \text{ and } \sum_{i=1}^{\infty} \ell(I_{k,i}) < m^*(E_k)+\epsilon/2^k.$$

Now, $\{I_{k,i}\}_{1\le k,i\le\infty}$ is a countable collection of open, bounded intervals that covers $\bigcup_{k=1}^{\infty} E_k$, and consequently,

$$\begin{aligned} m^*\left(\bigcup_{k=1}^{\infty} E_k\right) \le \sum_{1\le k,i<\infty} \ell(I_{k,i}) &= \sum_{k=1}^{\infty}\left[\sum_{i=1}^{\infty} \ell(I_{k,i})\right] \\ &< \sum_{k=1}^{\infty}\left[m^*(E_k)+\epsilon/2^k\right] \\ &= \left[\sum_{k=1}^{\infty} m^*(E_k)\right]+\epsilon. \end{aligned}$$

Since this holds for each $\epsilon>0$, it also holds for $\epsilon=0$. □

PROBLEMS

5. By using properties of outer-measure, prove that the interval $[0, 1]$ is not countable.
6. Let A be the set of irrational numbers in the interval $[0, 1]$. Prove that $m^*(A)=1$.
7. A set of real numbers is said to be a G_δ set provided that it is the intersection of a countable collection of open sets. Show that for any bounded set E, there is a G_δ set G for which

$$E \subseteq G \text{ and } m^*(G)=m^*(E).$$

8. (Jordan content) The Jordan content of a set is defined as is outer-measure m^*, except that only finite coverings of the set by open, bounded intervals are considered. Prove that if the set $Q \cap [0, 1]$ is covered by the finite collection $\{\ell(I_k)\}_{k=1}^n$, then $\sum_{k=1}^n \ell(I_k) \geq 1$.
9. Suppose that outer-measure is defined by covering sets by countable collections of closed, bounded intervals rather than coverings by open, bounded intervals. Show that the outer-measure remains unchanged.
10. Prove that if $m^*(A) = 0$, then, for any set B, $m^*(A \cup B) = m^*(B)$.
11. Let A and B be bounded sets for which there is an $\alpha > 0$ such that $|a - b| \geq \alpha$ for all $a \in A, b \in B$. Prove that $m^*(A \cup B) = m^*(A) + m^*(B)$.

2.3 THE σ-ALGEBRA OF LEBESGUE MEASURABLE SETS

Outer-measure has four virtues: (i) it is defined for all sets of real numbers; (ii) the outer-measure of an interval is its length, (iii) it is countably monotone, and (iv) it is translation invariant. But outer-measure fails even to be finitely additive. According to Corollary 24, there are disjoint sets A and B for which

$$m^*(A \cup B) < m^*(A) + m^*(B). \tag{2}$$

To ameliorate this fundamental defect, we select a σ-algebra of sets, called the Lebesgue measurable sets, which contains all open sets and all sets of outer-measure zero, and the restriction of the set-function outer-measure to the collection of Lebesgue measurable sets is countably additive. There are a number of ways to define what it means for a set to be measurable. We follow an approach due to Constantin Carathéodory.

Definition *A set E is said to be **Lebesgue measurable**, or simply **measurable**, provided that for any set A*[4],

$$m^*(A) = m^*(A \cap E) + m^*(A \sim E).$$

The collection of measurable sets is denoted by $\mathcal{M}$.

We immediately see one advantage possessed by measurable sets, namely, that the strict inequality (2) cannot occur if one of the sets is measurable. Indeed, if, say, A is measurable and B is any set disjoint from A, then

$$m^*(A \cup B) = m^*([A \cup B] \cap A)) + m^*([A \cup B)] \sim A) = m^*(A) + m^*(B).$$

So, if a set E is measurable, then outer-measure m^* is additive over particular partitions of any set A, namely, as $A = [A \cap E] \cup [A \sim E]$. Since outer-measure is finitely monotone and $A = [A \cap E] \cup [A \sim E]$, we always have

$$m^*(A) \leq m^*(A \cap E) + m^*(A \sim E).$$

[4]Recall that we denote by $A \sim B$ the set $\{x \in A \mid x \notin B\}$, the **relative complement** of B in A.

Therefore, E is measurable if and only if for each set A,

$$m^*(A) \geq m^*(A \cap E) + m^*(A \sim E). \tag{3}$$

This inequality trivially holds if $m^*(A) = \infty$. Consequently, it suffices to establish (3) for sets A for which $m^*(A) < \infty$.

Observe that the definition of measurability is symmetric in E and $\mathbf{R} \sim E$, and therefore a set is measurable if and only if its complement is measurable. Clearly, the set $\mathbf{R}$ is measurable.

Lebesgue outer-measure has the following **excision property**.

Proposition 6 *If $E_0 \subseteq E$, E_0 is measurable and $m^*(E_0) < \infty$, then*

$$m^*(E \sim E_0) = m^*(E) - m^*(E_0). \tag{4}$$

Proof Since E_0 is measurable,

$$m^*(E) = m^*(E_0) + m^*(E \sim E_0),$$

and therefore, since $m^*(E_0) < \infty$, (4) holds. □

Proposition 7 *Any set of outer-measure zero is measurable.*

Proof Assume that $m^*(E) = 0$. Let A be any set. Since

$$A \cap E \subseteq E \text{ and } A \sim E \subseteq A,$$

by the monotonicity of outer-measure,

$$m^*(A \cap E) \leq m^*(E) = 0 \text{ and } m^*(A \sim E) \leq m^*(A).$$

Therefore,

$$m^*(A) \geq m^*(A \sim E) = 0 + m^*(A \sim E) = m^*(A \cap E) + m^*(A \sim E),$$

and so E is measurable. □

Proposition 8 *The translate $E + c$ of a measurable set E is measurable.*

Proof By the translation invariance of outer-measure and the measurability of E, for any set A,

$$\begin{aligned} m^*(A) = m^*(A - c) &= m^*([A - c] \cap E) + m^*([A - c] \sim E) \\ &= m^*(A \cap [E + c]) + m^*(A \sim [E + c]). \end{aligned}$$

Therefore, $E + c$ is measurable. □

Proposition 9 *The union of a finite collection of measurable sets is measurable.*

Proof We show that the union of two measurable sets E_1 and E_2 is measurable. The general case follows by induction. Let A be any set. First using the measurability of E_1, and then the measurability of E_2, we have

$$m^*(A) = m^*(A \cap E_1) + m^*(A \sim E_1)$$

$$= m^*(A \cap E_1) + m^*([A \sim E_1] \cap E_2) + m^*([A \sim E_1] \sim E_2).$$

There are the following set identities:

$$[A \sim E_1] \sim E_2 = A \sim [E_1 \cup E_2]$$

and

$$[A \cap E_1] \cup [[A \sim E_1] \cap E_2] = A \cap [E_1 \cup E_2].$$

It follows from these identities and the finite monotonicity of outer-measure that

$$m^*(A) = m^*(A \cap E_1) + m^*([A \sim E_1] \cap E_2) + m^*([A \sim E_1] \sim E_2)$$

$$= m^*(A \cap E_1) + m^*([A \sim E_1] \cap E_2) + m^*(A \sim [E_1 \cup E_2])$$

$$\geq m^*(A \cap [E_1 \cup E_2]) + m^*(A \sim [E_1 \cup E_2]).$$

Therefore, $E_1 \cup E_2$ is measurable. □

A collection of subsets of **R** is called an **algebra** provided that it contains **R** and is closed with respect to relative complements and finite unions; by De Morgan's Identities, such a collection is also closed with respect to finite intersections. It follows from this proposition, together with the measurability of the complement of a measurable set, that the collection $\mathcal{M}$ of measurable sets is an algebra.

Proposition 10 *If A is any set and $\{E_k\}_{k=1}^n$ is a finite, disjoint collection of measurable sets, then*

$$m^*\left(A \cap \left[\bigcup_{k=1}^n E_k\right]\right) = \sum_{k=1}^n m^*(A \cap E_k).$$

In particular,

$$m^*\left(\bigcup_{k=1}^n E_k\right) = \sum_{k=1}^n m^*(E_k).$$

Proof The proof proceeds by induction on n. It is clearly true for $n = 1$. Assume that it is true for $n - 1$. Since the collection $\{E_k\}_{k=1}^n$ is disjoint,

$$A \cap \left[\bigcup_{k=1}^n E_k\right] \cap E_n = A \cap E_n$$

and

$$A \cap \left[\bigcup_{k=1}^n E_k\right] \sim E_n = A \cap \left[\bigcup_{k=1}^{n-1} E_k\right].$$

Consequently, by the measurability of E_n and the induction assumption,

$$m^*\left(A \cap \left[\bigcup_{k=1}^{n} E_k\right]\right) = m^*(A \cap E_n) + m^*\left(A \cap \left[\bigcup_{k=1}^{n-1} E_k\right]\right)$$

$$= m^*(A \cap E_n) + \sum_{k=1}^{n-1} m^*(A \cap E_k)$$

$$= \sum_{k=1}^{n} m^*(A \cap E_k).$$

□

Definition *A countable, disjoint collection $\{E_k\}_{k=1}^\infty$ of measurable subsets of E is called a* **measurable partition** *of E provided that $E = \bigcup_{k=1}^\infty E_k$.*

Lemma 11 *Let $E = \bigcup_{k=1}^\infty E_k$, a countable union of measurable sets. Then there is a measurable partition $\{E'_k\}_{k=1}^\infty$ of E for which each $E'_k \subseteq E_k$.*

Proof Define $E'_1 = E_1$ and for each $k \geq 2$, define

$$E'_k = E_k \sim \bigcup_{i=1}^{k-1} E_i.$$

Since the collection of measurable sets is an algebra, each E'_k is measurable, The collection $\{E'_k\}_{k=1}^\infty$ was constructed to be disjoint, and it is a measurable partition of E, since for each $x \in E$, there is a first index k for which $x \in E_k$. and so $x \in E'_k$. □

Proposition 12 *The union of a countable collection of measurable sets is measurable.*

Proof Let E be the union of a countable collection of measurable sets. By the preceding lemma, there is a measurable partition $\{E_k\}_{k=1}^\infty$ of E. Let $A \subseteq \mathbf{R}$ be any set. For each n, define $F_n = \bigcup_{k=1}^n E_k$. The measurable sets are an algebra, and therefore F_n is measurable, so that, by the inclusion $F_n \subseteq E$,

$$m^*(A) = m^*(A \cap F_n) + m^*(A \sim F_n) \geq m^*(A \cap F_n) + m^*(A \sim E).$$

By the preceding proposition,

$$m^*(A \cap F_n) = \sum_{k=1}^{n} m^*(A \cap E_k).$$

Therefore,

$$m^*(A) \geq \sum_{k=1}^{n} m^*(A \cap E_k) + m^*(A \sim E).$$

The left-hand side of this inequality is independent of n, so that

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*(A \sim E).$$

Consequently, by the countable monotonicity of outer-measure,

$$m^*(A) \geq m^*(A \cap E) + m^*(A \sim E),$$

and therefore E is measurable. □

Proposition 13 *Every interval is measurable.*

Proof Since the collection of measurable sets is a σ-algebra, to show that every interval is measurable, it suffices to show that every interval of the form (a, ∞) is measurable (see Problem 12). Consider such an interval. Let A be any set. Assume that a does not belong to A. Otherwise, replace A by $A \sim \{a\}$, and, of course, if $A \sim \{a\}$ is measurable, so is A. We must show that

$$m^*(A_1) + m^*(A_2) \leq m^*(A), \tag{5}$$

where

$$A_1 = A \cap (-\infty, a) \text{ and } A_2 = A \cap (a, \infty).$$

By the definition of $m^*(A)$ as an infimum, to verify (5), it is necessary and sufficient to show that for any countable collection $\{I_k\}_{k=1}^\infty$ of open, bounded intervals that covers A,

$$m^*(A_1) + m^*(A_2) \leq \sum_{k=1}^\infty \ell(I_k). \tag{6}$$

For such a covering and each index k, define

$$I_k' = I_k \cap (-\infty, a) \text{ and } I_k'' = I_k \cap (a, \infty).$$

Then I_k' and I_k'' are intervals and

$$\ell(I_k) = \ell(I_k') + \ell(I_k'').$$

Since $\{I_k'\}_{k=1}^\infty$ and $\{I_k''\}_{k=1}^\infty$ are countable collections of open, bounded intervals that cover A_1 and A_2, respectively,

$$m^*(A_1) \leq \sum_{k=1}^\infty \ell(I_k') \text{ and } m^*(A_2) \leq \sum_{k=1}^\infty \ell(I_k'').$$

Consequently,

$$\begin{aligned} m^*(A_1) + m^*(A_2) &\leq \sum_{k=1}^\infty \ell(I_k') + \sum_{k=1}^\infty \ell(I_k'') \\ &= \sum_{k=1}^\infty [\ell(I_k') + \ell(I_k'')] \\ &= \sum_{k=1}^\infty \ell(I_k). \end{aligned}$$

Therefore, (6) holds and the proof is complete. □

A collection of subsets of $\mathbf{R}$ is called an σ**-algebra** provided that it is an algebra and is closed with respect to countable unions; by De Morgan's Identities, such a collection is also closed with respect to countable intersections. We have shown that the collection of measurable sets, $\mathcal{M}$, is an algebra, and so it follows from Proposition 12 that it is a

σ-algebra. Since every open set is the union of a countable collection of open intervals[5], by the preceding two propositions, every open set is measurable. Every closed set, being the complement of an open set, is measurable. A set of real numbers is said to be a G_δ **set** provided that it is the intersection of a countable collection of open sets and said to be an F_σ **set** provided that it is the union of a countable collection of closed sets. Since $\mathcal{M}$ is a σ-algebra, every G_δ set and every F_σ set is measurable.

Let $\mathcal{S}$ be a collection of subsets of $\mathbf{R}$. Then $\mathcal{S}$ is contained in the σ-algebra of all subsets of $\mathbf{R}$. Define $\mathcal{A}$ to be the intersection of all σ-algebras that contain $\mathcal{S}$. Then $\mathcal{A}$ is the smallest σ-algebra that contains $\mathcal{S}$, smallest in the sense that $\mathcal{A}$ is a σ-algebra that contains $\mathcal{S}$ and is contained in any other σ-algebra that contains $\mathcal{S}$. The **Borel σ-algebra** $\mathcal{B}$ is defined to be the smallest σ-algebra that contains all open sets. Since $\mathcal{M}$ is such a σ-algebra, we have, by minimality, the inclusion $\mathcal{B} \subseteq \mathcal{M}$. According to Proposition 28, there are measurable sets that are not Borel sets. The following theorem has been proven.

Theorem 14 *The collection $\mathcal{M}$ of measurable sets is a σ-algebra that contains the Borel σ-algebra and all sets of outer-measure zero.*

PROBLEMS

12. Prove that if a σ-algebra of subsets of $\mathbf{R}$ contains intervals of the form (a, ∞), then it contains all intervals.
13. Show that (i) the translate of an F_σ set is also F_σ, (ii) the translate of a G_δ set is also G_δ, and (iii) the translate of a set of measure zero also has measure zero.
14. Show that if a set E has positive outer-measure, then there is a bounded subset of E that also has positive outer-measure.
15. Show that if $m(E) < \infty$ and $\epsilon > 0$, then E is the disjoint union of a finite number of measurable sets, each of which has measure at most ϵ.

2.4 FINER PROPERTIES OF MEASURABLE SETS

Definition *The restriction of the set-function outer-measure to the σ-algebra of measurable sets $\mathcal{M}$ is called* **Lebesgue measure** *or simply* **measure**. *It is denoted by m, so that if E is a measurable set, its Lebesgue measure, $m(E)$, is defined by*

$$m(E) = m^*(E).$$

Theorem 15 **(the Regularity of Lebesgue Measure)** *If E is Lebesgue measurable and $\epsilon > 0$, then there is a closed set F and an open set $\mathcal{O}$ for which*

$$F \subseteq E \subseteq \mathcal{O}, \;\; m(\mathcal{O} \sim E) < \epsilon \text{ and } m(E \sim F) < \epsilon. \tag{7}$$

Proof We establish the open outer-approximation in two steps. First, consider the case $m(E) < \infty$. By the definition of outer-measure, there is a countable collection of open,

[5] See page 17.

bounded intervals $\{I_k\}$ that covers E and for which $\sum_{k=1}^{\infty} \ell(I_k) < m(E) + \epsilon$. Define $\mathcal{O} = \cup_{k=1}^{\infty} I_k$. Then E is contained in the open set $\mathcal{O}$. Moreover, by the countable monotonicity of measure, $m(\mathcal{O}) \leq \sum_{k=1}^{\infty} \ell(I_k)$. Consequently, by the excision property of measure, since $m(E) < \infty$,

$$m(\mathcal{O} \sim E) = m(\mathcal{O}) - m(E) < \epsilon.$$

So the open outer-approximation holds if $m(E) < \infty$. Now consider the case $m(E) = \infty$. Fix k and define $E_k \equiv E \cap [-k, k]$. Then $m(E_k) < \infty$ and so, by what has just been established, there if an open set $\mathcal{O}_k$ for which

$$E_k \subseteq \mathcal{O}_k \text{ and } m(\mathcal{O}_k \sim E_k) < \epsilon/2^k.$$

Define $\mathcal{O} \equiv \bigcup_{k=1}^{\infty} \mathcal{O}_k$, so that $\mathcal{O}$ is open and $\mathcal{O} \sim E \subseteq \bigcup_{k=1}^{\infty} (\mathcal{O}_k \sim E_k)$. By the countable monotonicity of measure,

$$m(\mathcal{O} \sim E) \leq \sum_{k=1}^{\infty} m(\mathcal{O}_k \sim E_k) < \epsilon.$$

So the open outer-approximation property holds.

To establish the closed inner-approximation property of E, we use the open outer-approximation property of the measurable set $\mathbf{R} \sim E$. There is an open set $\mathcal{U}$ for which

$$\mathbf{R} \sim E \subseteq \mathcal{U} \text{ and } m(\mathcal{U} \sim (\mathbf{R} \sim E)) < \epsilon.$$

Define $F = \mathbf{R} \sim \mathcal{U}$. Then F is a closed subset of E and

$$E \sim F = E \sim (\mathbf{R} \sim \mathcal{U}) = E \cap U = \mathcal{U} \sim (\mathbf{R} \sim E).$$

Therefore,

$$m(E \sim F) = m(\mathcal{U} \sim (\mathbf{R} \sim E)) < \epsilon. \qquad \square$$

Corollary 16 *If E is measurable, then there is a G_δ set G and an F_σ set F for which*

$$F \subseteq E \subseteq G, \; m(G \sim E) = 0 \textit{ and } m(E \sim F) = 0. \tag{8}$$

Proof For each n, by the preceding theorem, there is a closed set F_n and an open set $\mathcal{O}_n$ for which

$$F_n \subseteq E \subseteq \mathcal{O}_n, \; m(\mathcal{O}_n \sim E) < 1/n \text{ and } m(E \sim F_n) < 1/n. \tag{9}$$

Define

$$G \equiv \bigcap_{n=1}^{\infty} \mathcal{O}_n \text{ and } F \equiv \bigcup_{n=1}^{\infty} F_n$$

Then G is a G_δ set, F is an F_σ set, and, by (9), $m(G \sim E) = m(E \sim F) = 0$. $\square$

We will frequently use the following elementary characterization of measurability.

Corollary 17 *A set of real numbers is measurable if and only if it is a G_δ set from which a set of measure zero has been excised.*

Proof Let E be measurable. According to the preceding corollary, there is a G_δ set G for which $E \subseteq G$ and $m(G \sim E) = 0$. Therefore, $E = G \sim [G \sim E]$, a G_δ set from which a set of measure zero has been excised. Since $\mathcal{M}$ is a σ-algebra, any such set is measurable. $\square$

Theorem 18 *The collection $\mathcal{M}$ of Lebesgue measurable, sets is the smallest σ-algebra that contains the Borel σ-algebra and all sets of outer-measure zero.*

Proof Let $\mathcal{S}$ be the union of the Borel σ-algebra and all sets of outer-measure zero, and let $\mathcal{A}$ be the smallest σ-algebra that contains $\mathcal{S}$. Since every open set and every set of outer-measure zero is measurable, by the minimality of $\mathcal{A}$, $\mathcal{A} \subseteq \mathcal{M}$. According to the preceding corollary, if E is a measurable set, then E is a G_δ set from which a set of measure zero has been excised and so E belongs to $\mathcal{A}$. Consequently, $\mathcal{M} \subseteq \mathcal{A}$, and so $\mathcal{M} = \mathcal{A}$. $\square$

The following property of measurable sets of finite measure asserts that such sets are, roughly speaking, "nearly" equal to the disjoint union of a finite collection of open, bounded intervals.

Theorem 19 *If $m(E) < \infty$ and $\epsilon > 0$, then there is a finite, disjoint collection of open, bounded intervals $\{I_k\}_{k=1}^n$ for which, if $\mathcal{U} \equiv \bigcup_{k=1}^n I_k$, then*[6]

$$m(\mathcal{U} \cup E) - m(\mathcal{U} \cap E) = m(\mathcal{U} \sim E) + m(E \sim \mathcal{U}) < \epsilon.$$

Proof According to Theorem 15, there is an open set $\mathcal{O}$ that contains E and

$$m(\mathcal{O} \sim E) < \epsilon/2. \tag{10}$$

Since $m(E) < \infty$, by the excision property of measure, $m(\mathcal{O}) < \infty$. Now $\mathcal{O}$, being open, is the disjoint union of a countable collection $\{I_k\}_{k=1}^\infty$ of open intervals and, by the countable additivity of measure,

$$\sum_{k=1}^{\infty} \ell(I_k) = m(\mathcal{O}) < \infty.$$

Choose an n for which

$$\sum_{k=n+1}^{\infty} \ell(I_k) < \epsilon/2.$$

Define $\mathcal{U} = \bigcup_{k=1}^n I_k$. Since $\mathcal{U} \sim E \subseteq \mathcal{O} \sim E$, it follows from (10) that

$$m(\mathcal{U} \sim E) < \epsilon/2.$$

On the other hand, since $E \subseteq \mathcal{O}$,

$$E \sim \mathcal{U} \subseteq \mathcal{O} \sim \mathcal{U} = \bigcup_{k=n+1}^{\infty} I_k,$$

[6] For two sets A and B, the **symmetric difference**, which is denoted by $A \Delta B$, is defined to be the set $[A \sim B] \cup [B \sim A]$. Using this notation, the conclusion is that $m(E \Delta \mathcal{U}) < \epsilon$.

so that

$$m(E \sim \mathcal{O}) \leq \sum_{k=n+1}^{\infty} \ell(I_k) < \epsilon/2.$$

Therefore,

$$m(\mathcal{U} \sim E) + m(E \sim \mathcal{U}) < \epsilon. \qquad \square$$

PROBLEMS

16. Show that a set E is measurable if for each $\epsilon > 0$, there is a closed set F and open set $\mathcal{O}$ for which $F \subseteq E \subseteq \mathcal{O}$ and $m(\mathcal{O} \sim F) < \epsilon$.
17. Assume that $m^*(E) < \infty$. Show that there is a G_δ set G that contains E and $m(G) = m^*(E)$. Show that E is measurable if and only if there is a G_δ set that contains G and $m^*(G \sim E) = 0$.
18. Assume that $m^*(E) < \infty$. Show that if E is not measurable, then there is an open set $\mathcal{O}$ containing E that has finite outer-measure and for which
$$m^*(\mathcal{O} \sim E) > m^*(\mathcal{O}) - m^*(E).$$
19. (Lebesgue's definition of measurability) Let E have finite outer-measure. Show that E is measurable if and only if for each open, bounded interval (a, b),
$$b - a = m^*((a, b) \cap E) + m^*((a, b) \sim E).$$
20. For any set A, define $m^{**}(A) \in [0, \infty]$ by
$$m^{**}(A) = \inf \left\{ m^*(\mathcal{O}) \;\middle|\; \mathcal{O} \supseteq A, \mathcal{O} \text{ open} \right\}.$$
How is this set-function m^{**} related to outer-measure m^*?
21. For any set A, define $m^{***}(A) \in [0, \infty]$ by
$$m^{***}(A) = \sup \left\{ m^*(F) \;\middle|\; F \subseteq A, F \text{ closed} \right\}.$$
How is this set-function m^{***} related to outer-measure m^*?

2.5 COUNTABLE ADDITIVITY AND CONTINUITY OF MEASURE, AND THE BOREL-CANTELLI LEMMA

Theorem 20 **(the Countable Additivity of Measure)** *Lebesgue measure is* **countably additive**, *in the sense that if $\{E_k\}_{k=1}^{\infty}$ is a measurable partition of E, then*

$$m(E) = \sum_{k=1}^{\infty} m(E_k).$$

Proof According to Proposition 12, $\bigcup_{k=1}^{\infty} E_k$ is measurable, and according to Proposition 5, outer-measure is countably monotone. Therefore,

$$m\left(\bigcup_{k=1}^{\infty} E_k\right) \leq \sum_{k=1}^{\infty} m(E_k). \tag{11}$$

It remains to prove this inequality in the opposite direction. By the finite additivity of measure, for each n,

$$m\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} m(E_k).$$

Since $\bigcup_{k=1}^{\infty} E_k$ contains $\bigcup_{k=1}^{n} E_k$, by the monotonicity of outer-measure and the preceding equality,

$$m\left(\bigcup_{k=1}^{\infty} E_k\right) \geq \sum_{k=1}^{n} m(E_k) \text{ for each } n.$$

The left-hand side of this inequality is independent of n. Consequently,

$$m\left(\bigcup_{k=1}^{\infty} E_k\right) \geq \sum_{k=1}^{\infty} m(E_k). \tag{12}$$

From the inequalities (11) and (12), it follows that these are equalities. □

According to Proposition 3, the outer-measure of an interval is its length while according to Proposition 4, outer-measure is translation invariant. Therefore, the preceding proposition completes the proof of the following theorem.

Theorem 21 *The set-function Lebesgue measure, defined on the σ-algebra of Lebesgue measurable sets $\mathcal{M}$, assigns length to any interval, is translation invariant, and is countably additive.*

A countable collection of sets $\{E_k\}_{k=1}^{\infty}$ is said to be **ascending** provided that for each k, $E_k \subseteq E_{k+1}$, and said to be **descending** provided that for each k, $E_{k+1} \subseteq E_k$.

Theorem 22 (**the Continuity of Measure**)

(i) If $\{A_k\}_{k=1}^{\infty}$ is an ascending collection of measurable sets, then

$$m\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k \to \infty} m(A_k). \tag{13}$$

(ii) If $\{B_k\}_{k=1}^{\infty}$ is a descending collection of measurable sets and $m(B_1) < \infty$, then

$$m\left(\bigcap_{k=1}^{\infty} B_k\right) = \lim_{k \to \infty} m(B_k). \tag{14}$$

Proof We first prove (i). If there is an index k_0 for which $m(A_{k_0}) = \infty$, then, by the monotonicity of measure, $m\left(\bigcup_{k=1}^{\infty} A_k\right) = \infty$ and $m(A_k) = \infty$ for all $k \geq k_0$. Therefore, (13) holds since each side equals ∞. It remains to consider the case that $m(A_k) < \infty$ for all k. Define $A_0 = \emptyset$ and then define $C_k = A_k \sim A_{k-1}$ for each $k \geq 1$. By construction, since the collection $\{A_k\}_{k=1}^{\infty}$ is ascending,

$$\{C_k\}_{k=1}^{\infty} \text{ is disjoint and } \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} C_k.$$

By the countable additivity of m,

$$m\left(\bigcup_{k=1}^{\infty} A_k\right) = m\left(\bigcup_{k=1}^{\infty} C_k\right) = \sum_{k=1}^{\infty} m(A_k \sim A_{k-1}). \tag{15}$$

The collection $\{A_k\}_{k=1}^{\infty}$ is ascending, and since each $m(A_k) < \infty$, it follows from the excision property of measure that

$$\begin{aligned} \sum_{k=1}^{\infty} m(A_k \sim A_{k-1}) &= \sum_{k=1}^{\infty} [m(A_k) - m(A_{k-1})] \\ &= \lim_{n \to \infty} \sum_{k=1}^{n} [m(A_k) - m(A_{k-1})] \\ &= \lim_{n \to \infty} [m(A_n) - m(A_0)]. \end{aligned} \tag{16}$$

Since $m(A_0) = m(\emptyset) = 0$, (13) follows from (15) and (16). To prove (ii), define $D_k = B_1 \sim B_k$ for each k. Since the collection $\{B_k\}_{k=1}^{\infty}$ is descending, the collection $\{D_k\}_{k=1}^{\infty}$ is ascending. By part (i),

$$m\left(\bigcup_{k=1}^{\infty} D_k\right) = \lim_{k \to \infty} m(D_k).$$

According to De Morgan's Identities,

$$\bigcup_{k=1}^{\infty} D_k = \bigcup_{k=1}^{\infty} [B_1 \sim B_k] = B_1 \sim \bigcap_{k=1}^{\infty} B_k.$$

On the other hand, by the excision property of measure, for each k, since $m(B_k) < \infty$, $m(D_k) = m(B_1) - m(B_k)$. Therefore,

$$m\left(B_1 \sim \bigcap_{k=1}^{\infty} B_k\right) = \lim_{n \to \infty} [m(B_1) - m(B_n)].$$

Once more using the excision property of measure, we obtain (14). □

The Borel-Cantelli Lemma *If $\{E_k\}_{k=1}^{\infty}$ is a countable collection of measurable sets for which $\sum_{k=1}^{\infty} m(E_k) < \infty$, then almost all $x \in \mathbf{R}$ belong to at most finitely many of the E_k's.*

Proof For each n, by the countable monotonicity of m,

$$m\left(\bigcup_{k=n}^{\infty} E_k\right) \le \sum_{k=n}^{\infty} m(E_k) < \infty.$$

Consequently, by the continuity of measure,

$$m\left(\bigcap_{n=1}^{\infty} \left[\bigcup_{k=n}^{\infty} E_k\right]\right) = \lim_{n \to \infty} m\left(\bigcup_{k=n}^{\infty} E_k\right) \le \lim_{n \to \infty} \sum_{k=n}^{\infty} m(E_k) = 0.$$

Therefore, almost all $x \in \mathbf{R}$ fail to belong to $\bigcap_{n=1}^{\infty} \left[\bigcup_{k=n}^{\infty} E_k\right]$, which means that they belong to at most finitely many E_k's. □

PROBLEMS

22. Show that if E_1 and E_2 are measurable sets of finite measure, then
$$m(E_1 \cup E_2) - m(E_1 \cap E_2) = m(E_1 \sim E_2) + m(E_2 \sim E_1).$$
23. Show that the assumption that $m(B_1) < \infty$ is necessary in part (ii) of the theorem regarding continuity of measure.
24. Let $\{E_k\}_{k=1}^{\infty}$ be a countable, disjoint collection of measurable sets. Prove that for any set A,
$$m^*\left(A \cap \bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} m^*(A \cap E_k).$$
25. Let $\mathcal{M}'$ be any σ-algebra of subsets of $\mathbf{R}$ and m' a set-function on $\mathcal{M}'$ that takes values in $[0, \infty)$, is countably additive, and such that $m'(\emptyset) = 0$.
    (i) Show that m' is finitely additive, monotone, countably monotone, and possesses the excision property.
    (ii) Show that m' possesses the same continuity properties as Lebesgue measure.
26. Show that the continuity of measure together with the finite additivity of measure are equivalent to the countable additivity of measure.
27. (Arzelà) Let $\{E_k\}_{k=1}^{\infty}$ be a collection of measurable subsets of $[a, b]$, and assume there is a $\delta > 0$ such that $m(E_k) \geq \delta$ for each k. Show that there is an $x \in [a, b]$ that belongs to infinitely many such E_k's.

2.6 VITALI'S EXAMPLE OF A NON-MEASURABLE SET

We have considered properties of individual measurable sets and of the collection of measurable sets. It is only natural to ask if, in fact, there are any sets that fail to be measurable. The answer is not at all obvious. If a set E has outer-measure zero, then it is measurable, and since any subset also has outer-measure zero, every subset of E is measurable. This is the best that can be said regarding the inheritance of measurability through the relation of set inclusion. We now show that if E has positive outer-measure, then there are subsets of E that fail to be measurable.

For any non-empty set E of real numbers, define two points in E to be **rationally equivalent** provided that their difference is a rational number. Clearly, this is an equivalence relation, that is, it is reflexive, symmetric, and transitive. We call it the rational equivalence relation on E. For this relation, there is the disjoint decomposition of E into the collection of equivalence classes. A **choice set** for the rational equivalence relation on E is a set $\mathcal{C}_E$ comprising exactly one member of each equivalence class. It follows from the Axiom of Choice[7] that there is such a set.

Theorem 23 (**Vitali**) *If E is a set of real numbers for which $m^*(E) > 0$, then there is a subset of E that is not measurable.*

Proof In view of the countable monotonicity of m^*, by possibly replacing E by a bounded subset of positive outer-measure, we assume that E is bounded. Choose $C_E \subseteq E$ to be a choice set for the rational equivalence relation on E, that is,

$$\text{the countable collection } \{C_E + q\}_{q \in \mathbf{Q}} \text{ is disjoint and } E \subseteq \bigcup_{q \in \mathbf{Q}} [C_E + q].$$

[7]See page 5.

We claim that C_E is not measurable, and to verify this, we assume otherwise and obtain a contradiction. Choose $\mathbf{Q}_0$ to be a countably infinite, bounded set of rational numbers. The set $\bigcup_{q \in \mathbf{Q}_0} [C_E + q]$ is bounded and measurable. Consequently, by the countable additivity of measure, since $\{C_E + q\}_{q \in \mathbf{Q}_0}$ is a countable, disjoint collection of measurable sets,

$$m\Big(\bigcup_{q \in \mathbf{Q}_0} [C_E + q] = \sum_{q \in \mathbf{Q}_0} m(C_E + q) < \infty.$$

However, by the translation invariance of measure, $m(C_E + q) = m(C_E)$, for every $q \in \mathbf{Q}_0$, and therefore, since $\mathbf{Q}_0$ is countably infinite, $m(C_E) = 0$. By the countable monotonicity of m^*, the countability of the rationals and the inclusion $E \subseteq \bigcup_{q \in \mathbf{Q}} [C_E + q]$, we obtain the following contradiction:

$$0 < m^*(E) \le \sum_{q \in \mathbf{Q}} m^*(C_E + q) = \sum_{q \in \mathbf{Q}} m(C_E + q) = 0.$$

$\square$

Corollary 24 *There are disjoint sets of real numbers A and B for which*

$$m^*(A \cup B) < m^*(A) + m^*(B).$$

Proof If there were not two such sets, then, by the very definition of measurable set, every set of real numbers would be measurable, which contradicts the preceding theorem. $\square$

PROBLEMS

28. (i) Show that rational equivalence defines an equivalence relation on any set.
    (ii) Explicitly find a choice set for the rational equivalence relation on $\mathbf{Q}$, and also on $\mathbf{R} \sim \mathbf{Q}$.
    (iii) Define two numbers to be irrationally equivalent provided that their difference is irrational or zero. Is this an equivalence relation on $\mathbf{R}$? Is this an equivalence relation on $\mathbf{Q}$?
29. Show that any choice set for the rational equivalence relation on a set of positive outer-measure must be uncountably infinite.

2.7 THE CANTOR SET AND THE CANTOR-LEBESGUE FUNCTION

In this section, we construct a set called the Cantor set and a function called the Cantor-Lebesgue function. These provide two interesting examples: a continuous, increasing function, which has a derivative that vanishes almost everywhere, and yet the function is not constant, and a measurable set that is not a Borel set.

Consider the closed, bounded interval $I = [0, 1]$. The first step in the construction of the Cantor set is to subdivide I into three intervals of equal length $1/3$ and remove the interior of the middle interval, that is, remove the interval $(1/3, 2/3)$ from the interval $[0, 1]$ to obtain the closed set C_1, which is the disjoint union of two disjoint, closed intervals, each of length $1/3$:

$$C_1 = [0, 1/3] \cup [2/3, 1].$$

Repeat this "open middle one-third removal" on each of the two intervals in C_1 to obtain a closed set C_2, which is the disjoint union of 2^2 closed intervals, each of length $1/3^2$:

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1].$$

Now repeat this "open middle one-third removal" on each of the four intervals in C_2 to obtain a closed set C_3, which is the disjoint union of 2^3 closed intervals, each of length $1/3^3$. Continue this removal operation countably many times to obtain a countable collection of sets $\{C_k\}_{k=1}^{\infty}$. Finally, define the Cantor set $\mathbf{C}$ by

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

The collection $\{C_k\}_{k=1}^{\infty}$ possesses the following two properties:

(i) $\{C_k\}_{k=1}^{\infty}$ is a countable, descending collection of closed subsets of $[0, 1]$, and
(ii) for each k, C_k is the disjoint union of 2^k closed intervals, each of length $1/3^k$.

Proposition 25 *The Cantor set* $\mathbf{C}$ *is a closed, uncountable set of measure zero.*

Proof The intersection of any collection of closed sets is closed. Therefore, $\mathbf{C}$ is closed. Each closed set is measurable, so that each C_k and $\mathbf{C}$ itself is measurable. Now each C_k is the disjoint union of 2^k intervals, each of length $1/3^k$, so that by the finite additivity of measure,

$$m(C_k) = (2/3)^k.$$

By the monotonicity of measure, since $m(\mathbf{C}) \leq m(C_k) = (2/3)^k$, for all k, $m(\mathbf{C}) = 0$. It remains to show that $\mathbf{C}$ is uncountable. To do so, suppose otherwise, and let $\{c_k\}_{k=1}^{\infty}$ be an enumeration of $\mathbf{C}$. One of the two disjoint Cantor intervals whose union is C_1 fails to contain the point c_1; denote it by F_1. One of the two disjoint Cantor intervals in C_2 whose union is F_1 fails to contain the point c_2; denote it by F_2. Continuing in this way, construct a countable collection of sets $\{F_k\}_{k=1}^{\infty}$, which, for each index k, possesses the following three properties: (i) F_k is closed and $F_{k+1} \subseteq F_k$; (ii) $F_k \subseteq C_k$; and (iii) $c_k \notin F_k$. From (i) and the Nested Set Theorem[8], it follows that the intersection $\bigcap_{k=1}^{\infty} F_k$ is non-empty. Let $x \in \bigcap_{k=1}^{\infty} F_k$. By property (ii),

$$\bigcap_{k=1}^{\infty} F_k \subseteq \bigcap_{k=1}^{\infty} C_k = \mathbf{C},$$

and therefore $x \in \mathbf{C}$. However, $\{c_k\}_{k=1}^{\infty}$ is an enumeration of $\mathbf{C}$, so that $x = c_n$ for some index n. Consequently, $c_n = x \in \bigcap_{k=1}^{\infty} F_k \subseteq F_n$. This contradicts the choice of c_n. Therefore, $\mathbf{C}$ is uncountable. □

We now define the Cantor-Lebesgue function, which is a continuous, increasing function $\varphi\colon [0, 1] \to \mathbf{R}$ with the remarkable property that, despite the fact that $\varphi(1) > \varphi(0)$, its derivative exists and is zero on a set of measure 1. For each k, let $\mathcal{O}_k$ be the union of the $2^k - 1$ intervals which have been removed during the first k stages of the Cantor deletion process. Therefore, $C_k = [0, 1] \sim \mathcal{O}_k$. Define $\mathcal{O} = \bigcup_{k=1}^{\infty} \mathcal{O}_k$. Then, by De Morgan's Identities, $\mathbf{C} = [0, 1] \sim \mathcal{O}$. We begin by defining φ on $\mathcal{O}$, and then define it on $\mathbf{C}$.

Fix k. Define φ on $\mathcal{O}_k$ to be the increasing function on $\mathcal{O}_k$ that is constant on each of its $2^k - 1$ open intervals and takes the $2^k - 1$ values

$$\{1/2^k, 2/2^k, 3/2^k, \ldots, [2^k - 1]/2^k\}.$$

[8]See page 18

On the three intervals that are removed in the first two stages, the prescription for φ is

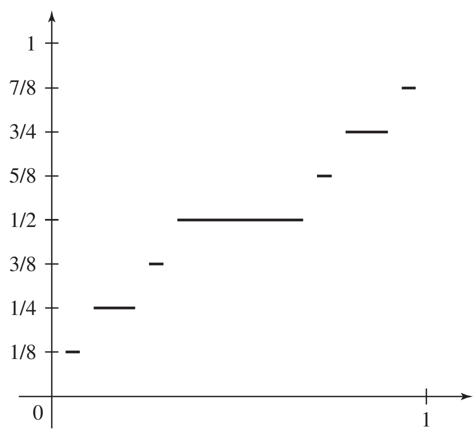
$$\varphi(x) = \begin{cases} 1/4 & \text{if } x \in (1/9,\, 2/9) \\ 2/4 & \text{if } x \in (3/9,\, 6/9) = (1/3, 2/3) \\ 3/4 & \text{if } x \in (7/9,\, 8/9) \end{cases}$$

Extend φ to all of $[0,\, 1]$ by defining it on $\mathbf{C}$ as follows:

$$\varphi(0) = 0 \text{ and } \varphi(x) = \sup\{\varphi(t) \mid t \in \mathcal{O} \cap [0,\, x)\} \text{ if } x \in \mathbf{C} \sim \{0\}.$$

Proposition 26 *The Cantor-Lebesgue function $\varphi\colon [0,\, 1] \to \mathbf{R}$ is an increasing, continuous function that maps $[0,\, 1]$ onto $[0,\, 1]$. Its derivative exists on the open set $\mathcal{O}$, the complement in $[0,\, 1]$ of the Cantor set,*

$$\varphi' = 0 \text{ on } \mathcal{O}, \text{ yet } m(\mathcal{O}) = 1 \text{ and } \varphi(1) > \varphi(0).$$



The graph of the Cantor-Lebesgue function on $\mathcal{O}_3 = [0,\, 1] \sim C_3$

Proof Since φ is increasing on $\mathcal{O}$, its extension to $[0, 1]$ also is increasing. As for continuity, φ certainly is continuous at each point in $\mathcal{O}$, since each such point belongs to an open interval on which it is constant. Now consider a point $x_0 \in \mathbf{C}$ with $x_0 \neq 0, 1$. Since $x_0 \in \mathbf{C}$, it is not a member of the $2^k - 1$ intervals removed in the first k stages of the removal process, the union of which is denoted by $\mathcal{O}_k$. Therefore, x_0 lies between two consecutive intervals in $\mathcal{O}_k$: choose a_k in the lower of these and b_k in the upper one. The function φ was defined to increase by $1/2^k$ across two consecutive intervals in $\mathcal{O}_k$. Therefore, for each k,

$$a_k < x_0 < b_k \text{ and } \varphi(b_k) - \varphi(a_k) = 1/2^k.$$

So the function φ fails to have a jump discontinuity at x_0. For an increasing function, a jump discontinuity is the only possible type of discontinuity. Therefore, φ is continuous at x_0. If x_0 is an end-point of $[0,\, 1]$, a similar argument establishes continuity at x_0.

Since φ is constant on each of the intervals removed at any stage of the removal process, its derivative exists and equals 0 at each point in $\mathcal{O}$. Since $\mathbf{C}$ has measure zero, its complement in $[0, 1]$, $\mathcal{O}$, has measure 1. Finally, since $\varphi(0) = 0, \varphi(1) = 1$ and φ is increasing and continuous, it follows from the Intermediate Value Theorem that φ maps $[0, 1]$ onto $[0, 1]$. $\square$

Proposition 27 *Let $\varphi\colon [0, 1] \to \mathbf{R}$ be the Cantor-Lebesgue function and define the function $\psi\colon [0, 1] \to \mathbf{R}$ by*

$$\psi(x) = \varphi(x) + x \text{ for all } x \in [0, 1].$$

Then ψ is a strictly increasing, continuous function that maps $[0, 1]$ onto $[0, 2]$,

(i) maps the Cantor set onto a measurable set of positive measure and

(ii) maps a measurable set, a subset of the Cantor set, onto a non-measurable set.

Proof The function ψ is continuous, since it is the sum of two continuous functions and is strictly increasing since it is the sum of an increasing and a strictly increasing function. Moreover, since $\psi(0) = 0$ and $\psi(1) = 2$, by the Intermediate Value Theorem, $\psi([0, 1]) = [0, 2]$. For $\mathcal{O} = [0, 1] \sim \mathbf{C}$, there is the disjoint decomposition of the domain

$$[0, 1] = \mathbf{C} \cup \mathcal{O},$$

which, since ψ is one-to-one, lifts to the disjoint decomposition of its image,

$$[0, 2] = \psi(\mathcal{O}) \cup \psi(\mathbf{C}). \tag{17}$$

A strictly increasing, continuous function defined on an interval has a continuous inverse. Therefore, $\psi(\mathbf{C})$ is closed and $\psi(\mathcal{O})$ is open, so both are measurable. We will show that $m(\psi(\mathcal{O})) = 1$ from which it will follow from (17) that $m(\psi(\mathbf{C})) = 1$ and thereby prove (i). Let $\{I_k\}_{k=1}^{\infty}$ be an enumeration (in any manner) of the collection of intervals that are removed in the Cantor removal process. We have $\mathcal{O} = \bigcup_{k=1}^{\infty} I_k$. Since φ is constant on each I_k, ψ maps I_k onto a translated copy of itself of the same length. Since ψ is one-to-one, the collection $\{\psi(I_k)\}_{k=1}^{\infty}$ is disjoint. By the countable additivity of measure,

$$m(\psi(\mathcal{O})) = \sum_{k=1}^{\infty} \ell(\psi(I_k)) = \sum_{k=1}^{\infty} \ell(I_k) = m(\mathcal{O}).$$

Therefore, $m(\psi(\mathcal{O})) = 1$ and so, by (17), $m(\psi(\mathbf{C})) = 1$. We have established (i). To verify (ii), observe that, according to Theorem 23, $\psi(\mathbf{C})$ contains a set W that is not measurable. The set $\psi^{-1}(W)$ is measurable and has measure zero, since it is a subset of the Cantor set. The set $\psi^{-1}(W)$, a subset C, is mapped by ψ to a non-measurable set. $\square$

Proposition 28 *There is a measurable set, a subset of the Cantor set, that is not a Borel set.*

Proof The function $\psi\colon [0, 1] \to \mathbf{R}$ defined in the preceding proposition maps a measurable set A onto a non-measurable set. A strictly increasing, continuous function defined on an interval maps Borel sets onto Borel sets, since ψ^{-1} is continuous. Therefore, the set A is not a Borel set, since otherwise its image under ψ would be a Borel set and therefore would be measurable. $\square$

PROBLEMS

30. Let f: $[a, b] \to \mathbf{R}$ be continuous and strictly increasing. Show that f^{-1}: $[f(a), f(b)] \to \mathbf{R}$ is continuous.

31. Let f: $[a, b] \to \mathbf{R}$ be continuous. Show that f maps closed sets to closed sets and maps F_σ sets to F_σ sets. Does f map measurable sets to measurable sets?

32. Let the function f: $[a, b] \to \mathbf{R}$ be Lipschitz, that is, there is a constant $c \geq 0$ such that for all $u, v \in [a, b]$, $|f(u) - f(v)| \leq c|u - v|$. Show that f is continuous and maps sets of measure zero to sets of measure zero. Conclude that f maps measurable set to measurable sets. (Suggestion: consider the preceding problem and the regularity of measure.)

33. Let F be the subset of $[0, 1]$ constructed in the same manner as the Cantor set except that each of the intervals removed at the nth deletion stage has length $\alpha 3^{-n}$ with $0 < \alpha < 1$. Show that F is a closed set, $[0, 1] \sim F$ is dense in $[0, 1]$, and $m(F) = 1 - \alpha$. Such a set F is called a generalized Cantor set.

34. Show that there is an open set of real numbers which, contrary to intuition, has a boundary of positive measure. (Suggestion: Consider the complement of the generalized Cantor set in the preceding problem.)

35. A subset A of $\mathbf{R}$ is said to be **nowhere dense** in $\mathbf{R}$ provided that the closure of every open set $\mathcal{O}$ has a non-empty open subset that is disjoint from A. Show that the Cantor set is nowhere dense in $[0, 1]$.

36. Let f: $[a, b] \to \mathbf{R}$ be continuous and B be a Borel set. Show that $f^{-1}(B)$ is a Borel set.

CHAPTER 3

Lebesgue Measurable Functions

Contents

3.1 Sums, Products, and Compositions 51
3.2 Sequential Pointwise Limits and Simple Approximation 55
3.3 Littlewood's Three Principles, Egoroff's Theorem, and Lusin's Theorem . . 59

3.1 SUMS, PRODUCTS, AND COMPOSITIONS

We continue to denote by $\mathcal{M}$ the σ-algebra of Lebesgue measurable sets and by m the set-function Lebesgue measure. In Part 11, we consider a general concept of measure. However, in the first eight chapters, we only consider Lebesgue measure for subsets of $\mathbf{R}$, and so, in these, we often drop the adjective Lebesgue. Recall that a property is said to hold **almost everywhere** on a measurable set E provided that it holds on $E_0 \subseteq E$, where $m(E \sim E_0) = 0$. The extended real numbers is the set $\mathbf{R} \cup \{\pm\infty\}$ which we denote by $\overline{\mathbf{R}}$. Given two functions h: $E \to \overline{\mathbf{R}}$ and g: $E \to \overline{\mathbf{R}}$, we write "$h \leq g$ on E" to mean that $h(x) \leq g(x)$ for all $x \in E$. We say that a sequence of functions $\{f_n: E \to \overline{\mathbf{R}}\}$ is increasing provided that $f_n \leq f_{n+1}$ on E for each index n.

Many arguments will depend on inverse images of functions. For any mapping f: $X \to Y$, if $\{Y_\lambda\}_{\lambda \in \Lambda}$ is a collection of subsets of Y, parametrized by a space Λ, then

$$f^{-1}\left(\bigcup_{\lambda \in \Lambda} Y_\lambda\right) = \bigcup_{\lambda \in \Lambda} f^{-1}(Y_\lambda) \text{ and } f^{-1}\left(\bigcap_{\lambda \in \Lambda} Y_\lambda\right) = \bigcap_{\lambda \in \Lambda} f^{-1}(Y_\lambda).$$

Definition *A function f: $E \to \overline{\mathbf{R}}$ is said to be* **Lebesgue measurable**, *or simply* **measurable**, *provided that its domain E is a measurable subset of $\mathbf{R}$ and for each real number c, the set $\{x \in E \mid f(x) < c\}$ is measurable.*

Proposition 1 *If f: $E \to \overline{\mathbf{R}}$ is measurable, then for every interval I of real numbers, $f^{-1}(I)$ is measurable.*

Proof Every bounded interval is the intersection of two unbounded intervals, and so it suffices to show that $f^{-1}(I)$ is measurable if I is unbounded. Fix a number c. We have

$$f^{-1}[c, \infty) = X \sim f^{-1}(-\infty, c) \text{ and } f^{-1}(c, \infty) = \bigcup_{1 \leq k < \infty} f^{-1}[c + 1/k, \infty).$$

Now, $\mathcal{M}$ is a σ-algebra, Therefore, $f^{-1}[c, \infty)$, being the complement of a measurable set, is measurable, and $f^{-1}(c, \infty)$, being the countable union of measurable sets, is measurable, as is the following countable intersection of measurable sets:

$$f^{-1}(-\infty, c] = \bigcap_{1 \leq k < \infty} f^{-1}(-\infty, c + 1/k). \qquad \square$$

Proposition 2 *A function $f: E \to \mathbf{R}$ is measurable if and only if for each open set $\mathcal{O}$, the inverse image of $\mathcal{O}$ under f, $f^{-1}(\mathcal{O})$, is a measurable set.*

Proof If the inverse image of each open set is measurable, then $E = f^{-1}(\mathbf{R})$ is measurable, and since each interval $(-\infty, c)$ is open, the function f is measurable. Conversely, suppose that f is measurable. Let $\mathcal{O}$ be open. According to Proposition 9 of Chapter 1, $\mathcal{O}$ is the union of a countable collection of intervals. Therefore, the measurable sets being a σ-algebra, it follows from the preceding proposition that $f^{-1}(\mathcal{O})$ is measurable. $\square$

The following proposition follows from its predecessor by recalling that, for a continuous function $f: E \to \mathbf{R}$ on a measurable set E, if $\mathcal{O}$ is open, then $f^{-1}(\mathcal{O}) = E \cap \mathcal{U}$, where U is an open subset of $\mathbf{R}$, and so $f^{-1}(\mathcal{O})$ is measurable.

Proposition 3 *If E is measurable, then every continuous function $f: E \to \mathbf{R}$ is measurable.*

Proposition 4 *If $E = A \cup B$, where A and B are measurable, then $f: E \to \overline{\mathbf{R}}$ is measurable if and only if its restrictions to A and B are measurable. In particular, if $E_0 \subseteq E$ and $m(E \sim E_0) = 0$, then*

$$f: E \to \overline{\mathbf{R}} \text{ is measurable if and only } f: E_0 \to \overline{\mathbf{R}} \text{ is measurable.} \tag{1}$$

Proof Let $f: E \to \overline{\mathbf{R}}$ be measurable. For each $c \in \mathbf{R}$,

$$\{x \in A \mid f(x) < c\} = \{x \in E \mid f(x) < c\} \cap A,$$

so that the restriction of f to A is measurable. Similarly, the restriction of f to B is measurable. On the other hand, if both restrictions are measurable, then for each $c \in \mathbf{R}$,

$$\{x \in E \mid f(x) < c\} = \{x \in A \mid f(x) < c\} \cup \{x \in B \mid f(x) < c\},$$

so that $f: E \to \overline{\mathbf{R}}$ is measurable. If $m(E_0) = 0$, then every subset of E_0 is measurable, and therefore every function on E_0 is measurable, and so (1) holds. $\square$

It follows from these two propositions that, for a real-valued function defined on a measurable set, if the set of points at which it fails to be continuous has measure zero, then the function is measurable. According to the first theorem of the following chapter, a monotone function on an interval is continuous except for at most a countable collection of points, and so is measurable.

Theorem 5 *If $f: E \to \mathbf{R}$ and $g: E \to \mathbf{R}$ are measurable functions, then for any α and β,*

(Linearity)

$$\alpha f + \beta g: E \to \mathbf{R} \text{ is measurable.}$$

(Products)

$$f \cdot g: E \to \mathbf{R} \text{ is measurable.}$$

Proof We consider the case $\alpha = \beta = 1$, and leave the case of general coefficients as an exercise. For $x \in E$, if $f(x) + g(x) < c$, then $f(x) < c - g(x)$ and so, by the density of the set of rational numbers $\mathbf{Q}$ in $\mathbf{R}$, there is a rational number q for which $f(x) < q < c - g(x)$. Consequently,

$$\{x \in E \mid f(x) + g(x) < c\} = \bigcup_{q \in \mathbf{Q}} \{x \in E \mid g(x) < c - q\} \cap \{x \in E \mid f(x) < q\}.$$

However, the rational numbers are countable, and so $\{x \in E \mid f(x) + g(x) < c\}$ is measurable, since $\mathcal{M}$ is a σ-algebra. Therefore, $f + g$ is measurable. To prove that the product of measurable functions is measurable, observe that

$$fg = \frac{1}{2}[(f+g)^2 - f^2 - g^2].$$

Since we have established linearity, to show that the product of two measurable functions is measurable, it suffices to show that the square of a measurable function is measurable. For $c > 0$,

$$\{x \in E \mid f^2(x) < c\} = \{x \in E \mid -\sqrt{c} < f(x) < \sqrt{c}\},$$

while for $c \leq 0$, $\{x \in E \mid f^2(x) < c\} = \emptyset$. Therefore, f^2 is measurable. □

We identify two measurable functions with a common domain that agree almost everywhere. In particular, in view of Proposition 4, we identify a measurable function that is finite almost everywhere with a real-valued measurable function. This identification permits the extension of measurability results for real-valued functions to hold for extended real-valued functions that are finite almost almost everywhere, regardless of the possibility that functional values of sums and products may not be properly defined at points at which one of the functions takes an infinite value. In particular, the above theorem extends to measurable functions that are finite almost everywhere.

If A is any set, the **characteristic function** of A, χ_A, is the function on $\mathbf{R}$ defined by

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

It is clear that the function χ_A is measurable if and only if the set A is measurable. Consequently, since there are non-measurable sets, there are non-measurable functions.

Many of the properties of functions considered in calculus, such as continuity and differentiability, are preserved under the operation of composition of functions. As the next example shows, the composition of measurable functions may not be measurable.

Example There are two measurable real-valued functions, each defined on all of $\mathbf{R}$, for which the composition fails to be measurable. Indeed, by Proposition 27 of the preceding chapter, there is a continuous, strictly increasing function $\psi \colon [0, 1] \to \mathbf{R}$ and a measurable subset A of $[0, 1]$ for which $\psi(A)$ is non-measurable. Extend ψ to a continuous, strictly increasing function that maps $\mathbf{R}$ onto $\mathbf{R}$. The function ψ^{-1} is continuous and therefore is measurable. On the other hand, A is a measurable set and so its characteristic function χ_A is a measurable function. However, the composition

$$f \equiv \chi_A \circ \psi^{-1} = \chi_{\psi(A)}$$

is not measurable, since the set $\psi(A)$ is not measurable.

There is the following useful result regarding the preservation of measurability under composition (also see Problem 12)[1].

Proposition 6 *If $f\colon E \to \mathbf{R}$ is a measurable function and $g\colon \mathbf{R} \to \mathbf{R}$ is continuous, then the composition $g \circ f\colon \mathbf{R} \to \mathbf{R}$ is measurable.*

Proof According to Proposition 2, a real-valued function is measurable if and only if the inverse image of each open set is measurable. Let $\mathcal{O}$ be open. Then

$$(g \circ f)^{-1}(\mathcal{O}) = f^{-1}(g^{-1}(\mathcal{O})).$$

Since g is continuous, the set $\mathcal{U} = g^{-1}(\mathcal{O})$ is an open subset of $\mathbf{R}$. By the measurability of f, $f^{-1}(\mathcal{U})$ is measurable. Therefore, $(g \circ f)^{-1}(\mathcal{O})$ is measurable, and so $g \circ f$ is measurable. $\square$

An immediate consequence of the above composition result is that if $f\colon E \to \mathbf{R}$ is measurable, then $|f|$ also is measurable, and indeed

$$|f|^p\colon E \to \mathbf{R} \text{ is measurable for each } p > 0.$$

For a finite collection of functions $\{f_k\colon E \to \mathbf{R}\}_{k=1}^n$, define

$$\max\{f_1, \ldots, f_n\}(x) \equiv \max\{f_1(x), \ldots, f_n(x)\} \text{ for } x \in E.$$

The function $\min\{f_1, \ldots, f_n\}$ is defined the same way. We leave the proof of the following result as an exercise.

Proposition 7 *For a finite collection $\{f_k\colon E \to \mathbf{R}\}_{k=1}^n$ of measurable functions, the functions $\max\{f_1, \ldots, f_n\}$ and $\min\{f_1, \ldots, f_n\}$ also are measurable.*

For a measurable function $f\colon E \to \mathbf{R}$, there are associated non-negative measurable functions f^+ and f^-, called the positive part and the negative part of f, defined on E by

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

These provide the following frequently used expressions of f and $|f|$ as the difference and sum, respectively, of non-negative measurable functions:

$$f = f^+ - f^- \text{ and } |f| = f^+ + f^- \text{ on } E.$$

[1]In general, it is important, but not always easy, to determine if the composition of measurable functions is measurable. In Chapter 6, we define what it means for a function $f\colon [a, b] \to R$ to be absolutely continuous; Lipschitz functions, for example, are absolutely continuous. We prove a theorem of von Neumann which asserts that if $f\colon [a, b] \to R$ is a strictly increasing, continuous function, then $g \circ f\colon [a, b] \to R$ is measurable whenever $g\colon R \to R$ is measurable if and only if its inverse $f^{-1}\colon [f(a), f(b)] \to R$ is absolutely continuous.

PROBLEMS

1. Let A and B be measurable sets. Is $f\colon A \cup B \to \mathbf{R}$ continuous if and only if its restrictions to A and B are continuous?
2. Show that if $h\colon E \to \mathbf{R}$ is measurable, then for any α, so is $\alpha \cdot h\colon E \to \mathbf{R}$.
3. Provide an example of a function $f\colon [a, b] \to \mathbf{R}$ that is not measurable, while both $|f|$ and f^2 are measurable.
4. Let $\{f_n\colon E \to \mathbf{R}\}$ be a sequence of measurable functions. Define E_0 to be the set of points x in E at which $\{f_n(x)\}$ converges to a real number. Is the set E_0 measurable?
5. Let E be measurable and $f\colon E \to \mathbf{R}$ be continuous except at a countable number of points. Show that f is measurable.
6. If E is measurable and the function $f\colon E \to \mathbf{R}$ has the property that $f^{-1}(c)$ is measurable for each number c, is f necessarily measurable?
7. If the function $f\colon E \to \mathbf{R}$ has the property that $\{x \in E \mid f(x) > c\}$ is a measurable set for each rational number c, is f necessarily measurable?
8. Show that $f\colon E \to \mathbf{R}$ is measurable if and only if the function $\hat{f}\colon \mathbf{R} \to \mathbf{R}$ is measurable, where $\hat{f}(x) = f(x)$ for $x \in E$ and $\hat{f}(x) = 0$ for $x \notin E$.
9. Show that $f\colon E \to \mathbf{R}$ is measurable if and only if for each Borel set A, $f^{-1}(A)$ is measurable.
10. (Borel measurability) A function $f\colon E \to \mathbf{R}$ is said to be **Borel measurable** provided that its domain E is a Borel set and for each c, the set $\{x \in E \mid f(x) < c\}$ is a Borel set. Verify that Proposition 1 and Theorem 5 remain valid if we replace "(Lebesgue) measurable set" by "'Borel set." Verify the following: (i) every Borel measurable function is Lebesgue measurable; (ii) if f is Borel measurable and B is a Borel set, then $f^{-1}(B)$ is a Borel set; (iii) if f and g are Borel measurable, so is $f \circ g$; and (iv) if f is Borel measurable and g is Lebesgue measurable, then $f \circ g$ is Lebesgue measurable.
11. Assume that $f\colon \mathbf{R} \to \mathbf{R}$ is measurable and $g\colon \mathbf{R} \to \mathbf{R}$ is continuous. Is the composition $f \circ g$ necessarily measurable?
12. Suppose that $f\colon [a, b] \to R$ is strictly increasing, continuous function that has a Lipschitz inverse. Show that $g \circ f\colon [a, b] \to R$ is measurable whenever $g\colon R \to R$ is measurable. (Suggestion: Show that a Lipschitz function maps measurable sets to measurable sets, by examining images of F_σ sets and sets of measure zero.)

3.2 SEQUENTIAL POINTWISE LIMITS AND SIMPLE APPROXIMATION

For a sequence of functions $\{f_n\colon E \to \overline{\mathbf{R}}\}$ and function $f\colon E \to \overline{\mathbf{R}}$, there are several ways in which it is necessary to consider what it means to state that

"the sequence $\{f_n\}$ converges to f."

In this chapter, we consider the concept of pointwise convergence. In later chapters, we consider many other modes of convergence.

Definition *A sequence of functions $\{f_n\colon E \to \overline{\mathbf{R}}\}$ is said to converge pointwise to the function $f\colon E \to \overline{\mathbf{R}}$ provided that*

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

The pointwise limit of continuous functions may not be continuous. The pointwise limit of Riemann integrable functions may not be Riemann integrable. However, for measurable functions, we have the following indispensable theorem.

Theorem 8 *If $\{f_n\colon E \to \overline{\mathbf{R}}\}$ is a sequence of measurable functions that converges pointwise almost everywhere to the function $f\colon E \to \overline{\mathbf{R}}$, then f is measurable.*

Proof In view of Proposition 4, we assume the convergence is on all of E. Fix a number c. It must be shown that $\{x \in E \mid f(x) < c\}$ is measurable. Observe that for a point $x \in E$, since $\lim_{n\to\infty} f_n(x) = f(x)$, $f(x) < c$ if and only if there are indices n and k for which

$$f_j(x) < c - 1/n \text{ for all } j \geq k.$$

But for any n and j, $\{x \in E \mid f_j(x) < c - 1/n\}$ is a measurable set, since f_j is a measurable function. Therefore, for any k, since $\mathcal{M}$ is a σ-algebra, the set

$$\bigcap_{j=k}^{\infty} \{x \in E \mid \ f_j(x) < c - 1/n\}$$

is measurable, as is the set

$$\{x \in E \mid \ f(x) < c\} = \bigcup_{1 \leq k, n < \infty} \left[\bigcap_{j=k}^{\infty} \{x \in E \mid \ f_j(x) < c - 1/n\} \right].$$

□

Linear combinations of characteristic functions of measurable sets play a role in Lebesgue integration that is foreshadowed by the role of step-functions in Riemann integration, and so we name these functions.

Definition *A real-valued function $\varphi\colon E \to \mathbf{R}$ is said to be* ***simple*** *provided that it is measurable and takes only a finite number of values.*

Clearly, linear combinations and products of characteristic functions are simple. If $\varphi\colon E \to \mathbf{R}$ is simple and takes the distinct values $c_1, \ldots, c_n$, then

$$\varphi = \sum_{k=1}^{n} c_k \cdot \chi_{E_k} \text{ on } E, \text{ where } E_k = \{x \in E \mid \ \varphi(x) = c_k\}.$$

This particular expression of φ as a linear combination of characteristic functions of measurable sets is called the **canonical representation of the simple function** φ. It is characterized by the E_k's being disjoint and the c_k's being distinct.

The Simple Approximation Lemma *If $f\colon E \to \mathbf{R}$ is a measurable, bounded function, then for each $\epsilon > 0$, there are simple functions $\varphi_\epsilon\colon E \to \mathbf{R}$ and $\psi_\epsilon\colon E \to \mathbf{R}$ for which*

$$\varphi_\epsilon \leq f \leq \psi_\epsilon \text{ and } 0 \leq \psi_\epsilon - \varphi_\epsilon < \epsilon \text{ on } E.$$

Proof Let $[c, d)$ be a bounded interval that contains the image of E, $f(E)$, and

$$c = y_0 < y_1 < \ldots < y_{n-1} < y_n = d$$

be a partition of the closed interval $[c, d]$ such that $y_k - y_{k-1} < \epsilon$ for $1 \leq k \leq n$. Define

$$I_k = [y_{k-1}, y_k) \text{ and } E_k = f^{-1}(I_k) \text{ for } 1 \leq k \leq n.$$

Since each I_k is an interval and the function f is measurable, each set E_k is measurable. Define the simple functions φ_ϵ and ψ_ϵ on E by

$$\varphi_\epsilon = \sum_{k=1}^{n} y_{k-1} \cdot \chi_{E_k} \text{ and } \psi_\epsilon = \sum_{k=1}^{n} y_k \cdot \chi_{E_k}.$$

Let $x \in E$. Since $f(E) \subseteq [c, d)$, there is a unique $k, 1 \leq k \leq n$, for which $y_{k-1} \leq f(x) < y_k$ and therefore,

$$\varphi_\epsilon(x) = y_{k-1} \leq f(x) < y_k = \psi_\epsilon(x).$$

But $y_k - y_{k-1} < \epsilon$, and therefore φ_ϵ and ψ_ϵ have the required approximation properties. □

The proof of the above lemma depends on a seminal idea of Henri Lebesgue. In the development of the Riemann integral, approximation of functions by step-functions is done by partitioning into subintervals the *domain* of a function and defining approximations with respect to such a partition. In contrast, the approximation of a bounded measurable function by simple functions is done by partitioning into subintervals an interval containing *image* of a function, and then considering preimages of these subintervals in order to obtain simple approximations.

Definition *A measurable function f: $E \to \mathbf{R}$ is said to be* **finitely supported** *provided that it vanishes on the complement of a set of finite measure.*

The Simple Approximation Theorem *If the function f: $E \to \overline{\mathbf{R}}$ is measurable, then there is a sequence $\{\varphi_n: E \to \mathbf{R}\}$ of finitely supported, simple functions that converges pointwise on E to f and has the property that*

$$|\varphi_n| \leq |f| \text{ on } E \text{ for all } n.$$

If $f \geq 0$, then, in addition, $\{\varphi_n\}$ is increasing and each $\varphi_n \geq 0$.

Proof Assume that $f \geq 0$ on E. The general case follows by expressing f as the difference of non-negative measurable functions. For each n, define $E_n = \{x \in E \mid f(x) \leq n\}$. Then E_n is a measurable set and the restriction f: $E_n \to \mathbf{R}$ is a non-negative, bounded, measurable function. By the Simple Approximation Lemma, there are simple functions φ_n: $E_n \to \mathbf{R}$ and ψ_n: $E_n \to \mathbf{R}$ for which

$$0 \leq \varphi_n \leq f \leq \psi_n \text{ and } 0 \leq \psi_n - \varphi_n < 1/n \text{ on } E_n.$$

Observe that

$$0 \leq \varphi_n \leq f \text{ and } 0 \leq f - \varphi_n \leq \psi_n - \varphi_n < 1/n \text{ on } E_n. \tag{2}$$

Extend φ_n to all of E by setting $\varphi_n(x) = n$ if $f(x) > n$. The function φ_n is a simple function defined on E and $0 \leq \varphi_n \leq f$ on E. We claim that the sequence $\{\varphi_n\}$ converges to f pointwise on E. Let $x \in E$.

Case 1: Assume that $f(x)$ is finite. Choose an N for which $f(x) < N$. Then

$$0 \leq f(x) - \varphi_n(x) < 1/n \text{ for } n \geq N,$$

and therefore $\lim_{n \to \infty} \varphi_n(x) = f(x)$.
Case 2: Assume that $f(x) = \infty$. Then $\varphi_n(x) = n$ for all n, so that $\lim_{n \to \infty} \varphi_n(x) = f(x)$.

Since each $\varphi_n \geq 0$, by replacing each φ_n with the product of $\max\{\varphi_1, \ldots, \varphi_n\}$ and $\chi_{[-n, n]}$, we obtain an increasing sequence of finitely supported, simple functions that converges pointwise to f. $\square$

PROBLEMS

13. (Dini's Theorem) Let $\{f_n: [a, b] \to \mathbf{R}\}$ be an increasing sequence of continuous functions that converges pointwise to the continuous function $f: [a, b] \to \mathbf{R}$. Show that the convergence is uniform on $[a, b]$. (Suggestion: Let $\epsilon > 0$. For each n, define $E_n = \{x \in [a, b] \mid f(x) - f_n(x) < \epsilon\}$. Show that $\{E_n\}$ is an open cover of $[a, b]$ and use the Heine-Borel Theorem.)
14. A real-valued measurable function is said to be *semisimple* provided that it takes only a countable number of values. Let $f: E \to \mathbf{R}$ be measurable. Show that there is a sequence of semisimple functions that converges to f uniformly on E.
15. Assume that $m(E) < \infty$ and let $f: E \to \mathbf{R}$ be measurable. Show that for each $\epsilon > 0$, there is a measurable subset E_0 of E such that f is bounded on E_0 and $m(E \sim E_0) < \epsilon$.
16. Assume that $m(E) < \infty$ and let $f: E \to \mathbf{R}$ be measurable. Show that for each $\epsilon > 0$, there is a measurable subset E_0 of E and a sequence $\{\varphi: E \to \mathbf{R}\}$ of simple functions for which $\{\varphi_n\} \to f$ uniformly on E_0 and $m(E \sim E_0) < \epsilon$. (Suggestion: See the preceding problem.)
17. Let E be a measurable subset of $[a, b]$, and define $f: [a, b] \to \mathbf{R}$ by $f = \chi_E$. For each $\epsilon > 0$, show that there is a measurable subset $E_0 \subseteq [a, b]$ and step-function $h: [a, b] \to \mathbf{R}$ for which

$$h = f \text{ on } E_0 \text{ and } m([a, b] \sim E_0) < \epsilon.$$

    (Suggestion: Use Theorem 19 of the preceding chapter.)
18. Let A and B be any sets. Show that

$$\begin{aligned}\chi_{A\cap B} &= \chi_A \cdot \chi_B\\ \chi_{A\cup B} &= \chi_A + \chi_B - \chi_A \cdot \chi_B\\ \chi_{A\sim B} &= [\chi_A - \chi_B] \cdot \chi_A.\end{aligned}$$

19. For a sequence $\{f_n: E \to \mathbf{R}\}$ of measurable functions, show that the functions $\inf\{f_n\}$ and $\sup\{f_n\}$ are measurable.
20. For $f: [a, b] \to \mathbf{R}$ measurable, let $E_+ = \{x \in E \mid f(x) \geq 0\}$ and $E_- = \{x \in E \mid f(x) < 0\}$. By considering the restriction of f to E_+ and E_-, prove the general Simple Approximation Theorem based on the special case of a non-negative measurable function.
21. Let $f: [a, b] \to \mathbf{R}$ be increasing. Show that f is measurable by first showing that, for each n, the strictly increasing function $x \mapsto f(x) + x/n$ is measurable, and then taking pointwise limits.

3.3 LITTLEWOOD'S THREE PRINCIPLES, EGOROFF'S THEOREM, AND LUSIN'S THEOREM

Speaking of the theory of functions of a real variable, J. E. Littlewood says[2], "The extent of knowledge required is nothing like so great as is sometimes supposed. There are three principles, roughly expressible in the following terms: Every [measurable] set is nearly a finite union of intervals; every [measurable] function is nearly continuous; every pointwise convergent sequence of [measurable] functions is nearly uniformly convergent. Most of the results of [the theory] are fairly intuitive applications of these ideas, and the student armed with them should be equal to most occasions when real variable theory is called for. If one of the principles would be the obvious means to settle the problem if it were 'quite' true, it is natural to ask if the 'nearly' is near enough, and for a problem that is actually solvable it generally is."

Theorem 19 of the preceding chapter is a precise formulation of Littlewood's first principle: it states that given a measurable set E of finite measure, for each $\epsilon > 0$, there is a finite, disjoint collection of open intervals for which the union $\mathcal{U}$ is "nearly equal to" E in the sense that $m(E \sim \mathcal{U}) + m(\mathcal{U} \sim E) < \epsilon$. A precise realization of the third of Littlewood's principle is the following theorem.

Egoroff's Theorem *Assume that $m(E) < \infty$. If $\{f_n \colon E \to \mathbf{R}\}$ is a sequence of measurable functions that converges pointwise on E to the function $f \colon E \to \mathbf{R}$, then for each $\epsilon > 0$, there is a closed set F for which*

$$F \subseteq E,\ \ m(E \sim F) < \epsilon \textit{ and } \{f_n\} \to f \textit{ uniformly on } F.$$

To prove Egoroff's Theorem, it is convenient to first establish the following lemma.

Lemma 9 *Under the assumptions of Egoroff's Theorem, for each $\eta > 0$ and $\delta > 0$, there is a measurable set A and an index N for which*

$$A \subseteq E,\ \ m(E \sim A) < \delta \textit{ and } |f_n - f| < \eta \textit{ on } A \textit{ for all } n \geq N.$$

Proof For each k, since the function $|f - f_k|$ is measurable, the set $\{x \in E \mid |f(x) - f_k(x)| < \eta\}$ is measurable. Since $\mathcal{M}$ is a σ-algebra, for each n,

$$E_n = \{x \in E \mid \ |f(x) - f_k(x)| < \eta \text{ for all } k \geq n\}$$

is a measurable set. Then $\{E_n\}_{n=1}^{\infty}$ is an ascending collection of measurable sets, and $E = \cup_{n=1}^{\infty} E_n$, since $\{f_n\}$ converges pointwise to f on E. It follows from the continuity of measure that

$$m(E) = \lim_{n \to \infty} m(E_n).$$

Since $m(E) < \infty$, we may choose an index N for which $m(E_N) > m(E) - \delta$. Define $A = E_N$ and observe that, by the excision property of measure, $m(E \sim A) = m(E) - m(E_N) < \delta$. □

Proof of Egoroff's Theorem By the preceding lemma, for each n, there is a measurable subset A_n of E and an index $N(n)$ for which

$$m(E \sim A_n) < \epsilon / 2^{n+1} \tag{3}$$

[2] Littlewood [Lit41], page 23.

and

$$|f_k - f| < 1/n \text{ on } A_n \text{ for all } k \geq N(n). \tag{4}$$

Define

$$A = \bigcap_{n=1}^{\infty} A_n.$$

By De Morgan's Identities, the countably monotonicity of measure and (3),

$$m(E \sim A) = m\left(\bigcup_{n=1}^{\infty} [E \sim A_n]\right) \leq \sum_{n=1}^{\infty} m(E \sim A_n) < \sum_{n=1}^{\infty} \epsilon/2^{n+1} = \epsilon/2.$$

We claim that $\{f_n\}$ converges to f uniformly on A. Indeed, let $\epsilon > 0$. Choose an index n_0 for which $1/n_0 < \epsilon$. Then, by (4),

$$|f_k - f| < 1/n_0 \text{ on } A_{n_0} \text{ for } k \geq N(n_0).$$

However, $A \subseteq A_{n_0}$ and $1/n_0 < \epsilon$ and therefore,

$$|f_k - f| < \epsilon \text{ on } A \text{ for } k \geq N(n_0).$$

Therefore, $\{f_n\}$ converges to f uniformly on A and $m(E \sim A) < \epsilon/2$. Finally, by the regularity of measure, there is a closed set F contained in A for which $m(A \sim F) < \epsilon/2$. Consequently, $m(E \sim F) < \epsilon$ and $\{f_n\} \to f$ uniformly on F. □

Lemma 10[3] *If F is a closed subset of* $\mathbf{R}$ *and $f\colon F \to \mathbf{R}$ is a continuous function, then it has a continuous extension to $f\colon \mathbf{R} \to \mathbf{R}$.*

Proof The complement of F, $\mathbf{R} \sim F$, is open, so $\mathbf{R} \sim F = \bigcup_{n=1}^{\infty} I_n$, a countable, disjoint union of open intervals. If $I_n = (a_n, b_n)$ is bounded, define f on $[a_n, b_n]$ to be any continuous function that agrees f at the end-points. If I_n is unbounded, define f on I_n to be the constant function that agrees with f at the finite end-point of I_n. We leave it as an exercise to verify the continuity of $f\colon \mathbf{R} \to \mathbf{R}$. □

The following theorem is a confirmation of Littlewood's second principle.

Lusin's Theorem *If $f\colon E \to \mathbf{R}$ is a measurable function, then for each $\epsilon > 0$, there is a continuous function $g\colon \mathbf{R} \to \mathbf{R}$ and a closed subset F of $\mathbf{R}$ for which*

$$F \subseteq E,\ m(E \sim F) < \epsilon \text{ and } f = g \text{ on } F.$$

Proof[4] Choose an enumeration $\{I_n\}_{n=1}^{\infty}$ of the countable collection of all open intervals that have rational end-points. Let $\epsilon > 0$. By the regularity of measure, for each n, there are closed subsets of $\mathbf{R}$, F_n, and H_n, for which $F_n \subseteq f^{-1}(I_n)$, $H_n \subseteq E \sim f^{-1}(I_n)$,

$$m(f^{-1}(I_n)) \sim F_n) < 1/\epsilon^{n+1} \text{ and } m((E \sim f^{-1}(I_n)) \sim H_n) < 1/\epsilon^{n+1}.$$

[3]This lemma is a very special case of the Tietze Extension Theorem, which is proven in Chapter 16.

[4]This elegant proof is due to Peter Loeb and Erik Talvila. We remark in passing that this theorem was first stated and proved by Vitali.

Define $F \equiv \cap_{n=1}^{\infty}(F_n \cup H_n)$. Then F, being the intersection of closed sets, is closed. By De Morgan's Identity and the countable monotonicity of measure,

$$m(E \sim F) = m\left(\bigcup_{n=1}^{\infty} E \sim (F_n \cup H_n)\right) < 2 \cdot \sum_{n=1}^{\infty} \epsilon/2^{n+1} = \epsilon.$$

We may assume that $F \neq \emptyset$, for otherwise, $m(E) < \epsilon$ and we simply take F to be a point in E and take g to be the constant function on $\mathbf{R}$ that agrees with f at that point.

We claim that $f\colon F \to R$ is continuous. To do so, for each $x \in F$ and open interval I containing $f(x)$, we need to show that there is an open subset $\mathcal{O}$ of $\mathbf{R}$ for which

$$f(\mathcal{O} \cap F) \subseteq I.$$

For such an x and I, since the rational numbers are a dense subset of $\mathbf{R}$, there is some k for which $f(x) \in I_k \subseteq I$. We claim that the above inclusion holds for the open set $\mathcal{O} = \mathbf{R} \sim H_k$. Indeed, observe that $f(F_k) \subseteq I_k \subseteq I$, and H_k and F_k are disjoint, so that $\mathcal{O} \cap F \subseteq (\mathbf{R} \sim H_k) \cap (F_k \cup H_k) = F_k$. According to the preceding lemma, f has a continuous extension to all of $\mathbf{R}$. $\square$

Corollary 11 *If* $f\colon E \to \mathbf{R}$ *is a measurable function, then there is a sequence of continuous functions* $\{f_n\colon \mathbf{R} \to \mathbf{R}\}$ *that converges pointwise almost everywhere on* E *to* f.

Proof According to Lusin's Theorem, for each n, there is a continuous function $f_n\colon \mathbf{R} \to \mathbf{R}$ for which

$$m\,\{x \in E \mid\ f_n(x) \neq f(x)\} < 1/2^n.$$

We deduce from the Borel-Cantelli Lemma that for almost all $x \in E$ there is an n such that $f_k(x) = f(x)$ for all $k \geq n$. $\square$

PROBLEMS

22. Verify the continuity of the function $f\colon \mathbf{R} \to \mathbf{R}$ defined in the proof of Lemma 10.
23. For the function $f\colon E \to \mathbf{R}$ and the set F in the statement of Lusin's Theorem, show that the restriction of f to F is continuous. Must there be any points in E at which $f\colon E \to \mathbf{R}$ is continuous?
24. Show that the conclusion of Egoroff's Theorem can fail if we drop the assumption that the domain has finite measure.
25. Show that Egoroff's Theorem continues to hold if f and each f_n is an extended real-valued function that is finite almost everywhere, and the convergence is pointwise almost everywhere on E.
26. Let $\{f_n\colon E \to \mathbf{R}\}$ be a sequence of measurable functions that converges pointwise on E to $f\colon E \to \mathbf{R}$. Show that $E = \bigcup_{k=1}^{\infty} E_k$, where for each E_k is measurable, and $\{f_n\}$ converges uniformly to f on each E_k if $k > 1$, and $m(E_1) = 0$.
27. Show that in Corollary 11, if f is bounded, then the sequence $\{f_n\}$ may be chosen to be uniformly bounded.

CHAPTER 4

Lebesgue Integration

Contents

4.1 Comments on the Riemann Integral 62
4.2 The Integral of a Bounded, Finitely Supported, Measurable Function 64
4.3 The Integral of a Non-negative Measurable Function 70
4.4 The General Lebesgue Integral 74
4.5 Countable Additivity and Continuity of Integration 78

We define the Lebesgue integral for a large linear space of measurable functions, and show that on this space the integral is a monotone, linear functional. A principal virtue of the Lebesgue integral, beyond the extent of the space of integrable functions, is the provision of quite general criteria under which, for a sequence of integrable function $\{f_n: E \to \mathbf{R}\}$, if $\{f_n\} \to f$ pointwise on E, then

$$\lim_{n\to\infty} \int_E f_n \, dm = \int_E [\lim_{n\to\infty} f_n] \, dm = \int_E f \, dm.$$

This is referred to as passage of the limit under the integral sign. In this chapter, criteria to justify this passage are provided for the Lebesgue integral by the Bounded Convergence Theorem, the Monotone Convergence Theorem, and the Dominated Convergence Theorem.

4.1 COMMENTS ON THE RIEMANN INTEGRAL

We recall a few definitions pertaining to the Riemann integral. Let $f: [a, b] \to \mathbf{R}$ be a bounded function and $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$, that is, $a = x_0 < x_1 < \ldots < x_n = b$. Define the **lower and upper Darboux sums** for f with respect to P, respectively, by

$$L(f, P) = \sum_{i=1}^{n} m_i \cdot (x_i - x_{i-1}) \text{ and } U(f, P) = \sum_{i=1}^{n} M_i \cdot (x_i - x_{i-1}),$$

where, for $1 \le i \le n$

$$m_i = \inf \{f(x) \mid x_{i-1} < x < x_i\} \text{ and } M_i = \sup \{f(x) \mid x_{i-1} < x < x_i\}.$$

A bounded function $f: [a, b] \to \mathbf{R}$ is defined to be Riemann integrable provided that

$$\sup_P L(f, P) = \inf_P U(f, P), \tag{1}$$

where the supremum and infimum are taken over all partitions P of $[a, b]$. The common value of this supremum and infimum is called the Riemann integral of f and denoted by $\int_a^b f(x)\,dx$. If $f: [a, b] \to \mathbf{R}$ is continuous, it is uniformly continuous and therefore Riemann integrable. A function η: $[a, b] \to \mathbf{R}$ is called a **step-function** provided that there is a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ and numbers $c_1, \ldots, c_n$ such that for $1 \le i \le n$,

$$\eta(x) = c_i \text{ if } x_{i-1} < x < x_i.$$

It is clear that this step-function is Riemann integrable and

$$\int_a^b \eta(x)\, dx = \sum_{i=1}^n c_i(x_i - x_{i-1}).$$

Consequently, the criterion for a function f: $[a, b] \to \mathbf{R}$ to be Riemann integrable may be recast as

$$\sup_{\varphi \le f} \int_a^b \varphi(x)\, dx = \inf_{f \le \psi} \int_a^b \psi(x)\, dx,$$

where the supremum and infimum are taken over step-functions on $[a, b]$.

Example (**Dirichlet's Function**) Define f: $[0, 1] \to \mathbf{R}$ to be the characteristic function of $E \equiv \mathbf{Q} \cap [0, 1]$. Let P be any partition of $[0, 1]$. Both the rational and the irrational numbers are dense in $[0, 1]$, so that

$$L(f, P) = 0 \text{ and } U(f, P) = 1.$$

Therefore, (1) does not hold, and so f is not Riemann integrable. Now, E is countable. Let $\{q_k\}_{k=1}^\infty$ be an enumeration of E. For each n, let f_n: $[0, 1] \to \mathbf{R}$ be the characteristic function of $\{q_1, \cdots, q_n\}$. Then each f_n is a step-function, so it is Riemann integrable. Moreover, $\{f_n\}$ is an increasing sequence of Riemann integrable functions,

$$|f_n| \le 1 \text{ on } [0, 1] \text{ for all } n \text{ and } \{f_n\} \to f \text{ pointwise on } [0, 1].$$

However, the pointwise limit function f fails to be Riemann integrable.

According to Theorem 15 in the next chapter, a bounded function f: $[a, b] \to \mathbf{R}$ is Riemann integrable if and only if the set of points at which they fails to be continuous has measure zero. Furthermore, in the next section, Arzelà's Bounded Convergence Theorem is proven, according to which if a sequence $\{f_n: [a, b] \to \mathbf{R}\}$ of Riemann integrable functions is uniformly pointwise bounded and converges pointwise to a function f: $[a, b] \to \mathbf{R}$ that is *assumed* to be Riemann integrable, then, as Riemann integrals,

$$\lim_{n \to \infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx.$$

PROBLEMS

1. Show that, in the above Dirichlet function example, $\{f_n\}$ fails to converge to f uniformly on $[0, 1]$.
2. A partition P' of $[a, b]$ is called a refinement of a partition P provided that each partition point of P is also a partition point of P'. For a bounded function f: $[a, b] \to \mathbf{R}$, show that under refinement lower Darboux sums do not decrease and upper Darboux sums do not increase.

3. Show that for a bounded function f: $[a, b] \to \mathbf{R}$, the left-hand side of (1) is no greater than the right-hand side.
4. Assume that f: $[a, b] \to \mathbf{R}$ is bounded. Show that f: $[a, b] \to \mathbf{R}$ is Riemann integrable if and only if there is a sequence $\{P_n\}$ of partitions of $[a, b]$, each P_{n+1} being a refinement of P_n, for which $\lim_{n \to \infty} [U(f, P_n) - L(f, P_n)] = 0$.
5. Use the preceding problem to show that a continuous function f: $[a, b] \to \mathbf{R}$, being uniformly continuous, is Riemann integrable.
6. Let f: $[0, 1] \to \mathbf{R}$ be increasing. For a natural number n, define P_n to be the partition of $[0, 1]$ into n subintervals of length $1/n$. Show that $U(f, P_n) - L(f, P_n) \leq 1/n[f(1) - f(0)]$. Use Problem 4 to show that f: $[0, 1] \to \mathbf{R}$ is Riemann integrable.
7. Let $\{f_n$: $[a, b] \to \mathbf{R}\}$ be a uniformly bounded sequence of Riemann integrable functions that converges uniformly to f: $[a, b] \to \mathbf{R}$. Show that f: $[a, b] \to \mathbf{R}$ also is Riemann integrable, and
$$\lim_{n \to \infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx.$$

4.2 THE INTEGRAL OF A BOUNDED, FINITELY SUPPORTED, MEASURABLE FUNCTION

The Dirichlet function exhibits one of the principal drawbacks of the Riemann integral: a uniformly bounded, increasing sequence of Riemann integrable functions on a closed, bounded interval may converge pointwise to a function that is not Riemann integrable. We will see that the Lebesgue integral does not have this shortcoming. Recall that a function is said to be finitely supported provided that it vanishes on the complement of a set of finite measure.

Definition *Let ψ: $E \to \mathbf{R}$ be a finitely supported, simple function. Then the integral of ψ over E, $\int_E \psi\, dm$, is defined as follows: if ψ is identically zero, define $\int_E \psi\, dm = 0$, and otherwise, let $\lambda_1, \ldots \lambda_n$ be the finite number of non-zero functional values taken by ψ and define*
$$\int_E \psi\, dm = \sum_{i=1}^{n} \lambda_i \cdot m(E_i) \text{ where each } E_i = \psi^{-1}(\lambda_i). \tag{2}$$

Observe that this sum is properly defined since ψ is finitely supported, so that each $m(E_i) < \infty$.

Proposition 1 *If the functions φ: $E \to \mathbf{R}$ and ψ: $E \to \mathbf{R}$ are finitely supported and simple, then for any α and β,*
$$\int_E (\alpha\varphi + \beta\psi)\, dm = \alpha \int_E \varphi\, dm + \beta \int_E \psi\, dm.$$

Proof Since φ and ψ are finitely supported, we may assume that $m(E) < \infty$. Consider the case $\alpha = \beta = 1$ and leave the case of general coefficients as an exercise. To begin, we establish the following special case: Let $\{A_j\}_{j=1}^m$ be a measurable partition of E and for $1 \leq j \leq m$, let $a_j \in \mathbf{R}$. We claim that

$$\text{if } \eta = \sum_{j=1}^{m} a_j \cdot \chi_{A_j} \text{ on } E, \text{ then } \int_E \eta \, dm = \sum_{j=1}^{m} a_j \cdot m(A_j). \tag{3}$$

Indeed, let $\{\lambda_1, \ldots, \lambda_n\}$ be the distinct values taken by η, and for each i, $1 \leq i \leq n$, define $E_i = \eta^{-1}(\lambda_i)$. Observe that each $j, 1 \leq j \leq m$, $\{E_i \cap A_j\}_{i=1}^n$ is a measurable partition of A_j, for each $i, 1 \leq i \leq n, \{E_i \cap A_j\}_{j=1}^m$ is a measurable partition of E_i, and furthermore, $a_j \cdot \chi_{E_i \cap A_j} = \lambda_i \cdot \chi_{E_i \cap A_j}$. Therefore, by twice using the finite additivity of measure,

$$\begin{aligned}
\sum_{j=1}^{m} a_j \cdot m(A_j) &= \sum_{j=1}^{m} a_j \cdot \left[\sum_{i=1}^{n} m(E_i \cap A_j) \right] \\
&= \sum_{i=1}^{n} \lambda_i \cdot \left[\sum_{j=1}^{m} m(E_i \cap A_j) \right] \\
&= \sum_{i=1}^{n} \lambda_i \cdot m(E_i) = \int_E \eta \, dm.
\end{aligned}$$

Therefore, (3) holds. By taking intersections of the level sets of φ and ψ, we obtain a measurable partition $\{A_j\}_{j=1}^m$ of E for which both ψ and φ are constant on each A_j. Linearity follows by using this partition and substituting ψ and φ for η in (3). □

The **linearity** of integration for finitely supported, simple functions has the following two important consequences. First, integration is **monotone**, meaning that

$$\text{if } \varphi \leq \psi, \text{ then } \int_E \varphi \, dm \leq \int_E \psi \, dm.$$

Indeed, by linearity, $\int_E \psi \, dm - \int_E \varphi \, dm = \int_E [\psi - \varphi] \, dm \geq 0$. Also, integration has an **additivity over domains** property, meaning that if A and B are disjoint measurable subsets of E, then

$$\int_{A \cup B} \varphi \, dm = \int_A \varphi \, dm + \int_B \varphi \, dm.$$

To see this, observe that, since A and B are disjoint, $\varphi \cdot \chi_{A \cup B} = \varphi \cdot \chi_A + \varphi \cdot \chi_B$, and so additivity over domains also follows from linearity.

Definition *A bounded function f: $E \to \mathbf{R}$ defined on a measurable set E is said to be* **Lebesgue integrable** *or simply* **integrable**, *provided that*

$$\sup_{\varphi \leq f} \int_E \varphi \, dm = \inf_{f \leq \psi} \int_E \psi \, dm$$

where the infimum and supremum are taken over finitely supported, simple functions on E. The common value of the infimum and supremum is called the **Lebesgue integral**, *or simply the* **integral** *of f over E, and denoted by $\int_E f \, dm$.*

We leave it as an exercise to prove the following theorem, by observing that each step-function is simple and has the same Riemann and Lebesgue integral.

Theorem 2 *If f: $[a, b] \to \mathbf{R}$ is a bounded, Riemann integrable function, then it is Lebesgue integrable and the two integrals are equal.*

Example The Dirichlet function is the function χ_E: $[0, 1] \to \mathbf{R}$, where $E = \mathbf{Q} \cap [0, 1]$. Since E is a measurable set, χ_E is a simple function and so is Lebesgue integrable, but it fails to be Riemann integrable.

Theorem 3 *If f: $E \to \mathbf{R}$ is a bounded, finitely supported, measurable function, then it is integrable.*

Proof Since f is finitely supported, we may assume that $m(E) < \infty$. Since f is measurable and bounded, for each n, by the Simple Approximation Lemma, there are simple functions φ_n: $E \to \mathbf{R}$ and ψ_n: $E \to \mathbf{R}$ for which

$$\varphi_n \leq f \leq \psi_n \text{ and } 0 \leq \psi_n - \varphi_n \leq 1/n \text{ on } E.$$

By the monotonicity and linearity of the integral for simple functions,

$$0 \leq \int_E \psi_n \, dm - \int_E \varphi_n \, dm = \int_E [\psi_n - \varphi_n] \, dm \leq 1/n \cdot m(E).$$

However,

$$\begin{aligned} 0 &\leq \inf\left\{\int_E \psi \, dm \;\middle|\; \psi \text{ simple}, \psi \geq f\right\} - \sup\left\{\int_E \varphi \, dm \;\middle|\; \varphi \text{ simple}, \varphi \leq f\right\} \\ &\leq \int_E \psi_n \, dm - \int_E \varphi_n \, dm \leq 1/n \cdot m(E). \end{aligned}$$

This inequality holds for every natural number n and $m(E)$ is finite. Therefore, f: $E \to \mathbf{R}$ is integrable. $\square$

The converse of this theorem holds: a bounded, finitely supported function is Lebesgue integrable if and only if it is measurable. This is Theorem 14 of the next chapter. Observe that if f: $E \to \mathbf{R}$ is bounded, finitely supported and measurable, and E_0 is measurable, then

$$\int_E f \, dm = \int_{E_0} f \, dm \text{ if } \{x \in E \mid f(x) \neq 0\} \subseteq E_0 \subseteq E.$$

Theorem 4 *If f: $E \to \mathbf{R}$ and g: $E \to \mathbf{R}$ are bounded, finitely supported, measurable functions, then for any α and β,*

$$\int_E (\alpha f + \beta g) \, dm = \alpha \int_E f \, dm + \beta \int_E g \, dm.$$

Proof In view of the preceding remark, since f and g are finitely supported, we may assume that $m(E) < \infty$. Consider the case $\alpha = \beta = 1$ and leave the case of general coefficients as an exercise. The sum of bounded, finitely supported, measurable functions has the same properties, and therefore, by Theorem 3, $f + g$ is integrable. For each n, according to the Simple Approximation Lemma, there are four simple functions $\psi_n \colon E \to \mathbf{R}$, $\varphi_n \colon E \to \mathbf{R}$, $\overline{\psi}_n \colon E \to \mathbf{R}$, and $\overline{\varphi}_n \colon E \to \mathbf{R}$ for which

$$\varphi_n \leq f \leq \psi_n \text{ and } \overline{\varphi}_n \leq g \leq \overline{\psi}_n \text{ on } E,$$

and

$$0 \leq \psi_n - \varphi_n \leq 1/n \text{ and } 0 \leq \overline{\psi}_n - \overline{\varphi}_n \leq 1/n \text{ on } E. \tag{4}$$

By the definition of the integrals of f and g,

$$\int_E \varphi_n \, dm \leq \int_E f \, dm \leq \int_E \psi_n \, dm \text{ and } \int_E dm \overline{\varphi}_n \leq \int_E g \, dm \leq \int_E \overline{\psi}_n \, dm,$$

and therefore, by the linearity of integration for simple functions,

$$\int_E [\varphi_n + \overline{\varphi}_n] \, dm \leq \int_E f \, dm + \int_E g \, dm \leq \int_E [\psi_n + \overline{\psi}_n] \, dm. \tag{5}$$

Now

$$\varphi_n + \overline{\varphi}_n \leq f + g \leq \psi_n + \overline{\psi}_n \text{ on } E.$$

and so, by the definition of the integral of $f + g$,

$$\int_E [\varphi_n + \overline{\varphi}_n] \, dm \leq \int_E [f + g] \, dm \leq \int_E [\psi_n + \overline{\psi}_n] \, dm. \tag{6}$$

By the linearity and monotonicity of integration for simple functions and the inequalities (4), we have

$$\int_E [\psi_n + \overline{\psi}_n] \, dm - \int_E [\varphi_n + \overline{\varphi}_n] \, dm = \int_E [\varphi_n - \psi_n] \, dm + \int_E [\varphi_n - \psi_n] \, dm \leq 2/n \cdot m(E)$$

We conclude from (5) and (6) that the numbers $\int_E [f + g] \, dm$ and $\int_E f \, dm + \int_E g \, dm$ belong to an interval of length at most $2/n \cdot m(E)$ and therefore,

$$\left| \int_E [f + g] \, dm - \left[\int_E f \, dm + \int_E g \, dm \right] \right| \leq 2/n \cdot m(E).$$

Since $m(E) < \infty$ and this inequality holds for all n, $\int_E [f + g] \, dm = \int_E f \, dm + \int_E g \, dm$.

□

As it did in the case of finitely supported, simple functions, the monotonicity and additivity over domains properties for the integral of a bounded, finitely supported, measurable function follow from linearity.

Corollary 5 *If $f \colon E \to \mathbf{R}$ is a bounded, finitely supported, measurable function, then*

$$\left| \int_E f \, dm \right| \leq \int_E |f| \, dm. \tag{7}$$

Proof The function $|f|$ is measurable and bounded and

$$-|f| \leq f \leq |f| \text{ on } E.$$

By the linearity and monotonicity of integration,

$$-\int_E |f|\, dm \leq \int_E f\, dm \leq \int_E |f|\, dm,$$

that is, (7) holds. □

Let $\{f_n \colon E \to \mathbf{R}\}$ be a sequence of bounded, finitely supported, measurable functions. We now provide the first criterion for justifying that if $\{f_n\} \to f$ pointwise on E, then

$$\lim_{n \to \infty} \int_E f_n\, dm = \int_E \lim_{n \to \infty} f_n\, dm = \int_E f\, dm.$$

We refer to this equality as passage of the limit under the integral sign. As the following example shows, this passage is not always justified.

Example For each n, let $I_n = [1/2n, 1/n]$ and define $f_n \colon [0, 1] \to \mathbf{R}$ by $f_n = 2n \cdot \chi_{I_n}$. Observe that $\int_0^1 f_n = 1$ for each n. Define $f \equiv 0$ on $[0, 1]$. Then $\{f_n\} \to f$ pointwise on $[0, 1]$ but $\lim_{n \to \infty} \int_0^1 f_n \neq \int_0^1 f$. For each n, let $E_n = [n, n + 1]$ and define $g_n \colon \mathbf{R} \to \mathbf{R}$ by $g_n = n \cdot \chi_{E_n}$. Again for this sequence, there is pointwise convergence of the functions without convergence of the integrals.

Definition *A sequence of functions* $\{f_n \colon E \to \mathbf{R}\}$ *is said to be* **uniformly pointwise bounded** *provided that there is an* $M \geq 0$ *for which*

$$|f_n| \leq M \ \text{ on } E \text{ for all } n.$$

The Bounded Convergence Theorem *Assume that* $m(E) < \infty$ *and* $\{f_n \colon E \to \mathbf{R}\}$ *is a uniformly pointwise bounded sequence of measurable functions. If* $\{f_n\} \to f$ *pointwise on* E, then

$$\lim_{n \to \infty} \int_E f_n\, dm = \int_E f\, dm.$$

Proof The function f, being the pointwise limit of a sequence of measurable functions, is measurable. Choose $M \geq 0$ for which $|f_n| \leq M$ on E for all n. Then $|f| \leq M$ on E. Let A be any measurable subset of E. For each n, by the linearity and additivity over domains properties of integration,

$$\int_E f_n\, dm - \int_E f\, dm = \int_E [f_n - f]\, dm = \int_A [f_n - f]\, dm + \int_{E \sim A} f_n\, dm + \int_{E \sim A} (-f)\, dm.$$

Therefore, by Corollary 5 and the monotonicity of integration,

$$\left| \int_E f_n\, dm - \int_E f\, dm \right| \leq \int_A |f_n - f|\, dm + 2M \cdot m(E \sim A). \tag{8}$$

To prove convergence of the integrals, let $\epsilon > 0$. Since $m(E) < \infty$ and f is real-valued, according to Egoroff's Theorem, there is a measurable subset A of E for which $\{f_n\} \to f$ uniformly on A and $m(E \sim A) < \epsilon/4M$. By uniform convergence, there is an index N for which

$$|f_n - f| < \frac{\epsilon}{2 \cdot m(E)} \text{ on } A \text{ for all } n \geq N.$$

Consequently, for $n \geq N$, it follows from (8) and the monotonicity of integration that

$$\left| \int_E f_n \, dm - \int_E f \, dm \right| \leq \frac{\epsilon}{2 \cdot m(E)} \cdot m(A) + 2M \cdot m(E \sim A) < \epsilon.$$

Therefore, $\lim_{n \to \infty} \{\int_E f_n\} \, dm = \int_E f \, dm.$ □

The Arzelà Bounded Convergence Theorem *Let $\{f_n\colon [a, b] \to \mathbf{R}\}$ be a uniformly pointwise bounded sequence of Riemann integrable functions that converges pointwise to a Riemann integrable function $f\colon [a, b] \to \mathbf{R}$. Then, as Riemann integrals,*

$$\lim_{n \to \infty} \int_a^b f_n(x) \, dx = \int_a^b f(x) \, dx.$$

Proof For each n, the function $f - f_n$, being the difference of Riemann integrable functions, is Riemann integrable and therefore also is Lebesgue integrable. Consequently, $\{f - f_n\colon [a, b] \to \mathbf{R}\}$ is a uniformly pointwise bounded sequence of Lebesgue integrable functions and for each $x \in [a, b]$, $\lim_{n \to \infty} [f(x) - f_n(x)] = 0$. According to the Bounded Convergence Theorem,

$$\lim_{n \to \infty} \int_{[a, b]} [f - f_n] \, dm = 0.$$

However, the Lebesgue integral of a Riemann integrable function agrees with its Riemann integral. Therefore, $\lim_{n \to \infty} \int_a^b f_n(x) \, dx = \int_a^b f(x) \, dx.$ □

Arzelà's Bounded Convergence Theorem is an immediate consequence of the Bounded Convergence Theorem. Nevertheless, Arzelà's Theorem is quite remarkable, in that it was proven almost two decades prior to the development of Lebesgue integration. The challenge of finding an elementary proof, meaning, roughly speaking, a proof that is independent of the countable additivity of Lebesgue measure, was accepted by many eminent mathematicians, including Bieberbach, Hausdorff, and Riesz[1].

PROBLEMS

8. Assume that $m(E) < \infty$ and let $\{f_n\colon E \to \mathbf{R}\}$ be a sequence of bounded, measurable functions. Without referring to the Bounded Convergence Theorem or Egoroff's Theorem, give a direct proof that if $\{f_n\} \to f$ uniformly on E, then $\lim_{n \to \infty} \int_E f_n \, dm = \int_E f \, dm$.

---

[1]In *Arzelà's Dominated Convergence Theorem for the Riemann Integral (the American Mathematical Monthly*, November, 1971), W. A. J. Luxemburg compares several elementary proofs, and presents such a proof of his own. What is particularly interesting are his observations regarding the shadows of the countable additivity of Lebesgue measure that he detects in these so-called elementary proofs.

9. Let $m(E) < \infty$ and $\{E_n\}_{n=1}^{\infty}$ be a measurable partition of E. Just using the linearity of integration and the Bounded Convergence Theorem, show that $m(E) = \sum_{n=1}^{\infty} m(E_n)$.
10. Let $f\colon E \to \mathbf{R}$ be bounded, finitely supported and measurable. For a measurable subset A of E, show that $\int_A f|_A \, dm = \int_E f \cdot \chi_A \, dm$.
11. Does the Bounded Convergence Theorem hold for the Riemann integral?
12. Let $f\colon E \to \mathbf{R}$ and $g\colon E \to \mathbf{R}$ be bounded, finitely supported and measurable, and assume that $f = g$ almost everywhere on E. Show that $\int_E f \, dm = \int_E g \, dm$.
13. Prove Theorem 2.

4.3 THE INTEGRAL OF A NON-NEGATIVE MEASURABLE FUNCTION

Definition *The* **integral** *of a non-negative, measurable function $f\colon E \to [0, \infty]$ is defined by*

$$\int_E f \, dm = \sup_{0 \le h \le f} \int_E h \, dm,$$

where the supremum is taken over bounded, finitely supported, measurable functions h.

Theorem 6 *If $f\colon E \to [0, \infty]$ and $g\colon E \to [0, \infty]$ are non-negative, measurable functions, then for any $\alpha > 0$ and $\beta > 0$,*

$$\int_E (\alpha f + \beta g) \, dm = \alpha \int_E f \, dm + \beta \int_E g \, dm.$$

Proof We consider the case $\alpha = \beta = 1$ and leave the case of general coefficients as an exercise. For brevity, let $\mathcal{F}$ denote the collection of bounded, finitely supported, non-negative, measurable functions on E. We first verify the inequality

$$\int_E f \, dm + \int_E g \, dm \le \int_E (f + g) \, dm, \tag{9}$$

and, in view of the definitions of $\int_E g \, dm$ and $\int_E f \, dm$ as suprema, in order to do so it is necessary and sufficient to show that if $h, k \in \mathcal{F}$, $0 \le h \le f$ and $0 \le k \le g$ on E, then

$$\int_E h \, dm + \int_E k \, dm \le \int_E (f + g) \, dm.$$

Let h and k be two such functions. We have $0 \le h + k \le f + g$ on E, and $h + k \in \mathcal{F}$. Therefore, by the linearity of integration for functions in $\mathcal{F}$,

$$\int_E h \, dm + \int_E k \, dm = \int_E (h + k) \, dm \le \int_E (f + g) \, dm,$$

and so (9) is verified. It remains to verify this inequality in the opposite direction, namely,

$$\int_E (f + g) \, dm \le \int_E f \, dm + \int_E g \, dm, \tag{10}$$

and to do so, in view of the definition of $\int_E (f+g)\,dm$ as a supremum, it is necessary and sufficient to show that if $\ell \in \mathcal{F}$ and $\ell \leq f+g$ on E, then

$$\int_E \ell\,dm \leq \int_E f\,dm + \int_E g\,dm. \tag{11}$$

For such a function ℓ, define the functions h and k on E by

$$h = \min\{f, \ell\} \text{ and } k = \ell - h \text{ on } E.$$

We claim that

$$0 \leq h \leq f,\ 0 \leq k \leq g \ \text{ and } \ell = h + k \text{ on } E.$$

It suffices to verify that $k \leq g$. Indeed, let $x \in E$. If $\ell(x) \leq f(x)$, then $k(x) = 0 \leq g(x)$; if $\ell(x) > f(x)$, then $k(x) = \ell(x) - f(x) \leq g(x)$. Therefore, $k \leq g$ on E. Both h and k belong to $\mathcal{F}$ and, again employing the linearity of integration for functions $\mathcal{F}$, we have

$$\int_E \ell\,dm = \int_E h\,dm + \int_E k\,dm \leq \int_E f\,dm + \int_E g\,dm.$$

Therefore, (11) holds and the proof of linearity is complete. □

As it did in the case of finitely supported, simple functions, linearity of integration for non-negative, measurable functions implies monotonicity and additivity over domains.

Fatou's Lemma *If $\{f_n\colon E \to [0, \infty]\}$ is a sequence of non-negative, measurable functions and $\{f_n\} \to f$ pointwise on E,* then

$$\int_E f\,dm \leq \liminf \int_E f_n\,dm. \tag{12}$$

Proof The function f is non-negative and measurable, since it is the pointwise limit of a sequence of such functions. To verify the inequality (12), in view of the definition of $\int_E f\,dm$ as a supremum, it is necessary and sufficient to show that if h is any finitely supported, bounded, measurable function for which $0 \leq h \leq f$ on E, then

$$\int_E h\,dm \leq \liminf \int_E f_n\,dm. \tag{13}$$

Let h be such a function. Since h is finitely supported, we may assume that $m(E) < \infty$. For each n, define the function $h_n\colon E \to \mathbf{R}$ by

$$h_n = \min\{h, f_n\} \text{ on } E.$$

Since h is bounded, $\{h_n\colon E \to [0, \infty]\}$ is a uniformly pointwise bounded sequence of measurable functions on a domain of finite measure. Furthermore, for each $x \in E$, since $h(x) \leq f(x)$ and $\{f_n(x)\} \to f(x)$, $\{h_n(x)\} \to h(x)$. According to the Bounded Convergence Theorem,

$$\lim_{n\to\infty} \int_E h_n\,dm = \int_E h\,dm.$$

However, for each n, $h_n \leq f_n$ on E and h_n is finitely supported, bounded, and measurable, so that, by the definition of the integral of f_n, $\int_E h_n\, dm \leq \int_E f_n\, dm$. Consequently,

$$\int_E h\, dm = \lim_{n\to\infty} \int_E h_n\, dm \leq \liminf \int_E f_n\, dm,$$

and so inequality (13) is verified. □

The inequality in Fatou's Lemma may be strict, as it is, for instance, if each f_n: $[0, 1] \to [0, \infty)$ is defined by $f_n = n \cdot \chi_{[0,\, 1/n]}$, or if each f_n: $\mathbf{R} \to [0, \infty)$ is defined by $f_n = \chi_{[n,\, n+1]}$. However, the inequality is an equality if the sequence of functions is increasing.

The Monotone Convergence Theorem *If $\{f_n\colon E \to [0, \infty]\}$ is an increasing sequence of non-negative, measurable functions and $\{f_n\} \to f$ pointwise on E, then*

$$\lim_{n\to\infty} \int_E f_n\, dm = \int_E f\, dm.$$

Proof According to Fatou's Lemma,

$$\int_E f\, dm \leq \liminf \int_E f_n\, dm.$$

However, for each n, $f_n \leq f$ on E, and so, by the monotonicity of integration, $\int_E f_n\, dm \leq \int_E f\, dm$, so that

$$\limsup \int_E f_n\, dm \leq \int_E f\, dm.$$

Therefore, $\lim_{n\to\infty} \int_E f_n\, dm = \int_E f\, dm$. □

Beppo Levi's Theorem *If $\{f_n\colon E \to [0, \infty]\}$ is an increasing sequence of non-negative, measurable functions and the sequence of integrals $\{\int_E f_n\, dm\}$ is bounded, then $\{f_n\}$ converges pointwise on E to a non-negative, measurable function $f\colon E \to [0, \infty]$ for which*

$$\lim_{n\to\infty} \int_E f_n\, dm = \int_E f\, dm < \infty.$$

Proof Since $\{f_n\colon E \to [0, \infty]\}$ is an increasing sequence, we may define the non-negative, measurable function $f\colon E \to [0, \infty]$ pointwise on E by

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

According to the Monotone Convergence Theorem, $\{\int_E f_n\, dm\} \to \int_E f\, dm$. Since the sequence of real numbers $\{\int_E f_n\, dm\}$ is bounded, $\int_E f\, dm < \infty$. □

Chebychev's Inequality *If $f\colon E \to [0, \infty]$ is a non-negative, measurable function, then for any $\lambda > 0$,*

$$m\,\{x \in E \mid f(x) \geq \lambda\} \leq \frac{1}{\lambda} \cdot \int_E f\, dm. \tag{14}$$

Proof Define $E_\lambda = \{x \in E \mid f(x) \geq \lambda\}$. First suppose $m(E_\lambda) = \infty$. For each n, define $E_{\lambda,\, n} = E_\lambda \cap [-n, n]$ and $\psi_n = \lambda \cdot \chi_{E_{\lambda,\, n}}$. Then ψ_n is a finitely supported, bounded, measurable function, and therefore,

$$\lambda \cdot m(E_{\lambda,\, n}) = \int_E \psi_n \, dm \text{ and } 0 \leq \psi_n \leq f \text{ on } E \text{ for all } n.$$

It follows from the continuity of measure that

$$\infty = \lambda \cdot m(E_\lambda) = \lambda \cdot \lim_{n \to \infty} m(E_{\lambda,\, n}) = \lim_{n \to \infty} \int_E \psi_n \, dm.$$

The inequality (14) holds since both sides equal ∞. Now consider the case $m(E_\lambda) < \infty$. Define $h = \lambda \cdot \chi_{E_\lambda}$. Then h is a finitely supported, bounded, measurable function and $0 \leq h \leq f$ on E. By the definition of the integral,

$$\lambda \cdot m(E_\lambda) = \int_E h \, dm \leq \int_E f \, dm.$$

Divide both sides of this inequality by λ to obtain Chebychev's Inequality. □

Definition *A non-negative, measurable function $f\colon E \to [0, \infty]$ is said to be* **integrable** *provided that*

$$\int_E f \, dm < \infty.$$

It follows from linearity of integration for non-negative, measurable functions that the sum of integrable functions is integrable.

Proposition 7 *If $f\colon E \to \mathbf{R}$ is a non-negative, integrable function, then it is finite almost everywhere.*

Proof For each n, it follows from the monotonicity of measure and Chebychev's Inequality that

$$m\,\{x \in E \mid \ f(x) = \infty\} \leq m\,\{x \in E \mid \ f(x) \geq n\} \leq \frac{1}{n} \int_E f \, dm.$$

But $\int_E f \, dm$ is finite and this holds for all n, and therefore $m\{x \in E \mid f(x) = \infty\} = 0$. □

We showed that if E is measurable, $E_0 \subseteq E$ and $m(E \sim E_0) = 0$, then a function $f\colon E \to \mathbf{R}$ is measurable if and only if $f\colon E_0 \to \mathbf{R}$ is measurable, and by additivity over domains for integration, if $f\colon E_0 \to \mathbf{R}$ is non-negative and measurable, then $\int_E f \, dm = \int_{E_0} f \, dm$. In view of this, in the assumptions of Fatou's Lemma, the Monotone Convergence Theorem and Beppo Levi's Theorem, pointwise convergence almost everywhere is sufficient, rather than pointwise convergence on the entire domain. Moreover, by this same property and Proposition 7, given a sequence $\{f_n\colon E \to [0, \infty]\}$ of integrable functions, without loss of generality, one may assume that it is a sequence of real-valued functions. All this is consistent with the identification of functions that are equal almost everywhere. Consideration of pointwise convergence almost everywhere is not made just for the sake of generality. It is necessary, and we provide two illustrations of this. First, Lebesgue's

Differentiation Theorem, which is proven in Chapter 6, asserts that if f: $[a, b] \to \mathbf{R}$ is increasing, then

$$\lim_{n \to \infty} \frac{f(x+1/n) - f(x)}{1/n} = f'(x) < \infty, \text{ for almost all } x \in [a, b],$$

and this refined Fatou's Lemma will be used to obtain an upper estimate for $\int_{[a, b]} f'\, dm$. Second, a theorem of Riesz, which is proven in the next chapter, asserts that if $\{f_n \colon E \to [0, \infty]\}$ is a sequence of non-negative, measurable functions and $\lim_{n\to\infty} \int_E f_n\, dm = 0$, then there is a subsequence that converges pointwise almost everywhere on E to $f = 0$. No conclusion can be drawn regarding pointwise convergence on all of E.

PROBLEMS

14. Let $\{E_n\}_{n=1}^{\infty}$ be a measurable partition of E. Just using the Monotone Convergence Theorem and the linearity of integration, show that $m(E) = \sum\limits_{n=1}^{\infty} m(E_n)$.
15. (i) Define $g(x) = x^{\alpha}$ for $0 < x \leq 1$, and $g(0) = 0$. Compute $\int_{[0, 1]} g\, dm$.

    (ii) Define $h(x) = x^{\alpha}$ for $x \geq 1$. Compute $\int_{[1, \infty]} h\, dm$.
16. If f: $[a, b] \to [0, \infty]$ is measurable, show that $\int_{[a, b]} f\, dm = \int_{(a, b)} f\, dm$.
17. Let $\{u_n \colon E \to [0, \infty]\}$ be a sequence of measurable functions. Define $s = \sum\limits_{n=1}^{\infty} u_n$. Show that $\int_E s\, dm = \sum_{n=1}^{\infty} \int_E u_n\, dm$.
18. If f: $[a, b] \to [0, \infty]$ is measurable, show that $\int_{[a, b]} f\, dm = 0$ if and only if $f = 0$ almost everywhere
19. Prove the linearity of integration for general coefficients α and β.
20. Let $\{a_n\}$ be a sequence of non-negative real numbers. Define the function f on $E = [1, \infty)$ by setting $f(x) = a_n$ if $n \leq x < n + 1$. Show that $\int_E f\, dm = \sum_{n=1}^{\infty} a_n$.
21. Let f: $E \to [0, \infty]$ be measurable. Show that

$$\int_E f\, dm = \sup_{0 \leq \varphi \leq f} \int_E \varphi\, dm,$$

    where the supremum is taken over finitely supported, simple functions.
22. Let $\{f_n \colon E \to [0, \infty]\}$ be a sequence of measurable functions that converges pointwise on E to f. Suppose $f_n \leq f$ on E for each n. Show that $\lim_{n\to\infty} \int_E f_n\, dm = \int_E f\, dm$.
23. Show that the Monotone Convergence Theorem does not hold for decreasing sequences of functions.
24. Prove the following generalization of Fatou's Lemma: If $\{f_n \colon E \to [0, \infty]\}$ is a sequence of measurable functions, then $\int_E [\liminf f_n]\, dm \leq \liminf \int_E f_n\, dm$.

4.4 THE GENERAL LEBESGUE INTEGRAL

For a measurable function f: $E \to \overline{\mathbf{R}}$, in the preceding chapter, we defined two non-negative, measurable functions, the positive and negative parts of f, f^+, and f^-, for which

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

Observe that if $|f|: E \to \overline{\mathbf{R}}$ is integrable, since $0 \le f^+ \le |f|$ and $0 \le f^- \le |f|$ on E, by the monotonicity of integration for non-negative, measurable functions, both f^+ and f^- are integrable.

Definition *A measurable function $f: E \to \overline{\mathbf{R}}$ is said to be* **integrable** *provided that $\int_E |f|\, dm < \infty$, and for such a function, the* **integral** *of f over E is defined by*

$$\int_E f\, dm = \int_E f^+\, dm - \int_E f^-\, dm.$$

Proposition 8 *If $f: E \to \overline{\mathbf{R}}$ is an integrable function, then it is finite almost everywhere.*

Proof According to Proposition 7, $|f|$ is finite almost everywhere on E, and therefore so is f. □

The following criterion for integrability is the Lebesgue integral correspondent of the comparison test for the absolute convergence of series.

Theorem 9 (**the Integral Comparison Test**) *If $f: E \to \mathbf{R}$ is a measurable function, and there is a non-negative, integrable function $g: E \to \mathbf{R}$ that dominates f on E, in the sense that*

$$|f| \le g \text{ on } E,$$

then $f: E \to \mathbf{R}$ is integrable and

$$\left| \int_E f\, dm \right| \le \int_E |f|\, dm.$$

Proof By the monotonicity of integration for non-negative, measurable functions, $|f|$ is measurable, and therefore, by definition, so is f. By the linearity of integration for non-negative, measurable functions,

$$\left| \int_E f\, dm \right| = \left| \int_E f^+\, dm - \int_E f^-\, dm \right| \le \int_E f^+\, dm + \int_E f^-\, dm = \int_E |f|\, dm. \qquad \square$$

According to the preceding proposition, an integrable function may be identified with a real-valued function, so that the sum of two such functions is properly defined. Also, the integral is unchanged by this identification. It follows from the integral comparison test that the sum of integrable functions is integrable.

Theorem 10 (**the Linearity of Integration**) *If $f: E \to \mathbf{R}$ and $g: E \to \mathbf{R}$ are integrable functions, then for any α and β, the function $\alpha f + \beta g: E \to \mathbf{R}$ also is integrable and*

$$\int_E (\alpha f + \beta g)\, dm = \alpha \int_E f\, dm + \beta \int_E g\, dm.$$

Proof Consider the case $\alpha = \beta = 1$, and leave the case of general coefficients as an exercise. As just observed, $f + g$ is integrable. To verify linearity is to show that

$$\int_E [f+g]^+\, dm - \int_E [f+g]^-\, dm = \left[\int_E f^+\, dm - \int_E f^-\, dm \right] + \left[\int_E g^+\, dm - \int_E g^-\, dm \right]. \tag{15}$$

But

$$(f+g)^+ - (f+g)^- = f+g = (f^+ - f^-) + (g^+ - g^-) \text{ on } E,$$

and therefore, since each of these six functions is real-valued,

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+ \text{ on } E.$$

It follows from linearity of integration for non-negative, measurable functions that

$$\int_E (f+g)^+ \, dm + \int_E f^- \, dm + \int_E g^- \, dm = \int_E (f+g)^- \, dm + \int_E f^+ \, dm + \int_E g^+ \, dm\,.$$

Since f, g, and $f+g$ are integrable over E, each of these six integrals is finite. Rearrange them to obtain (15). This completes the proof of linearity. □

The proof for finitely supported, simple functions that linearity of integration implies **monotonicity** holds for general integrable functions. The proof that linearity implies **additivity of domains** also holds in general, provided that one observes that if $f\colon E \to \mathbf{R}$ is integrable and $A \subseteq E$ is measurable, then $|f \cdot \chi_A| \le |f|$ on E and therefore, by the integral comparison test, the function $f \cdot \chi_A\colon E \to \mathbf{R}$ is integrable.

The Dominated Convergence Theorem *Let $\{f_n\colon E \to \mathbf{R}\}$ be a sequence of measurable functions. Assume that there is an integrable function $g\colon E \to [0, \infty]$ that dominates $\{f_n\}$ on E, in the sense that $|f_n| \le g$ on E for all n. If $\{f_n\} \to f$ pointwise on E, then $f\colon E \to \mathbf{R}$ is integrable and*

$$\lim_{n \to \infty} \int_E f_n \, dm = \int_E f \, dm.$$

Proof Since each $|f_n| \le g$ on E, $|f| \le g$ on E, and since g is integrable, by the integral comparison test, f and each f_n also are integrable. The function $g - f$ and for each n, the function $g - f_n$, are properly defined, non-negative and measurable. Moreover, the sequence $\{g - f_n\}$ converges pointwise on E to $g - f$. According to Fatou's Lemma,

$$\int_E (g-f) \, dm \le \liminf \int_E (g - f_n) \, dm.$$

Consequently, by the linearity of integration,

$$\int_E g \, dm - \int_E f \, dm = \int_E (g-f) \, dm \le \liminf \int_E (g-f_n) \, dm = \int_E g \, dm - \limsup \int_E f_n \, dm,$$

that is,

$$\limsup \int_E f_n \, dm \le \int_E f \, dm.$$

Similarly, considering the sequence $\{f_n + g\}$, we obtain

$$\int_E f \, dm \le \liminf \int_E f_n \, dm.$$

□

Example If $f\colon \mathbf{R} \to \overline{\mathbf{R}}$ is integrable, then $\lim_{n\to\infty} \int_{[-n,\,n]} f\, dm = \int_{\mathbf{R}} f\, dm$. To verify this, apply the Dominated Convergence Theorem to the sequence $\{f_n\colon \mathbf{R} \to \overline{\mathbf{R}}\}$, where, for each n, $f_n = f \cdot \chi_{[-n,\,n]}$.

The following generalization of the Dominated Convergence Theorem, the proof of which we leave as an exercise (see Problem 33), is often useful (see Problem 34).

Theorem 11 (**General Dominated Convergence Theorem**) *Let $\{f_n\colon E \to \mathbf{R}\}$ be a sequence of measurable functions that converges pointwise on E to f. Assume that there is a sequence $\{g_n\colon E \to [0,\, \infty]\}$ of integrable functions that converges pointwise on E to $g\colon E \to \mathbf{R}$ and dominates $\{f_n\}$ on E, in the sense that $|f_n| \le g_n$ on E for all n.*

$$\text{If } \lim_{n\to\infty} \int_E g_n\, dm = \int_E g\, dm < \infty, \text{ then } \lim_{n\to\infty} \int_E f_n\, dm = \int_E f\, dm.$$

For an integrable function $f\colon [a,\, b] \to \overline{\mathbf{R}}$, it is often convenient, recalling the notation in calculus, to define, since $m\{a\} = m\{b\} = 0$,

$$\int_a^b f\, dm \equiv \int_{[a,\,b]} f\, dm = \int_{(a,\,b)} f\, dm$$

and sometimes to use the Leibniz notation $\int_a^b f(x)\, dx$.

PROBLEMS

25. Prove the linearity of integration for general coefficients α and β.
26. Provide an example of a function $f\colon [0,\, 1] \to \mathbf{R}$ that is not integrable, while $|f|\colon [0,\, 1] \to \mathbf{R}$ is integrable. (Suggestion: Consider measurability.)
27. Provide an example of an integrable function $f\colon [0,\, 1] \to [0,\, \infty]$ for which $f^2 frm[o]--\colon [0,\, 1] \to [0,\, \infty]$ is not integrable.
28. Let $f\colon E \to \mathbf{R}$ and $g\colon E \to \mathbf{R}$ be measurable and $f^2\colon E \to \mathbf{R}$ and $g^2\colon E \to \mathbf{R}$ be integrable. Show that $f \cdot g\colon E \to \mathbf{R}$ is integrable. (Suggestion: $[a - b]^2 \ge 0$.)
29. Let $f\colon E \to \mathbf{R}$ be integrable and $g\colon E \to \mathbf{R}$ be bounded and measurable. Show that $f \cdot g\colon E \to \mathbf{R}$ is integrable.
30. Let $f\colon E \to \mathbf{R}$ be integrable and C a measurable subset of E. Show that the restriction of f to C is integrable and $\int_C f|_C\, dm = \int_E f \cdot \chi_C\, dm$.
31. For a measurable function $f\colon [1,\, \infty) \to \mathbf{R}$ that is bounded on bounded sets, for each n, define $a_n = \int_{[n,\, n+1)} f$. Is it true that f is integrable if and only if the series $\sum_{n=1}^{\infty} a_n$ converges? Is it true that f is integrable if and only if the series $\sum_{n=1}^{\infty} a_n$ converges absolutely?
32. Let $f\colon \mathbf{R} \to \mathbf{R}$ be integrable. Show that there is a sequence $\{\psi_n\colon \mathbf{R} \to \mathbf{R}\}$ of finitely supported, simple functions that converges pointwise to f and
$$\lim_{n\to\infty} \int_{\mathbf{R}} \psi_n\, dm = \int_{\mathbf{R}} f\, dm.$$
33. Prove the General Dominated Convergence Theorem by following the proof of the Dominated Convergence Theorem, but replacing the sequences $\{g - f_n\}$ and $\{g + f_n\}$, respectively, by $\{g_n - f_n\}$ and $\{g_n + f_n\}$.

34. Let $\{f_n: E \to \mathbf{R}\}$ be a sequence of integrable functions for which $f_n \to f$ pointwise on E and f is integrable over E. Show that $\int_E |f - f_n|\, dm \to 0$ if and only if $\lim_{n\to\infty} \int_E |f_n|\, dm = \int_E |f|\, dm$. (Suggestion: Use the General Dominated Convergence Theorem.)

35. Let f be a real-valued function of two variables (x, y) that is defined on the square $Q = \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1\}$ and is a measurable function of x for each fixed value of y. Suppose for each fixed value of x, $\lim_{y\to 0} f(x, y) = f(x)$ and that for all y, we have $|f(x, y)| \leq g(x)$, where g is integrable over $[0, 1]$. Show that

$$\lim_{y\to 0} \int_0^1 f(x, y)dx = \int_0^1 f(x)dx.$$

Also show that if the function $f(x, y)$ is continuous in y for each x, then

$$h(y) = \int_0^1 f(x, y)dx$$

is a continuous function of y.

36. Let f be a real-valued function of two variables (x, y) that is defined on the square $Q = \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1\}$ and is a measurable function of x for each fixed value of y. For each $(x, y) \in Q$, let the partial derivative $\partial f/\partial y$ exist. Suppose there is a function g that is integrable over $[0, 1]$ and such that

$$\left|\frac{\partial f}{\partial y}(x, y)\right| \leq g(x) \text{ for all } (x, y) \in Q.$$

Prove that

$$\frac{d}{dy}\left[\int_0^1 f(x, y)\, dx\right] = \int_0^1 \frac{\partial f}{\partial y}(x, y)\, dx \text{ for all } y \in [0, 1].$$

4.5 COUNTABLE ADDITIVITY AND CONTINUITY OF INTEGRATION

The linearity and monotonicity properties of Lebesgue integration are extensions of familiar properties of the Riemann integral. In this section, we establish two properties of the Lebesgue integral for which there is no corresponding property in Riemann integration. The following countable additivity property for Lebesgue integration is a companion of the countable additivity property for Lebesgue measure.

Theorem 12 (**the Countable Additivity of Integration**) *If $f: E \to \mathbf{R}$ is an integrable function and $\{E_n\}_{n=1}^{\infty}$ is a measurable partition of E, then*

$$\int_E f\, dm = \sum_{n=1}^{\infty} \int_{E_n} f\, dm. \tag{16}$$

Proof For each n, define $f_n = f \cdot \chi_n$ where χ_n is the characteristic function of the measurable set $\bigcup_{k=1}^{n} E_k$. Then f_n is a measurable function on E and

$$|f_n| \leq |f| \text{ on } E.$$

Observe that $\{f_n\} \to f$ pointwise on E. By the Dominated Convergence Theorem,

$$\int_E f\,dm = \lim_{n\to\infty} \int_E f_n\,dm.$$

On the other hand, since $\{E_n\}_{n=1}^{\infty}$ is disjoint, it follows from the additivity over domains property of the integral that for each n,

$$\int_E f_n\,dm = \sum_{k=1}^{n} \int_{E_k} f\,dm.$$

Therefore,

$$\int_E f\,dm = \lim_{n\to\infty} \int_E f_n\,dm = \lim_{n\to\infty} \left[\sum_{k=1}^{n} \int_{E_k} f\,dm \right] = \sum_{n=1}^{\infty} \int_{E_n} f\,dm. \qquad \square$$

We leave it as an exercise to use the countable additivity of integration to prove the following result regarding the continuity of integration, using as a pattern, the proof of continuity of measure as a consequence of the countable additivity of measure.

Theorem 13 (**the Continuity of Integration**) *Let $f\colon E \to \mathbf{R}$ be integrable.*

(i) If $\{E_n\}_{n=1}^{\infty}$ is an ascending collection of measurable subsets of E, then

$$\int_{\bigcup_{n=1}^{\infty} E_n} f\,dm = \lim_{n\to\infty} \int_{E_n} f\,dm.$$

(ii) If $\{E_n\}_{n=1}^{\infty}$ is a descending collection of measurable subsets of E, then

$$\int_{\bigcap_{n=1}^{\infty} E_n} f\,dm = \lim_{n\to\infty} \int_{E_n} f\,dm.$$

PROBLEMS

37. Prove the theorem regarding the continuity of integration.

38. Let $f\colon [0, \infty] \to \mathbf{R}$ be integrable. Show that

$$\lim_{n\to\infty} \int_{1/n}^{n} f\,dm = \int_0^{\infty} f\,dm.$$

39. Let $f\colon \mathbf{R} \to \mathbf{R}$ be integrable. Show that the function $F\colon \mathbf{R} \to \mathbf{R}$ defined by

$$F(x) = \int_{-\infty}^{x} f\,dm \text{ for all } x \in \mathbf{R}$$

is properly defined and continuous. Is it Lipschitz?

40. For each of the two functions f on $[1, \infty)$ defined below, show that $\lim_{n\to\infty} \int_{[1,\,n]} f\,dm$ exists while f is not integrable over $[1, \infty)$. Does this contradict the continuity of integration?
    (i) Define $f(x) = (-1)^k/k$, for $k \le x < k+1$.
    (ii) Define $f(x) = (\sin x)/x$ for $1 \le x < \infty$.

41. Let f: $\mathbf{R} \to \mathbf{R}$ be integrable. Show that the following four assertions are equivalent:
    (i) $f = 0$ almost everywhere on $\mathbf{R}$.
    (ii) $\int_{\mathbf{R}} fg\, dm = 0$ for every bounded, finitely supported, measurable function g.
    (iii) $\int_A f\, dm = 0$ for every measurable set A.
    (iv) $\int_{\mathcal{O}} f\, dm = 0$ for every open set $\mathcal{O}$.
42. Let f: $[a, b] \to \mathbf{R}$ be integrable and $\epsilon > 0$. Establish the following approximations.
    (i) There is a simple function η: $[a, b] \to \mathbf{R}$ for which $\int_{[a, b]} |f - \eta|\, dm < \epsilon$.
    (ii) There is a step-function s: $[a, b] \to \mathbf{R}$ for which $\int_{[a, b]} |f - s|\, dm < \epsilon$. (Suggestion: Consider Theorem 19 of Chapter 2.)
    (iii) There is a continuous function g: $[a, b] \to \mathbf{R}$ for which $\int_{[a, b]} |f - g|\, dm < \epsilon$.

C H A P T E R 5

Lebesgue Integration: Further Topics

Contents

5.1	Uniform Integrability and Tightness: The Vitali Convergence Theorems	81
5.2	Convergence in the Mean and in Measure: A Theorem of Riesz	85
5.3	Characterizations of Riemann and Lebesgue Integrability	88

5.1 UNIFORM INTEGRABILITY AND TIGHTNESS: THE VITALI CONVERGENCE THEOREMS

Lemma 1 *If $m(E) < \infty$ and $\delta > 0$, then there is a measurable partition $\{E_k\}_{k=1}^n$ of E which, for each k, has $m(E_k) < \delta$.*

Proof By the continuity of measure, $\lim_{n\to\infty} m(E \sim [-n, n]) = m(\emptyset) = 0$. Choose an n_0 for which $m(E \sim [-n_0, n_0]) < \delta$. By choosing a fine enough partition of $[-n_0, n_0]$, express $E \cap [-n_0, n_0]$ as the disjoint union of a finite collection of sets, each of measure less than δ. $\square$

Proposition 2 *If $f\colon E \to \mathbf{R}$ is an integrable function, then for each $\epsilon > 0$, there is a $\delta > 0$ for which*

$$\text{if } A \subseteq E \text{ is measurable and } m(A) < \delta, \text{ then } \int_A |f|\, dm < \epsilon. \tag{1}$$

Moreover, if $m(E) < \infty$ and f is measurable, and for each $\epsilon > 0$ there is a $\delta > 0$ for which (1) holds, then f is integrable.

Proof The theorem follows by establishing it separately for the positive and negative parts of f. We therefore suppose that $f \geq 0$ on E. Let $\epsilon > 0$. By the additivity over domains of integration and Chebychev's Inequality, if $A \subseteq E$ is measurable and $c > 0$, then

$$\int_A f\, dm = \int_{\{x\in A \,|\, f(x)<c\}} f\, dm + \int_{\{x\in A \,|\, f(x)\geq c\}} f\, dm \leq c \cdot m(A) + \frac{1}{c}\int_E f\, dm.$$

Choose $c > 0$ such that $1/c \cdot \int_E f\, dm < \epsilon/2$. Then

$$\int_A f\, dm < c \cdot m(A) + \epsilon/2.$$

Define $\delta = \epsilon/2c$. Then (1) holds for this choice of δ.

Now assume that $m(E) < \infty$ and f is measurable, and that for each $\epsilon > 0$, there is a $\delta > 0$ for which (1) holds. Let $\delta_0 > 0$ respond to the $\epsilon = 1$ challenge. Since $m(E) < \infty$, according to the preceding lemma, there is a measurable partition $\{E_k\}_{k=1}^N$ of E which, for each k, has $m(E_k) < \delta$. Therefore, by the additivity over domains of integration,

$$\int_E f\,dm = \sum_{k=1}^N \int_{E_k} f\,dm < N.$$

□

Definition *A collection $\mathcal{F}$ of integrable functions on E is said to be* **uniformly integrable** *provided that for each $\epsilon > 0$, there is a $\delta > 0$ such that for each $f \in \mathcal{F}$,*

$$\textit{if } A \subseteq E \textit{ is measurable and } m(A) < \delta, \textit{ then } \int_A |f|\,dm < \epsilon. \tag{2}$$

Example Let $g\colon E \to [0, \infty]$ be integrable. Define

$$\mathcal{F} = \{f \mid f\colon E \to \mathbf{R} \text{ is measurable and } |f| \leq g \text{ on } E\}.$$

Then $\mathcal{F}$ is uniformly integrable. This follows from Proposition 2, with f replaced by g, and the observation that for any measurable subset A of E, by the monotonicity of integration, if f belongs to $\mathcal{F}$, then

$$\int_A |f|\,dm \leq \int_A g\,dm.$$

We leave the proof of the following consequence of Proposition 2 as an exercise.

Proposition 3 *A finite collection $\{f_k\colon E \to \mathbf{R}\}_{k=1}^n$ of integrable functions is uniformly integrable.*

Proposition 4 *Assume that $m(E) < \infty$. If the sequence of functions $\{f_n\colon E \to \mathbf{R}\}$ is uniformly integrable and $\{f_n\} \to f$ pointwise on E, then f is integrable.*

Proof Let $\delta_0 > 0$ respond to the $\epsilon = 1$ challenge regarding the uniform integrability criteria for the sequence $\{f_n\}$. Since $m(E) < \infty$, by Lemma 1, there is a measurable partition $\{E_k\}_{k=1}^N$ of E which, for each k, has $m(E_k) < \delta_0$. It follows from the additivity over domains of integration that, for each n,

$$\int_E |f_n|\,dm = \sum_{k=1}^N \int_{E_k} |f_n|\,dm < N.$$

According to Fatou's Lemma,

$$\int_E |f|\,dm \leq \liminf \int_E |f_n|\,dm \leq N.$$

Therefore, $|f|$ is integrable. □

The Vitali Convergence Theorem *Assume that* $m(E) < \infty$. *If the sequence of functions* $\{f_n\colon E \to \mathbf{R}\}$ *is uniformly integrable and* $\{f_n\} \to f$ *pointwise on* E, *then* f *is integrable and*

$$\lim_{n \to \infty} \int_E f_n \, dm = \int_E f \, dm.$$

Proof According to the preceding proposition, f is integrable. Therefore, by Proposition 8 of the preceding chapter, we may assume that f is real-valued. It follows from the integral comparison test and the linearity of integration that, for any measurable subset A of E and any n,

$$\begin{aligned} \left| \int_E f_n \, dm - \int_E f \, dm \right| &= \left| \int_E (f_n - f) \, dm \right| \\ &\leq \int_E |f_n - f| \, dm \\ &= \int_{E \sim A} |f_n - f| \, dm + \int_A |f_n - f| \, dm \\ &\leq \int_{E \sim A} |f_n - f| \, dm + \int_A |f_n| \, dm + \int_A |f| \, dm. \end{aligned} \tag{3}$$

Let $\epsilon > 0$. By the uniform integrability of $\{f_n\}$, there is a $\delta > 0$ such that for any n, $\int_A |f_n| \, dm < \epsilon/3$ for any measurable subset A of E for which $m(A) < \delta$. Therefore, by Fatou's Lemma, we also have $\int_A |f| \, dm \leq \epsilon/3$ for any measurable subset of A of E for which $m(A) < \delta$. Since f is real-valued and $m(E) < \infty$, according to Egoroff's Theorem, there is a measurable subset E_0 of E for which $m(E_0) < \delta$ and $\{f_n\} \to f$ uniformly on $E \sim E_0$. Choose an index N for which $|f_n - f| < \epsilon/[3 \cdot m(E)]$ on $E \sim E_0$ for all $n \geq N$. Take $A = E_0$ in the integral inequality (3), so that if $n \geq N$, then

$$\begin{aligned} \left| \int_E f_n \, dm - \int_E f \, dm \right| &\leq \int_{E \sim E_0} |f_n - f| \, dm + \int_{E_0} |f_n| \, dm + \int_{E_0} |f| \, dm \\ &< \epsilon/[3 \cdot m(E)] \cdot m(E \sim E_0) + \epsilon/3 + \epsilon/3 < \epsilon. \end{aligned}$$

This completes the proof. □

Theorem 5 *Assume that* $m(E) < \infty$. *If* $\{f_n\colon E \to [0, \infty]\}$ *is a sequence of non-negative, integrable functions that converges pointwise on* E *to* $f \equiv 0$, *then*

$$\lim_{n \to \infty} \int_E f_n \, dm = 0 \text{ if and only if } \{f_n\colon E \to [0, \infty]\} \text{ is uniformly integrable.}$$

Proof If $\{f_n\}$ is uniformly integrable, then, according to the Vitali Convergence Theorem, $\lim_{n \to \infty} \int_E f_n \, dm = 0$. Conversely, suppose that $\lim_{n \to \infty} \int_E f_n \, dm = 0$. Let $\epsilon > 0$. There is an index N for which $\int_E f_n \, dm < \epsilon$ if $n \geq N$. Therefore, since each f_n is non-negative,

$$\text{if } A \subseteq E \text{ is measurable and } n \geq N, \text{then } \int_A f_n \, dm < \epsilon. \tag{4}$$

According to Proposition 3, the finite collection $\{f_n\}_{n=1}^{N-1}$ is uniformly integrable. Let δ respond to the ϵ challenge regarding the criterion for the uniform integrability of $\{f_n\}_{n=1}^{N-1}$. It follows from (4) that δ also responds to the ϵ challenge regarding the criterion for the uniform integrability of the sequence $\{f_n\}$. $\square$

The above Vitali Convergence Theorem does not extend to domains of infinite measure. Indeed, for each n, define $f_n\colon \mathbf{R} \to \mathbf{R}$ by $f_n = \chi_{[n,\, n+1]}$ and $f \equiv 0$ on $\mathbf{R}$. Then $\{f_n\}$ is uniformly integrable over $\mathbf{R}$ and converges pointwise on $\mathbf{R}$ to f. However,

$$\lim_{n \to \infty} \left[\int_{\mathbf{R}} f_n \, dm \right] = 1 \neq 0 = \int_{\mathbf{R}} \lim_{n \to \infty} f_n \, dm = \int_{\mathbf{R}} f \, dm.$$

The following property of functions that are integrable over a domain of infinite measure suggests an additional property that should accompany uniform integrability in order to extend the Vitali Convergence Theorem to such sets.

Proposition 6 *If $f\colon E \to \mathbf{R}$ is an integrable function, then for each $\epsilon > 0$, there is a measurable subset E_0 of E for which $m(E_0) < \infty$ and $\int_{E \sim E_0} |f| \, dm < \epsilon$.*

Proof Let $\epsilon > 0$. The non-negative, measurable function $|f|$ is integrable, and by the definition of the integral of such a function, there is a finitely supported, bounded, measurable function $h\colon E \to \mathbf{R}$ for which $0 \leq h \leq |f|$ on E and $\int_E |f| \, dm - \int_E h \, dm < \epsilon$. Choose a measurable subset E_0 of E for which $m(E_0) < \infty$ and $h = 0$ on $\mathbf{R} \sim E_0$. By the linearity of integration,

$$\int_{E \sim E_0} |f| \, dm = \int_{E \sim E_0} [|f| - h] \, dm \leq \int_E [|f| - h] \, dm < \epsilon. \qquad \square$$

Definition *A collection $\mathcal{F}$ of integrable functions on E is said to be* **tight** *provided that for each $\epsilon > 0$, there is a subset E_0 of E for which $m(E_0) < \infty$ and $\int_{E \sim E_0} |f| \, dm < \epsilon$ for all $f \in \mathcal{F}$.*

The General Vitali Convergence Theorem *If the sequence of functions $\{f_n\colon E \to \mathbf{R}\}$ is uniformly integrable and tight and $\{f_n\} \to f$ pointwise on E, then f is integrable and*

$$\lim_{n \to \infty} \int_E f_n \, dm = \int_E f \, dm.$$

Proof Let $\epsilon > 0$. By the tightness of the sequence $\{f_n\}$, there is a subset E_0 of E for which $m(E_0) < \infty$ and

$$\int_{E \sim E_0} |f_n| \, dm < \epsilon/4 \text{ for all } n.$$

It follows from Fatou's Lemma that $\int_{E \sim E_0} |f| \, dm \leq \epsilon/4$. In particular, f is integrable over $E \sim E_0$. Furthermore, by the integral comparison test and the linearity of integration,

$$\left| \int_{E \sim E_0} [f_n - f] \, dm \right| \leqslant \int_{E \sim E_0} |f_n| \, dm + \int_{E \sim E_0} |f| \, dm < \epsilon/2 \text{ for all } n.$$

But $m(E_0) < \infty$ and $\{f_n\}$ is uniformly integrable over E_0. Consequently, by the Vitali Convergence Theorem for functions on a domain of finite measure, f is integrable over E_0 and we may choose an index N for which if $n \geq N$, then

$$\left|\int_{E_0} [f_n - f]\, dm\right| < \epsilon/2.$$

Therefore, if $n \geq N$,

$$\left|\int_E [f_n - f]\, dm\right| \leq \left|\int_{E \sim E_0} [f_n - f]\, dm\right| + \left|\int_{E_0} [f_n - f]\, dm\right| < \epsilon.$$

The proof is complete. □

We leave the following extension of Theorem 5 as an exercise.

Theorem 7 *If $\{f_n\colon E \to [0, \infty]\}$ is a sequence of non-negative, integrable functions that converges pointwise on E to $f \equiv 0$, then*

$$\lim_{n \to \infty} \int_E f_n\, dm = 0 \text{ if and only if } \{f_n\colon E \to [0, \infty]\} \text{ is uniformly integrable and tight.}$$

PROBLEMS

1. Show that the Dominated Convergence Theorem follows from the Vitali Convergence Theorem.
2. Show that Theorem 5 is false without the assumption that the f_n's are non-negative.
3. Let $\mathcal{F}$ be a collection of integrable functions on E. Show that $\mathcal{F}$ is uniformly integrable over E if and only if for each $\epsilon > 0$, there is a $\delta > 0$ such that for each $f \in \mathcal{F}$,
$$\text{if } A \subseteq E \text{ is measurable and } m(A) < \delta, \text{ then } \left|\int_A f\, dm\right| < \epsilon.$$
4. Show that a finite collection of integrable functions is tight.
5. Prove Theorem 7.
6. Let the sequences of functions $\{f_n\colon E \to \mathbf{R}\}$ and $\{g_n\colon E \to \mathbf{R}\}$ be uniformly integrable and tight. Show that for any α and β, $\{\alpha f_n + \beta g_n\}$ also is uniformly integrable and tight.
7. Show that a sequence $\{f_n\colon E \to \mathbf{R}\}$ is uniformly integrable and tight if and only if for each $\epsilon > 0$, there is a measurable subset E_0 of E that has finite measure and a $\delta > 0$ such that for each measurable subset A of E and each n,
$$\text{if } m(A \cap E_0) < \delta, \text{ then } \int_A |f_n|\, dm < \epsilon.$$

5.2 CONVERGENCE IN THE MEAN AND IN MEASURE: A THEOREM OF RIESZ

Definition *A sequence $\{f_n\colon E \to \mathbf{R}\}$ of integrable functions is said to* **converge in the mean** *to the integrable function $f\colon E \to \mathbf{R}$ provided that*

$$\lim_{n \to \infty} \int_E |f_n - f|\, dm = 0.$$

Theorem 8 *If the sequence of integrable functions $\{f_n: E \to \mathbf{R}\}$ converges pointwise on E to the integrable function $f: E \to \mathbf{R}$, then*

$$\{f_n\} \to f \text{ in the mean if and only if } \{f_n\} \text{ is uniformly integrable and tight.}$$

Proof According to Propositions 2 and 6, respectively, the single integrable function f is uniformly integrable and tight. On the other hand, it follows from the triangle inequality that for all n,

$$|f_n| \le |f - f_n| + |f| \text{ and } |f - f_n| \le |f| + |f_n| \text{ on } E. \tag{5}$$

Consequently, $\{f_n\}$ is uniformly integrable and tight if and only if $\{|f - f_n|\}$ is uniformly integrable and tight. An appeal to Theorem 7 concludes the proof. □

Definition *For a sequence of measurable functions $\{f_n: E \to \mathbf{R}\}$ and a measurable function $f: E \to \mathbf{R}$, the sequence $\{f_n\}$ is said to* **converge in measure** *to f provided that for each $\eta > 0$,*

$$\lim_{n\to\infty} m\,\{x \in E \mid \ |f_n(x) - f(x)| > \eta\} = 0.$$

The following is just a rewording of Lemma 9 of Chapter 3, which preceded the proof of Egoroff's Theorem.

Proposition 9 *Assume that $m(E) < \infty$. If the sequence of measurable functions $\{f_n: E \to \mathbf{R}\}$ converges pointwise to $f: E \to \mathbf{R}$, then it converges in measure.*

Proposition 10 *If the sequence of integrable functions $\{f_n: E \to \mathbf{R}\}$ converges in the mean to the integrable function $f: E \to \mathbf{R}$, then it converges in measure.*

Proof According the Chebychev's Inequality, for each $\eta > 0$ and each n,

$$m\,\{x \in E \mid \ |f_n(x) - f(x)| > \eta\} \le 1/\eta \int_E |f_n - f|\, dm.$$

Therefore, $\{f_n\} \to f$ in measure. □

The following example shows that a sequence may converge in measure but not converge pointwise at any point.

Example Consider the sequence $\{I_n\}$ of open subintervals of $(0, 1)$ defined as follows: Define the first two terms of the sequence to be the two disjoint open subintervals of $(0, 1)$ of length $1/2$, indexed from left to right. Let the next three terms of the sequence be the three disjoint open subintervals of $(0, 1)$ of length $1/3$, indexed from left to right. Continuing in this way, a sequence $\{I_n\}$ of open subintervals of $(0, 1)$ is constructed for which each point in $(0, 1)$ belongs to infinitely many terms and fails to belong to infinitely many terms, and moreover $\lim_{n\to\infty} \ell(I_n) = 0$. For each n, define $f_n: (0, 1) \to \mathbf{R}$ by $f_n = \chi_{I_n}$. Then the sequence of integrable functions $\{f_n: (0, 1) \to \mathbf{R}\}$ converges in measure to $f \equiv 0$, but at no point x does $\lim_{n\to\infty} f_n(x) = 0$. Nevertheless, observe that for each k, the interval $(0, 1/k)$ is a term in the sequence $\{I_n\}$: denote its index by n_k. Then $\{f_{n_k}\}$ is a subsequence of $\{f_n\}$ that converges pointwise to $f \equiv 0$ on all of $(0, 1)$.

Theorem 11 (**Riesz**) *If a sequence of measurable functions $\{f_n\colon E \to \mathbf{R}\}$ converges in measure to the measurable function $f\colon E \to \mathbf{R}$, then there is a subsequence $\{f_{n_k}\}$ that converges pointwise almost everywhere on E to f.*

Proof By the definition of convergence in measure, there is a strictly increasing sequence of natural numbers $\{n_k\}$ for which

$$m\{x \in E \mid |f_j(x) - f(x)| > 1/k\} < 1/2^k \text{ for all } j \geq n_k.$$

In particular, if, for each k, $E_k = \{x \in E \mid |f_{n_k}(x) - f(x)| > 1/k\}$, then $m(E_k) < 1/2^k$. According to the Borel-Cantelli Lemma, for almost all $x \in E$, there is an index $K(x)$ such that $x \notin E_k$ if $k \geq K(x)$, that is,

$$|f_{n_k}(x) - f(x)| \leq 1/k \text{ for all } k \geq K(x).$$

Therefore, for almost all $x \in E$, $\lim_{k\to\infty} f_{n_k}(x) = f(x)$. □

The following corollary is a consequence of Proposition 10 and the preceding theorem.

.

Corollary 12 *If the sequence of integrable functions $\{f_n\colon E \to \mathbf{R}\}$ converges in the mean to the integrable function $f\colon E \to \mathbf{R}$, then there is a subsequence $\{f_{n_k}\}$ that converges pointwise almost everywhere on E to f.*

PROBLEMS

8. In the sequence $\{I_n\}$ of intervals constructed in the example after Proposition 10, for $k \geq 2$, what is the index n_k of the interval $(0, 1/k)$?
9. Let $\{f_n\colon E \to \mathbf{R}\}$ be a sequence of measurable functions and $f\colon E \to \mathbf{R}$ be measurable. For $p > 0$, assume that $\lim_{n\to\infty} \int_E |f_n - f|^p \, dm = 0$. Show that a subsequence of $\{f_n\}$ converges pointwise almost everywhere on E to f.
10. Show that Fatou's Lemma, the Monotone Convergence Theorem, the Dominated Convergence Theorem, and the Vitali Convergence Theorem hold if the assumption of pointwise convergence is replaced by the assumption of convergence in measure.
11. Show that Proposition 9 does not hold if $m(E) = \infty$.
12. Assume that $m(E) < \infty$. Let $\{f_n\colon E \to \mathbf{R}\}$ be a sequence of measurable functions and $f\colon E \to \mathbf{R}$ be measurable. Prove that $\{f_n\} \to f$ in measure if and only if every subsequence has in turn a further subsequence that converges to f pointwise almost everywhere on E.
13. Assume that $m(E) < \infty$. Show that linear combinations of sequences that converge in measure also converge in measure. Show that products of sequences that converge in measure also converge in measure. (Suggestion: See the preceding problem.)
14. Assume that $m(E) < \infty$. For two measurable functions $f\colon E \to \mathbf{R}$ and $g\colon E \to \mathbf{R}$, define

$$\rho(f, g) = \int_E \frac{|f - g|}{1 + |f - g|} \, dm.$$

Show that $\{f_n\colon E \to \mathbf{R}\} \to f$ in measure if and only if $\lim_{n\to\infty} \rho(f_n, f) = 0$.

15. Assume that $m(E) < \infty$. A sequence $\{f_n\colon E \to \mathbf{R}\}$ of measurable functions is said to be Cauchy in measure provided that for each $\eta > 0$ and $\epsilon > 0$, there is an index N such that for all $m, n \geq N$,
$$m\{x \in E \mid |f_n(x) - f_m(x)|\, dm \geq \eta\} < \epsilon.$$
Show that if $\{f_n\}$ is Cauchy in measure, then there is a measurable function $f\colon E \to \mathbf{R}$ to which the sequence $\{f_n\}$ converges in measure. (Suggestion: Choose a strictly increasing sequence of indices $\{n_j\}$ such that for each j, $m\{x \in E \mid |f_{n_{j+1}}(x) - f_{n_j}(x)| > 1/2^j\} < 1/2^j$.)

5.3 CHARACTERIZATIONS OF RIEMANN AND LEBESGUE INTEGRABILITY

Lemma 13 *Let $\{\varphi_n\colon E \to \mathbf{R}\}$ and $\{\psi_n\colon E \to \mathbf{R}\}$ be sequences of Lebesgue integrable functions for which $\{\varphi_n\}$ is increasing while $\{\psi_n\}$ is decreasing. Let $f\colon E \to \mathbf{R}$ be a function for which*
$$\varphi_n \leq f \leq \psi_n \text{ on } E \text{ for all } n.$$
If
$$\lim_{n \to \infty} \int_E [\psi_n - \varphi_n]\, dm = 0,$$
then
$$\{\varphi_n\} \to f \text{ and } \{\psi_n\} \to f \text{ pointwise almost everywhere on } E, f\colon E \to \mathbf{R} \text{ is integrable},$$
$$\lim_{n \to \infty} \int_E \varphi_n\, dm = \int_E f\, dm \text{ and } \lim_{n \to \infty} \int_E \psi_n\, dm = \int_E f\, dm.$$

Proof For $x \in E$, define
$$\varphi^*(x) = \lim_{n \to \infty} \varphi_n(x) \text{ and } \psi^*(x) = \lim_{n \to \infty} \psi_n(x).$$
Since the sequences of functions $\{\varphi_n\}$ and $\{\psi_n\}$ are monotone, the functions φ^* and ψ^* are properly defined, and they are measurable since each is the pointwise limit of a sequence of measurable functions. We have the inequalities
$$\varphi_n \leq \varphi^* \leq f \leq \psi^* \leq \psi_n \text{ on } E \text{ for all } n. \tag{6}$$
By the linearity of integration,
$$0 \leq \int_E (\psi^* - \varphi^*)\, dm \leq \int_E (\psi_n - \varphi_n)\, dm \text{ for all } n,$$
so that
$$0 \leq \int_E (\psi^* - \varphi^*)\, dm \leq \lim_{n \to \infty} \int_E (\psi_n - \varphi_n)\, dm = 0.$$
Since $\psi^* - \varphi^*$ is a non-negative, measurable function and $\int_E (\psi^* - \varphi^*)\, dm = 0$, according to Proposition 25 of Chapter 4, $\psi^* = \varphi^*$ almost everywhere on E. But $\varphi^* \leq f \leq \psi^*$ on E. Consequently,
$$\{\varphi_n\} \to f \text{ and } \{\psi_n\} \to f \text{ pointwise almost everywhere on } E.$$
Therefore, $f\colon E \to \mathbf{R}$ is measurable. Observe that since $0 \leq f - \varphi_1 \leq \psi_1 - \varphi_1$ on E and ψ_1 and φ_1 are integrable, it follows from the integral comparison test that $f\colon E \to \mathbf{R}$ is integrable. We deduce from inequality (6) that for all n,
$$0 \leq \int_E \psi_n\, dm - \int_E f\, dm = \int_E (\psi_n - f)\, dm \leq \int_E (\psi_n - \varphi_n)\, dm$$

and

$$0 \leq \int_E f\,dm - \int_E \varphi_n\,dm = \int_E (f - \varphi_n)\,dm \leq \int_E (\psi_n - \varphi_n)\,dm.$$

Consequently,

$$\lim_{n \to \infty} \int_E \varphi_n\,dm = \int_E f\,dm = \lim_{n \to \infty} \int_E \psi_n\,dm.$$

$\square$

Theorem 14 *Assume that $m(E) < \infty$. A bounded function $f\colon E \to \mathbf{R}$ is Lebesgue integrable if and only if it is measurable.*

Proof In the preceding chapter, we proved that, as a consequence of the Simple Approximation Lemma, a bounded, measurable function on a set of finite measure is Lebesgue integrable. It remains to prove the converse. Suppose that f is integrable. Directly from the definition of integrability, it follows that there are sequences of simple functions $\{\varphi_n\}$ and $\{\psi_n\}$ for which

$$\varphi_n \leq f \leq \psi_n \text{ on } E \text{ for all } n,$$

and

$$\lim_{n \to \infty} \int_E [\psi_n - \varphi_n]\,dm = 0.$$

By taking obvious maxima and minima, we may suppose that $\{\varphi_n\}$ is increasing and $\{\psi_n\}$ is decreasing. According to the preceding lemma, $\{\varphi_n\} \to f$ pointwise almost everywhere on E. Therefore f, being the pointwise limit almost everywhere of a sequence of measurable functions, is measurable. $\square$

Theorem 15 *A bounded function $f\colon [a, b] \to \mathbf{R}$ is Riemann integrable if and only if the set of points in $[a, b]$ at which it fails to be continuous has measure zero.*

Proof We first suppose that f is Riemann integrable. It follows directly from the definition that there are sequences of partitions $\{P_n\}$ and $\{P'_n\}$ of $[a, b]$ for which

$$\lim_{n \to \infty} [U(f, P_n) - L(f, P'_n)] = 0,$$

where $U(f, P_n)$ and $L(f, P'_n)$ are upper and lower Darboux sums. Since, under refinement, lower Darboux sums increase and upper Darboux sums decrease, by possibly replacing each P_n by a common refinement of the $2n$ partitions $P_1, \ldots, P_n$, $P'_1, \ldots, P'_n$, we may assume that each P_{n+1} is a refinement of P_n and $P_n = P'_n$. For each index n, define φ_n to be the lower step-function associated with f with respect to P_n, that is, which agrees with f at the partition points of P_n and which on each open interval determined by P_n has constant value equal to the infimum of f on that interval. We define the upper step-function ψ_n in a similar manner. By definition of the Darboux sums,

$$L(f, P_n) = \int_a^b \varphi_n\,dm \text{ and } U(f, P_n) = \int_a^b \psi_n\,dm \text{ for all } n.$$

Then $\{\varphi_n\}$ and $\{\psi_n\}$ are sequences of integrable functions such that for each index n, $\varphi_n \leq f \leq \psi_n$ on E. Moreover, the sequence $\{\varphi_n\}$ is increasing and $\{\psi_n\}$ is decreasing, because each P_{n+1} is a refinement of P_n. Finally,

$$\lim_{n \to \infty} \int_a^b [\psi_n - \varphi_n] \, dm = \lim_{n \to \infty} [U(f, P_n) - L(f, P_n)] = 0.$$

According to the preceding lemma, with the Lebesgue integral replaced by the Riemann integral,

$$\{\varphi_n\} \to f \text{ and } \{\psi_n\} \to f \text{ pointwise almost everywhere on } [a, b].$$

The set E of points x at which either $\{\psi_n(x)\}$ or $\{\varphi_n(x)\}$ fail to converge to $f(x)$ has measure 0. Let E_0 be the union of E and the set of all the partition points in the P_n's. As the union of a set of measure zero and a countable set, $m(E_0) = 0$. We claim that f is continuous at each point in $E \sim E_0$. Indeed, let x_0 belong to $E \sim E_0$. To show that f is continuous at x_0, let $\epsilon > 0$. Since $\{\psi_n(x_0)\}$ and $\{\varphi_n(x_0)\}$ converge to $f(x_0)$, we may choose an n_0 for which

$$f(x_0) - \epsilon < \varphi_{n_0}(x_0) \leq f(x_0) \leq \psi_{n_0}(x_0) < f(x_0) + \epsilon. \tag{7}$$

Since x_0 is not a partition point of P_{n_0}, we may choose $\delta > 0$ such that the open interval $(x_0 - \delta, x_0 + \delta)$ is contained in the open interval I_{n_0} determined by P_{n_0} that contains x_0. This containment implies that

$$\text{if } |x - x_0| < \delta, \text{ then } \varphi_{n_0}(x_0) \leq \varphi_{n_0}(x) \leq f(x) \leq \psi_{n_0}(x) \leq \psi_{n_0}(x).$$

From this inequality and inequality (7), it follows that

$$\text{if } |x - x_0| < \delta, \text{ then } |f(x) - f(x_0)| < \epsilon.$$

Therefore, f is continuous at x_0.

It remains to prove the converse. Assume that f is continuous at almost all points in $[a, b]$. Let $\{P_n\}$ be any sequence of partitions of $[a, b]$ for which[1]

$$\lim_{n \to \infty} \text{gap } P_n = 0.$$

To show that $f\colon [a, b] \to \mathbf{R}$ is Riemann integrable it suffices to show that

$$\lim_{n \to \infty} [U(f, P_n) - L(f, P_n)] = 0. \tag{8}$$

For each n, let φ_n and ψ_n be the lower and upper step-functions associated with f over the partition P_n. To prove (8) is to prove that

$$\lim_{n \to \infty} \int_a^b [\psi_n - \varphi_n] \, dm = 0. \tag{9}$$

[1]The gap of a partition P is defined to be the maximum distance between consecutive points of the partition.

The Riemann integral of a step-function equals its Lebesgue integral. Moreover, since the function f is bounded on the bounded set $[a, b]$, the sequences $\{\varphi_n\}$ and $\{\psi_n\}$ are uniformly pointwise bounded on $[a, b]$. Hence, by the Bounded Convergence Theorem, to verify (9) it suffices to show that $\{\varphi_n\} \to f$ and $\{\psi_n\} \to f$ pointwise on the set of points in (a, b) at which f is continuous and which are not partition points of any partition P_n. Let x_0 be such a point. We show that

$$\lim_{n\to\infty} \varphi_n(x_0) = f(x_0) \text{ and } \lim_{n\to\infty} \psi_n(x_0) = f(x_0). \tag{10}$$

Let $\epsilon > 0$. Let $\delta > 0$ be such that

$$f(x_0) - \epsilon/2 < f(x) < f(x_0) + \epsilon/2 \text{ if } |x - x_0| < \delta. \tag{11}$$

Choose an index N for which gap $P_n < \delta$ if $n \geq N$. If $n \geq N$ and I_n is the open partition interval determined by P_n, that contains x_0, then $I_n \subseteq (x_0 - \delta, x_0 + \delta)$. It follows from (11) that

$$f(x_0) - \epsilon/2 \leq \varphi_n(x_0) < f(x_0) < \psi_n(x_0) \leq f(x_0) + \epsilon/2$$

and so

$$0 \leq \psi_n(x_0) - f(x_0) < \epsilon \text{ and } 0 \leq f(x_0) - \varphi_n(x_0) < \epsilon \text{ for all } n \geq N.$$

Therefore, (10) holds and the proof is complete. □

PROBLEMS

16. Prove that the product of Riemann integrable functions on a closed, bounded interval is Riemann integrable, and the product of bounded Lebesgue integrable functions on a set of finite measure is Lebesgue integrable.
17. Let E be measurable and the set of points at which $f \colon E \to \mathbf{R}$ fails to be continuous have measure zero. Show that f is measurable.

CHAPTER 6

Differentiation and Integration

Contents

6.1	Continuity of Monotone Functions	92
6.2	Differentiability of Monotone Functions: Lebesgue's Theorem	94
6.3	Functions of Bounded Variation: Jordan's Theorem	99
6.4	Absolutely Continuous Functions	103
6.5	Integrating Derivatives: Differentiating Indefinite Integrals	107
6.6	Measurability: Images of Sets, Compositions of Functions	114
6.7	Convex Functions	118

We consider the fundamental theorems of integral and differential calculus, in the context of the Lebesgue integral, and the application of these theorems in making a change of variables in integration and, characterizing those strictly increasing, continuous functions f whose composition with g, $g \circ f$, are measurable, for any measurable function g. To do so, we introduce two new collections of functions, functions of bounded variation and a subset of these, those that are absolutely continuous. A centerpiece is the understanding of properties of a continuous function $f\colon [a, b] \to \mathbf{R}$ which imply that, as a Lebesgue integral,

$$\int_a^b f'\,dm = f(b) - f(a). \tag{*}$$

Extend f continuously to $[a, b+1]$. A change of variables by translation, and cancellation, provides the following discrete formulation of (*) for the Riemann integral:

$$\int_a^b \frac{f(x+h)-f(x)}{h}\,dx = \frac{1}{h}\int_b^{b+h} f(x)\,dx - \frac{1}{h}\int_a^{a+h} f(x)\,dx \text{ for all } 0 < h \leq 1.$$

The limit of the right-hand side as $h \to 0^+$ equals $f(b) - f(a)$. We prove a striking theorem of Lebesgue according to which an increasing function on (a, b) has a finite derivative almost everywhere that is integrable. We then define what it means for a function $f\colon [a, b] \to \mathbf{R}$ to be absolutely continuous and prove that such a function is the difference of increasing, absolutely continuous functions and, furthermore, the collection of divided difference functions, $\{\text{Diff}_h f\}_{0<h\leq 1}$, where $\text{Diff}_h f(x) = [f(x+h) - f(x)]/h$, is uniformly integrable. Consequently, if f is absolutely continuous, an appeal to Lebesgue's theorem and the Vitali Convergence Theorem permits passage of the limit under the integral sign as $h \to 0^+$ in the left-hand side of the elementary, discrete form of (*) to obtain (*) itself.

6.1 CONTINUITY OF MONOTONE FUNCTIONS

A function $f\colon E \to R$ is said to be increasing provided that $f(u) \leq f(v)$ whenever $u \leq v$, and decreasing is defined similarly. A function is said to be **monotone** if it is either increasing or decreasing.

Theorem 1 *A monotone function $f\colon I \to \mathbf{R}$ on an interval I is continuous except possibly at a countable number of points.*

Proof Assume that f is increasing. It suffices to prove this for $I = [a, b]$, a closed, bounded interval, and we need only consider continuity at points x_0 in (a, b). At such an x_0, f has a limit from the left and from the right. Define

$$\begin{aligned} f(x_0^-) &= \lim_{x \to x_0^-} f(x) = \sup\{f(x) \mid a < x < x_0\}, \\ f(x_0^+) &= \lim_{x \to x_0^+} f(x) = \inf\{f(x) \mid x_0 < x < b\}. \end{aligned}$$

Since f is increasing, $f(x_0^-) \leq f(x_0^+)$. The function f fails to be continuous at x_0 if and only if $f(x_0^-) < f(x_0^+)$, in which case we define the open "jump" interval $J(x_0)$ by

$$J(x_0) = \{y \mid f(x_0^-) < y < f(x_0^+)\}.$$

Each jump interval is contained in the bounded interval $[f(a), f(b)]$ and the collection of jump intervals is disjoint. Therefore, for each n, there are only a finite number of jump intervals of length greater than $1/n$. Consequently, the set of points of discontinuity of f is the union of a countable collection of finite sets and therefore is countable. □

Proposition 2 *If C is a countable subset of the open interval (a, b), then there is an increasing function on (a, b) that is continuous only at points in $(a, b) \sim C$.*

Proof If C is finite, the proof is clear. Assume that C is countably infinite, and choose an enumeration $\{q_n\}_{n=1}^{\infty}$ of C. Define the function f on (a, b) by setting[1]

$$f(x) = \sum_{\{n \mid q_n \leq x\}} \frac{1}{2^n} \text{ for all } a < x < b.$$

Since $\sum_{n=1}^{\infty} 1/2^n$ converges, f is properly defined. Moreover,

$$\text{if } a < u < v < b, \text{ then } f(v) - f(u) = \sum_{\{n \mid u < q_n \leq v\}} \frac{1}{2^n}, \tag{1}$$

and so f is increasing. Let $x_0 = q_k \in C$. Then, by (1),

$$f(x_0) - f(x) \geq \frac{1}{2^k} \text{ for all } x < x_0.$$

Therefore, f fails to be continuous at x_0. Now let $x_0 \in (a, b) \sim C$. Fix an index n. There is an open interval I containing x_0 for which $q_k \notin I$ for $1 \leq k \leq n$. We deduce from (1) that $|f(x) - f(x_0)| < 1/2^n$ for all $x \in I$. Consequently, f is continuous at x_0. □

Since a monotone function on a closed, bounded interval is continuous on the complement of a set of measure zero, and is bounded above and below by a constant, we have the following criterion for integrability:

Corollary 3 *A monotone function $f\colon [a, b] \to \mathbf{R}$ is integrable.*

[1] We use the convention that a sum over the empty-set is zero.

PROBLEMS

1. Show that there is a strictly increasing function f: $[0, 1] \to \mathbf{R}$ that is continuous only at the irrational numbers in $[0, 1]$.
2. Verify the assertion in the proof of the theorem that it suffices to only consider the case that I is closed and bounded.
3. Let E be any subset of $\mathbf{R}$ and C a countable subset of E. Is there a monotone function on E that is continuous only at points in $E \sim C$?

6.2 DIFFERENTIABILITY OF MONOTONE FUNCTIONS: LEBESGUE'S THEOREM

A closed, bounded interval $[c, d]$ is said to be non-degenerate provided that $c < d$.

Definition *A collection $\mathcal{F}$ of closed, bounded, non-degenerate intervals is said to cover a set E* **in the sense of Vitali** *provided that for each point x in E and $\epsilon > 0$, there is an interval I in $\mathcal{F}$ that contains x and has $\ell(I) < \epsilon$.*

The Vitali Covering Lemma *Assume that $m^*(E) < \infty$ and $\mathcal{F}$ is a collection of closed, bounded, non-degenerate intervals that covers E in the sense of Vitali. Then, for each $\epsilon > 0$, there is a subset E_0 of E, with $m(E_0) = 0$, together with a countable, disjoint subcollection $\{I_k\}_{k=1}^\infty$ of $\mathcal{F}$ for which*

$$E = E_0 \cup \bigcup_{k=1}^{\infty} [E \cap I_k] \text{ and } \sum_{k=1}^{\infty} \ell(I_k) < m^*(E) + \epsilon. \tag{2}$$

Proof Since $m^*(E) < \infty$, there is an open set $\mathcal{O}$ containing E for which $m(\mathcal{O}) < m^*(E) + \epsilon$. Because $\mathcal{F}$ is a Vitali covering of E, we may assume that each interval in $\mathcal{F}$ is contained in $\mathcal{O}$. By the countable additivity of measure,

$$\text{if } \{I_k\}_{k=1}^\infty \subseteq \mathcal{F} \text{ is disjoint, then } \sum_{k=1}^{\infty} \ell(I_k) \le m(\mathcal{O}) < m^*(E) + \epsilon < \infty. \tag{3}$$

Therefore, we need to only verify the first assertion of (2). If there is a finite, disjoint subcollection of $\mathcal{F}$ that covers E, there is nothing to prove. So assume that there is no such cover. We will inductively define a disjoint, countable subcollection $\{I_k\}_{k=1}^\infty$ of $\mathcal{F}$ that has the following property: for each $n > 1$,

$$\text{if } I \in \mathcal{F} \text{ and } I \cap \bigcup_{k=1}^{n-1} I_k = \emptyset, \text{ then } \ell(I) < 2 \cdot \ell(I_n). \tag{4}$$

Temporarily assume that such a collection has been selected. This collection has the required properties. Indeed, to prove this, in view of (3) and the continuity of measure, it suffices to show that, for each n,

$$m\left(E \sim \bigcup_{k=1}^{n} [E \cap I_k]\right) \le 5 \cdot \left[\sum_{k=n+1}^{\infty} \ell(I_k)\right]. \tag{5}$$

To verify this inequality, fix n, and let $x \in E \sim \bigcup_{k=1}^{n} [E \cap I_k]$. Since x does not belong to the closed set $\bigcup_{k=1}^{n} I_k$, and $\mathcal{F}$ is a Vitali cover of E, x belongs to an interval $I \in \mathcal{F}$ that is disjoint from I_k, $1 \le k \le n$.

But I cannot be disjoint from all the I_k, since otherwise $\ell(I) < 2 \cdot \ell(I_k)$ for all k which, since I is non-degenerate, contradicts (4). Let $N(n)$ be the first index for which $I \cap I_{N(n)} \neq \emptyset$. Then I is disjoint from each I_i for $1 \leq i < N(n)$ and so, again by (4), $\ell(I) < 2 \cdot \ell(I_{N(n)})$. Observe that this implies that I is contained in the interval with the same center as $I_{N(n)}$ and five times its length. For each $k > n$, let J_k be the interval with the same center at I_k and five times its length. We have shown

$$E \sim \bigcup_{k=1}^{n} [E \cap I_k] \subseteq \bigcup_{k=n+1}^{\infty} J_k \text{ and each } \ell(J_k) < 5 \cdot \ell(I_k)$$

from which, by the countable monotonicity of measure, we obtain (5).

It remains to select the countable, disjoint subcollection $\{I_k\}_{k=1}^{\infty}$ of $\mathcal{F}$ for which (4) holds. Choose any I_1 in $\mathcal{F}$. Fix an index $n > 1$, and suppose that a disjoint subcollection $\{I_k\}_{k=1}^{n}$ of $\mathcal{F}$ has been chosen with the property that, for $1 < k \leq n$,

$$\text{if } I \in \mathcal{F} \text{ and } I \cap \bigcup_{i=1}^{k-1} I_i = \emptyset, \text{ then } \ell(I) < 2 \cdot \ell(I_k).$$

The set $\bigcup_{k=1}^{n} I_k$ is closed, and no finite subcollection of $\mathcal{F}$ covers E, so that since $\mathcal{F}$ is a Vitali cover of E, there is an interval $I \in \mathcal{F}$ for which $I \cap \bigcup_{k=1}^{n} I_k = \emptyset$. Let s_n be the supremum of the lengths of all such intervals in $\mathcal{F}$ and choose I_{n+1} to be one of these for which $\ell(I_{n+1}) > s_n/2$. This inductively defines a sequence for which (4) holds. □

For a real-valued function f and an interior point x of its domain, the **upper derivative** of f at x, $\overline{D}f(x)$ and the **lower derivative** of f at x, $\underline{D}f(x)$ are defined as follows:

$$\overline{D}f(x) = \lim_{h \to 0} \left[\sup_{0<|t|\leq h} \frac{f(x+t) - f(x)}{t} \right];$$

$$\underline{D}f(x) = \lim_{h \to 0} \left[\inf_{0<|t|\leq h} \frac{f(x+t) - f(x)}{t} \right].$$

We have $\overline{D}f(x) \geq \underline{D}f(x)$. If $\overline{D}f(x)$ equals $\underline{D}f(x)$ and is finite, we say that f is **differentiable** at x and define $f'(x)$ to be the common value of the upper and lower derivatives. It is easy to see that f is differentiable at x if and only if, as in calculus,

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x),$$

so that linear combinations of function that are differentiable at x also are differentiable at x. These upper and lower derivatives, often called **Dini derivatives**, are, in conjunction with the Vitali Covering Lemma, essential tools for the study of differentiation in the context of Lebesgue integration.

The Mean Value Theorem of calculus implies that if a function f: $[c, d] \to \mathbf{R}$ is continuous, and differentiable on (c, d) with $f' \geq \alpha$ on (c, d), then

$$\alpha \cdot (d - c) \leq [f(d) - f(c)].$$

The proof of the following generalization of this inequality is a nice illustration of the use of the Vitali Covering Lemma in the analysis of monotone functions.

Lemma 4 *If the function* f: $[a, b] \to \mathbf{R}$ *is increasing, then for each* $\alpha > 0$,

$$m^*\{x \in (a, b) \,|\, \overline{D}f(x) > \alpha\} < \frac{1}{\alpha} \cdot [f(b) - f(a)]. \tag{6}$$

Proof Define $E_\alpha = \{x \in (a, b) \,|\, \overline{D}f(x) > \alpha\}$. Let $\mathcal{F}$ be the collection of closed intervals $[u, v] \subseteq (a, b)$ for which $f(v) - f(u) > \alpha \cdot [v - u]$. Since $\overline{D}f > \alpha$ on E_α, the collection $\mathcal{F}$ is a Vitali covering of E_α. By the Vitali Covering Lemma, there is a subset $E_0 \subseteq E_\alpha$, with $m(E_0) = 0$, together with a countable, disjoint subcollection $\{[a_k, b_k]\}_{k=1}^\infty$ of $\mathcal{F}$ for which

$$E_\alpha = E_0 \cup \bigcup_{k=1}^\infty [E_\alpha \cap [a_k, b_k]].$$

Since f is increasing and $\{[a_k, b_k]\}_{k=1}^\infty$ is disjoint,

$$\sum_{k=1}^\infty [f(b_k) - f(a_k)] \le f(b) - f(a).$$

But each $[a_k, b_k]$ belongs to $\mathcal{F}$ and $E_\alpha \cap [a_k, b_k] \subseteq [a_k, b_k]$, and therefore, by the countable monotonicity of outer measure,

$$m^*(E_\alpha) \le m(E_0) + \sum_{k=1}^\infty [b_k - a_k] < \frac{1}{\alpha} \cdot \sum_{k=1}^\infty [f(b_k) - f(a_k)] \le \frac{1}{\alpha} \cdot [f(b) - f(a)].$$

This proves (6). □

The Lebesgue Differentiation Theorem *If* f: $(a, b) \to \mathbf{R}$ *is a monotone function, then it is differentiable almost everywhere on* (a, b).

Proof Assume that f is increasing. If f is not differentiable at $x \in (a, b)$, then $\overline{D}f(x) > \underline{D}f(x)$. Since f is increasing, $\underline{D}f(x) \ge 0$ and so, since the rational numbers are dense in $\mathbf{R}$, there are rational numbers α and β, with $\alpha > \beta > 0$, for which

$$x \in E_\alpha^\beta = \left\{x \in (a, b) \;\middle|\; \overline{D}f(x) > \alpha > \beta > \underline{D}f(x)\right\}.$$

Since the rational numbers are countable, so is the collection of sets E_α^β as just defined. Consequently, by the countable monotonicity of outer-measure, to prove the theorem it suffices to fix $\alpha > \beta > 0$ and show that $m^*(E_\alpha^\beta) = 0$. Consider the collection $\mathcal{F}$ of closed intervals $[u, v] \subseteq (a, b)$ for which $f(v) - f(u) < \beta \cdot [v - u]$. Since $\underline{D}f(x) < \beta$ on E_α^β, $\mathcal{F}$ is a Vitali covering of E_α^β. Let $\epsilon > 0$. By the Vitali Covering Lemma, there is a subset E_0 of E_α^β, with $m(E_0) = 0$, together with a countable, disjoint subcollection $\{[a_k, b_k]\}_{k=1}^\infty$ of $\mathcal{F}$ for which

$$E_\alpha^\beta = E_0 \cup \bigcup_{k=1}^\infty [E_\alpha^\beta \cap [a_k, b_k]] \text{ and } \sum_{k=1}^\infty [b_k - a_k] < m^*(E_\alpha^\beta) + \epsilon.$$

For each k, employ the preceding lemma to the restriction of f to $[a_k, b_k]$ to conclude that

$$m^*(E_\alpha^\beta \cap [a_k, b_k]) \le m^*\left\{x \in (a_k, b_k) \;\middle|\; \overline{D}f(x) > \alpha\right\} \le \frac{1}{\alpha} \cdot [f(b_k) - f(a_k)].$$

Summing these inequalities, since $m(E_0) = 0$, we have, by the countable monotonicity of outer-measure,

$$m^*(E_\alpha^\beta) \le \frac{1}{\alpha} \cdot \sum_{k=1}^\infty [f(b_k) - f(a_k)].$$

On the other hand, since each $[a_k, b_k]$ belongs to $\mathcal{F}$,

$$\sum_{k=1}^{\infty}[f(b_k) - f(a_k)] < \beta \cdot \sum_{k=1}^{\infty}[b_k - a_k] < \beta \cdot [m^*(E_\alpha^\beta) + \epsilon].$$

Consequently,

$$m^*(E_\alpha^\beta) \leq \frac{\beta}{\alpha} \cdot [m^*(E_\alpha^\beta) + \epsilon].$$

This holds for all $\epsilon > 0$, and $0 < \beta < \alpha$, and therefore $m^*(E_\alpha^\beta) = 0$. □

Let $f\colon [a, b] \to \mathbf{R}$ be integrable. Extend f to take the value $f(b)$ on $(b, b+1]$. For $0 < h \leq 1$, define the **divided difference function** $\mathrm{Diff}_h\, f$ and **average value function** $\mathrm{Av}_h\, f$ of $[a, b]$ by

$$\mathrm{Diff}_h\, f(x) = \frac{f(x+h) - f(x)}{h} \text{ and } \mathrm{Av}_h\, f(x) = \frac{1}{h} \cdot \int_x^{x+h} f\, dm \text{ for all } x \in [a, b].$$

By a change of variables by translation, and cancellation, for all $a \leq u < v \leq b$,

$$\int_u^v \mathrm{Diff}_h\, f\, dm = \mathrm{Av}_h\, f(v) - \mathrm{Av}_h\, f(u). \tag{7}$$

Corollary 5 *If $f\colon [a, b] \to \mathbf{R}$ is an increasing function, then its derivative f' is integrable over $[a, b]$ and*

$$\int_a^b f'\, dm \leq f(b) - f(a). \tag{8}$$

Proof Extend f to take the value $f(b)$ on $(b, b+1]$. Since f is increasing on $[a, b+1]$, it is measurable and therefore the divided difference functions also are measurable, and they are non-negative. According to Lebesgue's Differentiation Theorem, f is differentiable almost everywhere on (a, b). Therefore, $\{\mathrm{Diff}_{1/n}\, f\}$ is a sequence of non-negative measurable functions that converges pointwise almost everywhere on $[a, b]$ to f'. According to Fatou's Lemma,

$$\int_a^b f'\, dm \leq \liminf \left[\int_a^b \mathrm{Diff}_{1/n}\, f\, dm\right]. \tag{9}$$

By (7), for each n, since f is increasing,

$$\int_a^b \mathrm{Diff}_{1/n}\, f\, dm = \frac{1}{1/n} \cdot \int_b^{b+1/n} f\, dm - \frac{1}{1/n} \cdot \int_a^{a+1/n} f\, dm \leq f(b) - f(a).$$

Consequently,

$$\limsup \left[\int_a^b \mathrm{Diff}_{1/n}\, f\, dm\right] \leq f(b) - f(a). \tag{10}$$

Inequality (8) follows from inequalities (9) and (10). □

Example For a continuous function $f\colon [a, b] \to \mathbf{R}$ that is differentiable on the open interval (a, b), if f fails to be monotone, its derivative f' may fail to be integrable. We leave it as an exercise to show that if $f\colon [0, 1] \to \mathbf{R}$ is defined by

$$f(x) = \begin{cases} x^2 \sin(1/x^2) & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0, \end{cases}$$

then f is continuous on $[0, 1]$ and differentiable on $(0, 1)$, but f' is not integrable over $(0, 1)$.

Lebesgue's Theorem is the best possible, in the sense that if $E \subseteq (a, b)$ is a set of measure zero, then there is an increasing function on (a, b) that fails to be differentiable at each point in E (see Problem 9)[2].

PROBLEMS

4. Show that the Vitali Covering Lemma does not extend to the case in which the covering collection has degenerate closed intervals.
5. Show that the Vitali Covering Lemma does extend to the case in which the covering collection comprises non-degenerate general intervals.
6. Is every continuous function $f\colon [0, 1] \to \mathbf{R}$ the difference of monotone functions?
7. Let I and J be closed, bounded intervals and $\gamma > 0$ be such that $\ell(I) > \gamma \cdot \ell(J)$. Assume that $I \cap J \neq \emptyset$. Show that if $\gamma \geq 1/2$, then $J \subseteq 3 * I$. Is the same true if $0 < \gamma < 1/2$?
8. Show that $m(E) = 0$ if and only if there is a countable collection of open intervals $\{I_k\}_{k=1}^{\infty}$ for which each point in E belongs to infinitely many of the I_k's and $\Sigma_{k=1}^{\infty} \ell(I_k) < \infty$.
9. (Riesz-Nagy) Let E be a set of measure zero contained in the open interval (a, b). Let $\{(c_k, d_k)\}_{k=1}^{\infty}$ be a cover of E as in the preceding problem. Define
$$f(x) = \sum_{k=1}^{\infty} \ell((c_k, d_k) \cap (-\infty, x)) \text{ for all } x \text{ in } (a, b).$$
Show that f is increasing and fails to be differentiable at each point in E.
10. For real numbers $\alpha < \beta$ and $\gamma > 0$, show that if $g\colon [\alpha + \gamma, \beta + \gamma] \to \mathbf{R}$ is integrable then
$$\int_{\alpha}^{\beta} g(t + \gamma)\, dt = \int_{\alpha+\gamma}^{\beta+\gamma} g(t)\, dt.$$
Prove this change of variables formula by successively considering simple functions, bounded, measurable functions, non-negative integrable functions, and general integrable functions. Use it to prove (7).
11. Compute the upper and lower derivatives of the characteristic function of the rationals.
12. Assume that $m^*(E) < \infty$ and let $\mathcal{F}$ be a collection of closed, bounded, non-degenerate intervals that cover E in the sense of Vitali. Show that there is a countable, disjoint collection $\{I_k\}_{k=1}^{\infty}$ of intervals in $\mathcal{F}$ for which
$$m^* \left[E \sim \bigcup_{k=1}^{\infty} I_k \right] = 0.$$

[2] In their book *Functional Analysis*, [RSN90], Frigyes Riesz and Béla Sz.-Nagy remark that Lebesgue's Differentiation Theorem is "one of the most striking and most important in real variable theory." A generalization of this theorem to functions on Euclidean space may be found in "Measure Theory and Finer Properties of Functions," by L.C. Evans and R.G. Gariepy.

13. Use the Vitali Covering Lemma to show that the union of any collection (countable or uncountable) of closed, bounded, non-degenerate intervals is measurable.
14. Define f: $\mathbf{R} \to \mathbf{R}$ by

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

Find the upper and lower derivatives of f at $x = 0$.

15. For an integrable function g: $[a, b] \to \mathbf{R}$, define the antiderivative of g to be the function f defined on $[a, b]$ by

$$f(x) = \int_a^x g\, dm \text{ for all } x \in [a, b].$$

Show that f is differentiable almost everywhere on (a, b).

16. Verify (7).
17. Show that for f: $(a, b) \to \mathbf{R}$ and $c \in (a, b)$ a local minimizer for f, $\underline{D}f(c) \leq 0 \leq \overline{D}f(c)$.
18. For f: $[a, b] \to \mathbf{R}$ assume that $\underline{D}f \geq 0$ on (a, b). Show that f is increasing on $[a, b]$. (Suggestion: First show this for a function g for which $\underline{D}g \geq \epsilon > 0$ on (a, b). Apply this to the function $g(x) = f(x) + \epsilon x$.)
19. For functions f: $[a, b] \to \mathbf{R}$ and g: $[a, b] \to \mathbf{R}$, show that

$$\underline{D}f + \underline{D}g \leq \underline{D}(f + g) \leq \overline{D}(f + g) \leq \overline{D}f + \overline{D}g \text{ on } (a, b).$$

20. Let f be defined on $[a, b]$ and g a continuous function on $[\alpha, \beta]$ that is differentiable at $\gamma \in (\alpha, \beta)$ with $g(\gamma) = c \in (a, b)$. Verify the following.
    (i) If $g'(\gamma) > 0$, then $\overline{D}(f \circ g)(\gamma) = \overline{D}f(c) \cdot g'(\gamma)$.
    (ii) If $g'(\gamma) = 0$ and the upper and lower derivatives of f at c are finite, then $\overline{D}(f \circ g)(\gamma) = 0$.
21. Show that a continuous function f: $[a, b] \to \mathbf{R}$ is Lipschitz if its upper and lower derivatives are bounded on (a, b).
22. Show that for f defined in the last example of this section, f': $[0, 1] \to \mathbf{R}$ is not integrable.

6.3 FUNCTIONS OF BOUNDED VARIATION: JORDAN'S THEOREM

In view of Lebesgue's Differentiation Theorem, the difference of two increasing functions on an open interval is differentiable almost everywhere. We now provide a characterization, due to Jordan, of the collection of functions on a closed, bounded interval that may be expressed as the difference of increasing functions, which shows that this class is surprisingly large; it includes, for instance, all Lipschitz functions.

For a function f: $[a, b] \to \mathbf{R}$ and a partition $P = \{x_0, \ldots, x_k\}$ of $[a, b]$, we define the **variation** of f: $[a, b] \to \mathbf{R}$ with respect to P by

$$V(f, P) = \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})|,$$

and the **total variation** of f: $[a, b] \to \mathbf{R}$ by

$$TV(f) = \sup\{V(f, P) \mid P \text{ a partition of } [a, b]\}.$$

For a subinterval $[c, d]$ of $[a, b]$, $TV(f_{[c, d]})$ denotes the total variation of the restriction of f to $[c, d]$.

Definition *A function f: $[a, b] \to \mathbf{R}$ is said to be of* **bounded variation** *provided that*

$$TV(f) < \infty.$$

Example Let f: $[a, b] \to \mathbf{R}$ be increasing. Then f is of bounded variation and

$$TV(f) = f(b) - f(a).$$

Indeed, for a partition $P = \{x_0, \ldots, x_k\}$ of $[a, b]$,

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

Example Let f: $[a, b] \to \mathbf{R}$ be Lipschitz. Then f is of bounded variation and

$$TV(f) \le c \cdot (b - a),$$

where

$$|f(u) - f(v)| \le c|u - v| \text{ for all } u, v \text{ in } [a, b].$$

Indeed, for a partition $P = \{x_0, \ldots, x_k\}$ of $[a, b]$,

$$V(f, P) = \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \le c \cdot \sum_{i=1}^{k} [x_i - x_{i-1}] = c \cdot [b - a].$$

Example Define the function f: $[0, 1] \to \mathbf{R}$ by

$$f(x) = \begin{cases} x \cos(\pi/2x) & \text{if } 0 < x \le 1 \\ 0 & \text{if } x = 0. \end{cases}$$

Then f is continuous, but, we claim, is not of bounded variation. Indeed, for each n, consider the partition $P_n = \{0, 1/2n, 1/[2n-1], \ldots, 1/3, 1/2, 1\}$ of $[0, 1]$. Then

$$V(f, P_n) = 1 + 1/2 + \cdots + 1/n,$$

so that f is not of bounded variation on $[0, 1]$, since the harmonic series diverges.

Observe that if $c \in (a, b)$, P is a partition of $[a, b]$, and P' is the refinement of P obtained by adjoining c to P, then, by the triangle inequality, $V(f, P) \le V(f, P')$. Consequently, in the definition of the total variation of a function on $[a, b]$, the supremum can be taken over partitions of $[a, b]$ that contain the point c. Now a partition P of $[a, b]$ that contains the point c induces, and is induced by, partitions P_1 and P_2 of $[a, c]$ and $[c, b]$, respectively, and for such partitions

$$V(f_{[a, b]}, P) = V(f_{[a, c]}, P_1) + V(f_{[c, b]}, P_2). \tag{11}$$

Take the supremum among such partitions to conclude that

$$TV(f_{[a,\,b]}) = TV(f_{[a,\,c]}) + TV(f_{[c,\,b]}). \tag{12}$$

Therefore, if f: $[a, b] \to \mathbf{R}$ is of bounded variation, then

$$TV(f_{[a,\,v]}) - TV(f_{[a,\,u]}) = TV(f_{[u,\,v]}) \geq 0 \text{ for all } a \leq u < v \leq b. \tag{13}$$

As a consequence, the function $x \mapsto TV(f_{[a,\,x]})$, which we call the **total variation function** for f, is an increasing function on $[a, b]$. Observe that for $a \leq u < v \leq b$, if we take the crudest partition of $[u, v]$, namely $P = \{u, v\}$, then

$$f(u) - f(v) \leq |f(v) - f(u)| = V(f_{[u,\,v]}, P) \leq TV(f_{[u,\,v]}) = TV(f_{[a,\,v]}) - TV(f_{[a,\,u]}),$$

so that

$$f(v) + TV(f_{[a,\,v]}) \geq f(u) + TV(f_{[a,\,u]}) \text{ for all } a \leq u < v \leq b.$$

This inequality together with inequality (13) establishes the following lemma.

Jordan's Lemma *If the function f: $[a, b] \to \mathbf{R}$ is of bounded variation, then it has the following explicit expression as the difference of two increasing functions on $[a, b]$:*

$$f(x) = [f(x) + TV(f_{[a,\,x]})] - TV(f_{[a,\,x]}) \textit{ for all } x \in [a, b]. \tag{14}$$

Jordan's Theorem *A function f: $[a, b] \to \mathbf{R}$ is of bounded variation if and only if it is the difference of two increasing functions.*

Proof Let f: $[a, b] \to \mathbf{R}$ be of bounded variation. The preceding lemma provides an explicit representation of f as the difference of increasing functions. To prove the converse, let $f = g - h$ on $[a, b]$, where g and h are increasing functions on $[a, b]$. For any partition $P = \{x_0, \ldots, x_k\}$ of $[a, b]$,

$$\begin{aligned}
V(f, P) &= \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \\
&= \sum_{i=1}^{k} |[g(x_i) - g(x_{i-1})] + [h(x_{i-1}) - h(x_i)]| \\
&\leq \sum_{i=1}^{k} |g(x_i) - g(x_{i-1})| + \sum_{i=1}^{k} |h(x_{i-1}) - h(x_i)| \\
&= \sum_{i=1}^{k} [g(x_i) - g(x_{i-1})] + \sum_{i=1}^{k} [h(x_i) - h(x_{i-1})] \\
&= [g(b) - g(a)] + [h(b) - h(a)].
\end{aligned}$$

This holds for every partition P and therefore f is of bounded variation. □

Corollary 6 *If the function f: $[a, b] \to \mathbf{R}$ is of bounded variation, then it is differentiable almost everywhere on (a, b) and its derivative f' is integrable over (a, b).*

Proof According to Jordan's Theorem, f is the difference of two increasing functions on $[a, b]$. Therefore, Lebesgue's Theorem implies that f is the difference of two functions that are differentiable almost everywhere on (a, b). As a consequence, f is differentiable almost everywhere on (a, b). The integrability of f' follows from Corollary 4. □

PROBLEMS

23. Let f be the Dirichlet function, the characteristic function of the rationals in $[0, 1]$. Is f: $[0, 1] \to \mathbf{R}$ of bounded variation?

24. Define $f(x) = \sin x$ on $[0, 2\pi]$. Find two increasing functions h and g for which $f = h - g$ on $[0, 2\pi]$.

25. Let f: $[a, b] \to \mathbf{R}$ be a step-function. Find a formula for its total variation.

26. (i) Define

$$f(x) = \begin{cases} x^2 \ \cos(1/x^2) & \text{if } x \neq 0, x \in [-1, 1] \\ 0 & \text{if } x = 0. \end{cases}$$

Is f: $[-1, 1] \to \mathbf{R}$ of bounded variation?

(ii) Define

$$g(x) = \begin{cases} x^2 \ \cos(1/x) & \text{if } x \neq 0, x \in [-1, 1] \\ 0 & \text{if } x = 0. \end{cases}$$

Is g: $[-1, 1] \to \mathbf{R}$ of bounded variation?

27. Suppose that f: $[0, 1] \to \mathbf{R}$ is continuous. Must there be a non-degenerate, closed subinterval $[a, b]$ of $[0, 1]$ for which the restriction of f to $[a, b]$ is of bounded variation?

28. Let P be a partition of $[a, b]$ that is a refinement of the partition P'. For f: $[a, b] \to \mathbf{R}$, show that $V(f, P') \leq V(f, P)$.

29. Assume that f: $[a, b] \to \mathbf{R}$ is of bounded variation. Show that there is a sequence of partitions $\{P_n\}$ of $[a, b]$ for which the sequence $\{V(f, P_n)\}$ is increasing and converges to $TV(f)$. (Suggestion: Consider the preceding problem.)

30. Let the sequence $\{f_n$: $[a, b] \to R\}$ converge pointwise to f: $[a, b] \to \mathbf{R}$. Show that

$$TV(f) \leq \liminf \ TV(f_n).$$

31. Let f: $[a, b] \to \mathbf{R}$ and g: $[a, b] \to \mathbf{R}$ be of bounded variation. Show that

$$TV(f + g) \leq TV(f) + TV(g) \text{ and } TV(\alpha f) = |\alpha| TV(f).$$

32. For α and β positive numbers, define the function f on $[0, 1]$ by

$$f(x) = \begin{cases} x^\alpha \sin(1/x^\beta) & \text{for } 0 < x \leq 1 \\ 0 & \text{for } x = 0. \end{cases}$$

Show that if $\alpha > \beta$, then f: $[0, 1] \to \mathbf{R}$ is of bounded variation, by showing that f' is integrable over $[0, 1]$. Then show that if $\alpha \leq \beta$, then f: $[0, 1] \to \mathbf{R}$ is not of bounded variation.

33. Let f: $[0, 1] \to \mathbf{R}$ fail to be of bounded variation. Show that there is a point x_0 in $[0, 1]$ such that there are subintervals of $[0, 1]$ that contain x_0 and have arbitrarily small length on which f fails to be of bounded variation.

6.4 ABSOLUTELY CONTINUOUS FUNCTIONS

Definition *A function $f\colon [a, b] \to \mathbf{R}$ is said to be* ***absolutely continuous*** *provided that for each $\epsilon > 0$, there is a $\delta > 0$ such that for every finite, disjoint collection $\{(a_k, b_k)\}_{k=1}^n$ of open subintervals of (a, b),*

$$\textit{if } \sum_{k=1}^{n} [b_k - a_k] < \delta, \ \textit{ then } \sum_{k=1}^{n} |f(b_k) - f(a_k)| < \epsilon.$$

The criterion for the absolute continuity of $f\colon [a, b] \to \mathbf{R}$ in the case that the finite collection of intervals comprises a single interval is the criterion for the uniform continuity of f on $[a, b]$. Consequently, absolutely continuous functions are continuous. We now show that the Cantor-Lebesgue function, which is increasing and continuous, fails to be absolutely continuous.

Example The Cantor-Lebesgue function $\varphi\colon [0, 1] \to \mathbf{R}$ is not absolutely continuous (also see Problems 42 and 48). Indeed, to verify this, observe that for each n, at the n-th stage of the construction of the Cantor set, a disjoint collection $\{[c_k, d_k]\}_{1 \le k \le 2^n}$ of 2^n subintervals of $[0, 1]$ have been constructed that cover the Cantor set, each of which has length $(1/3)^n$. The Cantor-Lebesgue function is constant on each of the intervals that comprise the complement in $[0, 1]$ of this collection of intervals. But

$$\sum_{1 \le k \le 2^n} [d_k - c_k] = (2/3)^n \text{ while } \sum_{1 \le k \le 2^n} [\varphi(d_k) - \varphi(c_k)] = \psi(1) - \psi(0) = 1,$$

and consequently there is no response to the $\epsilon = 1$ challenge regarding the criterion for φ to be absolutely continuous.

Linear combinations of absolutely continuous functions are absolutely continuous, and since such functions are bounded, so are products. However, the composition of absolutely continuous functions may fail to be absolutely continuous (see Problems 44–46).

Proposition 7 *If the function $f\colon [a, b] \to \mathbf{R}$ is Lipschitz, then it is absolutely continuous.*

Proof Let $c > 0$ be a Lipschitz constant for f on $[a, b]$, that is,

$$|f(u) - f(v)| \le c\,|u - v| \text{ for all } u, v \in [a, b].$$

Then, regarding the criterion for the absolute continuity of f, it is clear that $\delta = \epsilon/c$ responds to any $\epsilon > 0$ challenge. $\square$

We leave it as an exercise to show the function f on $[0, 1]$, defined by $f(x) = \sqrt{x}$ for $0 \le x \le 1$, is absolutely continuous but not Lipschitz.

Lemma 8 *If the function $f\colon [a, b] \to \mathbf{R}$ is absolutely continuous, then it is of bounded variation and its total variation function also is absolutely continuous.*

Proof First we show that f is of bounded variation. Indeed, let δ respond to the $\epsilon = 1$ challenge regarding the criterion for the absolute continuity of f. Let P be a partition of $[a, b]$

into N closed intervals $\{[c_k, d_k]\}_{k=1}^N$, each of length less than δ. Then, by the definition of δ in relation to the absolute continuity of f, it is clear that $TV(f_{[c_k, d_k]}) \le 1$, for $1 \le k \le N$. The additivity formula for total variation (12) extends to finite sums. Hence

$$TV(f) = \sum_{k=1}^{N} TV(f_{[c_k, d_k]}) \le N.$$

Consequently, f is of bounded variation, and it remains to show that its total variation function is absolutely continuous. Let $\epsilon > 0$. Choose δ as a response to the $\epsilon/2$ challenge regarding the criterion for the absolute continuity of f on $[a, b]$. Let $\{(c_k, d_k)\}_{k=1}^n$ be a disjoint collection of open subintervals of (a, b) for which $\sum_{k=1}^n [d_k - c_k] < \delta$. For $1 \le k \le n$, let P_k be a partition of $[c_k, d_k]$. By the choice of δ in relation to the absolute continuity of the function f

$$\sum_{k=1}^{n} V(f_{[c_k, d_k]}, P_k) < \epsilon/2.$$

Take the supremum as, for $1 \le k \le n$, P_k vary among partitions of $[c_k, d_k]$, to obtain

$$\sum_{k=1}^{n} TV(f_{[c_k, d_k]}) \le \epsilon/2 < \epsilon.$$

It follows from (12) that, for $1 \le k \le n$, $TV(f_{[c_k, d_k]}) = TV(f_{[a, d_k]}) - TV(f_{[a, c_k]})$. Hence

$$\text{if } \sum_{k=1}^{n} [d_k - c_k] < \delta, \text{ then } \sum_{k=1}^{n} \left[TV(f_{[a, d_k]}) - TV(f_{[a, c_k]})\right] < \epsilon. \tag{15}$$

So the total variation function of f is absolutely continuous. □

Theorem 9 *If the function $f\colon [a, b] \to \mathbf{R}$ is absolutely continuous, then it is the difference of increasing absolutely continuous functions.*

Proof The proof follows from Jordan's Lemma and the preceding one. □

There is the following refinement of the criterion for uniform integrability.

Lemma 10 *Let $\mathcal{F}$ be a collection of real-valued integrable functions on $[a, b]$. Then $\mathcal{F}$ is uniformly integrable if for each $\epsilon > 0$, there is a $\delta > 0$ such that for $E = \bigcup_{k=1}^n (a_k, b_k)$, the union of a finite, disjoint collection of open subintervals in (a, b),*

$$\textit{if } m(E) < \delta, \textit{ then } \int_E |f|\, dm < \epsilon \textit{ for all } f \in \mathcal{F}. \tag{16}$$

Proof Let $\epsilon > 0$. Choose $\delta > 0$ to be a response to the $\epsilon/2$ challenge regarding the proposed uniform integrability criterion in the statement in the lemma. We claim that (16) holds for any measurable subset E of $[a, b]$, and so $\mathcal{F}$ is uniformly integrable. First consider the case $E = \mathcal{O}$, an open subset of (a, b) for which $m(\mathcal{O}) < \delta$. Now $\mathcal{O} = \bigcup_{k=1}^{\infty}(a_k, b_k)$, the union of a countable, disjoint collection of open subintervals of (a, b). For each n, define $\mathcal{O}_n = \bigcup_{k=1}^{n}(a_k, b_k)$. Since $m(\mathcal{O}) < \delta$, $m(\mathcal{O}_n) < \delta$ for all n, and consequently, by the choice of δ together with the continuity of integration,

$$\int_{\mathcal{O}} |f|\, dm = \lim_{n\to\infty} \int_{\mathcal{O}_n} |f|\, dm \le \epsilon/2 \text{ for all } f \in \mathcal{F}. \tag{17}$$

Next consider the case $E = G$, a G_δ set for which $m(G) < \delta$. Now $G = \cap_{n=1}^{\infty} U_n$, the intersection of a descending, countable collection of open subsets of (a, b). It follows from (17), together with the continuity of both measure and integration, that

$$\int_G |f|\, dm = \lim_{n\to\infty} \int_{U_n} |f|\, dm \le \epsilon/2 < \epsilon \text{ for all } f \in \mathcal{F}.$$

From this we conclude that (16) holds for any measurable subset E of $[a, b]$, since such a set is a G_δ set from which a set of measure zero has been excised. □

Theorem 11 *If the function f: $[a, b+1] \to \mathbf{R}$ is absolutely continuous, then the collection of divided difference functions $\mathcal{F} = \{\text{Diff}_h\, f\colon [a, b] \to \mathbf{R}\}_{0<h\le 1}$ is uniformly integrable.*

Proof In view of Theorem 9, we may assume that f is increasing, so the divided difference functions are non-negative. Let $\epsilon > 0$. Choose $\delta > 0$ to be a response to the ϵ challenge regarding the criterion for the absolute continuity of f: $[a, b+1] \to \mathbf{R}$. We claim that δ is also a response to the $\epsilon > 0$ challenge in the criterion for uniform integrability of $\mathcal{F}$ stated in Lemma 10. To verify this, let $E = \bigcup_{k=1}^{n}(a_k, b_k)$ be the union of a finite, disjoint collection of open subintervals of (a, b) for which

$$m(E) = \sum_{k=1}^{n}[b_k - a_k] < \delta. \tag{18}$$

Fix $0 < h \le 1$. Observe that, by a change of variables,

$$\int_E \frac{f(x+h) - f(x)}{h}\, dx = \sum_{k=1}^{n} \int_{a_k}^{b_k} \frac{f(x+h) - f(x)}{h}\, dx = \frac{1}{h} \cdot \int_0^h \phi(t)\, dt$$

where

$$\phi(t) = \sum_{k=1}^{n}[f(d_k + t) - f(c_k + t)] \textit{ for } 0 \le t \le h.$$

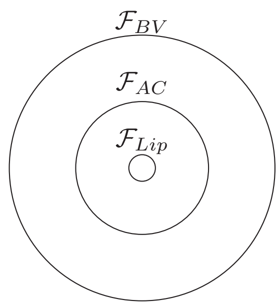
However, for each $0 \le t \le h$, since $\sum\limits_{k=1}^{n} [b_k + t] - [a_k + t] < \delta$, $\phi(t) < \epsilon$. Consequently,

$$\int_E \frac{f(x+h) - f(x)}{h}\, dx = \frac{1}{h} \cdot \int_0^h \phi(t)\, dt < \epsilon.$$

Since the divided difference functions are non-negative, $\mathcal{F}$ is uniformly integrable. □

For a closed, bounded, interval $[a, b]$, let $\mathcal{F}_{Lip}$, $\mathcal{F}_{AC}$, and $\mathcal{F}_{BV}$ denote the collection of functions on $[a, b]$ that are Lipschitz, absolutely continuous, and of bounded variation, respectively. We have the following strict inclusions:

$$\mathcal{F}_{Lip} \subseteq \mathcal{F}_{AC} \subseteq \mathcal{F}_{BV}. \tag{19}$$



Proposition 7 implies the first inclusion, while Lemma 8 implies the second. Each of these collections is a linear space. Moreover, a function in one of these spaces may be expressed as the difference of increasing functions in the same space. The Lipschitz functions are included since they are the most recognizable absolutely continuous functions. And, of course, a continuous function $f\colon [a, b] \to \mathbf{R}$ that has a bounded derivative on (a, b) is Lipschitz.

PROBLEMS

34. Let $f\colon [0, 1] \to \mathbf{R}$ be continuous, and be absolutely continuous on $[\epsilon, 1]$ for each $0 < \epsilon < 1$.
    (i) Show that $f\colon [0, 1] \to \mathbf{R}$ may not be absolutely continuous.
    (ii) Show that $f\colon [0, 1] \to \mathbf{R}$ is absolutely continuous if it is increasing.
    (iii) Show that the function $f\colon [0, 1] \to \mathbf{R}$ defined by $f(x) = \sqrt{x}$ for $0 \leq x \leq 1$ is absolutely continuous, but not Lipschitz.
35. Verify that the inclusions in (19) are strict.
36. Show that a Lipschitz function $f\colon [a, b] \to \mathbf{R}$ is the difference of increasing Lipschitz functions.
37. Assume that $f\colon [a, b] \to \mathbf{R}$ is increasing. Show that f is absolutely continuous if and only if for each $\epsilon > 0$, there is a $\delta > 0$ such that for every countable, disjoint collection $\{(a_k, b_k)\}_{k=1}^{\infty}$ of open intervals in (a, b),
$$\sum_{k=1}^{\infty}[f(b_k) - f(a_k)] < \epsilon \text{ if } \sum_{k=1}^{\infty}[b_k - a_k] < \delta.$$
38. Show that if $f\colon [a, b] \to \mathbf{R}$ is increasing and absolutely continuous, then it maps sets of measure zero to sets of measure zero. (Suggestion: Consider the preceding problem.)
39. Let $f\colon [a, b] \to \mathbf{R}$ be continuous. Conclude from the continuity of f and the compactness of $[a, b]$ that f maps closed sets to closed sets and therefore maps F_σ sets to F_σ sets.
40. Show that if $f\colon [a, b] \to \mathbf{R}$ is increasing and absolutely continuous, then f maps measurable sets to measurable sets. (Suggestion: Consider the preceding two problems.)

41. Show that if f: $[a, b] \to \mathbf{R}$ is continuous and increasing, then f is absolutely continuous if and only if for each ϵ, there is a $\delta > 0$ such that for each measurable subset E of $[a, b]$,

$$m(f(E)) < \epsilon \text{ if } m(E) < \delta.$$

42. Show that the Cantor-Lebesgue function φ is not absolutely continuous on $[0, 1]$ by arguing that the function ψ, defined by $\psi(x) = x + \varphi(x)$ for $0 \leq x \leq 1$, maps the Cantor, a set of measure zero, to a set of measure 1. (Suggestion: Consider the preceding problem.)
43. Show that both the sum and product of absolutely continuous functions are absolutely continuous.
44. Define the functions f and g on $[-1, 1]$ by $f(x) = x^{\frac{1}{3}}$ for $-1 \leq x \leq 1$ and

$$g(x) = \begin{cases} x^2 \ \cos(\pi/2x) & \text{if } x \neq 0, x \in [-1,\, 1] \\ 0 & \text{if } x = 0. \end{cases}$$

   (i) Show that both f: $[-1, 1] \to \mathbf{R}$ and g: $[-1, 1] \to \mathbf{R}$ are absolutely continuous.
   (ii) For the partition $P_n = \{-1, 0, 1/2n, 1/[2n-1], \ldots, 1/3, 1/2, 1\}$ of $[-1, 1]$, examine $V(f \circ g, P_n)$.
   (iii) Show that $f \circ g$: $[-1, 1] \to \mathbf{R}$ fails to be of bounded variation, and hence also fails to be absolutely continuous.
45. Let f: $\mathbf{R} \to \mathbf{R}$ be Lipschitz and g: $[a, b] \to \mathbf{R}$ be absolutely continuous. Show that the composition $f \circ g$: $[a, b] \to \mathbf{R}$ is absolutely continuous.
46. Let f: $\mathbf{R} \to \mathbf{R}$ be absolutely continuous and g: $[a, b] \to \mathbf{R}$ be absolutely continuous and strictly increasing. Show that the composition $f \circ g$: $[a, b] \to \mathbf{R}$ is absolutely continuous.
47. Prove the converse of the implication in Theorem 11.

6.5 INTEGRATING DERIVATIVES: DIFFERENTIATING INDEFINITE INTEGRALS

The following is the first version of the Fundamental Theorem of Calculus for the Lebesgue integral.

Theorem 12 *If the function f: $[a, b] \to \mathbf{R}$ is absolutely continuous, then it is differentiable almost everywhere on (a, b), its derivative f' is integrable, and*

$$\int_a^b f'\, dm = f(b) - f(a). \tag{20}$$

Proof[3] Extend f to take the value $f(b)$ on $[b, b+1]$. As already noted, there is the discrete formulation of (20) for the Riemann integral, namely, for each n,

$$\int_a^b \frac{f(x+1/n) - f(x)}{1/n}\, dx = \frac{1}{1/n}\int_b^{b+1/n} f(x)\, dx - \frac{1}{1/n}\int_a^{a+h} f(x)\, dx. \tag{21}$$

[3]This proof of the fundamental theorem of integral calculus for the Lebesgue integral is in the note *Absolute continuity of a function and uniform integrability of its divided differences* (The American Mathematical Monthly, January (2015)) by Patrick Fitzpatrick and Brian Hunt, in which Theorem 11 and its converse are proven.

Since $f\colon [a, b+1] \to \mathbf{R}$ is continuous,

$$\lim_{n\to\infty} \left[\frac{1}{1/n} \int_b^{b+1/n} f(x)\, dx - \frac{1}{1/n} \int_a^{a+h} f(x)\, dx \right] = f(b) - f(a). \tag{22}$$

Theorem 9 implies that f is the difference of increasing functions on $[a, b+1]$ and therefore, by Lebesgue's Differentiation Theorem, f is differentiable almost everywhere on (a, b). Consequently, $\{\mathrm{Diff}_{1/n} f\}$ converges pointwise almost everywhere on (a, b) to f'. On the other hand, according to Theorem 11, the sequence $\{\mathrm{Diff}_{1/n} f\}$ is uniformly integrable over $[a, b]$. The Vitali Convergence Theorem permits passage of the limit under the integral sign in order to conclude that

$$\lim_{n\to\infty} \left[\int_a^b \frac{f(x+1/n) - f(x)}{1/n}\, dx \right] = \int_a^b f'\, dm. \tag{23}$$

Formula (20) follows from (21), (22), and (23). □

In the study of calculus, indefinite integrals are defined with respect to the Riemann integral. We here call a function $f\colon [a, b] \to \mathbf{R}$ the **indefinite integral** of the Lebesgue integrable function $g\colon [a, b] \to \mathbf{R}$ provided that

$$f(x) = f(a) + \int_a^x g\, dm \text{ for all } x \in [a, b]. \tag{24}$$

Theorem 13 *A function $f\colon [a, b] \to \mathbf{R}$ is absolutely continuous if and only if it is an indefinite integral.*

Proof First suppose that $f\colon [a, b] \to \mathbf{R}$ is absolutely continuous. For each $x \in (a, b]$, the restriction $f\colon [a, x] \to \mathbf{R}$ is absolutely continuous and therefore, by the preceding theorem, in the case $[a, b]$ is replaced by $[a, x]$,

$$f(x) = f(a) + \int_a^x f'\, dm.$$

Consequently, f is the indefinite integral of f' over $[a, b]$.

Conversely, suppose that $f\colon [a, b] \to \mathbf{R}$ is the indefinite integral of $g\colon [a, b] \to \mathbf{R}$. For a disjoint collection $\{(a_k, b_k)\}_{k=1}^n$ of open intervals in (a, b), if we define $E = \bigcup_{k=1}^n (a_k, b_k)$, then, by the monotonicity and additivity over domains properties of the integral,

$$\sum_{k=1}^n |f(b_k) - f(a_k)| = \sum_{k=1}^n \left| \int_{b_k}^{a_k} g\, dm \right| \le \sum_{k=1}^n \int_{b_k}^{a_k} |g|\, dm = \int_E |g|\, dm. \tag{25}$$

Let $\epsilon > 0$. Now the single function $|g|\colon [a, b] \to \mathbf{R}$ is uniformly integrable, and so there is a $\delta > 0$ such that $\int_E |g|\, dm < \epsilon$ if $E \subseteq [a, b]$ is measurable and $m(E) < \delta$. It follows from (25) that this same δ responds to the ϵ challenge regarding the criterion for f to be absolutely continuous on $[a, b]$. □

Corollary 14 *A monotone function $f\colon [a, b] \to \mathbf{R}$ is absolutely continuous if and only if*

$$\int_a^b f'\, dm = f(b) - f(a). \tag{26}$$

Proof Theorem 12 is the assertion that (26) holds if f is absolutely continuous, irrespective of any monotonicity assumption. Conversely, assume that f is increasing and (26) holds. Let $x \in [a, b]$. By the additivity over domains of integration,

$$0 = \int_a^b f'\, dm - [f(b) - f(a)] = \left\{\int_a^x f'\, dm - [f(x) - f(a)]\right\} + \left\{\int_x^b f'\, dm - [f(b) - f(x)]\right\}.$$

According to Corollary 5,

$$\int_a^x f'\, dm - [f(x) - f(a)] \le 0 \text{ and } \int_x^b f'\, dm - [f(b) - f(x)] \le 0.$$

If the sum of two non-positive numbers is zero, then they both are zero. Therefore,

$$f(x) = f(a) + \int_a^x f'\, dm,$$

so that f is the indefinite integral of f'. The preceding theorem implies that f is absolutely continuous. □

Lemma 15 *If the function $f\colon [a, b] \to \mathbf{R}$ is integrable, then $f = 0$ almost everywhere on $[a, b]$ if*

$$\int_{x_1}^{x_2} f\, dm = 0 \textit{ for all } [x_1, x_2] \subseteq (a, b). \tag{27}$$

Proof We claim that

$$\int_E f\, dm = 0 \text{ for all measurable sets } E \subseteq [a, b]. \tag{28}$$

Indeed, (28) holds for all open subsets of (a, b), since integration is countably additive and every open set is the union of countable, disjoint collection of open intervals. The continuity of integration then implies that (28) also holds for all G_δ sets contained in (a, b). But every measurable set is a G_δ set from which a set of measure zero has been excised. We conclude from the additivity over domains of integration that (28) is verified. Define

$$E^+ = \{x \in [a, b] \mid f(x) \ge 0\} \text{ and } E^- = \{x \in [a, b] \mid f(x) \le 0\}.$$

These are two measurable subsets of $[a, b]$ and therefore, by (28),

$$\int_a^b f^+\, dm = \int_{E^+} f\, dm = 0 \text{ and } \int_a^b (f^-)\, dm = -\int_{E^-} f\, dm = 0.$$

According to Proposition 9 of Chapter 4, a non-negative integrable function with zero integral must vanish almost everywhere on its domain. Consequently, f^+ and f^- vanish almost everywhere on $[a, b]$, and therefore so does f. □

We have the following second version of the Fundamental Theorem of Calculus for the Lebesgue integral.

Theorem 16 *If the function $f\colon [a, b] \to \mathbf{R}$ is integrable, then*

$$\frac{d}{dx}\left[\int_a^x f\,dm\right] = f(x) \text{ for almost all } x \in (a, b). \tag{29}$$

Proof Define the function F on $[a, b]$ by $F(x) = \int_a^x f\,dm$ for all $x \in [a, b]$. Theorem 13 implies that since F is an indefinite integral, it is absolutely continuous. Therefore, by Theorem 12, F is differentiable almost everywhere on (a, b) and its derivative F' is integrable. According to the preceding lemma, to show that the integrable function $F' - f$ vanishes almost everywhere on $[a, b]$, it suffices to show that its integral over every closed subinterval of $[a, b]$ is zero. Let $[x_1, x_2]$ be such an interval. According to Theorem 12, in the case that $[a, b]$ is replaced by $[x_1, x_2]$, and the linearity and additivity over domains properties of integration,

$$\begin{aligned}\int_{x_1}^{x_2} [F' - f]\,dm &= \int_{x_1}^{x_2} F'\,dm - \int_{x_1}^{x_2} f\,dm = F(x_2) - F(x_1) - \int_{x_1}^{x_2} f\,dm \\ &= \int_a^{x_2} f\,dm - \int_a^{x_1} f\,dm - \int_{x_1}^{x_2} f\,dm = 0.\end{aligned}$$

□

There is the following extension of the change of variables formula for Riemann integration. An interesting aspect of its proof is that, by von Neumann's Composition Theorem in the next section, if g is strictly increasing and absolutely continuous, but its inverse is not absolutely continuous, then there are integrable functions f for which $f \circ g$ is not measurable. On the other hand, the product $[f \circ g].\ g'$ is always measurable.

Theorem 17 (**Change of Variables**) *Let the function $f\colon [a, b] \to \mathbf{R}$ be integrable. If the function $g\colon [c, d] \to \mathbf{R}$ is increasing and absolutely continuous, $a = g(c)$ and $b = g(d)$, then*

$$\int_a^b f\ dm = \int_c^d [f \circ g] \cdot g'\,dm. \tag{30}$$

We first prove the theorem in the case that the function f is the characteristic function of a measurable set $E \subseteq [a, b]$, so that the change of variables formula becomes

$$m(E) = \int_c^d [\chi_E \circ g] \cdot g'\,dm. \tag{31}$$

Lemma 18 *Let $g\colon [c, d] \to \mathbf{R}$ be increasing and absolutely continuous, $a = g(c)$ and $b = g(c)$. If $E \subseteq [a, b]$ is measurable, then* (31) *holds.*

Proof By possibly replacing E by $E \sim \{a, b\}$, and so leaving (31) unchanged, it is sufficient to consider E to be a measurable subset of the open interval (a, b).

Case 1: Assume that E is a G_δ set. First suppose that $E = (\alpha, \beta) \subseteq (a, b)$. Then $\chi_E \circ g = \chi_{(g^{-1}(\alpha),\, g^{-1}(\beta))}$. Consequently, by Theorem 12, since g is absolutely continuous,

$$\int_c^d [\chi_E \circ g] \cdot g' \, dm = \int_{g^{-1}(\alpha)}^{g^{-1}\beta} g' \, dm = \beta - \alpha = m(E),$$

so (31) holds if E is an open interval. Now suppose that $E = \mathcal{O}$, an open subset of (a, b), and $\mathcal{O} = \bigcup_{k=1}^{\infty} I_k$, the disjoint, countable union of open intervals. Observe that

$$[\chi_{\mathcal{O}} \circ g] \cdot g' = \sum_{k=1}^{\infty} [\chi_{I_k} \circ g] \cdot g' \text{ almost everywhere on } [c, d].$$

Since (31) holds for each interval I_k and $[\chi_{I_k} \circ g] \cdot g' \geq 0$ almost everywhere on $[c, d]$, it follows from the Monotone Convergence Theorem and the countable additivity of measure that (31) holds if E is open. Now consider E to be a general G_δ set, and $E = \bigcap_{n=1}^{\infty} \mathcal{O}_n$, where $\{\mathcal{O}_n\}$ is a countable, descending collection of open subsets of (a, b). Observe that $\{[\chi_{\mathcal{O}_n} \circ g] \cdot g'\} \to [\chi_E \circ g] \cdot g'$ pointwise almost everywhere on $[c, d]$. Moreover, for each n,

$$0 \leq [\chi_{\mathcal{O}_n} \circ g] \cdot g' \leq g' \text{ almost everywhere on } [c, d],$$

and g', being the derivative of an increasing function, is integrable over $[c, d]$. Since (31) holds for each $\mathcal{O}_n$, by the continuity of measure and the Dominated Convergence Theorem,

$$m(E) = \lim_{n \to \infty} m(\mathcal{O}_n) = \lim_{n \to \infty} \int_c^d [\chi_{\mathcal{O}_n} \circ g] \cdot g' \, dm = \int_c^d [\chi_E \circ g] \cdot g' \, dm.$$

Case 2: Assume that $m(E) = 0$. We claim that

$$[\chi_E \circ g] \cdot g' = 0 \text{ almost everywhere on } [c, d]. \tag{32}$$

If this claim is verified, then the function $[\chi_E \circ g] \cdot g'$ is non-negative and measurable, so the right-hand side of (31) is properly defined, and each side is zero so that (31) holds. To verify the claim, let $\lambda > 0$, and let $\mathcal{O}$ be open and $E \subseteq \mathcal{O} \subseteq (a, b)$. Observe that by Chebychev's Inequality and case 1,

$$m\{x \in [c, d] \mid \chi_{\mathcal{O}}(g(x)) \cdot g'(x) \geq \lambda\} \leq \frac{1}{\lambda} m(\mathcal{O}).$$

Also observe the inclusion

$$\{x \in [c, d] \mid \chi_E(g(x)) \cdot g'(x) \geq \lambda\} \subseteq \{x \in [c, d] \mid \chi_{\mathcal{O}}(g(x)) \cdot g'(x) \geq \lambda\}.$$

Therefore,

$$m^*\{x \in [c, d] \mid \chi_E(g(x)) \cdot g'(x) \geq \lambda\} \leq \frac{1}{\lambda} m(\mathcal{O}).$$

Let $\epsilon > 0$. Since $m(E) = 0$, there is an open set $\mathcal{O} \subseteq (a, b)$ containing E for which $m(\mathcal{O}) < \epsilon$, so that

$$m^*\{x \in [c, d] \mid \chi_E(g(x)) \cdot g'(x) \geq \lambda\} \leq \frac{1}{\lambda} \cdot \epsilon.$$

This holds for every $\epsilon > 0$, and therefore

$$m^* \{x \in [c, d] \mid \chi_E(g(x)) \cdot g'(x) \geq \lambda\} = 0.$$

This holds for every $\lambda > 0$ and consequently, by the countable monotonicity of outer-measure, claim (32) is verified. But each measurable set is a G_δ set from which a set of measure zero has been excised, and so the proof follows from consideration of the above two cases. $\square$

Proof of the Change of Variables Theorem In view of the linearity of integration, by the preceding lemma, (30) holds if $f\colon [a, b] \to \mathbf{R}$ is a simple function. Without loss of generality, we may assume that $f \geq 0$. According to the Simple Approximation Theorem, there is an increasing sequence of non-negative, simple functions $\{\psi_n\colon [a, b] \to [0, \infty]\}$ that converges pointwise on $[a, b]$ to f. We deduce from the preceding lemma that each function $[\psi_n \circ g] \cdot g'\colon [c, d] \to [0, \infty]$ is measurable. Moreover, the sequence $\{[\psi_n \circ g] \cdot g'\colon [a, b] \to \mathbf{R}\}$ is increasing, since the sequence $\{\psi_n \circ g\colon [c, d] \to \mathbf{R}\}$ is increasing and $g' \geq 0$ almost everywhere on $[c, d]$, and it converges pointwise almost everywhere on $[c, d]$ to $[f \circ g] \cdot g'$. We appeal to the Monotone Convergence Theorem to conclude that

$$\int_c^d [f \circ g] \cdot g' \, dm = \lim_{n\to\infty} \int_c^d [\psi_n \circ g] \cdot g' \, dm = \lim_{n\to\infty} \int_a^b \psi_n \, dm = \int_a^b f \, dm. \qquad \square$$

There is the following extension of integration by parts for the Riemann integration.

Theorem 19 (**Integration by Parts**) *If the functions $h\colon [a, b] \to \mathbf{R}$ and $g\colon [a, b] \to \mathbf{R}$ are absolutely continuous, then*

$$\int_a^b g \cdot h' \, dm = [g(b)h(b) - g(a)h(a)] - \int_a^b h \cdot g' \, dm. \tag{33}$$

Proof Since $g \cdot h\colon [a, b] \to \mathbf{R}$ is absolutely continuous, according to Theorem 12,

$$\int_a^b [g \cdot h]' \, dm = g(b)h(b) - g(a)h(a).$$

On the other hand,

$$\int_a^b [g \cdot h]' \, dm = \int_a^b [g \cdot h' + h \cdot g'] \, dm = \int_a^b g \cdot h' \, dm + \int_a^b h \cdot g' \, dm.$$

We deduce (33) from these two equalities. $\square$

In the last half of the nineteenth century, functions were discovered which had very strange and unexpected properties. For instance, in 1872, Karl Weierstrass presented a continuous function on an open interval that failed to be differentiable at any point[4]. Such discoveries were regarded by many mathematicians as the propagation of anarchy and chaos

[4]A geometric construction of such a function, as an infinite sum of saw-tooth functions, due to I. van der Waerden, may be found in Chapter 9 of Patrick Fitzpatrick's *Advanced Calculus*.

where past generations had sought order and harmony. In the first decade of the twentieth century, the Lebesgue integral, Henri Lebesgue's theorem on the differentiability of monotone functions, and Giuseppe Vitali's consideration of absolutely continuous functions made it possible to harmoniously reunite the two fundamental concepts of calculus, the definite integral and the antiderivative (Theorems 12 and 16), which had appeared to be forever separated as soon as integration went outside the domain of continuous functions[5].

PROBLEMS

48. The Cantor-Lebesgue function φ is continuous and increasing on $[0, 1]$. Conclude from Theorem 12 that φ is not absolutely continuous on $[0, 1]$. Compare this reasoning with that proposed in Problem 42.

49. Let $f\colon [a, b] \to \mathbf{R}$ be continuous and differentiable almost everywhere on (a, b). Show that
$$\int_a^b f'\, dm = f(b) - f(a)$$
if and only if
$$\int_a^b \left[\lim_{n \to \infty} \mathrm{Diff}_{1/n} f\right] dm = \lim_{n \to \infty} \left[\int_a^b \mathrm{Diff}_{1/n} f\right] dm.$$

50. Let $f\colon [a, b] \to \mathbf{R}$ be continuous and differentiable almost everywhere on (a, b). Show that if $\{\mathrm{Diff}_{1/n} f\}$ is uniformly integrable over $[a, b]$, then
$$\int_a^b f'\, dm = f(b) - f(a).$$

51. Let $f\colon [a, b] \to \mathbf{R}$ be continuous and differentiable almost everywhere on (a, b). Suppose that there is a non-negative integrable function $g\colon [a, b] \to \mathbf{R}$ for which
$$\left|\mathrm{Diff}_{1/n} f\right| \le g \text{ on } [a, b] \text{ for all } n.$$
Show that
$$\int_a^b f'\, dm = f(b) - f(a).$$

52. If $f\colon [a, b] \to \mathbf{R}$ is absolutely continuous, show that f is Lipschitz if and only if there is a $c \ge 0$ for which $|f'| \le c$ almost everywhere on $[a, b]$.

53. Let $f\colon [a, b] \to \mathbf{R}$ be of bounded variation, and define $v(x) = TV(f_{[a, x]})$ for all $x \in [a, b]$.
    (i) Show that $|f'| \le v'$ almost everywhere on $[a, b]$, and deduce from this that
$$\int_a^b |f'|\, dm \le TV(f).$$
    (ii) Show that the above is an equality if and only if f is absolutely continuous on $[a, b]$.
    (iii) Compare parts (i) and (ii) with Corollaries 4 and 12, respectively.

[5]An elaboration of these remarks and further historical commentary may be found in the introduction to Stanislaw Saks' classic book *Theory of the Integral.*

54. Construct an absolutely continuous, strictly increasing function f: $[0, 1] \to \mathbf{R}$ for which $f' = 0$ on a set of positive measure. (Suggestion: Let E be the relative complement in $[0, 1]$ of a generalized Cantor set of positive measure and f the indefinite integral of χ_E. See Problem 39 of Chapter 10 for the construction of such a Cantor set.)

55. For f and g as in the statement of Theorem 17, is it possible to establish the change of variables formula (30) by directly verifying that

$$\frac{d}{dx}\left[\int_a^{g(x)} f\,dm - \int_c^x [f \circ g] \cdot g'\,dm\right] = 0 \text{ for almost all } x \in (c, d)?$$

6.6 MEASURABILITY: IMAGES OF SETS, COMPOSITIONS OF FUNCTIONS

We proved that if f: $\mathbf{R} \to \mathbf{R}$ is continuous and g: $E \to \mathbf{R}$ is measurable, then the composition $f \circ g$ is measurable. Here, we consider the composition of continuous and measurable functions, but compose in the opposite order. The principal goal of this section is to prove the following theorem of von Neumann.[6] In doing so, we provide yet another characterization of absolute continuity: an increasing, continuous function f: $[a, b] \to \mathbf{R}$ is absolutely continuous if and only if maps measurable sets to measurable sets.

von Neumann's Composition Theorem *Let f: $[a, b] \to \mathbf{R}$ be a strictly increasing, continuous function. Then the composition $g \circ f$: $[a, b] \to \mathbf{R}$ is a measurable function whenever g: $\mathbf{R} \to \mathbf{R}$ is a measurable function if and only if f^{-1}: $[f(a), f(b)] \to \mathbf{R}$ is absolutely continuous.*

Lemma 20 *For a function f: $E \to \mathbf{R}$, the composition $g \circ f$: $E \to \mathbf{R}$ is a measurable function whenever g: $\mathbf{R} \to \mathbf{R}$ is a measurable function if and only if $f^{-1}(A)$ is a measurable set whenever A is a measurable set.*

Proof Observe that, for each c,

$$(g \circ f)^{-1}(E_c) = f^{-1}(g^{-1}(E_c)) \text{ where } E_c = (-\infty, c).$$

Therefore, if $f^{-1}(A)$ is a measurable set whenever A is measurable and g: $\mathbf{R} \to \mathbf{R}$ is measurable, then the composition $g \circ f$ also is measurable. On the other hand, if there is a measurable function g: $\mathbf{R} \to \mathbf{R}$ for which $g \circ f$ is not measurable, then there is a c for which $(g \circ f)^{-1}(E_c)$ is not a measurable set, in which case $g^{-1}(E_c)$ is a measurable set whose image under f^{-1} is not a measurable set. $\square$

A function f: $E \to \mathbf{R}$ is said to preserve measurable sets if it maps measurable sets to measurable sets, and said to preserve sets of measure zero if it maps sets of measure zero to sets of measure zero.

Lemma 21 *If the function f: $[a, b] \to \mathbf{R}$ preserves measurable sets, then it preserves sets of measure zero. Moreover, if f is continuous and preserves sets of measure zero, then it preserves measurable sets.*

[6] This is Theorem 9.27 in his book, *Functional Operators, Volume 1: Measure and Integrals.*

Proof First assume that f preserves measurable sets. To show that it preserves sets of measure zero, we argue by contradiction. Indeed, otherwise, there is a subset A of E with $m(A) = 0$ and $m^*(f(A)) > 0$. By Vitali's non-measurability theorem, there is a subset B of $f(A)$ that is non-measurable. But $f^{-1}(B) \cap A$ is a subset of a set of measure zero, so it is measurable, while its image under f is B, which is non-measurable. Therefore, f preserves sets of measure zero. Now, assume that f is continuous and preserves sets of measure zero. Since each measurable set is the union of a set of measure zero and an F_σ set, to show that f preserves measurable sets, it suffices to show it preserves F_σ sets. However, by the Bolzano-Weierstrass Theorem, since $[a, b]$ is closed and bounded, f preserves closed sets, and so also preserves F_σ sets. □

Lemma 22 *If f: $[a, b] \to \mathbf{R}$ is an increasing, absolutely continuous function, then it preserves measurable sets.*

Proof In view of the preceding lemma, since f is continuous, it suffices to show that f preserves sets of measure zero. Assume that $E \subseteq (a, b)$ has $m(E) = 0$. Let $\epsilon > 0$. It follows from the definition of absolute continuity that there is a $\delta > 0$ such that for $\{(a_k, b_k)\}_{k=1}^{\infty}$ a disjoint, countable collection of open subintervals of (a, b),

$$\text{if } \sum_{k=1}^{\infty}[b_k - a_k] < \delta, \text{ then } \sum_{k=1}^{\infty}|f(b_k) - f(a_k)| < \epsilon.$$

Since $m(E) = 0$, by the regularity of measure, there is an open set $\mathcal{O}$, with $E \subseteq \mathcal{O} \subseteq (a, b)$ and $m(\mathcal{O}) < \delta$. There is a disjoint, countable collection of open intervals, $\{(a_k, b_k)\}_{k=1}^{\infty}$, whose union is $\mathcal{O}$. Since f is increasing, for each k, $f(a_k, b_k) \subseteq [f(a_k), f(b_k)]$, and therefore, by the countable monotonicity of outer-measure and the choice of δ,

$$m^*(f(E)) \leq \bigcup_{k=1}^{\infty}[f(b_k) - f(a_k] < \epsilon.$$

This holds for all $\epsilon > 0$ and so $m(f(E)) = 0$. □

Lemma 23 *Let f: $[a, b] \to \mathbf{R}$ be an increasing function that preserves measurable sets. Then, if $\lambda > 0$ and E is a measurable subset of (a, b),*

$$m(f(E_\lambda)) \leq \lambda \cdot m(E_\lambda), \text{ where } E_\lambda = \{x \in E \mid f'(x) < \lambda\}. \tag{34}$$

Proof Since f' is a measurable function, E_λ is a measurable set, and therefore, by assumption, so is $f(E_\lambda)$. Let $\mathcal{F}$ be the collection of closed intervals $[u, v] \subseteq (a, b)$ for which $f(v) - f(u) < \lambda \cdot [v - u]$. Since $f' < \lambda$ on E_λ, $\mathcal{F}$ is a Vitali covering of E_λ. Let $\epsilon > 0$. According to the Vitali Covering Lemma, there is a subset E_0 of E_λ, with $m(E_0) = 0$, together with a countable, disjoint subcollection $\{[a_k, b_k]\}_{k=1}^{\infty}$ of $\mathcal{F}$ for which

$$E_\lambda = E_0 \cup \bigcup_{k=1}^{\infty}[E_\lambda \cap [a_k, b_k]] \text{ and } \sum_{k=1}^{\infty}[b_k - a_k] < m(E_\lambda) + \epsilon.$$

Observe that

$$f(E_\lambda) = f(E_0) \cup \bigcup_{k=1}^{\infty} f(E_\lambda \cap [a_k, b_k]).$$

Since f preserves measurable sets, it preserves sets of measure zero, so that $m(f(E_0)) = 0$. On the other hand, for each k, since f is increasing

$$f(E_\lambda \cap [a_k, b_k]) \subseteq f([a_k, b_k]) \subseteq [f(a_k), f(b_k)],$$

and therefore, since $[a_k, b_k]$ belongs to $\mathcal{F}$,

$$m(f(E_\lambda \cap [a_k, b_k])) \le f(b_k) - f(a_k) < \lambda \cdot [b_k - a_k].$$

By the countable monotonicity of m, we sum these inequalities to obtain

$$m(f(E_\lambda)) \le m(f(E_0)) + \sum_{k=1}^{\infty} m(f(E_\lambda \cap [a_k, b_k])) < \lambda \cdot \sum_{k=1}^{\infty} [b_k - a_k] < \lambda \cdot [m(E_\lambda) + \epsilon].$$

This holds for all $\epsilon > 0$, and so (34) holds. □

Lemma 24 *Let $f\colon [a, b] \to \mathbf{R}$ be an increasing, continuous function that preserves measurable sets. Then*

$$f(b) - f(a) \le \int_a^b f' \, dm. \tag{35}$$

Proof According to the Lebesgue Differentiation Theorem, if $E \subseteq (a, b)$ is the set of points at which f is differentiable, then $m([a, b] \sim E) = 0$, and so E is measurable and $\int_a^b f' \, dm = \int_E f' \, dm$. On the other hand, since f preserves measurable sets, it also preserves sets of measure zero, and therefore $m(f([a, b])) = m(f(E))$. But f is increasing and continuous, and therefore, by the Intermediate Value Theorem, $f([a, b]) = [f(a), f(b)]$, so that

$$f(b) - f(a) = m(f([a, b])) = m(f(E)).$$

Consequently, to verify (35) is to verify that

$$m(f(E)) \le \int_E f' dm. \tag{36}$$

Let $\epsilon > 0$. For each $k \ge 1$, define

$$E_k = \{x \in E \mid \ (k-1) \cdot \epsilon \le f'(x) < k \cdot \epsilon.\}$$

Since $f'\colon E \to \mathbf{R}$ is measurable, real-valued, and non-negative, $\{E_k\}_{k=1}^{\infty}$ is a measurable partition of E. By the preceding lemma, for each k,

$$m(f(E_k)) \le k \cdot \epsilon \cdot m(E_k).$$

Consequently, by the countable monotonicity of m,

$$m(f(E)) \le \sum_{k=1}^{\infty} m(f(E_k)) \le \sum_{k=1}^{\infty} k \cdot \epsilon \cdot m(E_k). \tag{37}$$

Now, for each k, by the monotonicity of integration,

$$k \cdot \epsilon \cdot m(E_k) - \epsilon \cdot m(E_k) \leq \int_{E_k} f' \, dm \leq k \cdot \epsilon \cdot m(E_k).$$

By the countable additivity of measure,

$$b - a = m([a, b] \sim E) + \sum_{k=1}^{\infty} m(E_k) = \sum_{k=1}^{\infty} m(E_k),$$

so that, summing the above inequalities and using the countable additivity over domains of integration, we have

$$\sum_{k=1}^{\infty} k \cdot \epsilon \cdot m(E_k) - \epsilon \cdot [b - a] \leq \int_E f' \, dm \leq \sum_{k=1}^{\infty} k \cdot \epsilon \cdot m(E_k).$$

Therefore,

$$\lim_{\epsilon \to 0^+} \sum_{k=1}^{\infty} k \cdot \epsilon \cdot m(E_k) = \int_E f' \, dm,$$

and so (36) follows from (37). □

Theorem 25 *An increasing, continuous function f: $[a, b] \to \mathbf{R}$ preserves measurable sets if and only if it is absolutely continuous.*

Proof We have shown that an increasing, absolutely continuous function preserves measurable sets. To prove the converse, assume that the increasing, continuous function f preserves measurable sets. According to Corollary 14, f is absolutely continuous if and only if

$$f(b) - f(a) = \int_a^b f' \, dm.$$

This equality follows from the inequalities provided by Corollary 5 and the preceding lemma. □

The proof of von Neumann's Composition Theorem follows from this theorem, applied to the increasing, continuous function f^{-1}: $[f(a), f(b)] \to \mathbf{R}$, together with Lemma 20.

Corollary 26 *If f: $[a, b] \to \mathbf{R}$ is a continuous function that has a continuous, positive derivative on (a, b), then the composition $g \circ f$ is a measurable function whenever the function g: $\mathbf{R} \to \mathbf{R}$ is measurable.*

Proof By the Mean Value Theorem, f: $[a, b] \to \mathbf{R}$ is strictly increasing, and by the Intermediate Value Theorem, $f([a, b]) = [f(a), f(b)]$. Moreover, f^{-1}: $[f(a), f(b)] \to \mathbf{R}$ continuous and has a continuous positive derivative on $(f(a), f(b))$. Let $f(a) < c < d < f(b)$. The function f^{-1}: $[c, d] \to \mathbf{R}$ has a continuous, bounded derivative and so it is Lipschitz, and therefore absolutely continuous. But f^{-1}: $[f(a), f(b)] \to \mathbf{R}$, being increasing, continuous, and absolutely continuous of each subinterval $[c, d]$ of $(f(a), f(b))$, is also absolutely continuous on $[f(a), f(b)]$. The conclusion follows from von Neumann's Composition Theorem. $\square$

Example Let φ: $[0, 1] \to \mathbf{R}$ be the Cantor-Lebesgue function and define h: $[0, 1] \to \mathbf{R}$ by $h(x) = x + \varphi(x)$. Then h is strictly increasing and continuous, and therefore so is its inverse h^{-1}: $[0, 2] \to \mathbf{R}$ which we denote by f. Since φ is not absolutely continuous, neither is $h = f^{-1}$. It follows from von Neumann's Composition Theorem that there is an integrable, not just mesurable (see Problem 59), function g: $\mathbf{R} \to \mathbf{R}$ for which the composition $g \circ f$ is not measurable. On the other hand, in the proof, in the preceding section, of the change of variables theorem for Lebesgue integration, we proved that the product $[g \circ f] \cdot g'$ is measurable.

PROBLEMS

56. Let f: $[a, b] \to \mathbf{R}$ preserve sets of measure zero. Show that the composition $g \circ f$ is measurable whenever g: $\mathbf{R} \to \mathbf{R}$ is measurable if this holds for simple functions g.
57. Let f: $[a, b] \to \mathbf{R}$ be an increasing, absolutely continuous function. Prove that for each measurable subset E of $[a, b]$,
$$m(f(E)) = \int_E f' \, dm.$$
58. Find an increasing function f: $[a, b] \to \mathbf{R}$ that preserves measurable sets for which inequality (35) does not hold.
59. Prove that if f: $[a, b] \to \mathbf{R}$ is not absolutely continuous, then there is an integrable (not just measurable) function g: $\mathbf{R} \to \mathbf{R}$ for which $g \circ f$ is not measurable. (Suggestion: consider the proof of Lemma 20)

6.7 CONVEX FUNCTIONS

Definition *A function φ: $(a, b) \to \mathbf{R}$ is said to be* **convex** *provided that for each $\lambda \in [0, 1]$ and $x_1, x_2 \in (a, b)$,*

$$\varphi(\lambda x_1 + (1 - \lambda)x_2) \le \lambda\varphi(x_1) + (1 - \lambda)\varphi(x_2). \tag{38}$$

By considering the graph of φ, this convexity inequality can be expressed geometrically by the observation that it means that each point on the chord joining the points $(x_1, \varphi(x_1))$ and $(x_2, \varphi(x_2)$ lies above the graph of φ. Observe that for $(x_1, x_2) \subseteq (a, b)$, each $x \in (x_1, x_2)$ may be expressed as

$$x = \lambda x_1 + (1 - \lambda)x_2 \text{ where } \lambda = \frac{x_2 - x}{x_2 - x_1}.$$

Therefore, the convexity inequality may be rewritten as

$$\varphi(x) \le \left[\frac{x_2 - x}{x_2 - x_1}\right]\varphi(x_1) + \left[\frac{x - x_1}{x_2 - x_1}\right]\varphi(x_2) \text{ for } x \in (x_1, x_2).$$

Regathering terms, this inequality, in turn, may be rewritten as

$$\frac{\varphi(x)-\varphi(x_1)}{x-x_1} \leq \frac{\varphi(x_2)-\varphi(x)}{x_2-x} \text{ for } x \in (x_1,\, x_2). \tag{39}$$

Consequently, convexity may be expressed geometrically by stating that for $x \in (x_1,\, x_2)$, the slope of the chord joining the points $(x_1,\, \varphi(x_1))$ to $(x,\, \varphi(x))$ is no greater than the slope of the chord joining the points $(x,\, \varphi(x))$ to $(x_2,\, \varphi(x_2))$.

Proposition 27 *If $\varphi\colon (a,\, b) \to \mathbf{R}$ is differentiable and its derivative φ' is increasing, then φ is convex. In particular, φ is convex if it has a non-negative second derivative φ'' on $(a,\, b)$.*

Proof Let $x_1, x_2 \in (a,\, b)$ with $x_1 < x_2$, and let $x \in (x_1,\, x_2)$. It must be verified that

$$\frac{\varphi(x)-\varphi(x_1)}{x-x_1} \leq \frac{\varphi(x_2)-\varphi(x)}{x_2-x}.$$

To do so, apply the Mean Value Theorem to the restriction of φ to each of the intervals $[x_1,\, x]$ and $[x,\, x_2]$ to choose points $c_1 \in (x_1,\, x)$ and $c_2 \in (x,\, x_2)$ for which

$$\varphi'(c_1) = \frac{\varphi(x)-\varphi(x_1)}{x-x_1} \text{ and } \varphi'(c_2) = \frac{\varphi(x_2)-\varphi(x)}{x_2-x}.$$

Consequently, since φ' is increasing,

$$\frac{\varphi(x)-\varphi(x_1)}{x-x_1} = \varphi'(c_1) \leq \varphi'(c_2) = \frac{\varphi(x_2)-\varphi(x)}{x_2-x}.$$

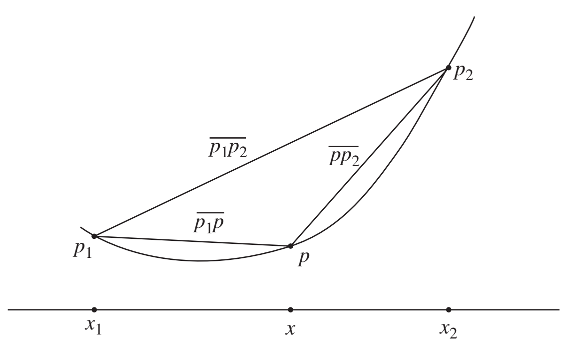
□

Example Each of the following three functions is convex since each has a non-negative second derivative:

$$\varphi(x) = x^p \text{ on } (0,\, \infty) \text{ for } p \geq 1;\;\; \varphi(x) = e^{ax} \text{ on } (-\infty,\, \infty);\;\; \varphi(x) = \ln(1/x) \text{ on } (0,\, \infty).$$

The following final geometric reformulation of convexity will be useful in establishing differentiability properties of convex functions.

The Chordal Slope Lemma *Let $\varphi\colon (a,\, b) \to \mathbf{R}$ be convex. If $x_1, x_2 \in (a,\, b)$ and $x_1 < x < x_2$, then for $p_1 = (x_1, \varphi(x_1)),\;\; p = (x, \varphi(x)),\;\; p_2 = (x_2, \varphi(x_2))$,*



$$\textit{Slope of } \overline{p_1 p} \le \textit{slope of } \overline{p_1 p_2} \le \textit{slope of } \overline{p p_2}.$$

Proof Regather terms in the inequality (39) to rewrite it in the following two equivalent forms:

$$\frac{\varphi(x_1)-\varphi(x)}{x_1-x} \le \frac{\varphi(x_2)-\varphi(x_1)}{x_2-x_1} \text{ for } x \in (x_1,\, x_2);$$
$$\frac{\varphi(x_2)-\varphi(x_1)}{x_2-x_1} \le \frac{\varphi(x_2)-\varphi(x)}{x_2-x} \text{ for } x \in (x_1,\, x_2). \qquad \square$$

For a function g on an open interval $(a,\, b)$, and point $x_0 \in (a,\, b)$, if

$$\lim_{h \to 0,\ h < 0} \frac{g(x_0+h)-g(x_0)}{h} \text{ exists and is finite,}$$

we denote this limit by $g'(x_0^-)$ and call it the left-hand derivative of g at x_0. Similarly, we define $g'(x_0^+)$ and call it the right-hand derivative of g at x_0. Of course, g is differentiable at x_0 if and only if it has left-hand and right-hand derivatives at x_0 that are equal. The continuity and differentiability properties of convex functions follow from the following lemma, whose proof follows directly from the Chordal Slope Lemma.

Lemma 28 *A convex function $\varphi\colon (a,\, b) \to \mathbf{R}$ has left-hand and right-hand derivatives at each point $x \in (a,\, b)$. Moreover, for points $u, v \in (a,\, b)$ with $u < v$, these one-sided derivatives satisfy the following inequality:*

$$\varphi'(u^-) \le \varphi'(u^+) \le \frac{\varphi(v)-\varphi(u)}{v-u} \le \varphi'(v^-) \le \varphi'(v^+). \tag{40}$$

Corollary 29 *If $\varphi\colon (a,\, b) \to \mathbf{R}$ is convex, then φ is Lipschitz, and therefore absolutely continuous, on each closed, bounded subinterval $[c,\, d]$ of (a, b).*

Proof According to the preceding lemma, for $c \le u < v \le d$,

$$\varphi'(c^+) \le \varphi'(u^+) \le \frac{\varphi(v)-\varphi(u)}{v-u} \le \varphi'(v^-) \le \varphi'(d^-) \tag{41}$$

and therefore

$$|\varphi(u)-\varphi(v)| \le M|u-v| \text{ for all } u, v \in [c, d],$$

where $M = \max\{|\varphi'(c^+)|, |\varphi'(d^-)|\}$. Consequently, the restriction of φ to $[u,\, v]$ is Lipschitz. A Lipschitz function on a closed, bounded interval is absolutely continuous. $\square$

It follows from the above corollary and Corollary 6 that a convex function defined on an open interval is differentiable almost everywhere on its domain. In fact, much more can be said.

Theorem 30 *If $\varphi\colon (a,\, b) \to \mathbf{R}$ is convex, then it is differentiable except at a countable number of points and its derivative φ' is an increasing function.*

Proof By the inequalities (40) we conclude that the functions

$$x \mapsto f'(x^-) \text{ and } x \mapsto f'(x^+)$$

are increasing real-valued functions on (a, b). But, according to Theorem 1, an increasing real-valued function is continuous except at a countable number of points. Therefore, except on a countable subset $\mathcal{C}$ of (a, b), both the left-hand and right-hand derivatives of φ are continuous. Let $x_0 \in (a, b) \sim \mathcal{C}$. Choose a sequence $\{x_n\}$ of points greater than x_0 that converges to x_0. Apply Lemma 28, with $x_0 = u$ and $x_n = v$, and take limits to conclude that

$$\varphi'(x_0^-) \leq \varphi'(x_0^+) \leq \varphi'(x_0^-).$$

Then $\varphi'(x_0^-) = \varphi'(x_0^+)$ so that φ is differentiable at x_0. To show that φ' is an increasing function on $(a, b) \sim \mathcal{C}$, let $u, v \in (a, b) \sim \mathcal{C}$ with $u < v$. Then by Lemma 28,

$$\varphi'(u) \leq \frac{\varphi(v) - \varphi(u)}{v - u} \leq \varphi'(v). \qquad \square$$

Let $\varphi(a, b) \to \mathbf{R}$ be convex and $x_0 \in (a, b)$. For a real number m, the line $y = m(x - x_0) + \varphi(x_0)$, which passes through the point $(x_0, \varphi(x_0))$, is called a **supporting line** at x_0 for the graph of φ provided that this line always lies below the graph of φ, that is, if

$$\varphi(x) \geq m(x - x_0) + \varphi(x_0) \text{ for all } x \in (a, b).$$

It follows from Lemma 28 that such a line is supporting if and only if its slope m lies between the left- and right-hand derivatives of φ at x_0. Therefore, in particular, there always is at least one supporting line at each point. This concept enables us to give a short proof of the following inequality.

Jensen's Inequality *Let $\varphi\colon \mathbf{R} \to \mathbf{R}$ be convex. Assume that both $f\colon [0, 1] \to \mathbf{R}$ and $\varphi \circ f\colon [0, 1] \to \mathbf{R}$ are integrable. Then*

$$\varphi\left(\int_0^1 f(x)\,dx\right) \leq \int_0^1 (\varphi \circ f)(x)\,dx. \tag{42}$$

Proof Define $\alpha = \int_0^1 f(x)\,dx$. Choose m to lie between the left–hand and right–hand derivative of φ at the point α. Then $y = m(t - \alpha) + \varphi(\alpha)$ is the equation of a supporting line at $(\alpha, \varphi(\alpha))$ for the graph of φ. Hence

$$\varphi(t) \geq m(t - \alpha) + \varphi(\alpha) \text{ for all } t \in \mathbf{R}.$$

Since f is integrable over $[0, 1]$, it is finite a.e. on $[0, 1]$ and therefore, substituting $f(x)$ for t in this inequality,

$$\varphi(f(x)) \geq m(f(x) - \alpha) + \varphi(\alpha) \text{ for almost all } x \in [0, 1].$$

Integrate across this inequality, using the monotonicity of the Lebesgue integral and the assumption that both f and $\varphi \circ f$ are integrable over $[a, b]$, to obtain

$$\begin{aligned} \int_0^1 \varphi(f(x))\,dx &\geq \int_0^1 [m(f(x) - \alpha) + \varphi(\alpha)]\,dx \\ &= m\left[\int_0^1 f(x)\,dx - \alpha\right] + \varphi(\alpha) = \varphi(\alpha). \end{aligned} \qquad \square$$

A few words are in order regarding the assumption, for Jensen's Inequality, of the integrability of $\varphi \circ f \colon [0,\, 1] \to \mathbf{R}$. We have shown that a convex function is continuous and therefore the composition $\varphi \circ f$ is measurable if φ is convex and f is integrable. If $\varphi \geq 0$, then it is unnecessary to assume that $\varphi \circ f$ is integrable, since equality (42) trivially holds if the right–hand integral equals $+\infty$. In the case that φ fails to be non-negative, if there are $c_1 \geq 0, c_2 \geq 0$ for which

$$|\varphi(s)| \leq c_1 + c_2\,|s| \text{ for all } s \in \mathbf{R}, \tag{43}$$

then it follows from the integral comparison test that $\varphi \circ f$ is integrable if f is integrable.

PROBLEMS

60. Show that $\varphi \colon (a,\, b) \to \mathbf{R}$ is convex if and only if for points $x_1, \ldots, x_n$ in $(a,\, b)$ and non-negative numbers $\lambda_1, \ldots, \lambda_n$ such that $\sum_{k=1}^{n} \lambda_k = 1$,
$$\varphi\left(\sum_{k=1}^{n} \lambda_k\, x_k\right) \leq \sum_{k=1}^{n} \lambda_k\, \varphi(x_k).$$
Use this to directly prove Jensen's Inequality for f a simple function.
61. Show that a continuous function $\varphi \colon (a,\, b) \to \mathbf{R}$ is convex if and only if
$$\varphi\left(\frac{x_1 + x_2}{2}\right) \leq \frac{\varphi(x_1) + \varphi(x_2)}{2} \text{ for all } x_1, x_2 \in (a,\, b).$$
62. A function $\varphi \colon I \to \mathbf{R}$ on a general interval I is said to be convex provided that it is continuous and (38) holds for all $x_1, x_2 \in I$. Is a convex function on a closed, bounded interval $[a,\, b]$ necessarily Lipschitz on $[a,\, b]$?
63. Let $\varphi \colon (a,\, b) \to \mathbf{R}$ have a second derivative at each point in $(a,\, b)$. Show that φ is convex if and only if φ'' is non-negative.
64. Let $a \geq 0$ and $b > 0$. Show that the function $\varphi(t) = (a + bt)^p$ is convex on $[0, \infty)$ for $1 \leq p < \infty$.
65. For what functions φ is Jensen's Inequality always an equality?
66. State and prove a version of Jensen's Inequality on a general closed, bounded interval $[a,\, b]$.
67. Let $f \colon [0, 1] \to \mathbf{R}$ be integrable. Show that
$$\exp\left[\int_0^1 f(x)\, dx\right] \leq \int_0^1 \exp(f(x))\, dx$$
68. Let $\{\alpha_n\}$ be a sequence of non-negative numbers whose sum is 1 and $\{\zeta_n\}$ a sequence of positive numbers. Show that
$$\prod_{n=1}^{\infty} \zeta_n^{\alpha_n} \leq \sum_{n=1}^{\infty} \alpha_n\, \zeta_n.$$
69. Let $g \colon [0, 1] \to \mathbf{R}$ be positive and measurable. Show that $\log\left(\int_0^1 g(x)\, dx\right) \geq \int_0^1 \log\left(g(x)\right) dx$ whenever each side is defined.
70. Let $\varphi \colon \mathbf{R} \to \mathbf{R}$ be continuous. Show that if there are $c_1 \geq 0, c_2 \geq 0$ for which (43) holds, then $\varphi \circ f \colon [0,\, 1] \to \mathbf{R}$ is integrable if $f \colon [0,\, 1] \to \mathbf{R}$ is integrable.

CHAPTER 7

The L^p Spaces: Completeness and Approximation

Contents

7.1 Normed Linear Spaces 123
7.2 The Inequalities of Young, Hölder, and Minkowski 125
7.3 L^p is Complete: Rapidly Cauchy Sequences and the Riesz-Fischer Theorem 130
7.4 Approximation and Separability 134

General normed linear spaces are considered, including those spaces that are complete and those that are separable. This provides a context in which to examine linear spaces of measurable spaces that possess certain integrability properties. For $E \subseteq \mathbf{R}$ measurable and $1 \leq p < \infty$, define $L^p(E)$ to be the collection of measurable functions f: $E \to \mathbf{R}$ for which $\int_E |f|^p \, dm < \infty$ and let $\| \cdot \|_p = \left[\int_E |f|^p \, dm\right]^{1/p}$. Hölder's Inequality is proven, and from it we derive the triangle inequality for the norm $\| \cdot \|_p$, which is called Minkowski's Inequality. Then the Riesz-Fischer Theorem is proven, which states that $L^p(E)$ is complete, and finally we prove that this space is separable.

7.1 NORMED LINEAR SPACES

On a linear space, in order to study such fruitful concepts as convergence of a sequence and continuity of functionals and operators, the absolute value that is defined for real numbers is extended to a corresponding concept called a norm.

Definition *Let X be a linear space. A real-valued functional $\| \cdot \|$ on X is called a* **norm** *provided that for each $u, v \in X$ and each real number α,*

(The Triangle Inequality)

$$\|u + v\| \leq \|u\| + \|v\|$$

(Positive Homogeneity)

$$\|\alpha u\| = |\alpha| \|u\|$$

(Non-negativity)

$$\|u\| \geq 0 \text{ and } \|u\| = 0 \text{ if and only if } u = 0.$$

By a **normed linear space** is meant a linear space together with a norm.

Definition *A subset A of a normed linear space X is said to be* **bounded** *in X provided that there is an $M \geq 0$ for which*

$$\|u\| \leq M \text{ for all } u \in A.$$

A sequence $\{u_n\}$ *in* X *said to* **converge** *in* X *to* $u \in X$ *provided that* $\lim_{n\to\infty} \|u_n - u\| = 0$.

Throughout this chapter, E denotes a measurable subset of $\mathbf{R}$, and X denotes a normed linear space, with norm $\|\cdot\|$.

Example [The Normed Linear Function Space $C(K)$] Let $K \subseteq \mathbf{R}$ be compact. The linear space of continuous real-valued functions on K is denoted by $C(K)$. Since a continuous function on a compact set takes a maximum value, for $f \in C(K)$, we may define

$$\|f\|_{\max} = \max_{x \in K} |f(x)|,$$

and leave it as an exercise to show that this a norm, called the **maximum norm**. Convergence with respect to this norm is uniform convergence.

Example [The Linear Function Spaces $L^p(E), 1 \le p < \infty$] Fix $1 \le p < \infty$ and define $L^p(E)$ to be the collection of measurable functions $f\colon E \to \overline{\mathbf{R}}$ for which $\int_E |f|^p \, dm < \infty$. Observe that for any two numbers a and b,

$$|a+b| \le |a| + |b| \le 2\max\{|a|, |b|\},$$

and therefore

$$|a+b|^p \le 2^p \cdot (|a|^p + |b|^p). \tag{1}$$

An integrable function is finite almost everywhere and so a function in $L^p(E)$ may be identified with a real-valued function. Therefore, by the above inequality, together with the linearity of integration, $L^p(E)$ is a linear space. For $f \in L^p(E)$, define

$$\|f\|_p = \left[\int_E |f|^p \, dm\right]^{1/p}.$$

It is clear that $\|\cdot\|_p$ satisfies the homogeneity property. By Chebychev's Inequality and the continuity of measure, if $\int_E |f|^p \, dm = 0$, then f may be identified with the function $f \equiv 0$ on E. The triangle inequality for $\|\cdot\|_1$ follows from the triangle inequality for real numbers. Observe that the space $L^1(E)$ is the space of integrable functions on E, and convergence of a sequence in $L^1(E)$ is what has been called mean convergence. For $1 < p < \infty$, the triangle inequality for the norm $\|\cdot\|_p$, which is established in the next section, is called Minkowski's Inequality.

A measurable function $f\colon E \to \overline{\mathbf{R}}$ is said to be **essentially bounded** provided that there is an $M \ge 0$, called an **essential upper bound** for f, for which

$$|f| \le M \text{ almost everywhere on } E.$$

Example [The Normed Linear Function Space $L^\infty(E)$] Define $L^\infty(E)$ to be the collection of essentially bounded functions on E. After appropriate identification of each function in $L^\infty(E)$ with a real-valued function, it is clear that $L^\infty(E)$ is a linear space. For a function f in this space, define $\|f\|_\infty$ to be the infimum of the essential upper bounds of f. We claim that $\|\cdot\|_\infty$ is a norm on $L^\infty(E)$. Indeed, non-negativity and positive homogeneity are clear. From the triangle inequality for real numbers and the observation that $\|f\|_\infty$ itself is actually the smallest essential upper bound for f, we obtain the triangle inequality for $\|\cdot\|_\infty$.

Example [The Linear Sequence Spaces $\ell^p, 1 \le p \le \infty$] Fix $1 \le p < \infty$. Let ℓ^p be the collection of sequences $\{a_k\}$ for which $\sum_{k=1}^\infty |a_k|^p < \infty$ and define

$$\|\{a_k\}\|_p \equiv \left[\sum_{k=1}^\infty |a_k|^p\right]^{1/p}.$$

It follows from (1) that ℓ^p is a linear space, and clearly $\|\cdot\|_p$ possesses the non-negativity and homogeneity properties. Define ℓ^∞ to be the collection of bounded sequences and for such a sequence $\{a_k\}$ define $\|\{a_k\}\|_\infty \equiv \sup_{1\le k<\infty} |a_k|$. Arguing as in the case of general function spaces, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ define norms on ℓ^1 and ℓ^∞, respectively.

PROBLEMS

1. If $K \subseteq \mathbf{R}$ is compact, show that $\|\cdot\|_{\max}$ is a norm on $C(K)$.
2. For $f \in C[a, b]$, define $\|f\|_1 = \int_a^b |f|\, dm$. Show that this is a norm on $C[a, b]$. Also show that there is a number $c \ge 0$ for which
$$\|f\|_1 \le c\|f\|_{\max} \text{ for all } f \in C[a, b],$$
but there is no $c \ge 0$ for which
$$\|f\|_{\max} \le c\|f\|_1 \text{ for all } f \in C[a, b].$$
3. Let X be the linear space of all polynomials with real coefficients defined on $\mathbf{R}$. For a polynomial p, define $\|p\|$ to be the sum of the absolute values of the coefficients. Is this a norm?
4. For $f \in L^1[a, b]$, define $\|f\| = \int_a^b x^2|f(x)|\, dx$. Show that this is a norm on $L^1[a, b]$.
5. For $f \in L^\infty(E)$, show that $\|f\|_\infty$ is the smallest essential upper bound for f and
$$\|f\|_\infty = \min\{M \mid m\{x \in E \mid |f(x)| > M\} = 0\}.$$
6. For $0 < p < 1$, define $L^p(E)$ to be the collection of measurable functions $f\colon E \to \overline{\mathbf{R}}$ for which $\int_E |f|^p\, dm < \infty$. Show that $L^p(E)$ is a linear space.
7. Show that in a normed linear space, a convergent sequence is bounded.
8. State and prove the linearity property of convergent sequences in a normed linear space.

7.2 THE INEQUALITIES OF YOUNG, HÖLDER, AND MINKOWSKI

Definition *The conjugate of a number $p \in (1, \infty)$ is the number $q = p/(p-1)$, which is the unique number $q \in (1, \infty)$ for which $1/p + 1/q = 1$. The conjugate of 1 is defined to be ∞ and the conjugate of ∞ defined to be 1.*

Young's Inequality *For $1 < p < \infty$, q the conjugate of p, and any $a > 0, b > 0$,*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Proof The function e^x has a positive second derivative and therefore is convex, that is, for any $\lambda \in [0, 1]$ and any numbers u and v,

$$e^{\lambda u + (1-\lambda)v} \leq \lambda e^u + (1-\lambda)e^v.$$

In particular, setting $\lambda = 1/p$, so $1 - \lambda = 1/q, u = \ln a^p$ and $v = \ln b^q$, we have

$$e^{\frac{1}{p}\cdot \ln a^p + \frac{1}{q}\cdot \ln b^q} \leq \frac{1}{p} \cdot e^{\ln a^p} + \frac{1}{q} \cdot e^{\ln b^q}.$$

Therefore,

$$ab = e^{\ln ab} = e^{\ln a + \ln b} = e^{\frac{1}{p}\cdot \ln a^p + \frac{1}{q}\cdot \ln b^q} \leq \frac{1}{p} \cdot e^{\ln a^p} + \frac{1}{q} \cdot e^{\ln b^q} = \frac{a^p}{p} + \frac{b^q}{q}. \qquad \square$$

Hölder's Inequality *For $1 \leq p \leq \infty$ and q the conjugate of p, if $f \in L^p(E)$ and $g \in L^q(E)$, then $f \cdot g \in L^1(E)$ and*

$$\int_E |f \cdot g| \, dm \leq \|f\|_p \cdot \|g\|_q. \tag{2}$$

Moreover, for $p < \infty$, if $f \neq 0$, the function[1] $f^* = \|f\|_p^{1-p} \cdot \operatorname{sgn}(f) \cdot |f|^{p-1}$ belongs to $L^q(E)$,

$$\int_E f \cdot f^* \, dm = \|f\|_p \textit{ and } \|f^*\|_q = 1. \tag{3}$$

Proof In the case $p = 1$ or $p = \infty$, inequality (2) is clear. Moreover, if $p = 1$, $f^* = \operatorname{sgn}(f)$ and therefore (3) holds. Now consider $1 < p < \infty$. Assume that $f \neq 0$ and $g \neq 0$, for otherwise there is nothing to prove. By the positive homogenity of the norm, it is clear that if Hölder's Inequality is true if f is replaced by $f/\|f\|_p$ and g is replaced by $g/\|g\|_q$, each of which has norm 1, then it is true for general f and g. Assume that $\|f\|_p = \|g\|_q = 1$, that is,

$$\int_E |f|^p \, dm = 1 \text{ and } \int_E |g|^q \, dm = 1,$$

in which case Hölder's Inequality becomes

$$\int_E |f \cdot g| \, dm \leq 1.$$

By Young's Inequality,

$$|f \cdot g| = |f| \cdot |g| \leq \frac{|f|^p}{p} + \frac{|g|^q}{q} \text{ on } E.$$

Therefore, $f \cdot g$ is integrable and

$$\int_E |f \cdot g| \, dm \leq \frac{1}{p} \int_E |f|^p \, dm + \frac{1}{q} \int_E |g|^q \, dm = \frac{1}{p} + \frac{1}{q} = 1.$$

[1]The function $\operatorname{sgn}(x)$ takes the value 1 if $x > 0$, -1 if $x < 0$ and $\operatorname{sgn}(0) = 0$. Therefore, if $f \in L^p(E)$, then $|f| = \operatorname{sgn}(f) \cdot f$ almost everywhere on E since f is finite almost everywhere on E.

It remains to prove (3). Observe that

$$f \cdot f^* = \|f\|_p^{1-p} \cdot |f|^p \text{ on } E.$$

Consequently,

$$\int_E f \cdot f^* \, dm = \|f\|_p^{1-p} \cdot \int_E |f|^p \, dm = \|f\|_p^{1-p} \cdot \|f\|_p^p = \|f\|_p.$$

Since $q(p-1) = p$, $\|f^*\|_q = 1$. □

For $1 \leq p < \infty$ and $f \in L^p(E)$, $f \neq 0$, we call the function $f^* \in L^q(E)$ for which (3) holds the **conjugate, or dual, function of** f.

Minkowski's Inequality *For $1 \leq p \leq \infty$, if $f, g \in L^p(E)$, then $f + g \in L^p(E)$ and*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof The case $p = \infty$ is left as an exercise. Assume that $1 \leq p < \infty$. It follows from (1) that $f + g \in L^p(E)$. Assume that $f + g \neq 0$, for otherwise there is nothing to prove. So the dual function of $f + g$ is defined. By Hölder's Inequality,

$$\begin{aligned}
\|f+g\|_p &= \int_E (f+g) \cdot (f+g)^* \\
&= \int_E f \cdot (f+g)^* + \int_E g \cdot (f+g)^* \\
&\leq \|f\|_p \cdot \|(f+g)^*\|_q + \|g\|_p \cdot \|(f+g)^*\|_q \\
&= \|f\|_p + \|g\|_p.
\end{aligned}$$

□

The case $p = q = 2$ is important (see Problem 14), since the space $L^2(E)$ is an example of a Hilbert space, a class of normed linear spaces that are modeled on Euclidean space with its inner-product, from which they inherit geometric structure. We consider Hilbert spaces in Chapter 20.

The Cauchy-Schwarz Inequality *If $f, g \in L^2(E)$, then $f \cdot g \in L^1(E)$ and*

$$\int_E |f \cdot g| \, dm \leq \sqrt{\int_E f^2 \, dm} \cdot \sqrt{\int_E g^2 \, dm}.$$

Corollary 1 *If $1 \leq p \leq \infty$, then ℓ^p is a normed linear space.*

Proof For a sequence $a = \{a_k\}$ in ℓ^p, define the measurable function $\mathcal{T}(a)$ on $E = [1, \infty]$ to take the value a_k on each interval $[k, k+1)$. The $\|\cdot\|_p$ norm of the sequence $a \in \ell^p$ is the same as the $\|\cdot\|_p$ norm of the function $\mathcal{T}(a) \in L^p(E)$. Therefore, the triangle inequality for ℓ^p is inherited from the same inequality for $L^p(E)$. □

Corollary 2 *Assume that $m(E) < \infty$. If $1 \leq p_1 < p_2 \leq \infty$, then $L^{p_2}(E) \subseteq L^{p_1}(E)$. Moreover, there is a $c \geq 0$ for which*

$$\|f\|_{p_1} \leq c\|f\|_{p_2} \textit{ for all } f \in L^{p_2}(E). \tag{4}$$

Proof The case $p_2 = \infty$ is left as an exercise. Assume $p_2 < \infty$. Define $p = p_2/p_1 > 1$ and let q be the conjugate of p. Let $f \in L^{p_2}(E)$. Observe that $f^{p_1} \in L^p(E)$ and $g = \chi_E \in L^q(E)$, since $m(E) < \infty$. Apply Hölder's Inequality to obtain

$$\int_E |f|^{p_1}\, dm = \int_E |f|^{p_1} \cdot g\, dm \leq [\|f\|_{p_2}]^{p_1} \cdot \left[\int_E |g|^q\, dm\right]^{1/q} = [\|f\|_{p_2}]^{p_1}\, [m(E)]^{1/q}.$$

Take the $1/p_1$-th power of each side to obtain (4), where $c = [m(E)]^{\frac{p_2 - p_1}{p_1 p_2}}$. □

In general, if $m(E) < \infty$ and $1 \leq p_1 < p_2 \leq \infty$, then $L^{p_2}(E)$ is a proper subspace of $L^{p_1}(E)$. For example, if $f(x) = x^{-1/2}$ for $0 < x < 1$, then $f \in L^1(0, 1) \sim L^2(0, 1)$.

Example In general, if $m(E) = \infty$, then there are no inclusion relationships among the $L^p(E)$ spaces. For instance, define

$$f(x) = \frac{x^{-1/2}}{1 + |\ln x|} \text{ for } x > 0.$$

We leave it as an exercise to show that $f \in L^p(0, \infty)$ if and only if $p = 2$.

As the following examples show, for a sequence $\{f_n\}$ in $L^p(E)$, pointwise convergence to a function in $L^p(E)$ does not imply convergence in $L^p(E)$.

Example Fix $1 \leq p < \infty$. For each n, define the function $g_n\colon [0, 1] \to \mathbf{R}$ by $g_n = n^{1/p}\chi_{(0,\, 1/n]}$. The sequence $\{g_n\colon [0, 1] \to \mathbf{R}\}$ converges pointwise on $[0, 1]$ to $g \equiv 0$ but does not converge to g in $L^p[0, 1]$. For each n, define the function $h_n\colon [0, \infty) \to \mathbf{R}$ by $h_n = \chi_{[n,\, n+1]}$. The sequence $\{h_n\colon [0, \infty) \to \mathbf{R}\}$ converges pointwise to $h \equiv 0$ but does not converge to h in $L^p[0, \infty)$.

Theorem 3 *Let $1 \leq p < \infty$. If $\{f_n\}$ is a sequence in $L^p(E)$ that converges pointwise almost everywhere on E to the function $f \in L^p(E)$, then $\{f_n\} \to f$ in $L^p(E)$ if and only if the sequence $\{|f_n|^p\}$ is uniformly integrable and tight.*

Proof The sequence $\{|f - f_n|^p\}$ converges pointwise almost everywhere to $g \equiv 0$. According to Corollary 7 of Chapter 5, a consequence of the Vitali Convergence Theorem, $\lim_{n \to \infty} \int_E |f_n - f|^p\, dm = 0$ if and only if $\{|f_n - f|^p\}$ is uniformly integrable and tight over E. However, by assumption, $|f|^p$ is integrable over E, and therefore, by Propositions 2 and 6, respectively, of Chapter 5, the single integrable function $|f|^p$ is uniformly integrable and tight. It follows from inequality (1) that for each n,

$$|f - f_n|^p \leq 2^p\{|f|^p + |f_n|^p\} \text{ and } |f_n|^p \leq 2^p\{|f_n - f|^p + |f|^p\} \text{ on } E.$$

Consequently, $\{|f_n - f|^p\}$ is uniformly integrable and tight if and only $\{|f_n|^p\}$ is uniformly integrable and tight. □

PROBLEMS

9. Show that if Hölder's Inequality holds for normalized functions, then it holds in general.
10. For $p_1 < p_2$ and $E = (0, 1)$, define $f(x) = x^\alpha$ for $0 < x < 1$, where $-1/p_1 < \alpha \le -1/p_2$. Show that $f \in L^{p_1}(E) \sim L^{p_2}(E)$.
11. Define $f(x) = \frac{x^{-1/2}}{1+|\ln x|}$ for $x > 0$. Verify that $f \in L^p(0, \infty)$ if and only if $p = 2$.
12. Define $f(t) = t$ for $0 \le t \le 1$. Fix $1 \le p < \infty$. Find the dual function of $f \in L^p[0, 1]$. Observe that it depends on p.
13. Define $f(t) = t$ for $0 \le t \le 1$. Then $f \in L^\infty[0, 1]$. Show that there is no function $f^* \in L^1[0, 1]$ for which
$$\int_E f \cdot f^* \, dm = \|f\|_\infty \text{ and } \|f^*\|_1 = 1.$$
Does this contradict (3)?
14. Let $f, g \in L^2(E)$. Show that for any number λ,
$$\lambda^2 \int_E f^2 \, dm + 2\lambda \int_E f \cdot g \, dm + \int_E g^2 \, dm = \int_E (\lambda f + g)^2 \, dm \ge 0,$$
and from this and the quadratic formula directly derive the Cauchy-Schwarz Inequality.
15. For $1 < p < \infty$, q the conjugate of p and $f \in L^p(E)$, show that if $A \subseteq E$ has finite measure, then
$$\int_A |f| \, dm \le \|f\|_p \cdot [m(A)]^{1/q}.$$
Use this to conclude that a bounded subset of $L^p(E)$ is uniformly integrable.
16. Show that in Young's Inequality there is equality if and only if $a^p = b^q$.
17. Show that in Hölder's Inequality there is equality if and only if there are constants α and β, not both zero, for which
$$\alpha |f|^p = \beta |g|^q \text{ almost everywhere on } E.$$
18. (Chebychev's Inequality for $L^p(E)$) For $1 \le p < \infty$ and $f \in L^p(E)$, show that for each $\lambda > 0$,
$$m\{x \in E \mid |f(x)| > \lambda\} \le \frac{1}{\lambda^p} \|f\|_p^p.$$
19. For $1 \le p < \infty$, suppose that $\{f_n\} \to f$ in $L^p(E)$. Show that $\{f_n\} \to f$ in measure, and therefore there is a subsequence of $\{f_n\}$ that converges pointwise almost everywhere on E to f.
20. Assume that $m(E) < \infty$. For $f \in L^\infty(E)$, show that $\lim_{p \to \infty} \|f\|_p = \|f\|_\infty$.
21. For $1 \le p < \infty$, q the conjugate of p, and $f \in L^p(E)$, show that
$$\|f\|_p = \max_{g \in L^q(E), \|g\|_q \le 1} \int_E f \cdot g \, dm.$$

22. For $1 \leq p < \infty$ and $f \in L^p(E)$, show that $f = 0$ if and only if $\int_E f \cdot \psi \, dm = 0$ for every finitely supported, simple function $\psi \colon E \to \mathbf{R}$.
23. For $1 \leq p < \infty$, q the conjugate of p and $f \in L^p(E)$, show that $f = 0$ if and only if $\int_E f \cdot g \, dm = 0$ for $g \in L^q(E)$.
24. For $1 \leq p < \infty$, find the values of the parameter λ for which

$$\lim_{\epsilon \to 0^+} \frac{1}{\epsilon^\lambda} \int_0^\epsilon f \, dm = 0 \text{ for all } f \in L^p[0, 1].$$

25. Let $1 \leq p \leq \infty$ and q be the conjugate of p. If $\{f_n\} \to f$ in $L^p(E)$ and $\{g_n\} \to g$ in $L^q(E)$, show that

$$\lim_{n \to \infty} \int_E f_n \cdot g_n \, dm = \int_E f \cdot g \, dm.$$

26. Assume that $m(E) < \infty$ and $1 \leq p_1 < p_2 \leq \infty$. Show that if $\{f_n\} \to f$ in $L^{p_2}(E)$, then $\{f_n\} \to f$ in $L^{p_1}(E)$.
27. (The L^p Dominated Convergence Theorem) Let $\{f_n \colon E \to \mathbf{R}\}$ be a sequence of measurable functions that converges pointwise almost everywhere on E to f. For $1 \leq p < \infty$, suppose that there is a function g in $L^p(E)$ such that for all n, $|f_n| \leq g$ almost everywhere on E. Prove that $\{f_n\} \to f$ in $L^p(E)$.
28. For $1 \leq p < \infty$ and $f \in L^p(\mathbf{R})$, show that $\{f \cdot \chi_{[-n, n]}\} \to f$ in $L^p(\mathbf{R})$. (Suggestion: Use the preceding problem.)
29. Assume that $m(E) < \infty$ and $1 \leq p < \infty$. Suppose that $\{f_n \colon E \to \mathbf{R}\}$ is a sequence of measurable functions that converges pointwise almost everywhere on E to f. For $1 \leq p < \infty$, show that $\{f_n\} \to f$ in $L^p(E)$ if there is a $\theta > 0$ for which $\{f_n\}$ is a bounded subset of $L^{p+\theta}(E)$.
30. (Riesz) For $1 \leq p < \infty$, show that if the absolutely continuous function f on $[a, b]$ is the indefinite integral of an $L^p[a, b]$ function, then there is an $M > 0$ such that for any partition $\{x_0, \ldots, x_n\}$ of $[a, b]$,

$$\sum_{k=1}^{n} \frac{|f(x_k) - f(x_{k-1})|^p}{|x_k - x_{k-1}|^{p-1}} \leq M.$$

7.3 L^p IS COMPLETE: RAPIDLY CAUCHY SEQUENCES AND THE RIESZ-FISCHER THEOREM

Definition *A sequence $\{f_n\}$ in a normed linear space X is said to be* ***Cauchy*** *provided that for each $\epsilon > 0$, there is an index N for which*

$$\|f_n - f_m\| < \epsilon \text{ for all } m, n \geq N.$$

A normed linear space X is said to be ***complete*** *provided that every Cauchy sequence in X converges to a member of X. A complete normed linear space is called a* ***Banach space.***

The completeness axiom for the real numbers is equivalent to the assertion that $\mathbf{R}$, normed by the absolute value, is complete. We leave it as an exercise to conclude from this that $C[a, b]$, normed by the maximum norm, also is complete.

Proposition 4 *Every convergent sequence in a normed linear space X is Cauchy. Moreover, a Cauchy sequence converges if it has a convergent subsequence.*

Proof Let $\{f_n\} \to f$ in X. By the triangle inequality,

$$\|f_n - f_m\| = \|[f_n - f] + [f - f_m]\| \leq \|f_n - f\| + \|f_m - f\| \text{ for all } m, n.$$

From this it follows that $\{f_n\}$ is Cauchy. Now let $\{f_n\}$ be a Cauchy sequence in X that has a subsequence $\{f_{n_k}\}$ that converges in X to f. Let $\epsilon > 0$. Since $\{f_n\}$ is Cauchy, there is an index N for which $\|f_n - f_m\| < \epsilon/2$ for all $n, m \geq N$. Since $\{f_{n_k}\} \to f$, there is an index k for which $n_k > N$ and $\|f_{n_k} - f\| < \epsilon/2$. Therefore, by the triangle inequality,

$$\begin{aligned}\|f_n - f\| &= \|[f_n - f_{n_k}] + [f_{n_k} - f]\| \\ &\leq \|f_n - f_{n_k}\| + \|f_{n_k} - f\| < \epsilon \text{ for all } n \geq N.\end{aligned}$$

Consequently, $\{f_n\} \to f$ in X □

In view of the above proposition, a strategy to establish the completeness of a particular normed linear space is to show that a particular type of Cauchy sequence, tailored to the properties of the space, converges and also show that every Cauchy sequence has a subsequence of this particular type. In the $L^p(E)$ spaces, rapidly Cauchy sequences play this role[2].

Definition *A sequence $\{f_n\}$ in a normed linear space X is said to be* **rapidly Cauchy** *provided that*

$$\sum_{k=1}^{\infty} \|f_{k+1} - f_k\| < \infty.$$

Observe that for any sequence $\{f_n\}$ in X and all n and k

$$f_{n+k} - f_n = \sum_{j=n}^{n+k-1} [f_{j+1} - f_j]$$

and therefore, by the triangle inequality,

$$\|f_{n+k} - f_n\| \leq \sum_{j=n}^{n+k-1} \|f_{j+1} - f_j\| \leq \sum_{j=n}^{\infty} \|f_{j+1} - f_j\|. \tag{5}$$

Proposition 5 *Every rapidly Cauchy sequence in a normed linear space X is Cauchy. Moreover, every Cauchy sequence in X has a rapidly Cauchy subsequence.*

---

[2] In his article "Rethinking the Lebesgue Integral" (*the American Mathematical Monthly, December, 2009*), Peter Lax singles out pointwise limits of sequences of continuous functions that are rapidly Cauchy (defined slightly differently) with respect to the L^1 norm as primary objects in the construction of the complete space L^1. He constructs functions in L^1 as limits of such sequences, without first making a separate study of measure theory.

Proof Let $\{f_n\}$ be a rapidly Cauchy sequence in X. It follows from inequality (5) that $\{f_n\}$ is Cauchy. Now, for any Cauchy sequence $\{f_n\}$, there is a strictly increasing sequence $\{n_k\}$ such that

$$\|f_j - f_{n_k}\| \leq 1/2^k \text{ for all } j \geq n_k.$$

In particular

$$\|f_{n_{k+1}} - f_{n_k}\| \leq 1/2^k \text{ for all } k.$$

Since $\sum_{k=1}^{\infty} 1/2^k < \infty$, the subsequence $\{f_{n_k}\}$ is rapidly Cauchy. □

Theorem 6 *Let $1 \leq p \leq \infty$. If the sequence $\{f_n\}$ in $L^p(E)$ is rapidly Cauchy, then it converges in the $L^p(E)$ norm and also pointwise almost everywhere on E to a function $f \in L^p(E)$. Moreover, $\{f_n\}$ is dominated by a function in $L^p(E)$, in the sense that there is a function $w \in L^p(E)$ for which $|f_n| \leq w$ almost everywhere on E for all n.*

Proof The case $p = \infty$ is left as an exercise. Assume that $1 \leq p < \infty$. Let $\{f_n\}$ be a rapidly convergent sequence in $L^p(E)$. For notational convenience, by replacing each f_n by $f_n - f_1$, assume that $f_1 = 0$. Observe that for each n,

$$f_n = \sum_{k=1}^{n-1} [f_{k+1} - f_k],$$

and define

$$w_n = \sum_{k=1}^{n-1} |f_{k+1} - f_k|.$$

Since $\{w_n\}$ is increasing, the measurable function $w\colon E \to \overline{\mathbf{R}}$ may be defined by

$$w(x) = \lim_{n\to\infty} w_n(x) \text{ for all } x \in E.$$

By Minkowski's Inequality, since $\{f_n\}$ is rapidly Cauchy in $L^p(E)$, for all n,

$$\|w_n\|_p \leq \sum_{k=1}^{\infty} \|f_{k+1} - f_k\|_p = C < \infty.$$

Therefore $\int_E w_n^p \, dm \leq C^p$ for all n. Fatou's Lemma implies that $\int_E w^p \, dm \leq C^p$. So $w \in L^p(E)$, and therefore w is finite almost everywhere on E. This means that for almost all $x \in E$, the series of real numbers $\sum_{k=1}^{\infty} |f_{k+1} - f_k|$ converges. Since a series of real numbers converges if it converges absolutely, for almost all $x \in E$, a real number $f(x)$ is defined by

$$f(x) \equiv \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \sum_{k=1}^{n-1} [f_{k+1}(x) - f_k(x)].$$

Define $f(x) = 0$ if this limit does not exist. For all n

$$|f_n| \leq w_n \leq w \text{ and } |f| \leq w \text{ on } E$$

and therefore, by inequality (1),

$$|f_n - f|^p \leq 2^p[|f_n|^p + |f|^p] \leq 2^{p+1}w^p \text{ on } E.$$

Since w^p is integrable and $\{[f_n - f]^p\}$converges pointwise almost everywhere on E to 0, by this inequality and the Dominated Convergence Theorem,

$$\lim_{n\to\infty} \int_E |f_n - f|^p \, dm = 0.$$

Therefore, the sequence $\{f_n\}$ converges both pointwise almost everywhere and in $L^p(E)$ to f, and is dominated by the function $w \in L^p(E)$. □

The Riesz-Fischer Theorem *For $1 \leq p \leq \infty$, $L^p(E)$ is complete. Moreover, if $\{f_n\} \to f$ in $L^p(E)$, then a subsequence of $\{f_n\}$ converges pointwise almost everywhere on E to f and is dominated by a function in $L^p(E)$.*

Proof Let $\{f_n\}$ be a Cauchy sequence $L^p(E)$. According to the preceding proposition, there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ that is rapidly Cauchy. The preceding theorem implies that $\{f_{n_k}\}$ converges in $L^p(E)$ to a function f in $L^p(E)$ and therefore, since $\{f_n\}$ is Cauchy and has a convergent subsequence, $\{f_n\}$ itself converges in $L^p(E)$ to f. Consequently, $L^p(E)$ is complete. According to the preceding theorem, this rapidly Cauchy subsequence has the required pointwise convergence and domination properties. □

Corollary 7 *For $1 \leq p \leq \infty$, ℓ^p is complete.*

Proof We leave the case $p = \infty$ as an exercise. Define $E = [1, \infty)$ and let X_p to be the subspace of functions in $L^p(E)$ that are, for each k, constant on $[k, k+1)$. We claim that X_p, with the norm $\|\cdot\|_p$, is complete. Indeed, by the Riesz-Fischer Theorem, a Cauchy sequence in X_p converges to a function in $L^p(E)$, and since a subsequence converges pointwise almost everywhere, the limit function belongs to X_p. So X_p is complete. For $a = \{a_k\} \in \ell^p$, define $\mathcal{T}(a) \in X_p$ to take, for each k, the value a_k on $[k, k+1)$. The mapping $\mathcal{T}\colon \ell^p \to X_p$ preserves norms and $\mathcal{T}(\ell^p) = X_p$. Consequently, ℓ^p inherits completeness from X_p. □

PROBLEMS

31. Prove that $L^\infty(E)$ is complete.
32. Provide an example of a Cauchy sequence of real numbers that is not rapidly Cauchy.
33. Show that a normed linear space is complete if and only if each rapidly Cauchy sequence converges.
34. For each n, define $f_n(x) = x^n$ for $0 \leq x \leq 1$. Is $\{f_n\colon [0,\, 1] \to \mathbf{R}\}$ rapidly Cauchy in $C[0,\, 1]$? For $1 \leq p < \infty$, is $\{f_n\}$ rapidly Cauchy in $L^p([0,\, 1])$?
35. Consider the linear space of polynomials restricted to $[a,\, b]$, normed by the maximum norm. Is this space complete?
36. Let $m(E) < \infty$ and $1 \leq p_1 < p_2 < \infty$. Consider the linear space $L^{p_2}(E)$ normed by $\|\cdot\|_{p_1}$. Is this normed linear space compete?
37. A subspace X_0 of a normed linear space X is said to be closed provided that the limit of a convergent sequence in X_0 also belongs to X_0. Assume that X is complete. Show that a subspace X_0 of X, with the norm it inherits from X, is complete if and only if it is closed.

7.4 APPROXIMATION AND SEPARABILITY

Definition *A subset $\mathcal{F}$ of a normed linear space X is said to be* **dense** *in X provided that for each $f \in X$ and $\epsilon > 0$, there is an $f \in \mathcal{F}$ for which $\|f - g\| < \epsilon$.*

We have already encountered dense sets. The rational numbers are dense in $\mathbf{R}$. Also, the Weierstrass Approximation Theorem may be stated in our present vocabulary of normed linear spaces as follows: the collection of polynomials restricted to $[a, b]$ is dense in the linear space $C[a, b]$, normed by the maximum norm[3]. Observe that a set $\mathcal{F}$ is dense in X if and only if each $f \in X$ is the limit of a sequence $\{f_n\}$ in $\mathcal{F}$. Also, if $\mathcal{F}$ is dense in X and each $f \in \mathcal{F}$ is the limit of a sequence $\{f_n\}$ in $\mathcal{S} \subseteq X$, then $\mathcal{S}$ also is dense in X.

Theorem 8 *Let E be a measurable set and $1 \leq p < \infty$.*

(i) *The subspace of finitely supported, simple functions is dense in $L^p(E)$.*
(ii) *The subspace of continuous functions is dense in $L^p(E)$.*
(iii) *The subspace of step-functions is dense in $L^p[a, b]$, for any closed, bounded interval $[a, b]$.*

Proof We first verify (i). Let $g \in L^p(E)$. The function g is measurable and therefore, by the Simple Approximation Theorem, there is a sequence $\{\varphi_n\}$ of finitely supported, simple functions on E such that $\{\varphi_n\} \to g$ pointwise on E and $|\varphi_n| \leq |g|$ on E for all n. We claim that $\{\varphi_n\} \to g$ in $L^p(E)$. Indeed, for all n, according to (1),

$$|\varphi_n - g|^p \leq 2^p \cdot (|\varphi_n|^p + |g|^p) \leq 2^{p+1}|g|^p \text{ on } E.$$

But $|g|^p$ is integrable, and so, by the Dominated Convergence Theorem, $\{\varphi_n\} \to g$ in $L^p(E)$.

We now verify (ii). In view of (i), to do so it suffices to let $f \in L^p(E)$ be finitely supported and simple and show that there is a sequence of continuous functions that converges in $L^p(E)$ to f. However, by Corollary 11 in Chapter 3, since f is measurable, there is a sequence $\{f_n: \mathbf{R} \to \mathbf{R}\}$ of continuous functions that converges to f pointwise almost everywhere on E, and since f is bounded, we may assume that this sequence is uniformly pointwise bounded on E by M. Now, again using inequality (1),

$$|f - f_n|^p \leq 2^p \cdot (|f|^p + |f_n|^p) \leq 2^{p+1}M^p \text{ almost everywhere on } E \text{ for all } n.$$

Since f is finitely supported, we may assume $m(E) < \infty$, and therefore, by the Bounded Convergence Theorem, $\{f_n\} \to f$ in $L^p(E)$.

Lastly, we verify (iii). In view of (ii), to do so it suffices to let $f: [a, b] \to \mathbf{R}$ be continuous and show that there is a sequence of step-functions that converges in $L^p[a, b]$ to f. Fix n. Let $P_n = \{x_0, \ldots, x_n\}$ be the partition of $[a, b]$ into n subintervals of equal length. Define $s_n: [a, b] \to \mathbf{R}$ to be a step-functions that, for $k = 0, \ldots n-1$, takes the value $f(x_k)$ on $[x_k, x_{k+1})$. Since f is continuous and $[a, b]$ is compact, f is uniformly continuous. Therefore $\{s_n\}$ converges in $L^\infty[a, b]$, and hence also in $L^p[a, b]$, to f. □

[3]A proof, due to Sergei Bernstein, of the Weierstrass Approximation Theorem is given in Chapter 8 of Patrick Fitzpatrick's *Advanced Calculus*. In Chapter 16, we prove a generalization called the Stone-Weierstrass Theorem.

Definition *A subset $\mathcal{S}$ of a normed linear space X is said to be* ***separable*** *provided that there is a countable subset of $\mathcal{S}$ that is dense in $\mathcal{S}$.*

The real numbers are separable, since the rational numbers are a countable, dense subset. The space $C[a, b]$, normed by the maximum norm, is separable, since it follows from the Weierstrass Approximation Theorem that the collections of polynomials with rational coefficients are a countable, dense subset. For $1 \leq p < \infty$, the sequence space ℓ^p is separable, since the collection of sequences in ℓ^p that have rational components is a countable, dense subset.

Theorem 9 *For $1 \leq p < \infty$ and E a measurable set, $L^p(E)$ is separable.*

Proof First consider the case $E = [a, b]$. For each n, let P_n be the partition of $[a, b]$ into n subintervals of equal length, and let $\mathcal{S}_n$ be the collection of rational-valued functions on $[a, b]$ that are constant on each open subinterval induced by this partition. Then $\mathcal{S}_n$ is countable, and therefore so is $\bigcup_{n=1}^{\infty} \mathcal{S}_n$. It is clear that each step-functions is the limit of a sequence in this union, and consequently, by part (iii) of the preceding theorem, $L^p([a, b])$ is separable.

Now consider the case $E = \mathbf{R}$. For each n, define $X_n^p = \{f \cdot \chi_{[-n, n]} \mid f \in L^p(\mathbf{R})\}$, and observe that, by the case just considered, X_n^p is separable, and therefore so is $\bigcup_{n=1}^{\infty} X_n^p$. Consequently, to show that $L^p(\mathbf{R})$ is separable, it suffices to let $f \in L^p(\mathbf{R})$ and show that $\{f \cdot \chi_{[-n, n]}\} \to f$ in $L^p(\mathbf{R})$. Now

$$|f - f \cdot \chi_{[-n, n]}|^p \leq |f|^p \text{ on } \mathbf{R} \text{ for all } n,$$

and so, by an appeal to the Dominated Convergence Theorem, $\{f \cdot \chi_{[-n, n]}\} \to f$ in $L^p(\mathbf{R})$.

Finally, for a general measurable set E, by extending functions in $L^p(E)$ to functions in $L^p(\mathbf{R})$ that take the value 0 on $\mathbf{R} \sim E$, it follows that $L^p(E)$ is separable. □

Example In general, $L^\infty(E)$ is not separable. Indeed, $L^\infty[0, 1]$ is not separable. To verify this, we argue by contradiction. Otherwise, there is a countable set $\{f_n\}_{n=1}^{\infty}$ that is dense in $L^\infty[0, 1]$. For each number $x \in [0, 1]$, select a natural number $\eta(x)$ for which

$$\|\chi_{[a, x]} - f_{\eta(x)}\|_\infty < 1/2.$$

Observe that

$$\|\chi_{[a, x_1]} - \chi_{[a, x_2]}\|_\infty = 1 \text{ if } 0 \leq x_1 < x_2 \leq 1,$$

and therefore, by the triangle inequality for the $\|\cdot\|_\infty$ norm, η is a one-to-one mapping of $[0, 1]$ onto a set of natural numbers. But any set of natural numbers is countable, while $[0, 1]$ is not countable. We conclude from this contradiction that $L^\infty[0, 1]$ is not separable.

PROBLEMS

38. Let $A \subseteq B \subseteq X$. Show that A is dense in B if and only if each $f \in B$ is the limit of a sequence $\{f_n\}$ in A.
39. Let $\mathcal{F}$ be a dense subset of X and suppose that $\mathcal{S} \subseteq X$ has the property that each $f \in \mathcal{F}$ is the limit of a sequence $\{f_n\}$ in $\mathcal{S}$. Show that $\mathcal{S}$ is dense in X. In particular, show that if $\mathcal{S}$ is countable, then X is separable.
40. Show that, for $1 \leq p < \infty$, ℓ^p is separable, but ℓ^∞ is not separable.
41. Show that the subspace of simple functions is dense in $L^\infty(E)$.
42. Show that $C[a, b]$ is not a dense subspace of $L^\infty[a, b]$.
43. Let $1 \leq p < \infty$, q be the conjugate of p, and $\mathcal{S}$ be a dense subset of $L^q(E)$. Show that if $g \in L^p(E)$ and $\int_E f \cdot g \, dm = 0$ for all $f \in \mathcal{S}$, then $g = 0$.
44. Exhibit a measurable set E for which $L^\infty(E)$ is separable. Show that $L^\infty(E)$ is not separable if the set E contains a non-degenerate interval.

CHAPTER 8

The L^p Spaces: Duality, Weak Convergence, and Minimization

Contents

8.1 Bounded Linear Functionals on a Normed Linear Space 137
8.2 The Riesz Representation Theorem for the Dual of $L^p, 1 \le p < \infty$ 144
8.3 Weak Sequential Convergence in L^p 147
8.4 The Minimization of Convex Functionals 150

If X is a normed linear space, we consider the dual of X, denoted by X^*, a normed linear space consisting of bounded linear functionals $\Phi\colon X \to \mathbf{R}$. After considering some examples, we prove if X is separable, then every bounded sequence in X^* has a subsequence that converges pointwise to a bounded linear functional. We prove a Uniform Boundedness Principle, according to which, if X is complete, a pointwise convergent sequence in X^* is bounded. Then the Riesz Representation Theorem is proven: it asserts that if $1 \le p < \infty$ and q is the conjugate of p, then every bounded linear functional $\Phi\colon L^p(E) \to \mathbf{R}$ is represented by a function $g \in L^q(E)$, in the sense that

$$\Phi(f) = \int_E f \cdot g \, dm \text{ for all } f \in L^p(E).$$

With these two theorems in hand, the concept of weak convergence is considered and it is shown that, for $1 < p < \infty$, every bounded sequence in $L^p(E)$ has a weakly convergent subsequence. As an application of this weak sequential compactness theorem, we prove that, for $1 < p < \infty$, a continuous, convex function defined on a closed, bounded, convex subset of $L^p(E)$ attains a minimum value.

8.1 BOUNDED LINEAR FUNCTIONALS ON A NORMED LINEAR SPACE

Definition *A* **linear functional** *on a normed linear space X is a real-valued function $\Phi\colon X \to \mathbf{R}$ such that for $u, v \in X$ and real numbers α and β,*

$$\Phi(\alpha \cdot u + \beta \cdot v) = \alpha \cdot \Phi(u) + \beta \cdot \Phi(v).$$

Such a functional Φ is said to be **bounded** *provided that there is an $M \ge 0$ for which*

$$|\Phi(u)| \le M \cdot \|u\| \text{ for all } u \in X. \tag{1}$$

The infimum of all such M is called the **norm** *of Φ and denoted by $\|\Phi\|_*$.*

It is clear that if $\Phi \in X^*$, then $\|\Phi\|_*$ is actually the smallest M for which (1) holds. Moreover, by the homogeneity of a linear functional and of the norm, we also have

$$\|\Phi\|_* = \sup\{|\Phi(x)| \mid \|x\| \le 1\}. \tag{2}$$

So for a linear functional to be bounded means that its restriction to $\{x \in X \mid \|x\| \le 1\}$ is a bounded function. A linear functional Φ cannot be a bounded function on all of X, unless $\Phi = 0$. A (not necessarily linear) functional $F\colon X \to \mathbf{R}$ is said to be continuous provided that if a sequence $\{u_n\}$ converges in X to $u \in X$, then $\{F(x_n)\} \to F(u)$.

Proposition 1 *A linear functional $\Phi\colon X \to \mathbf{R}$ on a normed linear space X is bounded if and only if it is continuous.*

Proof Assume that Φ is bounded, so that, by the linearity of Φ,

$$|\Phi(u) - \Phi(v)| = |\Phi(u - v)| \le \|\Phi\|_* \cdot \|u - v\| \text{ for all } u,\, v \in X.$$

Clearly, this Lipschitz property implies continuity. To prove the converse, we argue by contradiction. Suppose that Φ is continuous but is not bounded. Then, by (2), there is a sequence $\{u_n\}$ in X for which $\|u_n\| = 1$ for all n, while if $\alpha_n = \Phi(u_n)$, the $\lim_{n\to\infty} \alpha_n = \infty$. For each n, define $v_n = u_n/\alpha_n$. Then $\{v_n\} \to 0$, while $\Phi(v_n) = 1$ for all n, so that $\{\Phi(v_n)\}$ does not converge to $\Phi(0) = 0$. This contradicts the continuity of Φ at $u = 0$. □

Proposition 2 *Let X be a normed linear space. Then the collection of bounded linear functionals on X is a linear space on which $\|\cdot\|_*$ is a norm. This normed linear space is called the* **dual space** *of X and denoted by X^*.*

We leave the proof of this proposition as an exercise. Throughout this chapter, E denotes a measurable subset of $\mathbf{R}$, and X denotes a normed linear space, with norm $\|\cdot\|$. From the viewpoint of Lebesgue integration, the following result describes the most important bounded linear functionals.

Proposition 3 *For $1 \le p \le \infty$, q the conjugate of p, and g in $L^q(E)$, define the linear functional Φ on $L^p(E)$ by*

$$\Phi(f) = \int_E f \cdot g\, dm \textit{ for all } f \in L^p(E).$$

Then

$$\Phi \in [L^p(E)]^* \textit{ and } \|\Phi\|_* = \|g\|_q. \tag{3}$$

Proof For $f \in L^p(E)$ and $g \in L^q(E)$, by Hölder's Inequality,

$$|\Phi(f)| \le \|f\|_p \cdot \|g\|_q.$$

Therefore, Φ is properly defined, and is linear. Furthermore, $\Phi\colon L^p(E) \to \mathbf{R}$ is bounded and $\|\Phi\|_* \le \|g\|_q$. It remains to show that $\|\Phi\|_* = \|g\|_q$. First consider $1 < p \le \infty$. Assume that $g \ne 0$, for otherwise there is nothing to prove. Then $1 \le q < \infty$, and so (with p and q interchanged regarding the membership of the dual function), the dual function of g, g^* belongs to $L^p(E)$, $\Phi(g^*) = \|g\|_q$ and $\|g^*\|_p = 1$. Therefore, by (2), $\|\Phi\|_* = \|g\|_q$. Now consider $p = 1$. We argue by contradiction to establish the equality $\|\Phi\|_* = \|g\|_\infty$. Indeed, otherwise, $\|\Phi\|_* < \|g\|_\infty$, and so there is some $\epsilon > 0$ for which the set

$E_\epsilon = \{x \in E \mid |g(x)| > \|\Phi\|_* + \epsilon\}$ has non-zero measure. If we choose f be the characteristic function of a measurable subset of E_ϵ that has finite positive measure, we contradict the inequality $|\Phi(f)| \le \|\Phi\|_* \cdot \|f\|_1$. □

Example If f: $[0, 1] \to \mathbf{R}$ is integrable, define $\Phi(f) = \int_0^1 x f(x)\,dx$. We leave it as an exercise to show that for $\Phi\colon L^1[0, 1] \to \mathbf{R}$, $\|\Phi\|_* = 1$, while for $\Phi\colon L^\infty[0, 1] \to \mathbf{R}$, $\|\Phi\|_* = 1/2$.

Corollary 4 *If $1 \le p < \infty$ and $f \in L^p(E)$, $f \neq 0$, then there is a $\Phi \in [L^p(E)]^*$ for which*[1]

$$\Phi(f) = \|f\|_p \text{ and } \|\Phi\|_* = 1. \tag{4}$$

Proof Let $f^* \in L^q(E)$ be the dual function of f. According to the preceding proposition,

$$\Phi(g) = \int_E g \cdot f^*\,dm \text{ for all } g \in L^p(E)$$

defines $\Phi \in [L^p(E)]^*$, and $\|\Phi\|_* = \|f^*\|_q$. By the definition of dual function, $\Phi(f) = \|f\|_p$ and $\|f^*\|_q = 1$. □

For any set S, a sequence of functions $\{f_n\colon S \to \mathbf{R}\}$ is said to be **pointwise bounded** provided that for each s in S, $\{f_n(s)\}$ is bounded, and said to be **pointwise convergent** provided that there for each s in S, $\{f_n(s)\}$ converges to a real number. In the proof of the next theorem, we appeal to the Bolzano-Weierstrass Theorem, which we recall asserts that every bounded sequence of real numbers has a convergent subsequence.

Theorem 5 *If the set S is countable and the sequence $\{f_n\colon S \to \mathbf{R}\}$ is pointwise bounded, then there is a subsequence $\{f_{n_k}\}$ that converges pointwise on S to a function $f\colon S \to \mathbf{R}$.*

Proof Let $\{x_k\}_{k=1}^\infty$ be an enumeration of S. The first step is to construct a countable collection of subsequences of $\{f_n\colon S \to \mathbf{R}\}$ that, for each k, has the following two properties:

(i) The k-th subsequence, evaluated at the point x_k, converges to a real number c_k.
(ii) The $k+1$ subsequence is itself a subsequence of the k-th subsequence.

We construct this countable collection of subsequences by induction. By assumption, the sequence of real numbers $\{f_n(x_1)\}$ is bounded. Consequently, there is a subsequence of $\{f_n\colon S \to \mathbf{R}\}$ for which the evaluation at the point x_1 converges to a real number that we denote by c_1. Now this first subsequence, evaluated at the point x_2, is bounded. Therefore, this first subsequence has itself a subsequence for which the evaluation at the point x_2 converges to a number that we denote by c_2. And so the argument continues as follows: assume n is a natural number and we have chosen n subsequences, each a subsequence of it predecessor, and for $1 \le k \le n$, property (i) holds. Then the n-th subsequence of functions, evaluated at the point x_{n+1}, is bounded, and therefore there is an $n+1$ subsequence, a

[1] For any normed linear space X, if $f \in X$, $f \neq 0$, then there is a functional $\Phi \in X^*$ for which $\Phi(f) = \|f\|$ and $\|\Phi\|_* = 1$. This is a consequence of the Hahn-Banach Theorem, which we prove in Chapter 18. In particular, this corollary holds for $p = \infty$.

subsequence on the n-th sequence, for which the evaluation at the point x_{n+1} converges to c_{n+1}. We have constructed a countable collection of subsequences of $\{f_n\}$ that possess properties (i) and (ii).

At this point, it is convenient to label this countable collection of subsequences. For each n, label the n-th subsequence of functions selected above as $f_1^n, f_2^n f_3^n, \ldots$, that is, f_k^n is the k-th term of the n-th subsequence. We claim that

$$\lim_{k\to\infty} f_k^k(x_n) = c_n \text{ for all } n. \tag{5}$$

Indeed, if $n \leq m$, since $\{f_k^m\}$ is a subsequence of $\{f_k^n\}$, the j-th term of $\{f_k^m\}$ equals the ℓ-th term of $\{f_k^n\}$ for some $\ell \geq j$. Therefore if $k \geq n$, $f_k^k = f_\ell^n$ for some $\ell \geq k$ so that

$$\text{if } k \geq n,\ f_k^k(x_n) = f_\ell^n(x_n) \text{ for some } \ell \geq k.$$

Since $\lim_{\ell\to\infty} f_\ell^n(x_n) = c_n$, we have $\lim_{k\to\infty} f_k^k(x_n) = c_n$, and so (5) is verified. For each k, let n_k be the index of the function f_k^k as a term of the sequence $\{f_n\}$. Then for each k, $n_k \geq k$, $\{n_k\}$ is strictly increasing and $\{f_{n_k}\}$ converges pointwise on S to $f\colon S \to \mathbf{R}$ defined by $f(x_n) = c_n$ for all n. The proof is complete. □

Corollary 6 *Every bounded sequence $\{s_n\}$ in the space ℓ^2 has a subsequence that converges componentwise to a member of ℓ^2.*

Proof Let $\{s_n\}$ be a bounded sequence in ℓ^2. For each n, define the function $f_n\colon \mathbf{N} \to \mathbf{R}$ by setting, for each j, $f_n(j)$ equal to the j-th term of the sequence s_n. Since $\{s_n\}$ is bounded in ℓ^2, $\{f_n\colon \mathbf{N} \to \mathbf{R}\}$ is a pointwise bounded sequence of functions on the countable set $\mathbf{N}$. According to the preceding theorem, there is subsequence $\{f_{n_k}\colon \mathbf{N} \to \mathbf{R}\}$ that converges pointwise to $f\colon \mathbf{N} \to \mathbf{R}$. This means the sequence $\{s_{n_k}\}$ converges componentwise to the sequence s, the j-th term of which is $f(j)$. We leave it as an exercise to show that the boundedness in ℓ^2 of $\{s_n\}$ implies the convergence of the series $\sum_{j=1}^\infty [f(j)]^2$, and so $s \in \ell^2$. □

Helly's Theorem. *If X is a separable normed linear space and $\{\Phi_n\}$ is a bounded sequence in X^*, then a subsequence of $\{\Phi_n\}$ converges pointwise on X to $\Phi \in X^*$.*

Proof Let S be a countable dense subset on X, and $C > 0$ be such that

$$\|\Phi_n\| \leq C \text{ for all } n. \tag{6}$$

Since $\{\Phi_n\}$ is pointwise bounded on S, according to the preceding theorem, a subsequence of $\{\Phi_n\}$ converges pointwise on S to a function $\Phi\colon S \to \mathbf{R}$. For notational brevity, assume the whole sequence converges pointwise. Let $u \in X \sim S$. We claim that $\{\Phi_n(u)\}$ is Cauchy. Indeed, let $\epsilon > 0$. Since S is dense in X, there is a point $u_0 \in S$ for which $\|u - u_0\| < \epsilon/4C$. Observe that for any n and m

$$\Phi_n(u) - \Phi_m(v) = [\Phi_n(u) - \Phi_n(u_0)] + [\Phi_n(u_0) - \Phi_m(u_0)] + [\Phi_m(u_0) - \Phi_m(u)].$$

Therefore, by the choice of u_0, the triangle inequality and the estimate (6), for all n and m

$$|\Phi_n(u) - \Phi_m(u)| \leq \epsilon/2 + |\Phi_n(u_0) - \Phi_m(u_0)|.$$

Since $\{\Phi_n(u_0)\}$ converges, it is Cauchy, and so there is an N for which $|\Phi_n(u_0) - \Phi_m(u_0)| < \epsilon/2$ for all $n, m \geq N$. Consequently,

$$|\Phi_n(u) - \Phi_m(u)| \leq \epsilon/2 + |\Phi_n(u_0) - \Phi_m(u_0)| < \epsilon \text{ for all } n,\ m \geq N.$$

Consequently, the sequence of real numbers $\{\Phi_n(u)\}$ is Cauchy and therefore converges. Denote its limit by $\Phi(u)$, so we have

$$\lim_{n\to\infty} \Phi_n(u) = \Phi(u) \text{ for all } u \in X.$$

The functional $\Phi\colon X \to \mathbf{R}$ inherits linearity from the Φ_n's, and boundedness from (6). □

Definition *A subset S of a normed linear space X is said to be* **total** *provided that the subspace of finite linear combinations of members of S is dense in X.*

Example

(i) For E measurable and $1 \leq p < \infty$, let $S = \{\chi_A \mid A \subseteq E, m(A) < \infty\}$. Each simple function in $L^p(E)$ is a linear combination of functions in S. It follows from assertion (i) of Theorem 8 in the preceding chapter that S is a total subset of $L^p(E)$.

(ii) For $[a, b]$ a closed, bounded interval and $1 \leq p < \infty$, let $S = \{\chi_{[a,\, x]} \mid x \in [a, b]\}$. Each step-function is a linear combination of functions in S. It follows from assertion (iii) of Theorem 8 in the preceding chapter that S is a total subset of $L^p[a, b]$.

(iii) Let S be the collection of restrictions to $[a, b]$ of monomials x^n, for n a non-negative integer. The Weierstrass Approximation Theorem amounts to the assertion that S is a total subset of $C[a, b]$.

It is useful to observe that if S is a total subset of X and $\Phi \in X^*$, then, by the linearity and continuity of Φ,

$$\Phi = 0 \text{ if } \Phi(f) = 0 \text{ for all } f \in S. \tag{7}$$

Example Let g: $[a, b] \to \mathbf{R}$ and h: $[a, b] \to \mathbf{R}$ be continuous. Then $g = h$ if

$$\int_a^b g(x) \cdot x^n\, dx = \int_a^b h(x) \cdot x^n\, dx \text{ for all } n \geq 0.$$

To verify this, let $f = g - h$ and define $\Phi\colon C[a, b] \to \mathbf{R}$ by

$$\Phi(\eta) = \int_a^b f(x)\eta(x)\, dx \text{ for all } \eta \in C[a, b].$$

Observe that $|\Phi(\eta)| \leq \|f\|_1 \cdot \|\eta\|_{\max}$ for all $\eta \in C[a, b]$, so that $\Phi \in [C[a, b]]^*$. Also observe that $\Phi(x^n) = 0$ for all $n \geq 0$ and such monomials are a total subset in $C[a, b]$. Consequently, by (7), $\Phi = 0$. In particular, $\Phi(f) = 0$, so $f = g - h = 0$.

Proposition 7 *If S is a total subset of the normed linear space X, $\{\Phi_n\}$ is a bounded sequence in X^* and $\Phi \in X^*$, then $\{\Phi_n\}$ converges pointwise to Φ on X if and only if it converges pointwise to Φ on S.*

Proof Assume that $\{\Phi_n\}$ converges pointwise to Φ on S. Let X_0 be the subspace of finite linear combinations of members of S. By linearity, $\{\Phi_n\}$ converges pointwise to Φ on X_0. Now, since X_0 is dense in X, the proof proceeds as did the last part of the proof of Helly's Theorem. $\square$

Theorem 8 (**A Uniform Boundedness Principle**) *Let X be a complete normed linear space and $\{\Phi_n\}$ be a sequence in X^* that is pointwise bounded. Then $\{\Phi_n\}$ is bounded in X^*, that is, there is an $M \geq 0$ such that*

$$|\Phi_n(x)| \leq M \text{ for all } n \text{ and all } x \text{ with } ||x|| \leq 1.$$

Proof We argue by contradiction. Assume that $\{\Phi_n\}$ is not bounded in X^*. We claim that there is a sequence $\{x_k\}$ in X and a subsequence $\{\Phi_{n_k}\}$ such that, for each k,

$$||x_k|| \leq 1/2^k,\ \Phi_{n_k}(x_1 + \ldots + x_k) \geq k \text{ and } |\Phi_{n_k}(x_i)| \leq 1/2^i \text{ if } i > k. \tag{8}$$

Assume such a sequence and subsequence have been chosen. Each $||x_k|| \leq 1/2^k$, so that the sequence of partial sums of the series $\sum_{k=1}^{\infty} x_k$ is rapidly Cauchy. Therefore, by the completeness of X, the series $\sum_{k=1}^{\infty} x_k$ converges to a point $x \in X$. Fix k. By the continuity and linearity of Φ_{n_k},

$$\Phi_{n_k}(x) = \sum_{j=1}^{\infty} \Phi_{n_k}(x_j) = \Phi_{n_k}(x_1 + \ldots + x_k) + \sum_{j=k+1}^{\infty} \Phi_{n_k}(x_j).$$

Now, for each k,

$$\Phi_{n_k}(x_1 + \ldots + x_k) \geq k \text{ and } \left|\sum_{j=k+1}^{\infty} \Phi_{n_k}(x_j)\right| \leq \sum_{j=k+1}^{\infty} 1/2^j \leq 1.$$

Consequently, $\Phi_{n_k}(x) \geq k - 1$, for all k, which contradicts the boundedness of $\{\Phi_n(x)\}$.

It remains to define the sequence and subsequence for which (8) holds. We do so inductively, using the following observation: for each $u \in X$ and $r_k > 0$ since, by assumption, $\{\Phi_n(u)\}$ is bounded, but $\{\Phi_n\}$ is not bounded in X^*, there is an index n_k and $v \in X$ for which $||v|| \leq r_k$ and $\Phi_{n_k}(v) \geq k - \Phi_n(u)$ for all n, so that, in particular,

$$||v|| \leq r_k \text{ and } \Phi_{n_k}(u + v) \geq k. \tag{9}$$

Since $\{\Phi_n\}$ is not bounded in X^*, there is point x_1 and index n_1 such that $||x_1|| \leq 1/2$ and $\Phi_{n_1}(x_1) \geq 1$. Suppose that $x_1, \ldots x_{k-1}$ and $\Phi_{n_1}, \ldots \Phi_{n_{k-1}}$ have been chosen such that, for $1 \leq i \leq k-1$,

$$||x_i|| \leq 1/2^i,\ \Phi_{n_i}(x_1 + \ldots x_i) \geq i, \text{ and } |\Phi_{n_i}(x_j)| \leq 1/2^j \text{ if } 1 \leq i < j \leq k-1.$$

Choose $r_k > 0$ such that $r_k < 1/2^k$ and if $||x|| \leq r_k$, then $|\Phi_{n_i}(x)| < 1/2^k$ for $1 \leq i \leq k-1$. For this choice of r_k and $u = x_1 + \ldots + x_{k-1}$ choose $v \in X$ for which (9) holds, and define $x_k = v$. This inductively defines a sequence and subsequence for which (8) holds. $\square$

In the elegant proof of the above Uniform Boundedness Principle, which is due to Hans Hahn, linearity of the functionals is essential. However, this principle is a special case of one that applies to both linear and non-linear functionals. We prove such a result in Chapter 14, a special case of which asserts the following: For any pointwise bounded sequence of real-valued continuous (not necessarily linear) functions on a complete normed linear space, there is a ball on which the sequence is uniformly pointwise bounded.

PROBLEMS

1. For $1 \leq p \leq \infty$, define $\Phi\colon L^p[0, 1] \to \mathbf{R}$ by $\Phi(f) = \int_0^1 xf(x)\,dx$. Show that $\Phi \in [L^p[0, 1]]^*$ and determine $\|\Phi\|_*$.
2. For $x_0 \in [a, b]$, define the functional $\Phi\colon C[a, b] \to \mathbf{R}$ by $\Phi(f) = f(x_0)$ for $f \in C[a, b]$. Prove that $\Phi \in [C[a, b]]^*$ and determine $\|\Phi\|_*$.
3. Define the functional $\Phi\colon C[a, b] \to \mathbf{R}$ by $\Phi(f) = \int_a^b f(x)\,dx$ for $f \in C[a, b]$. Prove that $\Phi \in [C[a, b]]^*$ and determine $\|\Phi\|_*$.
4. Show that a linear functional $\Phi\colon X \to \mathbf{R}$ is bounded if and only if there is an $r > 0$ and $x_0 \in X$ such that the restriction Φ to $\{x \in X \mid \|x - x_0\| \leq r\}$ is bounded.
5. Show that a sequence $\{\Phi_n\}$ in X^* fails to be bounded if and only if for each index k, $u \in X$, and $r > 0$, there is an index n_k and $v \in X$ such that
$$\Phi_{n_k}(u + v) > k \text{ and } \|v\| \leq r.$$
6. For $f \in L^p(E), 1 \leq p < \infty$, show that $f = 0$ if $\Phi(f) = 0$ for all $\Phi \in [L^p(E)]^*$.
7. Verify (7).
8. Let $\mathcal{S}$ be a dense subset of X. Show that if $\Phi \in X^*$, then $\|\Phi\|_* = \sup\{\Phi(f) \mid f \in \mathcal{S}, \|f\| \leq 1\}$.
9. For $g \in L^p(E), 1 \leq p < \infty$, show that $g = 0$ if $\int_A g\,dm = 0$ for all $A \subseteq E$ with $m(A) < \infty$.
10. For $g \in C[a, b]$, show that $g = 0$ if $\int_a^x g(t)\,dt = 0$ for all $x \in [a, b]$.
11. Provide an example of a sequence of functionals in X^* that converges pointwise on X to a functional in X^*, but not as a sequence in the normed linear space X^*.
12. Prove Proposition 2.
13. Prove the assertion in the proof of Corollary 6 regarding the convergence of the series $\sum_{j=1}^{\infty}[f(j)]^2$.
14. Extend Corollary 6 from ℓ^2 to ℓ^p, for $1 \leq p < \infty$.
15. For X a normed linear space, show that its dual space X^* is a complete normed linear space.
16. A general functional F: $X \to \mathbf{R}$ on a normed linear space X is said to be Lipschitz provided that there is a $c \geq 0$ such that $|F(u) - F(v)| \leq c\|u - v\|$ for all $u, v \in X$. The infimum of such c's is called the Lipschitz constant for F. Show that a linear functional Φ is bounded if and only if it is Lipschitz, in which case its Lipschitz constant is $\|\Phi\|_*$.
17. Prove that a general functional F: $X \to \mathbf{R}$ is continuous if whenever a rapidly Cauchy sequence $\{u_n\}$ converges in X to $u \in X$, then $\{F(x_n)\} \to F(u)$.
18. Let $\varphi\colon \mathbf{R} \to \mathbf{R}$ be continuous and there is a $c \geq 0$ for which $|\varphi(s)| \leq c + c \cdot |s|^2$ for all $s \in \mathbf{R}$. Define $F(f) = \int_E \varphi \circ f\,dm$ *for* $f \in L^2[a, b]$. Show that F: $L^2[a, b] \to \mathbf{R}$ is continuous. (Suggestion: Use the preceding problem.)

19. Define the functional F: $L^2[a, b] \to \mathbf{R}$ by $F(f) = \int_{[a, b]} \varphi \circ f \, dm$ for $f \in L^2[a, b]$. For the following choices of φ, show that F is continuous: (i) $\varphi(s) = \cos s$; (ii) $\varphi(s) = s^2/(1 + |s|)$.

8.2 THE RIESZ REPRESENTATION THEOREM FOR THE DUAL OF L^p, $1 \leq p < \infty$

There is the following useful criterion for the membership of a function in $L^q(E)$ for $1 < q \leq \infty$.

Lemma 9 *For $1 \leq p < \infty$ and $g \in L^1(E)$, assume that there is an $M \geq 0$ for which*

$$\left| \int_E \varphi \cdot g \, dm \right| \leq M \cdot \|\varphi\|_p \text{ for every finitely supported, simple function } \varphi. \tag{10}$$

Then $g \in L^q(E)$, where q is the conjugate of p. Moreover, $\|g\|_q \leq M$.

Proof Observe that if φ is finitely supported and simple, then so is the product $\operatorname{sgn}(g) \cdot \varphi$,

$$|g| \cdot \varphi = g \cdot \operatorname{sgn}(g) \cdot \varphi \text{ and } \|\operatorname{sgn}(g) \cdot \varphi\|_p = \|\varphi\|_p.$$

Therefore, (10) holds if g is replaced by $|g|$. So we may assume $g \geq 0$. First consider the case $p = 1$. It must be shown that $g \in L^\infty(E)$ and M is an essential upper bound for g. Indeed, otherwise, there is some $\epsilon > 0$ for which the set $E_\epsilon = \{x \in E \mid |g(x)| > M + \epsilon\}$ has non-zero measure. Define φ to be the characteristic function of a measurable subset of E_ϵ that has finite positive measure, and so contradict (10). Now consider $p > 1$. Since g is non-negative and measurable, by the Simple Approximation Theorem, there is an increasing sequence $\{\varphi_k \colon E \to \mathbf{R}\}$ of non-negative, finitely supported, simple functions that converges pointwise on E to g. Observe that, since $1 \leq q < \infty$, $\{\varphi_k^q\}$ is an increasing sequence of non-negative, measurable functions that converges pointwise to g^q and therefore, by the Monotone Convergence Theorem, to verify that $g \in L^q(E)$ and $\|g\|_q \leq M$, it suffices to show that

$$\int_E [\varphi_k]^q \, dm \leq M^q \text{ for all } k. \tag{11}$$

Fix k. Observe that

$$\varphi_k^q = \varphi_k \cdot \varphi_k^{q-1} \leq g \cdot \varphi_k^{q-1},$$

and φ_k^{q-1} is finitely supported and simple. Therefore, by assumption (10), since $(q-1)p = q$,

$$\int_E [\varphi_k]^q \, dm \leq M \cdot \left[\int_E [\varphi_k]^{(q-1)p} \, dm \right]^{1/p} = M \cdot \left[\int_E [\varphi_k]^q \, dm \right]^{1/p}.$$

Since $1 - 1/p = 1/q$, this is precisely the inequality (11). □

Definition *For $1 \leq p \leq \infty$ and q the conjugate of p, a functional $\Phi \in [L^p(E)]^*$ is said to be* **represented** *by the function $g \in L^q(E)$, written $\mathcal{R}(\Phi) = g$, provided that*

$$\Phi(f) = \int_E g \cdot f \, dm \text{ for all } f \in L^p(E). \tag{12}$$

Theorem 10 *If $1 \leq p < \infty$ and q is the conjugate of p, then every $\Phi \in [L^p[a, b]]^*$ is represented by a function $g \in L^q[a, b]$.*

Proof Consider the case that $p > 1$. The proof for the case $p = 1$ is similar (see Problem 20). For $x \in [a, b]$, define $\eta(x) = 0$ if $x = a$ and otherwise

$$\eta(x) = \Phi(\chi_{[a, x)}).$$

We claim that $\eta\colon [a, b] \to \mathbf{R}$ is absolutely continuous. Indeed, by the linearity of Φ, for each $[c, d] \subseteq [a, b]$, since $\chi_{[c, d)} = \chi_{[a, d)} - \chi_{[a, c)}$,

$$\eta(d) - \eta(c) = \Phi(\chi_{[a, d)}) - \Phi(\chi_{[a, c)}) = \Phi(\chi_{[c, d)}).$$

Now, if $\{(a_k, b_k)\}_{k=1}^n$ is a finite disjoint collection of intervals in (a, b), for each k, define $\epsilon_k = \operatorname{sgn}(\eta(b_k) - \eta(a_k))$, so that $|\eta(b_k) - \eta(a_k)| = \epsilon_k \cdot [\eta(b_k) - \eta(a_k)]$. Since Φ is linear,

$$\sum_{k=1}^{n} |\eta(b_k) - \eta(a_k)| = \sum_{k=1}^{n} \epsilon_k \cdot \Phi(\chi_{[a_k, b_k)}) = \Phi\left(\sum_{k=1}^{n} \epsilon_k \cdot \chi_{[a_k, b_k)}\right). \tag{13}$$

Now, for the simple function $f \equiv \sum_{k=1}^{n} \epsilon_k \cdot \chi_{[a_k, b_k)}$,

$$|\Phi(f)| \leq \|\Phi\|_* \cdot \|f\|_p \text{ and } \|f\|_p = \left[\sum_{k=1}^{n} (b_k - a_k)\right]^{1/p}.$$

Therefore, by (13),

$$\sum_{k=1}^{n} |\eta(b_k) - \eta(a_k)| = \Phi(f) \leq \|\Phi\|_* \cdot \|f\|_p = \|\Phi\|_* \cdot \left[\sum_{k=1}^{n} (b_k - a_k)\right]^{1/p}.$$

Consequently, $\delta = (\epsilon/\|\Phi\|_*)^p$ responds to any $\epsilon > 0$ challenge regarding the criterion for η to be absolutely continuous on $[a, b]$.

According to Theorem 12 of Chapter 6, the function η is differentiable almost everywhere on (a, b), $g \equiv \eta'$ is integrable, and moreover, since $\eta(a) = 0$,

$$\eta(x) = \int_a^b g(t)\, dt \text{ for all } x \in [a, b].$$

Therefore, for each $[c, d] \subseteq (a, b)$,

$$\Phi(\chi_{[c, d)}) = \eta(d) - \eta(c) = \int_a^b g \cdot \chi_{[c, d)}\, dm.$$

Since the functional Φ and the functional $f \mapsto \int_a^b g \cdot f\, dm$ are linear on the linear space of step-functions, it follows that (12) holds for any step-function f on $[a, b]$. According to Theorem 8 in the preceding chapter, the step-functions are dense in $L^p[a, b]$. Let $\varphi\colon [a, b] \to \mathbf{R}$ be simple. There is a sequence of step-functions that converges in $L^p[a, b]$ to φ.

By the continuity of Φ and the integrability of g, taking limits we have $\Phi(\varphi) = \int_a^b g \cdot \varphi \, dm$. Consequently, for all such φ,

$$\left| \int_a^b g \cdot \varphi \, dm \right| \leq \|\Phi\|_* \cdot \|\varphi\|_q,$$

and we conclude from the preceding lemma that $g \in L^q[a, b]$. Therefore, (12) holds for all simple functions, and such functions are dense in $L^p[a, b]$. Let $f \in L^p[a, b]$. Choose a sequence of simple functions $\{\varphi_n\}$ that converges to f in $L^p[a, b]$. Then

$$\Phi(f) = \lim_{n \to \infty} \Phi(\varphi_n) = \lim_{n \to \infty} \int_a^b \varphi_n \cdot g \, dm = \int_a^b f \cdot g \, dm.$$

This completes the proof. □

The Riesz Representation Theorem for the Dual of $\mathbf{L}^p(E)$ *If $E \subseteq \mathbf{R}$ is measurable, $1 \leq p < \infty$ and q is the conjugate of p, then every functional $\Phi \in [L^p(E)]^*$ is represented by a function g in $L^q(E)$.*

Proof First consider the case that $E = \mathbf{R}$. Choose an ascending collection $\{I_n\}_{n=1}^\infty$ of closed, bounded intervals whose union is $\mathbf{R}$. We consider $L^p(I_n)$ to be the subspace of $L^p(\mathbf{R})$ comprising functions that vanish outside I_n. According to the preceding theorem, for each n, there is a $f_n \in L^q(I_n)$ that represents Φ_n, the restriction of Φ to $L^p(I_n)$, that is,

$$\Phi_n(g) = \int_{I_n} f_n \cdot g \, dm \text{ for all } g \in L^p(I_n). \tag{14}$$

It follows from Proposition 3 that, after the usual identification of functions that are equal almost everywhere, a bounded linear functional has at most one representative. Therefore, we may define $f \colon \mathbf{R} \to \mathbf{R}$ so that its restriction to each I_n is f_n. It follows from (3) that for each n,

$$\|f_n\|_q = \|\Phi_n\|_* \leq \|\Phi\|_*.$$

Therefore, since $\{|f_n|^q\}$ converges pointwise on $\mathbf{R}$ to $|f|^q$, by Fatou's Lemma, f belongs to $L^q(\mathbf{R})$. Again by (3), f represents a continuous functional on $L^p(\mathbf{R})$, as is the functional Φ. Let $g \in L^p(\mathbf{R})$. For each n, let χ_n be the characteristic function of I_n, and observe that $|g - g \cdot \chi_n|^q \leq |g|^q$. Consequently, by the Dominated Convergence Theorem, $\{g \cdot \chi_n\} \to g$ in $L^p(\mathbf{R})$. In view of (14), taking limits, we have, for all $g \in L^p(\mathbf{R})$,

$$\Phi(g) = \lim_{n \to \infty} \Phi(g \cdot \chi_n) = \lim_{n \to \infty} \int_{\mathbf{R}} f \cdot (g \cdot \chi_n) \, dm = \int_{\mathbf{R}} f \cdot g \, dm.$$

Therefore Φ is represented by f.

Now consider a general measurable set E. Define the functional $\widehat{\Phi} \in [L^p(\mathbf{R})]^*$ by $\widehat{\Phi}(f) = \Phi(f|_E)$. Then $\widehat{\Phi} \in [L^p(\mathbf{R})]^*$. It has just been shown that there is a function $\hat{g} \in L^q(\mathbf{R})$ that represents $\widehat{\Phi}$. Define g to be the restriction of $\hat{g}$ to E. Then $g \in L^q(E)$ represents Φ. □

Corollary 11 *For $1 \leq p < \infty$ and q the conjugate of p, the Riesz representation operator*

$$\mathcal{R}\colon [L^p(E)]^* \to L^q(E) \text{ is one-to-one and onto, and moreover,}$$

$$\|\mathcal{R}(\Phi)\|_q = \|\Phi\|_* \text{ for all } \Phi \in [L^p(E)]^*.$$

Proof According to the Riesz Representation Theorem, linear operator $\mathcal{R}$ is properly defined and, according to Proposition 3, it is onto and preserves norms. Being linear and preserving norms, $\mathcal{R}$ is one-to-one. □

In the preceding representation theorem, it is assumed that $p \neq \infty$, and this is a necessary restriction. In Chapter 12, a theorem of Kantorovitch is proven, which provides a representation for all the bounded linear functionals on $L^\infty(E)$.

PROBLEMS

20. Establish the Riesz Representation Theorem in the case $p = 1$ by first showing, in the notation of the proof of the theorem, that the function η is Lipschitz and therefore it is absolutely continuous. Then follow the proof for $p > 1$.
21. For $1 \leq p < \infty$, by mapping a sequence $\{a_k\}$ in ℓ^p to a measurable function in $L^p[1, \infty)$ that takes the value a_k on each $[k, k+1)$, extend Hölder's Inequality to the ℓ^p spaces. Use this to state and prove an ℓ^p version of Lemma 9. Then state and prove a representation theorem for the dual of $\ell^p, 1 \leq p < \infty$.
22. Show that if every bounded linear functional on $L^\infty[a, b]$ were represented by a function in $L^1[a, b]$, then the dual of $L^\infty[a, b]$ would be separable.
23. Let $\Phi \in [C[a, b]]^*$. For $x \in [a, b]$, define $g_x\colon [a, b] \to \mathbf{R}$ by $g(t) = a - t$ for $a \leq t \leq x$ and $g_x(t) = a - x$ for $x \leq t \leq b$. Define $\eta(x) = \Phi(g_x)$. Show that $\eta\colon [a, b] \mapsto \mathbf{R}$ is Lipschitz.

8.3 WEAK SEQUENTIAL CONVERGENCE IN L^p

Definition *A sequence $\{u_n\}$ in the normed linear space X is said to* **converge weakly** *in X to $u \in X$ provided that*

$$\lim_{n \to \infty} \Phi(u_n) = \Phi(u) \text{ for all } \Phi \in X^*.$$

It is convenient to say that a sequence in X **converges strongly** in X if it converges with respect to the norm. For a sequence $\{u_n\}$ in X and $u \in X$,

$$|\Phi(u_n) - \Phi(u)| = |\Phi(u_n - u)| \leq \|\Phi\|_* \cdot \|u_n - u\| \text{ for all } \Phi \in X^*.$$

Consequently, if a sequence $\{u_n\}$ converges strongly to u, then it converges weakly to u. The converse is false.

Proposition 12 *Let $1 \leq p < \infty$ and q be the conjugate of p. Then $\{f_n\} \to f$ weakly in $L^p(E)$ if and only if*

$$\lim_{n \to \infty} \int_E g \cdot f_n \, dm = \int_E g \cdot f \, dm \text{ for all } g \in L^q(E).$$

Proof This follows immediately from Corollary 11, which characterizes all the bounded linear functionals on $L^p(E)$. $\square$

Theorem 13 *If $1 \leq p < \infty$, then every weakly convergent sequence in $L^p(E)$ is bounded in $L^p(E)$.*

Proof Let q be the conjugate of p. Assume $\{f_n\} \to f$ weakly in $L^p(E)$. For each n, let $\Phi_n \in [L^q(E)]^*$ be represented by f_n. Fix $g \in L^q(E)$ and let $\Psi \in [L^p(E)]^*$ be represented by g. Observe that

$$\Phi_n(g) = \int_E f_n \cdot g \, dm = \Psi(f_n) \text{ for all } n.$$

By weak convergence $\{\Psi(f_n)\} \to \Psi(f)$, and therefore, in particular $\{\Phi_n(g)\}$ is bounded. Since $L^q(E)$ is complete, according to the Uniform Boundedness Principle, $\{\Phi_n\}$ is bounded in $[L^q(E)]^*$. It follows from (3) (with p and q interchanged) that $\{f_n\}$ is bounded in $L^p(E)$. $\square$

Theorem 14 *If $1 \leq p < \infty$, then every bounded sequence in $L^p(E)$ has a subsequence that converges weakly to a function in $L^p(E)$.*

Proof Let q be the conjugate of p. Let $\{f_n\}$ be a bounded sequence in $L^p(E)$. For each n, let the functional $\Phi_n \in [L^q(E)]^*$ be represented by f_n. It follows from (3) (with p and q interchanged) that $\{\Phi_n\}$ is a bounded sequence in $[L^q(E)]^*$. Since $1 < q < \infty$, according to Theorem 9 in the preceding chapter, $L^q(E)$ is separable. By Helly's Theorem, there is a subsequence $\{\Phi_{n_k}\}$ that converges pointwise on $L^q(E)$ to a functional $\Phi \in [L^q(E)]^*$. The Riesz Representation Theorem (with p and q interchanged) implies that Φ is represented by a function $f \in L^p(E)$. Observe that

$$\lim_{k \to \infty} \int_E g \cdot f_{n_k} \, dm = \lim_{k \to \infty} \Phi_{n_k}(g) = \Phi(g) = \int_E g \cdot f \, dm \text{ for all } g \in L^q(E).$$

According to the preceding proposition, $\{f_{n_k}\} \to f$ weakly in $L^p(E)$. $\square$

Proposition 15 *For $1 \leq p < \infty$, if $\{f_n\} \to f$ weakly in $L^p(E)$, then*

$$\|f\|_p \leq \liminf \|f_n\|_p. \tag{15}$$

Proof Let q be the conjugate of p and $f^* \in L^q(E)$ be the dual function of f, so that $\|f\|_p = \int_E f^* \cdot f \, dm$ and $\|f^*\|_q = 1$. By Hölder's Inequality,

$$\int_E f^* \cdot f_n \, dm \leq \|f^*\|_q \cdot \|f_n\|_p = \|f_n\|_p \text{ for all } n.$$

Since $\{f_n\} \to f$ weakly

$$\|f\|_p = \int_E f^* \cdot f = \lim_{n \to \infty} \int_E f^* \cdot f_n \, dm \leq \liminf \|f_n\|_p.$$

$\square$

Definition *A subset K of a normed linear space X is said to be* **weakly sequentially compact** *provided that every sequence $\{u_n\}$ in K has a subsequence that converges weakly to a point u in K.*

Example If $1 < p < \infty$, then $B \equiv \{f \in L^p(E) \mid \|f\|_p \leq 1\}$ is a weakly sequentially compact subset of $L^p(E)$. To verify this, let $\{f_n\}$ be a sequence in B. The preceding theorem implies that there is a subsequence $\{f_{n_k}\}$ that converges weakly to $f \in L^p(E)$. Moreover, $\|f\|_p \leq 1$ since, by (15), $\|f\|_p \leq \liminf \|f_{n_k}\|_p \leq 1$.

We conclude this section by providing two criteria for justifying weak convergence.

Theorem 16 *For $1 < p < \infty$, assume $\{f_n\}$ is a bounded sequence in $L^p(E)$ and $f \in L^p(E)$. Then $\{f_n\} \to f$ weakly in $L^p(E)$ if and only if for every subset A of E of finite measure,*

$$\lim_{n \to \infty} \int_A f_n \, dm = \int_A f \, dm. \tag{16}$$

Proof Let $\Phi \in [L^q(E)]^*$ be represented by f and for each n, $\Phi_n \in [L^q(E)]^*$ be represented by f_n. It follows from Proposition 12 that $\{f_n\} \to f$ weakly in $L^p(E)$ if and only if the sequence of functionals $\{\Phi_n\}$ converges pointwise L^q to Φ. Now $1 < q < \infty$ and therefore, by assertion (i) of Theorem 8 in the preceding chapter, $S = \{\chi_A \mid A \subseteq E, m(a) < \infty\}$ is a total subset of $L^q(E)$. According to Proposition 7, a bounded sequence of functionals on $L^q(E)$ converges pointwise on $L^q(E)$ to a bounded linear functional if and only if it converges pointwise on a total subset of $L^q(E)$. This completes the proof. □

By appealing to assertion (iii), rather than assertion (i), of Theorem 8 in the preceding chapter, the proof of the above theorem provides a proof of the following.

Theorem 17 *For $1 < p < \infty$, suppose $\{f_n\}$ is a bounded sequence in $L^p[a, b]$ and $f \in L^p[a, b]$. Then $\{f_n\} \to f$ weakly in $L^p[a, b]$ if and only if*

$$\lim_{n \to \infty} \left[\int_a^x f_n \, dm \right] = \int_a^x f \, dm \text{ for all } x \in [a, b]. \tag{17}$$

Lemma 18 (the Riemann-Lebesgue Lemma) *Let $I = [-\pi, \pi]$ and for each n, define $f_n(x) = \sin nx$ for $x \in I$. Then $\{f_n\} \to f \equiv 0$ weakly in $L^2(I)$, that is,*

$$\lim_{n \to \infty} \int_{-\pi}^x \sin nt \cdot g(t) \, dt = 0 \text{ for all } g \in L^2(I).$$

Proof Indeed, observe that for each n, $|f_n| \leq 1$ on I, and

$$\lim_{n \to \infty} \int_{-\pi}^x \sin nt \, dt = \lim_{n \to \infty} -\frac{1}{n}[\cos nx - (-1)^n] = 0 \text{ for all } x \in I.$$

It follows from the preceding theorem that $\{f_n\} \to f$ weakly in $L^2(I)$. □

The functions defined in the following example are called **Rademacher functions**.

Example For $I = [0, 1]$ and each n, define the step-function f_n on I by

$$f_n(x) = (-1)^k \text{ for } k/2^n \leq x < (k+1)/2^n \text{ where } 0 \leq k < 2^n - 1.$$

We leave it as an exercise to deduce from Theorem 17 that $\{f_n\} \to f \equiv 0$ weakly in $L^2(I)$. However, no subsequence converges either strongly or pointwise almost everywhere on $[0, 1]$ to $f \equiv 0$.

PROBLEMS

24. Characterize the bounded linear functionals on Euclidean space $\mathbf{R}^n$. Show that for a sequence in $\mathbf{R}^n$, the concepts of strong convergence, weak convergence, and componentwise convergence are the same.
25. For $1 \leq p < \infty$, assume that both $\{f_n\} \to h$ and $\{f_n\} \to g$ weakly $L^p(E)$. Show that $g = h$.
26. For $1 \leq p \leq \infty$, and q the conjugate of p, assume that $\{f_n\} \to f$ weakly in $L^p(E)$ and $\{g_n\} \to g$ strongly in $L^q(E)$. Show that

$$\lim_{n\to\infty} \int_E f_n \cdot g_n \, dm = \int_E f \cdot g \, dm.$$

27. At what point does the proof of Theorem 14 fail in the case $p = 1$ and also in the case $p = \infty$.
28. Show that if a sequence in $C[a, b]$ converges weakly then it converges pointwise. Find a bounded sequence $C[a, b]$ that fails to have any weakly convergent subsequence.
29. Verify the assertions made regarding the convergence of the Radamacher functions.
30. If $\{f_n\} \to f$ weakly in $L^\infty[a, b]$, show that

$$\lim_{n \to \infty} \int_a^x f_n \, dm = \int_a^x f \, dm \text{ for all } x \in [a, b].$$

31. Fix real numbers α and β. For each natural numbers k and n, consider the step-function f_n defined on $I = [0, 1]$ by

$$f_n(x) = (1 - (-1)^k)\alpha/2 + (1 + (-1)^k)\beta/2 \text{ for } k/2^n \leq x < (k+1)/2^n, \ 0 \leq k < 2^n - 1.$$

    Show that $\{f_n\}$ converges weakly in $L^2(I)$ to the constant function that takes the value $(\alpha + \beta)/2$. For $\alpha \neq \beta$, show that no subsequence of $\{f_n\}$ converges strongly in $L^2(I)$.
32. (Radon-Riesz) Show that $\{f_n\} \to f$ strongly in $L^2(E)$ if and only $\{f_n\} \to f$ weakly in $L^2(E)$ and $\{\|f_n\|_2\} \to \|f\|_2$.
33. Use Problem 20 to show that, for $1 \leq p < \infty$, a bounded sequence in ℓ^p converges weakly if and only if it converges componentwise.

8.4 THE MINIMIZATION OF CONVEX FUNCTIONALS

A subset C of a normed linear space X is said to be **closed** provided that if $\{u_n\}$ is a sequence in C that converges strongly to u, then $u \in C$, and said to be **convex** provided that whenever $u, v \in C$ and $\lambda \in [0, 1]$, then $\lambda u + (1 - \lambda)v \in C$.

Example For $1 \leq p \leq \infty$ and g a non-negative function in $L^p(E)$, the set

$$C = \{f\colon E \to \mathbf{R} \text{ measurable} \mid \ |f| \leq g \text{ almost everywhere on } E\}$$

is a closed, bounded, convex subset of $L^p(E)$. Indeed, we see from the integral comparison test that C is a bounded and convex subset of $L^p(E)$. To verify that C is closed, let $\{f_n\}$ be a sequence in $L^p(E)$ that converges in $L^p(E)$ to f. By the Riesz-Fischer Theorem, there is a subsequence of $\{f_n\}$ that converges pointwise almost everywhere on E to f. From this pointwise convergence, it follows that $f \in C$.

Example For $1 \leq p \leq \infty$, the set $B = \{f \in L^p(E) \mid \|f\|_p \leq 1\}$ is a closed, bounded, convex subset of $L^p(E)$. Clearly C is convex and bounded. To see that B is closed, observe that if $\{f_n\}$ is a sequence in B that converges in $L^p(E)$ to f, then, by the Minkowski Inequality, $|\|f_n\|_p - \|f\|_p| \leq \|f_n - f\|_p$ for all n. Therefore $f \in B$.

We have seen examples, for instance the sequence of Radamacher functions, of sequences that converge weakly but have no strongly convergent subsequence. In view of this, the following theorem is rather surprising.

The Banach-Saks Theorem *If $1 < p < \infty$ and $\{f_n\} \to f$ weakly in $L^p(E)$, then there is a subsequence $\{f_{n_k}\}$ for which the sequence of arithmetic means converges strongly in $L^p(E)$ to f, that is,*

$$\lim_{k \to \infty} \frac{f_{n_1} + f_{n_2} + \ldots + f_{n_k}}{k} = f \text{ strongly in } L^p(E).$$

Proof We present the proof for the case $p = 2$ [2]. By replacing each f_n with $f_n - f$, we suppose $f \equiv 0$. According to Proposition 13, $\{f_n\}$ is bounded in $L^2(E)$. Choose $M \geq 0$ for which

$$\int_E f_n^2 \, dm \leq M \text{ for all } n.$$

We will inductively choose a subsequence $\{f_{n_k}\}$ such that

$$\int_E (f_{n_1} + \ldots + f_{n_k})^2 \, dm \leq (2 + M)k \text{ for all } k.$$

Indeed, define $n_1 = 1$. Suppose we have chosen indices $n_1 < n_2 < \ldots < n_k$ such that

$$\int_E (f_{n_1} + \ldots + f_{n_j})^2 \, dm \leq (2 + M)j \text{ for } j = 1, \ldots, k.$$

Since $f_{n_1} + \ldots + f_{n_k}$ belongs to $L^2(E)$ and $\{f_n\}$ converges weakly in $L^2(E)$ to 0, we may choose an index $n_{k+1} > n_k$ for which

$$\int_E (f_{n_1} + \ldots + f_{n_k}) \cdot f_{n_{k+1}} \, dm \leq 1. \tag{18}$$

[2] For $p \neq 2$, the proof is similar in spirit but technically more challenging.

However,

$$\int_E (f_{n_1}+\ldots+f_{n_{k+1}})^2\,dm = \int_E (f_{n_1}+\ldots+f_{n_k})^2\,dm$$
$$+\,2\int_E (f_{n_1}+\ldots+f_{n_k})\cdot f_{n_{k+1}}\,dm + \int_E f_{n_{k+1}}^2\,dm,$$

and therefore

$$\int_E (f_{n_1}+\ldots+f_{n_{k+1}})^2\,dm \le (2+M)k+2+M=(2+M)(k+1).$$

The subsequence $\{f_{n_k}\}$ has been inductively chosen so that

$$\int_E \left[\frac{f_{n_1}+f_{n_2}+\ldots+f_{n_k}}{k}\right]^2 dm \le \frac{(2+M)}{k} \text{ for all } k.$$

Therefore, the sequence of arithmetic means of $\{f_{n_k}\}$ converges strongly in $L^2(E)$ to $f\equiv 0$. $\square$

Corollary 19 *If $1<p<\infty$ and C is a closed, bounded, convex subset of $L^p(E)$, then C is weakly sequentially compact.*

Proof Let $\{f_n\}$ be a sequence in C. Since C is bounded, according to Theorem 14, a subsequence of $\{f_n\}$ converges weakly to $f\in L^p(E)$. For notational convenience, assume $\{f_n\}\to f$ weakly. According to the Banach-Saks Theorem, there is a subsequence of $\{f_n\}$ for which the sequence of arithmetic means converges strongly to f. Since C is convex, each arithmetic mean belongs to C, and therefore, since, by assumption, C is closed, $f\in C$. $\square$

We now consider functionals that are not necessarily linear. A functional $\Phi\colon C\to\mathbf{R}$ defined on a convex subset C of a normed linear space X is said to be **convex** provided that for $u,v\in C$ and $0\le\lambda\le 1$,

$$\Phi(\lambda u+(1-\lambda)v)\le\lambda\Phi(u)+(1-\lambda)\Phi(v).$$

Lemma 20 *Let $1<p<\infty$, C be a closed, bounded, convex subset of $L^p(E)$ and the functional $\Phi\colon C\to\mathbf{R}$ be continuous and convex. If $\{f_n\}$ is a sequence in C that converges weakly to $f\in C$, then*

$$\Phi(f)\le\liminf\Phi(f_n).$$

Proof By possibly taking a subsequence, we may suppose that $\{\Phi(f_n)\}\to\liminf\Phi(f_n)$. The arithmetic means of a convergent sequence of real numbers converge to the same limit and so

$$\lim_{k\to\infty}\frac{\Phi(f_1)+\Phi(f_2)+\cdots+\Phi(f_k)}{k}=\liminf\Phi(f_n).$$

According to the Banach-Saks Theorem, since C is convex, by possibly taking a subsequence of $\{f_n\}$, we may assume that the sequence of arithmetic means of $\{f_n\}$ converges strongly in $L^p(E)$ to f. Since the functional Φ is continuous,

$$\Phi(f) = \lim_{k \to \infty} \Phi\left(\frac{f_1 + f_2 + \cdots + f_k}{k}\right).$$

On the other hand, since Φ is convex, for each k,

$$\Phi\left(\frac{f_1 + f_2 + \cdots + f_k}{k}\right) \leq \frac{\Phi(f_1) + \Phi(f_2) + \cdots + \Phi(f_k)}{k}.$$

Consequently,

$$\begin{aligned}\Phi(f) &= \lim_{k \to \infty} \Phi\left(\frac{f_1 + f_2 + \cdots + f_k}{k}\right) \\ &\leq \lim_{k \to \infty} \frac{\Phi(f_1) + \Phi(f_2) + \cdots + \Phi(f_k)}{k} = \liminf \Phi(f_n).\end{aligned}$$

□

Theorem 21 *Let $1 < p < \infty$, C be a closed, bounded, convex subset of $L^p(E)$ and the functional $\Phi\colon C \to \mathbf{R}$ be continuous and convex. Then $\Phi\colon C \to \mathbf{R}$ has a minimum value.*

Proof Since C is weakly sequentially compact, by possibly taking a subsequence, we may assume that $\{\Phi(f_n)\} \to \inf\{\Phi(f) \mid f \in C\}$, $\{f_n\} \to f$ weakly, and $f \in C$. According to the preceding lemma,

$$\Phi(f) \leq \liminf \Phi(f_n) = \inf\{\Phi(f) \mid f \in C\}.$$

The functional Φ attains a minimum value on C at f. □

Corollary 22 (**Beppo Levi**) *Let $1 < p < \infty$ and C be a closed, bounded, convex subset of $L^p(E)$. For each $g \in L^p(E)$, there is a function in $f_0 \in C$ that is closest to g, in the sense that*

$$\|f_0 - g\|_p \leq \|f - g\|_p \text{ for all } f \in C.$$

Proof Apply the preceding theorem to the continuous, convex functional $\Phi\colon C \to \mathbf{R}$ defined by $\Phi(f) = \|f - g\|_p$ for $f \in C$. □

We conclude Part 1 with an application of Theorem 21 to a non-linear integral functional.

Lemma 23 *Let $\varphi\colon \mathbf{R} \to \mathbf{R}$ be a continuous function. For $1 \leq p < \infty$, assume there are $c_1, c_2 \geq 0$ for which*

$$|\varphi(s)| \leq c_1 + c_2 \cdot |s|^p \text{ for all } s. \tag{19}$$

If $m(E) < \infty$, then

$$\Phi(f) \equiv \int_E \varphi \circ f \, dm \text{ for } f \in L^p(E) \tag{20}$$

defines a continuous functional $\Phi\colon L^p(E) \to \mathbf{R}$.

Proof Fix $f \in L^p(E)$. Since φ is continuous, the composition $\varphi \circ f$ is measurable. Observe that $\varphi \circ f \in L^p(E)$ since $m(E) < \infty$ and $|\varphi \circ f| \leq c_1 + c_2 \cdot |f|^p$ almost everywhere on E. According to Proposition 5 of the preceding chapter, every convergent sequence contains a rapidly Cauchy subsequence. Consequently, to verify the continuity of Φ at f it suffices to consider a sequence $\{f_n\}$ in $L^p(E)$ that converges in $L^p(E)$ to f and also is rapidly Cauchy, and verify that $\{\Phi(f_n)\}$ converges to $\Phi(f)$. Consider such a sequence. According to Theorem 4 of the preceding chapter, since $\{f_n\}$ is rapidly Cauchy, $\{f_n\}$ converges pointwise almost everywhere on E to f, and is dominated by a function in $L^p(E)$, in the sense that there is a $g \in L^p(E)$ for which

$$|f_n| \leq g \text{ almost everywhere on } E \text{ for all } n.$$

Since φ is continuous, $\{\varphi \circ f_n\}$ converges pointwise almost everywhere on E to $\varphi \circ f$. Moreover, it follows from inequality (19) that

$$|\varphi(f_n(x))| \leq c_1 + c_2 \cdot |f_n(x)|^p \leq c_1 + c_2 \cdot g(x)^p \text{ for almost all } x \in E.$$

However, since $m(E) < \infty$, $c_1 + c_2 \cdot g^p$ is integrable and therefore, by the Dominated Convergence Theorem,

$$\lim_{n \to \infty} \int_E \varphi \circ f_n \, dm = \int_E \varphi \circ f \, dm.$$

Consequently, Φ is continuous at f. $\square$

Theorem 24 *Let the function $\varphi\colon \mathbf{R} \to \mathbf{R}$ be both continuous and convex. For $1 < p < \infty$, assume there are $c_1, c_2 \geq 0$ for which*

$$|\varphi(s)| \leq c_1 + c_2 \cdot |s|^p \text{ for all } s.$$

Assume that $m(E) < \infty$ and let C be a closed, bounded, convex subset of $L^p(E)$. Then there is an $f_0 \in C$ for which

$$\int_E \varphi \circ f_0 \, dm \leq \int_E \varphi \circ f \, dm \text{ for all } f \in C.$$

Proof Define the function $\Phi\colon C \to \mathbf{R}$ by $\Phi(f) = \int_E \varphi \circ f \, dm$ for all $f \in C$. According to the preceding lemma, Φ is continuous. The functional Φ inherits convexity from the function φ. An appeal to the preceding theorem implies that $\Phi\colon C \to \mathbf{R}$ takes a minimum value. $\square$

The proof that a continuous real-valued function on a closed, bounded subset of Euclidean space takes a minimum value depends on the Bolzano-Weierstrass Theorem which, as already noted, does not extend, with respect to strong convergence, to any infinite dimensional space normed linear space[3]. Nevertheless, up until the late-19th century, it had been uncritically assumed that a similar minimization principle was valid for establishing

[3] Yet another theorem of Riesz, which is proven in Chapter 17, implies that in every infinite dimensional normed linear space there is a bounded sequence for which no subsequence converges strongly. In Chapter 13, we prove that every continuous real-valued function on a metric space has a minimum value if and only if the space is compact.

minimum values of real-valued functionals that were bounded below and defined on infinite dimensional spaces of functions. Karl Weierstrass observed the fallacy in this assumption. He exhibited an explicit integral functional of the type considered above, which, while bounded below, did not take a minimum value. Many mathematicians turned their attention to investigating classes of functionals for which it is possible to prove the existence of minimizers and new methods were introduced. Among these, David Hilbert showed that every bounded sequence in the space ℓ^2 has a subsequence that converges componentwise to a member of ℓ^2, that is, he proved Corollary 6, a forerunner of weak sequential compactness[4]. Weak sequential compactness was the basis of the proof of Theorem 21. In Chapter 14, we prove the Dunford-Pettis Theorem, which asserts that a bounded, uniformly integrable sequence in $L^1(E)$ has a weakly convergent subsequence. A different approach to the study of compactness is to characterize, in a specific normed linear space X, those closed, bounded subsets that are sequentially compact with respect to strong sequential convergence. Also in Chapter 14, the Arzelá-Ascoli Theorem is proven, which provides such a characterization in, say, $C[a, b]$, normed with the maximum norm.

PROBLEMS

34. Show that if a sequence of real numbers $\{a_n\}$ converges to $a \in \overline{\mathbf{R}}$, then its sequence of arithmetic means also converges to a.
35. If $\Phi \in X^*$, prove that
$$\inf\{\Phi(x) \mid \|x\| \leq 1\} = -\|\Phi\|_*.$$
36. For $1 < p < \infty$ define $C = \{f \in L^p(E) \mid \|f\|_p \leq 1.\}$ For $\Phi \in [L^p(E)]^*$, identify a function $f_0 \in C$ at which the functional $\Phi\colon C \to \mathbf{R}$ has a minimum value.
37. Let $C = \{f \in L^1[0, 1] \mid \|f\|_1 \leq 1\}$, and define $\Phi\colon C \to \mathbf{R}$ by $\Phi(f) = \int_0^1 x f(x)\, dx$. Show that C is a closed, bounded, convex subset of $L^1[0, 1]$, that $\Phi\colon C \to \mathbf{R}$ is continuous and convex, but does not have a minimum value.
38. Let $C = \{f \in C[0, 1 \mid \|f\|_{\max} \leq 1\}$, and define $\Phi\colon C \to \mathbf{R}$ by $\Phi(f) = \int_0^{1/2} f(x)\, dx - \int_{1/2}^1 f(x)\, dx$. Show that C is a closed, bounded, convex subset of $C[0, 1]$, that $\Phi\colon C \to \mathbf{R}$ is continuous and convex, but does not have a minimum value.
39. State and prove the Banach-Saks Theorem in ℓ^2.
40. (Nemytskii) Let $m(E) < \infty$ and $p_1, p_2 \in [1, \infty)$. Suppose $\varphi\colon \mathbf{R} \to \mathbf{R}$ is continuous and there are c_1, $c_2 \geq 0$ for which $|\varphi(s)| \leq c_1 + c_2|s|^{p_1/p_2}$ for all s. Show that
$$\text{if } \{f_n\} \to f \text{ in } L^{p_1}(E), \text{ then } \{\varphi \circ f_n\} \to \varphi \circ f \text{ in } L^{p_2}.$$
41. Let C be a subset of a normed linear space X with the property that each sequence in C has a subsequence that converges strongly to a point in C. Show that C is a closed and bounded subset of X.

[4]Hilbert's article *On the Dirichlet Principle* is translated in *A Source Book in Classical Analysis* by Garrett Birkhoff (Harvard University Press, 1973).

This page intentionally left blank

PART TWO

MEASURE AND INTEGRATION: GENERAL THEORY

This page intentionally left blank

CHAPTER 9

General Measure Spaces: Their Properties and Construction

Contents

9.1 Measurable Sets and Measure Spaces 159
9.2 Measures Induced by an Outer-Measure 164
9.3 The Carathéodory-Hahn Theorem 168

A measure is a set-function $\mu\colon \mathcal{M} \to [0, \infty]$, defined on a σ-algebra $\mathcal{M}$ of subsets of a general set X, which is countably additive over countable, disjoint unions of sets in $\mathcal{M}$, and has $\mu(\emptyset) = 0$. The set X is not assumed to be ordered nor to possess a metric or topological structure. In the first section, we establish the properties of general measure spaces. In the following two sections, we show that the strategy we pursued for the construction of Lebesgue measure on $\mathbf{R}$ is feasible in general. First, we prove that any non-negative set-function μ defined on a collection of subsets of X induces an outer-measure μ^*, which, in turn, determines a measure we call the Carathéodory measure induced by μ. Then, we prove the Carathéodory-Hahn Theorem, which provides criteria for the Carathéodory measure to be an extension of μ, as it was in the case of the set-function length on the collection of bounded, open intervals of real numbers.

9.1 MEASURABLE SETS AND MEASURE SPACES

Recall that a σ-algebra of subsets of a set X is a collection of subsets of X that contains the empty-set and is closed with respect to complements in X and to countable unions, and therefore, by the De Morgan Identities, with respect to countable intersections.

Definition *A set-function $\mu\colon \mathcal{M} \to [0, \infty]$ on a σ-algebra $\mathcal{M}$ of subsets of X is called a* **measure** *provided that $\mu(\emptyset) = 0$ and it is* **countably additive**, *in the sense that for any countable, disjoint collection $\{E_k\}_{k=1}^{\infty}$ of sets in $\mathcal{M}$*

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k).$$

The triple $(X, \mathcal{M}, \mu)$ is called a **measure space** *and a set in $\mathcal{M}$ is said to be* **measurable**. *For a measurable set E, a countable, disjoint collection $\{E_k\}_{k=1}^{\infty}$ of measurable sets is called a* **measurable partition** *of E if $E = \bigcup_{n=1}^{\infty} E_n$. The measure μ, or the measure space $(X, \mathcal{M}, \mu)$, is said to be* **finite** *provided that $\mu(X) < \infty$, and said to be* **σ-finite** *provided that $X = \bigcup_{n=1}^{\infty} X_k$, where each X_k is a measurable set of finite measure.*

Of course, the primary example we have so far of a measure space is $(\mathbf{R}, \mathcal{L}, m)$, called **Lebesgue measure** on $\mathbf{R}$, where $\mathcal{L}$ is the collection of Lebesgue measurable subsets of $\mathbf{R}$, and m is Lebesgue measure. This is a σ-finite measure space. If $E \subseteq \mathbf{R}$ is Lebesgue measurable, and $\mathcal{M}$ is the σ-algebra of Lebesgue measurable subsets of E, we call $(E, \mathcal{M}, m)$ Lebesgue measure on E. Another example of a measure space is $(\mathbf{R}, \mathcal{B}, m)$, called **Borel measure on** $\mathbf{R}$, where $\mathcal{B}$ is the collection of Borel subsets of $\mathbf{R}$ and again m is Lebesgue measure. For any σ-algebra $\mathcal{M}$ of subsets of X, the **counting measure** $c\colon \mathcal{M} \to [0, \infty]$ assigns the measure of a finite set to be the number of elements in the set and otherwise assigns the value ∞. The counting measure on $\mathbf{R}$ is not σ-finite. For $x_0 \in X$, the **Dirac measure** $\delta_{x_0}\colon \mathcal{M} \to [0, \infty)$ assigns 1 to a set that contains x_0 and otherwise assigns the value 0.

Example Consider a non-negative, Lebesgue measurable function $f\colon \mathbf{R} \to [0, \infty]$. Define $\nu\colon \mathcal{L} \to [0, \infty]$ by[1]

$$\nu(E) = \int_E f\, dm \text{ for all } E \in \mathcal{L}.$$

The countable additivity over domains of Lebesgue integration may be reworded by stating that ν is a measure. It has the property that if $m(E) = 0$, then $\nu(E) = 0$, which is called the **absolute continuity** of ν with respect to m. According to the Radon-Nikodym Theorem, which is proved in the next chapter, a σ-finite measure ν of $\mathcal{L}$ that is absolutely continuous with respect to Lebesgue measure has the above integral representation by a non-negative Lebesgue measurable function.

Example Let $\mathcal{M}$ be the collection of all subsets of $\mathbf{R}^n$ and define the measure $\mu^{int}\colon \mathcal{M} \to [0, \infty]$, called the **integral count**, that assigns to a set the number of points in the set that have integer coordinates, if there are a finite number of such points, and otherwise assigns the value ∞. For $\epsilon > 0$ and $E \subseteq \mathbf{R}^n$, define the ϵ-dilation of E, $\epsilon \cdot E = \{\epsilon \cdot x \mid x \in E\}$, and $\mu^{int}_\epsilon(E) = \epsilon^{-n} \mu^{int}(\epsilon \cdot E)$. Using a short, elegant proof of von Neumann, in the next chapter we show that if $I \subseteq \mathbf{R}^n$ is the n-fold product of bounded intervals, then

$$\lim_{\epsilon \to \infty} \mu^{int}_\epsilon(I) = \text{vol}(I)$$

where $\text{vol}(I)$ is the product of the lengths of the edges. As we did in the case $n = 1$, in this way, side-step several bothersome details that often occur in the construction of Lebesgue measure on $\mathbf{R}^n$.

Proposition 1 *Let $(X, \mathcal{M}, \mu)$ be a measure space.*

(Finite Additivity) For any finite, disjoint collection $\{E_k\}_{k=1}^n$ of measurable sets,

$$\mu\left(\bigcup_{k=1}^n E_k\right) = \sum_{k=1}^n \mu(E_k).$$

(Monotonicity) If A and B are measurable sets and $A \subseteq B$, then

$$\mu(A) \leq \mu(B).$$

[1] The integral over $\emptyset$ is defined to be zero.

(Excision) Moreover, if $A \subseteq B$ and $\mu(A) < \infty$, then

$$\mu(B \sim A) = \mu(B) - \mu(A),$$

so that if $\mu(A) = 0$, then

$$\mu(B \sim A) = \mu(B).$$

(Countable Monotonicity) For any countable collection $\{E_k\}_{k=1}^{\infty}$ of measurable sets that covers a measurable set E,

$$\mu(E) \leq \sum_{k=1}^{\infty} \mu(E_k).$$

Proof Finite additivity follows from countable additivity by setting $E_k = \emptyset$, so that $\mu(E_k) = 0$, for $k > n$. By finite additivity, since $A \cup B = A \cup (B \sim A)$, a disjoint union of measurable sets,

$$\mu(A \cup B) = \mu(A) + \mu(B \sim A),$$

which immediately implies monotonicity and excision. To verify countable monotonicity, let $\{E_k\}_{k=1}^{\infty}$ be a collection of measurable sets that covers the measurable set E. We first show that there is a measurable partition $\{E'_k\}_{k=1}^{\infty}$ of $\bigcup_{k=1}^{\infty} E_k$ for which each $E'_k \subseteq E_k$. Indeed, define $E'_1 = E_1$ and

$$E'_k = E_k \sim \left[\bigcup_{i=1}^{k-1} E_i\right] \quad \textit{for } k \geq 2.$$

This defines a countable, disjoint collection of measurable sets, and it also covers E, since for each $x \in E$ there is a first index k for which $x \in E_k$, and so $x \in E'_k$. It follows from the countable additivity and monotonicity of μ that

$$\mu(E) \leq \mu\left(\bigcup_{k=1}^{\infty} E'_k\right) = \sum_{k=1}^{\infty} \mu(E'_k) \leq \sum_{k=1}^{\infty} \mu(E_k).$$

□

A countable collection of sets $\{E_k\}_{k=1}^{\infty}$ is said to be **ascending** provided that for each index k, $E_k \subseteq E_{k+1}$, and said to be **descending** provided that for each k, $E_{k+1} \subseteq E_k$.

Theorem 2 (Continuity of Measure) *Let $(X, \mathcal{M}, \mu)$ be a measure space.*

(i) If $\{A_k\}_{k=1}^{\infty}$ is an ascending collection of measurable sets, then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k \to \infty} \mu(A_k). \tag{1}$$

(ii) If $\{B_k\}_{k=1}^{\infty}$ is a descending collection of measurable sets, and $\mu(B_1) < \infty$, then

$$\mu\left(\bigcap_{k=1}^{\infty} B_k\right) = \lim_{k \to \infty} \mu(B_k). \tag{2}$$

Proof We first verify (1). If there is an index k_0 for which $\mu(A_{k_0}) = \infty$, then, by the monotonicity of μ, $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \infty$. Therefore, (1) holds, since each side equals ∞. It remains to consider the case that $\mu(A_k) < \infty$ for all k. Define $A_0 = \emptyset$ and then define $C_k = A_k \sim A_{k-1}$ for each $k \geq 1$. Since the collection $\{A_k\}_{k=1}^{\infty}$ is ascending, the collection $\{C_k\}_{k=1}^{\infty}$ is a measurable partition of $\bigcup_{k=1}^{\infty} A_k$. By the countable additivity of μ,

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \mu\left(\bigcup_{k=1}^{\infty} C_k\right) = \sum_{k=1}^{\infty} \mu(A_k \sim A_{k-1}). \tag{3}$$

By the excision property of measure,

$$\begin{aligned} \sum_{k=1}^{\infty} \mu(A_k \sim A_{k-1}) &= \sum_{k=1}^{\infty} [\mu(A_k) - \mu(A_{k-1})] \\ &= \lim_{n \to \infty} \sum_{k=1}^{n} [\mu(A_k) - \mu(A_{k-1})] \\ &= \lim_{n \to \infty} [\mu(A_n) - \mu(A_0)]. \end{aligned} \tag{4}$$

But $\mu(A_0) = \mu(\emptyset) = 0$. As a consequence, (1) follows from (3) and (4).

To prove (2), for each index k, define $D_k = B_1 \sim B_k$. Since the collection $\{B_k\}_{k=1}^{\infty}$ is descending, the collection $\{D_k\}_{k=1}^{\infty}$ is ascending. By part (i),

$$\mu\left(\bigcup_{k=1}^{\infty} D_k\right) = \lim_{k \to \infty} \mu(D_k).$$

According to De Morgan's Identities,

$$\bigcup_{k=1}^{\infty} D_k = \bigcup_{k=1}^{\infty} [B_1 \sim B_k] = B_1 \sim \bigcap_{k=1}^{\infty} B_k.$$

On the other hand, by the excision property of measure, for each k, since $\mu(B_k) < \infty$, $\mu(D_k) = \mu(B_1) - \mu(B_k)$. Therefore,

$$\mu\left(B_1 \sim \bigcap_{k=1}^{\infty} B_k\right) = \lim_{n \to \infty} [\mu(B_1) - \mu(B_n)].$$

Once more using the excision property of measure, we obtain (2). □

For a measure space $(X, \mathcal{M}, \mu)$ and a measurable set E, a property is said to hold **almost everywhere** on E, or it holds **for almost all** x in E, provided that it holds on $E \sim E_0$, where E_0 is a measurable subset of E for which $\mu(E_0) = 0$.

The Borel-Cantelli Lemma *Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{E_k\}_{k=1}^{\infty}$ be a countable collection of measurable sets for which $\sum_{k=1}^{\infty} \mu(E_k) < \infty$. Then almost all points in X belong to at most a finite number of the E_k's.*

Proof By the countable monotonicity of μ, $\mu\left(\bigcup_{k=1}^{\infty} E_k\right) \le \sum_{k=1}^{\infty} \mu(E_k) < \infty$. Therefore, by the continuity of μ and again by its countable monotonicity,

$$\mu\left(\bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} E_k\right]\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=n}^{\infty} E_k\right) \le \lim_{n\to\infty} \sum_{k=n}^{\infty} \mu(E_k) = 0.$$

Observe that $\bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} E_k\right]$ is the set of points in X that belong to an infinite number of the E_k's. □

The Ascoli-Young Lemma *Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{E_k\}_{k=1}^{\infty}$ be a countable collection of measurable sets for which $\mu\left(\bigcup_{k=1}^{\infty} E_k\right) < \infty$. Assume that there is an $\epsilon > 0$ such that $\mu(E_k) \ge \epsilon$ for all k. Then the set of points in X that belong to infinitely many of the E_k's has measure at least ϵ.*

Proof By continuity and monotonicity properties of measure

$$\mu\left(\bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} E_k\right]\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=n}^{\infty} E_k\right) \ge \epsilon.$$

Again observe that $\bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} E_k\right]$ is the set of points in X that belong to an infinite number of the E_k's. □

Definition *A measure space $(X, \mathcal{M}, \mu)$ is said to be* **complete** *provided that if $E \in \mathcal{M}$ and $\mu(E) = 0$, then every subset of E also belongs to $\mathcal{M}$.*

In Chapter 2, we proved that Lebesgue measure on $\mathbf{R}$ is complete. And, as we saw by employing the Cantor-Lebesgue function, the Cantor set which, being closed, is a Borel set, contains a subset that is not a Borel set. Consequently, Borel measure on $\mathbf{R}$ is not complete. Completeness plays an important part in the extension of the Tonelli and Fubini Theorems from the Riemann to the Lebesgue integral for functions on $\mathbf{R}^k \times \mathbf{R}^m$.

PROBLEMS

1. Define $\mu\colon 2^{\mathbf{N}} \to [0, \infty)$ by $\mu(E) = \sum_{n\in E} 1/2^n$. Show that this is a finite measure.
2. For each k, let $(X, \mathcal{M}, \mu_k)$ be a measure space and suppose that there is a $c \ge 0$ for which $\mu_k(X) \le c$ for all k. Show that $\mu = \sum_{k=1}^{\infty} \frac{\mu_k}{2^k}$ also is a measure.
3. Let $\mathcal{M}$ be a σ-algebra and the set-function $\mu\colon \mathcal{M} \to [0, \infty)$ be finitely additive. Prove that μ is a measure if and only if whenever $\{A_k\}_{k=1}^{\infty}$ is an ascending collection of sets in $\mathcal{M}$, then (1) holds.
4. Let $(X, \mathcal{M}, \mu)$ be a finite measure space. The symmetric difference, $A \,\Delta\, B$, of two measurable sets A and B is defined by

$$A \,\Delta\, B = [A \sim B] \cup [B \sim A] = [A \cup B] \sim [A \cap B].$$

Identify two measurable sets whose symmetric difference has measure zero, and prove the following triangle inequality on $\mathcal{M}$: For measurable sets A, B and C,

$$\mu(A \Delta C) \leq \mu(A \Delta B) + \mu(B \Delta C).$$

The σ-algebra $\mathcal{M}$, with the metric $\rho(A,\ B) = \mu(A \Delta B)$, is called the Nikodym metric space.

5. Let $\mathcal{M}$ be the collection of subsets of X that are either countable or have a countable complement in X. Show that $\mathcal{M}$ is a σ-algebra. For $E \in \mathcal{M}$, define $\mu(E) = 0$ if E is countable and $\mu(E) = 1$, if E has a countable complement. Show that μ is a measure. Is this measure space complete?
6. Let $\mathcal{S}$ be any collection of subsets of X and define $\mathcal{M}$ to be the intersection of all σ-algebras of subsets of X that contain $\mathcal{S}$. Show that $\mathcal{M}$ is a σ-algebra, and it is the smallest σ-algebra that contains $\mathcal{S}$, in the sense that it is contained in any σ-algebra that contains $\mathcal{S}$. For $X = \mathbf{R}$, find the smallest σ-algebra that contains all singleton sets.
7. Let $\mathcal{M}$ be a σ-algebra. Verify the following:
   (i) If μ and ν are measures on $\mathcal{M}$, then the set-function λ defined on $\mathcal{M}$ by $\lambda(E) = \mu(E) + \nu(E)$ also is a measure.
   (ii) If μ and ν are measures on $\mathcal{M}$ and $\mu \geq \nu$, then there is a measure λ on $\mathcal{M}$ for which $\mu = \nu + \lambda$.
   (iii) If ν is σ-finite, the measure λ in (ii) is unique.
   (iv) Show that in general the measure λ need not be unique but that there is always a smallest such λ.

9.2 MEASURES INDUCED BY AN OUTER-MEASURE

Here, we introduce the general notion of a set-function called outer-measure, and of measurability of a set with respect to such a function. The Carathéodory strategy for the construction of Lebesgue measure on $\mathbf{R}$ as the restriction of Lebesgue outer-measure is feasible in general.

Definition *A set-function μ^*: $2^X \to [0, \infty]$ is said to be an* **outer-measure** *provided that $\mu^*(\emptyset) = 0$, and it is countably monotone, in the sense that if E is covered by the countable collection $\{E_k\}_{k=1}^\infty$ of subsets of X, then*

$$\mu^*(E) \leq \sum_{k=1}^{\infty} \mu^*(E_k).$$

A subset E of X is said to be measurable with respect to μ^* provided that for every $A \subseteq X$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \sim E).$$

So, measurability of a set E means that μ^* is additive over particular disjoint decompositions of any subset A of X, specifically, as $A = (A \cap E) \cup (A \sim E)$. Clearly the empty-set and the whole space X are always measurable with respect to an outer-measure. As we saw with Vitali's example for Lebesgue outer-measure on $X = \mathbf{R}$, there may be subsets of X that are not measurable. Since $\mu^*(\emptyset) = 0$, an outer-measure is finitely monotone and, in particular, is monotone. Consequently, to verify that $E \subseteq X$ is measurable with respect to μ^*, it is sufficient to show that

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \sim E) \text{ for all } A \subseteq X \text{ for which } \mu^*(A) < \infty.$$

Obviously a subset E of X is measurable if and only if its complement in X is measurable and, by the monotonicity of μ^*, every set of outer-measure zero is measurable.

The considerations in Chapter 2 may be divided into two. On the one hand, there are those based on particular properties of $\mathbf{R}$. We considered open subsets of $\mathbf{R}$, each of which has a countable, measurable partition into open intervals, and the definition of an interval depends on the order structure. That the outer-measure of a bounded interval is equal to its length was based on the Heine-Borel Theorem. On the other hand, other considerations depended solely on the De Morgan Identities: if $\{E_\lambda\}_{\lambda \in \Lambda}$ is any collection of subsets of a set X, parametrized by a set Λ, then

$$X \sim \bigcup_{\lambda \in \Lambda} E_\lambda = \bigcap_{\lambda \in \Lambda} (X \cap E_\lambda) \text{ and } X \sim \bigcap_{\lambda \in \Lambda} E_\lambda = \bigcup_{\lambda \in \Lambda} (X \cup E_\lambda).$$

The proofs in Chapter 2 that Lebesgue outer-measure on $\mathbf{R}$ is countably monotone, and the measurable subsets of $\mathbf{R}$ are a σ-algebra on which outer-measure is countably additive, depended solely on these identities. By repeating these same arguments, which, of course, are important, but not particularly illuminating, we have the following theorem.

Theorem 3 *Let $\mu^*: 2^X \to [0, \infty]$ be an outer-measure. Then the collection $\mathcal{M}$ of sets that are measurable with respect to μ^* is a σ-algebra. If μ is the restriction of μ^* to $\mathcal{M}$, then $(X, \mathcal{M}, \mu)$ is a complete measure space, and every set of outer-measure zero is measurable.*

The construction of Lebesgue outer-measure on $\mathbf{R}$ proceeded by first defining the set-function length on the collection of open, bounded intervals, and then defining the outer-measure of a set to be the infimum of sums of lengths of countable collections of open, bounded intervals that cover the set. A similar construction of outer-measure works in general.

Theorem 4 *Let the set-function $\mu: \mathcal{S} \to [0, \infty]$ be defined on a non-empty collection $\mathcal{S}$ of subsets of X. Define the set-function $\mu^*: 2^X \to [0, \infty]$ by $\mu^*(\emptyset) = 0$ and for $E \subseteq X, E \neq \emptyset$,*

$$\mu^*(E) = \inf \sum_{k=1}^{\infty} \mu(E_k), \tag{5}$$

where the infimum is taken over all collections $\{E_k\}_{k=1}^{\infty}$ of sets in $\mathcal{S}$ that cover E[2]. Then $\mu^*: 2^X \to [0, \infty]$ is an outer-measure, which is called the outer-measure induced by μ.

[2] If a subset E of X cannot be covered by a countable subcollection of $\mathcal{S}$, then, by definition, it has outer-measure equal to ∞.

Proof To verify countable monotonicity, let $\{E_k\}_{k=1}^\infty$ be a collection of subsets of X that covers a set E. If $\mu^*(E_k)=\infty$ for some k, then $\mu^*(E) \le \sum\limits_{k=1}^\infty \mu^*(E_k)=\infty$. Therefore, it may be assumed that each $\mu^*(E_k)<\infty$. Let $\epsilon>0$. For each k, there is a countable collection $\{E_{i,\,k}\}_{i=1}^\infty$ of sets in $\mathcal{S}$ that covers E_k and

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

Then $\{E_{i,\,k}\}_{1\le k,i<\infty}$ is a countable collection of sets in $\mathcal{S}$ that covers $\bigcup_{k=1}^\infty E_k$ and so also covers E. Consequently,

$$\begin{aligned}\mu^*(E) \le \sum_{1\le k,i<\infty} \mu(E_{i,\,k}) &= \sum_{k=1}^\infty \left[\textstyle\sum_{i=1}^\infty \mu(E_{i,\,k})\right]\\ &\le \sum_{k=1}^\infty \mu^*(E_k) + \sum_{k=1}^\infty \epsilon/2^k\\ &= \sum_{k=1}^\infty \mu^*(E_k) + \epsilon.\end{aligned}$$

Since this holds for all $\epsilon>0$, it also holds for $\epsilon=0$. $\square$

Definition *Let the set-function $\mu\colon \mathcal{S}\to[0,\,\infty]$ be defined on a non-empty collection $\mathcal{S}$ of subsets of X and μ^* be the outer-measure induced by μ. The measure $\bar{\mu}$ that is the restriction of μ^* to the σ-algebra $\overline{\mathcal{M}}$ of μ^*-measurable sets is called the* **Carathéodory measure** *induced by μ, and $(X, \overline{\mathcal{M}}, \bar{\mu})$ is called the Carathéodory measure space induced by μ.*

$$\begin{tikzcd}[column sep=small] & \begin{array}{c}\mu^*\colon 2^X \to [0,\,\infty]\\ \textit{(the induced outer-measure)}\end{array} \arrow[dr] & \\ \begin{array}{c}\mu\colon \mathcal{S} \to [0,\,\infty]\\ \textit{(a general set-function)}\end{array} \arrow[ur] & & \begin{array}{c}\bar{\mu}\colon \overline{\mathcal{M}} \to [0,\,\infty]\\ \textit{(the induced Carathéodory measure)}\end{array}\end{tikzcd}$$

The Carathéodory Construction

Definition *Let $\mathcal{S}$ be a collection of subsets of X. By $\mathcal{S}_\sigma$ we denote the collection of sets that are countable unions of sets of $\mathcal{S}$ and by $\mathcal{S}_{\sigma\delta}$ the collection of sets that are countable intersections of sets in $\mathcal{S}_\sigma$.*

Observe that if $\mathcal{S}$ is the collection of open intervals of real numbers, then $\mathcal{S}_\sigma$ is the collection of open subsets of $\mathbf{R}$ and $\mathcal{S}_{\sigma\delta}$ is the collection of G_δ subsets of $\mathbf{R}$. We proved in Chapter 2 that for a set $E\subseteq\mathbf{R}$, E is Lebesgue measurable if and only if E is a G_δ set from which a set of Lebesgue measure zero has been excised. The proof used the σ-finiteness of Lebesgue measure on $\mathbf{R}$. We have the following variation of this outer-approximation property for any Carathéodory measure space induced by a general set-function $\mu\colon \mathcal{S} \to [0, \infty]$, provided that $\bar{\mu}$ is an extension of μ, that is $\mathcal{S} \subseteq \overline{\mathcal{M}}$ and $\mu = \bar{\mu}$ on $\mathcal{S}$.

Lemma 5 *Let the set-function $\mu\colon \mathcal{S} \to [0, \infty]$ be defined on a non-empty collection $\mathcal{S}$ of subsets of X and $\mu^*\colon 2^X \to [0, \infty]$ be the outer-measure measure induced by μ. If $E \subseteq X$ and $\mu^*(E) < \infty$, then there is a subset A of X for which*

$$A \in \mathcal{S}_{\sigma\delta},\ E \subseteq A \text{ and } \mu^*(E) = \mu^*(A). \tag{6}$$

Moreover, if the Carathéodory measure $\bar{\mu}$ induced by μ is an extension of μ and $E \in \overline{\mathcal{M}}$ then, in addition,

$$\bar{\mu}(A \sim E) = 0. \tag{7}$$

Proof Let $\epsilon > 0$. Since $\mu^*(E) < \infty$, there is a cover of E by a collection $\{E_k\}_{k=1}^{\infty}$ of sets in $\mathcal{S}$ for which

$$\sum_{k=1}^{\infty} \mu(E_k) < \mu^*(E) + \epsilon.$$

Define $A_\epsilon = \bigcup_{k=1}^{\infty} E_k$, so that by the countable monotonicity of μ^*

$$A_\epsilon \in \mathcal{S}_\sigma, E \subseteq A_\epsilon \text{ and } \mu^*(A_\epsilon) < \mu^*(E) + \epsilon.$$

Define $A = \bigcap_{k=1}^{\infty} A_{1/k}$. By the monotonicity of μ^*, for all k,

$$\mu^*(A) \leq \mu^*(A_{1/k}) \leq \mu^*(E) + 1/k,$$

and so

$$A \in \mathcal{S}_{\sigma\delta},\ E \subseteq A \text{ and } \mu^*(A) \leq \mu^*(E).$$

However, $E \subseteq A$, so that $\mu^*(E) \leq \mu^*(A)$. Therefore, (6) holds for this choice A. Now, assume that $\bar{\mu}$ is an extension of μ and $E \in \overline{\mathcal{M}}$. Since $\mathcal{S}$ is contained in the σ-algebra $\overline{\mathcal{M}}$, so is $S_{\sigma\delta}$. Both E and A belong to $\overline{\mathcal{M}}$, and so $\mu^*(A) = \bar{\mu}(A)$ and $\mu^*(E) = \bar{\mu}(E)$. Since $\bar{\mu}(A) = \bar{\mu}(E) < \infty$, by the excision property of $\bar{\mu}$, $\bar{\mu}(A \sim E) = 0$. □

Since $E = A \sim (A \sim E)$, if (7) holds for $E \in \overline{\mathcal{M}}$, then E is a set in $S_{\sigma\delta}$ from which a set of $\bar{\mu}$-measure zero has been excised.

PROBLEMS

8. Show that in the construction of Lebesgue measure on $\mathbf{R}$, the outer-measure is unchanged if covers by open, bounded intervals are replaced by covers by general, bounded intervals.
9. Let $\mu^*\colon 2^X \to [0, \infty]$ be an outer-measure. Let $A \subseteq X$, $\{E_k\}_{k=1}^{\infty}$ be a countable, disjoint collection of measurable sets and $E = \bigcup_{k=1}^{\infty} E_k$. Show that
$$\mu^*(A \cap E) = \sum_{k=1}^{\infty} \mu^*(A \cap E_k).$$
10. Show that any measure that is induced by an outer-measure is complete.
11. Define $\mu\colon 2^{\mathbf{R}} \to \mathbf{R}$ by defining $\mu(A)$ to be the number of integers in A. Determine the outer-measure μ^* induced by μ and the σ-algebra of measurable sets.

9.3 THE CARATHÉODORY-HAHN THEOREM

Lebesgue measure on $\mathbf{R}$ was constructed as a Carathéodory extension. In view of this, the following question occurs: what properties must a set-function $\mu\colon \mathcal{S} \to [0, \infty]$ possess in order that the induced Carathéodory measure be an extension of μ? The following is a simple example in which there are sets in $\mathcal{S}$ that are not μ^*-measurable.

Example Let $X = \{1, 2, 3\}$ and define $\mu^*\colon 2^X \to [0, \infty)$ by $\mu^*(\emptyset) = 0, \mu^*(X) = 2$, and for every proper, non-empty subset E of X, $\mu^*(E) = 1$. We leave it as an exercise to show that μ^* is an outer-measure and that the only sets that are measurable with respect to μ^* are $\emptyset$ and X.

Now, assume that all the sets in 2^X are μ^*-measurable. Since $\bar{\mu} = \mu^*$ on $\mathcal{S}$ and E is covered by itself,

$$\bar{\mu}(E) \leq \mu(E) \text{ for all } E \in \mathcal{S}.$$

The following is an example in which there is a set E in $\mathcal{S}$ for which $\bar{\mu}(E) < \mu(E)$.

Example Let $X = \mathbf{N}$ and $\mathcal{S}$ be the collection of singleton sets $\{n\}$ together with the whole space $\mathbf{N}$. For each n, define $\mu\{n\} = 1/2^n$ and define $\mu(\mathbf{N}) = 2$. Since $\mathbf{N} \subseteq \bigcup_{n=1}^{\infty}\{n\}$,

$$\mu^*(E) = \sum_{n\in E} 1/2^n \text{ if } E \text{ is finite, and } \mu^*(\mathbf{N}) = \sum_{n=1}^{\infty} 1/2^n = 1.$$

Clearly, every subset of $\mathbf{N}$ is μ^*-measurable, so $\mu^* = \bar{\mu}$, but $\bar{\mu}(\mathbf{N}) < \mu(\mathbf{N})$.

Definition *A set-function $\mu\colon \mathcal{S} \to [0, \infty]$ on a collection $\mathcal{S}$ of subsets of X that contains the empty-set is said to be a* **premeasure** *provided that it is finitely additive, countably monotone and $\mu(\emptyset) = 0$.*

Since a measure is both finitely additive and countably monotone, if a set-function is extendable to a measure, it must possess these same two properties. Being a premeasure is a necessary but not sufficient condition for the Carathéodory measure induced by μ to extend μ. However, if finer set-theoretic structure is imposed on $\mathcal{S}$, this necessary condition is also sufficient. We impose this structure in two steps, the first being the following.

Theorem 6 *Let $\mu\colon \mathcal{S} \to [0, \infty]$ be a premeasure on a collection $\mathcal{S}$ of subsets of X for which*

$$\textit{if } A \textit{ and } B \textit{ belong to } \mathcal{S}, \textit{then so does } A \sim B.$$

Then the Carathéodory measure $\bar{\mu}\colon \overline{\mathcal{M}} \to [0, \infty]$ induced by μ is an extension of μ, called the **Carathéodory extension** *of μ.*

Proof Let $E \in \mathcal{S}$. To show that E is measurable with respect to the outer-measure induced by μ, it suffices to let $A \subseteq X$, with $\mu^*(A) < \infty$, let $\epsilon > 0$ and verify that

$$\mu^*(A) + \epsilon \geq \mu^*(A \cap E) + \mu^*(A \sim E). \tag{8}$$

Since $\mu^*(A) < \infty$, there is a collection $\{E_k\}_{k=1}^{\infty}$ of sets in $\mathcal{S}$ that covers A and

$$\sum_{k=1}^{\infty} \mu(E_k) \leq \mu^*(A) + \epsilon. \tag{9}$$

However, for each k, since $\mathcal{S}$ is closed with respect to relative complements, $E_k \sim E$ belongs to $\mathcal{S}$ and so does $E_k \cap E = E_k \sim [E_k \sim E]$. A premeasure is finitely additive. Therefore,

$$\mu(E_k) = \mu(E_k \cap E) + \mu(E_k \sim E).$$

Sum these equalities to conclude that

$$\sum_{k=1}^{\infty} \mu(E_k) = \sum_{k=1}^{\infty} \mu(E_k \cap E) + \sum_{k=1}^{\infty} \mu(E_k \sim E). \tag{10}$$

Observe that $\{E_k \cap E\}_{k=1}^{\infty}$ and $\{E_k \sim E\}_{k=1}^{\infty}$ are countable collections of sets in $\mathcal{S}$ that cover $A \cap E$ and $A \sim E$, respectively. Consequently,

$$\sum_{k=1}^{\infty} \mu(E_k \cap E) \geq \mu^*(A \cap E) \text{ and } \sum_{k=1}^{\infty} \mu(E_k \sim E) \geq \mu^*(A \sim E).$$

The desired inequality (8) follows from the these two inequalities together with (9) and (10). Clearly $\mu(E) = \mu^*(E)$ for each set $E \in \mathcal{S}$ if and only if μ is countable monotone, which, by assumption, it is. Consequently, $\mu(E) = \mu^*(E) = \bar{\mu}(E)$, since the outer-measure $\mu^*(E)$ extends the measure $\bar{\mu}$. □

Observe the distinct roles played in the proof of this theorem by the two properties of a premeasure. Finite additivity implied that every set in $\mathcal{S}$ is μ^*-measurable. Countable monotonicity is equivalent to the equality $\mu(E) = \mu^*(E)$ for each $E \in \mathcal{S}$. The assumption that the domain of a premeasure is closed with respect to relative complements is too restrictive: few collections of sets possess this property. Certainly, the collection of bounded intervals of real numbers does not possess this property. However, the collection of unions of finite, disjoint collections of bounded intervals does possess it. We now take the second step in selecting the properties of the domain of a premeasure which ensure that it is extended by its induced Carathéodory measure.

Definition *A collection $\mathcal{S}$ of subsets of X is called a* **semi-ring** *provided that if A and B belong to $\mathcal{S}$, then so does $A \cap B$, and if $A \subseteq B$, then there is a finite, disjoint collection $\{C_k\}_{k=1}^{n}$ of sets in $\mathcal{S}$ for which*

$$B \sim A = \bigcup_{k=1}^{n} C_k.$$

Verification that the following collections of sets are semi-rings is left as an exercise.

Example

(i) The collection $\mathcal{S}$ of all intervals of real numbers is a semi-ring.

(ii) For $a < b$, consider the collection $\mathcal{S}$ of subsets of $[a, b]$ comprising the singleton set $\{b\}$ and all sets of the form $[c, d)$ for $a \leq c \leq d \leq b$. Then $\mathcal{S}$ is a semi-ring.

(iii) The collection $\mathcal{S}$ of all subsets of the plane $\mathbf{R}^2$ of the form $I \times J$, where I and J are bounded intervals of real numbers is a semi-ring.

The following lemma bundles all the set-theoretic properties of a semi-ring that we will employ. The collections $\mathcal{S}_\sigma$ and $\mathcal{S}_{\sigma\delta}$ were defined in the preceding section.

Lemma 7 *Let $\mathcal{S}$ be a semi-ring of subsets of X. Then every set in $\mathcal{S}_\sigma$ is the disjoint, countable union of sets in $\mathcal{S}$ and every set in $\mathcal{S}_{\sigma\delta}$ is the intersection of a descending, countable collection of sets in $\mathcal{S}_\sigma$. Moreover, if $\mathcal{S}'$ is the collection of unions of finite, disjoint collections of sets in $\mathcal{S}$, then*

$$\text{if } A \text{ and } B \text{ belong to } \mathcal{S}', \text{ so does } A \sim B. \tag{11}$$

Proof First, consider the special case that $\mathcal{S}$ is an algebra, that is, if A and B belong to $\mathcal{S}$, then so do $A \sim B$, $A \cap B$ and $A \cup B$. Let $E = \bigcup_{k=1}^\infty E_k$, the union of a countable collection of sets in $\mathcal{S}$. Define $E'_1 = E_1$ and $E'_k = E_k \sim \bigcup_{i=1}^{k-1} E_i$ if $k \geq 2$. Then the collection $\{E'_k\}_{k=1}^\infty$, is disjoint and also, $\mathcal{S}$ being an algebra, each E'_k belong to $\mathcal{S}$. Furthermore, $\bigcup_{k=1}^\infty E_k = \bigcup_{k=1}^\infty E'_k$ because, for each $x \in E$, there is a first index k such that $x \in E_k$, and therefore $x \in E'_k$. Consequently, $E = \bigcup_{k=1}^\infty E'_k$, a countable, disjoint union of sets in $\mathcal{S}$. To prove the descending intersection property for $\mathcal{S}_{\sigma\delta}$, we first show that $\mathcal{S}_\sigma$ is closed with respect to intersection. Indeed, if $B = \bigcup_{j=1}^\infty B_j$ and $C = \bigcup_{k=1}^\infty C_k$, each a collection of sets in $\mathcal{S}$, then

$$B \cap C = \bigcup_{1 \leq j, k < \infty} (B_j \cap C_k).$$

Therefore, since $\mathcal{S}$ is an algebra, $B \cap C$ belongs to $\mathcal{S}_\sigma$. To prove the descending intersection property of $\mathcal{S}_{\sigma\delta}$, observe that if $D = \bigcap_{k=1}^\infty D_k$, a general intersection of sets in $\mathcal{S}_\sigma$, then we also have $D = \bigcap_{k=1}^\infty D'_k$, where each D'_k is the intersection of $\{D_k\}_{i=1}^k$, so, D is the intersection of a descending collection in $\mathcal{S}_\sigma$.

Having proved the result if $\mathcal{S}$ is an algebra, to prove the general result it suffices to show that a semi-ring is an algebra, since a countable, disjoint union of disjoint unions is also a disjoint union. Let $A = \bigcup_{k=1}^n A_k$ and $B = \bigcup_{j=1}^m B_j$, each a disjoint union of sets in $\mathcal{S}'$. Observe that

$$A \cap B = \bigcup_{1 \leq i \leq n, 1 \leq j \leq m} A_i \cap B_j \text{ and } A \sim B = \bigcup_{k=1}^n \left[\bigcap_{j=1}^m (A_k \sim B_j) \right].$$

Consequently, $\mathcal{S}$ being a semi-ring, $A \cap B$ and $A \sim B$ belong to $\mathcal{S}'$ and, since $A \cup B = A \cup (B \sim A)$, so does $A \cup B$. □

Lemma 8 *Let $\mathcal{S}$ be a semi-ring of subsets of X, and $\mathcal{S}'$ be the collection of unions of finite, disjoint collections of sets in $\mathcal{S}$. Then a premeasure on $\mathcal{S}$ has a unique extension to a premeasure μ' on $\mathcal{S}'$.*

Proof Let $\mu\colon \mathcal{S} \to [0, \infty]$ be a premeasure on $\mathcal{S}$. For $E = \{A_k\}_{k=1}^n$, the disjoint union of sets in $\mathcal{S}$, define $\mu'(E) = \sum_{k=1}^n \mu(A_k)$. To verify that $\mu'(E)$ is properly defined, let E also be the disjoint union of the finite collection $\{B_j\}_{j=1}^m$ of sets in $\mathcal{S}$. It is necessary to show that

$$\sum_{j=1}^m \mu(B_j) = \sum_{k=1}^n \mu(A_k).$$

However, by the finite additivity of the premeasure μ on $\mathcal{S}$,

$$\sum_{j=1}^m \mu(B_j) = \sum_{j=1}^m \left[\sum_{k=1}^n \mu(B_j \cap A_k)\right] = \sum_{k=1}^n \left[\sum_{j=1}^m \mu(B_j \cap A_k)\right] = \sum_{k=1}^n \mu(A_k).$$

It remains to show that μ' is a premeasure on $\mathcal{S}'$. Since μ' is properly defined, it inherits finite additivity from the finite additivity possessed by μ. To establish the countable monotonicity of μ', let $E \in \mathcal{S}'$ be covered by a collection $\{E_k\}_{k=1}^\infty$ of sets in $\mathcal{S}'$. By the monotonicity of μ' and the preceding lemma, we may assume that the collection $\{E_k\}_{k=1}^\infty$ is disjoint. Let $E = \bigcup_{j=1}^m A_j$, a disjoint union of sets in $\mathcal{S}$. For each j, A_j is covered by $\bigcup_{k=1}^\infty (A_j \cap E_k)$, and so, by the countable monotonicity of μ, $\mu(A_j) \le \sum_{k=1}^\infty \mu(A_j \cap E_k)$. For each k, since $\{E_k\}_{k=1}^m$ is disjoint, $\mu(E \cap E_k) = \sum_{j=1}^m \mu(A_j \cap E_k)$, and so, by the monotonicity of μ',

$$\begin{aligned}
\mu'(E) = \sum_{j=1}^m \mu(A_j) \quad &\le \sum_{j=1}^m \left[\sum_{k=1}^\infty \mu(A_j \cap E_k)\right] \\
&= \sum_{k=1}^\infty \left[\sum_{j=1}^m \mu(A_j \cap E_k)\right] \\
&= \sum_{k=1}^\infty \mu(E \cap E_k) \\
&\le \sum_{k=1}^\infty \mu'(E_k).
\end{aligned}$$

Consequently, μ' is countably monotone. □

The Carathéodory-Hahn Theorem *If $\mu\colon \mathcal{S} \to [0, \infty]$ is a premeasure on a semi-ring $\mathcal{S}$ of subsets of X, then the Carathéodory measure $\bar{\mu}\colon \overline{\mathcal{M}} \to [0, \infty]$ induced by μ is an extension of μ, which is complete and every set of μ^* outer-measure zero is measurable. Moreover, if μ is σ-finite, then $\bar{\mu}$ is the only measure on $\overline{\mathcal{M}}$ that extends μ, that is, for any measure $\nu\colon \overline{\mathcal{M}} \to [0, \infty]$,*

$$\text{if } \nu = \mu \textit{ on } \mathcal{S}, \textit{ then } \nu = \bar{\mu} \textit{ on } \overline{\mathcal{M}}. \tag{12}$$

Proof By the preceding lemma, μ is uniquely extendable to a premeasure $\mu'\colon \mathcal{S}' \to [0, \infty]$, where $\mathcal{S}'$ is the collection of finite, disjoint unions of sets in $\mathcal{S}$. According to Lemma 7, the collection $\mathcal{S}'$ is closed with respect to complements. Consequently, by an appeal to Theorem 6, the Carathéodory measure induced by $\mu'\colon S' \to [0, \infty]$ is an extension of μ'. However, clearly $\mathcal{S}_\sigma = \mathcal{S}'_\sigma$, and so the Carathéodory measure induced by μ is the same as that induced by μ'. Therefore, the Carathéodory measure induced by μ is an extension of μ. To prove uniqueness, assume that μ is σ-finite, in which case its extension $\bar{\mu}$ also is σ-finite. Let $\nu\colon \overline{\mathcal{M}} \to [0, \infty]$ be a measure that extends μ. It suffices to let $E \in \overline{\mathcal{M}}$ with $\bar{\mu}(E) < \infty$, and show that $\bar{\mu}(E) = \nu(E)$. First, observe that since this extension ν is countably monotone,

$$\nu(E) \leq \mu^*(E) = \bar{\mu}(E) \text{ if } E \in \overline{\mathcal{M}} \text{ and } \bar{\mu}(E) < \infty. \tag{13}$$

Therefore, $\nu(E) < \infty$. According to Lemma 5, there is a set A for which

$$A \in \mathcal{S}_{\sigma\delta},\ E \subseteq A \text{ and } \bar{\mu}(A \sim E) = 0.$$

It follows from (13) that

$$\nu(A \sim E) = \bar{\mu}(A \sim E) = 0.$$

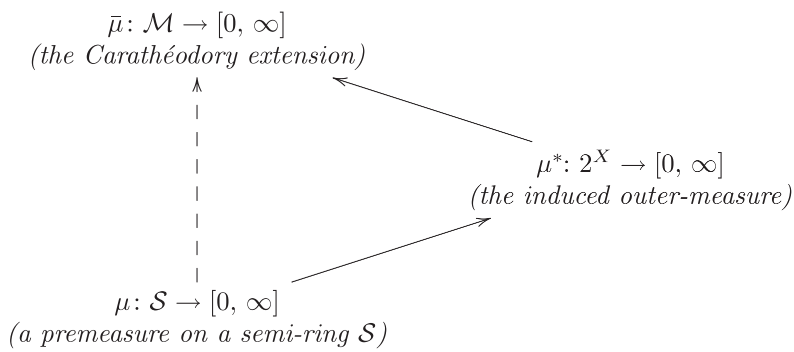
By Lemma 7, each set in $\mathcal{S}_\sigma$ is the union of a countable, disjoint collection of sets in $\mathcal{S}$ and so $\nu = \bar{\mu}$ on $\mathcal{S}_\sigma$. Also, the set A is the intersection of a descending, countable collection of sets in $\mathcal{S}_\sigma$. Consequently, since $\nu(E) < \infty$ and $\bar{\mu}(E) < \infty$, by the continuity of the measures ν and $\bar{\mu}$,

$$\nu(A) = \bar{\mu}(A).$$

By the excision property of these measures,

$$\nu(E) = \nu(A) - \nu(A \sim E) = \bar{\mu}(A) - \bar{\mu}(A \sim E) = \bar{\mu}(E).$$

□



The Carathéodory Construction Extends a Premeasure on a Semi-ring to a Measure

PROBLEMS

12. Let $S \subseteq 2^X$ and $\mu^*\colon 2^X \to [0, \infty]$ be the outer-measure induced by $\mu\colon S \to [0, \infty]$. Show that for a set $A \subseteq X$ to be measurable with respect to μ^* it suffices to verify that

$$\mu^*(E) = \mu^*(A \cap E) + \mu^*(A \sim E) \text{ for all } E \in \mathcal{S},$$

not for all of the subsets E of X. This was the definition of measurable set used by Lebesgue for the set-function length defined on the collection of open, bounded intervals in $\mathbf{R}$.

13. Show that a non-negative set-function on a σ-algebra is a measure if and only if it is a premeasure.
14. Let $\mathbf{Q}$ be the set of rational numbers and $\mathcal{S}$ the collection of all finite unions of intervals of the form $(a, b] \cap \mathbf{Q}$, where $a, b \in \mathbf{Q}$ and $a \leq b$. Define $\mu(a, b] = \infty$ if $a < b$ and $\mu(\emptyset) = 0$. Show that $\mathcal{S}$ is closed with respect to relative complements and $\mu \colon \mathcal{S} \to [0, \infty]$ is a premeasure. Then show that the extension of μ to the smallest σ-algebra containing $\mathcal{S}$ is not unique.
15. Verify the assertions made in this section's first and second example.
16. Show that the three collections of sets in this section's third example are semi-rings.
17. Let $\mathcal{S}$ be a σ-algebra of subsets of X and $\mu \colon \mathcal{S} \to [0, \infty]$ be a measure. Let $\bar{\mu} \colon \overline{\mathcal{M}} \to [0, \infty]$ be the Carathéodory measure induced by μ. Show that $\mathcal{S}$ is a subcollection of $\overline{\mathcal{M}}$ and it may be a proper subcollection.
18. Show that Lebesgue on $\mathbf{R}$ is the Carathéodory extension of the premeasure Lebesgue measure on the semi-ring of Borel sets.

CHAPTER 10

Particular Measures

Contents

10.1 Lebesgue Measure on Euclidean Space 174
10.2 Lebesgue Measurability and Measure of Images of Mappings 182
10.3 Borel Measures on $\mathbf{R}^n$ and Regularity 190
10.4 Carathéodory Outer-Measures and Hausdorff Measures 193
10.5 Signed Measures: The Hahn and Jordan Decompositions 197

We first consider Lebesgue measure on $\mathbf{R}^n$, which is constructed as the Carathéodory extension of the set-function volume defined on the collection of n-fold products of bounded intervals of real numbers. We prove regularity and show that the collection of Lebesgue measurable sets is the smallest σ-algebra that contains the Borel σ-algebra and all set of Lebesgue outer-measure zero. We then consider mappings $\mathcal{N}\colon \mathbf{R}^n \to \mathbf{R}^n$ which preserve Lebesgue measurability, show that both Lipschitz and continuously differentiable mappings defined on open sets do so, and show that for a linear operator, multiplication by the absolute value of its determinant of the measure of a set gives the Lebesgue measure of the its image. The proof of this determinant property is based on the Vitali Partition Theorem, according to which an open subset of $\mathbf{R}^n$ is, after the excision of a set of Lebesgue measure zero, the countable, disjoint union of open balls. We then prove that any finite measure on the Borel subsets of $\mathbf{R}^n$ is regular. The chapter concludes with the Hahn Decomposition and Jordan Decompositions Theorems, which provide expressions of a signed measure as the difference of measures.

10.1 LEBESGUE MEASURE ON EUCLIDEAN SPACE

For a natural number n, $\mathbf{R}^n$ denotes the collection of ordered n-tuples of real numbers $x = (x_1, \ldots, x_n)$. We denote points in $\mathbf{R}^n$ by x and y. Of course, $\mathbf{R}^n$ is a linear space, and we define the mapping $\langle \cdot, \cdot \rangle \colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ by

$$\langle x, y \rangle = \sum_{k=1}^{n} x_k \cdot y_k \text{ for all } x, y.$$

This mapping is **symmetric** in x and y, is **bilinear**, meaning that if one of the variables is fixed, it is linear in the other, and is **positive-definite**, meaning that $\langle x, x \rangle$ is non-negative and only vanishes at $x = 0$. A mapping with these three properties is called an inner-product on $\mathbf{R}^n$. We use just this one and call it the **(Euclidean) inner-product.** We define the mapping $\|\cdot\|\colon \mathbf{R}^n \to \mathbf{R}^n$ by

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{k=1}^{n} x_k^2} \text{ for all } x,$$

and call it the **(Euclidean) norm.**

The Cauchy-Schwarz Inequality in $\mathbf{R}^n$.

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \text{ for all } x, y.$$

To verify this inequality, fix x and y, and observe that, by bilinearity of the inner-product,

$$p(t) = \|tx + y\|^2 = \|x\|^2 \cdot t^2 + 2\langle x, y\rangle \cdot t + \|y\|^2 \geq 0 \text{ for all } t \in \mathbf{R}.$$

Then $p(t)$ is a quadratic polynomial which fails to have distinct real roots, so that its discriminant is not positive, that is, the Cauchy-Schwarz Inequality holds. By this inequality and bilinearity, for all x, y,

$$\|x + y\|^2 = \langle x + y, x + y\rangle \leq [\|x\| + \|y\|]^2.$$

Therefore, the triangle inequality holds for $\|\cdot\|$, so it indeed is a norm. The linear space $\mathbf{R}^n$, considered with this inner-product and norm, is called ***n*-dimensional Euclidean space**[1].

For $x_0 \in \mathbf{R}^n$ and $r > 0$, we call $\{x \in \mathbf{R}^n \mid \|x - x_0\| < r\}$ the **open ball** centered at x_0 of radius r. We call an open set that contains a subset E of $\mathbf{R}^n$ a **neighborhood** of E. For $E \subseteq \mathbf{R}^n$, define the distance function $\operatorname{dist}_E \colon \mathbf{R}^n \to \mathbf{R}$ by

$$\operatorname{dist}_E(x) = \inf\{\|x - x'\| \mid x' \in E\} \quad \text{for all } x.$$

Then

$$|\operatorname{dist}_E(x) - \operatorname{dist}_E(y)| \leq \|x - y\| \quad \text{for all } x, y.$$

Indeed, to verify this, let $x, y \in \mathbf{R}^n$. For $x' \in E$, we have, by the triangle inequality,

$$\|x - x'\| \leq \|x - y\| + \|y - x'\|.$$

Take the infimum over $x' \in E$, to obtain the inequality $\operatorname{dist}_E(x) \leq \|x - y\| + \operatorname{dist}_E(y)$, and then interchange x and y to get the stated Lipschitz inequality. Since the distance function is continuous, for each $\delta > 0$, $\mathcal{N}_\delta(E) = \{x \in \mathbf{R}^n \mid \operatorname{dist}_E(x) < \delta\}$ is a neighborhood of E, which will play the role for a set that is played by an open ball for a point. Observe that if E is closed, then $\operatorname{dist}_E(x) = 0$ if and only if $x \in E$. A continuous function on a subset of $\mathbf{R}^n$ is said to be **compactly supported** provided that it vanishes on the complement of a compact set.

Urysohn's Lemma in $\mathbf{R}^n$. *If A and B are closed, disjoint subsets of $\mathbf{R}^n$, then there is a continuous function $f \colon \mathbf{R}^n \to \mathbf{R}$ for which*

$$f = 1 \text{ on } A, f = 0 \text{ on } B \text{ and } 0 \leq f \leq 1 \text{ on } \mathbf{R}^n. \tag{1}$$

Moreover, if A is compact then, in addition, the function f is compactly supported.

---

[1] Here, we use results always seen in advanced calculus courses: the elementary properties of open and closed subsets of $\mathbf{R}^n$, and three theorems: the Heine-Borel Theorem (A subset K of $\mathbf{R}^n$ is closed and bounded if and only if it is compact, in the sense that every cover of K by a collection of open sets has a finite subcover) and the equivalent Bolzano-Weierstrass Theorem (A subset K of $\mathbf{R}^n$ is closed and bounded if and only if it is sequentially compact, in the sense that every sequence in K has a subsequence that converges to a point in K) and the Mean Value Theorem for continuously differentiable functions.

Proof Distance functions are Lipschitz, and therefore are continuous. Since A and B are closed and disjoint, the continuous function $\text{dist}_A + \text{dist}_B$ takes positive values. Therefore, (1) holds for

$$f = \frac{\text{dist}_B}{\text{dist}_A + \text{dist}_B}. \tag{2}$$

Now assume that A is compact. Since B is closed and disjoint from A, there is a $\delta > 0$ for which $B \subseteq \mathbf{R}^n \sim \mathcal{N}_\delta(A)$. Replace B in (2) by the closed set $B' = \mathbf{R}^n \sim \mathcal{N}_\delta(A)$, and then the function f has compact support, since it vanishes on the complement of the bounded set $\mathcal{N}_\delta(A)$. $\square$

The above lemma is needed in the proof we provide, in Chapter 12, that the space of compactly supported, smooth functions are, for $1 \leq p < \infty$, dense in $L^p(\mathbf{R}^n)$.

Definition *A subset I of $\mathbf{R}^n$ that is the n-fold Cartesian product of bounded intervals of real numbers is called an* **interval**. *If $I = I_1 \times I_2 \times \cdots \times I_n$, the* **volume** *of I, $\text{vol}(I)$, is defined by*

$$\text{vol}(I) = \ell(I_1) \cdot \ell(I_2) \cdot \cdots \cdot \ell(I_n).$$

A point in $\mathbf{R}^n$ is called an **integral point** provided that each of its coordinates is an integer, and for an interval I in $\mathbf{R}^n$, the **integral count** of I, $\mu^{int}(I)$, is defined to be the number of integral points in I.

Lemma 1 *For each $\epsilon > 0$, define the ϵ-dilation $T_\epsilon \colon \mathbf{R}^n \to \mathbf{R}^n$ by $T_\epsilon(x) = \epsilon \cdot x$. Then for each bounded interval I in $\mathbf{R}^n$,*

$$\lim_{\epsilon \to \infty} \frac{\mu^{int}(T_\epsilon(I))}{\epsilon^n} = \text{vol}(I). \tag{3}$$

Proof For an interval I in $\mathbf{R}$ with end-points a and b, there is the following estimate for its integral count:

$$(b-a) - 1 \leq \mu^{int}(I) \leq (b-a) + 1.$$

Therefore, for the interval $I = I_1 \times I_2 \times \cdots \times I_n$ in $\mathbf{R}^n$, since

$$\mu^{int}(I) = \mu^{int}(I_1) \cdots \mu^{int}(I_n),$$

if each I_k has end-points a_k and b_k, there is the estimate

$$[(b_1 - a_1) - 1] \cdot \cdots \cdot [(b_n - a_n) - 1] \leq \mu^{int}(I) \leq [(b_1 - a_1) + 1] \cdot \cdots \cdot [(b_n - a_n) + 1].$$

For $\epsilon > 0$ replace the interval I by the dilated interval $T_\epsilon(I)$ and obtain the estimate

$$[\epsilon \cdot (b_1 - a_1) - 1] \cdot \cdots \cdot [\epsilon \cdot (b_n - a_n) - 1] \leq \mu^{int}(T_\epsilon(I)) \leq [\epsilon \cdot (b_1 - a_1) + 1] \cdot \cdots \cdot [\epsilon \cdot (b_n - a_n) + 1].$$

Divide this inequality by ϵ^n and take the limit as $\epsilon \to \infty$ to obtain (3). $\square$

We employ this lemma to provide a quite direct proof of the following fundamental theorem.

Theorem 2 *The set-function volume,* vol: $\mathcal{I}_n \to [0, \infty)$, *is a premeasure on the semi-ring* $\mathcal{I}_n$ *of intervals in* $\mathbf{R}^n$.

Proof (**von Neumann**) The proof that $\mathcal{I}_n$ is a semi-ring is left as an exercise in induction. We begin by showing that volume is finitely additive. Let $\{I^k\}_{k=1}^m$ be a partition of the interval I by intervals. Then, for each $\epsilon > 0$, $\{T_\epsilon(I^k)\}_{k=1}^m$ is a partition by intervals of the interval $T_\epsilon(I)$. It is clear that the integral count μ^{int} is finitely additive. Therefore,

$$\mu^{int}(T_\epsilon(I)) = \sum_{k=1}^m \mu^{int}(T_\epsilon(I^k)) \text{ for all } \epsilon > 0.$$

Divide each side by ϵ^n and take the limit as $\epsilon \to \infty$ to obtain, by (3),

$$\text{vol}(I) = \sum_{k=1}^m \text{vol}(I^k),$$

so that volume is finitely additive. To establish countable monotonicity, let I be an interval that is covered by the countable collection of intervals $\{I^k\}_{k=1}^\infty$. First, consider the case that I is closed and each I^k is open. By the Heine-Borel Theorem, there is an index m for which I is covered by the finite subcollection $\{I^k\}_{k=1}^m$. Clearly, the integral count is finitely monotone, and so

$$\mu^{int}(T_\epsilon(I)) \leq \sum_{k=1}^m \mu^{int}(T_\epsilon(I^k)) \text{ for all } \epsilon > 0.$$

Divide each side by ϵ^n and take the limit as $\epsilon \to \infty$ to obtain, by (3),

$$\text{vol}(I) \leq \sum_{k=1}^m \text{vol}(I^k) \leq \sum_{k=1}^\infty \text{vol}(I^k).$$

It remains to consider the case of general intervals. Let $\epsilon > 0$. Choose a closed interval $\hat{I}$ contained in I and a collection $\{\hat{I}^k\}_{k=1}^\infty$ of open intervals for which each $I^m \subseteq \hat{I}^m$ and, moreover,

$$\text{vol}(I) - \text{vol}(\hat{I}) < \epsilon \text{ and } \text{vol}(\hat{I}^m) - \text{vol}(I^m) < \epsilon/2^m \text{ for all } m.$$

By the case just considered,

$$\text{vol}(\hat{I}) \leq \sum_{k=1}^\infty \text{vol}(\hat{I}^k),$$

and so

$$\text{vol}(I) \leq \sum_{k=1}^\infty \text{vol}(I^k) + 2\epsilon.$$

Consequently, volume is a premeasure. □

In view of this theorem, by an appeal to the Carathéodory-Hahn Theorem, Lebesgue measure on $\mathbf{R}^n$ may be defined as the Carathéodory extension of volume.

Definition *The outer-measure μ_n^* induced by the premeasure volume on the semi-ring $\mathcal{I}_n$ of intervals in $\mathbf{R}^n$ is called* **Lebesgue outer-measure** *on $\mathbf{R}^n$. The collection of μ_n^*-measurable sets is denoted by $\mathcal{L}^n$, and the restriction of μ_n^* to $\mathcal{L}^n$ is denoted by μ_n. The measure space $(\mathbf{R}^n, \mathcal{L}^n, \mu_n)$ and the measure μ_n are referred to as* **Lebesgue measure** *on $\mathbf{R}^n$.*

Theorem 3 *The Lebesgue measure space $(\mathbf{R}^n, \mathcal{L}^n, \mu_n)$ is both σ-finite and complete, every interval is Lebesgue measurable, every set of Lebesgue outer-measure zero in Lebesgue measurable, and for any Lebesgue measurable set E,*

$$\mu_n(E) = \inf \sum_{k=1}^{\infty} \operatorname{vol} I_k, \tag{4}$$

where the infimum is taken over coverings $\{I_k\}_{k=1}^{\infty}$ of E by bounded intervals.

Proof According to the preceding theorem, volume is a premeasure on the semi-ring of intervals in $\mathbf{R}^n$. According to the Carathéodory-Hahn Theorem, Lebesgue measure is an extension of volume, sets of Lebesgue outer-measure zero are Lebesgue measurable, and the measure space $(\mathbf{R}^n, \mathcal{L}^n, \mu_n)$ is complete. It clearly is σ-finite. Since μ_n is the restriction of μ_n^*, we have (4). $\square$

Every open subset $\mathcal{O}$ of $\mathbf{R}^n$ is Lebesgue measurable. Indeed, let $\mathcal{F}$ be the collection of all open, bounded intervals that are contained in $\mathcal{O}$ and are n-fold products of open, bounded intervals in $\mathbf{R}$ with rational end-points. Since $\mathbf{Q}$ is a dense subset of $\mathbf{R}$, $\mathcal{O}$ is the union of sets in $\mathcal{F}$, and since $\mathbf{Q}$ is countable, so is $\mathcal{F}$. Each closed set, being the complement of an open set, also is Lebesgue measurable. The intersection of a countable collection of open subsets of $\mathbf{R}^n$ is called a G_δ set, and the union of a countable collection of closed sets is called an F_σ set. Consequently, $\mathcal{L}^n$ being a σ-algebra, each G_δ set and F_σ set is Lebesgue measurable.

Theorem 4 (**The Regularity of Lebesgue Measure on Euclidean Space**) *If $E \subseteq \mathbf{R}^n$ is Lebesgue measurable and $\epsilon > 0$, then there is a closed set $F \subseteq \mathbf{R}^n$ and an open set $\mathcal{O} \subseteq \mathbf{R}^n$ for which*

$$F \subseteq E \subseteq \mathcal{O}, \;\; \mu_n(\mathcal{O} \sim E) < \epsilon \text{ and } \mu_n(E \sim F) < \epsilon.$$

If $\mu_n(E) < \infty$, then, in addition, F is compact.

Proof Since μ_n is σ-finite, to prove the open outer-approximation property it suffices to consider $E \in \mathcal{L}^n$ for which $\mu_n(E) < \infty$. For such an E, let $\epsilon > 0$. It follows from (4) that there is a countable collection $\{I^k\}_{k=1}^{\infty}$ of intervals that cover E and

$$\sum_{k=1}^{\infty} \operatorname{vol}(I^k) < \mu_n(E) + \epsilon/2.$$

Choose a collection of open intervals $\{\hat{I}^k\}_{k=1}^{\infty}$ for which

$$I^k \subseteq \hat{I}^k \text{ and } \operatorname{vol}(\hat{I}^k) - \operatorname{vol} I^k < \epsilon/2^{k+1} \text{ for all } k,$$

so that

$$E \subseteq \mathcal{O} = \bigcup_{k=1}^{\infty} \hat{I}^k \text{ and } \sum_{k=1}^{\infty} \operatorname{vol}(\hat{I}^k) < \mu_n(E) + \epsilon.$$

Then $\mathcal{O}$, being the union of open sets, is open and, by the excision property of measure,

$$\mu_n(\mathcal{O} \sim E) = \mu_n(\mathcal{O}) - \mu_n(E) \leq \sum_{k=1}^{\infty} \operatorname{vol}(\hat{I}^k) - \mu_n(E) < \epsilon.$$

In order to establish the closed inner-approximation for a set, we use the open outer-approximation property of its measurable complement. There is an open set $\mathcal{U}$ for which

$$\mathbf{R}^n \sim E \subseteq \mathcal{U} \text{ and } \mu_n(\mathcal{U} \sim (\mathbf{R}^n \sim E)) < \epsilon.$$

Define $F = \mathbf{R}^n \sim \mathcal{U}$. Then F is a closed subset of E and

$$E \sim F = E \sim (\mathbf{R}^n \sim \mathcal{U}) = E \cap U = \mathcal{U} \sim (\mathbf{R}^n \sim E).$$

Therefore,

$$\mu_n(E \sim F) = \mu_n(\mathcal{U} \sim (\mathbf{R} \sim E)) < \epsilon.$$

It remains only to establish the compact inner-approximation property. Assume that $\mu_n(E) < \infty$, and choose $\{E_k\}_{k=1}^{\infty}$ to be any measurable partition of E into bounded sets. We have

$$\mu_n(E) = \sum_{k=1}^{\infty} \mu_n(E_k) < \infty.$$

Choose an index k_0 so that $\mu_n(E) - \sum_{k=1}^{k_0} \mu_n(E_k) < \epsilon$. Then $\bigcup_{k=1}^{k_0} E_k$ is a bounded set, and so its closure K is compact. We have, by the excision property of measure,

$$\mu_n(E \sim K) = \mu_n(E) - \mu_n(K) \leq \mu_n(E) - \sum_{k=1}^{k_0} \mu_n(E_k) < \epsilon.$$

□

In the next section, we establish the regularity of every finite measure on the Borel σ-algebra $\mathcal{B}(\mathbf{R}^n)$, a measure that is not constructed as a Carathéodory extension, since it is not complete.

Theorem 5 *A subset set E of $\mathbf{R}^n$ is Lebesgue measurable if and only if it is a G_δ set from which a set of Lebesgue measure zero has been excised.*

Proof Assume that E is Lebesgue measurable. By the preceding theorem, for each index k, there is an open set $\mathcal{O}_k$ that contains E and has $\mu_n(\mathcal{O}_k \sim E) < 1/k$. Define $G = \bigcap_{k=1}^{\infty} \mathcal{O}_k$. Since $G \sim E \subseteq \mathcal{O}_k \sim E$ for each k, $\mu_n(G \sim E) = 0$. Therefore, $E = G \sim (G \sim E)$, a G_δ set from which a set of measure zero has been excised. Since $\mathcal{L}^n$ is a σ-algebra, every such set is Lebesgue measurable. □

Corollary 6 *A subset set E of $\mathbf{R}^n$ is Lebesgue measurable if and only if it is the disjoint union of an F_σ set and a set of Lebesgue measure zero.*

Proof If E is Lebesgue measurable, then so is its complement and, by this theorem, $\mathbf{R}^n \sim E = G \sim E_0$, where E_0 is a subset of a G_δ set G for which $\mu_n(E_0) = 0$. Therefore, $E = (\mathbf{R}^n \sim G) \cup (E_0)$, the disjoint union of the complement of a G_δ set, which is F_σ, and a set of Lebesgue measure zero. Since $\mathcal{L}^n$ is a σ-algebra, any such set is Lebesgue measurable. $\square$

For any collection $\mathcal{S}$ of subsets of $\mathbf{R}^n$, the intersection of all σ-algebras of subsets of $\mathbf{R}^n$ that contain $\mathcal{S}$ is the smallest σ-algebra that contains $\mathcal{S}$, in the sense that it is a σ-algebra that contains $\mathcal{S}$ and is contained in any other σ-algebra that contains $\mathcal{S}$. The smallest σ-algebra that contains the open subsets of $\mathbf{R}^n$ is called the Borel σ-algebra and denoted by $\mathcal{B}(\mathbf{R}^n)$. Since every open set is in the σ-algebra $\mathcal{L}^n$, by minimality, $\mathcal{B}(\mathbf{R}^n) \subseteq \mathcal{L}^n$. The measure space $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mu_n)$ is called the **Borel measure space** on $\mathbf{R}^n$. We have the following consequence of the preceding lemmas.

Corollary 7 *The collection $\mathcal{L}^n$ of Lebesgue measurable subsets of $\mathbf{R}^n$ is the smallest σ-algebra that contains the Borel σ-algebra and all sets of Lebesgue outer-measure zero.*

It is natural to inquire if, as in the case $n = 1$, in general, there are subsets of $\mathbf{R}^n$ that are not Lebesgue measurable. A very minor variation of Vitali's construction of such subsets of $\mathbf{R}$ provides a similar construction for $\mathbf{R}^n$.

Theorem 8 (**Vitali**) *If $E \subseteq \mathbf{R}^n$ and $\mu_n^*(E) > 0$, then there is a subset of E that is non-measurable.*

Proof In view of the countable monotonicity of μ_n^*, we may assume that E is bounded. Let Λ be the countable subset of $\mathbf{R}^n$ of all points with rational coordinates. For $u, v \in E$, define $u \approx v$ provided that $u - v \in \Lambda$. This is an equivalence relation on E, and therefore E is the disjoint union of equivalence classes. By the Axiom of Choice, there is a choice set $C_E \subseteq E$ for this relation, that is, C_E comprises exactly one point from each equivalence class, that is,

$$\text{the countable collection of translates } \{C_E + \lambda\}_{\lambda \in \Lambda} \text{ is disjoint and } E \subseteq \bigcup_{\lambda \in \Lambda} [C_E + \lambda].$$

We claim that C_E is not Lebesgue measurable, and to verify this, we assume otherwise and obtain a contradiction. Choose Λ_0 to be a countably infinite, bounded subset of Λ. Since E is bounded, the set $\bigcup_{\lambda \in \Lambda_0} [C_E + \lambda]$ is bounded, and, by the translation invariance of measure, is Lebesgue measurable. Consequently, by the countable additivity of measure,

$$\mu_n\Big(\bigcup_{\lambda \in \Lambda_0} [C_E + \lambda]\Big) = \sum_{\lambda \in \Lambda_0} \mu_n(C_E + \lambda) < \infty.$$

However, $\mu_n(C_E + \lambda) = \mu_n(C_E)$, for every λ in the countably infinite set Λ_0, and therefore $\mu_n(C_E) = 0$. By the countable monotonicity of μ_n^*, the inclusion $E \subseteq \bigcup_{\lambda \in \Lambda} [C_E + \lambda]$ and the countability of Λ, there is the following contradiction:

$$0 < \mu_n^*(E) \leq \sum_{\lambda \in \Lambda} \mu_n^*(C_E + \lambda) = \sum_{\lambda \in \Lambda} \mu_n(C_E + \lambda) = 0. \qquad \square$$

Corollary 9 *There are disjoint subsets A and B of $\mathbf{R}^n$ for which*

$$\mu_n^*(A \cup B) < \mu_n^*(A) + \mu_n^*(B).$$

Proof If there were not two such subsets, then every subset of $\mathbf{R}^n$ would be Lebesgue measurable, a contradiction of the preceding theorem. □

PROBLEMS

1. Show that the Cauchy-Schwarz Inequality is an equality if and only if the points are linearly dependent.
2. Verify the following generalization of the Pythagorean Identity called the Parallelogram Law:
$$\|x + y\|^2 + \|x - y\|^2 = 2[\|x\|^2 + \|y\|^2 \text{ for all } x, y \in \mathbf{R}^n.$$
3. For $1 \le p \le \infty$, define the $\|\cdot\|_p$ norm on $\mathbf{R}^n$ and prove the Hölder and Minkowski Inequalities. (Suggestion: Identify points in $\mathbf{R}^n$ with functions $[1, n+1)$ that are constant on each subinterval $[k, k+1)$.) Show that for all $1 \le \alpha, \beta \le \infty$, there is a constant $c = c(\alpha, \beta)$ for which
$$\|x\|_\alpha \le c \cdot \|x\|_\beta \text{ for all } x \in \mathbf{R}^n.$$
(Suggestion: A continuous function on a compact set has a minimum value.)
4. Show that, in Urysohn's Lemma, if A fails to be compact, there may not be a compactly supported, separating function f.
5. Show that Lebesgue outer-measure on $\mathbf{R}$ is changed if, following Jordan, countable covers are restricted to finite covers. (Suggestion: Consider $E = \mathbf{Q} \cap [0, 1]$.)
6. For the cube $Q \subseteq \mathbf{R}^3$ with each side $[0, 1]$, let Q be the disjoint union of k intervals. Prove additivity of volume for this partition without using Lemma 1.
7. Let A and B be of $\mathbf{R}^n$ for which there is a $c > 0$ such that $\|u - v\| \ge c$ for all $u \in A, v \in B$. Show that
$$\mu_n^*(A \cup B) = \mu_n^*(A) + \mu_n^*(B).$$
8. Prove that $\mathcal{L}^n$ is the smallest σ-algebra that contains all Borel sets and all sets of Lebesgue measure zero.
9. If $E \in \mathcal{L}^n$ and $\epsilon > 0$, is there a compact set F such that $F \subseteq E$ and $\mu_n(E \sim F) < \epsilon$?
10. Let E be any set of real numbers. Show that $E \times \mathbf{Q}$ is a Lebesgue measurable subset of $\mathbf{R}^2$, regardless of the Lebesgue measurability of $E \subseteq \mathbf{R}$.
11. Show that if the mapping $\mathcal{N}$: $\mathbf{R}^n \to \mathbf{R}^n$ is continuous and maps sets of Lebesgue measure zero to sets of Lebesgue measure zero, then it maps Lebesgue measurable sets to Lebesgue measurable sets.
12. Consider the triangle $\Delta = \{(x, y) \in \mathbf{R}^2 \,|\, 0 \le x \le a, 0 \le y \le [b/a]x\}$. By covering Δ with finite collections of rectangles and using the continuity of measure, determine the Lebesgue measure of Δ.
13. Let $[a, b]$ be a closed, bounded interval. Suppose that f: $[a, b] \to \mathbf{R}$ is bounded and Lebesgue measurable. Show that the graph of f has measure zero with respect to Lebesgue measure on the plane. Generalize this to bounded, real-valued measurable functions of several real variables.

10.2 LEBESGUE MEASURABILITY AND MEASURE OF IMAGES OF MAPPINGS

A mapping $\mathcal{N}: E \to \mathbf{R}^n$ defined on a Lebesgue measurable subset of $\mathbf{R}^n$ is said to preserve Lebesgue measurability if it maps measurable sets to measurable sets. In Section 6.6, in order to prove the von Neumann Composition Theorem, we first proved that an increasing, absolutely continuous function on a closed, bounded interval preserves Lebesgue measurability.

Lemma 10 *A continuous mapping $\mathcal{N}: E \to \mathbf{R}^n$ on a measurable subset of $\mathbf{R}^n$ preserves Lebesgue measurability if and only if it preserves sets of Lebesgue measure zero.*

Proof First assume that $\mathcal{N}$ preserves sets of Lebesgue measure zero. By the Bolzano-Weierstrass Theorem, since $\mathcal{N}$ is continuous, it maps closed, bounded subsets of E to closed sets, and therefore maps F_σ subsets of E to F_σ sets. According to Corollary 6, a set is Lebesgue measurable if and only if it is the union of an F_σ set and a set of Lebesgue measure zero. Therefore, $\mathcal{N}$ preserves Lebesgue measurability. For the converse, suppose that $\mathcal{N}$ does not preserve sets of Lebesgue measure zero. Then there is a subset E_0 of E with $\mu_n(E_0) = 0$, while $\mu_n^*(\mathcal{N}(E_0)) > 0$. According to Vitali's non-measurability theorem, there is a non-measurable subset A of $\mathcal{N}(E_0)$. Then $E_0 \cap \mathcal{N}^{-1}(A)$ is measurable, since Lebesgue measure is complete, while its image under $\mathcal{N}$ is non-measurable. $\square$

Definition *A mapping $\mathcal{N}: E \to \mathbf{R}^n$ on a subset E of $\mathbf{R}^n$ is said to be* **Lipschitz** *provided that there is a $c \geq 0$ for which*

$$\|\mathcal{N}(x) - \mathcal{N}(y)\| \leq c \cdot \|x - y\| \textit{ for all } x, y \in E.$$

There is a smallest such c, called the **Lipschitz constant** *for $\mathcal{N}$.*

By a cube $Q_\ell \subseteq \mathbf{R}^n$ of side-length ℓ we mean the n-fold product of intervals of equal length ℓ, and by $\overline{B}_r$ we denote a closed ball of radius r. Observe that each cube Q_ℓ is contained in a closed ball $\overline{B}_r$, where $r = (\sqrt{(n)}/2) \cdot \ell$, and each closed ball $\overline{B}_r$ is contained in a cube Q_ℓ, for $\ell = 2 \cdot r$. A Lipschitz mapping with Lipschitz constant c maps a closed ball of radius r into a closed ball of radius $c \cdot r$. Consequently, if $\mathcal{N}: \mathbf{R}^n \to \mathbf{R}^n$ has Lipschitz constant c, $\ell > 0$, and a cube Q_ℓ is given, setting $r = (\sqrt{(n)}/2) \cdot \ell$, there is a ball $\overline{B}_r$, a ball $\overline{B}_{2\cdot c\cdot r}$ and a cube $Q_{2\cdot c\cdot r}$ for which $\mathcal{N}(Q_\ell) \subseteq \mathcal{N}(\overline{B}_r) \subseteq \overline{B}_{c\cdot r} \subseteq Q_{2\cdot c\cdot r}$. Therefore, if $C = c^n \cdot n^{n/2}$, then

$$\mu_n^*(\mathcal{N}(Q)) \leq C \cdot \mu_n(Q) \text{ for each cube } Q. \tag{5}$$

Theorem 11 *A Lipschitz mapping $\mathcal{N}: \mathcal{O} \to \mathbf{R}^n$ on an open subset $\mathcal{O}$ of $\mathbf{R}^n$ preserves Lebesgue measurability, and in addition, if it has Lipschitz constant c, then, for each Lebesgue measurable subset of E of $\mathcal{O}$,*

$$\mu_n(\mathcal{N}(E)) \leq C \cdot \mu_n(E) \textit{ where } C = c^n \cdot n^{n/2}. \tag{6}$$

Proof It is clear that if I is an interval and $\epsilon > 0$, then there is a finite collection $\{Q_K\}_{k=1}^m$ of cubes that cover I and has $\sum_{k=1}^m \text{vol}(Q_k) < \text{vol}(I) + \epsilon$. Therefore, Lebesgue outer-measure may be defined by coverings by cubes rather than by general intervals, and, since $\mathcal{O}$ is open, we can cover subsets of $\mathcal{O}$ by cubes in $\mathcal{O}$. It immediately follows from (5) that, for each measurable subset E of $\mathbf{R}^n$,

$$\mu_n^*(\mathcal{N}(E)) \leq C \cdot \mu_n(E).$$

In particular, $\mathcal{N}$ preserves sets of measure zero, and since it is continuous, by the preceding lemma, it preserves Lebesgue measurability. Since Lebesgue measure extends outer-measure, (6) follows. □

Corollary 12 *Lebesgue measure is translation invariant, that is, if $E \subseteq \mathbf{R}^n$ is Lebesgue measurable and $z \in \mathbf{R}^n$, then $E + z = \{x + z \,|\, x \in E\}$ also is Lebesgue measurable and*

$$\mu_n(E + z) = \mu_n(E).$$

Proof Define the translation mapping $\mathcal{N} \colon \mathbf{R}^n \to \mathbf{R}^n$ by $\mathcal{N}(x) = x + z$. If I is an interval, then so is $\mathcal{N}(I)$ and obviously $\mu_n(\mathcal{N}(I)) = \mu_n(I)$. Consequently, by the countable monotonicity of μ_n^*,

$$\mu_n^*(\mathcal{N}(E)) \leq \mu_n^*(E) \text{ for all } E \subseteq \mathbf{R}^n.$$

But observe that $\mathcal{N}$ is invertible and its inverse is also a translation mapping, so that, arguing as above,

$$\mu_n^*(E) = \mu_n^*((\mathcal{N}^{-1} \circ \mathcal{N})(E)) \leq \mu_n^*(\mathcal{N}(E)) \text{ for all } E \subseteq \mathbf{R}^n.$$

Therefore,

$$\mu_n^*(\mathcal{N}(E)) = \mu_n^*(E) \text{ for all } E \subseteq \mathbf{R}^n.$$

Consequently, $\mathcal{N}$ preserves sets of Lebesgue measure zero, and since it is continuous, by the preceding lemma theorem, $\mathcal{N}$ maps Lebesgue measurable sets to Lebesgue measurable sets. Lebesgue measure is the restriction of Lebesgue outer-measure, so we have established translation invariance. □

The **Euclidean basis** for $\mathbf{R}^n$ is the subset $\{e_i\}_{i=1}^n$, where, for $1 \leq i \leq n$, e_i has i-th coordinate 1 and other coordinates zero, so that for $x \in \mathbf{R}^n$, $x = \sum_{i=1}^n x_i e_i$. For a linear operator[2] $L \colon \mathbf{R}^n \to \mathbf{R}^n$, define

$$a_{i,j} = \langle L(e_i),\, e_j \rangle \text{ for } 1 \leq i, j \leq n.$$

For $1 \leq j \leq n$, let C_j be the j-the column of the $n \times n$ matrix $(a_{i,j})$ and observe that, by linearity,

$$\langle L(x),\, e_j \rangle = \langle x,\, C_j \rangle \text{ for all } x. \tag{7}$$

We say that L is represented by the $n \times n$ matrix $(a_{i,j})$. Here, we only use representatives of a linear operator with respect to the Euclidean basis. Define $\|L\|$ by

$$\|L\|^2 = \sum_{1 \leq i,j \leq n} a_{i,j}^2. \tag{8}$$

[2]It is customary to call a linear mapping a linear operator. Here, we suggest the distinction between the linear and non-linear by using the symbols L and $\mathcal{N}$.

By applying the Cauchy-Schwarz Inequality to the coordinates of $L(x)$ determined by (7), it follows that $\|L\|$ is a Lipschitz constant for L, and we leave it as an exercise to show that it is the smallest such constant.

Corollary 13 *A proper subspace of* $\mathbf{R}^n$ *has Lebesgue measure zero.*

Proof Let V be a proper subspace of $\mathbf{R}^n$. Choose a k-dimensional subspace W of $\mathbf{R}^n$ for which there is the direct sum $\mathbf{R}^n = W \oplus V$ and $\dim W = k > 0$. By suitably labeling coordinates, we view $\mathbf{R}^n$ as being $\mathbf{R}^k \times \mathbf{R}^{n-k}$. Choose a linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$ that maps $\{0\} \times \mathbf{R}^{n-k}$ onto V. A linear operator is Lipschitz, and consequently, according to the preceding theorem, L preserves sets of Lebesgue measure zero. But V is the image under L of $\{0\} \times \mathbf{R}^{n-k}$, a set of Lebesgue measure zero. □

For a mapping $\mathcal{N}\colon E \to \mathbf{R}^n$ defined on $E \subseteq \mathbf{R}^n$ and $1 \leq i \leq n$, define the real-valued component functions $\mathcal{N}_i\colon \mathbf{R}^n \to \mathbf{R}$ by $\mathcal{N}_i(x) = \langle \mathcal{N}(x), e_i \rangle$, so that

$$\mathcal{N}(x) = (\mathcal{N}_1(x), \ldots, \mathcal{N}_n(x)) \text{ for all } x \in E.$$

We call a mapping on an open subset of $\mathbf{R}^n$ **continuously differentiable** provided that each component function has continuous first-order partial derivatives.

Theorem 14 *A continuously differentiable mapping* $\mathcal{N}\colon \mathcal{O} \to \mathbf{R}^n$ *on an open subset* $\mathcal{O}$ *of* $\mathbf{R}^n$ *preserves Lebesgue measurability.*

Proof Let $\mathcal{F}$ be the collection of all closed, bounded intervals that are contained in $\mathcal{O}$ and are n-fold products of closed, bounded intervals in $\mathbf{R}$ with rational end-points. Since $\mathbf{Q}$ is a dense subset of $\mathbf{R}$, $\mathcal{O}$ is the union of sets in $\mathcal{F}$, and since $\mathbf{Q}$ is countable, so is $\mathcal{F}$. Let $\{I_k\}_{k=1}^{\infty}$ be an enumeration of $\mathcal{F}$. Clearly, $\mathcal{N}$ preserves Lebesgue measurability if its restriction to each I_k does. Let I be a closed, bounded interval in $\mathcal{O}$. We will show that the restriction to I is Lipschitz and so, by an appeal to the preceding theorem, it preserves Lebesgue measurability. For such an I, being continuous functions on I, the first-order partial derivatives of the component functions are uniformly pointwise bounded on I. Choose M for which

$$\left|\frac{\partial \mathcal{N}_i}{\partial x_j}(x)\right| \leq M \text{ for all } 1 \leq i, j \leq n, x \in I.$$

We will show that, for $c = n \cdot M$,

$$\|\mathcal{N}(x) - \mathcal{N}(y)\| \leq c \cdot \|x - y\| \text{ for all } x, y \in I.$$

To verify this Lipschitz inequality, it suffices to show that for each $x, y \in I$, there is a linear operator $L_{x,y}\colon \mathbf{R}^n \to \mathbf{R}^n$ such that

$$\mathcal{N}(x) - \mathcal{N}(y) = L_{x,y}(x - y) \text{ and } \|L_{x,y}\| \leq c. \tag{9}$$

By the Mean Value Theorem for continuously differentiable functions, for $1 \leq i \leq n$, there is a point z_i on the segment joining x to y, and therefore in the convex set I, for which

$$\mathcal{N}_i(x) - \mathcal{N}_i(y) = \langle \nabla \mathcal{N}_i(z_i), x - y \rangle.$$

Define $L_{x,y}$ to be the linear operator that is represented by the $n \times n$ matrix whose i-th row is $\nabla \mathcal{N}_i(z_i)$. Then

$$\mathcal{N}(x) - \mathcal{N}(y) = L_{x,y}(x - y).$$

By the definition of M and the formula for the Lipschitz constant of a linear operator, $\|L_{x,y}\| \leq n \cdot M = c$. Consequently, (9) is established. $\square$

Definition *A linear operator* $L \colon \mathbf{R}^n \to \mathbf{R}^n$ *is said to be* **diagonal** *with diagonal entries* $(\lambda_1, \ldots, \lambda_n) \in \mathbf{R}^n$ *provided that*

$$L(e_i) = \lambda_i \cdot e_i \text{ for } 1 \leq i \leq n.$$

It is said to be **orthogonal** *provided that*

$$\langle L(x),\, L(y)\rangle = \langle x,\, y\rangle \textit{ for all } x, y. \tag{10}$$

Other formulations of orthogonality are given in Problem 21. We will prove that a linear operator is orthogonal, that is, it preserves the inner-product, if and only if it preserves Lebesgue measure. Only representing linear operators with respect to the Euclidean basis, we make the following definition.

Definition *The determinant of a linear operator* $L \colon \mathbf{R}^n \to \mathbf{R}^n$ *that is represented by the* $n \times n$ *matrix* $(a_{i,j})$ *is given by*

$$\det L = \det(a_{i,j}).$$

We now specify the properties of the determinant of a linear operator on $\mathbf{R}^n$ that we need, and, of course, these are inherited from those of the determinant of an $n \times n$ matrix. Of course, the determinant of a diagonal matrix is the product of the diagonal entries. The representing matrix of the composition of operators is the matrix product of the matrix representatives of the composites, and so the following composition property for the determinant of linear operators is inherited from the product property for the determinant of matrices:

$$\det(S \circ L) = \det S \cdot \det L.$$

The operator that is represented by the adjoint matrix of the representative of L is called the adjoint of L and denoted by L^*. The determinant of the adjoint matrix equals the determinant of the matrix, and so:

$$\langle L(x),\, y\rangle = \langle x,\, L^*(y)\rangle \text{ for all } x, y \text{ and } \det L = \det L^*.$$

Lemma 15 *If the linear operator* $L \colon \mathbf{R}^n \to \mathbf{R}^n$ *is orthogonal, then* $|\det L| = 1$.

Proof Since

$$\|L(x)\|^2 = \langle L(x),\, L(x)\rangle = \langle x,\, x\rangle = \|x\|^2 \text{ for all } x,$$

$\ker L = \{0\}$, and so L is invertible. Observe that for $1 \leq i, j \leq n$,

$$\langle L^{-1}(e_i),\, e_j\rangle = \langle (L \circ L^{-1})(e_i),\, L(e_j)\rangle = \langle e_i,\, L(e_j)\rangle = \langle L^*(e_i),\, e_j\rangle,$$

that is, the matrix that represents L^{-1} is the same as the one that represents L^*. Consequently, if Id is the identity operator on $\mathbf{R}^n$, then

$$1 = \det \mathrm{Id} = \det L \circ L^{-1} = \det L \circ L^* = \det L \cdot \det L,$$

that is, $|\det L| = 1$. □

Theorem 16 *Let $L\colon \mathbf{R}^n \to \mathbf{R}^n$ be a diagonal linear operator. Then*

$$\mu_n(L(E)) = |\det L| \cdot \mu_n(E) \text{ for all } E \in \mathcal{L}^n. \tag{11}$$

Proof We have $\det L = \lambda_1 \cdots \lambda_n$. If some $\lambda_k = 0$, then L is not invertible, so its image is a proper subspace of $\mathbf{R}^n$. According to Corollary 13, (11) holds since each side is zero. Assume that $\lambda_k \neq 0$, for all k. Define $\gamma = |\lambda_1 \cdots \lambda_n|$. If I is an interval, obviously (11) holds, so that, by the countable monotonicity of μ_n^*,

$$\mu_n^*(L(E) \leq \gamma \cdot \mu_n^*(E) \text{ for all } E \subseteq \mathbf{R}^n.$$

Since L is diagonal and invertible, L^{-1} is also an diagonal operator, with γ replaced by $1/\gamma$. Therefore, as argued above,

$$\mu_n^*(E) = \mu_n^*((L^{-1} \circ L)(E)) \leq 1/\gamma \cdot \mu_n^*(L(E)) \text{ for all } E \subseteq \mathbf{R}^n.$$

Consequently,

$$\mu_n^*(L(E)) = \gamma \cdot \mu_n^*(E) \text{ for all } E \subseteq \mathbf{R}^n.$$

Since L maps Lebesgue measurable sets to Lebesgue measurable, and Lebesgue measure is the restriction of Lebesgue outer-measure, (11) is established. □

Taking $r > 0$ and $L = r \cdot \mathrm{Id}$, we have the following formula for the Lebesgue measure of a closed ball.

Corollary 17 *For $r > 0$,*

$$\mu_n \{x \in \mathbf{R}^n \mid \|x\| \leq r\} = r^n \cdot \mu_n \{x \in \mathbf{R}^n \mid \|x\| \leq 1\}.$$

By the continuity of measure, the measure of an open ball is equal to the measure of its closure, and, by the translation invariance of Lebesgue measure and this corollary, the measure of any ball of radius r is r^n times the measure of any ball of radius 1.

Lemma 18 *A linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$ is orthogonal if and only if it maps open balls onto open balls of the same radius.*

Proof Let B be the open ball centered at the origin that has radius 1. Then a linear operator maps open balls onto open balls of the same radius if and only if it maps B onto B. Let $x, y \in \mathbf{R}^n$. Assume that L is orthogonal. Then

$$\|L(x)\|^2 = \langle L(x),\, L(x)\rangle = \langle x,\, x\rangle = \|x\|^2.$$

Therefore, $\ker L = \{0\}$, so L is invertible, and it preserves norms, so that it maps B onto B. To prove the converse, assume L maps B onto B. So L is invertible, and its inverse has the same property. So L preserves norms, and we claim it preserves inner-products. Indeed, expand, using bilinearity, the identity

$$\langle L(x+y),\, L(x+y)\rangle = \langle x+y,\, x+y\rangle,$$

and substitute $\langle L(x),\, L(x)\rangle = \langle x,\, x\rangle$ and $\langle L(y),\, L(y)\rangle = \langle y,\, y\rangle$, to obtain

$$\langle L(x),\, L(y)\rangle = \langle x,\, y\rangle \text{ for all } x, y,$$

that is, L is orthogonal. □

Every open subset of $\mathbf{R}$ is the countable, disjoint union of open intervals, and therefore every open, bounded subset is the countable, disjoint union of open balls. Every open, unbounded subset of $\mathbf{R}$ is, after the excision of a collection of integers, the countable, disjoint union of open balls. Now, the open, bounded subset of the plane $(0,\,1)\times(0,\,1)$ is not the countable, disjoint union of open balls. However, there is the following theorem of Vitali for general Euclidean spaces.

The Vitali Partition Theorem. *Every non-empty open subset of $\mathbf{R}^n$ is, after the excision of a set of Lebesgue measure zero, the union of a countable, disjoint collection of open balls.*

Proof For $r>0$, let $\overline{B}_r \subseteq \mathbf{R}^n$ denote a closed ball of radius r, B_r denote its interior, and observe that, by Corollary 17 and the continuity and translation invariance of Lebesgue measure, $\mu_n(\overline{B}_r) = \mu_n(B_r) = c\cdot r^n$, where $c = \mu_n(\overline{B}_1)$. Let $\mathcal{O} \subseteq \mathbf{R}^n$ be non-empty and open, and we may assume that it is bounded. Let $\mathcal{F}$ be the collection of closed balls that are contained in $\mathcal{O}$. Observe that if $\{\overline{B}_{r_i}\}_{i=1}^{\infty}$ is a disjoint subcollection of $\mathcal{F}$, then

$$\sum_{i=1}^{\infty} r_i^n \leq 1/c\cdot\mu_n(\mathcal{O}) < \infty. \tag{12}$$

First, we inductively define a disjoint, countable subcollection $\{\overline{B}_{r_i}\}_{i=1}^{\infty}$ of $\mathcal{F}$ such that, for each index $k>1$,

$$\text{if } \overline{B}_r \in \mathcal{F} \text{ and } \overline{B}_r \cap \bigcup_{i=1}^{k-1} \overline{B}_{r_i} = \emptyset, \text{ then } r < 2\cdot r_k. \tag{13}$$

To do so, let $\overline{B}_{r_1}$ be any ball in $\mathcal{F}$. For an index m, assume that a disjoint subcollection $\{\overline{B}_{r_i}\}_{i=1}^{m}$ of $\mathcal{F}$ has been chosen so that (13) holds for $1\leq k\leq m$. Since $\mathbf{R}^n$ is connected, the closed set $\bigcup_{i=1}^{m} \overline{B}_{r_i}$ must be a proper subset of the open set $\mathcal{O}$. Therefore, there is a ball in $\mathcal{F}$ that is disjoint from $\bigcup_{i=1}^{m} \overline{B}_{r_i}$, and, since $\mathcal{O}$ is bounded, by choosing r_{m+1} greater that half the supremum of the radii of all such balls, we complete the inductive definition of a subcollection $\{\overline{B}_{r_i}\}_{i=1}^{\infty}$ of $\mathcal{F}$ for which (13) holds. We claim that for each k,

$$\mu_n\left(\mathcal{O} \sim \bigcup_{i=1}^{k} \overline{B}_{r_i}\right) \leq 5^n\cdot c\cdot\left[\sum_{i=k+1}^{\infty} r_i^n\right]. \tag{14}$$

Once this is verified, by replacing each $\overline{B}_{r_i}$ by its interior and using (12), the proof is concluded. To verify the claim, let $x \in \mathcal{O} \sim \bigcup_{i=1}^{k} \overline{B}_{r_i}$. There is a ball $\overline{B}_r \in \mathcal{F}$ centered at x that is disjoint from each $\overline{B}_{r_i}$, $1 \leq i \leq k$. But $\overline{B}_r$ cannot be disjoint from every $\overline{B}_{r_i}$, since otherwise, by (13), $r < 2 \cdot r_i$ for all i, which contradicts the convergence of $\{r_i\}$ to zero. Let $K(k) > k$ be the first index for which $\overline{B}_r \cap \overline{B}_{r_{N(k)}} \neq \emptyset$. Then $\overline{B}_r$ is disjoint from each $\overline{B}_{r_i}$ for $1 \leq i < K(k) - 1$, and so, again by (13), $r < 2 \cdot r_{K(k)}$. Therefore, by the triangle inequality, $\overline{B}_r$ is contained in a closed ball with the same center as $\overline{B}_{r_{K(k)}}$ and five times its radius. Consequently, taking unions,

$$\mathcal{O} \sim \bigcup_{i=1}^{k} \overline{B}_{r_i} \subseteq \bigcup_{i=k+1}^{\infty} \overline{B}_{5 \cdot r_i},$$

so that, by the monotonicity and countable additivity of μ_n, (14) holds. $\square$

Theorem 19 *A linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$ is orthogonal if and only if it preserves Lebesgue measure, that is,*

$$\mu_n(L(E)) = \mu_n(E) \text{ for all } E \in \mathcal{L}^n. \tag{15}$$

Proof First, assume that L preserves Lebesgue measure. Let B be the open ball centered at the origin of radius 1. Since $\mu_n(L(B)) = 1$, it follows from Corollary 13 that $L(\mathbf{R}^n)$ is not contained in a proper subspace, so L is invertible. The operators L and L^{-1} are continuous, so $L(B)$ is open, and, by linearity, it is an open ball centered at the origin of radius r. It follows from Corollary 17 that $r = 1$. Therefore, L maps B onto B, and, again by linearity, it maps open balls onto open balls of the same radius. According to the preceding lemma, L is orthogonal. To prove the converse, assume that L is orthogonal. By the preceding lemma and the translation invariance of Lebesgue measure, (15) holds for each open ball. According to Lemma 10, L maps sets of Lebesgue measure zero to sets of Lebesgue measure zero. Consequently, by an appeal to the Vitali Partition Theorem, (15) holds for each open set. By the continuity of Lebesgue measure, (15) also holds for each bounded G_δ set. According to Theorem 5, each Lebesgue measurable set is a G_δ set from which a set of Lebesgue measure zero has been excised, so that (15) holds for each bounded, Lebesgue measurable set. Therefore, it holds in general, since each Lebesgue measurable set has a countable Lebesgue measurable partition into bounded, Lebesgue measurable sets. $\square$

For linear operators, in addition to the properties of the determinant, we need the following composition theorem.

Theorem 20 *Each linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$ may be expressed as a composition of linear operators,*

$$L = A \circ D \circ B, \tag{16}$$

where A and B are orthogonal and D is diagonal.

Theorem 21 *A linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$ maps Lebesgue measurable sets to Lebesgue measurable sets, and*

$$\mu_n(L(E)) = |\det L| \cdot \mu_n(E) \text{ for all } E \in \mathcal{L}^n.$$

Proof Write $L = A \circ D \circ B$, as in the above theorem. According to Theorems 16 and 19, $\mu_n(L(E)) = \mu_n((A \circ D \circ B)(E)) = \mu_n((D \circ B)(E)) = |\det D| \cdot \mu_n(B(E)) = |\det D| \cdot \mu_n(E)$. On the other hand, by Lemma 15,

$$|\det L| = |\det[A \circ D \circ B]| = |\det A \cdot \det D \cdot \det B| = |\det A| \cdot |\det D| \cdot |\det B| = |\det D|.$$

□

For $\{z_1, \ldots, z_n\} \subseteq \mathbf{R}^n$, define $\mathcal{P}(z_1, \ldots, z_n)$ to be the set of points $\sum_{i=1}^{n} x_i \cdot z_i$, with coefficients $0 \leq x_i \leq 1$, for $1 \leq i \leq n$. Such a set is called a **parallelepiped**. Define $\det(z_1, \ldots, z_n)$ to be the determinant of the $n \times n$ matrix whose i-th column is z_i.

Corollary 22 *For* $\{z_1, \ldots, z_n\} \subseteq \mathbf{R}^n$,

$$\mu_n(\mathcal{P}(z_1, \ldots, z_n)) = |\det(z_1, \ldots, z_n)|.$$

Proof Define the linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$ by $L(x) = \sum_{i=1}^{n} x_i \cdot z_i$. Then the matrix that represents L is the $n \times n$ matrix whose i-th column is z_i. If I is the n-fold product of $[0, 1]$, then L maps I onto $\mathcal{P}$. An appeal to the preceding theorem establishes the formula. □

PROBLEMS

14. Show that a continuously differentiable mapping $\mathcal{N}\colon \mathbf{R}^n \to \mathbf{R}^n$ may not be Lipschitz, but its restriction to each closed, bounded set is.
15. For a linear operator, compare the estimate (6) with the actual measure.
16. Let $R \subseteq \mathbf{R}^n$ be a bounded rectangle and $\epsilon > 0$. Show that there is a finite cover of R by cubes, the sum of whose volumes is less than $\text{vol}(R) + \epsilon$. Conclude from this that Lebesgue outer-measure on $\mathbf{R}^n$ is unaltered if coverings by intervals are restricted to coverings by cubes.
17. Show that each open subset of $\mathbf{R}^n$ is an F_σ set.
18. Show that the open, bounded subset of the plane $(0, 1) \times (0, 1)$ is not the union of a countable, disjoint collection of open balls in $\mathbf{R}^2$. Does this contradict the Vitali Partition Theorem?
19. Show that for $n = 1$ the constant C in (6) cannot be improved.
20. Show that the Cauchy-Schwarz Inequality is an equality if and only if x and y are linearly dependent. Use this to show that $\|L\|$ is the smallest Lipchitz constant for a linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$.
21. A pair of points in $\mathbf{R}^n$ is said to be orthogonal provided that their inner-product is zero. A set $\{v_1, \ldots v_n\} \subseteq \mathbf{R}^n$ is said to be an orthonormal basis for $\mathbf{R}^n$ provided that each pair of points in the set is orthogonal, and each point has norm 1. For a linear operator $L\colon \mathbf{R}^n \to \mathbf{R}^n$, prove the equivalence of the following assertions:
    (i) L is orthogonal
    (ii) L is an isometry, that is
    $$\|L(x) - L(y)\| = \|x - y\| \text{ for all } x, y \in \mathbf{R}^n.$$

(iii) L maps the Euclidean basis $\{e_1, \ldots, e_n\}$ onto an orthonormal basis.

(iv) L is invertible and its inverse equals its adjoint.

22. Show that, in $\mathbf{R}^3$, Corollary 22 reduces to the triple-product formula for volume.
23. (Lebesgue) Prove that for any open cover of a closed, bounded subset of $\mathbf{R}^n$ there is an $\alpha > 0$ (called the Lebesgue number of the cover) such that every ball centered in K and of radius α is contained in one of the covering sets.
24. The collection of invertible linear operators on $\mathbf{R}^n$ is called the general linear group and denoted by $GL(n)$; the collection of orthogonal operators is denoted by $\mathcal{O}(n)$ and called the orthogonal group. Show that with the group operation of composition, $GL(n)$ is a group that has $\mathcal{O}(n)$ as a subgroup.

10.3 BOREL MEASURES ON $\mathbf{R}^n$ AND REGULARITY

In the preceding section, we proved that Lebesgue measure on Euclidean space is regular. We now establish a regularity theorem for finite Borel measures on Euclidean space. For a Borel measure $\mu\colon \mathcal{B}(\mathbf{R}^n) \to [0, \infty]$, here, it is convenient to say that a set $E \in \mathcal{B}(\mathbf{R}^n)$ is regular provided that for every $\epsilon > 0$, there is a compact set K and open set $\mathcal{O}$ for which

$$K \subseteq E \subseteq \mathcal{O},\ \mu(E \sim K) < \epsilon \text{ and } \mu(\mathcal{O} \sim E) < \epsilon. \tag{17}$$

If every set in $\mathcal{B}(\mathbf{R}^n)$ is regular, then the measure μ is said to be regular. In the proofs of the following lemma and theorem, the assumption that $\mu(\mathbf{R}^n)$ is finite permits use of the excision and intersection-continuity properties of measure.

Lemma 23 *If $\mu\colon \mathcal{B}(\mathbf{R}^n) \to [0, \infty)$ is a finite Borel measure, then the set $\mathbf{R}^n$ is regular.*

Proof Let $\epsilon > 0$. It suffices to show that there is a compact set K for which $\mu(\mathbf{R}^n \sim K) < \epsilon$. Choose $\{E_k\}_{k=1}^{\infty}$ to be any measurable partition of $\mathbf{R}^n$ into bounded Borel sets. We have

$$\mu(\mathbf{R}^n) = \sum_{k=1}^{\infty} \mu(E_k) < \infty.$$

Choose an index k_0 so that $\mu(\mathbf{R}^n) - \sum_{k=1}^{k_0} \mu(E_k) < \epsilon$. Then $\bigcup_{k=1}^{k_0} E_k$ is a bounded set, and so its closure K is compact. By the excision property of measure,

$$\mu(\mathbf{R}^n \sim K) = \mu(\mathbf{R}^n) - \mu(K) \leq \mu(\mathbf{R}^n) - \sum_{k=1}^{k_0} \mu(E_k) < \epsilon. \qquad \square$$

Theorem 24 *Every finite Borel measure $\mu\colon \mathcal{B}(\mathbf{R}^n) \to [0, \infty)$ is regular.*

Proof Denote the collection of regular sets by $\mathcal{S}$. We will show that $\mathcal{S}$ is a σ-algebra that contains all closed sets, and therefore, by minimality, $\mathcal{S} = \mathcal{B}(\mathbf{R}^n)$. According to the above lemma, the whole space $\mathbf{R}^n$ is regular. First, we show that the complement of a regular set is regular. Let $E \in \mathcal{S}$ and $\epsilon > 0$. There is a closed set F and open set $\mathcal{O}$ for which

$$F \subseteq E \subseteq \mathcal{O},\ \mu(E \sim K) < \epsilon \text{ and } \mu(\mathcal{O} \sim E) < \epsilon/2.$$

Define $\mathcal{O}' = \mathbf{R}^n \sim F$ and $F' = \mathbf{R}^n \sim \mathcal{O}$. Then $\mathcal{O}'$ is open, F' is closed, and

$$F' \subseteq \mathbf{R}^n \sim E \subseteq \mathcal{O}', \;\; \mu((\mathbf{R}_n \sim E) \sim F') < \epsilon/2 \text{ and } \mu(\mathcal{O}' \sim (\mathbf{R}^n \sim E)) < \epsilon.$$

Since $\mathbf{R}^n$ is regular, there is a compact set K for which $\mu(\mathbf{R}^n \sim K) < \epsilon/2$. Now, $K' = F' \cap K$, being the intersection of a closed set and compact set, is compact. Since

$$(\mathbf{R}^n \sim E) \sim K' \subseteq ((\mathbf{R}^n \sim E) \sim F) \cup (\mathbf{R}^n \sim K),$$

$\mu((\mathbf{R}^n \sim E) \sim K') < \epsilon$ and we already have $\mu(\mathcal{O}' \sim (\mathbf{R}^n \sim E)) < \epsilon$, so that $\mathbf{R}^n \sim E$ is regular. We now show that $\mathcal{S}$ is a σ-algebra. Let $E = \bigcup_{k=1}^{\infty} E_k$, the countable union of regular sets. Let $\epsilon > 0$. For each k, define $E_k' = \mathbf{R}^n \sim E_k$, so that since E_k and E_k' are inner-regular, there are compact sets K_k and K_k' for which

$$K_k \subseteq E_k, \;\; \mu(E_k \sim K_k) < \epsilon/2^{k+1} \text{ and } K_k' \subseteq E_k', \;\; \mu(E_k' \sim K_k') < \epsilon/2^k.$$

Since $E \sim \bigcup_{k=1}^{\infty} K_k \subseteq \bigcup_{k=1}^{\infty}(E_k \sim K_k)$, $\mu(E \sim \bigcup_{k=1}^{\infty} K_k) < \epsilon/2$, and therefore, by the continuity of μ, there is an index k_0 for which,

$$\mu\left(\bigcup_{k=1}^{k_0} K_k\right) > \mu(E) - \epsilon.$$

Define $K = \bigcup_{k=1}^{k_0} K_k$. Then K, being the finite union of compact sets, is compact and K is contained in E. We have $\mu(E \sim K) = \mu(E) - \mu(K) < \epsilon$. This establishes the inner-regularity property for E. Define $K' = \bigcap_{k=1}^{\infty} K_k'$. Then K', being the intersection of compact sets, also is compact. Now, by the De Morgan's Identities,

$$(\mathbf{R}^n \sim E) \sim K' = \bigcup_{k=1}^{\infty}((\mathbf{R}^n \sim E) \sim K_k').$$

By the countable monotonicity of μ, $\mu((\mathbf{R}^n \sim E) \sim K') < \epsilon$. Define $\mathcal{O} = \mathbf{R}^n \sim K'$. Being the complement of a closed set, $\mathcal{O}$ is open and $\mathcal{O} \sim E = (\mathbf{R}^n \sim E) \sim K'$ so that $\mu(\mathcal{O} \sim E) < \epsilon$. Therefore, E is regular. So the regular sets are a σ-algebra. To conclude the proof, it suffices to show that the closed sets are regular. Since every closed set is the union of a countable collection of compact sets and $\mathcal{S}$ is closed with respect to countable unions, it suffices to let K be compact and $\epsilon > 0$ and show that there is an open set $\mathcal{O}$ that contains K and $\mu(\mathcal{O} \sim K) < \epsilon$. For such a compact set K, the distance function $\operatorname{dist}_K\colon \mathbf{R}^n \to \mathbf{R}$ function is continuous, since it is Lipschitz and clearly, since K is closed, $x \in K$ if and only if $\operatorname{dist}_K(x) = 0$. For each k, define

$$\mathcal{O}_k = \{x \in \mathbf{R}^n \mid \operatorname{dist}_K(x) < 1/k\}.$$

Since dist_K is continuous, $\mathcal{O}_k$ is open, and $K = \bigcap_{k=1}^{\infty} \mathcal{O}_k$. So $\{\mathcal{O}_k\}_{k=1}^{\infty}$ is a descending, countable collection of measurable sets whose intersection is K. Let $\epsilon > 0$. By the intersection-continuity and excision properties of measure, there is an index k_0 for which $\mu(\mathcal{O}_{k_0}) - \mu(K) < \epsilon$, and so $\mu(\mathcal{O}_{k_0} \sim K) < \epsilon$. Therefore, $F \in \mathcal{S}$. Consequently, the σ-algebra $\mathcal{S}$ is $\mathcal{B}(\mathbf{R}^n)$. $\square$

In Chapter 13, we prove Ulam's Theorem, which extends the above theorem by asserting that a finite Borel measure on a complete, separable metric space X is regular. The proof of the theorem is very similar, but the proof of the lemma regarding the regularity of the whole space X requires ingenuity.

Let $I = [a, b]$ be a closed, bounded interval in $\mathbf{R}$ and $\mathcal{B}(I)$ the collection of Borel subsets of I. A measure μ on $\mathcal{B}(I)$ is called a **Borel measure**. For such a measure, define the function $g_\mu \colon I \to \mathbf{R}$ by

$$g_\mu(x) = \mu[a, x] \text{ for all } x \text{ in } I.$$

The function g_μ is called the **cumulative distribution function** of μ.

Proposition 25 *Let I be a closed, bounded interval and μ be a finite Borel measure on $\mathcal{B}(I)$. Then its cumulative distribution function g_μ is increasing and continuous on the right. Conversely, each function $g \colon I \to \mathbf{R}$ that is increasing and continuous on the right is the cumulative distribution function of a unique finite Borel measure μ_g on $\mathcal{B}(I)$.*

Proof First, let μ be a finite Borel measure on $\mathcal{B}(I)$. Its cumulative distribution function is certainly increasing and bounded. Let x_0 belong to $[a, b)$ and $\{x_k\}$ be a decreasing sequence in $(x_0, b]$ that converges to x_0. Then $\bigcap_{k=1}^\infty (x_0, x_k] = \emptyset$ so that, since μ is finite, by the continuity of measure,

$$0 = \mu(\emptyset) = \lim_{k\to\infty} \mu(x_0, x_k] = \lim_{k\to\infty} [g_\mu(x_k) - g_\mu(x_0)].$$

Thus g_μ is continuous on the right at x_0. To prove the converse, let $g \colon I \to \mathbf{R}$ be an increasing function that is continuous on the right. Consider the collection $\mathcal{S}$ of subsets of I consisting of the empty-set, the singleton set $\{a\}$, and all subintervals of I of the form $(c, d]$. Then $\mathcal{S}$ is a semi-ring. Consider the set-function $\mu \colon \mathcal{S} \to \mathbf{R}$ defined by setting $\mu(\emptyset) = 0, \mu\{a\} = g(a)$ and

$$\mu(c, d] = g(d) - g(c) \text{ for } (c, d] \subseteq I.$$

We leave it as an exercise (see Problem 25) to verify that if $(c, d] \subseteq I$ is the union of finite, disjoint collection $\bigcup_{k=1}^n (c_k, d_k]$, then

$$g(d) - g(c) = \sum_{k=1}^n [g(d_k) - g(c_k)]$$

and that if $(c, d] \subseteq I$ is covered by the countable collection $\bigcup_{k=1}^\infty (c_k, d_k]$, then

$$g(d) - g(c) \leq \sum_{k=1}^\infty [g(d_k) - g(c_k)]. \tag{18}$$

This means that μ is a premeasure on the semi-ring $\mathcal{S}$. By the Carathéodory-Hahn Theorem, the Carathéodory measure $\overline{\mu}$ induced by μ is an extension of μ. In particular, each open subset of $[a, b]$ is μ^*-measurable. By the minimality of the Borel σ-algebra, the σ-algebra of μ^*-measurable sets contains $\mathcal{B}(I)$. The function g is the cumulative distribution function for the restriction of $\overline{\mu}$ to $\mathcal{B}(I)$ since for each $x \in [a, b]$,

$$\mu[a, x] = \mu\{a\} + \mu(a, x] = g(a) + [g(x) - g(a)] = g(x). \qquad \square$$

PROBLEMS

25. Prove the inequality (18). (Hint: Choose $\epsilon > 0$. By the continuity on the right of g, choose $\eta_i > 0$ so that $g(b_i + \eta_i) < g(b_i) + \epsilon 2^{-i}$, and choose $\delta > 0$ so that $g(a+\delta) < g(a) + \epsilon$. Then the open intervals $(a_i, b_i + \eta_i)$ cover the closed interval $[a+\delta, b]$.)
26. For a finite measure μ on $\mathcal{B}(\mathbf{R})$, define $g\colon \mathbf{R} \to \mathbf{R}$ by setting $g(x) = \mu(-\infty, x]$. Show that each bounded, increasing function $g\colon \mathbf{R} \to \mathbf{R}$ that is continuous on the right and $\lim_{x \to -\infty} g(x) = 0$ is the cumulative distribution function of a unique finite Borel measure on $\mathcal{B}(\mathbf{R})$.
27. Show that every closed subset of $\mathbf{R}^n$ is a G_δ set, and every open subset of $\mathbf{R}^n$ is an F_σ set.
28. Show that Theorem 24 may not hold if $\mu(\mathbf{R}^n) = \infty$.
29. Show that the regularity of Lebesgue measure on $\mathbf{R}^n$ is a consequence of Theorem 24.

10.4 CARATHÉODORY OUTER-MEASURES AND HAUSDORFF MEASURES

Lebesgue outer-measure on Euclidean space $\mathbf{R}^n$ has the property that if A and B are subsets of $\mathbf{R}^n$ and there is a $\delta > 0$ for which $\|u - v\| \geq \delta$ for all $u \in A$ and $v \in B$, then

$$\mu_n^*(A \cup B) = \mu_n^*(A) + \mu_n^*(B).$$

In this brief section, we consider measures induced by outer-measures on a metric space that possess this property and a particular class of such measures called Hausdorff measures. We only refer to the most elementary properties of metric spaces.

Let X be a set and Γ a collection of real-valued functions on X. It is often of interest to know conditions under which an outer-measure μ^* has the property that every function in Γ is measurable with respect to the measure induced by μ^* through the Carathéodory construction. We present a sufficient criterion for this. Two subsets A and B of X are said to be separated by the real-valued function f on X provided that there are real numbers a and b with $a < b$ for which $f \leq a$ on A and $f \geq b$ on B.

Proposition 26 *Let φ be a real-valued function on a set X and $\mu^*\colon 2^X \to [0, \infty]$ an outer-measure with the property that whenever two subsets A and B of X are separated by φ, then*

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

Then φ is measurable with respect to the measure induced by μ^*.

Proof Let a be a real number. We must show that the set

$$E = \{x \in X \mid \varphi(x) > a\}$$

is μ^*-measurable, that is, that for any $\epsilon > 0$ and any subset A of X of finite outer-measure,

$$\mu^*(A) + \epsilon \geq \mu^*(A \cap E) + \mu^*(A \sim E). \tag{19}$$

Define $B = A \cap E$ and $C = A \sim E$. For each index n, define

$$B_n = \{x \in B \mid \varphi(x) > a + 1/n\} \text{ and } R_n = B_n \sim B_{n-1}.$$

We have

$$B = B_n \cup \left[\bigcup_{k=n+1}^{\infty} R_k \right].$$

Now on B_{n-2} we have $\varphi > a + 1/(n-2)$, while on R_n we have $\varphi \leq a + 1/(n-1)$. Thus φ separates R_n and B_{n-2} and so separates R_{2k} and $\bigcup_{j=1}^{k-1} R_{2j}$, since the latter set is contained in B_{2k-2}. Consequently, we argue by induction to show that for each k,

$$\mu^* \left[\bigcup_{j=1}^{k} R_{2j} \right] = \mu^*(R_{2k}) + \mu^* \left[\bigcup_{j=1}^{k-1} R_{2j} \right] = \sum_{j=1}^{k} \mu^*(R_{2j}).$$

Since $\sum_{j=1}^{k} R_{2j} \subseteq B \subseteq A$, we have $\sum_{j=1}^{k} \mu^*(R_{2j}) \leq \mu^*(A)$, and so the series $\sum_{j=1}^{\infty} \mu^*(R_{2j})$ converges. Similarly, the series $\sum_{j=1}^{\infty} \mu^*(R_{2j+1})$ converges, and so does the series $\sum_{k=1}^{\infty} \mu^*(R_k)$. Choose n so large that $\sum_{k=n+1}^{\infty} \mu^*(R_k) < \epsilon$. Then by the countable monotonicity of μ^*,

$$\mu^*(B) \leq \mu^*(B_n) + \sum_{k=n+1}^{\infty} \mu^*(R_k) < \mu^*(B_n) + \epsilon$$

or

$$\mu^*(B_n) > \mu^*(B) - \epsilon.$$

Now

$$\mu^*(A) \geq \mu^*(B_n \cup C) = \mu^*(B_n) + \mu^*(C)$$

since φ separates B_n and C. Consequently,

$$\mu^*(A) \geq \mu^*(B) + \mu^*(C) - \epsilon.$$

We have established the desired inequality (19). □

Let (X, ρ) be a metric space. Recall that for two subsets A and B of X, we define the distance between A and B, which we denote by $\rho(A, B)$, by

$$\rho(A, B) = \inf_{u \in A, v \in B} \rho(u, v).$$

By the Borel σ-algebra associated with this metric space, which we denote by $\mathcal{B}(X)$, we mean the smallest σ-algebra that contains all open sets induced by the metric.

Definition *Let (X, ρ) be a metric space. An outer-measure $\mu^*\colon 2^X \to [0, \infty]$ is called a* **Carathéodory outer-measure** *provided that whenever A and B are subsets of X for which $\rho(A, B) > 0$, then*

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

Theorem 27 *Let μ^* be a Carathéodory outer-measure on a metric space (X, ρ). Then every Borel subset of X is measurable with respect to μ^*.*

Proof The collection of Borel sets is the smallest σ-algebra containing the closed sets, and the measurable sets are a σ-algebra. Therefore, it suffices to show that each closed set is measurable. However, each closed subset F of X can be expressed as $F = f^{-1}(0)$, where f is the continuous function on X defined by $f(x) = \rho(F, \{x\})$. It therefore suffices to show that every continuous function is measurable. To do so, we apply Proposition 26. Indeed, let A and B be subsets of X for which there is a continuous function on X and real numbers $a < b$ such that $f \leq a$ on A and $f \geq b$ on B. By the continuity of $f, \rho(A, B) > 0$. Hence, by assumption, $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$. According to Proposition 26, each continuous function is measurable. The proof is complete. $\square$

We now turn our attention to a particular family of Carathéodory outer-measures on the metric space (X, ρ). First, recall that we define the diameter of a subset A of X, $\operatorname{diam}(A)$, by

$$\operatorname{diam}(A) = \sup_{u,v \in A} \rho(u, v).$$

For each positive real number α, we define a measure H_α on the Borel σ-algebra $\mathcal{B}(X)$ called the Hausdorff measure on X of dimension α. These measures are particularly important for the Euclidean spaces $\mathbf{R}^n$, in which case they provide a gradation of size among sets of n-dimensional Lebesgue measure zero.

Fix $\alpha > 0$. Take $\epsilon > 0$ and for a subset E of X, define

$$H_\alpha^{(\epsilon)}(E) = \inf \sum_{k=1}^{\infty} [\operatorname{diam}(A_k)]^\alpha,$$

where $\{A_k\}_{k=1}^\infty$ is a countable collection of subsets of X that covers E and each A_k has a diameter less than ϵ. Observe that $H_\alpha^{(\epsilon)}$ increases as ϵ decreases. Define

$$H_\alpha^*(E) = \sup_{\epsilon > 0} H_\alpha^{(\epsilon)}(E) = \lim_{\epsilon \to 0} H_\alpha^{(\epsilon)}(E).$$

Proposition 28 *Let (X, ρ) be a metric space and α a positive real number. Then the set-function $H_\alpha^* \colon 2^X \to [0, \infty]$ is a Carathéodory outer-measure.*

Proof It is readily verified that H_α^* is a countably monotone set-function on 2^X and $H_\alpha^*(\emptyset) = 0$. Therefore, H_α^* is an outer-measure on 2^X. We claim it is a Carathéodory outer-measure. Indeed, let E and F be two subsets of X for which $\rho(E, F) > \delta$. Then

$$H_\alpha^{(\epsilon)}(E \cup F) \geq H_\alpha^{(\epsilon)}(E) + H_\alpha^{(\epsilon)}(F)$$

as soon as $\epsilon < \delta$: For if $\{A_k\}$ is a countable collection of sets, each of diameter at most ϵ, that covers $E \cup F$, no A_k can have non-empty intersection with both E and F. Taking limits as $\epsilon \to 0$, we have

$$H_\alpha^*(E \cup F) \geq H_\alpha^*(E) + H_\alpha^*(F). \qquad \square$$

It follows from Theorem 27 that H_α^* induces a measure on a σ-algebra that contains the Borel subsets of X. We denote the restriction of this measure to $\mathcal{B}(X)$ by H_α and call it **Hausdorff α-dimensional measure** on the metric space X.

Proposition 29 *Let (X, ρ) be a metric space, A be a Borel subset of X, and $0 < \alpha < \beta$. If $H_\alpha(A) < \infty$, then $H_\beta(A) = 0$.*

Proof Let $\epsilon > 0$. Choose $\{A_k\}_{k=1}^\infty$ to be a covering of A by sets of diameter less than ϵ for which

$$\sum_{k=1}^\infty [\operatorname{diam}(A_k)]^\alpha \leq H_\alpha(A) + 1.$$

Then

$$H_\beta^{(\epsilon)}(A) \leq \sum_{k=1}^\infty [\operatorname{diam}(A_k)^\beta] \leq \epsilon^{\beta-\alpha} \cdot \sum_{k=1}^\infty [\operatorname{diam}(A_k)]^\alpha \leq \epsilon^{\beta-\alpha} \cdot [H_\alpha(A) + 1].$$

Take the limit as $\epsilon \to 0$ to conclude that $H_\beta(A) = 0$. $\square$

For a Borel subset E of $\mathbf{R}^n$, we define the Hausdorff dimension of E, $\dim_H(E)$, by

$$\dim_H(E) = \inf \{\beta \geq 0 \mid H_\beta(E) = 0\}.$$

Hausdorff measures are particularly significant for Euclidean space $\mathbf{R}^n$. In the case $n = 1$, H_1 equals Lebesgue measure. To see this, let $I \subseteq \mathbf{R}$ be an interval. Given $\epsilon > 0$, the interval I may be expressed as the disjoint union of subintervals of length less than ϵ and the diameter of each subinterval is its length. Thus, H_1 and Lebesgue measure agree on the semi-ring of intervals of real numbers. Therefore, by the construction of these measures from outer-measures, these measures also agree on the Borel sets. Thus, $H_1^{(\epsilon)}(E)$ is the Lebesgue outer-measure of E. For $n > 1$, H_n is not equal to Lebesgue measure, but it can be shown that it is a constant multiple of n-dimensional Lebesgue measure. It follows from the above proposition that if A is a Borel subset of $\mathbf{R}^n$ that has positive Lebesgue measure, then $\dim_H(A) = n$. There are many specific calculations of Hausdorff dimension of subsets of Euclidean space. For instance, it can be shown that the Hausdorff dimension of the Cantor set is $\log 2 / \log 3$[3].

PROBLEMS

30. Directly from the definition of Lebesgue outer-measure, show that if A and B are subsets of $\mathbf{R}^n$ and there is a $\delta > 0$ for which $\|u - v\| \geq \delta$ for all $u \in A$ and $v \in B$, then
$$\mu_n^*(A \cup B) = \mu_n^*(A) + \mu_n^*(B).$$
31. Show that in the definition of Hausdorff measure one can take the coverings to be by open sets or by closed sets.
32. Show that the set-function outer Hausdorff measure H_α^* is countably monotone.
33. In the plane $\mathbf{R}^2$ show that a bounded set may be enclosed in a ball of the same diameter. Use this to show that for a bounded subset A of $\mathbf{R}^2$, $H_2(A) \geq 4/\pi \cdot \mu_2(A)$, where μ_2 is Lebesgue measure on $\mathbf{R}^2$.

[3]Further results on Hausdorff measure, including specific calculations of Hausdorff dimensions, may be found in *Measure Theory and Finer Properties of Functions* by L.C. Evans and R.G. Gariepy.

34. Let (X, ρ) be a metric space and $\alpha > 0$. For $E \subseteq X$, define

$$H'_\alpha(E) = \inf \sum_{k=1}^{\infty} [\text{diam}(A_k)]^\alpha,$$

where $\{A_k\}_{k=1}^\infty$ is a countable collection of subsets of X that covers E: there is no restriction regarding the size of the diameters of the sets in the cover. Compare the set-functions H'_α and H_α.

35. Show that each Hausdorff measure H_α on Euclidean space $\mathbf{R}^n$ is invariant with respect to constant perturbations of orthogonal operators.

36. Give a direct proof to show that if I is a nontrivial interval in $\mathbf{R}^n$, then $H_n(I) > 0$.

37. Show that in any metric space, H_0 is counting measure.

38. Let $[a, b]$ be a closed, bounded interval on real numbers and consider the subset of the plane defined by $\mathbf{R} = \{(x, y) \in \mathbf{R}^2 \mid a \leq x \leq b, y = 0\}$. Show that $H_2(R) = 0$. Then show that $H_1(R) = b - a$. Conclude that the Hausdorff dimension of $\mathbf{R}$ is 1.

39. Let f: $[a, b] \to \mathbf{R}$ be a continuous, bounded function on the closed, bounded interval $[a, b]$ that has a continuous, bounded derivative on the open interval (a, b). Consider the graph G of f as a subset of the plane. Show that $H_1(G) = \int_a^b \sqrt{1 + |f'(x)|^2}\, dx$.

40. Let J be an interval in $\mathbf{R}^n$, each of whose sides has length 1. Define $\gamma_n = H_n(J)$. Show that if I is any bounded interval in $\mathbf{R}^n$, then $H_n(I) = \gamma_n \cdot \mu_n(I)$. From this conclude, using the uniqueness assertion of the Carathéodory-Hahn Theorem, that $H_n = \gamma_n \cdot \mu_n$ on the Borel subsets of $\mathbf{R}^n$.

10.5 SIGNED MEASURES: THE HAHN AND JORDAN DECOMPOSITIONS

Definition *A set-function ν: $\mathcal{M} \to \overline{\mathbf{R}}$ on a σ-algebra $\mathcal{M}$ is said to be a* **signed measure** *provided that it possesses the following two properties:*

(i) $\nu(\emptyset) = 0$ and ν assumes that at most one of the values $+\infty, -\infty$.

(ii) For any countable, disjoint collection $\{E_k\}_{k=1}^\infty$ of sets in $\mathcal{M}$,

$$\nu\Big(\bigcup_{k=1}^{\infty} E_k\Big) = \sum_{k=1}^{\infty} \nu(E_k), \tag{20}$$

and the series $\sum_{k=1}^\infty \nu(E_k)$ converges absolutely if $\nu(\bigcup_{k=1}^\infty E_k)$ is finite[4].

[4]In order that countable additivity be properly defined, it is necessary that if $E = \bigcup_{k=1}^\infty E_k$ and p: $\mathbf{N} \to \mathbf{N}$ is a permutation then, since $E = \bigcup_{k=1}^\infty E_{p(k)}$,

$$\sum_{k=1}^{\infty} \nu(E_{p(k)}) = \nu(E) = \sum_{k=1}^{\infty} \nu(E_k).$$

According to the Riemann Rearrangement Theorem (on page 24) if $\sum_{k=1}^\infty \nu(E_k)$ converges, this property holds if and only if the series $\sum_{k=1}^\infty \nu(E_k)$ is absolutely convergent.

Example Consider $(\mathbf{R}, \mathcal{L}, m)$, Lebesgue measure on $\mathbf{R}$, and let $f\colon \mathbf{R} \to \mathbf{R}$ be Lebesgue integrable. Define $\nu\colon \mathcal{L} \to \mathbf{R}$ by

$$\nu(E) = \int_E f\, dm \text{ for all } E \in \mathcal{L}.$$

Then ν is a signed measure. Indeed, to verify this, first let f^+ and f^- be the positive and negative parts of f, and define the set-functions ν^+ and ν^- on $\mathcal{L}$ by

$$\nu^+(E) = \int_E f^+\, dm \text{ and } \nu^-(E) = \int_E f^-\, dm \text{ for all } E \in \mathcal{L}.$$

By the countable additivity of Lebesgue integration, $\nu^+\colon \mathcal{L} \to \mathbf{R}$ and $\nu^-\colon \mathcal{L} \to \mathbf{R}$ are finite measures. Let $\{E_k\}_{k=1}^\infty$ be a measurable partition of E. But $f = f^+ - f^-$ and therefore

$$\nu = \nu^+ - \nu^-,$$

the difference of finite measures, so that $\nu(E) = \sum_{k=1}^\infty \nu(E_k)$. Moreover, since $|f| = f^+ + f^-$, $|\nu| = \nu^+ + \nu^-$, the sum of two measures, so that the series $\sum_{k=1}^\infty \nu(E_k)$ converges absolutely.

The main goal of this section is to show that, as in this example, a general signed measure ν is the difference of measures, which is called a Jordan Decomposition of ν. For such a signed measure ν, a set A is said to be **positive** (with respect to ν) provided that A is measurable and for every measurable subset E of A, $\nu(E) \geq 0$. The restriction of ν to the measurable subsets of a positive set is a measure. Similarly, a set B is called **negative** (with respect to ν) provided that it is measurable and every measurable subset of B has non-positive ν measure. The restriction of $-\nu$ to the measurable subsets of a negative set also is a measure.

Hahn's Lemma *Let $\nu\colon \mathcal{M} \to \overline{\mathbf{R}}$ be a signed measure, $E \in \mathcal{M}$ and $0 < \nu(E) < \infty$. Then there is a positive subset E_+ of E for which $\nu(E_+) > 0$.*

Proof If E itself is a positive set, then the proof is complete. Otherwise, E has subsets of negative measure. Let m_1 be the smallest natural number for which there is a measurable set of measure less than $-1/m_1$. Choose a measurable set $E_1 \subseteq E$ with $\nu(E_1) < -1/m_1$. Let n be a natural number for which natural numbers $m_1, \ldots, m_n$ and measurable sets $E_1, \ldots, E_n$ have been chosen such that for $1 \leq k \leq n$, m_k is the smallest natural number for which there is a measurable subset of $E \sim \bigcup_{j=1}^{k-1} E_j$ of measure less than $-1/m_k$ and E_k is a subset of $[E \sim \bigcup_{j=1}^{k-1} E_j]$ for which $\nu(E_k) < -1/m_k$. If this selection process terminates, then the proof is complete. Otherwise, define

$$A = E \sim \bigcup_{k=1}^\infty E_k, \text{ so that } E = A \cup \left[\bigcup_{k=1}^\infty E_k\right] \text{ is a measurable partition of } E.$$

Since a signed measure ν does not take the values ∞ and $-\infty$, for B and D measurable sets,

$$\text{if } B \subseteq D \text{ and } |\nu(D)| < \infty, \text{ then } |\nu(B)| < \infty.$$

Since $\bigcup_{k=1}^{\infty} E_k$ is a measurable subset of E and $|\nu(E)| < \infty$ and

$$-\infty < \nu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \nu(E_k) \leq \sum_{k=1}^{\infty} -1/m_k.$$

Consequently, $\lim_{k\to\infty} m_k = \infty$. We claim that A is a positive set. Indeed, if B is a measurable subset of A, then, for each k,

$$B \subseteq A \subseteq E \sim \left[\bigcup_{j=1}^{k-1} E_j\right],$$

and so, by the minimal choice of m_k, $\nu(B) \geq -1/(m_k - 1)$. Since $\lim_{k\to\infty} m_k = \infty$, $\nu(B) \geq 0$. So A is a positive set. It remains only to show that $\nu(A) > 0$. But this follows from the finite additivity of ν since $\nu(E) > 0$ and $\nu(E \sim A) = \nu(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \nu(E_k) < 0$. $\square$

The Hahn Decomposition Theorem *Let $\nu\colon \mathcal{M} \to \overline{\mathbf{R}}$ be a signed measure. Then there is a positive set A for ν and a negative set B for ν for which*

$$X = A \cup B \text{ and } A \cap B = \emptyset.$$

Proof Assume that $+\infty$ is the infinite value omitted by ν. Let $\mathcal{P}$ be the collection of positive subsets of X and define $\lambda = \sup\{\nu(E) \mid E \in \mathcal{P}\}$. Then $\lambda \geq 0$ since $\mathcal{P}$ contains the empty-set. Let $\{A_k\}_{k=1}^{\infty}$ be a countable collection of positive sets for which $\lambda = \lim_{k\to\infty} \nu(A_k)$. Define $A = \bigcup_{k=1}^{\infty} A_k$. The set A is itself a positive set, so that $\lambda \geq \nu(A)$. On the other hand, for each k, $A \sim A_k \subseteq A$, and so $\nu(A \sim A_k) \geq 0$. Therefore,

$$\nu(A) = \nu(A_k) + \nu(A \sim A_k) \geq \nu(A_k).$$

So $\nu(A) \geq \lambda$. Consequently, $\nu(A) = \lambda$ and $\lambda < \infty$ since ν does not take the value ∞.

Let $B = X \sim A$. We argue by contradiction to show that B is negative. Assume otherwise. Then there is a subset E of B with positive measure and therefore, by Hahn's Lemma, a subset E_0 of B that is both positive and of positive measure. Then $A \cup E_0$ is a positive set and

$$\nu(A \cup E_0) = \nu(A) + \nu(E_0) > \lambda,$$

a contradiction to the choice of λ. $\square$

The decomposition of X described in this theorem is called a **Hahn Decomposition** for ν. This decomposition may not be unique. If $\{A, B\}$ is a Hahn decomposition for ν, two measures ν^+ and ν^- with $\nu = \nu^+ - \nu^-$ are defined by setting

$$\nu^+(E) = \nu(E \cap A) \text{ and } \nu^-(E) = -\nu(E \cap B).$$

Two measures ν_1 and ν_2 on $\mathcal{M}$ are said to be **mutually singular** if there are disjoint measurable sets A and B with $X = A \cup B$ for which $\nu_1(A) = \nu_2(B) = 0$. The measures ν^+ and ν^- defined above are mutually singular. The existence part of the following theorem has been established, while the uniqueness part is left as an exercise.

The Jordan Decomposition Theorem *Let $\nu\colon \mathcal{M} \to \overline{\mathbf{R}}$ be a signed measure. There are two mutually singular measures ν^+ and ν^- on $\mathcal{M}$ for which $\nu = \nu^+ - \nu^-$. Moreover, there is only one such pair of mutually singular measures.*

The decomposition of a signed measure $\nu\colon \mathcal{M} \to \overline{\mathbf{R}}$ given by this theorem is called the **Jordan Decomposition** of ν. The measures ν^+ and ν^- are called the positive and negative parts (or variations) of ν. Since ν assumes at most one of the values $+\infty$ and $-\infty$, either ν^+ or ν^- must be finite. If they are both finite, ν is called a finite signed measure. The measure $|\nu|$ is defined on $\mathcal{M}$ by

$$|\nu|(E) = \nu^+(E) + \nu^-(E) \text{ for all } E \in \mathcal{M}.$$

It is left as an exercise to show that

$$|\nu|(X) = \sup \sum_{k=1}^{n} |\nu(E_k)|,$$

where the supremum is taken over all finite, disjoint collections $\{E_k\}_{k=1}^n$ of measurable subsets of X. For this reason, $|\nu|(X)$ is called the **total variation** of ν and denoted by $\|\nu\|_{var}$.

PROBLEMS

41. Show that the difference of two measures, one of which is finite, is a signed measure.
42. Show that the Hahn Decomposition may not be unique.
43. Let μ be a measure and μ_1 and μ_2 be mutually singular measures on a σ-algebra $\mathcal{M}$ for which $\mu = \mu_1 - \mu_2$. Show that $\mu_2 = 0$. Use this to establish the uniqueness of the Jordan Decomposition.
44. Show that if ν_1 and ν_2 are any two finite signed measures, then for any α and β, so is $\alpha\nu_1 + \beta\nu_2$. Show that

$$|\alpha\nu| = |\alpha||\nu| \text{ and } |\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|,$$

where $\nu \leq \mu$ means $\nu(E) \leq \mu(E)$ for all measurable sets E.

45. Let μ and ν be finite signed measures. Define $\mu \wedge \nu = \frac{1}{2}(\mu + \nu - |\mu - \nu|)$ and $\mu \vee \nu = \mu + \nu - \mu \wedge \nu$.
    (i) Show that the signed measure $\mu \wedge \nu$ is smaller than μ and ν but larger than any other signed measure that is smaller than μ and ν.
    (ii) Show that the signed measure $\mu \vee \nu$ is larger than μ and ν but smaller than any other measure that is larger than μ and ν.
    (iii) If μ and ν are positive measures, show that they are mutually singular if and only if $\mu \wedge \nu = 0$.

CHAPTER 11

Integration over General Measure Spaces

Contents

11.1 Measurable Functions: The Egoroff and Lusin Theorems and Sequential Approximation 201
11.2 Integration of Non-negative Measurable Functions: Fatou's Lemma, the Monotone Convergence Theorem, and Beppo Levi's Theorem 207
11.3 Integration of General Measurable Functions: The Dominated Convergence Theorem and the Vitali Convergence Theorem 213
11.4 The Radon-Nikodym Theorem 219
11.5 Product Measures: The Tonelli and Fubini Theorems 225
11.6 Products of Lebesgue Measure on Euclidean Spaces: Cavalieri's Principle . . 233

The first three sections provide variations of results seen in the first part for Lebesgue measure and integration on $\mathbf{R}$. Much of this is the same, but differences occur; some due to a general measure failing to be being complete, or failing to be σ-finite, or failing to have sets of arbitrarily small, but positive, measure. The integral of a non-negative, measurable function is directly defined to be the supremum of the integrals of the non-negative, simple functions that it dominates. In the last three sections, new ideas are considered. In Section 4, the concept of absolute continuity of one measure with respect to another is introduced and the Radon-Nikodym Theorem, a far-reaching generalization of the representation of an absolutely continuous function of a real variable as an indefinite integral, is proven. In the following section, product measures are defined as a Carathéodory extensions and two criteria for verifying iterated integration, Tonelli's Theorem and Fubini's Theorem, are proven, based on what we call the Fubini-Tonelli Lemma, a very special case of both theorems. Finally, the product structure of Lebesgue measure on Euclidean space is established, and we obtain the classic Cavalieri's Principle.

11.1 MEASURABLE FUNCTIONS: THE EGOROFF AND LUSIN THEOREMS AND SEQUENTIAL APPROXIMATION

Recall that for any mapping $f\colon X \to Y$, if $\{Y_\lambda\}_{\lambda\in\Lambda}$ is a collection of subsets of Y, parametrized by a space Λ, we have the following properties of inverses:

$$f^{-1}\left(\bigcup_{\lambda\in\Lambda} Y_\lambda\right) = \bigcup_{\lambda\in\Lambda} f^{-1}(Y_\lambda) \text{ and } f^{-1}\left(\bigcap_{\lambda\in\Lambda} Y_\lambda\right) = \bigcap_{\lambda\in\Lambda} f^{-1}(Y_\lambda),$$

and moreover, for another mapping $g\colon Y \to Z$,

$$(g\circ f)^{-1}(Z_0) = f^{-1}(g^{-1}(Z_0)) \textit{ for } Z_0 \subseteq Z.$$

Throughout this section, a standing assumption is that $(X, \mathcal{M}, \mu)$ *is a measure space,* and a set is called measurable provided that it belongs to $\mathcal{M}$. The general concept of a measurable function is identical with that for Lebesgue measure on $\mathbf{R}$.

Definition *An extended real-valued function $f\colon X \to \overline{\mathbf{R}}$ is said to be* **measurable** *provided that for each real number c, the set $\{x \in X \mid f(x) < c\}$ is measurable. Such a function is said to be* **finitely supported** *provided that $\mu\{x \in X \mid f(x) \neq 0\} < \infty$.*

For a set X and the σ-algebra $\mathcal{M} = 2^X$ of all subsets of X, every extended real-valued function on X is measurable with respect to $\mathcal{M}$. At the opposite extreme, consider the σ-algebra $\mathcal{M} = \{X, \emptyset\}$, with respect to which the only measurable functions are those that are constant. The proof of the following proposition is, as it was for functions of a real variable: an application of the properties of inverses, that each open subset of $\mathbf{R}$ is the countable union of open intervals, and that for continuous functions, the inverse image of open sets are open.

Proposition 1 *Each of the following two properties is equivalent to the measurability of a function $f\colon X \to \mathbf{R}$.*

(i) For every interval I of real numbers, $f^{-1}(I)$ is measurable.

(ii) For each open set $\mathcal{O}$ of real numbers, $f^{-1}(\mathcal{O})$ is measurable.

Moreover, if $f\colon X \to \mathbf{R}$ is measurable and $g\colon \mathbf{R} \to \mathbf{R}$ is continuous, then the composition $g \circ f\colon X \to \mathbf{R}$ is measurable.

Let $f\colon E \to \mathbf{R}$ and $g\colon E \to \mathbf{R}$ be measurable. Since $\mathbf{Q}$ is dense in $\mathbf{R}$, we have

$$(f+g)^{-1}(-\infty, c) = \bigcap_{q \in \mathbf{Q}} f^{-1}(-\infty, q) \cap g^{-1}(-\infty, q-c).$$

so that, since $\mathbf{Q}$ is countable, $f+g$ is measurable. Clearly, the square of a measurable function is measurable, and hence so is $f \cdot g$, since

$$f \cdot g = 1/4[(f+g)^2 - (f-g)^2].$$

Proposition 2 *If $f\colon X \to \mathbf{R}$ and $g\colon X \to \mathbf{R}$ are measurable functions, then, for any α and β,*

$$\alpha f + \beta g\colon X \to \mathbf{R} \text{ also is measurable, and so is}$$

$$f \cdot g\colon X \to \mathbf{R}.$$

As in the case of Lebesgue measure on $\mathbf{R}$, measurable functions on X that agree on the complement of a set of measure zero are identified. Accordingly, we identify a function that is finite almost everywhere with a real-valued function. In view of this, we often state results only for real-valued functions, although it is important, say in the Monotone Convergence and Tonelli Theorems, to consider extended real-valued functions. Just as in the special case of Lebesgue measure on $\mathbf{R}$, we will ubiquitously employ the preservation of measurability for functions under sequential pointwise limits.

Theorem 3 *If $\{f_n: X \to \overline{\mathbf{R}}\}$ is a sequence of measurable functions that converges pointwise on X to the function $f: X \to \overline{\mathbf{R}}$, then f is measurable.*

Proof Fix a real number c. It must be shown that the set $\{x \in X \mid f(x) < c\}$ is measurable. Observe that for $x \in X$, since $\lim_{n\to\infty} f_n(x) = f(x)$, $f(x) < c$ if and only if there are indices n and k such that for all $j \geq k$, $f_j(x) < c - 1/n$. But for two such n and j, since the function f_j is measurable, the set $\{x \in X \mid f_j(x) < c - 1/n\}$ is measurable. Since $\mathcal{M}$ is a σ-algebra, for any k,

$$\bigcap_{j=k}^{\infty} \{x \in X \mid \ f_j(x) < c - 1/n\}$$

also is measurable, and therefore so is

$$\{x \in X \mid \ f(x) < c\} = \bigcup_{1 \leq k, n < \infty} \left[\bigcap_{j=k}^{\infty} \{x \in X \mid \ f_j(x) < c - 1/n\} \right].$$

□

Definition *A real-valued function $\varphi: X \to \mathbf{R}$ is said to be **simple** provided that it is measurable and takes only a finite number of values.*

Clearly, a linear combination of characteristic functions of measurable sets is simple, and the converse holds. Indeed, if $\varphi: X \to \mathbf{R}$ is simple and takes the distinct values $c_1, \ldots, c_n$, then

$$\varphi = \sum_{k=1}^{n} c_k \cdot \chi_{X_k} \text{ on } X, \text{ where } X_k = \{x \in X \mid \ \varphi(x) = c_k\}.$$

This particular expression of φ as a linear combination of characteristic functions of measurable sets is called the **canonical representation of the simple function** φ; it is distinguished by the E_k's being disjoint and the c_k's being distinct.

The Simple Approximation Lemma *Let $f: X \to \mathbf{R}$ be a bounded, measurable function. Then for each $\epsilon > 0$, there are simple functions $\varphi_\epsilon: X \to \mathbf{R}$ and $\psi_\epsilon: X \to \mathbf{R}$ that have the following approximation properties:*

$$\varphi_\epsilon \leq f \leq \psi_\epsilon \textit{ and } 0 \leq \psi_\epsilon - \varphi_\epsilon < \epsilon \textit{ on } X.$$

Proof Let $[c, d)$ be a bounded interval that contains the image of X, $f(X)$, and

$$c = y_0 < y_1 < \ldots < y_{n-1} < y_n = d$$

be a partition of the closed, bounded interval $[c, d]$ such that $y_k - y_{k-1} < \epsilon$ for $1 \leq k \leq n$. Define

$$I_k = [y_{k-1}, y_k) \text{ and } X_k = f^{-1}(I_k) \text{ for } 1 \leq k \leq n.$$

Since each I_k is an interval and the function f is measurable, each set X_k is measurable. Define the simple functions φ_ϵ and ψ_ϵ on X by

$$\varphi_\epsilon = \sum_{k=1}^{n} y_{k-1} \cdot \chi_{X_k} \text{ and } \psi_\epsilon = \sum_{k=1}^{n} y_k \cdot \chi_{X_k}.$$

Let $x \in X$. Since $f(X) \subseteq [c, d)$, there is a unique $k, 1 \leq k \leq n$, for which $y_{k-1} \leq f(x) < y_k$ and therefore

$$\varphi_\epsilon(x) = y_{k-1} \leq f(x) < y_k = \psi_\epsilon(x).$$

But $y_k - y_{k-1} < \epsilon$, and therefore φ_ϵ and ψ_ϵ have the required approximation properties. □

The Simple Approximation Theorem *If $f\colon X \to \overline{\mathbf{R}}$ is a measurable function, then there is a sequence $\{\psi_n\colon X \to \mathbf{R}\}$ of simple functions that converge pointwise on X to f,*

$$|\psi_n| \leq |f| \text{ on } X \text{ for all } n, \text{ and}$$

(i) *if X is σ-finite, then, in addition, each ψ_n is finitely supported;*
(ii) *if $f \geq 0$, then, in addition, $\{\psi_n\}$ is increasing and each $0 \leq \psi_n \leq f$.*

Proof Fix n, and define $E_n = \{x \in X \,|\, |f(x)| \leq n\}$. By the Simple Approximation Lemma, applied to the bounded, measurable function $f\colon E_n \to \mathbf{R}$, there are simple functions $h_n\colon E_n \to \mathbf{R}$ and $g_n\colon E_n \to \mathbf{R}$ for which

$$h_n \leq f \leq g_n \text{ and } 0 \leq g_n - h_n < 1/n \text{ on } E_n.$$

For $x \in E_n$, define $\psi_n(x) = 0$ if $f(x) = 0$, $\psi_n(x) = \max\{h_n(x), 0\}$ if $f(x) > 0$ and $\psi_n(x) = \min\{g_n(x), 0\}$ if $f(x) < 0$. Extend ψ_n to all of X by setting $\psi_n(x) = n$ if $f(x) > n$ and $\psi_n(x) = -n$ if $f(x) < -n$. This defines a sequence $\{\psi_n\}$ of simple functions for which $|\psi_n| \leq |f|$ on X for all n, and the sequence $\{\psi_n\}$ converges pointwise on X to f. If X is σ-finite, express X as the union of a countable, ascending collection $\{X_n\}_{n=1}^\infty$ of measurable subsets, each of which has finite measure. Replace each ψ_n by $\psi_n \cdot \chi_{X_n}$ and (i) is verified. If $f \geq 0$, replace each ψ_n by $\max_{1 \leq i \leq n} |\psi_i|$ and (ii) is verified. □

Egoroff's Theorem for a general measure space is further confirmation of Littlewood's second principle, namely, pointwise convergence is "almost" uniform.

Egoroff's Theorem *Assume that $\mu(X) < \infty$. If $\{f_n\colon X \to \mathbf{R}\}$ is a sequence of measurable functions that converges pointwise on X to the function $f\colon X \to \mathbf{R}$, then for each $\epsilon > 0$, there is a measurable subset X_0 of X for which*

$$\mu(X \sim X_0) < \epsilon \text{ and } \{f_n\} \to f \text{ uniformly on } X_0.$$

We base the proof of this theorem on the following lemma.

Lemma 4 *Under the assumptions of Egoroff's Theorem, for each $\eta > 0$ and $\delta > 0$, there is a measurable set $E \subseteq X$ and an index N for which*

$$\mu(X \sim E) < \delta \text{ and } |f_n - f| < \eta \text{ on } E \text{ for all } n \geq N.$$

Proof For each k, the set $\{x \in X \,|\, |f(x) - f_k(x)| < \eta\}$ is measurable, and therefore so is

$$E_n = \{x \in X \mid |f(x) - f_k(x)| < \eta \text{ for all } k \geq n\}.$$

Then $\{E_n\}_{n=1}^{\infty}$ is an ascending collection of measurable sets, and $X = \cup_{n=1}^{\infty} E_n$, since $\{f_n\}$ converges pointwise to f on X. It follows from the continuity of measure that

$$\mu(X) = \lim_{n \to \infty} \mu(E_n).$$

Since $\mu(X) < \infty$, there is an index N for which $\mu(E_N) > \mu(X) - \delta$. Define $E = E_N$ and observe that, by the excision property of measure, $\mu(X \sim E) = \mu(X) - \mu(E_N) < \delta$. $\square$

Proof of Egoroff's Theorem It follows from the preceding lemma that for each n, there is a set $E_n \in \mathcal{M}$ and an index $N(n)$ for which

$$\mu(X \sim E_n) < \epsilon/2^n \tag{1}$$

and

$$|f_k - f| < 1/n \text{ on } E_n \text{ for } k \geq N(n). \tag{2}$$

Define

$$X_0 = \bigcap_{n=1}^{\infty} E_n.$$

By De Morgan's Identities, the countably monotonicity of μ and (1),

$$\mu(X \sim X_0) = \mu\left(\bigcup_{n=1}^{\infty} [X \sim E_n]\right) \leq \sum_{n=1}^{\infty} \mu(X \sim E_n) < \sum_{n=1}^{\infty} \epsilon/2^n = \epsilon.$$

To see that $\{f_n\} \to f$ uniformly on X_0, let $\epsilon > 0$. Choose an index n_0 for which $1/n_0 < \epsilon$. Then, by (2),

$$|f_k - f| < 1/n_0 \text{ on } E_{n_0} \text{ for } k \geq N(n_0).$$

However, $X_0 \subseteq E_{n_0}$ and $1/n_0 < \epsilon$ and so

$$|f_k - f| < \epsilon \text{ on } X_0 \text{ for } k \geq N(n_0).$$

Consequently, $\{f_n\} \to f$ uniformly on X_0 and $\mu(X \sim X_0) < \epsilon$. $\square$

In order to consider a general form of Littlewood's third principle, namely, that a measurable function is "almost" continuous, it is necessary to impose further structure on X. We extend Lusin's Theorem to the Lebesgue measure space $(\mathbf{R}^n, \mathcal{L}^n, \mu_n)$, which was constructed in the preceding chapter, only using the regularity of this space. In Chapter 3, we proved the following lemma in the case that $n = 1$. This lemma for Euclidean space is a consequence of the Tietze Extension Theorem, which we prove in Chapter 16.

Lemma 5 *If F is a closed subset of $\mathbf{R}^n$ and the function f: $F \to \mathbf{R}$ is continuous, then it has a continuous extension to f: $\mathbf{R}^n \to \mathbf{R}$.*

Lusin's Theorem *If $E \subseteq \mathbf{R}^n$ and the function f: $E \to \mathbf{R}$ is Lebesgue measurable, then for each $\epsilon > 0$, there is a continuous function g: $\mathbf{R}^n \to \mathbf{R}$ and a closed subset F of $\mathbf{R}^n$ for which*

$$F \subseteq E,\ \mu_n(E \sim F) < \epsilon \text{ and } f = g \text{ on } F.$$

Proof[1] Choose an enumeration $\{I_k\}_{k=1}^{\infty}$ of the countable collection of all open intervals in $\mathbf{R}$ that have rational end-points. Let $\epsilon > 0$. By the regularity of Lebesgue measure μ_n, for each k, since $f^{-1}(I_k)$ is measurable, there are closed subsets of $\mathbf{R}^n$, F_k and H_k, for which

$$F_k \subseteq f^{-1}(I_k), \;\; H_k \subseteq E \sim f^{-1}(I_k),$$

$$\mu_n(f^{-1}(I_k) \sim F_k) < \epsilon/2^{k+1} \;\text{ and }\; \mu_n([E \sim f^{-1}(I_k)] \sim H_k) < \epsilon/2^{k+1}.$$

Define $F \equiv \cap_{k=1}^{\infty}(F_k \cup H_k)$. Being the intersection of closed subsets of $\mathbf{R}^n$, F is closed and, by De Morgan's Identities and the countable monotonicity of μ_n,

$$\mu_n(E \sim F) = \mu_n(\cup_{k=1}^{\infty}(E \sim (F_k \cup H_k))) < 2 \cdot \sum_{k=1}^{\infty} \epsilon/2^{k+1} = \epsilon.$$

Assume that $F \neq \emptyset$, for otherwise $\mu_n(E) < \epsilon$ and choose the closed subset F to be any point in E and take g to be the constant function on $\mathbf{R}^n$ that agrees with f at that point. To verify that $f\colon F \to R$ is continuous, it is necessary to show that for each $x \in F$ and open interval I containing $f(x)$, there is an open subset $\mathcal{O}$ of $\mathbf{R}^n$ for which

$$x \in \mathcal{O} \text{ and } f(\mathcal{O} \cap F) \subseteq I.$$

For such an x and I, since the rational numbers are dense in $\mathbf{R}$, there is a k for which $f(x) \in I_k \subseteq I$. We claim the above inclusion holds for the open set $\mathcal{O} = \mathbf{R}^n \sim H_k$. Indeed, observe that $f(F_k) \subseteq I_k \subseteq I$, and since H_k and F_k are disjoint,

$$\mathcal{O} \cap F \subseteq (\mathbf{R}^n \sim H_k) \cap (F_k \cup H_k) = F_k.$$

Therefore, the function $f\colon F \to \mathbf{R}$ is continuous and, by the preceding lemma, it has a continuous extension to all of $\mathbf{R}^n$. □

Corollary 6 *If $E \subseteq \mathbf{R}^n$ and the function $f\colon E \to \mathbf{R}$ is Lebesgue measurable, then there is a sequence of continuous functions $\{f_k\colon \mathbf{R}^n \to \mathbf{R}\}$ that converges pointwise almost everywhere on E to f.*

Proof According to Lusin's Theorem, for each k, there is a continuous function $f_k\colon \mathbf{R}^n \to \mathbf{R}$ for which

$$\mu_n \{x \in E \mid \; f_k(x) \neq f(x)\} < 1/2^k.$$

It follows from the Borel-Cantelli Lemma that for almost all $x \in E$, there is a k such that $f_j(x) = f(x)$ for all $j \geq k$. □

PROBLEMS

1. Let $E \in \mathcal{M}$. Show that a function $f\colon X \to \overline{\mathbf{R}}$ is measurable if and only if its restrictions to E and $X \sim E$ are measurable.
2. Assume that $(X, \mathcal{M}, \mu)$ is not complete. Show that there are functions $f\colon X \to \mathbf{R}$ and $g\colon X \to \mathbf{R}$ for which $f = g$ almost everywhere, and f is measurable but g is not.

[1]This elegant proof is due to Peter Loeb and Erik Talvila.

3. Assume that the function $f: X \to \mathbf{R}$ is measurable. Show that $f^{-1}(E)$ is measurable if $E \subseteq R$ is a Borel set. Is $f^{-1}(E)$ measurable for every Lebesgue measurable set $E \subseteq R$?
4. Show that the function $f: X \to \overline{\mathbf{R}}$ is measurable if for each rational number $c, \{x \in X \mid f(x) < c\}$ is a measurable set.
5. (i) Show that every function $f: X \to \overline{\mathbf{R}}$ is measurable with respect to the σ-algebra $\mathcal{M} = 2^X$.

   (ii) Let $x_0 \in X$ and δ_{x_0} be the Dirac measure at x_0 on 2^X. Show that two functions on X are equal almost everywhere $[\delta_{x_0}]$ if and only if they take the same value at x_0.

   (iii) Let c be the counting measure on 2^X. Show that two functions on X are equal almost everywhere $[c]$ if and only if they take the same value at every point in X.
6. If $f: \mathbf{R} \to \mathbf{R}$ is measurable with respect to the σ-algebra of Lebesgue measurable sets, is it necessarily measurable with respect to the Borel σ-algebra?
7. Let $\{f_n: X \to \mathbf{R}\}$ be a sequence of measurable functions such that, for each n, $\mu\{x \in X \mid |f_n(x) - f_{n+1}(x)| > 1/2^n\} < 1/2^n$. Show that $\{f_n\}$ is pointwise convergent almost everywhere on X. (Suggestion: Use the Borel-Cantelli Lemma.)
8. Under the assumptions of Egoroff's Theorem, show that $X = \bigcup_{k=0}^{\infty} X_k$, where each X_k is measurable, $\mu(X_0) = 0$ and, for $k \geq 1$, $\{f_n\}$ converges uniformly to f on X_k.
9. A sequence $\{f_n: X \to \mathbf{R}\}$ of measurable functions is said to **converge in measure** to a measurable function $f: X \to \mathbf{R}$ provided that for each $\eta > 0$,
$$\lim_{n \to \infty} \mu \{x \in X \mid \ |f_n(x) - f(x)| > \eta\} = 0.$$
   (i) Show that if $\mu(X) < \infty$ and $\{f_n: X \to \mathbf{R}\}$ converges pointwise almost everywhere on X to a measurable function $f: X \to \mathbf{R}$, then $\{f_n\}$ converges to f in measure.

   (ii) Show that if $\{f_n: X \to \mathbf{R}\}$ converges to $f: X \to \mathbf{R}$ in measure, then there is a subsequence of $\{f_n\}$ that converges pointwise almost everywhere on X to f. (Suggestion: Use the Borel-Cantelli Lemma.)
10. Assume that $\mu(X) < \infty$. Show that $\{f_n: X \to \mathbf{R}\} \to f: X \to \mathbf{R}$ in measure if and only if each subsequence of $\{f_n\}$ has a further subsequence that converges pointwise almost everywhere on X to f. Use this to show that for two sequences that converge in measure, the product sequence also converges in measure to the product of the limits. (Suggestion: Use (ii) in the preceding problem.)

11.2 INTEGRATION OF NON-NEGATIVE MEASURABLE FUNCTIONS: FATOU'S LEMMA, THE MONOTONE CONVERGENCE THEOREM, AND BEPPO LEVI'S THEOREM

The standing assumption continues to be that $(X, \mathcal{M}, \mu)$ *is a measure space.*

Definition *Let $\psi: X \to [0, \infty)$ be a non-negative, simple function. Define the integral of ψ over X, $\int_X \psi\, d\mu$, as follows: if $\psi \equiv 0$ on X, define $\int_X \psi\, d\mu = 0$. Otherwise, let $c_1, c_2, ..., c_n$ be the distinct positive values taken by ψ and, for $1 \leq k \leq n$, let $E_k = \{x \in X \mid \psi(x) = c_k\}$. Define*
$$\int_X \psi\, d\mu = \sum_{k=1}^{n} c_k \cdot \mu(E_k),$$
using the convention that the right-hand side is ∞ if, for some k, $\mu(E_k) = \infty$.

Proposition 7 *If $\varphi\colon X \to [0, \infty)$ and $\psi\colon X \to [0, \infty)$ are simple functions and $\alpha, \beta > 0$, then*

$$\int_X [\alpha \cdot \psi + \beta \cdot \varphi]\, d\mu = \alpha \cdot \int_X \psi\, d\mu + \beta \cdot \int_X \psi\, d\mu.$$

Proof If either ψ or φ is positive on a set of infinite measure, then the linear combination $\alpha \cdot \psi + \beta \cdot \varphi$ has the same property and therefore each side of the integral equality is infinite. Assume that both ψ and φ are finitely supported, in which case, the proof is exactly the same as that for Lebesgue measure on **R**. It follows from the provision of a finite, measurable partition of X such that on each set in the partition, the two functions are constant, obtained by intersecting the level sets of the functions. □

Linearity of integration for non-negative, simple functions implies **monotonicity** and **additivity over disjoint domains**. Indeed, if $\psi \le \varphi$ on X, then

$$\int_X \psi\, d\mu \le \int_X \varphi\, d\mu.$$

This is a consequence of linearity, since

$$\int_X \varphi\, d\mu - \int_X \psi\, d\mu = \int_X [\varphi - \psi]\, d\mu \ge 0.$$

Also, if A and B are disjoint, measurable subsets of X, then

$$\int_{A \cup B} \psi\, d\mu = \int_A \psi\, d\mu + \int_B \psi\, d\mu.$$

This also a consequence of linearity since, A and B being disjoint,

$$\psi \cdot \chi_{A \cup B} = \psi \cdot \chi_A + \psi \cdot \chi_{AB}.$$

Definition *If the non-negative function $f\colon X \to [0, \infty]$ is measurable, then the* **integral** *of f over X with respect to μ is defined by*

$$\int_X f\ d\mu = \sup_{0 \le \varphi \le f} \int_X \varphi\ d\,\mu,$$

where the supremum is taken over simple functions.

Chebychev's Inequality *If the non-negative function $f\colon X \to [0, \infty]$ is measurable and $\lambda > 0$, then*

$$\mu\,\{x \in X \mid\ f(x) \ge \lambda\} \le \frac{1}{\lambda} \int_X f\, d\mu.$$

Proof Define $X_\lambda = \{x \in X \mid f(x) \ge \lambda\}$ and $\varphi = \lambda \cdot \chi_{X_\lambda}$. Observe that $0 \le \varphi \le f$ on X and φ is a simple function. Therefore, by definition,

$$\lambda \cdot \mu(X_\lambda) = \int_X \varphi\, d\mu \le \int_X f\, d\mu.$$

Divide this inequality by λ to obtain Chebychev's Inequality. □

Proposition 8 *If $f\colon X \to [0, \infty]$ is a non-negative, measurable function, and $\int_X f\,d\mu < \infty$, then f is finite almost everywhere on X and has σ-finite support.*

Proof Define $X_\infty = \{x \in X \mid f(x) = \infty\}$. For each n, consider the simple function $\psi_n = n \cdot \chi_{X_\infty}$. By definition, $\int_X \psi_n\,d\mu = n \cdot \mu(X_\infty)$, and since $0 \le \psi_n \le f$ on X, $n \cdot \mu(X_\infty) \le \int_X f\,d\mu < \infty$. Therefore, $\mu(X_\infty) = 0$. Now define $X_n = \{x \in X \mid f(x) \ge 1/n\}$. By Chebychev's Inequality,

$$\mu(X_n) \le n \cdot \int_X f\,d\mu < \infty.$$

Moreover,

$$\{x \in X \mid\ f(x) > 0\} = \bigcup_{n=1}^{\infty} X_n.$$

Consequently, the set $\{x \in X \mid f(x) > 0\}$ is σ-finite. □

Fatou's Lemma *If $\{f_n\colon X \to [0, \infty]\}$ is a sequence of non-negative, measurable functions that converges pointwise on X to the function f, then*

$$\int_X f\ d\mu \le \liminf \int_X f_n\ d\mu.$$

Proof According to Theorem 3, the function f is measurable, and, of course, is non-negative. By the definition of $\int_X f\ d\mu$ as a supremum, to verify the integral inequality it is necessary and sufficient to show that if φ is any simple function for which $0 \le \varphi \le f$ on X, then

$$\int_X \varphi\ d\mu \le \liminf \int_X f_n\ d\mu. \tag{3}$$

Let φ be such a function. This inequality clearly holds if $\int_X \varphi\,d\mu = 0$. Assume $\int_X \varphi\,d\mu > 0$.

Case 1: Assume that $\int_X \varphi\,d\mu = \infty$. Then there is a measurable set $X_\infty \subseteq X$ and $a > 0$ for which $\mu(X_\infty) = \infty$ and $\varphi = a$ on X_∞. For each n, define

$$A_n = \{x \in X \mid\ f_k(x) \ge a/2 \text{ for all } k \ge n\}.$$

Then $\{A_n\}_{n=1}^{\infty}$ is an ascending sequence of measurable subsets of X. Since $X_\infty \subseteq \bigcup_{n=1}^{\infty} A_n$, by the continuity and monotonicity of measure,

$$\lim_{n\to\infty} \mu(A_n) = \mu(\bigcup_{n=1}^{\infty} A_n) \ge \mu(X_\infty) = \infty.$$

However, by Chebychev's Inequality, for each n,

$$\mu(A_n) \le \frac{2}{a}\int_{A_n} f_n\,d\mu \le \frac{2}{a}\int_X f_n\,d\mu.$$

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

Case 2: Assume that $0 < \int_X \varphi\, d\mu < \infty$. By excising from X the set where φ takes the value 0, the left-hand side of (3) remains unchanged and the right-hand side does not increase. Therefore, assume that $\varphi > 0$ on X, and so $\mu(X) < \infty$ since $\int_X \varphi\, d\mu < \infty$. To verify (3), choose $\epsilon > 0$. For each n, define

$$X_n = \{x \in X \mid \ f_k(x) > (1-\epsilon)\,\varphi(x) \text{ for all } k \geq n\}.$$

Then $\{X_n\}$ is an ascending sequence of measurable subsets of X whose union is X. Therefore, $\{X \sim X_n\}$ is a descending sequence of measurable subsets of X, which has empty intersection. Since $\mu(X) < \infty$, by the continuity of measure, $\lim_{n\to\infty} \mu(X \sim X_n) = 0$. Choose an index N for which $\mu(X \sim X_n) < \epsilon$ for all $n \geq N$. Define $M > 0$ to be the maximum of the finite number of values taken by φ on X. It follows from the linearity of integration for non-negative, simple function and the finiteness of $\int_X \varphi\, d\mu$ that, for $n \geq N$,

$$\begin{aligned}
\int_X f_n\, d\mu \geq \int_{X_n} f_n\, d\mu \quad &\geq (1-\epsilon) \textstyle\int_{X_n} \varphi\, d\mu \\
&= (1-\epsilon)\int_X \varphi\, d\mu - (1-\epsilon)\int_{X \sim X_n} \varphi\, d\mu \\
&\geq (1-\epsilon)\int_X \varphi\, d\mu - \int_{X \sim X_n} \varphi\, d\mu \\
&\geq (1-\epsilon)\int_X \varphi\, d\mu - \epsilon \cdot M \cdot \mu(X) \\
&= \int_X \varphi\, d\mu - \epsilon\left[\int_X \varphi\, d\mu + M\mu(X)\right].
\end{aligned}$$

Consequently,

$$\liminf \int_X f_n\, d\mu \geq \int_X \varphi\, d\mu - \epsilon\left[\int_X \varphi\, d\mu + M\mu(X)\right].$$

This inequality holds for all $\epsilon > 0$ and so, since $\int_X \varphi\, d\mu + M\mu(X)$ is finite, it also holds for $\epsilon = 0$. Therefore, (3) is verified. □

For a sequence of non-negative, measurable functions that converges pointwise, there is the following criterion, the first for a general measure space, for justification of passage of the limit under the integral sign.

The Monotone Convergence Theorem *Let $\{f_n \colon X \to [0, \infty]\}$ be an increasing sequence of non-negative, measurable functions. Define $f(x) = \lim_{n\to\infty} f_n(x)$ for each $x \in X$. Then*

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

Proof According to Fatou's Lemma,

$$\int_X f\, d\mu \leq \liminf \int_X f_n\, d\mu.$$

However, for each n, $f_n \leq f$ on X, and so, by the monotonicity of integration, $\int_X f_n \, d\mu \leq \int_X f \, d\mu$. Therefore,

$$\limsup \int_X f_n \, d\mu \leq \int_X f \, d\mu.$$

Consequently,

$$\limsup \int_X f_n \, d\mu \leq \int_X f \, d\mu \leq \liminf \int_X f_n \, d\mu. \qquad \square$$

Beppo Levi's Theorem *Let $\{f_n \colon X \to [0, \infty]\}$ be an increasing sequence of non-negative, measurable functions. If the sequence of integrals $\{\int_X f_n \, d\mu\}$ is bounded, then $\{f_n\}$ converges pointwise on X to a non-negative, measurable function f that is finite almost everywhere on X, has σ-finite support, and*

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

Proof Define $f(x) = \lim_{n \to \infty} f_n(x)$ for each $x \in X$. The Monotone Convergence Theorem implies that $\left\{\int_X f_n \, d\mu\right\} \to \int_X f \, d\mu$. Therefore, since the sequence of real numbers $\left\{\int_X f_n \, d\mu\right\}$ is bounded, its limit is finite and so $\int_X f \, d\mu < \infty$. According to Proposition 8, f is finite almost everywhere on X and has σ-finite support. $\square$

Proposition 9 *If $f \colon X \to [0, \infty]$ is a non-negative, measurable function, then there is an increasing sequence $\{\psi_n \colon X \to [0, \infty]\}$ of simple functions that converges pointwise on X to f and*

$$\lim_{n \to \infty} \int_X \psi_n \, d\mu = \int_X f \, d\mu. \tag{4}$$

Proof This follows from conclusion (ii) of the Simple Approximation Theorem and the Monotone Convergence Theorem. $\square$

Proposition 10 *If $f \colon X \to [0, \infty]$ and $g \colon X \to [0, \infty]$ are non-negative, measurable functions, and $\alpha, \beta > 0$, then*

$$\int_X [\alpha \cdot f + \beta \cdot g] \, d\mu = \alpha \int_X f \, d\mu + \beta \int_X g \, d\mu. \tag{5}$$

Proof Consider the case that $\alpha = \beta = 1$, and leave the case of general coefficients as an exercise. According to the preceding proposition, there are increasing sequences $\{\psi_n\}$ and $\{\varphi_n\}$ of non-negative, simple functions on X that converge pointwise on X to g and f, respectively,

$$\lim_{n \to \infty} \int_X \psi_n \, d\mu = \int_X g \, d\mu \text{ and } \lim_{n \to \infty} \int_X \varphi_n \, d\mu = \int_X f \, d\mu.$$

Then $\{\varphi_n + \psi_n\}$ is an increasing sequence of non-negative, simple functions that converges pointwise on X to $f+g$. By the linearity of integration for non-negative, simple functions and the Monotone Convergence Theorem,

$$\begin{aligned}\int_X [f+g]\,d\mu &= \lim_{n\to\infty}\int_X [\varphi_n + \psi_n]\,d\mu \\ &= \lim_{n\to\infty}\left[\int_X \varphi_n\,d\mu + \int_X \psi_n\,d\mu\right] \\ &= \lim_{n\to\infty}\int_X \varphi_n\,d\mu + \lim_{n\to\infty}\int_X \psi_n\,d\mu \\ &= \int_X f\,d\mu + \int_X g\,d\mu. \end{aligned}$$

□

As they did in the case of non-negative, simple functions, the monotonicity and additivity over domains properties for integration of non-negative, measurable functions are a consequence of linearity.

Definition *A non-negative, measurable function $f\colon X \to [0,\infty]$ is said to be* **integrable** *provided that $\int_X f\,d\mu < \infty$.*

According to the preceding proposition, the sum of non-negative, integrable functions is integrable, while, by Proposition 8, a non-negative, integrable function is finite almost everywhere and has σ-finite support.

PROBLEMS

11. Let $\{u_n\colon X \to [0,\infty]\}$ be a sequence of measurable functions. For $x \in X$, define $f(x) = \sum_{n=1}^{\infty} u_n(x)$. Show that
$$\int_X f\,d\mu = \sum_{n=1}^{\infty}\left[\int_X u_n\,d\mu\right].$$

12. Show that if $f\colon X \to [0,\infty]$ is measurable, then
$$\int_X f\,d\mu = 0 \text{ if and only if } f = 0 \text{ almost everywhere on } X.$$

13. Let $g\colon X \to [0,\infty]$ be integrable. Define
$$\nu(E) = \int_E g\,d\mu \text{ for all } E \in \mathcal{M}.$$

    (i) Show that ν is a measure on $\mathcal{M}$.

(ii) Let $f\colon X \to [0, \infty]$ be measurable. Show that

$$\int_X f\, d\nu = \int_X fg\, d\mu.$$

(Suggestion: First establish this for the case that f is simple and then use the Simple Approximation Theorem and the Monotone Convergence Theorem.)

14. Let $f\colon X \to [0, \infty]$ and $g\colon X \to [0, \infty]$ be measurable functions for which f^2 and g^2 are integrable. Show that $f \cdot g$ also is integrable. (Suggestion: $(a-b)^2 \geq 0$.)
15. Let X be the union of a countable, ascending sequence of measurable sets $\{X_n\}$ and $f\colon X \to [0, \infty]$ be measurable. Show that f is integrable over X if and only if there is an $M \geq 0$ for which $\int_{X_n} f\, d\mu \leq M$ for all n.
16. Let η be the counting measure on $2^{\mathbf{N}}$. Characterize the non-negative, real-valued functions (that is, sequences) that are integrable over $\mathbf{N}$ with respect to η and the value of $\int_{\mathbf{N}} f\, d\eta$.
17. Let $x_0 \in X$ and δ_{x_0} be the Dirac measure concentrated at x_0. Characterize the non-negative, real-valued functions on X that are integrable over X with respect to δ_{x_0} and the value of $\int_X f\, d\delta_{x_0}$.

11.3 INTEGRATION OF GENERAL MEASURABLE FUNCTIONS: THE DOMINATED CONVERGENCE THEOREM AND THE VITALI CONVERGENCE THEOREM

In this section, the standing assumption continues to be that $(X, \mathcal{M}, \mu)$ is a measure space. Let the function $f\colon X \to \overline{\mathbf{R}}$ be measurable. We define the **positive part** and the **negative part** of f, f^+ and f^- by

$$f^+ = \max\{f, 0\} \text{ and } f^- = \max\{-f, 0\} \text{ on } X.$$

Both f^+ and f^- are non-negative, measurable,

$$f = f^+ - f^- \text{ and } |f| = f^+ + f^- \text{ on } X.$$

Since $0 \leq f^+ \leq |f|$ and $0 \leq f^- \leq |f|$ on X, it follows from the linearity of integration for non-negative, measurable functions that f^+ and f^- are integrable if and only if $|f|$ is integrable.

Definition *A measurable function $f\colon X \to \overline{\mathbf{R}}$ is said to be* **integrable** *provided that $|f|\colon X \to \overline{\mathbf{R}}$ is integrable. For such a function, the integral of f over X with respect to μ is defined by*

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

Observe that, as in the case of non-negative, measurable functions, if $f\colon X \to \overline{\mathbf{R}}$ is integrable, then it is finite almost everywhere and has σ-finite support. Moreover, the excision of a set of measure zero from the domain of an integrable function leaves the integral unchanged. Consequently, for an integrable function, without loss of generality, we may assume that it is real-valued.

The Integral Comparison Test *Let the function $f\colon X \to \mathbf{R}$ be measurable. If there is a non-negative, integrable function $g\colon X \to [0, \infty)$ for which $|f| \leq g$ on X, then $f\colon X \to \mathbf{R}$ is integrable and*

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

Proof By the monotonicity of integration for non-negative, measurable functions, $|f|$ is integrable. Therefore,

$$\left|\int_X f\,d\mu\right| = \left|\int_X f^+\,d\mu - \int_X f^-\,d\mu\right| \leq \int_X f^+\,d\mu + \int_X f^-\,d\mu = \int_X |f|\,d\mu. \qquad \square$$

Theorem 11 (Linearity of Integration) *If the functions $f\colon X \to \mathbf{R}$ and $g\colon X \to \mathbf{R}$ are integrable, then for any α and β, the function $\alpha f + \beta g\colon X \to \mathbf{R}$ also is integrable and*

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

Proof Consider the case that $\alpha = \beta = 1$, and leave the case of general coefficients as an exercise. Both $|f|$ and $|g|$ are integrable, and so, by the linearity of integration for non-negative, measurable functions, the sum $|f| + |g|$ also is integrable. Since $|f + g| \leq |f| + |g|$ on X, by the integral comparison test, the sum $f + g$ is integrable. Therefore, the positive and negative parts of f, g and $f + g$ are integrable. To verify linearity is to show that

$$\int_X [f+g]^+\,d\mu - \int_X [f+g]^-\,d\mu = \left[\int_X f^+\,d\mu - \int_X f^-\,d\mu\right] + \left[\int_X g^+\,d\mu - \int_X g^-\,d\mu\right].$$

But

$$(f+g)^+ - (f+g)^- = f + g = (f^+ - f^-) + (g^+ - g^-) \text{ on } X,$$

and therefore, since each of these six functions is real-valued,

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+ \text{ on } X.$$

By the linearity of integration for non-negative, measurable functions,

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

Since f, g and $f + g$ are integrable, each of these six integrals is finite; rearrange them to conclude the proof of linearity of integration. $\square$

Theorem 12 (Countable Additivity over Domains of Integration) *If the function $f\colon X \to \mathbf{R}$ is integrable and $\{X_n\}_{n=1}^\infty$ is a measurable partition of X, then*

$$\int_X f\,d\mu = \sum_{n=1}^\infty \int_{X_n} f\,d\mu. \tag{6}$$

Proof Assume that $f \geq 0$. The general case follows by considering the positive and negative parts of f. For each n, define

$$f_n = \sum_{k=1}^{n} f \cdot \chi_{X_k} \text{ on } X.$$

The summation formula (6) now follows from the Monotone Convergence Theorem and the linearity of integration. $\square$

For an integrable function $f\colon X \to [0, \infty)$, this theorem implies that the set-function $E \mapsto \int_E f\,d\mu$ defines a finite measure on $\mathcal{M}$[2] and so has the continuity properties possessed by measures. This observation, applied to the positive and negative parts of an integrable function, provides the proof of the following theorem.

Theorem 13 (Continuity of Integration) *Let the function $f\colon X \to \mathbf{R}$ be integrable.*

(i) If $\{X_n\}_{n=1}^{\infty}$ is a countable, ascending collection of measurable subsets of X, then

$$\int_{\bigcup_{n=1}^{\infty} X_n} f\,d\mu = \lim_{n\to\infty} \int_{X_n} f\,d\mu.$$

(ii) If $\{X_n\}_{n=1}^{\infty}$ is a countable, descending collection of measurable subsets of X, then

$$\int_{\bigcap_{n=1}^{\infty} X_n} f\,d\mu = \lim_{n\to\infty} \int_{X_n} f\,d\mu.$$

The Dominated Convergence Theorem *Let $\{f_n\colon X \to \mathbf{R}\}$ be a sequence of measurable functions that converges pointwise on X to the function f. Assume that there is a non-negative, integrable function $g\colon X \to [0, \infty)$ that dominates the sequence $\{f_n\}$, in the sense that*

$$|f_n| \leq g \text{ on } X \text{ for all } n.$$

Then $f\colon X \to \mathbf{R}$ is integrable and

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

Proof By the integral comparison test, f and each f_n are integrable. Apply Fatou's Lemma and the linearity of integration to the two sequences of non-negative, measurable functions $\{g - f_n\}$ and $\{g + f_n\}$ in order to conclude that

$$\int_X g\,d\mu - \int_X f\,d\mu = \int_X [g - f]\,d\mu \leq \liminf \int_X [g - f_n]\,d\mu = \int_X g\,d\mu - \limsup \int_X f_n\,d\mu;$$

$$\int_X g\,d\mu + \int_X f\,d\mu = \int_X [g + f]\,d\mu \leq \liminf \int_X [g + f_n]\,d\mu = \int_X g\,d\mu + \liminf \int_X f_n\,d\mu.$$

Consequently,

$$\limsup \int_X f_n\,d\mu \leq \int_X f\,d\mu \leq \liminf \int_X f_n\,d\mu.$$

$\square$

[2] The integral over the empty-set is defined to be zero.

Proposition 14 *Let the function $f\colon X \to \mathbf{R}$ be integrable. Then for each $\epsilon > 0$, there is a $\delta > 0$ such that for any $E \in \mathcal{M}$,*

$$\text{if } \mu(E) < \delta, \text{ then } \int_E |f|\, d\mu < \epsilon. \tag{7}$$

Furthermore, for each $\epsilon > 0$, there is an $X_0 \in \mathcal{M}$ for which $\mu(X_0) < \infty$ and

$$\int_{X \sim X_0} |f|\, d\mu < \epsilon. \tag{8}$$

Proof It suffices to consider $f \geq 0$ on X. By the definition of the integral of a non-negative function, there is a simple function ψ on X for which

$$0 \leq \psi \leq f \text{ on } X \text{ and } 0 \leq \int_X f\, d\mu - \int_X \psi\, d\mu < \epsilon/2.$$

Choose $M > 0$ such that $0 \leq \psi \leq M$ on X. By the linearity and monotonicity of integration, if $E \subseteq X$ is measurable, then

$$\int_E f\, d\mu = \int_E \psi\, d\mu + \int_E [f - \psi]\, d\mu \leq \int_E \psi\, d\mu + \epsilon/2 \leq M \cdot m(E) + \epsilon/2.$$

Consequently, (7) holds for $\delta = \epsilon/2M$. Since the non-negative, simple function ψ is integrable, the measurable set $X_0 = \{x \in X \mid \psi(x) > 0\}$ has finite measure. Moreover, since $\psi = 0$ on $X \sim X_0$ and $f - \psi \geq 0$,

$$\int_{X \sim X_0} f\, d\mu = \int_{X \sim X_0} [f - \psi]\, d\mu \leq \int_X [f - \psi]\, d\mu < \epsilon.$$

Therefore, (8) holds. □

Definition *A collection $\mathcal{F}$ of integrable functions $f\colon X \to \mathbf{R}$ is said to be* **uniformly integrable** *provided that for each $\epsilon > 0$, there is a $\delta > 0$ such that for any measurable subset E of X and for any $f \in \mathcal{F}$,*

$$\text{if } \mu(E) < \delta, \text{ then } \int_E |f|\, d\mu < \epsilon.$$

The collection $\mathcal{F}$ is said to be **tight** *provided that for each $\epsilon > 0$, there is a subset X_0 of X for which $\mu(X_0) < \infty$ and for any $f \in \mathcal{F}$,*

$$\int_{X \sim X_0} |f|\, d\mu < \epsilon.$$

It follows from the preceding proposition that a finite collection of integrable functions is both uniformly integrable and tight. The Vitali Convergence Theorem, a generalization of the Dominated Convergence Theorem, has been established for the Lebesgue integral of a function of a real variable. We now provide a slight variation of this theorem for integration over general measure spaces.

The Vitali Convergence Theorem *Let the sequence of functions* $\{f_n\colon X \to \mathbf{R}\}$ *be uniformly integrable and tight. Assume that* $\{f_n\} \to f$ *pointwise on* X *and that the function* $f\colon X \to \mathbf{R}$ *is integrable. Then*

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

Proof First consider the case that $\mu(X) < \infty$. It follows from the integral comparison test and the linearity of integration, together with the assumption that f and each f_n is integrable, that, for any $E \in \mathcal{M}$ and any index n,

$$\begin{aligned} \left| \int_X f_n \, d\mu - \int_X f \, d\mu \right| &= \left| \int_X (f_n - f) \, d\mu \right| \\ &\leq \int_X |f_n - f| \, d\mu \\ &= \int_{X \sim E} |f_n - f| \, d\mu + \int_E |f_n - f| \, d\mu \\ &\leq \int_{X \sim E} |f_n - f| \, d\mu + \int_E |f_n| \, d\mu + \int_E |f| \, d\mu. \end{aligned} \tag{9}$$

Let $\epsilon > 0$. By the uniform integrability of $\{f_n\}$, there is a $\delta > 0$ such that for any n, $\int_E |f_n| \, d\mu < \epsilon/3$ if $E \in \mathcal{M}$ and $\mu(E) < \delta$. Therefore, by Fatou's Lemma, $\int_E |f| \, d\mu \leq \epsilon/3$ if $E \in \mathcal{M}$ and $m(E) < \delta$. Since $\mu(X) < \infty$, according to Egoroff's Theorem, there is a measurable subset E_0 of X for which $\mu(E_0) < \delta$ and $\{f_n\} \to f$ uniformly on $X \sim E_0$. Choose an index N for which $|f_n - f| < \epsilon/[3 \cdot \mu(X)]$ on $X \sim E_0$, for all $n \geq N$. Take $E = E_0$ in the integral inequality (9). If $n \geq N$, then

$$\begin{aligned} \left| \int_X f_n \, d\mu - \int_X f \, d\mu \right| &\leq \int_{X \sim E_0} |f_n - f| \, d\mu + \int_{E_0} |f_n| \, d\mu + \int_{E_0} |f| \, d\mu \\ &< \epsilon/[3 \cdot m(X)] \cdot m(X \sim E_0) + \epsilon/3 + \epsilon/3 \leq \epsilon. \end{aligned}$$

This completes the proof in the case that $\mu(X) < \infty$. Now consider the general case. Let $\epsilon > 0$. Since $\{f_n\}$ is tight, there is an $X_0 \in \mathcal{M}$ for which $\mu(X_0) < \infty$ and $\int_{X \sim X_0} |f_n| \, d\mu < \epsilon/3$ for all n. According to Fatou's Lemma, $\int_{X \sim X_0} |f| \, d\mu \leq \epsilon/3$. Consequently, for all n,

$$\left| \int_{X \sim X_0} f_n \, d\mu - \int_{X \sim X_0} f \, d\mu \right| \leq \int_{X \sim X_0} |f_n| \, d\mu + \int_{X \sim X_0} |f| \, d\mu < 2/3 \cdot \epsilon.$$

By the first case, considering restrictions of functions to X_0, there is an index N for which

$$\left| \int_{X_0} f_n \, d\mu - \int_{X_0} f \, d\mu \right| < \epsilon/3 \text{ for all } n \geq N.$$

Therefore,

$$\left|\int_X f_n\,d\mu - \int_X f\,d\mu\right| < \epsilon \text{ for all } n \geq N. \qquad \square$$

The Vitali Convergence Theorem for a general measure space differs from the special case of Lebesgue measure on $\mathbf{R}$. Unlike this special case, in the general case, it needs to be *assumed* that the limit function $f\colon X \to \mathbf{R}$ is integrable (see Problems 24 and 25). The proof of the following corollary is left as an exercise.

Corollary 15 *Let $\{f_n\colon X \to [0, \infty)\}$ be a sequence of non-negative, integrable functions. Suppose that $\{f_n(x)\} \to 0$ for all $x \in X$. Then*

$$\lim_{n\to\infty} \int_X f_n\,d\mu = 0 \textit{ if and only if } \{f_n\} \textit{ is uniformly integrable and tight.}$$

PROBLEMS

18. Let $\mathcal{M}$ be the σ-algebra of all subsets of $\mathbf{N}$ and η be the counting measure of $\mathcal{M}$. Let $\mathcal{F}$ be a uniformly pointwise bounded collection of functions that are integrable over $\mathbf{N}$ with respect to η
    (i) Show that $\mathcal{F}$ is uniformly integrable.
    (ii) Show that $\mathcal{F}$ is tight if and only if for each $\epsilon > 0$, there is an N such that if $\{a_k\} \in \mathcal{F}$, then $\sum_{n=N}^{\infty} |a_n| < \epsilon$. This property is called equisummability. .
    (iii) State the Vitali Convergence Theorem for the measure space $(\mathbf{N}, \mathcal{M}, \eta)$.
19. Show that if the function $f\colon X \to \mathbf{R}$ is integrable and $E \in \mathcal{M}$, then $f\colon E \to \mathbf{R}$ is integrable.
20. Assume that $\mu(X) = 0$ and the function $f\colon X \to \mathbf{R}$ take the constant value ∞. Show that $\int_X f\,d\mu = 0$.
21. Provide an example of a function $f\colon X \to \mathbf{R}$ for which $|f|\colon X \to \mathbf{R}$ is integrable but $f\colon X \to \mathbf{R}$ is not integrable. (Suggestion: Consider measurability.)
22. Let the function $f\colon X \to \mathbf{R}$ be finitely supported, bounded, and measurable. Show that

$$\int_X f\,d\mu = \sup \int_X \psi\,d\mu = \inf \int_X \varphi\,d\mu,$$

    where ψ and φ range over all simple functions on X for which $\psi \leq f \leq \varphi$ on X.
23. Assume that $(X, \mathcal{M}, , \mu)$ is complete, the function $f\colon X \to \mathbf{R}$ is finitely supported and bounded, and

$$\sup \int_X \psi\,d\mu = \inf \int_X \varphi\,d\mu,$$

    where ψ and φ range over all simple functions on X for which $\psi \leq f \leq \varphi$ on X. Prove that f is measurable.
24. Assume that $(X, \mathcal{M}, \mu)$ has the property that for each $\epsilon > 0$, X is the union of a finite collection of measurable sets, each of measure at most ϵ. Let the sequence $\{f_n\colon X \to [0, \infty]\}$ be uniformly integrable and converge pointwise on X to $f\colon X \to [0, \infty]$. Show that the function f is integrable.

25. On a set X, consider σ-algebra $\mathcal{M} = \{\emptyset, X\}$, and define $\mu(\emptyset) = 0, \mu(X) = 1$. Show that every collection of integrable functions on X is uniformly integrable and tight. Find a sequence of simple functions that converges pointwise to a non-integrable function.

26. Prove Corollary 15.

27. Deduce the Dominated Convergence Theorem from the Vitali Convergence Theorem.

28. Show that a collection $\mathcal{F}$ of integrable functions over X is uniformly integrable if and only if for each $\epsilon > 0$, there is a $\delta > 0$ such that for any $E \in \mathcal{M}$ and any $f \in \mathcal{F}$,
$$\text{if } \mu(E) < \delta, \text{ then } \left| \int_E f \, d\mu \right| < \epsilon.$$

29. Let η be another measure on $\mathcal{M}$. Assume that the function $f: X \to [0, \infty]$ is measurable. Is
$$\int_X f \, d[\mu + \eta] = \int_X f \, d\mu + \int_X f \, d\eta?$$

30. Let $\mathcal{M}_0$ be a σ-algebra that is contained in $\mathcal{M}$, μ_0 be the restriction of μ to $\mathcal{M}_0$, and the function $f: X \to [0, \infty]$ be measurable with respect to $\mathcal{M}_0$. Show that f is measurable with respect to $\mathcal{M}$ and
$$\int_X f \, d\mu_0 \leq \int_X f \, d\mu.$$
Provide an example for which this inequality is strict.

31. Let ν be a signed measure on a σ-algebra of subsets of X that has Jordan decomposition $\nu = \nu^+ - \nu^-$. For $f: X \to \mathbf{R}$, define
$$\int_X f \;\; d\nu = \int_X f \;\; d\nu^+ - \int_X f \;\; d\nu^-,$$
provided that f is integrable over X with respect to both ν^+ and ν^-. Show that if $|f| \leq M$ on X, then
$$\left| \int_X f \;\; d\nu \right| \leq M|\nu|(X).$$
Moreover, if $|\nu|(X) < \infty$, show that there is a measurable function f with $|f| \leq 1$ on X for which
$$\int_X f \;\; d\nu = |\nu|(X).$$

11.4 THE RADON-NIKODYM THEOREM

Let $(X, \mathcal{M}, \mu)$ be a measure space. For a non-negative, integrable function $f: X \to [0, \infty]$ define the set-function ν on $\mathcal{M}$ by
$$\nu(E) = \int_E f \, d\mu \text{ for all } E \in \mathcal{M}. \tag{10}$$
It follows from the countable additivity over domains of integration that ν is a measure, and
$$\text{if } E \in \mathcal{M} \text{ and } \mu(E) = 0, \text{ then } \nu(E) = 0. \tag{11}$$

The theorem named in the title of this section asserts that if μ is σ-finite, then every σ-finite measure ν on $\mathcal{M}$ for which (11) holds is represented by (10) for some non-negative, measurable function $f\colon X \to [0, \infty]$. A measure ν is said to be **absolutely continuous** with respect to the measure μ provided that (11) holds. The following proposition recasts absolute continuity as a more familiar continuity criterion.

Proposition 16 *Let $(X, \mathcal{M}, \mu)$ be a measure space and ν be a finite measure on $\mathcal{M}$. Then ν is absolutely continuous with respect to μ if and only if for each $\epsilon > 0$, there is a $\delta > 0$ such that for any set $E \in \mathcal{M}$,*

$$\textit{if } \mu(E) < \delta, \textit{ then } \nu(E) < \epsilon.$$

Proof It is clear that the ϵ-δ criterion implies that ν is absolutely continuous with respect to μ, independently of any finiteness assumption on ν. To prove the converse, we argue by contradiction. Indeed, otherwise, there is an $\epsilon_0 > 0$ and a sequence of sets in $\mathcal{M}$, $\{E_n\}$, such that for each n, $\mu(E_n) < 1/2^n$ while $\nu(E_n) \geq \epsilon_0$. For each n, define $A_n = \bigcup_{k=n}^{\infty} E_k$. Then $\{A_n\}$ is a descending sequence of sets in $\mathcal{M}$. By the monotonicity of ν and the countable monotonicity of μ,

$$\nu(A_n) \geq \epsilon_0 \text{ and } \mu(A_n) \leq 1/2^{n-1} \text{ for all } n.$$

Define $A_\infty = \bigcap_{k=1}^{\infty} A_n$. By the monotonicity of μ, $\mu(A_\infty) = 0$. It follows from the continuity ν that, since $\nu(A_1) \leq \nu(X) < \infty$ and $\nu(A_n) \geq \epsilon_0$ for all n, $\nu(A_\infty) \geq \epsilon_0$. This contradicts the absolute continuity of ν with respect to μ. $\square$

The Radon-Nikodym Theorem *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and ν be a σ-finite measure on $\mathcal{M}$ that is absolutely continuous with respect to μ. Then there is a non-negative, measurable function $f\colon X \to [0, \infty]$ for which*

$$\nu(E) = \int_E f\, d\mu \text{ for all } E \in \mathcal{M}. \tag{12}$$

The function f is unique, in the sense that if the non-negative, measurable function $g\colon X \to [0, \infty]$ also has this property, then $g = f$ almost everywhere with respect to μ.

Proof First consider the case that $\mu(X) < \infty$ and $\nu(X) < \infty$. If $\nu(X) = 0$, then (12) holds for $f \equiv 0$ on X. Assume that $\nu(X) > 0$, and therefore, by the absolute continuity of ν with respect to μ, $\mu(X) > 0$ also. We claim that there is a non-negative, measurable function $f\colon X \to [0, \infty]$ for which

$$\int_X f\, d\mu > 0 \text{ and } \int_E f\, d\mu \leq \nu(E) \text{ for all } E \in \mathcal{M}. \tag{13}$$

Indeed, since $\nu(X) > 0$ and $\mu(X) < \infty$, $\lambda > 0$ can be chosen sufficiently small so that

$$(\nu - \lambda\mu)(X) > 0.$$

Define the signed measure η on $\mathcal{M}$ by $\eta = \nu - \lambda\mu$. According to the Hahn Decomposition Theorem, there is a Hahn decomposition $\{P_\lambda, N_\lambda\}$ for η. Then $\mu(P_\lambda) > 0$, since otherwise, by the absolute continuity of ν with respect to μ, $\eta(P_\lambda) = 0$, which contradicts the assumption that $\eta(X) > 0$. Define $f = \lambda \cdot \chi_{P_\lambda}$. Then $\int_X f\, d\mu > 0$. Since $\eta(E \cap P_\lambda) \geq 0$ for all $E \in \mathcal{M}$,

$$\int_E f\, d\mu = \lambda \cdot \mu(E \cap P_\lambda) \leq \nu(E \cap P_\lambda) \leq \nu(E) \text{ for all } E \in \mathcal{M}.$$

The function f satisfies (13). Let $\mathcal{F}$ be the collection of all such functions, and define

$$M = \sup_{f \in \mathcal{F}} \int_X f\, d\mu.$$

Choose an increasing sequence $\{f_n\}$ in $\mathcal{F}$ for which $\lim_{n\to\infty} \int_X f_n\, d\mu = M$. According to the Monotone Convergence Theorem, $\{f_n\}$ converges pointwise to a non-negative, measurable function $f\colon X \to [0, \infty]$, $\int_X f\, d\mu = M$ and f belongs to $\mathcal{F}$. We claim that (12) holds for this choice of f. We argue by contradiction. Indeed, otherwise, there is a set $X_0 \in \mathcal{M}$ for which

$$\int_{X_0} f\, d\mu < \nu(X_0).$$

Consider the σ-algebra of subsets of X_0, $\mathcal{M}_0 = \{E \cap X_0 \mid E \in \mathcal{M}\}$. Choose $\lambda_0 > 0$ sufficiently small so that

$$\nu(X_0) - \int_{X_0} f\, d\mu - \lambda_0 \cdot \mu(X_0) > 0,$$

and define the signed measure $\eta_0\colon \mathcal{M}_0 \to \mathbf{R}$ by

$$\eta_0(E_0) = \nu(E_0) - \int_{E_0} f\, d\mu - \lambda_0 \cdot \mu(E_0) \text{ for all } E_0 \in \mathcal{M}_0.$$

There is a Hahn decomposition $\{P_0, N_0\}$ for η_0. Then $\mu(P_0) > 0$, since otherwise, by the absolute continuity of ν with respect to μ, $\eta_0(P_0) = 0$, which contradicts $\eta_0(X_0) > 0$. Define $f_0\colon X \to [0, \infty)$ by $f_0 = \lambda_0 \cdot \chi_{P_0}$. Then $\int_X f_0\, d\mu > 0$. Since $\eta_0(E \cap P_0) \geq 0$ for all $E \in \mathcal{M}$,

$$\int_{E \cap X_0} [f_0 + f]\, d\mu \leq \nu(E \cap X_0) \text{ for all } E \in \mathcal{M}.$$

Also, since $f \in \mathcal{F}$, and $P_0 \subseteq X_0$,

$$\int_{E \sim X_0} [f_0 + f]\, d\mu = \int_{E \sim X_0} f\, d\mu \leq \nu(E \sim X_0) \text{ for all } E \in \mathcal{M}.$$

Therefore, $f + f_0 \in \mathcal{F}$ and

$$\int_E [f_0 + f]\, d\mu = \int_E f_0\, d\mu + \int_E f\, d\mu = \lambda_0 \cdot \mu(P_0) + M > M,$$

which contradicts the definition of M. This completes the proof if μ and ν are both finite.

Now consider the general case, namely, μ and ν are σ-finite. There are two countable, measurable partitions of X, on one of which each set has μ finite and, on the other, each set has ν finite. By intersecting sets in these partitions, a countable, measurable partition of X is obtained for which each set has both μ and ν finite. An application of the special case just considered on each set in this partition, together with the countable additivity of integration and measure, establishes the general case. It remains to prove uniqueness. Suppose that f_1 and f_2 are measurable functions, each of which are integral representations with respect to μ of ν. Then

$$\int_E [f_1 - f_2]\, d\mu = 0 \text{ for all } E \in \mathcal{M}.$$

Therefore, $f_1 = f_2$ almost everywhere $[\mu]$ on X. $\square$

The function f for which (12) holds is called the **Radon-Nikodym derivative** of ν with respect to μ. It is often denoted by $\frac{d\nu}{d\mu}$, so that

$$\nu(E) = \int_E \frac{d\nu}{d\mu}\, d\mu \text{ for all } E \in \mathcal{M}.$$

In Chapter 20, we consider Hilbert spaces, and provide an elegant proof by von Neumann of the Radon-Nikodym Theorem.

Now, in Chapter 9, we proved that a signed measure $\nu\colon \mathcal{M} \to \mathbf{R}$ has a Jordan decomposition $\nu = \nu_1 - \nu_2$, where ν_1 and ν_2 are measures, one of which is finite, and observed that $|\nu| = \nu_1 + \nu_2$ also defines a measure. The signed measure ν is defined to be absolutely continuous with respect to a measure $\mu\colon \mathcal{M} \to [0, \infty]$ provided that $|\nu|$ is absolutely continuous with respect to μ, which is equivalent to the absolute continuity of both ν_1 and ν_2 with respect to μ. By this decomposition of signed measures and the Radon-Nikodym Theorem, there is the following variation of this same theorem for finite signed measures.

Corollary 17 *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and $\nu\colon \mathcal{M} \to \mathbf{R}$ be a finite signed measure that is absolutely continuous with respect to μ. Then there is a function $f\colon X \to \mathbf{R}$ that is integrable over X with respect to μ and*

$$\nu(E) = \int_E f\, d\mu \textit{ for all } E \in \mathcal{M}.$$

Two measures μ and ν on a σ-algebra $\mathcal{M}$ are said to be **mutually singular**, or singular with respect to one other, provided that there is a measurable partition $\{A, B\}$ of X for which $\nu(A) = \mu(B) = 0$.

The Lebesgue Decomposition Theorem *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and ν be a σ-finite measure on $\mathcal{M}$. Then there is a measure ν_0 on $\mathcal{M}$, singular with respect to μ, and a measure ν_1 on $\mathcal{M}$, absolutely continuous with respect to μ, for which $\nu = \nu_0 + \nu_1$. The measures ν_0 and ν_1 are unique.*

Proof Define $\lambda = \mu + \nu$. We leave it as an exercise to show that if $g\colon X \to [0, \infty]$ is measurable, then

$$\int_E g\, d\lambda = \int_E g\, d\mu + \int_E g\, d\nu \text{ for all } E \in \mathcal{M}.$$

Since μ and ν are σ-finite measures, so is the measure λ. Moreover, μ is absolutely continuous with respect to λ. According to the Radon-Nikodym Theorem, there is a measurable function $f\colon X \to [0, \infty]$ for which

$$\mu(E) = \int_E f\,d\lambda = \int_E f\,d\mu + \int_E f\,d\nu \text{ for all } E \in \mathcal{M}. \tag{14}$$

Define $X_+ = \{x \in X \mid f(x) > 0\}$ and $X_0 = \{x \in X \mid f(x) = 0\}$. Since f is a measurable function, $X = X_0 \cup X_+$ is a measurable partition of X, so that $\nu = \nu_0 + \nu_+$ is the sum of mutually singular measures, where

$$\nu_0(E) = \nu(E \cap X_0) \text{ and } \nu_+(E) = \nu(E \cap X_+) \text{ for all } E \in \mathcal{M}.$$

Now $\mu(X_0) = \int_{X_0} f\,d\lambda = 0$, since $f = 0$ on X_0, and $\nu_0(X_+) = \nu(X_+ \cap X_0) = \nu(\emptyset) = 0$. Therefore, μ and ν_0 are mutually singular. It remains only to show that ν_+ is absolutely continuous with respect to μ. Let $\mu(E) = 0$. Then $\int_E f\,d\mu = 0$. Consequently, by (14) and the additivity of integration over domains,

$$\int_E f\,d\nu = \int_{E\cap X_0} f\,d\nu + \int_{E\cap X_+} f\,d\nu = 0.$$

But $f = 0$ on $E \cap X_0$ and $f > 0$ on $E \cap X_+$ and so $\nu(E \cap X_+) = 0$, that is, $\nu_+(E) = 0$. The proof of uniqueness is left as an exercise. $\square$

The Radon-Nikodym Theorem for the measure space $([a, b], \mathcal{M}, m)$, where m is Lebesgue measure on the σ-algebra $\mathcal{M}$ of Lebesgue measurable subsets of $[a, b]$, is a consequence of the representation of an absolutely continuous function as the indefinite integral of its derivative. To see this, let the measure $\nu\colon \mathcal{M} \to [0, \infty)$ be absolutely continuous with respect to m, and define the function $h\colon [a, b] \to \mathbf{R}$

$$h(x) = \nu([a, x]) \text{ for all } x \in [a, b].$$

It follows from Proposition 16 that the function h inherits absolute continuity from the absolute continuity of the measure ν with respect to m. Consequently, since $m(\{x\}) = \nu(\{x\}) = 0$ for all $x \in [a, b]$,

$$\nu((c, d)) = h(d) - h(c) = \int_{(c,\,d)} h'\,dm \text{ for all } (c, d) \subseteq (a, b).$$

It is left as an exercise to conclude from this that, since a Lebesgue measurable subset of $[a, b]$ is a G_δ set from which a set of Lebesgue measure zero has been excised,

$$\nu(E) = \int_E h'\,dm \text{ for all } E \in \mathcal{M}.$$

PROBLEMS

32. Let $f\colon [a, b] \to \mathbf{R}$ be of bounded variation. Show that there is an absolutely continuous functions g on $[a, b]$ and a function h on $[a, b]$ that is of bounded variation and has $h' = 0$ almost everywhere on $[a, b]$, for which $f = g + h$ on $[a, b]$. Then show that this decomposition is unique except for addition of constants.

33. For the Lebesgue measure space $(\mathbf{R}, \mathcal{M}, m)$, let $\nu\colon \mathcal{M} \to [0, \infty)$ be a measure and define $h\colon [a, b] \to \mathbf{R}$ $h(x) = \nu([a, x])$ for all $x \in [a, b]$. Show that the function h is absolutely continuous if and only if the measure ν is absolutely continuous with respect to m.

34. Let $\{\mu_n\}$ be a sequence of measures on $\mathcal{M}$ for which there is a $c > 0$ such that $\mu_n(X) \leq c$ for all n. Define $\mu\colon \mathcal{M} \to [0, \infty]$ by

$$\mu = \sum_{n=1}^{\infty} \frac{\mu_n}{2^n}.$$

Show that μ is a measure on $\mathcal{M}$ and each μ_n is absolutely continuous with respect to μ.

35. Characterize the measure spaces $(X, \mathcal{M}, \mu)$ for which the counting measure on $\mathcal{M}$ is absolutely continuous with respect to μ and those for which, given $x_0 \in X$, the Dirac measure δ_{x_0} on $\mathcal{M}$ is absolutely continuous with respect to μ.

36. Let $X = [0, 1]$, $\mathcal{M}$ be the collection of Lebesgue measurable subsets of $[0, 1]$, m be Lebesgue measure, and c be the counting measure of $\mathcal{M}$. Show that m is absolutely continuous with respect to c, but there is no function $f\colon X \to [0, \infty]$ that is integrable over $[0, 1]$ with respect to c for which $m(E) = \int_E f\, dc$ for all $E \in \mathcal{M}$.

37. Let $f\colon X \to [0, \infty)$ be integrable. Find the Lebesgue decomposition with respect to μ of the measure ν defined by $\nu(E) = \int_E f\, d\mu$ for $E \in \mathcal{M}$.

38. Let μ, ν, and λ be σ-finite measures on $\mathcal{M}$.
    (i) If $\nu \ll \mu$ (meaning ν is absolutely continuous with respect to μ) and $f\colon X \to [0, \infty]$ is measurable, show that
    $$\int_X f\, d\nu = \int_X f \left[\frac{d\nu}{d\mu}\right] d\mu.$$
    (ii) If $\nu \ll \mu$ and $\lambda \ll \mu$, show that
    $$\frac{d(\nu + \lambda)}{d\mu} = \frac{d\nu}{d\mu} + \frac{d\lambda}{d\mu} \text{ almost everywhere } [\mu].$$
    (iii) If $\nu \ll \mu \ll \lambda$, show that
    $$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda} \text{ almost everywhere } [\lambda].$$
    (iv) If $\nu \ll \mu$ and $\mu \ll \nu$, show that
    $$\frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\nu} = 1 \text{ almost everywhere } [\mu].$$

39. Let μ, , ν_1, and ν_2 be measures on $\mathcal{M}$.
    (i) Show that if ν_1 and ν_2 are singular with respect to μ, then, for any $\alpha \geq 0, \beta \geq 0$, so is the measure $\alpha\nu_1 + \beta\nu_2$.
    (ii) Show that if ν_1 and ν_2 are absolutely continuous with respect to μ, then, for any $\alpha \geq 0, \beta \geq 0$, so is the measure $\alpha\nu_1 + \beta\nu_2$.
    (iii) Prove the uniqueness assertion of the Lebesgue decomposition.

40. Let μ and ν be measures on $\mathcal{M}$, define $\lambda = \mu + \nu$, and let $f\colon X \to [0, \infty]$ be measurable. Show that f is integrable with respect to λ if and only if it is integrable with respect to both μ and ν. Also show that if f is integrable with respect to λ, then

$$\int_E g\,d\lambda = \int_E g\,d\mu + \int_E g\,d\nu \text{ for all } E \in \mathcal{M}.$$

11.5 PRODUCT MEASURES: THE TONELLI AND FUBINI THEOREMS

Throughout this section, $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are two reference measure spaces. Consider the Cartesian product $X \times Y$ of X and Y. If $A \in \mathcal{A}$ and $B \in \mathcal{B}$, $\mu(A) < \infty$ and $\nu(B) < \infty$, then $A \times B \subseteq X \times Y$ is called a **measurable rectangle**.

Lemma 18 *Let $\{A_k \times B_k\}_{k=1}^{\infty}$ be a disjoint collection of measurable rectangles, whose union also is the measurable rectangle $A \times B$. Then*

$$\mu(A) \times \nu(B) = \sum_{k=1}^{\infty} \mu(A_k) \times \nu(B_k).$$

Proof Fix $x \in A$. For each $y \in B$, the point (x, y) belongs to exactly one $A_k \times B_k$. Therefore, there is the following disjoint union:

$$B = \bigcup_{\{k \mid x \in A_k\}} B_k.$$

By the countable additivity of the measure ν,

$$\nu(B) = \sum_{\{k \mid x \in A_k\}} \nu(B_k).$$

Rewrite this equality in terms of characteristic functions:

$$\nu(B) \cdot \chi_A(x) = \sum_{k=1}^{\infty} \nu(B_k) \cdot \chi_{A_k}(x) \text{ for all } x \in A.$$

Since each A_k is contained in A, this equality also clearly holds for $x \in X \setminus A$. Consequently,

$$\nu(B) \cdot \chi_A = \sum_{k=1}^{\infty} \nu(B_k) \cdot \chi_{A_k} \text{ on } X.$$

By the Monotone Convergence Theorem,

$$\mu(A) \times \nu(B) = \int_X \nu(B) \cdot \chi_A d\mu = \sum_{k=1}^{\infty} \int_X \nu(B_k) \cdot \chi_{A_k} d\mu = \sum_{k=1}^{\infty} \mu(A_k) \times \nu(B_k). \qquad \square$$

Theorem 19 *Let $\mathcal{R} = \mathcal{R}(X \times Y)$ be the collection of measurable rectangles in $X \times Y$, and for $A \times B \in \mathcal{R}$, define*

$$\lambda(A \times B) = \mu(A) \cdot \nu(B).$$

Then $\mathcal{R}$ is a semi-ring and $\lambda\colon \mathcal{R} \to [0, \infty]$ is a premeasure.

Proof We first show that $\mathcal{R}$ is a semi-ring. It is clear that the intersection of two measurable rectangles is a measurable rectangle. The relative complement of two measurable rectangles is the disjoint union of two measurable rectangles. To see this, let A and B be measurable subsets of X and C and D be measurable subsets of Y. Observe that

$$(A \times C) \sim (B \times D) = [(A \sim B) \times C] \cup [(A \cap B) \times (C \sim D)],$$

and the right-hand union is the disjoint union of two measurable rectangles. It remains to show that λ is a premeasure. The finite additivity of λ follows from the preceding lemma. It is also clear that λ is monotone. To establish the countable monotonicity of λ, let the measurable rectangle E be covered by the collection $\{E_k\}_{k=1}^{\infty}$ of measurable rectangles. Since $\mathcal{R}$ is a semi-ring, according to Lemma 7 in Chapter 9, it may be assumed that $\{E_k\}_{k=1}^{\infty}$ is disjoint. Therefore,

$$E = \bigcup_{k=1}^{\infty} E \cap E_k,$$

this union being disjoint and each $E \cap E_k$ being a measurable rectangle. According to the preceding lemma and the monotonicity of λ,

$$\lambda(E) = \sum_{k=1}^{\infty} \lambda(E \cap E_k) \leq \sum_{k=1}^{\infty} \lambda(E_k).$$

Consequently, λ is countably monotone. □

In view of this proposition, by an appeal to the Carathéodory-Hahn Theorem, the product measure may be defined as a Carathéodory extension.

Definition *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces, $\mathcal{R} = \mathcal{R}(X \times Y)$ be the semi-ring of measurable rectangles contained in $X \times Y$, and λ be the premeasure defined on $\mathcal{R}$ by*

$$\lambda(A \times B) = \mu(A) \cdot \nu(B) \text{ for } A \times B \in \mathcal{R}.$$

By the **product measure** *$\mu \times \nu$ is meant the Carathéodory extension of $\lambda \colon \mathcal{R} \to [0, \infty]$ defined on the σ-algebra of λ^* measurable subsets of $X \times Y$.*

The following result supports the intuitive understanding in beginning calculus of the Riemann integral of a non-negative function as being the area under the graph.

Theorem 20 *If $(X, \mathcal{A}, \mu)$ is a general measure space and the non-negative function $f \colon X \to [0, \infty)$ is integrable, then*

$$\int_X f(x)\, d\mu(x) = (\mu \times m)(E) \text{ where } E = \{(x, t) \in X \times \mathbf{R} \mid 0 \leq t \leq f(x)\}. \tag{15}$$

Proof By considering the canonical representation of a simple function and only the definitions of the integral and product measure, (15) clearly holds if f is non-negative, finitely supported, and simple. Since f is non-negative and measurable, by conclusion (ii) of the Simple

Approximation Theorem, it is the pointwise limit of an increasing sequence $\{\varphi_n\colon X \to [0, \infty]\}$ of non-negative, simple functions. Consequently, by the Monotone Convergence Theorem and the continuity of measure, if, for each n, $E_n = \{(x, t) \in X \times \mathbf{R} \,|\, 0 \le t \le \varphi_n(x)\}$, then

$$\int_X f(x)\,d\mu(x) = \lim_{n\to\infty} \int_X \varphi_n(x)\,d\mu(x) = \lim_{n\to\infty} (\mu \times m)(E_n) = (\mu \times m)(E).$$ □

It is important to observe that the product measure, being a Carathéodory extension, is complete, irrespective of completeness of the factor measure spaces[3]. Our aim now is to extend to a general setting the familiar theorems in several-variable calculus for verifying iterated integration with respect to the Riemann integral. Given $E \subseteq X \times Y$ and $f\colon X \times Y \to \overline{\mathbf{R}}$, for $x \in X$, the set $E_x = \{y \in Y \,|\, (x, y) \in E\} \subseteq Y$ is called the **x-section of E** and the function $f(x, \cdot)$ defined on E_x by $f(x, \cdot)(y) = f(x, y)$ is called the **x-section of f**. And similarly for y-sections E_y. The following is the first of two fundamental results regarding iterated integration.

Tonelli's Theorem *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite, complete measure spaces, and the non-negative function $k\colon X \times Y \to [0, \infty]$ be measurable with respect to $\mu \times \nu$. Then the function $x \mapsto \int_Y k(x, y)\,d\nu(y)$ is defined almost everywhere on X and is μ-measurable, and similarly for the function $y \mapsto \int_X k(x, y)\,d\mu(x)$, and*

$$\int_{X \times Y} k \; d(\mu \times \nu) = \int_X \left[\int_Y k(x, y)\; d\nu(y)\right] d\mu(x) = \int_Y \left[\int_X k(x, y)\; d\mu(x)\right] d\nu(y). \tag{16}$$

Observe that there is no assumption regarding integrability in Tonelli's Theorem: consider, for example, the function k that is identically equal to ∞. The following examples are instructive regarding the conclusions and assumptions of this theorem. In each of the three examples, $(X, \mathcal{A}, \mu)$ is the complete, σ-finite Lebesgue measure space $(\mathbf{R}, \mathcal{L}, m)$, and for $(Y, \mathcal{B}, \nu)$ we consider the spaces $(\mathbf{R}, \mathcal{L}, m)$, Lebesgue measure on the Borel sets, $(\mathbf{R}, \mathcal{B}(\mathbf{R}), m)$, and the counting measure on $\mathcal{L}$, respectively, the second of which is not complete, and the third is not σ-finite.

Example Consider $(X, \mathcal{A}, \mu) = (Y, \mathcal{B}, \nu) = (\mathbf{R}, \mathcal{L}, m)$. We showed that there is a non-measurable subset A of $[0, 1]$. The set $E = \mathbf{Q} \times [0, 1]$ is a measurable rectangle and, since $m(Q) = 0$, $(m \times m)(E) = 0$. Every product measure, being a Carétheodory extension, is complete. Since $m \times m$ is complete, the set $E_0 = \mathbf{Q} \times A$ is $m \times m$ measurable. Therefore, the function $k = \chi_{E_0}\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is non-negative and $m \times m$ measurable. However, for every x in $\mathbf{Q}$, the function $y \mapsto k(x, y)$ fails to be Lebesgue measurable, since its domain $E_x = A$ is not a Lebesgue measurable set. Consequently, the restriction of "almost everywhere" cannot be removed from the conclusion of the theorem.

[3]The reader should be aware that there are other definitions of product measure for which the product measure space is not complete. Sometimes, the actual product measure is as defined above, but the collection of measurable sets is defined to be the smallest σ-algebra containing all measurable rectangles. This measure space may fail to be complete. For instance, with this definition, the product of $(\mathbf{R}^1, \mathcal{L}^1, \mu_1)$ with itself is the Borel measure space $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2), \mu_2)$, which is not complete. In the next section, we show that for Lebesgue measure on Euclidean space, the product of $(\mathbf{R}^k, \mathcal{L}^k, \mu_k)$ with $(\mathbf{R}^m, \mathcal{L}^m, \mu_m)$ is $(\mathbf{R}^{m+k}, \mathcal{L}^{m+k}, \mu_{m+k})$.

The following two examples show that neither the assumption of completeness nor σ-finiteness can be omitted in this theorem.

Example Consider $(X, \mathcal{A}, \mu) = (\mathbf{R}, \mathcal{L}, m)$ and $(Y, \mathcal{B}, \nu) = (\mathbf{R}, \mathcal{B}(\mathbf{R}), m)$, where $\mathcal{B}(\mathbf{R})$ is the Borel σ-algebra for $\mathbf{R}$. We showed that there is a subset A of the Cantor set C that is is not a Borel set, and, of course, $m(C) = 0$. Then $E = [0, 1] \times C$ is an $m \times m$-measurable rectangle, and $(m \times m)(E) = 0$. Since $m \times m$ is complete, $E_0 = [0, 1] \times A$ is $m \times m$-measurable, so that the function $k = \chi_{E_0}\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is non-negative and $m \times m$ measurable. However, for every x in $\mathbf{R}$, the function $y \mapsto k(x, y)$ fails to be Borel measurable, since its domain $E_x = A$ is not a Borel set. Observe that $(X, \mathcal{A}, \mu)$ is σ-finite and complete, while $(Y, \mathcal{B}, \nu)$ is σ-finite, but not complete.

Example Consider $(X, \mathcal{A}, \mu) = (\mathbf{R}, \mathcal{L}, m)$ and $(Y, \mathcal{B}, \nu) = (\mathbf{R}, \mathcal{L}, c)$, where c is the counting measure. Define $\Delta = \{(x, y) \mid x = y\}$. Since measurable rectangles have finite product measure, the set Δ cannot be covered by a countable collection of $(m \times c)$-measurable rectangles, and so, by the definition of the outer-measure $(m \times c)^*$, $(m \times c)^*(\Delta) = \infty$. By the definition of a $(\mu \times \nu)^*$-measurable set, Δ is $(m \times c)^*$-measurable. Therefore, $k = \chi_\Delta\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is a non-negative and $m \times c$ measurable function. For every $x \in \mathbf{R}$, $m(\Delta_x) = 0$. Consequently,

$$\int_{\mathbf{R}\times\mathbf{R}} k \, d(m \times c) = \infty \text{ and } \int_{\mathbf{R}} \left[\int_{\mathbf{R}} k(x, y) \, dm(x) \right] dc(y) = 0.$$

Observe that $(X, \mathcal{A}, \mu)$ is σ-finite and complete, and $(Y, \mathcal{B}, \nu)$ is complete, but not σ-finite,

There is the following companion of Tonelli's Theorem, in which the assumptions that $k \geq 0$ and the factor measure spaces are σ-finite are replaced by the single assumption that $k\colon X \times Y \to \mathbf{R}$ is integrable.

Fubini's Theorem *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be complete measure spaces, and the function $k\colon X \times Y \to \mathbf{R}$ be integrable. Then the function $x \mapsto \int_Y k(x, y) \, d\nu(y)$ is defined almost everywhere on X and is μ-integrable, similarly for the function $y \mapsto \int_X k(x, y) \, d\mu(x)$, and*

$$\int_{X\times Y} k \, d(\mu \times \nu) = \int_X \left[\int_Y k(x, y) \, d\nu(y) \right] d\mu(x) = \int_Y \left[\int_X k(x, y) \, d\mu(x) \right] d\nu(y). \quad (17)$$

We now show that the Tonelli and Fubini Theorems are consequences of the following lemma, which is the very special case of both, namely, for k the characteristic function of a set that has finite product measure. We then turn to the proof of the lemma itself. The assumptions and conclusions of these two theorems are symmetric with respect to x and y. Therefore, it suffices to establish the left-hand side of the integral equalities, and so the following lemmas address just these left-hand equalities.

The Fubini-Tonelli Lemma *Assume that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are measure spaces and $(Y, \mathcal{B}, \nu)$ is complete. Let $E \subseteq X \times Y$ be measurable with respect to the product measure and $(\mu \times \nu)(E) < \infty$. Then for almost all x, the section E_x is ν-measurable, the function $x \mapsto \nu(E_x)$ is μ-measurable, and*

$$(\mu \times \nu)(E) = \int_X \nu(E_x) \, d\mu(x). \quad (18)$$

Proof of Tonelli's Theorem Let $\varphi\colon X \times Y \to \mathbf{R}$ be a finitely supported simple function. By the Fubini-Tonelli Lemma and the linearity of integration, the function $x \mapsto \int_Y \varphi(x, y)\, d\nu(y)$ is defined almost everywhere and is μ-integrable, and

$$\iint_{X\times Y} \varphi \, d(\mu\times\nu) = \int_X \left[\int_Y \varphi(x, y)\, d\nu(y)\right] d\mu(x). \tag{19}$$

Since μ and ν are σ-finite, so is the product measure. Therefore, being non-negative and measurable, by conclusion (i) and (ii) of the Simple Approximation Theorem, k is the pointwise limit of an increasing sequence $\{\varphi_n\colon X\times Y \to [0, \infty)\}$ of non-negative, finitely supported, simple functions. We have

$$\iint_{X\times Y} \varphi_n \, d(\mu\times\nu) = \int_X \left[\int_Y \varphi_n(x, y)\, d\nu(y)\right] d\mu(x) \text{ for each } n. \tag{20}$$

By the Monotone Convergence Theorem with respect to the product measure $\mu\times\nu$,

$$\iint_{X\times Y} k \, d(\mu\times\nu) = \lim_{n\to\infty} \iint_{X\times Y} \varphi_n \, d(\mu\times\nu). \tag{21}$$

On the other hand, again by the Fubini-Tonelli Lemma, there is a subset $X_0 \subseteq X$ for which $\mu(X \sim X_0) = 0$ and

$$y \mapsto \varphi_n(x, y) \text{ is } \nu\text{-measurable for each } x\in X_0 \text{ and each } n.$$

Therefore, by the Monotone Convergence Theorem with respect to ν,

$$\int_Y k(x, y)\, d\nu(y) = \lim_{n\to\infty} \int_Y \varphi_n(x, y)\, d\nu(y) \text{ for each } x \in X_0.$$

For $x \in X_0$ and each n, define $g_n(x) = \int_Y \varphi_n(x, y)\, d\nu(y)$. Again by the Fubini-Tonelli Lemma, each function $g_n\colon X \to [0, \infty]$ is μ-measurable. Then $\{g_n\colon X \to [0, \infty]\}$ is an increasing sequence of μ-measurable functions, and so, by Monotone Convergence Theorem with respect to μ,

$$\int_X \left[\int_Y k(x, y)\, d\nu(y)\right] d\mu(x) = \lim_{n\to\infty} \int_X \left[\int_Y \varphi_n(x, y)\, d\nu(y)\right] d\mu(x).$$

This equality, together with (20) and (21), concludes the proof of Tonelli's Theorem. $\square$

Proof of Fubini's Theorem By expressing k as $k = k^+ - k^-$, where $k^\pm$ are non-negative and integrable, the proof reduces to the case that $k \geq 0$. By conclusion (ii) of the Simple Approximation Theorem, there is an increasing sequence $\{\varphi_n\colon X \times Y \to [0, \infty]\}$ of non-negative, simple functions that converges pointwise to k. Since k is integrable, each φ_n is finitely supported. An appeal to (19) and the Monotone Convergence Theorem concludes the proof, exactly as it did in the proof of Tonelli's Theorem. $\square$

Recall that $\mathcal{R}_\sigma$ denotes the collection of countable unions of measurable rectangles, and $\mathcal{R}_{\sigma\delta}$ denotes the collection of countable intersections of sets in $\mathcal{R}_\sigma$. Rewording Lemma 7 of the Chapter 9, we now describe the properties of product measurable sets that we will use to prove the Fubini-Tonelli Lemma, a description which suggests its proof.

Proposition 21 *If $E \subseteq X \times Y$ is measurable with respect to the product measure and $(\mu \times \nu)(E) < \infty$, then E has the following approximation property: there is a measurable set $E' \subseteq X \times Y$ for which*

$$E \subseteq E' \text{ and } (\mu \times \nu)(E' \sim E) = 0, \text{ where} \tag{22}$$

$E' = \bigcap_{k=1}^\infty E_n$, the intersection of a descending collection of sets, each of which is the countable, disjoint union of measurable rectangles.

Lemma 22 *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be general measure spaces. If $E \in \mathcal{R}_{\sigma\delta}$ and $(\mu \times \nu)(E) < \infty$, then for all $x \in X$, the set E_x is ν-measurable, the function $x \mapsto \nu(E_x)$ is μ-measurable, and (18) holds.*

Proof First consider the case that $E = A \times B$, a measurable rectangle. For $x \in X$,

$$E_x = \begin{cases} B & \text{for } x \in A \\ \emptyset & \text{for } x \notin A. \end{cases}$$

Consequently,

$$(\mu \times \nu)(E) = \mu(A) \cdot \nu(B) = \nu(B) \cdot \int_X \chi_A \, d\mu = \int_X \nu(E_x) \, d\mu(x).$$

Now consider a set E in $\mathcal{R}_\sigma$. There is a measurable partition of $\{A_n \times B_n\}_{n=1}^\infty$ of E into measurable rectangles. Fix $x \in X$. Observe that

$$E_x = \bigcup_{n=1}^\infty (A_n \times B_n)_x.$$

Therefore, E_x is ν-measurable, since it is the countable union of ν-measurable sets, and because this union is disjoint, by the countable additivity of ν,

$$\nu(E_x) = \sum_{n=1}^\infty \nu((A_n \times B_n)_x).$$

By the Monotone Convergence Theorem, the validity of (18) for each measurable rectangle $A_n \times B_n$ and the countable additivity of the measure $\mu \times \nu$,

$$\begin{aligned} \int_X \nu(E_x) \, d\mu(x) &= \sum_{n=1}^\infty \int_X \nu((A_n \times B_n)_x) d\mu \\ &= \sum_{n=1}^\infty \mu(A_n) \times \nu(B_n) \\ &= (\mu \times \nu)(E). \end{aligned}$$

Therefore, (18) holds if E is an $\mathcal{R}_\sigma$ set. Now consider a set E in $\mathcal{R}_{\sigma\delta}$. There is a descending collection $\{E_n\}_{n=1}^\infty$ of sets in $\mathcal{R}_\sigma$ whose intersection is E. For each x, E_x is ν-measurable, since it is the intersection of the countable collection of ν-measurable sets $\{(E_n)_x\}_{n=1}^\infty$. It remains to establish the integral equality (18). Since $(\mu\times\nu)^*(E)=(\mu\times\nu)(E)<\infty$, we may assume $(\mu\times\nu)(E_1)<\infty$, and therefore, by the continuity of the measure $\mu\times\nu$,

$$\lim_{n\to\infty}(\mu\times\nu)(E_n)=(\mu\times\nu)(E). \tag{23}$$

Since $E_1\in\mathcal{R}_\sigma$, by the preceding case, the function $x\mapsto\nu((E_1)_x)$ is defined on all of X and is integrable. Consequently, for almost all x, $\nu((E_1)_x)<\infty$ and so, by the continuity of the measure ν,

$$\lim_{n\to\infty}\nu((E_n)_x)=\nu(E_x).$$

The function $x\mapsto\nu((E_1)_x)$ is a non-negative, μ-integrable function that, for each n, dominates the function $x\mapsto\nu((E_n)_x)$. Therefore, by the Dominated Convergence Theorem, the validity of (18) for each $\mathcal{R}_\sigma$ set E_n, and (23),

$$\begin{aligned}\int_X \nu(E_x)\,d\mu(x) &= \lim_{n\to\infty}\int_X \nu((E_n)_x)d\mu\\ &= \lim_{n\to\infty}(\mu\times\nu)(E_n)\\ &= (\mu\times\nu)(E).\end{aligned}$$

□

Lemma 23 *Let $(X,\mathcal{A},\mu)$ and $(Y,\mathcal{B},\nu)$ be measure spaces and $(Y,\mathcal{B},\nu)$ be complete. If $E\subseteq X\times Y$ and $(\mu\times\nu)(E)=0$, then for almost all x, E_x is ν-measurable, $\nu(E_x)=0$, and* (18) *holds.*

Proof Since $(\mu\times\nu)(E)<\infty$, there is a set $E'\subseteq X\times Y$ for which (22) holds. By the excision property of measure, $(\mu\times\nu)(E')=(\mu\times\nu)(E)=0$. According to the preceding lemma, for all $x\in X$, E'_x is ν-measurable, the function $x\to\nu(E_x)$ is a μ-measurable, non-negative and

$$(\mu\times\nu)(E')=\int_X \nu(E'_x)\,d\mu(x)=0.$$

Therefore, $\nu(E'_x)=0$ for almost all $x\in X$. However, for all $x\in X$, $E_x\subseteq E'_x$. Therefore, by the completeness of ν, for almost all $x\in X$, E_x is ν-measurable and $\nu(E_x)=0$. Consequently, (18) holds. □

Proof of the Fubini-Tonelli Lemma Let $E'\in\mathcal{R}_{\sigma,\delta}$ be a set for which (22) holds. Since $(\mu\times\nu)(E'\sim E)=0$, by the preceding lemma, for almost all x, $\nu((E'\sim E)_x)=0$, and so

$$\nu(E'_x)=\nu(E_x)+\nu((E'\sim E)_x)=\nu(E_x) \text{ for almost all } x\in X.$$

The function $x\mapsto\nu(E_x)$ is μ-integrable, since it agrees almost everywhere with the function $x\mapsto\nu(E'_x)$, which, according to Lemma 22 is μ-measurable. By the excision property of measure, $(\mu\times\nu)(E)=(\mu\times\nu)(E')$, and therefore

$$(\mu\times\nu)(E)=(\mu\times\nu)(E')=\int_X \nu(E'_x)\,d\mu(x)=\int_X \nu(E_x)\,d\mu(x).$$

□

By the Tonelli and Fubini Theorems for the product of $(\mathbf{N},\, 2^{\mathbf{N}},\, c)$ with itself, we have the following corollary.

Corollary 24 *Let $\{a_{i,\,j}\}_{1\leq j,k<\infty}$ be a collection of real-numbers. The following switch in the order of summation holds provided that either each $a_{i,\,j}\geq 0$ or $\sum\limits_{j=1}^{\infty}\left(\sum\limits_{i=1}^{\infty}|a_{i,\,j}|\right)<\infty$:*

$$\sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty}a_{i,\,j}\right)=\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty}a_{i,\,j}\right).$$

PROBLEMS

In the following problems, for $X=Y=\mathbf{N}$, with respect to measurability and integration, only the counting measure space $(\mathbf{N},\, 2^{\mathbf{N}},\, c)$ is considered.

41. Let $(X,\,\mathcal{A},\,\mu)=(Y,\,\mathcal{B},\,\nu)=(\mathbf{N},\,2^{\mathbf{N}},\,c)$. Show that the product measure is the counting measure. State the Fubini and Tonelli Theorems explicitly for this case.
42. Show that the smallest σ-algebra that contains all products of Lebesgue measurable subsets of $\mathbf{R}$ is the Borel σ-algebra $\mathcal{B}(\mathbf{R}^2)$.
43. Assume $(X,\,\mathcal{A},\mu)$ and $(Y,\,\mathcal{B},\nu)$ are complete, $h\colon X\to[0,\,\infty]$ and $g\colon Y\to[0,\,\infty]$ are nonnegative and integrable, and define $k(x,y)=h(x)\cdot g(y)$. Show that k is a $(\mu\times\nu)$-integrable and
$$\int_{X\times Y}k\;d(\mu\times\nu)=\int_X h\;d\mu\cdot\int_Y g\;d\nu.$$
44. Define $f\colon\mathbf{N}\times\mathbf{N}\to\mathbf{R}$ by setting
$$f(m,\,n)=\begin{cases}2-2^{-m} & \text{if } m=n\\ -2+2^{-m} & \text{if } m=n+1\\ 0 & \text{otherwise.}\end{cases}$$
Show that f is measurable with respect to the product measure space. Also show that
$$\int_{\mathbf{N}}\left[\int_{\mathbf{N}}f(m,\,n)\;dc(m)\right]dc(n)\neq\int_{\mathbf{N}}\left[\int_{\mathbf{N}}f(m,\,n)\;dc(n)\right]dc(m).$$
Is this a contradiction either of Fubini's Theorem or Tonelli's Theorem?
45. Let $(X,\,\mathcal{A},\,\mu)$ be a complete measure space. Consider $\mathbf{N}\times X$ with the product measure $c\times\mu$.
    (i) Show that a subset E of $\mathbf{N}\times X$ is measurable with respect to the product measure space if and only if for each k, $E_k=\{x\in X\,|\,(k,\,x)\in E\}$ is measurable with respect to μ.
    (ii) Show that a function $f\colon\mathbf{N}\times X\to\mathbf{R}$ is measurable with respect to the product measure space if and only if for each k, $f(k,\,\cdot)\colon X\to\mathbf{R}$ is measurable with respect to μ.

(iii) Show that a function $f\colon \mathbf{N}\times X \to \mathbf{R}$ is integrable over $\mathbf{N}\times X$ with respect to $c\times\mu$ if and only if for each k, $f(k,\,\cdot)\colon X\to\mathbf{R}$ is integrable over X with respect to μ and

$$\sum_{k=1}^{\infty}\int_X |f(k,\,x)|\,d\mu(x) < \infty.$$

(iv) Show that if the function $f\colon \mathbf{N}\times X\to\mathbf{R}$ is integrable over $\mathbf{N}\times X$ with respect to $c\times\mu$, then

$$\int_{\mathbf{N}\times X} f\,d(c\times\mu) = \sum_{k=1}^{\infty}\int_X f(k,\,x)\,d\mu(x) < \infty.$$

46. Prove that the conclusion of Tonelli's Theorem holds if one of the spaces is the space $(\mathbf{N},\,2^{\mathbf{N}},\,c)$ and the other space is a complete measure space that need not be σ-finite.

47. If $\{(X_k, \mathcal{A}_k, \mu_k)\}_{k=1}^n$ is a finite collection of measure spaces, define the product measure $\mu_1\times\cdots\times\mu_n$ on the space $X_1\times\cdots\times X_n$ by starting with the semi-ring of rectangles of the form $R = A_1\times\cdots\times A_n$, define $\mu(R)=\prod\mu_k(A_k)$, show that μ is a premeasure and define the product measure to be the Carathéodory extension of μ. Show that if $(X_1\times\cdots\times X_p)\times(X_{p+1}\times\cdots\times X_n)$ is identified with $(X_1\times\cdots\times X_n)$, then

$$(\mu_1\times\cdots\times\mu_p)\times(\mu_{p+1}\times\cdots\times\mu_n) = \mu_1\times\cdots\times\mu_n.$$

48. A measure space $(X,\,\mathcal{M},\,\mu)$ such that $\mu(X)=1$ is called a probability measure space. Let $\{(X_\lambda, \mathcal{A}_\lambda, \mu_\lambda)\}_{\lambda\in\Lambda}$ be a collection of probability measure spaces parametrized by the set Λ. Show that it is possible to define a probability measure

$$\mu = \prod_{\lambda\in\Lambda}\mu_\lambda$$

on a suitable σ-algebra on the Cartesian product $\prod_{\lambda\in\Lambda} X_\lambda$ so that

$$\mu(A) = \prod_{\lambda\in\Lambda}\mu_\lambda(A_\lambda)$$

when $A=\prod_{\lambda\in\Lambda} A_\lambda$. (Note that $\mu(A)$ can only be non-zero if all but a countable number of the A_λ have $\mu(A_\lambda)=1$.)

11.6 PRODUCTS OF LEBESGUE MEASURE ON EUCLIDEAN SPACES: CAVALIERI'S PRINCIPLE

Recall from the preceding chapter that for each n, $(\mathbf{R}^n,\,\mathcal{L}^n,\,\mu_n)$ denotes the Lebesgue measure space on $\mathbf{R}^n$. Denote the collection of bounded intervals in $\mathbf{R}^n$ by $\mathcal{I}_n$. For each m and k, by suitably labeling coordinates, we identify $\mathbf{R}^m\times\mathbf{R}^k$ with $\mathbf{R}^{m+k}$. Moreover, each interval in $\mathcal{I}_{m+k}$ is of the form $I\times J$, where $I\in\mathcal{I}_m$ and $J\in\mathcal{I}_k$, and

$$\mu_{m+k}(I\times J) = \mu_m(I)\cdot\mu_k(J). \tag{24}$$

Lemma 25 *If $A\in\mathcal{L}^m$, $B\in\mathcal{L}^k$, $\mu_m(A)<\infty$, and $\mu_k(B)<\infty$, then $A\times B\in\mathcal{L}^{m+k}$ and*

$$\mu_{m+k}(A\times B) = \mu_m(A)\cdot\mu_k(B). \tag{25}$$

Proof First consider the case that either $\mu_m(A) = 0$ or $\mu_k(B) = 0$. Assume, say, that $\mu_m(B) = 0$. The three Lebesgue measures are restrictions of corresponding outer-measures. Let $\epsilon > 0$. Choose covers $\{I^i\}_{i=1}^{\infty} \subseteq \mathcal{I}_m$ of A and $\{J^j\}_{j=1}^{\infty} \subseteq \mathcal{I}_k$ to B for which

$$\sum_{i=1}^{\infty} \mu_m(I^i) < \mu_m(A) + \epsilon \text{ and } \sum_{j=1}^{\infty} \mu_k(J^j) < \epsilon.$$

The countable collection of measurable rectangles $\{I^i \times J^j\}_{1 \leq i,j < \infty}$ covers $A \times B$, and therefore, by the definition of outer-measure,

$$\begin{aligned} \mu^*_{m+k}(A \times B) &\leq \textstyle\sum_{1 \leq, i,j < \infty} \mu_{m+k}(I^1 \times J^j) \\ &= \textstyle\sum_{1 \leq i,j < \infty} \mu_m(I^i) \cdot \mu_k(J^j) \\ &= \sum_{j=1}^{\infty} \left[\sum_{i=1}^{\infty} \mu_m(I^i) \right] \cdot \mu_k(J^j) \\ &\leq (\mu_m(A) + \epsilon) \cdot \epsilon. \end{aligned}$$

Consequently, $\mu^*_{m+k}(A \times B) = 0$, so that $A \times B \in \mathcal{L}^{m+k}$ and (25) holds. Now consider general sets A and B. To show that $A \times B \in \mathcal{L}^{m+k}$, recall that, according to Theorem 5 of the preceding chapter, a subset of Euclidean space is Lebesgue measurable if and only if it is a G_δ set from which a set of Lebesgue measure zero has been excised. Let G' and G'' be G_δ subsets on $\mathbf{R}^m$ and $\mathbf{R}^k$, respectively, and E' and E'' be subsets of $\mathbf{R}^m$ and $\mathbf{R}^k$, respectively, of Lebesgue outer-measure zero, for which

$$A = G' \sim E' \text{ and } B = G'' \sim E''.$$

Then

$$A \times B = G' \times G'' \sim [(G' \times E'') \cup (E' \times G'')].$$

Therefore, $A \times B$ is a G_δ subset of $\mathbf{R}^{m+k}$, $G' \times G''$, from which, by the first case, a set of product measure zero, $(G' \times E'') \cup (E' \times G'')$, has been excised, and consequently $A \times B \in \mathcal{L}^{m+k}$. Fix a rectangle $J \in \mathcal{I}_k$. The map $A \mapsto \mu_{m+k}(A \times J)$ is properly defined on $\mathcal{L}^m$, and is a measure that agrees with the measure $A \mapsto \mu_m(A) \cdot \times \mu_k(J)$ on $\mathcal{I}_m$. Since Lebesgue measure on a Euclidean space is σ-finite, by the uniqueness assertion of the Carathéodory-Hahn Theorem, these two measure are equal on $\mathcal{L}^k$, that is, (25) holds for $A \in \mathcal{L}_m$ and $B = J \in \mathcal{I}_k$. Similarly, fixing $A \in \mathcal{L}_m$ and defining the measure $B \mapsto \mu_{m+k}(A \times B)$ for $B \in \mathcal{L}_k$, it follows that (25) holds in general. □

Theorem 26 *The Lebesgue measure space* $(\mathbf{R}^{m+k}, \mathcal{L}^{m+k}, \mu_{m+k})$ *is the product of the Lebesgue measure spaces* $(\mathbf{R}^m, \mathcal{L}^m, \mu_m)$ *and* $(\mathbf{R}^k, \mathcal{L}^k, \mu_k)$.

Proof By an appeal to the preceding lemma, the two maps $\mu_m \times \mu_k$ and μ_{k+m} are measures on $\mathcal{L}^{m+k}$ that agree on

$$\mathcal{M} = \left\{ A \times B \;\middle|\; A \in \mathcal{L}^m, B \in \mathcal{L}^k, \mu_m(A) < \infty, \mu_k(B) < \infty \right\}.$$

But $\mu_{m+k}\colon \mathcal{L}_{m+k} \to [0, \infty]$ is the Carathéodory extension of $\mu_{m+k}\colon \mathcal{M} \to [0, \infty]$. The measure space $(\mathbf{R}^{m+k}, \mathcal{L}^{m+k}, \mu_{m+k})$ is σ-finite. Therefore, by the uniqueness assertion of the Carathéodory-Hahn Theorem, $\mu_m \times \mu_k = \mu_{k+m}$. $\square$

A simple criterion for verifying that a function on Euclidean space is measurable is to show that it is continuous on the complement of a set measure zero. Of course, to verify the integrability of a slice means one has to verify integrability of its absolute value. The conclusion of Tonellli's Theorem is clear: if one of the three integrals is finite, then so are the other two, and all three are equal. The following two examples show that care is needed in applying Fubini's Theorem. The first shows that one of the iterated integrals may be properly defined and finite for each value of a variable, but the other iterated integral may fail to be defined for any value of the variable.

Example Define $f\colon [-1, 1] \times [-1, 1]$ by

$$f(x, y) = \begin{cases} x/y & \text{if } y \neq 0 \\ 0 & \text{if } y = 0. \end{cases}$$

This function is measurable, since it is continuous except on a set of μ_2-measure zero. Fix y. The function $x \mapsto f(x, y)$ is odd and continuous, so that

$$\int_{-1}^{1} \left[\int_{-1}^{1} f(x, y)\, dx \right] dy = 0.$$

On the other hand, for each $x \in [-1, 1]$, the function $y \mapsto |f(x, y)|$ is not integrable over $[-1, 1]$.

The second example shows that both iterated integrals may be properly defined and equal, and yet the function is not integrable.

Example Define $f\colon [-1, 1] \times [-1, 1]$ by

$$f(x, y) = \begin{cases} 6xy/(x^2+y^2)^4 & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

This function is measurable, since it is continuous except at $(0, 0)$. Fixing one of the variables, the function is odd in the other, and one checks all of the slices of $|f(x, y)|$ are integrable. Therefore,

$$\int_{-1}^{1} \left[\int_{-1}^{1} f(x, y)\, dx \right] dy = 0 = \int_{-1}^{1} \left[\int_{-1}^{1} f(x, y)\, dy \right] dx.$$

On the other hand, f is not integrable over $[-1, 1] \times [-1, 1]$, since if it were, it would be integrable over $[0, 1] \times [0, 1]$. However, for each $y \in [0, 1]$, by choosing an anti derivative, one sees that $\int_0^1 f(x, y)\, dx = y^{-5} - y(1+y^2)^{-3}$, and the integral of this function is ∞. It follows from Tonelli's Theorem that f is not integrable over $[0, 1] \times [0, 1]$.

Theorem 27 (**Cavalieri's Principle**) *Let E' and E'' be bounded, Lebesgue measurable subsets of* $\mathbf{R}^3$. *Assume that*

$$\text{area}(E'_z) = \text{area}(E''_z) \textit{ for all } z.$$

Then

$$\text{vol}(E') = \text{vol}(E'').$$

Proof Assume that E' and E'' are contained in $\mathbf{R}^2 \times [c, d]$. By the preceding theorem, Lebesgue measure on $\mathbf{R}^3$ is the product of Lebesgue measure on $\mathbf{R}^2$ and on $\mathbf{R}^1$. Therefore, by the Fubini-Tonelli Lemma,

$$\text{vol}(E') = \int_c^d \text{area}(E'_z)\, dz = \int_c^d \text{area}(E''_z)\, dz = \text{vol}(E''). \qquad \square$$

We use the informal words "area" for Lebesgue measure on $\mathbf{R}^2$ and "volume" for Lebesgue measure on $\mathbf{R}^3$ in order to be reminded of the ancient geometric intuition behind this formula. This principle[4] is named after Bonaventura Cavalieri, a student of Galileo, who, in the seventeenth century, considered concepts of infinitesmal approximation, one reason being to clarify the geometric understanding of this principle by Archimedes who, two millennia earlier, concluded from it that the volume of a ball of radius r in $\mathbf{R}^3$ is $\frac{4}{3}\pi r^3$.

PROBLEMS

49. Is every continuous real-valued function $k\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ measurable with respect to the product measure space? Show that for any subset A of $\mathbf{R}$, the set $A \times Q$ is measurable with respect to the product measure, irrespective of the measurability of A.
50. Suppose that the function $f\colon \mathbf{R}^2 \to \mathbf{R}$ is continuous with respect to each variable separately, show that it is Lebesgue measurable with respect to μ_2.
51. If $f\colon \mathbf{R}^m \to \mathbf{R}$ and $g\colon \mathbf{R}^k \to \mathbf{R}$ are Lebesgue measurable, show that the function $(x, y) \mapsto f(x) \cdot g(y)$ is Lebesgue measurable on $\mathbf{R}^{m+k}$.
52. Verify that

$$\int_{-1}^{1} \left[\int_{-1}^{1} \frac{x^2 - y^2}{(x^2+y^2)^2}\, dx\right] dy = -\frac{\pi}{4} \text{ and } \int_{-1}^{1} \left[\int_{-1}^{1} \frac{x^2 - y^2}{(x^2+y^2)^2}\, dy\right] dx = \frac{\pi}{4},$$

by checking that, for $(x, y) \neq (0, 0)$,

$$\frac{\partial}{\partial y}\left[\frac{y}{x^2+y^2}\right] = \frac{x^2-y^2}{(x^2+y^2)^2} = \frac{\partial}{\partial x}\left[\frac{-x}{x^2+y^2}\right].$$

Explain why this example does not contradict either Tonelli's Theorem or Fubini's Theorem.

53. Show that

$$\int_{-1}^{1} \left[\int_{-1}^{1} \frac{xy}{(x^2+y^2)^2}\, dx\right] dy \neq \int_{-1}^{1} \left[\int_{-1}^{1} \frac{xy}{(x^2+y^2)^2}\, dy\right] dx.$$

Explain why this does not contradict either Tonelli's Theorem or Fubini's Theorem.

[4] As one of many very interesting topics, the history and applications of this principle, including Archimedes' geometric arguments, may be found in George Simmons' *Gems of Calculus: Brief Lives and Memorable Mathematics.*

54. Define $f(x, y) = y\exp(-(1+x^2)y^2)$ for $x \geq 0, y \geq 0$. Show that

$$\int_0^\infty \left[\int_0^\infty f(x, y)\,dy\right] dx = \int_0^\infty \left[\int_0^\infty f(x, y)dx\right] dy,$$

and conclude from this that
$$\int_0^\infty \exp(-x^2)dx = \sqrt{\pi}/2.$$

55. Define $f(x, y) = \exp(-xy^2) \cdot \sin x$ for $x \geq 0, y \geq 0$. Show that for each $b > 0$

$$\int_0^b \left[\int_0^\infty f(x, y)\,dy\right] dx = \int_0^\infty \left[\int_0^b f(x, y)\,dx\right] dy,$$

and conclude from this that, as an improper Riemann integral,

$$\int_0^\infty \frac{\sin x}{x^{1/2}}\,dx \equiv \lim_{b\to\infty} \int_0^b \frac{\sin x}{x^{1/2}}\,dx = \sqrt{\pi/2}.$$

56. Assume that $f\colon \mathbf{R}^n \to [0, \infty)$ is finitely supported and Lebesgue integrable, and that $g\colon \mathbf{R} \to \mathbf{R}$ is increasing, bounded, absolutely continuous and $g(0) = 0$. Show that $g \circ f\colon \mathbf{R}^n \to \mathbf{R}$ is Lebesgue integrable, and that

$$\int_{\mathbf{R}^n} g(f(x))\,d\mu_n(x) = \int_0^\infty \left[\mu_n \left\{x \in \mathbf{R}^n \mid 0 \leq t \leq f(x)\right\} \cdot g'(t)\right] dt.$$

Suggestion: Define $E = \{(x, t) \in \mathbf{R}^n \times [0, \infty) \mid 0 \leq t \leq g(f(x))\}$, and verify the following:

$$\int_{\mathbf{R}^n} g(f(x))\,d\mu_n(x) = (\mu_n \times m)(E) = \int_{\mathbf{R}^n} m(E_x)\,d\mu_n(x);$$

$$m(E_x) = g(f(x)) = \int_0^{f(x)} g'(t)\,dt = \int_0^\infty \chi_{[0,\, f(x)]} \cdot g'(t)\,dt$$

57. Show that the product of semi-rings is a semi-ring. Use this to show that the collection of bounded intervals in $\mathbf{R}^n$ is a semi-ring.
58. Let the mapping $\mathcal{N}\colon \mathbf{R}^n \to \mathbf{R}^n$ be one-to-one and onto and there is a $c > 0$ such that

$$\|\mathcal{N}(x) - \mathcal{N}(y)\| \geq c \cdot \|x - y\| \;\forall x, y \in \mathbf{R}^n.$$

Show that if $f\colon \mathbf{R}^n \to \mathbf{R}$ is Lebegue measurable, so is $f \circ \mathcal{N}$.

CHAPTER 12

General L^p Spaces: Completeness, Convolution, and Duality

Contents

12.1 The Spaces $L^p(X, \mu)$, $1 \le p \le \infty$ 238
12.2 Convolution, Smooth Approximation, and a Smooth Urysohn's Lemma . . . 243
12.3 The Riesz Representation Theorem for the Dual of $L^p(X, \mu)$, $1 \le p < \infty$ . . 252
12.4 Weak Sequential Compactness in $L^p(X, \mu)$, $1 < p < \infty$ 255
12.5 The Kantorovitch Representation Theorem for the Dual of $L^\infty(X, \mu)$ 257

For a measure space $(X, \mathcal{M}, \mu)$ and $1 \le p \le \infty$, the normed linear spaces $L^p(X, \mu)$ are defined just as they were in Chapter 7 for the case of Lebesgue measure on $\mathbf{R}$. Arguments very similar to those used in this special case show that the Hölder and Minkowski Inequalities hold and that $L^p(X, \mu)$ is complete. In the first section, these and related topics are covered. In the second section, we use Tonelli's Theorem to consider the convolution of two functions on $\mathbf{R}^n$, and prove Young's Convolution Inequality, from which we conclude that for $1 \le p < \infty$, the subspace of smooth, compactly supported functions is dense in $L^p(\mathbf{R}^n, \mu_n)$. If $(X, \mathcal{M}, \mu)$ is σ-finite and $1 \le p < \infty$, the Riesz Representation Theorem for the dual of $L^p(X, \mu)$ is proved, which makes it possible to consider weak sequential convergence in these spaces.

12.1 THE SPACES $L^p(X, \mu)$, $1 \le p \le \infty$

Let $(X, \mathcal{M}, \mu)$ be a measure space and $1 \le p < \infty$. Define $L^p(X, \mu)$ to be the collection of measurable functions $f\colon X \to \overline{\mathbf{R}}$ for which

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

According to Proposition 8 in the preceding chapter, a function in $L^p(X, \mu)$ is finite almost everywhere and therefore, by suitable identification, may be regarded as being real-valued. As observed earlier, for real numbers a, b, since $|a+b| \le |a| + |b| \le 2 \cdot \max\{|a|, |b|\}$ if $1 \le p < \infty$, then

$$|a+b|^p \le 2^p[|a|^p + |b|^p]. \tag{1}$$

Consequently, $L^p(X, \mu)$ is a linear space. For $f \in L^p(X, \mu)$ define

$$\|f\|_p = \left[\int_X |f|^p \, d\mu\right]^{1/p}.$$

A measurable function $f\colon X \to \overline{\mathbf{R}}$ is said to be **essentially bounded** provided that there is an $M \ge 0$, called an **essential upper bound** for f, for which

$$|f| \le M \text{ almost everywhere on } X.$$

An essentially bounded function may be identified with a measurable real-valued function that is bounded on all of X. Define $L^\infty(X, \mu)$ to be the linear space of essentially bounded functions, and for $f \in L^\infty(X, \mu)$, define $\|f\|_\infty$ to be the infimum of the essential upper bounds for f, which, in fact, is a minimum. The triangle inequality for real numbers implies the triangle inequality for $L^\infty(X, \mu)$. Recall that the conjugate q of a number p in $(1, \infty)$ is the unique number q in $(1, \infty)$ for which $1/p + 1/q = 1$; 1 is called the conjugate of ∞ and ∞ the conjugate of 1.

Theorem 1 *Let $(X, \mathcal{M}, \mu)$ be a measure space, $1 \leq p \leq \infty$, and q be the conjugate of p. If $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$, then $f \cdot g \in L^1(X, \mu)$ and*

Hölder's Inequality

$$\int_X |f \cdot g| \, d\mu = \|f \cdot g\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

Moreover, if $p < \infty$ and $f \neq 0$, the function $f^* = \|f\|_p^{1-p} \cdot \operatorname{sgn}(f) \cdot |f|^{p-1}$, called the dual or conjugate of f, belongs to $L^q(X, \mu)$,

$$\int_X f \cdot f^* \, d\mu = \|f\|_p \textit{ and } \|f^*\|_q = 1. \tag{2}$$

If $f, g \in L^p(X, \mu)$, then $f + g \in L^p(X, \mu)$, and

Minkowski's Inequality

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof In the case $p = 1$ or $p = \infty$, Hölder's Inequality is clear. Moreover, if $p = 1$, $f^* = \operatorname{sgn}(f)$ and therefore (2) holds. Now consider $1 < p < \infty$. Assume that $f \neq 0$ and $g \neq 0$, for otherwise there is nothing to prove. By the positive homogeneity of the norm, it is clear that if Hölder's Inequality holds if f is replaced by its normalization $f/\|f\|_p$ and g is replaced by its normalization $g/\|g\|_q$, then it holds in general. Assume that $\|f\|_p = \|g\|_q = 1$, that is,

$$\int_X |f|^p \, d\mu = 1 \text{ and } \int_X |g|^q \, d\mu = 1,$$

in which case Hölder's Inequality becomes

$$\int_X |f \cdot g| \, d\mu \leq 1.$$

Assume that f and g are finite on X, so that, by Young's Inequality,

$$|f \cdot g| = |f| \cdot |g| \leq \frac{|f|^p}{p} + \frac{|g|^q}{q} \text{ on } X.$$

It follows from the linearity of integration and the integral comparison test that $f \cdot g$ is integrable and

$$\int_X |f \cdot g| \, d\mu \leq \frac{1}{p} \int_X |f|^p \, d\mu + \frac{1}{q} \int_X |g|^q \, d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

To prove (2), observe that

$$f \cdot f^* = \|f\|_p^{1-p} \cdot |f|^p \text{ on } X.$$

Therefore,

$$\int_X f \cdot f^* \, d\mu = \|f\|_p^{1-p} \cdot \int_X |f|^p \, d\mu = \|f\|_p^{1-p} \cdot \|f\|_p^p = \|f\|_p.$$

Since $q(p-1) = p$, $\|f^*\|_q = 1$. To verify the triangle inequality for $L^p(X, \mu)$, that is, Minkowski's Inequality, first observe that it follows from inequality (1) that $f + g \in L^p(X, \mu)$. If $f + g = 0$, there is nothing to prove. So assume that $f + g \neq 0$, in which case, since $1 \leq p < \infty$, its dual function $(f+g)^* \in L^q(X, \mu)$ is defined. By the definition of dual function and Hölder's Inequality,

$$\begin{aligned}
\|f+g\|_p &= \int_X (f+g) \cdot (f+g)^* \, d\mu \\
&= \int_X f \cdot (f+g)^* \, d\mu + \int_X g \cdot (f+g)^* \, d\mu \\
&\leq \|f\|_p \cdot \|(f+g)^*\|_q + \|g\|_p \cdot \|(f+g)^*\|_q \\
&= \|f\|_p + \|g\|_p.
\end{aligned}$$

□

Proposition 2 *Let $(X, \mathcal{M}, \mu)$ be a finite measure space. If $1 \leq p_1 < p_2 \leq \infty$, then $L^{p_2}(X) \subseteq L^{p_1}(X)$. Moreover, there is a $c \geq 0$ for which*

$$\|f\|_{p_1} \leq c\|f\|_{p_2} \text{ for all } f \in L^{p_2}(E). \tag{3}$$

Proof If $p_2 = \infty$, define $c = \mu(X)$. Assume that $p_2 < \infty$. Define $p = p_2/p_1 > 1$ and let q be the conjugate of p. Let $f \in L^{p_2}(X)$. Observe that $f^{p_1} \in L^p(X)$, and $g = \chi_X \in L^q(X)$, since $\mu(X) < \infty$. Apply Hölder's Inequality to obtain

$$\int_X |f|^{p_1} d\mu = \int_X [|f|^{p_1} \cdot g] d\mu \leq [\|f\|_{p_2}]^{p_1} \cdot \left[\int_X |g|^q \, d\mu\right]^{1/q} = [\|f\|_{p_2}]^{p_1} \, [m(X)]^{1/q}.$$

Take the $1/p_1$-th power of each side to obtain (3), where $c = [m(X)]^{\frac{p_2 - p_1}{p_1 p_2}}$. □

Proposition 3 *Let $(X, \mathcal{M}, \mu)$ be a measure space and $1 \leq p \leq \infty$. Then the subspace of simple functions is dense in $L^p(X, \mu)$.*

Proof The case $p = \infty$ follows from the Simple Approximation Lemma. Assume that $1 \leq p < \infty$ and let $f \in L^p(X, \mu)$. Since f is measurable, according to the Simple Approximation Theorem, there is a sequence $\{\psi_n\}$ of simple functions that converges pointwise on X to f and for which $|\psi_n| \leq |f|$ on X for all n. Observe that

$$|\psi_n - f|^p \leq 2^{p+1} \cdot |f|^p \text{ on } X \text{ for all } n.$$

Since $|f|^p$ is integrable, according to the Dominated Convergence Theorem, $\{\psi_n\} \to f$ in $L^p(X, \mu)$. □

The proof of the following consequence of the Vitali Convergence Theorem is left as an exercise.

The Vitali $\mathbf{L^p}$ Convergence Criterion *Let $(X, \mathcal{M}, \mu)$ be a measure space and $1 \le p < \infty$. Let $\{f_n\}$ be a sequence in $L^p(X, \mu)$ that converges pointwise to $f \in L^p(X, \mu)$. Then $\{f_n\} \to f$ in $L^p(X, \mu)$ if and only if $\{|f_n|^p\}$ is uniformly integrable and tight.*

The above criterion for convergence in $L^p(X, \mu)$ has the following companion criterion for membership in this space.

Theorem 4 *Let $(X, \mathcal{M}, \mu)$ be σ-finite, $1 \le p < \infty$, q be the conjugate of p, and the function $f\colon X \to \mathbf{R}$ be integrable. Assume that for some $M \ge 0$,*

$$\left| \int_X f \cdot \varphi \, d\mu \right| \le M \|\varphi\|_q \text{ if } \varphi \text{ is finitely supported and simple.} \tag{4}$$

Then $f \in L^p(X, \mu)$ and $\|f\|_p \le M$. Furthermore, if $f \ge 0$, then it suffices to assume that f is measurable and that (4) holds for non-negative functions.

Proof If φ is finitely supported and simple, then so is $\operatorname{sgn}(f) \cdot \varphi$, $\|\operatorname{sgn}(f) \cdot \varphi\|_q = \|\varphi\|_q$ and $|f| \cdot \varphi = f \cdot [\operatorname{sgn}(f) \cdot \varphi]$. Therefore, (4) holds if f is replaced by $|f|$, so we may assume that $f \ge 0$. First consider the case $p = 1$. Assume the conclusion does not hold, in which case there is an $\alpha > 0$ for which if $E_\alpha = \{x \in X \mid f(x) \ge M + \alpha\}$, then $\mu(E_\alpha) > 0$. Since μ is σ-finite, we may assume that $\mu(E_\alpha) < \infty$. Letting $\varphi = \chi_{E_\alpha}$, we contradict (4) for $q = \infty$. Now consider $p > 1$. Since f is non-negative and measurable, and μ is σ-finite, by conclusions (i) and (ii) of the Simple Approximation Theorem, there is an increasing sequence $\{\varphi_k\colon X \to [0, \infty)\}$ of non-negative, finitely supported, simple functions that converges pointwise to f. Observe that $\{\varphi_k^p\}$ is an increasing sequence of non-negative, measurable functions that converges pointwise to f^p and therefore, by the Monotone Convergence Theorem, to verify that $f \in L^p(X, \mu)$ and $\|f\|_p \le M$, it is sufficient to show that

$$\int_X [\varphi_k]^p \, d\mu \le M^p \text{ for all } k. \tag{5}$$

Fix k. Observe that

$$\varphi_k^p = \varphi_k \cdot \varphi_k^{p-1} \le f \cdot \varphi_k^{p-1},$$

and φ_k^{p-1} is non-negative, finitely supported and simple. Therefore, by assumption (4), since $p > 1$, and $(p-1)q = p$,

$$\int_X [\varphi_k]^p \, d\mu \le M \cdot \left[\int_X [\varphi_k]^{(p-1)q} \, d\mu \right]^{1/q} = M \cdot \left[\int_X [\varphi_k]^p \, d\mu \right]^{1/q}.$$

Since $1 - 1/q = 1/p$, this is precisely the inequality (5). □

We now turn to the proof that each normed linear space $L^p(X, \mu)$, $1 \le p \le \infty$, is complete, that is, every Cauchy sequence in $L^p(X, \mu)$ converges. To do so, recall that a sequence $\{v_k\}$ in a normed linear space V is said to be rapidly Cauchy provided that

$$\sum_{k=1}^{\infty} \|v_{k+1} - v_k\| < \infty.$$

The following theorem is the centerpiece of the proof that the spaces $L^p(X, \mu)$, $1 \le p \le \infty$, are complete.

Theorem 5 *Let $(X, \mathcal{M}, \mu)$ be a measure space and $1 \le p \le \infty$. If the sequence $\{f_n\}$ in $L^p(X, \mu)$ is rapidly Cauchy, then it converges in the $L^p(X, \mu)$ norm and also pointwise almost everywhere on X to a function f in $L^p(X, \mu)$. Moreover, $\{f_n\}$ is dominated by a function in $L^p(X, \mu)$, in the sense that there is a function w in $L^p(X, \mu)$ for which $|f_n| \le w$ almost everywhere on X for all n.*

Proof The case $p = \infty$ is left as an exercise. Assume that $1 \le p < \infty$. Let $\{f_n\}$ be a rapidly Cauchy sequence in $L^p(X, \mu)$. For notational convenience, by replacing each f_n by $f_n - f_1$, assume that $f_1 = 0$. Observe that for each n,

$$f_n = \sum_{k=1}^{n-1} [f_{k+1} - f_k],$$

and define

$$w_n = \sum_{k=1}^{n-1} |f_{k+1} - f_k|.$$

Since $\{w_n\}$ is increasing, the measurable function $w\colon X \to \overline{\mathbf{R}}$ is properly defined by

$$w(x) = \lim_{n\to\infty} w_n(x) \text{ for all } x \in X.$$

By Minkowski's Inequality, since $\{f_n\}$ is rapidly Cauchy in $L^p(X)$, for all n,

$$\|w_n\|_p \le \sum_{k=1}^{\infty} \|f_{k+1} - f_k\|_p = C < \infty.$$

Therefore, $\int_X w_n^p \le C^p$ for all n. According to Fatou's Lemma, $\int_X w^p \le C^p$. So $w \in L^p(X)$. In particular, w is finite almost everywhere on X. This means that for almost all $x \in X$, the series of real numbers $\sum_{k=1}^{\infty} |f_{k+1}(x) - f_k(x)|$ converges. Since a series of real numbers converges if it converges absolutely, for almost all $x \in X$, a real number $f(x)$ is defined by

$$f(x) \equiv \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \sum_{k=1}^{n-1} [f_{k+1}(x) - f_k(x)].$$

Define $f(x) = 0$ if this limit does not exist. Being the pointwise limit almost everywhere of a sequence of measurable functions, f is measurable. For all n,

$$|f_n| \le w_n \le w \text{ and } |f| \le w,$$

so that

$$|f_n - f|^p \leq 2^p[|f_n|^p + |f|^p] \leq 2^{p+1} \cdot w^p.$$

Since w^p is integrable and $\{[f_n - f]^p\}$converges pointwise almost everywhere on X to 0, it follows from this inequality and the Dominated Convergence Theorem that $\{f_n\} \to f$ in $L^p(X, \mu)$. Moreover, $\{f_n\}$ converges pointwise almost everywhere to f and is dominated by the function $w \in L^p(X, \mu)$. □

In Chapter 8, we proved that, in a normed linear space, each Cauchy sequence has a rapidly Cauchy subsequence and that a Cauchy sequence converges if it has a convergent subsequence. These two properties of Cauchy sequences, together with this theorem, suffice to prove the following fundamental result.

The Riesz-Fischer Theorem *If $(X, \mathcal{M}, \mu)$ is a measure space and $1 \leq p \leq \infty$, then the normed linear space $L^p(X, \mu)$ is complete. Moreover, if $\{f_n\} \to f$ in $L^p(X, \mu)$, then a subsequence of $\{f_n\}$ converges pointwise almost everywhere on X to f and is dominated by a function in $L^p(X, \mu)$.*

PROBLEMS

1. Consider the measure space $\{X, \emptyset\}$ with $\mu(X) = \infty$. For $1 \leq p \leq \infty$, characterize the space $L^p(X, \mu)$.
2. For $1 \leq p < \infty$ and an index n, define $f_n(x) = n^{1/p}$ if $0 \leq x \leq 1/n$ and $f_n(x) = 0$ if $1/n < x \leq 1$. Let f be identically zero on $[0, 1]$. Show that $\{f_n\}$ converges pointwise to f but does not converge in $L^p([0, 1], m)$. Why does the Vitali L^p Convergence Criterion not apply?
3. For $1 \leq p < \infty$ and an index n, let $f_n \colon \mathbf{R} \to \mathbf{R}$ be the characteristic function of $[n, n+1]$. Let f be identically zero on $\mathbf{R}$. Show that $\{f_n\}$ converges pointwise to f but does not converge in $L^p(\mathbf{R}, m)$. Why does the Vitali L^p Convergence Criterion not apply?
4. Prove the Vitali L^p Convergence Criterion.
5. For a measure space $(X, \mathcal{M}, \mu)$ and $0 < p < 1$, define $L^p(X, \mu)$ to be the collection of measurable functions on X for which $|f|^p$ is integrable. Show that $L^p(X, \mu)$ is a linear space. For $f \in L^p(X, \mu)$, define $\|f\|_p^p = \int_X |f|^p \, d\mu$. (In general, $\|\cdot\|_p$ is not a norm, since Minkowski's Inequality may fail.)
6. Let $\{f_n\}$ a Cauchy sequence in $L^\infty(X, \mu)$. Show that there is a measurable subset X_0 of X for which $\mu(X \sim X_0) = 0$ and for each $\epsilon > 0$, there is an index N for which

   $$|f_n - f_m| \leq \epsilon \text{ on } X_0 \text{ for all } n, m \geq N.$$

   Use this to show that $L^\infty(X, \mu)$ is complete.
7. Let c be the counting measure on $\mathbf{N}$. For $1 \leq p \leq \infty$, show that $L^p(\mathbf{N}, c) = \ell^p$. Then show that ℓ^p is complete.

12.2 CONVOLUTION, SMOOTH APPROXIMATION, AND A SMOOTH URYSOHN'S LEMMA

In the context of Lebesgue integration on Euclidean spaces, we now consider an operation on pairs of functions called convolution, and use this operation to prove, among other things, that, for $1 \leq p < \infty$, the compactly supported, infinitely differentiable functions are a dense subspace of $L^p(\mathbf{R}^n, \mu_n)$. For notational clarity, because we consider iterated integrals, we here denote $d\mu_n$ by dx and also by dy. Lebesgue measure is σ-finite, and it is complete, since it is constructed as a Carathéodory extension. According to Theorem 26 of the preceding chapter, $\mu_{n+n} = \mu_n \times \mu_n$, as a product measure. We record the following special case of Tonelli's Theorem that is needed in this section to justify switching the order of integration.

Theorem 6 *If the non-negative function $k\colon \mathbf{R}^n \times \mathbf{R}^n \to [0, \infty]$ is Lebesgue measurable, then the functions $x \mapsto \int_{\mathbf{R}^n} k(x,y)dy$ and $y \mapsto \int_{\mathbf{R}^n} k(x,y)\,dx$ are defined almost everywhere on $\mathbf{R}^n$, are Lebesgue measurable, and*

$$\int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,y)\,dx\right]dy = \int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,y)\,dy\right]dx.$$

Minowski's Integral Inequality *Let $1 \leq p < \infty$ and the non-negative function $k\colon \mathbf{R}^n \times \mathbf{R}^n \to [0, \infty)$ be Lebesgue measurable. Then*

$$\left[\int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,y)\,dy\right]^p dx\right]^{1/p} \leq \int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,\,y)^p\,dx\right]^{1/p} dy.$$

Proof According to Tonelli's Theorem, the non-negative function $x \mapsto \int_{\mathbf{R}^n} k(x,\,y)\,dy$ is defined almost everywhere and is Lebesgue measurable and so, by Theorem 4, to verify this inequality, it is sufficient to show that for each non-negative, finitely supported, simple function $\varphi\colon \mathbf{R}^n \to [0, \infty)$,

$$\left|\int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,y)\,dy\right]\varphi(x)dx\right| \leq \int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,\,y)^p\,dx\right]^{1/p} dy\cdot\|\varphi\|_q. \tag{6}$$

However, for such a φ, by switching the order of integration and Hölder's Inequality,

$$\begin{aligned}\int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,y)\,dy\right]\varphi(x)dx &= \int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,y)\varphi(x)\,dy\right]dx\\ &= \int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,y)\varphi(x)\,dx\right]dy\\ &\leq \int_{\mathbf{R}^n}\left[\int_{\mathbf{R}^n} k(x,\,y)^p\,dx\right]^{1/p} dy\cdot\|\varphi\|_q.\end{aligned}$$

Therefore, (6) holds. □

Littlewood suggested that a measurable function is "nearly" continuous. One confirmation of this is Lusin's Theorem for Lebesgue measurable functions on Euclidean space. As an immediate corollary to this theorem (Corollary 6 in the preceding chapter), we proved that a Lebesgue measurable function on a subset of $\mathbf{R}^n$ is the pointwise limit almost everywhere of a sequence of continuous functions. This will be the basis of the proof of the following lemma. We denote by $C_c(\mathbf{R}^n)$ the space of continuous functions on $\mathbf{R}^n$ that are compactly supported.

Lemma 7 *If $1 \leq p < \infty$, then the subspace $C_c(\mathbf{R}^n)$ is dense in $L^p(\mathbf{R}^n, \mu_n)$.*

Proof According to Proposition 3, the simple functions are dense in $L^p(\mathbf{R}^n, \mu_n)$. Moreover, if $f \in L^p(\mathbf{R}^n, \mu_n)$, then $\lim_{k\to\infty} \int_{\|x\|\geq k} |f|^p \, d\mu = 0$. Therefore, to prove the result it suffices to let $\psi \in L^p(\mathbf{R}^n, \mu_n)$ be simple and such that there is a $c > 0$ for which $\psi(x) = 0$ if $\|x\| \geq c$, and show that there is a sequence in $C_c(\mathbf{R}^n)$ that converges in $L^p(\mathbf{R}^n, \mu_n)$ to ψ. For such a ψ, choose a sequence of continuous functions $\{h_k\colon [0, \infty) \to \mathbf{R}\}$ such that, for each k,

$$0 \leq h_k \leq 1, h(t) = 1 \text{ if } 0 \leq t \leq c \text{ and } h(t) = 0 \text{ if } t \geq c + 1/k.$$

As observed above, there is a sequence of continuous functions $\{g_k\colon \mathbf{R}^n \to \mathbf{R}\}$ that converges pointwise almost everywhere on $\mathbf{R}^n$ to ψ. Since ψ is bounded, we may assume that the sequence $\{g_k\}$ is uniformly pointwise bounded by M. For each k, define $f_k\colon \mathbf{R}^n \to \mathbf{R}$ by

$$f_k(x) = h_k(\|x\|) \cdot g_k(x) \text{ for all } x \in \mathbf{R}^n.$$

Observe that $\{f_k\colon \mathbf{R}^n \to \mathbf{R}\}$ is a sequence in $C_c(\mathbf{R}^n)$, also is uniformly pointwise bounded by M, and converges pointwise almost everywhere on $\mathbf{R}^n$ to ψ. For each k, if $A_k = \{x \in \mathbf{R}^n \mid c \leq \|x\| \leq c + 1/k\}$, then, by the choice of h_k,

$$\int_{\mathbf{R}^n} |f_k - \psi|^p \, d\mu_n \leq \int_{\|x\|\leq c} |f_k - \psi|^p \, d\mu_n + M^p \cdot \mu_n(A_k).$$

By the Dominated Convergence Theorem and continuity of measure, $\{f_k\} \to \psi$ in $L^p(\mathbf{R}^n, \mu_n)$. $\square$

We need the following change of variables theorem[1].

Lemma 8 *Let $f \in L^1(\mathbf{R}^n, \mu_n)$, $L\colon \mathbf{R}^n \to \mathbf{R}^n$ be an invertible linear operator, and h be a point in $\mathbf{R}^n$. Then*

$$\int_{\mathbf{R}^n} f(x) \, dx = |\det L| \cdot \int_{\mathbf{R}^n} f(L(x) - h) \, dx. \tag{7}$$

Proof By the translation invariance of Lebesgue measure on $\mathbf{R}^n$ and Theorem 22 in Chapter 10, for each Lebesgue measurable subset E of $\mathbf{R}^n$,

$$\mu_n(\{L(x) - h \mid x \in E\}) = |\det L| \cdot \mu_n(E) \text{ and } \det L^{-1} = [\det L]^{-1}.$$

Therefore, (7) holds for f the characteristic functions of Lebesgue measurable sets of finite Lebesgue measure, and so also for finitely supported, simple functions φ. Since Lebesgue measure is σ-finite, by an appeal to conclusion (i) of the Simple Approximation Theorem and the Dominated Convergence Theorem, (7) holds for all integrable functions. $\square$

For $h \in \mathbf{R}^n$, define the translation operator $\mathcal{T}_h\colon L^p(\mathbf{R}^n, \mu_n) \to L^p(\mathbf{R}^n, \mu_n)$ by

$$\mathcal{T}_h(f)(x) = f(x - h) \text{ for } x \in \mathbf{R}^n \text{ and } f \in L^p(\mathbf{R}^n, \mu_n).$$

[1] A general non-linear change of variables formula, an extension to Euclidean space of Theorem 16 in Chapter 6, is proven in *Measure Theory and Fine Properties of Functions* by L.C. Evans and R.G. Gariepy.

By (7), with $L = \text{Id}$, $\mathcal{T}_h(f) \in L^p(\mathbf{R}^n, \mu_n)$ and $\|\mathcal{T}_h(f)\|_p = \|f\|_p$.

Proposition 9 *If $1 \le p < \infty$ and $f \in L^p(\mathbf{R}^n, \mu_n)$, then*

$$\lim_{h \to 0,\, h \in \mathbf{R}^n} \|\mathcal{T}_h(f) - f\|_p = 0.$$

Proof Let $\epsilon > 0$. According to Lemma 7, there is a compactly supported, continuous function $g\colon \mathbf{R}^n \to \mathbf{R}$ for which $\|f - g\|_p < \epsilon/3$. Observe that

$$\mathcal{T}_h(f) - f = [\mathcal{T}_h(f - g)] + [\mathcal{T}_h(g) - g] + [g - f],$$

and therefore, by Minkowski's Inequality and the choice of g, since $\|\mathcal{T}_h(f-g)\|_p = \|f-g\|_p$,

$$\|\mathcal{T}_h(f) - f\|_p \le \|\mathcal{T}_h(f - g)\|_p + \|\mathcal{T}_h(g) - g\|_p + \|g - f\| \le 2/3 \cdot \epsilon + \|\mathcal{T}_h(g) - g\|_p. \tag{8}$$

Since g is compactly supported, choose $r > 0$ for which $\mathcal{T}_h(g)(y) - g(y) = 0$ for $\|y\| \ge r$ and $\|h\| \le 1$. Therefore, if $\|h\| \le 1$, then

$$\|\mathcal{T}_h(g) - g\|_p^p = \int_{\|y\| \le r} |\mathcal{T}_h(g)(y) - g(y)|^p \, dy.$$

Being continuous and compactly supported, g is uniformly continuous on all of $\mathbf{R}^n$. Therefore, there is a $\delta > 0$ for which

$$\|\mathcal{T}_h(g) - g\|_p^p = \int_{\|y\| \le r} |\mathcal{T}_h(g)(y) - g(y)|^p \, dy < \epsilon/3 \text{ if } \|h\| < \delta.$$

It follows from (8) that $\|\mathcal{T}_h(f) - f\|_p < \epsilon$ if $\|h\| < \delta$. □

Lemma 10 *For Lebesgue measurable functions $f\colon \mathbf{R}^n \to \mathbf{R}$ and $g\colon \mathbf{R}^n \to \mathbf{R}$, define*

$$k(x, y) = f(x - y)\, g(y) \text{ for } (x, y) \in \mathbf{R}^n \times \mathbf{R}^n. \tag{9}$$

Then $k\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ also is Lebesgue measurable.

Proof The product of measurable functions is measurable, and so it suffices to show that the functions $F\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ and $H\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ are Lebesgue measurable, where

$$F(x, y) = f(x - y) \text{ and } H(x, y) = g(y) \quad \text{for } (x, y) \in \mathbf{R}^n \times \mathbf{R}^n.$$

The function H is Lebesgue measurable, since for each c, $H^{-1}(-\infty, c) = \mathbf{R}^n \times g^{-1}(-\infty, c)$, and the Cartesian product of measurable sets is measurable. Define $G(x, y) = f(y)$, so that, as was just argued, G is Lebesgue measurable, and define $S(x, y) = (x + y, x - y)$. Observe that $F = G \circ S$. According to Theorem 21 of Chapter 10, the linear operator S^{-1} maps Lebesgue measurable sets to Lebesgue measurable sets. For each c,

$$\{u \in \mathbf{R}^n \times \mathbf{R}^n \mid F(u) = (G \circ S)(u) < c\} = S^{-1}\{v \in \mathbf{R}^n \times \mathbf{R}^n \mid G(v) < c\},$$

and so, since the function G is Lebesgue measurable, so also is $F = G \circ S$. □

The product of two integrable functions may not be integrable. For example, consider $f(x) = g(x) = x^{-1/2}$ on $(0, 1)$. However, it follows from the next theorem that if f, g: $\mathbf{R}^n \to \mathbf{R}$ are integrable, then, for almost all $x \in \mathbf{R}^n$, the "shifted product" function $y \mapsto f(y-x)g(y)$ is integrable. Let f: $\mathbf{R}^n \to \mathbf{R}$ and g: $\mathbf{R}^n \to \mathbf{R}$ be Lebesgue measurable. For $x \in \mathbf{R}^n$, the function $y \mapsto f(x-y)g(y)$, by the translate invariance of Lebesgue measure, is the product of Lebesgue measurable functions, and so is Lebesgue measurable. Define the **convolution** of f with g at x by

$$(f * g)(x) = \int_{\mathbf{R}^n} f(x-y)g(y)\,dy,$$

provided that the integral is defined. This is defined if the function $y \mapsto f(x-y)g(y)$ is integrable and also if $f \geq 0$ and $g \geq 0$. Observe that if f, g, and h are Lebesgue measurable functions on $\mathbf{R}^n$, and both $(f*g)(x)$ and $(f*h)(x)$ are defined, then so is $(g*f)(x)$, and we have the following commutative and distributive properties of convolution:

$$(f*g)(x) = (g*f)(x) \text{ and } (f*(g+h))(x) = (f*g)(x) + (f*h)(x).$$

The first equality follows from a change of variables; the second follows from the linearity of integration.

Young's Convolution Inequality *If $1 \leq p < \infty$, $f \in L^p(\mathbf{R}^n, \mu_n)$ and $g \in L^1(\mathbf{R}^n, \mu_n)$, then $f * g \in L^p(\mathbf{R}^n, \mu_n)$, and*

$$\|f*g\|_p \leq \|f\|_p \cdot \|g\|_1.$$

Proof First consider the case $f \geq 0$ and $g \geq 0$. By Lemma 10, $k(x, y) = f(x-y)g(y)$ defines a Lebesgue measurable function on $\mathbf{R}^n \times \mathbf{R}^n$, and since it is non-negative, we may employ Minkowski's Integral Inequality. For each $y \in \mathbf{R}^n$, $\|\mathcal{T}_y(f)\|_p = \|f\|_p$, and so

$$\begin{aligned}
\|f*g\|_p &= \left[\int_{\mathbf{R}^n} \left|\int_{\mathbf{R}^n} f(x-y)g(y)\,dy\right|^p dx\right]^{1/p} \\
&\leq \int_{\mathbf{R}^n} \left[\int_{\mathbf{R}^n} f(x-y)^p g(y)^p\,dx\right]^{1/p} dy \\
&= \int_{\mathbf{R}^n} \left[\int_{\mathbf{R}^n} f(x-y)^p\,dx\right]^{1/p} g(y)\,dy \\
&= \int_{\mathbf{R}^n} \|\mathcal{T}_y(f)\|_p\, g(y) dy \\
&= \|f\|_p \|g\|_1.
\end{aligned}$$

Now consider general functions $f \in L^p(\mathbf{R}^n, \mu_n)$ and $g \in L^1(\mathbf{R}^n, \mu_n)$. We first show that $f * g \in L^p(\mathbf{R}^n, \mu_n)$. To do so, express f and g as differences of non-negative functions, $f = f^+ - f^-$ and $g = g^+ - g^-$, where $f^\pm \in L^p(\mathbf{R}^n, \mu_n)$ and $g^\pm \in L^1(\mathbf{R}^n, \mu_n)$. Then

$$f * g = (f^+ - f^-) * (g^+ - g^-),$$

so that, by the case of non-negative functions and the distributive and commutative properties of convolution, $f * g \in L^p(\mathbf{R}^n, \mu_n)$. Therefore, $f * g$ is defined and finite almost everywhere, and so, by the integral comparison test, for almost all x,

$$|(f * g)(x)|^p \leq (|f| * |g|)(x))^p.$$

Consequently, by the non-negative case,

$$\|f * g\|_p^p \leq \||f| * |g|\|_p^p \leq \|f\|_p^p \cdot \|g\|_1^p$$

Take a p-th root to complete the proof. □

One consequence of Young's Convolution Inequality for $p = 1$ is that if f: $\mathbf{R}^n \to \mathbf{R}$ and g: $\mathbf{R}^n \to \mathbf{R}$ are integrable, then, by replacing the function f by the function $x \mapsto f(-x)$, for almost all $x \in \mathbf{R}^n$, the "shifted product" function $y \mapsto f(y - x)g(y)$ also is integrable. For instance, we see that for the integrable function f: $\mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^{-1/2} \cdot \chi_{(0,1)}(x)$, the convolution $(f * f)(x)$ is defined for all x and is finite for $x \neq 0$.

Proposition 11 *Let $1 \leq p < \infty$, q be the conjugate to p, $f \in L^p(\mathbf{R}^n, \mu_n)$ and $g \in L^q(\mathbf{R}^n, \mu_n)$. Then the convolution $f * g$: $\mathbf{R}^n \to \mathbf{R}$ is defined and finite on all of $\mathbf{R}^n$ and is uniformly continuous.*

Proof By Hölder's Inequality and a change of variables, for each $x \in \mathbf{R}^n$, the function $y \mapsto f(x - y)g(y)$ is integrable over $\mathbf{R}^n$. Therefore, the convolution $f * g$ is defined and finite on all of $\mathbf{R}^n$. Let $x, h \in \mathbf{R}^n$. By the integral comparison test, one more use of the Hölder's Inequality, and a change of variables,

$$\begin{aligned}
|(f * g)(x + h) - (f * g)(x)| &= \left| \int_{\mathbf{R}^n} [f(x + h - y) - f(x - y)]\, g(y)\, dy \right| \\
&\leq \left[\int_{\mathbf{R}^n} |f(x + h - y) - f(x - y)|\, |g(y)|\, dy \right] \\
&\leq \left[\int_{\mathbf{R}^n} |f(x + h - y) - f(x - y)|^p dy \right]^{1/p} \cdot \|g\|_q \\
&= \left[\int_{\mathbf{R}^n} |f(y - h) - f(y)|^p dy \right]^{1/p} \cdot \|g\|_q \\
&= \|T_h(f) - f\|_p \cdot \|g\|_q.
\end{aligned}$$

It follows from Proposition 9 that $f * g$: $\mathbf{R}^n \to \mathbf{R}$ is uniformly continuous on $\mathbf{R}^n$. □

Lemma 12 *Let $1 \leq p < \infty$, $f \in L^p(\mathbf{R}^n, \mu_n)$ and g: $\mathbf{R}^n \to \mathbf{R}$ be compactly supported and continuously differentiable. Then the convolution $f * g$ belongs to $L^p(\mathbf{R}^n, \mu_n)$, is continuously differentiable, and*

$$\partial(f * g)/\partial x_i = f * \partial g/\partial x_i \text{ on } \mathbf{R}^n \text{ for } 1 \leq i \leq n. \tag{10}$$

Proof According to the preceding proposition, $f * g$ is defined and finite on all $\mathbf{R}^n$. Fix $x \in \mathbf{R}^n$ and $1 \leq i \leq n$. For each $t \neq 0$, since $f * g = g * f$,

$$\frac{(f * g)(x + te_i) - (f * g)(x)}{t} = \int_{\mathbf{R}^n} \frac{g(x + te_i - y) - g(x - y)}{t} \cdot f(y)\, dy. \tag{11}$$

Consider the family of functions $\{h^t \colon \mathbf{R}^n \to \mathbf{R}\}_{t \neq 0}$ defined for $t \neq 0$ by

$$h^t(y) = \frac{g(x + te_i - y) - g(x - y)}{t} \text{ for } y \in \mathbf{R}^n.$$

Since $\partial g / \partial x_i$ is compactly supported and continuous, it is uniformly continuous on $\mathbf{R}^n$, and consequently, by the Mean Value Theorem,

$$\lim_{t \to 0} h^t(y) = \partial g / \partial x_i (x - y) \text{ uniformly over } y \in \mathbf{R}^n.$$

Moreover, the family $\{h^t \colon \mathbf{R}^n \to \mathbf{R}^n\}_{0 < |t| \leq 1}$ has uniform compact support. Therefore, as $t \to 0$, h^t converges to the function $y \to \partial g / \partial x_i (x - y)$ in $L^q(\mathbf{R}^n)$, where q is the conjugate of p. Since $f \in L^p(\mathbf{R}^n, \mu_n)$, by Hölder's Inequality, it represents a continuous linear functional on $L^q(\mathbf{R}^n, \mu_n)$. Take the limit as $t \to 0$ in (11) to verify (10). And since, by the preceding proposition, $f * \frac{\partial g}{\partial x_i}$ is continuous, $f * g$ is continuously differentiable. □

A function $g \colon \mathbf{R}^n \to \mathbf{R}$ is said to be infinitely differentiable provided that it has continuous partial derivatives of all orders. Denote by $C^\infty(\mathbf{R}^n)$ the space of such functions, and by $C_c^\infty(\mathbf{R}^n)$ the subspace of compactly supported functions. For an n-tuple $\alpha = (\alpha_1, \ldots \alpha_n)$ of non-negative integers define the differential operator

$$D^\alpha = [\partial / \partial_{x_1}]^{\alpha_1} \cdots [\partial / \partial_{x_n}]^{\alpha_n}.$$

Define the order of D^α to be $\alpha_1 + \cdots + \alpha_n$. From the preceding lemma and an induction argument, based on the order of D^α, we have the following result.

Theorem 13 *Let $1 \leq p < \infty$, $f \in L^p(\mathbf{R}^n, \mu_n)$ and $g \in C_c^\infty(\mathbf{R}^n)$. Then $f * g$ to belongs $L^p(\mathbf{R}^n, \mu_n)$, is infinitely differentiable, and*

$$D^\alpha(f * g) = f * D^\alpha(g) \text{ for all } \alpha = (\alpha_1, \ldots \alpha_n).$$

It is not obvious that $C_c^\infty(\mathbf{R}^n)$ contains any non-zero functions. Indeed, observe that if $f \in C_c^\infty(\mathbf{R}^n)$ and $f \neq 0$, then f, while being infinitely differentiable, fails, since it compactly supported, to be analytic, in that it is not given on all of $\mathbf{R}^n$ by a Taylor series (see Problem 8). We now construct a family of such functions. Define the function $w \colon \mathbf{R}^n \to \mathbf{R}$ by

$$w(x) = \begin{cases} c \exp\left(\frac{1}{\|x\|^2 - 1}\right) & \text{for } \|x\| < 1 \\ 0 & \text{for } \|x\| \geq 1, \end{cases}$$

where the constant $c > 0$ is chosen so that $\int_{\mathbf{R}^n} w(x)\, dx = 1$. We leave it as an exercise (see Problem 9) to show that $w \in C^\infty(\mathbf{R}^n)$. For $\epsilon > 0$, define

$$w_\epsilon(x) = \epsilon^{-n} \cdot w(x / \epsilon) \text{ for } x \in \mathbf{R}^n.$$

Then, by the change of variables (7), with $h = 0$ and $L = 1/\epsilon \cdot \mathrm{Id}$, $\int_{\mathbf{R}^n} w_\epsilon(x)\,dx = 1$. The family of functions $\{w_\epsilon\}_{\epsilon>0}$ has the following properties: for each $\epsilon > 0$,

$$w_\epsilon \geq 0,\ w_\epsilon \in C^\infty(\mathbf{R}^n),\ w_\epsilon(x) = 0 \text{ if } \|x\| \geq \epsilon, \text{ and } \int_{\mathbf{R}^n} w_\epsilon(y)\,dy = 1. \tag{12}$$

Such a family is called a **smooth approximation of the identity,** a name justified by the following result, whose proof uses exactly these four properties.

Theorem 14 *If $1 \leq p < \infty$ and $f \in L^p(\mathbf{R}^n, \mu_n)$, then for each $\epsilon > 0$, $f * w_\epsilon \in C^\infty(\mathbf{R}^n)$ and*

$$\lim_{\epsilon \to 0^+} \|f - f * w_\epsilon\|_p = 0. \tag{13}$$

Proof Let $\epsilon > 0$. According to the preceding theorem, $f * w_\epsilon \in C^\infty(\mathbf{R}^n)$. By Young's Convolution Inequality, the function $f - f * w_\epsilon$ belongs to $L^p(\mathbf{R}^n, \mu_n)$, so it is finite almost everywhere For almost all $x \in \mathbf{R}^n$, since $\int_{\mathbf{R}^n} w_\epsilon(y)\,dy = 1$,

$$f(x) - (f * w_\epsilon)(x) = \int_{\mathbf{R}^n} [f(x) - f(x-y)]w_\epsilon(y)\,dy.$$

By the integral comparison test and Minkowski's Integral Inequality, since $w_\epsilon \geq 0$ and since $w_\epsilon(y) = 0$ if $\|y\| \geq \epsilon$,

$$\begin{aligned}
\|f - f * w_\epsilon\|_p &= \left[\int_{\mathbf{R}^n} \left|\int_{\mathbf{R}^n} [f(x) - f(x-y)]w_\epsilon(y)\,dy\right|^p dx\right]^{1/p} \\
&\leq \left[\int_{\mathbf{R}^n} \left[\int_{\mathbf{R}^n} |f(x) - f(x-y)|w_\epsilon(y)\,dy\right]^p dx\right]^{1/p} \\
&\leq \int_{\mathbf{R}^n} \left[\int_{\mathbf{R}^n} |f(x) - f(x-y)|^p [w_\epsilon(y)]^p\,dx\right]^{1/p} dy \\
&= \int_{\mathbf{R}^n} \left[\int_{\mathbf{R}^n} |f(x) - f(x-y)|^p\,dx\right]^{1/p} w_\epsilon(y)\,dy \\
&= \int_{\overline{B}(0,\epsilon)} \|f - T_y(f)\|_p \cdot w_\epsilon(y)\,dy.
\end{aligned}$$

According to Proposition 9, $\lim_{y\to 0} \|f - T_y(f)\|_p = 0$, and so (13) holds. □

Theorem 15 *If $1 \leq p < \infty$, then the subspace $C_c^\infty(\mathbf{R}^n)$ is dense in $L^p(\mathbf{R}^n, \mu_n)$.*

Proof According to Lemma 7, $C_c(\mathbf{R}^n)$ is a dense subspace of $L^p(\mathbf{R}^n, \mu_n)$. Therefore, to prove this result if suffices to let $f \in C_c(\mathbf{R}^n)$ and $\epsilon > 0$, and show that there is $g \in C_c^\infty(\mathbf{R}^n)$ for which $\|f - g\|_p < \epsilon$. For such an f, by the preceding theorem, there is an $\alpha > 0$ for which $\|f - f * w_\alpha\|_p < \epsilon$. Then, by Theorem 13, $f * w_\alpha$ is infinitely differentiable and, being the convolution of functions with compact support, has compact support. □

In Chapter 10 we defined, for a subset E of $\mathbf{R}^n$, the continuous distance function $\text{dist}_E\colon \mathbf{R}^n \to \mathbf{R}$ and for $\delta > 0$, the δ-neighborhood of E, $\mathcal{N}_\delta(E) = \{x \in \mathbf{R}^n \mid \text{dist}_E(x) < \delta\}$ of E, and proved Urysohn's Lemma in $\mathbf{R}^n$.

A Smooth Urysohn's Lemma *If A and B are disjoint, closed subsets of $\mathbf{R}^n$, and A is compact, then there is an infinitely differentiable, compactly supported function $f\colon \mathbf{R}^n \to \mathbf{R}$ for which*

$$f = 1 \text{ on } A,\quad f = 0 \text{ on } B,\quad \text{and } 0 \le f \le 1 \text{ on } \mathbf{R}^n. \tag{14}$$

Proof Since A and B are disjoint, both are closed, and A is compact, there is a $\delta > 0$ for which the closures of the neighborhoods $\mathcal{N}_\delta(A)$ and $\mathcal{N}_\delta(B)$ are disjoint. By an appeal to Urysohn's Lemma in $\mathbf{R}^n$ with respect to the closures of these neighborhoods, there is a compactly supported, continuous function $g\colon \mathbf{R}^n \to [0,1]$ with the property that

$$\text{if } x \in A, \text{ then } g = 1 \text{ on } B(x, \delta) \text{ and if } x \in B, \text{ then } g = 0 \text{ on } B(x, \delta). \tag{15}$$

Consider the function $f = g * w_\delta\colon \mathbf{R}^n \to \mathbf{R}^n$. According to Theorem 13, f is infinitely differentiable and, being the convolution of functions with compact support, f itself has compact support. Let $x \in A$. If $\|y\| \ge \delta$, then $w_\delta(y) = 0$, while if $\|y\| \le \delta$, then $g(x-y) = 1$. Consequently,

$$f(x) = \int_{\mathbf{R}^n} g(x-y)\cdot w_\delta(y)\,dy = \int_{\|y\|\le\delta} g(x-y)\cdot w_\delta(y)\,dy = \int_{\|y\|\le\delta} w_\delta(y)\,dy = 1.$$

Similarly, $f(x) = 0$ if $x \in B$. $\square$

PROBLEMS

8. Define $f\colon \mathbf{R} \to \mathbf{R}$ by $f(0) = 0$ and otherwise $f(x) = \exp(-1/x^2)$. Use the Taylor expansion for $\exp(x)$ at $x = 0$ to show that for each polynomial $p(x)$,
$$\lim_{x\to 0} p(1/x) f(x) = 0.$$
Use this to show that $f \in C^\infty(\mathbf{R})$, and $f^{(k)}(0) = 0$ for all k. For what $x \in \mathbf{R}^n$ does the Taylor expansion of f at $x = 0$ agree with $f(x)$?
9. Use the preceding problem to verify that the function w belongs to $C_c^\infty(\mathbf{R})$.
10. For $1 \le p < \infty$, let $f \in L^p(\mathbf{R}^n)$. Show that if $\int_{\mathbf{R}^n} f(x)g(x)\,dx = 0$ for all $g \in C_c^\infty(\mathbf{R}^n)$, then $f = 0$.
11. For the function $f\colon \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^{-1/2}\cdot \chi_{(0,1)}(x)$, show that $(f * f)(x)$ is defined for all x and is finite for $x \ne 0$.
12. Show that if $f\colon \mathbf{R}^n \to \mathbf{R}$ is bounded and continuous, then, as $\epsilon \to 0$, $\{f * w_\epsilon\} \to f$ pointwise, uniformly on compact sets.
13. Let $k \in L^2(\mathbf{R}^n \times \mathbf{R}^n)$, and define the integral operator $T\colon L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^n)$ by
$$T(f)(x) = \int_{\mathbf{R}^n} k(x, y) f(y)\,dy \text{ for } f \in L^2(\mathbf{R}^n) \text{ and } x \in \mathbf{R}^n.$$
Show that the linear operator T is properly defined, and determine a $c \ge 0$ for which
$$\|T(f)\|_2 \le c\cdot \|f\|_2 \text{ for all } f \in L^2(\mathbf{R}^n).$$

12.3 THE RIESZ REPRESENTATION THEOREM FOR THE DUAL OF $L^p(X, \mu)$, $1 \le p < \infty$

Recall that in Chapter 8 the dual of a normed linear space V, denoted by V^*, was defined to be the normed linear space of bounded linear functionals Φ on V, where the norm $\|\Phi\|_*$ is the smallest c for which

$$|\Phi(v)| \le c \cdot \|v\| \text{ for all } v \in V,$$

so that, by the homogeneity of the norm and the linear functional,

$$\|\Phi\|_* = \sup\{|\Phi(v)| \mid \|v\| \le 1\}. \tag{16}$$

In this section, we prove a representation theorem for the dual $L^p(X, \mu)$ for $1 \le p < \infty$, where $(X, \mathcal{M}, \mu)$ is a general σ-finite measure space, which directly generalizes the representation theorem that was established in Chapter 8 for the Lebesgue integral for functions of a real variable.

Proposition 16 *For $1 \le p \le \infty$, q the conjugate of p, and f in $L^q(X, \mu)$, define the linear functional Φ on $L^p(X, \mu)$ by*

$$\Phi(g) = \int_X f \cdot g \, d\mu \textit{ for all } g \in L^p(X, \mu). \tag{17}$$

Then

$$\Phi \in [L^p(X, \mu)]^* \textit{ and } \|\Phi\|_* = \|f\|_q. \tag{18}$$

Proof For $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$, by Hölder's Inequality, $f \cdot g$ is integrable, and therefore Φ is properly defined, and is linear. Furthermore, $\Phi\colon L^p(X, \mu) \to \mathbf{R}$ is bounded and $\|\Phi\|_* \le \|g\|_q$. It remains to show that this is an equality. First consider the case that $1 < p \le \infty$. Assume that $g \ne 0$, since otherwise there is nothing to prove. According to Theorem 1 (with p and q interchanged), since $1 \le q < \infty$, the dual function of g, g^*, belongs to $L^p(X, \mu)$, $\Phi(g^*) = \|g\|_q$ and $\|g^*\|_p = 1$. Therefore, $\Phi(g^*) = \|g\|_q \cdot \|g^*\|_p$, so that, by (16), $\|\Phi\|_* = \|g\|_q$. Now consider the case that $p = 1$. We argue by contradiction to establish the equality $\|\Phi\|_* = \|g\|_\infty$. Indeed, otherwise $\|\Phi\|_* < \|g\|_\infty$, and so there is some $\epsilon > 0$ for which the set $X_\epsilon = \{x \in E \mid |g(x)| > \|\Phi\|_* + \epsilon\}$ has non-zero measure. If we choose f be the characteristic function of a measurable subset of X_ϵ that has finite positive measure, we contradict the inequality $|\Phi(f)| \le \|\Phi\|_* \cdot \|f\|_1$. □

Corollary 17 *If $1 \le p < \infty$ and $f \in L^p(X, \mu), f \ne 0$, then there is a $\Phi \in [L^p(X, \mu)]^*$ for which*[2]

$$\Phi(f) = \|f\|_p \textit{ and } \|\Phi\|_* = 1. \tag{19}$$

[2]For any normed linear space X, if $f \in X, f \ne 0$, then there is a functional $\Phi \in X^*$ for which $\Phi(f) = \|f\|$ and $\|\Phi\|_* = 1$. This is a consequence of the Hahn-Banach Theorem, which is proven in Chapter 18. In particular, this corollary holds for $p = \infty$.

Proof Let $f^* \in L^q(X, \mu)$ be the dual function of f. According to the preceding proposition,

$$\Phi(g) = \int_X g \cdot f^* \, d\mu \text{ for all } g \in L^p(X, \mu)$$

defines $\Phi \in [L^p(X, \mu)]^*$ and $\|\Phi\|_* = \|f^*\|$. By the definition of dual function, $\|\Phi\|_* = \|f^*\|_q = 1$ and $\Phi(f) = \|f\|_p$. □

A bounded functional $\Phi\colon L^p(X, \mu) \to \mathbf{R}$ is said to be **represented** by a function $f \in L^q(X, \mu)$ provided that (17) holds. It follows from $\|\Phi\|_* = \|f\|_q$ that if $\Phi = 0$, then $f = 0$, so that if there is a representative of a functional, there is only one representative.

The Riesz Representation Theorem for the Dual of $L^p(X, \mu)$ *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, $1 \leq p < \infty$, and q be the conjugate to p. Then every bounded, linear functional $\Phi \in [L^p(X, \mu)]^*$ is represented by a unique function f in $L^q(X, \mu)$.*

Proof The case $p = 1$ is left as an exercise (see Problem 14). Assume that $1 < p < \infty$. First consider the case $\mu(X) < \infty$. Let $\Phi \in [L^p(X, \mu)]^*$. Define a set-function ν on $\mathcal{M}$ by setting

$$\nu(E) = \Phi(\chi_E) \text{ for } E \in \mathcal{M}.$$

This is properly defined since $\mu(X) < \infty$, and so $\chi_E \in L^p(X, \mu)$. We claim that ν is a signed measure. Indeed, if $\{E_k\}_{k=1}^\infty$ is a measurable partition of $E \in \mathcal{M}$, then, by the countable additivity of μ,

$$\mu(E) = \sum_{k=1}^\infty \mu(E_k) < \infty,$$

and so $\lim_{n\to\infty} \sum_{k=n+1}^\infty \mu(E_k) = 0$. Consequently,

$$\lim_{n\to\infty} \|\chi_E - \sum_{k=1}^n \chi_{E_k}\|_p = \lim_{n\to\infty} \left[\sum_{k=n+1}^\infty \mu(E_k)\right]^{1/p} = 0. \tag{20}$$

But Φ is both linear and continuous on $L^p(X, \mu)$ and so $\Phi(\chi_E) = \sum_{k=1}^\infty \Phi(\chi_{E_k})$, that is,

$$\nu(E) = \sum_{k=1}^\infty \nu(E_k).$$

To verify that ν is a signed measure, it must be shown that the series on the right converges absolutely. For each k, define $c_k = \text{sgn}(\Phi(\chi_{E_k}))$, and argue as above to conclude that the series $\sum_{k=1}^\infty \Phi(c_k \cdot \chi_{E_k})$ is Cauchy and therefore, since $L^p(X, \mu)$ is complete,

$$\sum_{k=1}^\infty |\nu(E_k)| = \sum_{k=1}^\infty \Phi(c_k \cdot \chi_{E_k}) \text{ converges.}$$

Consequently, ν is a signed measure. The measure ν is absolutely continuous with respect to μ. Indeed, if $E \in \mathcal{M}$ has $|\mu(E)| = 0$, then χ_E may be identified with the function $f \equiv 0$ and therefore, since Φ is linear, $\nu(E) = \Phi(\chi_E) = \Phi(f \equiv 0) = 0$. According to Corollary 17 in the preceding chapter, a consequence of the Hahn Decomposition and Radon-Nikodym Theorems, there is a function $f \in L^1(X, \mu)$ for which

$$\Phi(\chi_E) = \nu(E) = \int_E f \, d\mu \text{ for all } E \in \mathcal{M}.$$

By the linearity of Φ and of integration,

$$\Phi(\varphi) = \int_X f \cdot \varphi \, d\mu \text{ for each simple function } \varphi. \tag{21}$$

Since the functional Φ is bounded on $L^p(X, \mu)$,

$$\left| \int_X f \cdot \varphi \, d\mu \right| = |\Phi(\varphi)| \leq \|\Phi\|_* \cdot \|\varphi\|_p \text{ for all simple functions } \varphi. \tag{22}$$

Consequently, since $\mu(X) < \infty$, by Theorem 4, $f \in L^q$. According to Proposition 16, f represents a continuous linear functional on $L^p(E, \mu)$, and it agrees with the continuous functional Φ on the simple functions. Since the simple functions are dense in $L^p(E, \mu)$, by continuity, Φ is represented by f. This concludes the proof in the case that $\mu(X) < \infty$.

$$\Phi(g) = \lim_{n\to\infty} \Phi(\varphi_n) = \lim_{n\to\infty} \int_X f \cdot \varphi_n \, d\mu = \int_X f \cdot g \, d\mu \text{ for all } g \in L^p(X, \mu).$$

This conclude the proof in the case $\mu(X) < \infty$.

Now consider the general case, namely, that X is σ-finite. Choose an ascending collection $\{X_n\}_{n=1}^{\infty}$ of sets of finite measure whose union is X. For each n, we consider $L^p(X_n, \mu)$ to be the subspace of $L^p(X, \mu)$ comprising functions that vanish outside X_n, and, since $\mu(X_n) < \infty$, let $f_n \in L^q(X_n, \mu)$ represent Φ_n, the restriction of Φ to $L^p(X_n, \mu)$, that is

$$\Phi_n(g) = \int_{X_n} f_n \cdot g \, d\mu \text{ for all } g \in L^p(X_n, \mu). \tag{23}$$

It follows from Proposition 16 that representatives of bounded functionals are unique, and so we may define the measurable function $f\colon X \to \mathbf{R}$ so that its restriction to each X_n is f_n. Again by Proposition 16, for each n,

$$\|f_n\|_q = \|\Phi_n\|_* \leq \|\Phi\|_*.$$

Therefore, since $\{|f_n|^q\}$ converges pointwise on X to $|f|^q$, by Fatou's Lemma, f belongs to $L^q(X, \mu)$, and so f represents a continuous functional $L^p(X, \mu)$, as is the functional Φ. Let $g \in L^p(X, \mu)$. For each n, let χ_n be the characteristic function of X_n, and observe that $|g - g \cdot \chi_n|^q \leq |g|^q$. Consequently, by the Dominated Convergence Theorem, $\{g \cdot \chi_n\} \to g$ in $L^p(X, \mu)$. In view of (23), taking limits, we have, for all $g \in L^p(X, \mu)$,

$$\Phi(g) = \lim_{n\to\infty} \Phi(g \cdot \chi_n) = \lim_{n\to\infty} \int_X f \cdot (g \cdot \chi_n) \, d\mu = \int_X f \cdot g \, d\mu.$$

Therefore, $f \in L^q(X, \mu)$ represents Φ. $\square$

PROBLEMS

14. Prove the Riesz Representation Theorem for the case $p = 1$ by adapting the proof for the case $p > 1$.
15. For $1 \le p \le \infty$ and q the conjugate of p, let $\{f_n\} \to f$ in $L^q(X, \mu)$ and let $\{g_n\} \to g$ in $L^p(X, \mu)$. Show that
$$\lim_{n\to\infty} \int_X f_n \cdot g_n \, d\mu = \int_X f \cdot g \, d\mu.$$
16. Find a measure space $(X, \mathcal{M}, \mu)$ for which the Riesz Representation Theorem does extend to the case $p = \infty$.
17. For $1 \le p < \infty$, find a representation for the bounded, linear functionals on ℓ^p. (Suggestion: Consider the counting measure on $\mathbf{N}$.)
18. Find a representation for the bounded, linear functionals on a Dirac measure space.
19. Find a representation for the bounded, linear functionals on $\mathbf{R}^n$.

12.4 WEAK SEQUENTIAL COMPACTNESS IN $L^p(X, \mu), 1 < p < \infty$

Recall that a normed linear space V is said to be separable provided that there is a countable, dense subset. If $\mathcal{F}$ is dense in V and $\mathcal{S} \subseteq V$ is a countable subset of V and each point in $\mathcal{F}$ is the limit of a sequence in $\mathcal{S}$, then V is separable. Also recall that a sequence $\{v_k\}$ in a normed linear space V is said to converge weakly to $v \in V$ provided that $\{\Phi(v_k)\} \to \Phi(v)$ for all Φ in the dual space V^*. We considered weak convergence in Chapter 8, in which Helly's Theorem was proven, according to which if V is a separable normed linear space, then for every bounded sequence $\{\Phi_k\}$ in the dual space V^*, there is a subsequence $\{\Phi_{k_j}\}$ that converges pointwise on V to a functional $\Phi \in V^*$. This theorem, together with the Riesz Representation Theorem, was used to show that if $E \subseteq \mathbf{R}$ is Lebesgue measurable, then, for $1 < p < \infty$, every bounded sequence in $L^p(E, \mu_1)$ has a weakly convergent subsequence. Here, Helly's Theorem and the general Riesz Representation Theorem will be used in order to extend this weak sequential compactness result to functions defined on a subset of Euclidean space.

Theorem 18 *If $E \subseteq \mathbf{R}^n$ is Lebesgue measurable and $1 \le p < \infty$, then $L^p(E, \mu_n)$ is separable.*

Proof The space $L^p(E, \mu_n)$ may be identified with the subspace of $L^p(\mathbf{R}^n, \mu_n)$ comprising functions that vanish outside E. Consequently, it suffices to show that $L^p(\mathbf{R}^n, \mu_n)$ is separable. Since $C_c(\mathbf{R}^n)$ is dense in $L^p(\mathbf{R}^n, \mu_n)$, to do so, it suffices to select a countable set $\mathcal{F} \subseteq L^p(\mathbf{R}^n, \mu_n)$ for which each function in $C_c(\mathbf{R}^n)$ is the limit in $L^p(\mathbf{R}^n, \mu_n)$ of a sequence in $\mathcal{F}$. Fix an index k. Let $\mathcal{P}_k$ be a finite measurable partition of the ball $\overline{B}(0, k)$ such that if δ_k is the maximum of the diameters of the sets in $\mathcal{P}_k$, then $\delta_k < 1/k$. Define $\mathcal{F}_k$ to be the space of rational-valued simple functions on $\mathbf{R}^n$ that vanish outside $\overline{B}(0, k)$ and are constant on each set in the partition $\mathcal{P}_k$. Then $\mathcal{F}_k$ is countable and hence so is $\mathcal{F} = \bigcup_{k=1}^{\infty} \mathcal{F}_k$. Let $f \in C_c(\mathbf{R}^n)$. Choose a function $f_k \in \mathcal{F}_k$ with the property that in each set in the partition $\mathcal{P}_k$ there is a point x at which $|f(x) - f_k(x)| < 1/k$. Since f has compact support, there is an index K such that f, and therefore each f_k vanishes outside $\overline{B}(0, K)$

for all $k \geq K$. Since f: $\mathbf{R}^n \to \mathbf{R}$ is uniformly continuous and $\lim_{k\to\infty} \delta_k = 0$, $\{f_k\} \to f$ in $L^\infty(\overline{B}(0, K), \mu_n)$ and therefore also in $L^p(\mathbf{R}^n, \mu_n)$. □

In view of the characterization of the dual of $L^p(E, \mu_n)$ for $1 \leq p < \infty$, we have the following result.

Proposition 19 *If $1 \leq p < \infty$ and q is the conjugate of p, then $\{f_k\} \to f$ weakly in $L^p(E, \mu_n)$ if and only if*

$$\lim_{k\to\infty} \int_X f_k \cdot g \, d\mu = \int_X f \cdot g \, d\mu \textit{ for all } g \in L^q(E, \mu_n). \tag{24}$$

Theorem 20 *If E is a Lebesgue measurable subset of $\mathbf{R}^n$ and $1 < p < \infty$, then every bounded sequence in $L^p(E, \mu_n)$ has a weakly convergent subsequence.*

Proof Let $\{f_k\}$ be a bounded sequence in $L^p(E, \mu_n)$. For each k, let the functional $\Phi_k \in [L^q(E)]^*$ be represented by f_k. It follows from Proposition 16 (with p and q interchanged) that $\{\Phi_k\}$ is a bounded sequence in $[L^q(E, \mu_n)]^*$. Since $1 < q < \infty$, according to the preceding theorem, $L^q(E, \mu_n)$ is separable. By Helly's Theorem, there is a subsequence $\{\Phi_{k_j}\}$ that converges pointwise on $L^q(E)$ to a functional $\Phi \in [L^q(E)]^*$. The Riesz Representation Theorem (with p and q interchanged) implies that Φ is represented by a function $f \in L^p(E)$. Observe that

$$\lim_{j\to\infty} \int_E f_{k_j} \cdot g \, d\mu_n = \lim_{j\to\infty} \Phi_{k_j}(g) = \Phi(g) = \int_E f \cdot g \, d\mu_n \text{ for all } g \in L^q(E)$$

Therefore, in view of (24), $\{f_{k_j}\} \to f$ weakly in $L^p(E, \mu_n)$. □

Recall that a subset K of a normed linear space is said to be *weakly sequentially compact* provided that any sequence in K has a subsequence that converges weakly to a member of K.

Proposition 21 *Let E be a Lebesgue measurable subset of $\mathbf{R}^n$ and $1 < p < \infty$. Then the set $K = \{f \in L^p(E, \mu_n) \,|\, \|f\|_p \leq 1\}$ is weakly sequentially compact.*

Proof Let $\{f_k\}$ be a sequence in K. The preceding theorem implies that there is a subsequence $\{f_{k_j}\}$ of $\{f_k\}$ and a function f in $L^p(E, \mu_n)$ such that $\{f_{k_j}\} \to f$ weakly. It remains to show that $f \in K$. Consider the function $f^* \in L^q(E, \mu_n)$, the dual of f. By the definition of weak convergence,

$$\|f\|_p = \int_E f \cdot f^* \, d\mu_n = \lim_{j\to\infty} \int_E f_{k_j} \cdot f^* \, d\mu_n.$$

On the other hand, for each k, by Hölder's Inequality,

$$\int_E f_{k_j} \cdot f^* \, d\mu_n \leq \|f_{k_j}\|_p \cdot \|f^*\|_q = \|f_{k_j}\|_p \leq 1.$$

Therefore, $\|f\|_p \leq 1$, that is, $f \in K$. □

The proof of Theorem 20 directly provides a proof of the following theorem.

Theorem 22 *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, $1 < p < \infty$ and q be the conjugate of p. Assume that $L^q(X, \mu)$ is separable. Then every bounded sequence in $L^p(X, \mu)$ has a weakly convergent subsequence.*

The separability assumption in the above theorem is superflous, but the proof of this requires considerations beyond Helly's Theorem. In Chapter 18, the concept of reflexive normed linear space is introduced. In such a space, we prove that every bounded sequence has a weakly convergent subsequence. Moreover, we also prove that if the measure space $(X, \mathcal{M}, \mu)$ is σ-finite and $1 < p < \infty$, then $L^p(X, \mu)$ is reflexive.

PROBLEMS

20. Extend the Banach-Saks Theorem to $L^2(\mathbf{R}^n, \mu_n)$, and use this to prove that a closed, bounded, convex subset of $L^2(\mathbf{R}^n, \mu_n)$ is weakly sequentially compact.
21. Prove that a convex, continuous real-valued function on a closed, bounded, convex subset $L^2(\mathbf{R}^n, \mu_n)$ takes a minimum value. (Suggestion: Consider the corresponding material in Chapter 8 and the preceding problem.)
22. Show that $L^\infty(\mathbf{R}^n, \mu_n)$ is not separable.
23. For $1 < p < \infty$, show that a weakly convergent sequence in $L^p(X, \mu)$ is bounded. (Suggestion: Consider the Uniform Boundedness Principle.)
24. Let c be the counting measure on the set of natural numbers $\mathbf{N}$. For $1 \le p < \infty$ prove that a bounded sequence in $L^p(\mathbf{N}, c)$ converges weakly if and only if it converges pointwise.
25. A linear functional $\Psi\colon L^p(X, \mu) \to \mathbf{R}$ is said to be positive provided that $\Psi(g) \ge 0$ for each non-negative function g in $L^p(X, \mu)$. For $1 \le p < \infty$ and μ σ-finite, show that each bounded linear functional on $L^p(X, \mu)$ is the difference of bounded, positive linear functionals.

12.5 THE KANTOROVITCH REPRESENTATION THEOREM FOR THE DUAL OF $L^\infty(X, \mu)$

In the preceding section, the dual of $L^p(X, \mu)$ for $1 \le p < \infty$ and $(X, \mathcal{M}, \mu)$ a σ-finite measure space was characterized. The dual space of $L^\infty(X, \mu)$ is now considered.

Definition *Let the set-function $\nu\colon \mathcal{M} \to \mathbf{R}$ be finitely additive. The total variation of ν over X, $\|\nu\|_{var}$, is defined by*

$$\|\nu\|_{var} = \sup \sum_{k=1}^{n} |\nu(E_k)|, \tag{25}$$

where the supremum is taken over finite, disjoint collections $\{E_k\}_{k=1}^n$ of sets in $\mathcal{M}$. The function ν is called a **bounded, finitely additive, signed measure** *provided that $\|\nu\|_{var} < \infty$.*

If $\nu\colon \mathcal{M} \to [0, \infty]$ is a measure, then $\|\nu\|_{\text{var}} = \nu(X)$. If $\nu\colon \mathcal{M} \to \mathbf{R}$ is a signed measure, it is not difficult to see that the total variation $\|\nu\|_{\text{var}}$ is given by

$$\|\nu\|_{\text{var}} = |\nu|(X) = \nu^+(X) + \nu^-(X),$$

where $\nu = \nu^+ - \nu^-$ is the Jordan Decomposition of ν as the difference of measures. If $\nu\colon \mathcal{M} \to \mathbf{R}$ is a bounded, finitely additive, signed measure and $\varphi = \sum_{k=1}^{n} c_k \cdot \chi_{E_k}$ is a simple function, the integral of φ over X with respect to ν is defined by

$$\int_X \varphi \, d\nu = \sum_{k=1}^{n} c_k \cdot \nu(E_k).$$

The integral is properly defined, linear with respect to the integrand and

$$\left| \int_X \varphi \, d\nu \right| \leq \|\nu\|_{\text{var}} \cdot \|\varphi\|_\infty. \tag{26}$$

Indeed, the arguments used in considering the integral of simple functions with respect to a measure (only the finite additivity of measure was needed) are sufficient to show that the above integral is properly defined and linear. Let $f\colon X \to \mathbf{R}$ be a bounded measurable function. According to the Simple Approximation Lemma, there are sequences $\{\psi_n\}$ and $\{\varphi_n\}$ of simple functions on X for which

$$\varphi_n \leq f \leq \psi_n, \text{ and } 0 \leq \psi_n - \varphi_n \leq 1/n \text{ on } X \text{ for all } n.$$

It follows from (26) that

$$\left| \int_X \varphi_{n+k} \, d\nu - \int_X \varphi_n \, d\nu \right| \leq \|\nu\|_{\text{var}} \cdot \|\varphi_{n+k} - \varphi_n\|_\infty \text{ for all } n \text{ and } k.$$

Define the integral of f over X with respect to ν by

$$\int_X f \, d\nu = \lim_{n \to \infty} \int_X \varphi_n \, d\nu$$

This does not depend on the choice of the sequence of simple functions that converges uniformly on X to f.

Now, for the measure space $(X, \mathcal{M}, \mu)$, in order to properly define the integral $\int_X f \, d\nu$ for $f \in L^\infty(X, \mu)$, and respect the standard identification of functions that are equal almost everywhere, it is necessary that $\int_X f \, d\nu = \int_X f_1 \, d\nu$ if $f = f_1$ almost everywhere $[\mu]$ on X. If there is a set $E \in \mathcal{M}$ for which $\mu(E) = 0$, but $\nu(E) \neq 0$, then clearly this does not hold. This prompts the following definition.

Definition *Let $(X, \mathcal{M}, \mu)$ be a measure space. By $\mathcal{BFA}(X, \mathcal{M}, \mu)$ is denoted the normed linear space of bounded, finitely additive, signed measures ν on $\mathcal{M}$ that are absolutely continuous with respect to μ in the sense that if $E \in \mathcal{M}$ and $\mu(E) = 0$, then $\nu(E) = 0$. The norm of $\nu \in \mathcal{BFA}(X, \mathcal{M}, \mu)$ is the total variation norm $\|\nu\|_{var}$.*

By the absolute continuity of ν with respect to μ, it is clear that if $\nu \in \mathcal{BFA}(X, \mathcal{M}, \mu)$ and φ and ψ are simple functions that are equal almost everywhere $[\mu]$ on X, then $\int_X \varphi \, d\nu = \int_X \psi \, d\nu$ and so the same is true for essentially bounded measurable functions that are equal almost everywhere $[\mu]$ on X. Therefore, the integral of an $L^\infty(X, \mu)$ function over X with respect to ν is properly defined, since it has been defined above for a bounded measurable function, and

$$\left| \int_X f \, d\nu \right| \leq \|\nu\|_{\text{var}} \cdot \|f\|_\infty \text{ for all } f \in L^\infty(X, \mu) \text{ and } \nu \in \mathcal{BFA}(X, \mathcal{M}, \mu). \tag{27}$$

It follows that integration with respect to ν is a bounded, linear functional on $L^\infty(X, \mu)$. A functional $\Phi \in [L^\infty(X, \mu)]^*$ is said to be **represented** by a set-function $\nu \in \mathcal{BFA}(X, \mathcal{M}, \mu)$ provided that

$$\Phi(f) = \int_X f \, d\nu \text{ for all } f \in L^\infty(X, \mu).$$

Theorem 23 (the Kantorovitch Representation Theorem) *Every $\Phi \in [L^\infty(X, \mu)]^*$ is represented by a set-function $\nu \in \mathcal{BFA}(X, \mathcal{M}, \mu)$.*

Proof Let $\Phi \in [L^\infty(X, \nu)]^*$. Define $\nu \colon \mathcal{M} \to \mathbf{R}$ by

$$\nu(E) = \Phi(\chi_E) \text{ for all } E \in \mathcal{M}. \tag{28}$$

Now, $\chi_E \in L^\infty(X, \nu)$ and therefore ν is properly defined. Moreover, ν is finitely additive, since Φ is linear. The measure ν is absolutely continuous with respect to μ. Indeed, to verify this, let $E \in \mathcal{M}$ have $\mu(E) = 0$. The function χ_E is identified with the function $f \equiv 0$, and therefore, since Φ is linear, $\nu(E) = \Phi(\chi_E) = \Phi(f \equiv 0) = 0$. It follows from the linearity of Φ and of integration with respect to ν that

$$\Phi(\varphi) = \int_X \varphi \, d\nu \text{ for all simple functions } \varphi$$

The Simple Approximation Lemma implies that the simple functions are dense in $L^\infty(X, \mu)$. Therefore, since both Φ and integration with respect to ν are continuous on $L^\infty(X, \mu)$, that agree on a dense subset, they are equal. □

For the Lebesgue measure space $([0, 1], \mathcal{L}, m)$, in Chapter 8, is was argued that there is a bounded, linear functional $\Phi \in [L^\infty([0, 1], m)]^*$ that is not represented by a function in $L^1([0, 1], m)$. However, by the Kantorovitch Representation Theorem, for such a functional, there is a set-function $\nu \in \mathcal{BFA}([0, 1], \mathcal{L}, m)$ for which

$$\Phi(f) = \int_X f \, d\nu \text{ for all } f \in L^\infty([0, 1], m).$$

But there is no function $g \in L^1([0, 1], m)$ for which

$$\Phi(f) = \int_{[0, 1]} f \, d\nu = \int_{[0, 1]} g \cdot f \, dm \text{ for all } f \in L^\infty([0, 1], m). \tag{29}$$

The set-function ν cannot be countably additive, that is, it cannot be a signed measure. Indeed, if it were, by the Radon-Nikodym Theorem applied to the terms in the Jordan Decomposition of ν, there would be an $L^1([0, 1], m)$ function g for which (29) holds. Therefore, ν is a bounded set-function on the Lebesgue measurable subsets of $[a, b]$, is absolutely continuous with respect to Lebesgue measure, is finitely additive but not countably additive. No such set-function has been explicitly exhibited.

PROBLEMS

26. Verify (26) and (27).
27. Show that $\mathcal{BFA}(X, \mathcal{M}, \mu)$ is a linear space on which $\|\cdot\|_{\mathrm{var}}$ is a norm. Then show that this normed linear space is complete.
28. For $\Phi \in [L^\infty(X, \mu)]^*$, let $\mathcal{R}(\Phi) \in \mathcal{BFA}(X, \mathcal{M}, \mu)$ be the representative of Φ. Prove that $\mathcal{R}\colon [L^\infty(X, \mu)]^* \to \mathcal{BFA}(X, \mathcal{M}, \mu)$ is an isometry.

PART THREE

ABSTRACT SPACES: METRIC, TOPOLOGICAL, BANACH, AND HILBERT SPACES

This page intentionally left blank

CHAPTER 13

Metric Spaces: General Properties

Contents

13.1 Examples of Metric Spaces 263
13.2 Open Sets, Closed Sets, and Convergent Sequences 265
13.3 Continuous Mappings Between Metric Spaces 269
13.4 Complete Metric Spaces 271
13.5 Compact Metric Spaces 274
13.6 Separable Metric Spaces: Ulam's Regularity Theorem 280

In Chapter 1, three types of properties of the real numbers were considered. The first were the algebraic properties related to addition and multiplication. The second were the properties of the positive numbers, by way of which the concepts of order and absolute value were defined. Using the algebraic and order properties, the distance between two real numbers was defined to be the absolute value of their difference. The final property possessed by the real numbers was completeness: the Completeness Axiom for the real numbers was equivalent to the property that every Cauchy sequence of real numbers converges to a real number. In the study of normed linear spaces, which was begun in Chapter 7, the algebraic structure of the real numbers was extended to that of a linear space; the absolute value was extended to the concept of a norm, which induces a concept of distance between points; and the order properties of the real numbers were left aside. We now proceed one step further in generalization. The object of the present chapter is to consider general spaces called metric spaces for which the notion of distance between two points is fundamental. There is no linear structure. The concepts of open set and closed set in Euclidean space extend naturally to general metric spaces, as do the concepts of convergence of a sequence and continuity of a function or mapping. We first consider these general concepts, and then consider metric spaces that possess finer structure: those that are complete, compact, or separable.

13.1 EXAMPLES OF METRIC SPACES

Definition *Let X be a non-empty set. A function $\rho\colon X \times X \to \mathbf{R}$ is called a* **metric** *provided that for all x, y, and z in X,*

(i) $\rho(x,y) \geq 0$;
(ii) $\rho(x,y) = 0$ *if and only if* $x = y$;
(iii) $\rho(x,y) = \rho(y,x)$;
(iv) $\rho(x,y) \leq \rho(x,z) + \rho(z,y)$.

The set X together with the metric ρ is called a **metric space** *and denoted by (X, ρ).*

Property (iv) is known as the **triangle inequality** for the metric. The quintessential example of a metric space is the set $\mathbf{R}$ of all real numbers with $\rho(x,y) = |x - y|$.

Normed Linear Spaces Recall that a non-negative real-valued function $\|\cdot\|$ on a linear space X is called a **norm** provided that for each $u, v \in X$ and real number α,

(i) $\|u\| = 0$ if and only if $u = 0$.
(ii) $\|u + v\| \leq \|u\| + \|v\|$.
(iii) $\|\alpha u\| = |\alpha|\|u\|$.

A norm $\|\cdot\|$ on a linear space X induces a metric ρ on X by defining

$$\rho(x, y) = \|x - y\| \text{ for all } x, y \in X. \tag{1}$$

Property (ii) of a norm is called the triangle inequality for the norm. It is equivalent to the triangle inequality for the induced metric. Indeed, for $x, y, z \in X$, set $u = x - z$ and $v = z - y$ and observe that

$$\|u + v\| \leq \|u\| + \|v\| \text{ if and only if } \rho(x, y) \leq \rho(x, z) + \rho(z, y).$$

Among the normed linear spaces that have been considered are: (i) Euclidean space $\mathbf{R}^n$; (ii) $L^p(X, \mu)$ and ℓ^p, for $1 \leq p \leq \infty$; (iii) $C[a, b]$, with the maximum norm. Each subset of these spaces is a metric space.

The Discrete Metric For any non-empty set X, the discrete metric ρ is defined by setting $\rho(x, y) = 0$ if $x = y$ and $\rho(x, y) = 1$ if $x \neq y$.

Definition *A mapping f from a metric space (X, ρ) to a metric space (Y, σ) is said to be an* **isometry** *provided that it maps X onto Y and for all $x_1, x_2 \in X$,*

$$\sigma(f(x_1), f(x_2)) = \rho(x_1, x_2).$$

Two metric spaces are called **isometric** provided that there is an isometry from one onto the other. To be isometric is an equivalence relation among metric spaces. From the viewpoint of metric spaces, two isometric metric spaces are exactly the same, an isometry amounting merely to a relabeling of the points and the metrics.

PROBLEMS

1. *The Nikodym Metric.* Let $(X, \mathcal{M}, \mu)$ be a finite measure space. For $A, B \in \mathcal{M}$, define $\rho(A, B) = m(A \Delta B)$, where $A \Delta B = [A \sim B] \cup [B \sim A]$, the symmetric difference of A and B. Define two measurable sets to be equivalent provided that their symmetric difference has measure zero. Show that ρ induces a metric on the collection of equivalence classes. Finally, show that for $A, B \in \mathcal{M}$,
$$\rho(A, B) = \int_X |\chi_A - \chi_B| \, d\mu,$$
where χ_A and χ_B are the characteristic functions of A and B, respectively.
2. Show that for $a, b, c \geq 0$,
$$\text{if } a \leq b + c, \text{ then } \frac{a}{1+a} \leq \frac{b}{1+b} + \frac{c}{1+c}.$$
3. Let $(X, \mathcal{M}, \mu)$ be a finite measure space and $\mathcal{F}$ be the collection of measurable real-valued functions on X. For $f, g \in \mathcal{F}$, define
$$\rho(f, g) = \int_E \frac{|f - g|}{1 + |f - g|} \, d\mu.$$

Define two measurable functions to be equivalent provided that they are equal almost everywhere on X. Show that ρ induces a metric on the collection of equivalence classes. (Suggestion: Consider the preceding problem.)

4. For $0 < p < 1$, show that
$$(a+b)^p \leq a^p + b^p \text{ for all } a, b \geq 0.$$

5. Let $(X, \mathcal{M}, \mu)$ be measure space and $0 < p < 1$. For measurable functions $g\colon X \to \mathbf{R}$ and $h\colon X \to \mathbf{R}$ that have integrable p-th powers, define
$$\rho_p(h, g) = \int_E |g-h|^p \, d\mu.$$
Define two such functions to be equivalent provided that they are equal almost everywhere on X. Show that $\rho_p(\cdot, \cdot)$ induces a metric on the collection of equivalence classes. (Suggestion: Consider the preceding problem.)

6. Show that the triangle inequality for Euclidean space $\mathbf{R}^n$ follows from the triangle inequality for $L^2\,[0, 1]$.

13.2 OPEN SETS, CLOSED SETS, AND CONVERGENT SEQUENCES

Many concepts studied for subsets of Euclidean spaces and general normed linear spaces can be directly extended to general metric spaces: they do not depend on the ambient linear structure.

Definition *Let (X, ρ) be a metric space. For a point x in X and $r > 0$, the set*
$$B(x, r) \equiv \{x' \in X \mid \rho(x', x) < r\}$$
is called the **open ball** *centered at x of radius r. A subset $\mathcal{O}$ of X is said to be* **open** *provided that for every point $x \in \mathcal{O}$, there is an open ball centered at x that is contained in $\mathcal{O}$. For a point $x \in X$, an open set that contains x is called a* **neighborhood** *of x.*

Proposition 1 *Let X be a metric space. The whole set X and $\emptyset$ are open; the intersection of a finite collection of open subsets of X is open; and the union of any collection of open subsets of X is open.*

Proof It is clear that X and $\emptyset$ are open and the union of a collection of open sets is open. Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be open subsets of X. If these two sets are disjoint, then the intersection is the empty-set, which is open. Otherwise, let x belong to $\mathcal{O}_1 \cap \mathcal{O}_2$. Since $\mathcal{O}_1$ and $\mathcal{O}_2$ are open sets containing x, there are positive numbers δ_1 and δ_2 for which $B(x, \delta_1) \subseteq \mathcal{O}_1$ and $B(x, \delta_2) \subseteq \mathcal{O}_2$. Define $\delta = \min\{\delta_1, \delta_2\}$. Then the open ball $B(x, \delta)$ is contained $\mathcal{O}_1 \cap \mathcal{O}_2$. Therefore, $\mathcal{O}_1 \cap \mathcal{O}_2$ is open. $\square$

The following proposition, the proof of which is left as an exercise, provides a description, in the case the metric space X is a subspace of the metric space Y, of the open subsets of X in terms of the open subsets of Y.

Proposition 2 *Let X be a subspace of the metric space Y and E a subset of X. Then E is open in X if and only if $E = X \cap \mathcal{O}$, where $\mathcal{O}$ is open in Y.*

Definition *For a subset E of a metric space X, a point $x \in X$ is called a* **point of closure** *of E provided that every neighborhood of x contains a point in E. The collection of points of closure of E is called the* **closure** *of E and is denoted by $\overline{E}$.*

It is clear that $E \subseteq \overline{E}$. If E contains all of its points of closure, that is, $E = \overline{E}$, then the set E is said to be **closed.** For a point x in the metric space (X, ρ) and $r > 0$, the set $\overline{B}(x, r) \equiv \{x' \in X \mid \rho(x', x) \leq r\}$ is called the **closed ball** centered at x of radius r. It follows from the triangle inequality for the metric that $\overline{B}(x, r)$ is closed[1].

Proposition 3 *For E a subset of a metric space X, its closure $\overline{E}$ is closed. Moreover, $\overline{E}$ is the smallest closed subset of X containing E, in the sense that if F is closed and $E \subseteq F$, then $\overline{E} \subseteq F$.*

Proof The set $\overline{E}$ is closed if it contains all its points of closure. Let x be a point of closure of $\overline{E}$. Consider a neighborhood $\mathcal{U}_x$ of x. There is a point $x' \in \overline{E} \cap \mathcal{U}_x$. Since x' is a point of closure of E and $\mathcal{U}_x$ is a neighborhood of x', there is a point $x'' \in E \cap \mathcal{U}_x$. Therefore, every neighborhood of x contains a point of E and hence $x \in \overline{E}$. So the set $\overline{E}$ is closed. It is clear that if $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$, and hence if F is closed and contains E, then $\overline{E} \subseteq \overline{F} = F$. $\square$

Proposition 4 *A subset of a metric space X is open if and only if its complement in X is closed.*

Proof First suppose E is an open subset of X. Let x be a point of closure of $X \sim E$. Then x cannot belong to E, because otherwise there would be a neighborhood of x that is contained in E and thus disjoint from $X \sim E$. Therefore, x belongs to $X \sim E$, so that $X \sim E$ is closed. Now suppose $X \sim E$ is closed. Let x belong to E. Then there must be a neighborhood of x that is contained in E, since otherwise, x is a point of closure of the closed set $X \sim E$, and so $x \in X \sim E$. This is a contradiction. $\square$

The proof of the following proposition is left as an exercise in the use of the De Morgan's Identities.

Proposition 5 *Let X be a metric space. Then whole set X and $\emptyset$ are closed; the union of any finite collection closed subsets of X is closed; and the intersection of any collection of closed subsets of X is closed.*

Convergence of a sequence in a normed linear space has been defined. The following is the natural generalization of sequential convergence to metric spaces.

Definition *A sequence $\{x_n\}$ in a metric space (X, ρ) is said to* **converge** *to the point $x \in X$ provided that*

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

The point to which the sequence converges is called the **limit** *of the sequence and $\{x_n\} \to x$ in X denotes the convergence of $\{x_n\}$ to x.*

[1]The notation needs clarification. If the metric is induced by a norm, the closed ball $\overline{B}(x, r)$ is the closure of the open ball $B(x, r)$, but this is not so in general (see Problem 12).

A sequence in a metric space can converge to at most one point. Indeed, given two points u, v in a metric space X, set $r = \rho(u, v)/2$. It follows from the triangle inequality for the metric ρ that $B(u, r)$ and $B(v, r)$ are disjoint. So it is not possible for a sequence to converge to both u and v.

Proposition 6 *For a subset E of a metric space X, a point $x \in X$ is a point of closure of E if and only if x is the limit of a sequence in E. Therefore, E is closed if and only if whenever a sequence in E converges to a limit $x \in X$, the limit x belongs to E.*

Proof It suffices to prove the first assertion. First suppose x belongs to $\overline{E}$. For each natural number n, since $B(x, 1/n) \cap E \neq \emptyset$, there is a point, which we label x_n, that belongs to $B(x, 1/n) \cap E$. Then $\{x_n\}$ is a sequence in E that we claim converges to x. Indeed, let $\epsilon > 0$. There is an index N for which $1/N < \epsilon$. Then

$$\rho(x_n, x) < 1/n < 1/N < \epsilon \text{ if } n \geq N.$$

Thus, $\{x_n\}$ converges to x. Conversely, if a sequence in E converges to x, then every ball centered at x contains infinitely many terms of the sequence and therefore contains points in E. So $x \in \overline{E}$. $\square$

Proposition 7 *Let $X = \prod_{n=1}^{\infty} X_n$ be the Cartesian product of the countable collection of metric spaces $\{(X_n, \rho_n)\}_{n=1}^{\infty}$. For points $u = \{u_n\}$ and $v = \{v_n\}$ in X, define*

$$\rho(u, v) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \frac{\rho_n(u_n, v_n)}{1 + \rho_n(u_n, v_n)}.$$

Then ρ is a metric on X for which sequential convergence is componentwise convergence in each component metric space (X_n, ρ_n).

Proof For non-negative numbers a, b, c with $a < b + c$,

$$\frac{a}{1+a} \leq \frac{b}{1+b} + \frac{c}{1+c}.$$

Therefore, for each n,

$$\tau_n(u, v) = \frac{\rho_n(u, v)}{1 + \rho_n(u, v)} \text{ for } u, v \in X_n$$

defines a metric on X_n, and sequential convergence with respect τ_n is the same as with respect to ρ_n. We leave it as an exercise to conclude from this that ρ is a metric. $\square$

PROBLEMS

7. Show that the following induce the same convergent sequences in Euclidean space:

$$\rho^*(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|;$$

$$\rho^+(x, y) = \max\{|x_1 - y_1|, \ldots, |x_n - y_n|\}.$$

8. Find a metric on $\mathbf{R}^n$ that does not induce the same convergent sequences as the metric induced by the Euclidean norm.
9. Show that in a metric space an open ball is open.
10. In a metric space X, is it possible for $r > 0$ and two distinct points u and v in X to have $B(u, r) = B(v, r)$? Is this possible in Euclidean space $\mathbf{R}^n$? Is it possible in a normed linear space?
11. Let (X, ρ) be a metric space in which $\{u_n\} \to u$ and $\{v_n\} \to v$. Show that $\{\rho(u_n, v_n)\} \to \rho(u, v)$.
12. Let X be a metric space, x belong to X and $r > 0$.
    (i) Show that $\overline{B}(x, r)$ is closed and contains $B(x, r)$.
    (ii) Show that in a normed linear space X the closed ball $\overline{B}(x, r)$ is the closure of the open ball $B(x, r)$, but this is not so in a general metric space.
13. Prove Proposition 2.
14. Let X be a subspace of the metric space Y and A a subset of X. Show that A is closed in X if and only if $A = X \cap F$, where F is closed in Y.
15. Let X be a subspace of the metric space Y.
    (i) If $\mathcal{O}$ is an open subset of the metric subspace X, is $\mathcal{O}$ an open subset of Y? What if X is an open subset of Y?
    (ii) If F is a closed subset of the metric subspace X, is F a closed subset of Y? What if X is a closed subset of Y?
16. For a subset E of a metric space X, a point $x \in X$ is called an interior point of E provided that there is an open ball centered at x that is contained in E: the collection of interior points of E is called the interior of E and denoted by $\operatorname{int} E$. Show that $\operatorname{int} E$ is always open and E is open if and only if $E = \operatorname{int} E$.
17. For a subset E of a metric space X, a point $x \in X$ is called an exterior point of E provided that there is an open ball centered at x that is contained in $X \sim E$: the collection of exterior points of E is called the exterior of E and denoted by $\operatorname{ext} E$. Show that $\operatorname{ext} E$ is always open. Show that E is closed if and only if $X \sim E = \operatorname{ext} E$.
18. For a subset E of a metric space X, a point $x \in X$ is called a boundary point of E provided that every open ball centered at x contains points in E and points in $X \sim E$: the collection of boundary points of E is called the boundary of E and denoted by $\operatorname{bd} E$. Show (i) that $\operatorname{bd} E$ is always closed, (ii) that E is open if and only if $E \cap \operatorname{bd} E = \emptyset$, and (iii) that E is closed if and only if $\operatorname{bd} E \subseteq E$.
19. Let A and B be subsets of a metric space X. Show that if $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$. Also, show that $\overline{(A \cup B)} = \overline{A} \cup \overline{B}$ and $\overline{(A \cap B)} \subseteq \overline{A} \cap \overline{B}$.
20. Show that for a subset E of a metric space X, the closure of E is the intersection of all closed subsets of X that contain E.
21. Let ρ be a metric on a set X. Define

$$\tau(u, v) = \frac{\rho(u, v)}{1 + \rho(u, v)} \text{ for all } u, v \in X.$$

Verify that τ is a bounded metric on X and convergence of sequences with respect to the ρ metric and the τ metric is the same. Conclude that sets that are closed with respect to the ρ metric are closed with respect to the τ metric and that sets that are open with respect to the ρ metric are open with respect to the τ metric.

13.3 CONTINUOUS MAPPINGS BETWEEN METRIC SPACES

Definition *A mapping $f: X \to Y$ between metric spaces X and Y is said to be continuous at the point $x \in X$ provided that for any sequence $\{x_n\}$ in X,*

$$\text{if } \{x_n\} \to x, \text{ then } \{f(x_n)\} \to f(x).$$

The mapping f is said to be **continuous** *provided that it is continuous at every point in X.*

The following three propositions are generalizations of corresponding results for real-valued functions of a real variable, and the proofs of the general results are essentially the same as the proofs in this special case.

The ϵ-δ Criterion for Continuity *A mapping f from a metric space (X, ρ) to a metric space (Y, σ) is continuous at the point $x \in X$ if and only if for every $\epsilon > 0$, there is a $\delta > 0$ for which if $\rho(x, x') < \delta$, then $\sigma(f(x), f(x')) < \epsilon$, that is,*

$$f(B(x, \delta)) \subseteq B(f(x), \epsilon).$$

Proof First suppose that $f: X \to Y$ is continuous at x. We establish the ϵ-δ criterion for continuity by arguing by contradiction. Suppose there is some $\epsilon_0 > 0$ for which there is no positive number δ for which $f(B(x, \delta)) \subseteq B(f(x), \epsilon_0)$. In particular, if n is a natural number, it is not true that $f(B(x, 1/n)) \subseteq B(f(x), \epsilon_0)$. This means that there is a point in X, which we label x_n, such that $\rho(x, x_n) < 1/n$ while $\sigma(f(x), f(x_n)) \geq \epsilon_0$. This defines a sequence $\{x_n\}$ in X that converges to x, but for which the image sequence $\{f(x_n)\}$ does not converge to $f(x)$. This contradicts the continuity of the mapping $f: X \to Y$ at the point x.

To prove the converse, suppose the ϵ-δ criterion holds. Let $\{x_n\}$ be a sequence in X that converges to x. We must show that $\{f(x_n)\}$ converges to $f(x)$. Let $\epsilon > 0$. We can choose a positive number δ for which $f(B(x, \delta)) \subseteq B(f(x), \epsilon)$. Moreover, since the sequence $\{x_n\}$ converges to x, we can select an index N such that $x_n \in B(x, \delta)$ for $n \geq N$. Hence, $f(x_n) \in B(f(x), \epsilon)$ for $n \geq N$. Thus, the sequence $\{f(x_n)\}$ converges to $f(x)$ and therefore $f: X \to Y$ is continuous at the point x. □

Proposition 8 *A mapping $f: X \to Y$ between metric spaces X and Y is continuous if and only if for each open subset $\mathcal{O}$ of Y, the inverse image under f of $\mathcal{O}$, $f^{-1}(\mathcal{O})$, is an open subset of X.*

Proof First assume that the mapping f is continuous. Let $\mathcal{O}$ be an open subset of Y. Let x be a point in $f^{-1}(\mathcal{O})$; we must show that an open ball centered at x is contained in $f^{-1}(\mathcal{O})$. But $f(x)$ is a point in $\mathcal{O}$, which is open in Y, so there is some positive number r for which $B((f(x), r)) \subseteq \mathcal{O}$. Since $f: X \to Y$ is continuous at the point x, by the ϵ-δ criterion for continuity at a point, we can select a positive number δ for which $f(B(x, \delta)) \subseteq B(f(x), r) \subseteq \mathcal{O}$. Thus, $B(x, \delta) \subseteq f^{-1}(\mathcal{O})$ and therefore, $f^{-1}(\mathcal{O})$ is open in X.

To prove the converse, suppose the inverse image under f of each open set is open. Let x be a point in X. To show that f is continuous at x, we use the ϵ-δ criterion for continuity.

Let $\epsilon > 0$. The open ball $B(f(x), \epsilon)$ is an open subset of Y. Thus, $f^{-1}(B(f(x), \epsilon))$ is open in X. Therefore, we can choose a positive number δ with $B(x, \delta) \subseteq f^{-1}(B(f(x), \epsilon))$, that is, $f(B(x, \delta)) \subseteq B(f(x), \epsilon)$. □

Proposition 9 *The composition of continuous mappings between metric spaces is continuous.*

Proof Let $f\colon X \to Y$ be continuous and $g\colon Y \to Z$ be continuous, where X, Y, and Z are metric spaces. We use the preceding proposition. Let $\mathcal{O}$ be open in Z. Since g is continuous, $g^{-1}(\mathcal{O})$ is open in Y and therefore, since f is continuous, $f^{-1}(g^{-1}(\mathcal{O})) = (g \circ f)^{-1}(\mathcal{O})$ is open in X. Therefore, $g \circ f$ is continuous. □

Definition *A mapping from a metric space (X, ρ) to a metric space (Y, σ) is said to be* **uniformly continuous** *provided that for every $\epsilon > 0$, there is a $\delta > 0$ such that for $u, v \in X$,*

$$\textit{if } \rho(u, v) < \delta, \textit{ then } \sigma(f(u), f(v)) < \epsilon.$$

It follows from the ϵ-δ criterion for continuity at a point that a uniformly continuous mapping is continuous. The converse is false.

Example A mapping f from a metric space (X, ρ) to a metric space (Y, σ) is said to be **Lipschitz** provided that there is a $c \geq 0$ such that for all $u, v \in X$,

$$\sigma(f(u), f(v)) \leq c \cdot \rho(u, v).$$

A Lipschitz mapping is uniformly continuous since, regarding the criterion for uniform continuity, $\delta = \epsilon/c$ responds to any $\epsilon > 0$ challenge.

PROBLEMS

22. Exhibit a continuous mapping that is not uniformly continuous and a uniformly continuous mapping that is not Lipschitz.
23. Show that every mapping from a metric space (X, ρ) to a metric space (Y, σ) is continuous if ρ is the discrete metric.
24. Suppose there is a continuous, one-to-one mapping from a metric space (X, ρ) to a metric space (Y, σ), where σ is the discrete metric. Show that every subset of X is open.
25. For a metric space (X, ρ), show that the metric $\rho\colon X \times X \to \mathbf{R}$ is continuous, where $X \times X$ has the product metric.
26. Let z be a point in the metric space (X, ρ). Define the function $f\colon X \to \mathbf{R}$ by $f(x) = \rho(x, z)$. Show that f is uniformly continuous.
27. For a non-empty subset E of the metric space (X, ρ) and a point $x \in X$, define the distance from x to E, $dist_E(x)$, as follows:

$$dist_E(x) = \inf\{\rho(x, y) \mid y \in E\}.$$

   (i) Show that $|dist_E(x) - dist_E(y)| \leq \rho(x, y)$ for all $x, y \in X$.
   (ii) Show that $\{x \in X \mid dist_E(x) = 0\} = \overline{E}$.

28. Show that a subset E of a metric space X is open if and only if there is a continuous real-valued function f on X for which $E = \{x \in X \mid f(x) > 0\}$.
29. Show that a subset E of a metric space X is closed if and only if there is a continuous real-valued function f on X for which $E = f^{-1}(0)$.
30. Let $X = C[a, b]$. Define the function $\psi\colon X \to \mathbf{R}$ by

$$\psi(f) = \int_a^b f(x)\,dx \quad \text{for each } f \text{ in } X.$$

Show that ψ is Lipschitz on the metric space X, where X has the metric induced by the maximum norm.

13.4 COMPLETE METRIC SPACES

The completeness of the real numbers is an indispensable ingredient in the proof of the most basic theorems of analysis. Completeness has been defined for normed linear spaces, and is now extended to metric spaces. In the next chapter, three fundamental theorems for complete metric spaces are considered.

Definition *A sequence $\{x_n\}$ in a metric space (X, ρ) is said to be a* **Cauchy sequence** *provided that for each $\epsilon > 0$, there is an index N for which*

$$\textit{if } n, m \geq N, \ \textit{then } \rho(x_n, x_m) < \epsilon.$$

The space (X, ρ) is said to be **complete** *provided that every Cauchy sequence in X converges to a point in X.*

In general, a subspace of a complete metric space is not complete. For instance, an open, bounded interval of real numbers is not complete, while $\mathbf{R}$ is complete. However, there is the following simple characterization of those subspaces that are complete.

Proposition 10 *Let E be a subset of the complete metric space X. Then the metric subspace E is complete if and only if E is a closed subset of X.*

Proof First suppose E is a closed subset of X. Let $\{x_n\}$ be a Cauchy sequence in E. Then $\{x_n\}$ can be considered as a Cauchy sequence in X and X is complete. Thus, $\{x_n\}$ converges to a point x in X. According to Proposition 6, since E is a closed subset of X, the limit of a convergent sequence in E belongs to E. Thus, x belongs to E and hence E is a complete metric space.

To prove the converse, suppose E is complete. According to Proposition 6, to show that E is a closed subset of X, it must be verified that the limit of a convergent sequence in E also belongs to E. Let $\{x_n\}$ be a sequence in E that converges to $x \in X$. But a convergent sequence is Cauchy. Thus, by the completeness of E, $\{x_n\}$ converges to a point in E. But a convergent sequence in a metric space has only one limit. Thus, x belongs to E. $\square$

Euclidean spaces and $C[a, b]$, with the maximum metric, inherit completeness from completeness of the real numbers. In the preceding chapter, we proved the Riesz-Fischer Theorem, according to which, for any measure space (X, M, μ) and $1 \leq p \leq \infty$, $L^p(X, \mu)$ is complete. Therefore, by the preceding proposition, we have the following collection of completed metric spaces.

Corollary 11 *The following are complete metric spaces:*

(i) Each closed subset of Euclidean space $\mathbf{R}^n$.
(ii) Each closed subset of $L^p(X, \mu)$, for $1 \leq p \leq \infty$.
(iii) Each closed subset of $C[a, b]$.

Definition *For a non-empty subset E of a metric space (X, ρ), the* **diameter** *of E, $\operatorname{diam} E$, is defined by*

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

The set E is said to be **bounded** *provided that it has finite diameter. A descending sequence $\{E_n\}_{n=1}^{\infty}$ of non-empty subsets of X is called a* **contracting sequence** *provided that*

$$\lim_{n \to \infty} \operatorname{diam}(E_n) = 0.$$

The Nested Set Theorem of Chapter 1 implies that the intersection of a contracting sequence of non-empty, closed sets of real numbers consists of a single point. This generalizes as follows.

The Cantor Intersection Theorem *Let X be a metric space. Then X is complete if and only if whenever $\{F_n\}_{n=1}^{\infty}$ is a contracting sequence of non-empty, closed subsets of X, there is a point $x \in X$ for which $\bigcap_{n=1}^{\infty} F_n = \{x\}$.*

Proof First assume that X is complete. Let $\{F_n\}_{n=1}^{\infty}$ be a contracting sequence of non-empty, closed subsets of X. For each index n, select $x_n \in F_n$. We claim that $\{x_n\}$ is a Cauchy sequence. Indeed, let $\epsilon > 0$. There is an index N for which $\operatorname{diam} F_N < \epsilon$. Since $\{F_n\}_{n=1}^{\infty}$ is descending, if $n, m \geq N$, then x_n and x_m belong to F_N and therefore $\rho(x_n, x_m) \leq \operatorname{diam} F_N < \epsilon$. Thus, $\{x_n\}$ is a Cauchy sequence. Since X is complete, this sequence converges to some $x \in X$. However, for each index n, F_n is closed and $x_k \in F_n$ for $k \geq n$ so that x belongs to F_n. Thus, x belongs to $\bigcap_{n=1}^{\infty} F_n$. It is not possible for the intersection to contain two distinct points for, if it did, $\lim_{n \to \infty} \operatorname{diam} F_n \neq 0$.

To prove the converse, suppose that for any contracting sequence $\{F_n\}_{n=1}^{\infty}$ of non-empty, closed subsets of X, there is a point $x \in X$ for which $\bigcap_{n=1}^{\infty} F_n = \{x\}$. Let $\{x_n\}$ be a Cauchy sequence in X. For each index n define F_n to be the closure of the non-empty set $\{x_k \mid k \geq n\}$. Then $\{F_n\}$ is a descending sequence of non-empty closed sets. Since $\{x_n\}$ is Cauchy, the sequence $\{F_n\}$ is contracting. Thus, by assumption, there is a point x in X for which $\{x\} = \bigcap_{n=1}^{\infty} F_n$. For each index n, x is a point of closure of $\{x_k \mid k \geq n\}$ and therefore any ball centered at x has non-empty intersection with $\{x_k \mid k \geq n\}$. Hence, we may inductively select a strictly increasing sequence of natural numbers $\{n_k\}$ such that for each index k, $\rho(x, x_{n_k}) < 1/k$. The subsequence $\{x_{n_k}\}$ converges to x. Since $\{x_n\}$ is Cauchy, the whole sequence $\{x_n\}$ converges to x (see Problem 32). Therefore, X is complete. $\square$

A very rough geometric interpretation of the Cantor Intersection Theorem is that a metric space fails to be complete because it has "holes." If X is an incomplete metric space, it can always be suitably minimally enlarged to become complete. For example, the set of rational numbers is not complete, but it is a dense metric subspace of the complete space $\mathbf{R}$. As a further example, let $X = C[a, b]$, now considered with the norm $\|\cdot\|_1$, which it inherits from $L^1[a, b]$. The metric space (X, ρ_1) is not complete. But it is a dense metric subspace of the complete metric space $L^1[a, b]$. These are two specific examples of a construction that has a quite abstract generalization. We outline a proof of the following theorem in Problem 40.

Theorem 12 *Let (X, ρ) be a metric space. Then there is a complete metric space $(\widetilde{X}, \widetilde{\rho})$ for which X is a dense subset of $\widetilde{X}$ and*

$$\rho(u, v) = \widetilde{\rho}(u, v) \text{ for all } u, v \in X.$$

We call the metric space described above the **completion** of (X, ρ). The completion is unique, in the sense that any two completions are isometric by way of an isometry that is the identity mapping on X.

PROBLEMS

31. In a metric space X, show (i) that a convergent sequence is Cauchy and (ii) that a Cauchy sequence is bounded.
32. In a metric space X, show that a Cauchy sequence converges if and only if it has a convergent subsequence.
33. Define the concept of rapidly Cauchy sequence in a metric space, and show that a space is complete if and only if every rapidly Cauchy sequence converges.
34. Provide an example of a descending, countable collection of closed, non-empty sets of real numbers for which the intersection is empty. Does this contradict the Cantor Intersection Theorem?
35. For a mapping f of the metric space (X, ρ) to the metric space (Y, σ), show that f is uniformly continuous if and only if for any two sequences $\{u_n\}$ and $\{v_n\}$ in X,
$$\text{if } \lim_{n\to\infty} \rho(u_n, v_n) = 0, \text{ then } \lim_{n\to\infty} \sigma(f(u_n), f(v_n))) = 0.$$
36. For the countable Cartesian product of metric spaces defined in Proposition 7, show that the product is complete if and only if each factor space is complete.
37. For each index n, define $f_n(x) = \alpha x^n + \beta \cos(x/n)$ for $0 \le x \le 1$. For what values of the parameters α and β is the sequence $\{f_n\}$ a Cauchy sequence in the metric space $C[0, 1]$?
38. Let $\mathcal{D}$ be the subspace of $C[0, 1]$ consisting of the continuous functions $f\colon [0, 1] \to \mathbf{R}$ that are differentiable on $(0, 1)$. Is $\mathcal{D}$ complete?
39. Define $\mathcal{L}$ to be the subspace of $C[0, 1]$ consisting of the functions $f\colon [0, 1] \to \mathbf{R}$ that are Lipschitz. Is $\mathcal{L}$ complete?
40. For a metric space (X, ρ), complete the following outline of a proof of Theorem 12:
    (i) If $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences in X, show that $\{\rho(x_n, y_n)\}$ is a Cauchy sequence of real numbers and therefore converges.
    (ii) Define X' to be the set of Cauchy sequences in X. For two Cauchy sequences in X, $\{x_n\}$ and $\{y_n\}$, define $\rho'(\{x_n\}, \{y_n\}) = \lim \rho(x_n, y_n)$.

(iii) Define two members of X', that is, two Cauchy sequences $\{x_n\}$ and $\{y_n\}$ in X, to be equivalent, provided that $\rho'(\{x_n\}, \{y_n\}) = 0$. Show that this is an equivalence relation in X' and denote by $\widehat{X}$ the set of equivalence classes. Define the distance $\widehat{\rho}$ between two equivalence classes to be the ρ' distance between representatives of the classes. Show that $\widehat{\rho}$ is properly defined and is a metric on $\widehat{X}$.

(iv) Show that the metric space $(\widehat{X}, \widehat{\rho})$ is complete. (Suggestion: If $\{x_n\}$ is a Cauchy sequence from X, we may assume [by taking subsequences] that $\rho(x_n, x_{n+1}) < 2^{-n}$ for all n. If $\{\{x_{n,m}\}_{n=1}^{\infty}\}_{m=1}^{\infty}$ is a sequence of such Cauchy sequences that represents a Cauchy sequence in $\widehat{X}$, then the sequence $\{x_{n,n}\}_{n=1}^{\infty}$ is a Cauchy sequence from X that represents the limit of the Cauchy sequences from $\widehat{X}$.)

(v) Define the mapping h from X to $\widehat{X}$ by defining, for $x \in X$, $h(x)$ to be the equivalence class of the constant sequence, each term of which is x. Show that $h(X)$ is dense in $\widehat{X}$ and that $\widehat{\rho}(h(u), h(v)) = \rho(u, v)$ for all $u, v \in X$.

(vi) Define the set $\widetilde{X}$ to be the disjoint union of X and $\widehat{X} \sim h(X)$. For $u, v \in \widetilde{X}$, define $\widetilde{\rho}(u, v)$ as follows: $\widetilde{\rho}(u, v) = \rho(u, v)$ if $u, v \in X$; $\widetilde{\rho}(u, v) = \widehat{\rho}(u, v)$ for $u, v \in \widehat{X} \sim h(X)$; and $\widetilde{\rho}(u, v) = \widehat{\rho}(h(u), v)$ for $u \in X$, $v \in \widehat{X} \sim h(X)$. From the preceding two parts conclude that the metric space $(\widetilde{X}, \widetilde{\rho})$ is a complete metric space containing (X, ρ) as a dense subspace.

13.5 COMPACT METRIC SPACES

Recall that a collection of sets $\{E_\lambda\}_{\lambda \in \Lambda}$ is said to be a **cover** of a set E provided that $E \subseteq \bigcup_{\lambda \in \Lambda} E_\lambda$. By a subcover of a cover of E is meant a subcollection of the cover which itself also is a cover of E. If E is a subset of a metric space X, by an **open cover** of E is meant a cover of E consisting of open subsets of X. In Chapter 10, we defined what it means for a subset of $\mathbf{R}^n$ to be compact and generalize this as follows.

Definition *A metric space X is called* **compact** *provided that every open cover of X has a finite subcover. A subset K of X is called compact provided that K, considered as a metric subspace of X, is compact.*

An open subset of the subspace K of a metric space X is the intersection of K with an open subset of X. Therefore a subset K of a metric space X is compact if and only if each cover of K by a collection of open subsets of X has a finite subcover.

If $\mathcal{T}$ is a collection of open subsets of a metric space X, then the collection $\mathcal{F}$ of complements of sets in $\mathcal{T}$ is a collection of closed sets. Moreover, $\mathcal{T}$ is a cover if and only if $\mathcal{F}$ has an empty intersection. Therefore, by De Morgan's Identities, a metric space X is compact if and only if every collection of closed sets with an empty intersection has a finite subcollection for which the intersection also is empty. A collection $\mathcal{F}$ of sets in X is said to have the **finite intersection property** provided that any finite subcollection of $\mathcal{F}$ has a non-empty intersection. Therefore, the concept of compactness has the following formulation in terms of closed sets.

Proposition 13 *A metric space X is compact if and only if every collection $\mathcal{F}$ of closed subsets of X with the finite intersection property has a non-empty intersection.*

Definition *By an* ϵ-**net** *for a metric space* X *is meant a finite collection of open balls* $\{B(x_k, \epsilon)\}_{k=1}^n$ *with centers* x_k *in* X *that covers* X. *A metric space* X *is said to be* **totally bounded** *provided that for each* $\epsilon > 0$, *there is an* ϵ*-net for* X.

If a metric space X is totally bounded, then it is bounded, in the sense that its diameter is finite. Indeed, if X is covered by a finite number of balls of radius r, then it follows from the triangle inequality that $\operatorname{diam} X \leq c$, where $c = 2 \cdot r + d$, d being the maximum distance between the centers of the covering balls. However, as is seen in the following example, a bounded metric space may not be totally bounded.

Example Let X be the Banach space ℓ^2 of square summable sequences of real numbers. Consider the closed unit ball $B = \{\{x_n\} \in \ell^2 \mid ||\{x_n\}||_2 \leq 1\}$. Then B is bounded. We claim that B is not totally bounded. Indeed, for each natural number n, let e_n have n-th component 1 and other components 0. Then $||e_n - e_m||_2 = \sqrt{2}$ if $m \neq n$. Then B cannot be contained in a finite number of balls of radius $r < 1/2$ since one of these balls would contain two of the e_n's, which are distance $\sqrt{2}$ apart and yet the ball has diameter less than 1.

Proposition 14 *A subset of Euclidean space* $\mathbf{R}^n$ *is bounded if and only if it is totally bounded.*

Proof It is always the case that a totally bounded metric space is bounded. To prove the converse, let E be a bounded subset of $\mathbf{R}^n$. For simplicity take $n = 2$. Let $\epsilon > 0$. Since E is bounded, for $a > 0$ sufficiently large, E is contained in the square $[-a, a] \times [-a, a]$. Let P_k be a partition of $[-a, a]$ for which each partition interval has length less than $1/k$. Then $P_k \times P_k$ induces a partition of $[-a, a] \times [-a, a]$ into closed rectangles of diameter at most $\sqrt{2}/k$. Choose k such that $\sqrt{2}/k < \epsilon$. Consider the finite collection of balls of radius ϵ with centers (x, y) where x and y are partition points of P_k. Then this finite collection of balls of radius ϵ covers the square $[-a, a] \times [-a, a]$ and therefore also covers E. Therefore, E is totally bounded. $\square$

Definition *A metric space* X *is said to be* **sequentially compact** *provided that every sequence in* X *has a subsequence that converges to a point in* X.

Theorem 15 (**Characterization of Compactness for a Metric Space**) *For a metric space* X, *the following three assertions are equivalent:*

(i) X *is complete and totally bounded;*
(ii) X *is compact;*
(iii) X *is sequentially compact.*

The proof is separated into three propositions.

Proposition 16 *If a metric space* X *is complete and totally bounded, then it is compact.*

Proof We argue by contradiction. Suppose $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda}$ is an open cover of X for which there is no finite subcover. Since X is totally bounded, there is a finite collection of open balls of radius less than $1/2$ that cover X. There must be one of these balls that cannot be covered by a finite subcollection of $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda}$. Select such a ball and label its closure F_1. Then F_1

is closed and $\operatorname{diam} F_1 \leq 1$. Once more using the total boundedness of X, there is a finite collection of open balls of radius less than $1/4$ that cover X. This collection also covers F_1. There must be one of these balls for which the intersection with F_1 cannot be covered by a finite subcollection of $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$. Define F_2 to be the closure of the intersection of such a ball with F_1. Then F_1 and F_2 are closed, $F_2 \subseteq F_1$, and $\operatorname{diam} F_1 \leq 1, \operatorname{diam} F_2 \leq 1/2$. Continuing in this way, a contracting sequence of non-empty, closed sets $\{F_n\}$ is obtained, with the property that each F_n cannot be covered by a finite subcollection of $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$. But X is complete. According to the Cantor Intersection Theorem there is a point x_0 in X that belongs to the intersection $\bigcap_{n=1}^\infty F_n$. There is some λ_0 such that $\mathcal{O}_{\lambda_0}$ contains x_0 and since $\mathcal{O}_{\lambda_0}$ is open, there is a ball centered at x_0, $B(x_0, r)$, such that $B(x_0, r) \subseteq \mathcal{O}_{\lambda_0}$. Since $\lim_{n\to\infty} \operatorname{diam} F_n = 0$ and $x_0 \in \bigcap_{n=1}^\infty F_n$, there is an index n such that $F_n \subseteq \mathcal{O}_{\lambda_0}$. This contradicts the choice of F_n as being a set that cannot be covered by a finite subcollection of $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$. This contradiction shows that X is compact. $\square$

Proposition 17 *If a metric space X is compact, then it is sequentially compact.*

Proof Let $\{x_n\}$ be a sequence in X. For each index n, let F_n be the closure of the non-empty set $\{x_k \mid k \geq n\}$. Then $\{F_n\}$ is a descending sequence of non-empty closed sets. It follows from Proposition 13 that there is a point x_0 in X which belongs to the intersection $\bigcap_{n=1}^\infty F_n$. Since for each n, x belongs to the closure of $\{x_k \mid k \geq n\}$, the ball $B(x_0, 1/k)$ has non-empty intersection with $\{x_k \mid k \geq n\}$. By induction, there is a strictly increasing sequence of indices $\{n_k\}$ such that for each index k, $\rho(x_0, x_{n_k}) < 1/k$. The subsequence $\{x_{n_k}\}$ converges to x_0. Thus, X is sequentially compact. $\square$

Proposition 18 *If a metric space X is sequentially compact, then it is complete and totally bounded.*

Proof We argue by contradiction to establish total boundedness. Suppose X is not totally bounded. Then there is an $\epsilon > 0$ for which X cannot be covered by a finite number of open balls of radius ϵ. Select a point x_1 in X. Since X is not contained in $B(x_1, \epsilon)$, there is an $x_2 \in X$ for which $\rho(x_1, x_2) \geq \epsilon$. Now since X is not contained in $B(x_1, \epsilon) \cup B(x_2, \epsilon)$, there is an $x_3 \in X$ for which $\rho(x_3, x_2) \geq \epsilon$ and $\rho(x_3, x_1) \geq \epsilon$. Continuing in this way, a sequence $\{x_n\}$ in X is defined with the property that $\rho(x_n, x_k) \geq \epsilon$ for $n > k$. Then the sequence $\{x_n\}$ can have no convergent subsequence, since any two different terms of any subsequence are a distance ϵ or more apart. Thus, X is not sequentially compact. This contradiction shows that X must be totally bounded.

To show that X is complete, let $\{x_n\}$ be a Cauchy sequence in X. Since X is sequentially compact, a subsequence of $\{x_n\}$ converges to a point $x \in X$. Using the Cauchy property it is not difficult to see that the whole sequence converges to x. Thus, X is complete. $\square$

These three propositions complete the proof of the Characterization of Compactness Theorem. Since Euclidean space $\mathbf{R}^n$ is complete, in view of Proposition 10, each closed subset is complete as a metric subspace. Moreover, Proposition 14 asserts that a subset of Euclidean space is bounded if and only if it is totally bounded. Therefore, by the

Characterization of Compactness Theorem for metric spaces, there is the following characterization of compactness for a subspace of Euclidean space.

Theorem 19 (**Characterization of Compactness in $\mathbf{R}^n$**) *For a subset K of $\mathbf{R}^n$, the following three assertions are equivalent:*

(i) K is closed and bounded;

(ii) K is compact;

(iii) K is sequentially compact.

Regarding this theorem, the equivalence of (i) and (ii) is known as the Heine-Borel Theorem and that of (i) and (iii) as the Bolzano-Weierstrass Theorem.

Proposition 20 *If X is a compact metric space, Y is a metric space, and the mapping $f\colon X \to Y$ is continuous, then its image $f(X)$ also is compact.*

Proof Let $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$ be an open covering of $f(X)$. Then, by the continuity of f, $\{f^{-1}(\mathcal{O}_\lambda)\}_{\lambda\in\Lambda}$ is an open cover of X. By the compactness of X, there is a finite subcollection $\{f^{-1}(\mathcal{O}_{\lambda_1}), \ldots, f^{-1}(\mathcal{O}_{\lambda_n})\}$ that also covers X. Since f maps X onto $f(X)$, the finite collection $\{\mathcal{O}_{\lambda_1}, \ldots, \mathcal{O}_{\lambda_n}\}$ covers $f(X)$. □

One of the first properties of functions of a real variable that is established in a calculus course, and which is proven in Chapter 1, is that a continuous function on a closed, bounded interval takes maximum and minimum values, and this extends to continuous functions on a closed, bounded subset of $\mathbf{R}^n$. It is natural to attempt to classify the metric spaces for which the extreme value property holds.

Theorem 21 (**Extreme Value Theorem**) *A metric space X is compact if and only if every continuous function $f\colon X \to \mathbf{R}$ has a maximum and a minimum value.*

Proof First assume that X is compact. Let the function $f\colon X \to \mathbf{R}$ be continuous. The preceding proposition implies that $f(X)$ is a compact set of real numbers. According to Theorem 19, $f(X)$ is closed and bounded. It follows from the completeness of $\mathbf{R}$ that a closed, bounded, non-empty set of real numbers has a largest and smallest member.

To prove the converse, assume that every continuous real-valued function on X takes a maximum and minimum value. According to Theorem 19, to show that X is compact it is necessary and sufficient to show it is totally bounded and complete. We argue by contradiction to show that X is totally bounded. If X is not totally bounded, then there is an $r > 0$ and a countably infinite subset of X, enumerated as $\{x_n\}_{n=1}^\infty$, for which the collection of open balls $\{B(x_n, r)\}_{n=1}^\infty$ is disjoint. For each natural number n, define the function $f_n\colon X \to \mathbf{R}$ by

$$f_n(x) = \begin{cases} r/2 - \rho(x, x_n) & \text{if } \rho(x, x_n) \le r/2 \\ 0 & \text{otherwise.} \end{cases}$$

Define the function $f\colon X \to \mathbf{R}$ by

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

Since each f_n is continuous and vanishes outside $B(x_n, r/2)$ and the collection $\{B(x_n, r)\}_{n=1}^{\infty}$ is disjoint, f is properly defined and continuous. But for each natural number n, $f(x_n) = n \cdot r/2$, and hence f is unbounded above and therefore does not take a maximum value. This is a contradiction. Therefore, X is totally bounded. It remains to show that X is complete. Let $\{x_n\}$ be a Cauchy sequence in X. Then for each $x \in X$, it follows from the triangle inequality that $\{\rho(x, x_n)\}$ is a Cauchy sequence of real numbers that, since $\mathbf{R}$ is complete, converges to a real number. Define the function $f\colon X \to \mathbf{R}$ by

$$f(x) = \lim_{n \to \infty} \rho(x, x_n) \text{ for all } x \in X.$$

Again by use of the triangle inequality, we conclude that f is continuous. By assumption, there is a point x in X at which f takes a minimum value. Since $\{x_n\}$ is Cauchy, the infimum of f on X is 0. Therefore, $f(x) = 0$ and hence $\{x_n\}$ converges to x. Thus, X is complete. □

If $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda}$ is an open cover of a metric space X, then each point $x \in X$ is contained in a member of the cover, $\mathcal{O}_\lambda$, and since $\mathcal{O}_\lambda$ is open, there is some $\epsilon > 0$ such that

$$B(x, \epsilon) \subseteq \mathcal{O}_\lambda. \tag{2}$$

In general, the ϵ depends on the choice of x. The following proposition implies that for a compact metric space this containment holds uniformly. in the sense that there is an ϵ independently of $x \in X$ for which the inclusion (2) holds. A positive number ϵ with this property is called a **Lebesgue number** for the cover $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda}$.

The Lebesgue Covering Lemma *Let $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda}$ be an open cover of a compact metric space X. Then there is a number $\epsilon > 0$ such that for each $x \in X$, the open ball $B(x, \epsilon)$ is contained in some member of the cover.*

Proof We argue by contradiction. Assume that there is no such positive Lebesgue number. Then for each natural number n, $1/n$ fails to be a Lebesgue number. Thus, there is a point in X, which we label x_n, for which $B(x_n, 1/n)$ fails to be contained in a single member of the cover. This defines a sequence $\{x_n\}$ in X. By the Characterization of Compactness Theorem, X is sequentially compact. Thus, a subsequence $\{x_{n_k}\}$ converges to a point $x_0 \in X$. Now there is some $\lambda_0 \in \Lambda$ for which $\mathcal{O}_{\lambda_0}$ contains x_0 and since $\mathcal{O}_{\lambda_0}$ is open, there is a ball centered at x_0, $B(x_0, r_0)$, for which

$$B(x_0, r_0) \subseteq \mathcal{O}_{\lambda_0}.$$

Choose an index k for which $\rho(x_0, x_{n_k}) < r_0/2$ and $1/n_k < r_0/2$. By the triangle inequality, $B(x_{n_k}, 1/n_k) \subseteq \mathcal{O}_{\lambda_0}$ and this contradicts the choice of x_{n_k} as being a point for which $B(x_{n_k}, 1/n_k)$ fails to be contained in a single member of the cover. □

Proposition 22 *If X is a compact metric space, Y is a metric space, and the mapping $f\colon X \to Y$ is continuous, then f is uniformly continuous.*

Proof Let $\epsilon > 0$. By the ϵ-δ criterion for continuity at a point, for each $x \in X$, there is a $\delta_x > 0$ for which if $\rho(x, x') < \delta_x$, then $\sigma(f(x), f(x')) < \epsilon/2$. Therefore, setting $\mathcal{O}_x = B(x, \delta_x)$, by the triangle inequality for σ,

$$\sigma(f(u), f(v)) \leq \sigma(f(u), f(x)) + \sigma(f(x), f(v)) < \epsilon \text{ if } u, v \in \mathcal{O}_x. \quad (3)$$

Let δ be a Lebesgue number for the open cover $\{\mathcal{O}_x\}_{x \in X}$. Then for $u, v \in X$, if $\rho(u, v) < \delta$ there is some x for which $u \in B(v, \delta) \subseteq \mathcal{O}_x$ and therefore, by (3), $\sigma(f(u), f(v)) < \epsilon$. $\square$

The proof of the following proposition is left as an exercise.

Proposition 23 *If a subspace K of a metric space X is compact, then K is a closed, bounded subset of X.*

In view of this proposition, it is interesting to determine the closed, bounded subsets of a metric space that are compact. As we have seen, all such subsets of Euclidean space compact. It follows from a theorem of Riesz that we will prove in Chapter 17, that in every infinite dimensional normed linear space, there are closed, bounded subsets that fail to be compact.

PROBLEMS

41. Consider the metric space $\mathbf{Q}$ consisting of the rational numbers with the metric induced by the absolute value. Which subspaces of $\mathbf{Q}$ are complete and which are compact?
42. Let $B = B(x, r)$ be an open ball in Euclidean space $\mathbf{R}^n$. Show that B fails to be compact by (i) showing B is not sequentially compact, (ii) finding an open cover of B without any finite subcover, and (iii) showing B is not closed.
43. When is a set X with the discrete metric a compact metric space?
44. Prove Proposition 23.
45. Prove that if X is compact, then $K \subseteq X$ is compact if and only if it is closed subset of X.
46. For the countable Cartesian product of metric spaces considered in Proposition 7, show that the product is compact if and only if each factor space is compact.
47. Let E be a subset of the compact metric space X. Show the metric subspace E is compact if and only if E is a closed subset of X.
48. (Fréchet Intersection Theorem). Let $\{F_n\}_{n=1}^{\infty}$ be a descending countable collection of non-empty closed subsets of a compact metric space X. Show that $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$.
49. For a subset E of a metric space X, show that E is totally bounded if and only if its closure $\overline{E}$ is totally bounded.
50. For a subset E of a complete metric space X, show that E is totally bounded if and only if its closure $\overline{E}$ is compact.
51. Let $B = \{\{x_n\} \in \ell^2 \mid \sum_{n=1}^{\infty} x_n^2 \leq 1\}$ be the closed unit ball in ℓ^2. Show that B fails to be compact by (i) showing B is not sequentially compact, (ii) finding an open cover of B without any finite subcover, and (iii) showing B is not totally bounded.
52. Let $B = \{f \in L^2[a, b] \mid ||f||_2 \leq 1\}$ be the closed unit ball in $L^2[a, b]$. Show that B fails to be compact by (i) showing B is not sequentially compact, (ii) finding an open cover of B without any finite subcover, and (iii) showing B is not totally bounded.

53. Let X be a totally bounded metric space.
   (i) If f is a uniformly continuous mapping from X to a metric space Y, show that $f(X)$ is totally bounded.
   (ii) Is (i) still true if f is only required to be continuous?
54. Let ρ be a metric on a set X. Define

$$\tau(u, v) = \frac{\rho(u, v)}{1 + \rho(u, v)} \text{ for all } u, v \in X.$$

   Verify that τ is a bounded metric on X and convergence of sequences with respect to the ρ metric and the τ metric is the same. Conclude that sets that are closed with respect to the ρ metric are closed with respect to the τ metric and that sets that are open with respect to the ρ metric are open with respect to the τ metric.
55. Let E be a subset of Euclidean space $\mathbf{R}^n$. Assume that every continuous real-valued function on E takes a minimum value. Is E necessarily closed and bounded?
56. Suppose f is a continuous real-valued function on Euclidean space $\mathbf{R}^n$ with the property that there is a positive number c such that $|f(x)| \geq c \cdot ||x||$ for all $x \in \mathbf{R}^n$. Show that if K is a compact set of real numbers, then its inverse image under f, $f^{-1}(K)$, also is compact. (Mappings with this property are called **proper**.)
57. For a compact metric space (X, ρ), show that there are points $u, v \in X$ for which $\rho(u, v) = \operatorname{diam} X$.
58. Let K be a compact subset of the metric space (X, ρ) and x_0 belong to X. Show that there is a point $z \in K$ for which

$$\rho(z, x_0) \leq \rho(x, x_0) \text{ for all } x \in K.$$

59. Let K be a compact subset of the metric space X. For a point $x \in X \sim K$, show that there is an open set $\mathcal{U}$ containing K and an open set $\mathcal{O}$ containing x for which $\mathcal{U} \cap \mathcal{O} = \emptyset$.
60. Let A and B be subsets of a metric space (X, ρ). Define

$$\operatorname{dist}(A, B) = \inf \{\rho(u, v) \mid u \in A, v \in B\}.$$

   If A is compact and B is closed, show that $A \cap B = \emptyset$ if and only if $\operatorname{dist}(A, B) > 0$.
61. Let K be a compact subset of a metric space X and $\mathcal{O}$ an open set containing K. Use the preceding problem to show that there is an open set $\mathcal{U}$ for which $K \subseteq \mathcal{U} \subseteq \overline{\mathcal{U}} \subseteq \mathcal{O}$.

13.6 SEPARABLE METRIC SPACES: ULAM'S REGULARITY THEOREM

Definition *A subset D of a metric space X is said to be* **dense** *in X provided that every non-empty open subset of X contains a point of D. A metric space X is said to be* **separable** *provided that there is a countable subset of X that is dense in X.*

Every Euclidean space is separable. The Weierstrass Approximation Theorem states that the polynomials are dense in $C[a, b]$. Therefore, the collection of polynomials with rational coefficients is countable and dense in $C[a, b]$, and so $C[a, b]$ is separable. In Chapter 12, we proved that for E a Lebesgue measurable subset of $\mathbf{R}^n$ and $1 \leq p < \infty$, the normed linear space $L^p(E, \mu_n)$ is separable. We also proved that $L^\infty[0, 1]$ is not separable.

Proposition 24 *A compact metric space is separable.*

Proof Let X be a compact metric space. Then X is totally bounded. For each natural number n, cover X by a finite number of balls of radius $1/n$. Let D be the collection of points that are centers of one of this countable collection of covers. Then D is countable and dense. □

Proposition 25 *A metric space X is separable if and only if there is a countable collection $\{\mathcal{O}_n\}_{n=1}^{\infty}$ of open subsets of X such that any open subset of X is the union of a subcollection of $\{\mathcal{O}_n\}_{n=1}^{\infty}$.*

Proof First suppose X is separable. Let D be a countable dense subset of X. If D is finite, then $X = D$. Assume that D is countably infinite. Let $\{x_n\}$ be an enumeration of D. Then $\{B(x_n, 1/m)\}_{n,m \in N}$ is a countable collection of open subsets of X. We claim that every open subset of X is the union of a subcollection of $\{B(x_n, 1/m)\}_{n,m \in N}$. Indeed, let $\mathcal{O}$ be an open subset of X. Let x belong to $\mathcal{O}$. We must show there are natural numbers n and m for which

$$x \in B(x_n, 1/m) \subseteq \mathcal{O}. \tag{4}$$

Since $\mathcal{O}$ is open, there is a natural number m for which $B(x, 2/m)$ is contained in $\mathcal{O}$. Since x is a point of closure of D, we may choose a natural number n for which x_n belongs to $D \cap B(x, 1/m)$. Thus, (4) holds for this choice of n and m.

To prove the converse, suppose there is a countable collection $\{\mathcal{O}_n\}_{n=1}^{\infty}$ of open sets such that any open subset of X is the union of a subcollection of $\{\mathcal{O}_n\}_{n=1}^{\infty}$. For each index n, choose a point in $\mathcal{O}_n$ and label it x_n. Then the set $\{x_n\}_{n=1}^{\infty}$ is countable and is dense since every non-empty open subset of X is the union of a subcollection of $\{\mathcal{O}_n\}_{n=1}^{\infty}$ and therefore contains points in the set $\{x_n\}_{n=1}^{\infty}$. □

Proposition 26 *Every subspace of a separable metric space is separable.*

Proof Let E be a subspace of the separable metric space X. By the preceding proposition, there is a countable collection $\{\mathcal{O}_n\}_{n=1}^{\infty}$ of open sets in X for which each open set in X is a union of some subcollection of $\{\mathcal{O}_n\}_{n=1}^{\infty}$. Thus, $\{\mathcal{O}_n \cap E\}_{n=1}^{\infty}$ is a countable collection of subsets of E, each one of which, by Proposition 2, is open. Since each open subset of E is the intersection of E with an open subset of X, every open subset of E is a union of a subcollection of $\{\mathcal{O}_n \cap E\}_{n=1}^{\infty}$. It follows from the preceding proposition that E is separable. □

Corollary 27 *The following are separable metric spaces:*

(i) Each subset of Euclidean space $\mathbf{R}^n$.
(ii) For E a Lebesgue measurable subset of $\mathbf{R}^n$ and $1 \le p < \infty$, each subset of $L^p(E, \mu_n)$.
(iii) Each subset of $C[a, b]$.

We close this section with an interesting consequence of the characterization of compactness for subsets of a complete metric space X, regarding regularity of Borel measures on such a space. Recall that the smallest σ-algebra of subsets of X that contains the open sets is denoted by $\mathcal{B}(X)$ and a measure $\mu \colon \mathcal{B}(X) \to [0, \infty)$ is called a Borel measure.

Lemma 28 *Let X be a separable, complete metric space, and $\mu\colon \mathcal{B}(X) \to [0, \infty)$ be a finite Borel measure. Then for each $\epsilon > 0$, there is a compact subset K of X for which $\mu(X \sim K) < \epsilon$.*

Proof Let $\epsilon > 0$. Let $D = \{x_n\}_{n=1}^{\infty}$ be an enumeration of a dense subset of X. Since D is dense, for each n, $\{\overline{B}(x_k, 1/n)\}_{k=1}^{\infty}$ is a countable, cover of X by closed, and hence measurable, sets. Since $\mu(X) < \infty$, by the continuity and excision properties of μ, there is an index $N(n)$ for which

$$\text{if } F_n = \bigcup_{k=1}^{N(n)} \overline{B}(x_k, 1/n), \text{ then } \mu(X \sim F_n) = \mu(X) - \mu(F_n) \le \epsilon/2^n. \tag{5}$$

Define $K = \bigcap_{n=1}^{\infty} F_n$. Then K is closed, since it is the intersection of closed sets, and therefore, being a closed subset of a complete metric space, is complete. Also, by construction, K is totally bounded. According to Theorem 15, K is compact. Moreover, by the De Morgan's Identities, and the countable monotonicity of μ,

$$\mu(X \sim K) = \mu\left(\bigcup_{k=1}^{\infty}(X \sim F_n)\right) \le \sum_{k=1}^{\infty} \mu(X \sim F_n) \le \sum_{k=1}^{\infty} \epsilon/2^n = \epsilon.$$

$\square$

For a Borel measure $\mu\colon \mathcal{M}(X) \to [0, \infty)$ it is convenient to call a set $E \in \mathcal{B}(X)$ regular provided that for every $\epsilon > 0$, there is a compact set K and open set $\mathcal{O}$.

$$K \subseteq E \subseteq \mathcal{O},\ \mu(E \sim K) < \epsilon \text{ and } \mu(\mathcal{O} \sim E) < \epsilon. \tag{6}$$

If every set in $B(X)$ is regular, then the measure $\mu\colon \mathcal{M}(X) \to [0, \infty)$ is said to be regular.

Ulam's Regularity Theorem *Let X be a separable, complete metric space. Then every finite Borel measure $\mu\colon \mathcal{B}(X) \to [0, \infty)$ is regular.*

Proof Denote the collection of regular sets by $\mathcal{R}$. We will show that $\mathcal{R}$ is a σ-algebra that contains all closed sets, and therefore, by minimality, $\mathcal{R} = \mathcal{B}(X)$. According to the preceding lemma, the whole space X is regular First, we show that the complement of a regular set is regular. Let $E \in \mathcal{R}$ and $\epsilon > 0$. There is a closed set F and open set $\mathcal{O}$ for which

$$F \subseteq E \subseteq \mathcal{O},\ \mu(E \sim K) < \epsilon \text{ and } \mu(\mathcal{O} \sim E) < \epsilon/2.$$

Define $\mathcal{O}' = X \sim F$ and $F' = X \sim \mathcal{O}$. Then $\mathcal{O}'$ is open, F' is closed, and

$$F' \subseteq X \sim E \subseteq \mathcal{O}',\ \mu((\mathbf{R}_n \sim E) \sim F') < \epsilon/2 \text{ and } \mu(\mathcal{O}' \sim (X \sim E)) < \epsilon.$$

By the preceding lemma, there is a compact set K for which $\mu(X \sim K) < \epsilon/2$. Then $F' \cap K$, being the intersection of a closed set and compact set, is compact. Since

$$(X \sim E) \sim (F \cap K) \subseteq ((X \sim E) \sim F) \cup (X \sim K),$$

$\mu((X \sim E) \sim (F \cap K)) < \epsilon$ and we already have $\mu(\mathcal{O}' \sim (X \sim E)) < \epsilon$, so that $X \sim E$ is regular. We now show that $\mathcal{R}$ is a σ-algebra. Let $E = \bigcup_{k=1}^{\infty} E_k$, the countable union of regular sets. Let $\epsilon > 0$. For each k, define $E'_k = X \sim E_k$, so that since E_k and E'_k are inner-regular, there are compact sets K_k and K'_k for which

$$K_k \subseteq E_k,\ \ \mu(E_k \sim K_k) < \epsilon/2^{k+1} \text{ and } K'_k \subseteq E'_k,\ \ \mu(E'_k \sim K'_k) < \epsilon/2^k.$$

Since $E \sim \bigcup_{k=1}^{\infty} K_k \subseteq \bigcup_{k=1}^{\infty}(E_k \sim K_k)$, $\mu(E \sim \bigcup_{k=1}^{\infty} K_k) < \epsilon/2$, and therefore, by the continuity of μ, there is an index k_0 for which,

$$\mu\left(\bigcup_{k=1}^{k_0} K_k\right) > \mu(E) - \epsilon.$$

Define $K = \bigcup_{k=1}^{k_0} K_k$. Then K, being the finite union of compact sets, is compact and K is contained in E. We have $\mu(E \sim K) = \mu(E) - \mu(K) < \epsilon$. This establishes the inner-regularity property for E. Define $K' = \bigcap_{k=1}^{\infty} K'_k$. Then K', being the intersection of compact sets, also is compact. Now, by the De Morgan's Identities,

$$(X \sim E) \sim K' = \bigcup_{k=1}^{\infty}((X \sim E) \sim K'_k).$$

By the countable monotonicity of μ, $\mu((X \sim E) \sim K') < \epsilon$. Define $\mathcal{O} = X \sim K'$. Being the complement of a closed set, $\mathcal{O}$ is open and $\mathcal{O} \sim E = (X \sim E) \sim K'$ so that $\mu(\mathcal{O} \sim E) < \epsilon$. Therefore, E is regular. So the regular sets are a σ-algebra. To conclude the proof, it suffices to show that the closed sets are regular. Let $F \subseteq X$ be closed and $\epsilon > 0$. Being a closed subset of a complete metric space, F is a complete metric space, and it inherits separability from X. Therefore, by the preceding lemma, applied with X replaced by F, there is a compact subset K of F for which $\mu(F \sim K) < \epsilon$. Observe that K is a closed subset, and therefore, a compact subset, of X. To establish outer regularity, define the distance function $\text{dist}_F \colon X \to \mathbf{R}$ by

$$\text{dist}_F(x) = \inf\{\rho(x, x') \mid x' \in F\} \text{ for } x \in X.$$

We leave it as an exercise to show, as we did in Chapter 10 in the case $X = \mathbf{R}^n$, that

$$|\text{dist}_F(x) - \text{dist}_F(y)| \leq \rho(x, y) \text{ for all } x, y \in X,$$

and therefore the distance function is continuous. For each index k, define

$$\mathcal{O}_k = \{x \in X \mid |\text{dist}_F(x) < 1/k\}.$$

Since dist_F is continuous, $\mathcal{O}_k$ is open. Since F is closed, $x \in F$ if and only if $\text{dist}_F(x) = 0$, and so $F = \bigcap_{k=1}^{\infty} \mathcal{O}_k$. Now, $\{\mathcal{O}_k\}_{k=1}^{\infty}$ is a descending, countable collection of measurable sets whose intersection is F. Since $\mu(X) < \infty$, by the continuity and excision properties of measure, there is an index k_0 for which $\mu(\mathcal{O}_{k_0} \sim F) < \epsilon$. Therefore, $F \subseteq \mathcal{R}$. Consequently, $\mathcal{B}(X) \subseteq \mathcal{R}$. $\square$

PROBLEMS

62. Let X be a metric space that contains a finite dense subset D. Show that $X = D$.
63. Show that if two continuous mappings defined on a metric space X take the same values on a dense subset, then they are equal.
64. Prove Lusin's Theorem for a finite Borel measure space $(X, \mathcal{B}(X), \mu)$ where X is a separable, complete metric space. (Suggestion: Consider the proof for Lebesgue measure on a subset of Euclidean space.)

C H A P T E R 14

Metric Spaces: Three Fundamental Theorems and Applications

Contents

14.1 The Arzelà-Ascoli Theorem 285
14.2 The Banach Contraction Principle and Picard's Theorem 289
14.3 The Baire Category Theorem 295
14.4 The Nikodym Metric Space: The Vitali-Hahn-Saks Theorem and the Dunford-Pettis Theorem 299

In this chapter, three theorems that are widely used in mathematical analysis are proven. Furthermore, an application of the Banach Contraction Principle to the existence of solutions of differential equations is provided. We employ the Baire Category Theorem to (i) prove a uniform boundedness principle for non-linear functionals, (ii) establish continuity properties of functions that are the pointwise limit of a sequence of continuous functions, and (iii) together with the Arzelà-Ascoli Lemma, to consider weak sequential compactness in $L^1(X, \mu)$.

14.1 THE ARZELÀ-ASCOLI THEOREM

In many important problems in analysis, given a sequence of continuous real-valued functions it is useful to know that there is a subsequence that converges uniformly. In this section, our main result is the Arzelà-Ascoli Theorem, which provides a criterion for a uniformly bounded sequence of continuous real-valued functions on a compact metric space X to have a uniformly convergent subsequence. We then consider this theorem as a response, for the normed linear space space $C(X)$, to the question of which closed, bounded subsets of an infinite dimensional normed linear space are compact.

For a compact metric space X, $C(X)$ denotes the collection of continuous real-valued functions on X, and the maximum metric is defined for $g, h \in C(X)$, by

$$\rho_{\max}(g, h) = \max\{|g(x) - h(x)| \mid x \in X\}.$$

According to Theorem 21 of the preceding chapter, every continuous function on X takes a maximum value, and so this metric is properly defined. This metric called the **uniform metric** because a sequence in $C(X)$ converges with respect to this metric if and only if it converges uniformly on X. The proof of the following result is left as an exercise, since it is no different than in the case $X = [a, b]$.

Proposition 1 *If X is a compact metric space, then $C(X)$ is complete.*

Definition *A collection $\mathcal{F}$ of real-valued functions on a metric space X is said to be* **equicontinuous** *at the point $x \in X$ provided that for each $\epsilon > 0$, there is a $\delta > 0$ such that for every $f \in \mathcal{F}$ and $x' \in X$,*

$$\text{if } \rho(x', x) < \delta, \text{ then } |f(x') - f(x)| < \epsilon.$$

The collection $\mathcal{F}$ is said to be equicontinuous on X provided that it is equicontinuous at every point in X.

Of course, each function in an equicontinuous collection of functions is continuous and any finite collection of continuous functions is equicontinuous. In general, an infinite collection of continuous functions may not be equicontinuous. For instance, for each n, define $f_n(x) = x^n$ for $0 \le x \le 1$. Then $\{f_n\}$ is a countable collection of continuous functions on $[0, 1]$ that is not equicontinuous at $x = 1$, but is equicontinuous at the other points in $[0, 1]$.

Example For $M \ge 0$, let $\mathcal{F}$ be the collection of real-valued functions on the closed, bounded interval $[a, b]$ that are differentiable on the open interval (a, b) and for which

$$|f'| \le M \text{ on } (a, b).$$

It follows from the Mean Value Theorem that

$$|f(u) - f(v)| \le M \cdot |u - v| \text{ for all } u, v \in [a, b].$$

Therefore, $\mathcal{F}$ is equicontinuous since, regarding the criterion for equicontinuity at each point in X, $\delta = \epsilon/M$ responds to the $\epsilon > 0$ challenge.

A sequence $\{f_n \colon X \to \mathbf{R}\}$ of functions on a set X is said to be **pointwise bounded** provided that for each $x \in X$, the sequence $\{f_n(x)\}$ is bounded and is said to be **uniformly pointwise bounded** on X provided that there is some $M \ge 0$ for which

$$|f_n| \le M \text{ on } X \text{ for all } n.$$

Lemma 2 (**the Arzelà-Ascoli Lemma**) *Let X be a separable metric space and the sequence of functions $\{f_n \colon X \to \mathbf{R}\}$ be equicontinuous and pointwise bounded. Then a subsequence of $\{f_n\}$ converges pointwise on all of X to a function $f \colon X \to \mathbf{R}$.*

Proof Let $\{x_j\}_{j=1}^{\infty}$ be an enumeration of a dense subset D of X. Then $\{f_n \colon D \to \mathbf{R}\}$ is a pointwise bounded sequence of real-valued functions on a countable set. According to Theorem 5 of Chapter 8, a subsequence converges pointwise to a real-valued function $f \colon D \to \mathbf{R}$.

For notational convenience, assume the whole sequence of $\{f_n\}$ converges pointwise on D to f. Let $x_0 \in X$. Then $\{f_n(x_0)\}$ is Cauchy. Indeed, let $\epsilon > 0$. By the equicontinuity of $\{f_n\}$ at x_0, there is a $\delta > 0$ such that $|f_n(x) - f_n(x_0)| < \epsilon/3$ for all indices n and all $x \in X$ for which $\rho(x, x_0) < \delta$. Since D is dense, there is a point $x \in D$ such that $\rho(x, x_0) < \delta$. Moreover, since $\{f_n(x)\}$ converges, it must be a Cauchy sequence, and so there is an N for which

$$|f_n(x) - f_m(x)| < \epsilon/3 \text{ for all } m, n \ge N.$$

For all $m, n \geq N$,

$$\begin{aligned}|f_n(x_0) - f_m(x_0)| \leq\ & |f_n(x_0) - f_n(x)| + |f_n(x) - f_m(x) + |f_m(x_0) - f_m(x)|| \\ & < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.\end{aligned}$$

Therefore, $\{f_n(x_0)\}$ is a Cauchy sequence of real numbers. Since $\mathbf{R}$ is complete, $\{f_n(x_0)\}$ converges. Denote the limit by $f(x_0)$. The sequence $\{f_n\}$ converges pointwise on all of X to $f\colon X \to \mathbf{R}$. $\square$

According to Proposition 22 of the preceding chapter, a continuous real-valued function on a compact metric space is uniformly continuous. The exact same proof, which depends on the existence of a Lebesgue number for a cover, shows that if X is a compact metric space and $\mathcal{F}$ is an equicontinuous collection of real-valued functions on X, then $\mathcal{F}$ is **uniformly equicontinuous,** in the sense that for each $\epsilon > 0$, there is a $\delta > 0$ such that for $u, v \in X$ and any $f \in \mathcal{F}$,

$$\text{if } \rho(u, v) < \delta, \text{ then } |f(u) - f(v)| < \epsilon.$$

The Arzelà-Ascoli Theorem *Let X be a compact metric space and the sequence of functions $\{f_n\colon X \to \mathbf{R}\}$ be uniformly pointwise bounded and equicontinuous. Then $\{f_n\}$ has a subsequence that converges uniformly on X to a continuous function $f\colon X \to \mathbf{R}$.*

Proof Since X is a compact metric space, according to Proposition 24 of the preceding chapter, it is separable. The Arzelà-Ascoli Lemma implies that a subsequence of $\{f_n\}$ converges pointwise on all of X to a real-valued function f. For notational convenience, assume the whole sequence $\{f_n\}$ converges pointwise on X to f. Therefore, in particular, for each x in X, $\{f_n(x)\}$ is a Cauchy sequence of real numbers. We use this, together with equicontinuity, to show that $\{f_n\}$ is a Cauchy sequence in $C(X)$.

Let $\epsilon > 0$. By the uniform equicontinuity of $\{f_n\}$ on X, there is a $\delta > 0$ such that for all n,

$$|f_n(u) - f_n(v)| < \epsilon/3 \text{ for all } u, v \in X \text{ such that } \rho(u, v) < \delta. \tag{1}$$

Since X is a compact metric space, according to Theorem 15 of the preceding chapter, it is totally bounded. Therefore there are a finite number of points $x_1, \ldots, x_k$ in X for which X is covered by $\{B(x_i, \delta)\}_{i=1}^k$. For $1 \leq i \leq k$, $\{f_n(x_i)\}$ is Cauchy, so there is an index N such that

$$|f_n(x_i) - f_m(x_i)| < \epsilon/3 \text{ for } 1 \leq i \leq k \text{ and all } n, m \geq N. \tag{2}$$

Now for any x in X, there is an $i, 1 \leq i \leq k$, such that $\rho(x, x_i) < \delta$, and therefore for $n, m \geq N$,

$$\begin{aligned}|f_n(x) - f_m(x)| \leq\ & |f_n(x) - f_n(x_i)| + |f_n(x_i) - f_m(x_i)| + |f_m(x_i) - f_m(x)| \\ & < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.\end{aligned}$$

Therefore, $\{f_n\}$ is uniformly Cauchy. Consequently, since $C(X)$ is complete, $\{f_n\}$ converges uniformly on X to a continuous function. $\square$

Theorem 3 *Let X be a compact metric space and $\mathcal{F}$ be a closed, bounded subset of $C(X)$. Then $\mathcal{F}$ is a compact subspace of $C(X)$ if and only if it is equicontinuous.*

Proof First suppose that $\mathcal{F}$ is closed, bounded, and equicontinuous. Let $\{f_n\}$ be a sequence in $\mathcal{F}$. According to the Arzelà-Ascoli Theorem, a subsequence of $\{f_n\}$ converges uniformly to a function in $f \in C(X)$. Since $\mathcal{F}$ is closed, f belongs to $\mathcal{F}$. It follows that $\mathcal{F}$ is a sequentially compact metric space and therefore is compact.

Now assume $\mathcal{F}$ is compact. According to Proposition 23 of the preceding chapter, $\mathcal{F}$ is a closed, bounded subset of $C(X)$. We argue by contradiction to show that $\mathcal{F}$ is equicontinuous. Otherwise, $\mathcal{F}$ is not equicontinuous at a point x in X. Then there is an $\epsilon_0 > 0$ such that for each n, there is a function in $\mathcal{F}$ that we label f_n and a point in X that we label x_n for which

$$|f_n(x_n) - f_n(x)| \geq \epsilon_0 \text{ while } \rho(x_n, x) < 1/n. \tag{3}$$

Since $\mathcal{F}$ is a compact metric space, it is sequentially compact. Therefore, there is a subsequence $\{f_{n_k}\}$ that converges uniformly on X to a continuous function f. Choose an index K such that $\rho_{\max}(f, f_{n_k}) < \epsilon_0/3$ for $k \geq K$. It follows from (3) that for $k \geq K$,

$$|f(x_{n_k}) - f(x)| > \epsilon_0/3 \text{ while } \rho(x_{n_k}, x) < 1/n_k. \tag{4}$$

This contradicts the continuity of f at the point x. Therefore, $\mathcal{F}$ is equicontinuous. $\square$

The forthcoming Riesz's Theorem of Chapter 17 asserts that the closed unit ball of a normed linear space is compact if and only if the linear space is finite dimensional. Therefore, given a particular infinite dimensional normed linear space, it is interesting to characterize the closed, bounded subsets that are compact. The compactness criterion provided by the Arzelà-Ascoli Theorem for subsets of $C(X)$ has a ℓ^p counterpart. It is not difficult to show, by using, say, the Vitali Convergence Theorem that for $1 \leq p < \infty$, a closed, bounded subset $\mathcal{S}$ of ℓ^p is compact if and only if it is **equisummable**, in the sense that for each $\epsilon > 0$, there is an index N for which

$$\sum_{k=N}^{\infty} |x_k|^p < \epsilon \text{ for all } x = \{x_n\} \in \mathcal{S}.$$

PROBLEMS

1. Let E be a compact subspace of a metric space Y. Show that E is a closed, bounded subset of Y.
2. Show that an equicontinuous sequence of real-valued functions on a compact metric space is pointwise bounded if and only if it is uniformly bounded.
3. Show that an equicontinuous collection of continuous functions on a compact metric space is uniformly equicontinuous.
4. Let X be a metric space and $\{f_n\}$ a sequence in $C(X)$ that converges uniformly on X to $f \in C(X)$. Show that $\{f_n\}$ is equicontinuous.
5. A real-valued function f on $[0, 1]$ is said to be Hölder continuous of order α provided that there is a constant C for which

   $$|f(x) - f(y)| \leq C|x - y|^{\alpha} \text{ for all } x, y \in [0, 1].$$

   Define the Hölder norm

   $$\|f\|_{\alpha} = \max\left\{|f(x)| + |f(x) - f(y)|/|x - y|^{\alpha} \mid x, y \in [0, 1], x \neq y\right\}.$$

Show that for $0 < \alpha \leq 1$, the set of functions for which $\|f\|_\alpha \leq 1$ has compact closure as a subset of $C[0, 1]$.

6. Let X be a compact metric space and $\mathcal{F}$ a subset of $C(X)$. Show that $\mathcal{F}$ is equicontinuous if and only if its closure in $C(X)$, $\overline{\mathcal{F}}$, is equicontinuous. Conclude that a subset of $C(X)$ has compact closure if and only if it is equicontinuous and uniformly bounded.
7. For a closed, bounded interval $[a, b]$, let $\{f_n\}$ be a sequence in $C[a, b]$. If $\{f_n\}$ is equicontinuous, does $\{f_n\}$ necessarily have a uniformly convergent subsequence? If $\{f_n\}$ is uniformly bounded, does $\{f_n\}$ necessarily have a uniformly convergent subsequence?
8. Let X be a compact metric space and Y be a general metric space. Denote by $C(X, Y)$ the set of continuous mappings from X to Y. State and prove a version of the Arzelà-Ascoli Theorem for a sequence in $C(X, Y)$ in which the assumption that $\{f_n\}$ is pointwise bounded is replaced by the assumption that for each $x \in X$, the closure of the set $\{f_n(x) \mid n \text{ a natural number}\}$ is a compact subspace of Y.
9. Let $\{f_n\}$ be an equicontinuous, uniformly bounded sequence of continuous real-valued functions on $\mathbf{R}$. Show that there is a subsequence of $\{f_n\}$ that converges pointwise on $\mathbf{R}$ to a continuous function on $\mathbf{R}$ and that the convergence is uniform on each bounded subset of $\mathbf{R}$.
10. For $1 \leq p < \infty$, show that a subspace of ℓ^p is compact if and only if it is closed, bounded, and equisummable.
11. For a sequence of non-negative real numbers $\{c_n\}$, let $\mathcal{S}$ be the subset of ℓ^2 consisting of those $x = \{x_n\} \in \ell^2$ such that $|x_n| \leq c_n$ for all n. Show that $\mathcal{S}$ is equisummable if $\{c_n\}$ belongs to ℓ^2.
12. For $1 \leq p \leq \infty$, show that the closed unit ball in ℓ^p is not compact.
13. For $1 \leq p \leq \infty$, show that the closed unit ball in the space $L^p[0, 1]$ is not compact.
14. Let S be a countable set and $\{f_n\}$ a sequence of real-valued functions on S that is pointwise bounded on S. Show that there is a subsequence of $\{f_n\}$ that converges pointwise on S to a real-valued function.

14.2 THE BANACH CONTRACTION PRINCIPLE AND PICARD'S THEOREM

Definition *A point x in X is called a* **fixed-point** *of the mapping T: $X \to X$ provided that $T(x) = x$.*

We are interested here in finding assumptions on a mapping which ensure that it has a fixed-point. A fixed-point of a real-valued function of a real variable corresponds to a point in the plane at which the graph of the function intersects the diagonal line $y = x$. This observation provides the geometric insight for the most elementary result regarding the existence of fixed-points: Let $[a, b]$ be a closed, bounded interval in $\mathbf{R}$ and suppose that the image of the continuous function f: $[a, b] \to \mathbf{R}$ is contained in $[a, b]$. Then f: $[a, b] \to \mathbf{R}$ has a fixed-point. This follows from the Intermediate Value Theorem by observing that if we define $g(x) = f(x) - x$ for x in $[a, b]$, then $g(a) \geq 0$ and $g(b) \leq 0$, so that $g(x_0) = 0$ for some x_0 in $[a, b]$, which means that $f(x_0) = x_0$. This generalizes to mappings on subsets of Euclidean spaces as follows: If K is a compact, convex subset of $\mathbf{R}^n$ and the mapping T: $K \to K$ is continuous, then T has a fixed-point. This is called Brouwer's Fixed-Point Theorem. An elementary fixed-point result called the Banach Contraction Principle is now proven, in which there is a much more restrictive assumption on the mapping but a very general assumption on the underlying space.

Definition *A mapping T from a metric space (X, ρ) into itself is said to be* **Lipschitz** *provided that there is a number $c \geq 0$, called a Lipschitz constant for the mapping, for which*

$$\rho(T(u), T(v)) \leq c\,\rho(u, v) \text{ for all } u, v \in X.$$

If $c < 1$, the Lipschitz mapping is called a **contraction**.

The Banach Contraction Principle *Let X be a complete metric space and the mapping $T\colon X \to X$ be a contraction. Then $T\colon X \to X$ has exactly one fixed-point.*

Proof Let c be a number with $0 \leq c < 1$ that is a Lipschitz constant for the mapping T. Select a point in X and label it x_0. Now define the sequence $\{x_k\}$ inductively by defining $x_1 = T(x_0)$ and, if k is a natural number such that x_k is defined, defining $x_{k+1} = T(x_k)$. The sequence $\{x_n\}$ is properly defined since $T(X)$ is a subset of X. We will show that this sequence converges to a fixed-point of T. Indeed, observe that for each k,

$$\rho(x_{k+1}, x_k) = \rho(T(x_k), T(x_{k-1})) \leq c\,\rho(x_k, x_{k-1}),$$

and so, by an induction argument,

$$\rho(x_{k+1}, x_k) \leq c^k \rho(T(x_0), x_0).$$

Therefore by the triangle inequality and, by the geometric sum formula[1], for $m > k$

$$\begin{aligned}
\rho(x_m, x_k) &\leq \rho(x_m, x_{m-1}) + \rho(x_{m-1}, x_{m-2}) + \cdots + \rho(x_{k+1}, x_k) \\
&\leq [c^{m-1} + c^{m-2} + \cdots + c^k]\rho(T(x_0), x_0) \\
&= c^k[1 + c + \cdots + c^{m-1-k}]\rho(T(x_0), x_0) \\
&= c^k \cdot \frac{1 - c^{m-k}}{1 - c} \cdot \rho(T(x_0), x_0).
\end{aligned}$$

Consequently, since $0 \leq c < 1$,

$$\rho(x_m, x_k) \leq \frac{c^k}{1 - c} \cdot \rho(T(x_0), x_0) \text{ if } m > k.$$

But $\lim_{k \to \infty} c^k = 0$, and so, from the preceding inequality it follows that $\{x_k\}$ is a Cauchy sequence.

---

[1]

$$\sum_{k=0}^{n} c^k = \frac{1 - c^{n+1}}{1 - c} \text{ if } c \neq 1.$$

By assumption, the metric space X is complete and so there is a point x in X to which the sequence $\{x_k\}$ converges. Since T is Lipschitz, it is continuous. It follows that

$$T(x) = \lim_{k \to \infty} T(x_k) = \lim_{k \to \infty} x_{k+1} = x.$$

Therefore, the mapping $T: X \to X$ has at least one fixed-point. It remains to check that there is only one fixed-point. But if u and v are points in X such that $T(u) = u$ and $T(v) = v$, then

$$0 \leq \rho(u, v) = \rho(T(u), T(v)) \leq c\,\rho(u, v),$$

so that since $0 \leq c < 1$, $\rho(u, v) = 0$, that is, $u = v$. □

The above proof of the Banach Contraction Principle actually proves substantially more than the *existence* of a unique fixed-point. *It provides an algorithm for approximating the fixed-point.* Indeed, under the assumptions of the Banach Contraction Principle, what has been proven is that if c is a number with $0 \leq c < 1$ that is a Lipschitz constant for the mapping $T: X \to X$, and x_0 is any point in X, then (i) the sequence $\{x_k\}$ defined recursively by setting $x_1 = T(x_0)$ and $x_{k+1} = T(x_k)$ for $k \geq 1$ converges to a fixed-point x_* of T and (ii)

$$\rho(x_*, x_k) \leq \frac{c^k}{1-c} \cdot \rho(T(x_0), x_0) \text{ for every } k.$$

The Banach Contraction Principle is widely used in the study of non-linear differential equations. We provide just one application. Suppose $\mathcal{O}$ is an open subset of the plane $\mathbf{R}^2$ that contains the point (x_0, y_0). Given a function $g: \mathcal{O} \to \mathbf{R}$, the problem we pose is to find an open interval of real numbers I containing the point x_0 and a differentiable function $f: I \to \mathbf{R}$ such that

$$\begin{aligned} f'(x) &= g(x, f(x)) \text{ for all } x \in I \\ f(x_0) &= y_0. \end{aligned} \tag{5}$$

A very special case of the above equation occurs if g is independent of its second variable, so $g(x, y) = h(x)$. Even in this case, if the image of the function $h: I \to \mathbf{R}$ fails to be an interval, there is no solution of equation (5) (see Problems 28 and 29). On the other hand, if h is continuous, then it follows from the Fundamental Theorem of Calculus for the Riemann integral that equation (5) has a unique solution given by

$$f(x) = y_0 + \int_{x_0}^{x} h(t)\,dt \text{ for all } x \in I.$$

Therefore, for a general continuous real-valued function of two variables g, if a continuous function $f: I \to \mathbf{R}$ has the property that $(x, f(x)) \in \mathcal{O}$ for each $x \in I$, then f is a solution of (5) if and only if

$$f(x) = y_0 + \int_{x_0}^{x} g(t, f(t))\,dt \text{ for all } x \in I. \tag{6}$$

As we will see in the proof of the next theorem, this equivalence between solutions of the differential equation (5) and those of the *integral equation* (6) facilitates the use of fixed-point theorems in the study of differential equations.

The Picard Local Existence Theorem *Let $\mathcal{O}$ be an open subset of the plane $\mathbf{R}^2$ containing the point (x_0, y_0). Assume that the function $g\colon \mathcal{O} \to \mathbf{R}$ is continuous and there is a positive number M for which the following Lipschitz property in the second variable holds, uniformly with respect to the first variable:*

$$|g(x,\, y_1) - g(x,\, y_2)| \leq M|y_1 - y_2| \text{ for all points } (x,\, y_1) \text{ and } (x,\, y_2) \text{ in } \mathcal{O}. \tag{7}$$

Then there is an open interval I containing x_0 on which the differential equation (5) has a unique solution.

Proof For ℓ a positive number, define I_ℓ to be the closed interval $[x_0 - \ell, x_0 + \ell]$. In view of the equivalence noted above between solutions of (5) and (6), it suffices to show that ℓ can be chosen so that there is exactly one continuous function $f\colon I_\ell \to \mathbf{R}$ having the property that

$$f(x) = y_0 + \int_{x_0}^{x} g(t, f(t))\, dt \text{ for all } x \in I_\ell.$$

Since $\mathcal{O}$ is open, there are $a, b > 0$ for which the closed rectangle $R = [x_0 - a, x_0 + a] \times [y_0 - b, y_0 + b]$ is contained in $\mathcal{O}$. Now for each positive number ℓ with $\ell \leq a$, define X_ℓ to be the subspace of the metric space $C(I_\ell)$ consisting of those continuous functions $f\colon I_\ell \to \mathbf{R}$ that have the property that

$$|f(x) - y_0| \leq b \text{ for all } x \in I_\ell;$$

that is, the continuous functions on I_ℓ that have a graph contained in the rectangle $I_\ell \times [y_0 - b, y_0 + b]$.

For $f \in X_\ell$, define the function $T(f) \in C(I_\ell)$ by

$$T(f)(x) = y_0 + \int_{x_0}^{x} g(t, f(t))dt \text{ for all } x \in I_\ell.$$

A solution of the integral equation (6) is simply a fixed-point of the mapping $T\colon X_\ell \to C(I_\ell)$. The strategy of the proof is as follows: Since $C(I_\ell)$ is a complete metric space and X_ℓ is a closed subset of $C(I_\ell)$, X_ℓ also is a complete metric space. We will show that if ℓ is chosen sufficiently small, then

$$T(X_\ell) \subseteq X_\ell \text{ and } T\colon X_\ell \to X_\ell \text{ is a contraction.}$$

It follows from the Banach Contraction Principle that $T\colon X_\ell \to X_\ell$ has a unique fixed-point.

In order to choose ℓ so that $T(X_\ell) \subseteq X_\ell$ we first use the compactness of the closed, bounded rectangle R together with the continuity of g to choose a positive number K such that

$$|g(x, y)| \leq K \text{ for all points } (x, y) \text{ in } R.$$

Now for $f \in X_\ell$ and $x \in I_\ell$,

$$|T(f)(x) - y_0| = \left| \int_{x_0}^{x} g(t, f(t))\, dt \right| \leq \ell K,$$

so that

$$T(X_\ell) \subseteq X_\ell \text{ provided that } \ell K \leq b.$$

Observe that for functions $f_1, f_2 \in X_\ell$, and $x \in I_\ell$, it follows from (7) that

$$|g(x, f_1(x)) - g(x, f_2(x))| \leq M\rho_{\max}(f_1, f_2).$$

Consequently, using the linearity and monotonicity properties of the integral,

$$\begin{aligned} |T(f_1)(x) - T(f_2)(x)| &= \left| \int_{x_0}^{x} [g(t, f_1(t)) - g(t, f_2(t))] dt \right| \\ &\leq |x - x_0| M\rho_{\max}(f_1, f_2) \\ &\leq \ell M \rho_{\max}(f_1, f_2). \end{aligned}$$

This inequality, together with the inclusion $T(X_\ell) \subseteq X_\ell$ provided that $\ell K \leq b$, implies that

$$T\colon X_\ell \to X_\ell \text{ is a contraction provided that } \ell K \leq b \text{ and } \ell M < 1.$$

Define $\ell = \min\{b/K, 1/2M\}$. The Banach Contraction Principle implies that the mapping $T\colon X_\ell \to X_\ell$ has a unique fixed-point. □

PROBLEMS

15. Let p be a polynomial. Show that $p\colon \mathbf{R} \to \mathbf{R}$ is Lipschitz if and only if the degree of p is less than 2.
16. Fix $\alpha > 0$, define $f(x) = \alpha x(1 - x)$ for x in $[0, 1]$.
    (i) For what values of α is $f([0, 1]) \subseteq [0, 1]$?
    (ii) For what values of α is $f([0, 1]) \subseteq [0, 1]$ and $f\colon [0, 1] \to [0, 1]$ a contraction?
17. Does a mapping of a metric space X into itself that is Lipschitz with Lipschitz constant less than 1 necessarily have a fixed-point?
18. Does a mapping of a complete metric space into itself that is Lipschitz with Lipschitz constant 1 necessarily have a fixed-point?
19. Let X be a compact metric space and T a mapping from X into itself such that

    $$\rho(T(u), T(v)) < \rho(u, v) \text{ for all } u, v \in X, u \neq v.$$

    Show that T has a unique fixed-point.
20. Define $f(x) = \pi/2 + x - \arctan x$ for all real numbers x. Show that

    $$|f(u) - f(v)| < |u - v| \text{ for all } u, v \in \mathbf{R}, u \neq v.$$

    Show that f does not have a fixed-point. Does this contradict the preceding problem?
21. In Euclidean space $\mathbf{R}^n$ consider the closed unit ball $B = \{x \in \mathbf{R}^n \mid \|x\| \leq 1\}$. Let f map B into B and be Lipschitz with Lipschitz constant 1. Without using the Brouwer Fixed-Point Theorem, show that f has a fixed-point.

22. Suppose that the mapping $f\colon \mathbf{R}^n \to \mathbf{R}^n$ is a contraction. Define $g(x) = x - f(x)$ for all x in $\mathbf{R}^n$. Show that the mapping $g\colon \mathbf{R}^n \to \mathbf{R}^n$ is both one-to-one and onto. Also show that g and its inverse are continuous.

23. Let X be a complete metric space containing the point x_0 and let r be a positive real number. Define $K = \{x \text{ in } X \mid \rho(x, x_0) \le r\}$. Suppose that the mapping $T\colon K \to X$ is Lipschitz with Lipschitz constant c. Suppose also that $cr + \rho(T(x_0), x_0) \le r$. Prove that $T(K) \subseteq K$ and that $T\colon K \to X$ has a fixed-point.

24. Show that if the function $g\colon \mathbf{R}^2 \to \mathbf{R}$ has continuous first-order partial derivatives, then for each point (x_0, y_0) in $\mathbf{R}^2$ there is a neighborhood $\mathcal{O}$ of (x_0, y_0) on which the Lipschitz assumption (7) holds.

25. In case the function $g\colon \mathcal{O} \to \mathbf{R}$ has the form $g(x, y) = h(x) + by$, where the function $h\colon \mathbf{R} \to \mathbf{R}$ is continuous, prove that the following is an explicit formula for the solution of (5):

$$f(x) = e^{b(x-x_0)}y_0 + \int_{x_0}^{x} e^{b(x-t)}h(t)\,dt \text{ for all } x \text{ in } I.$$

26. Consider the differential equation

$$\begin{aligned} f'(x) &= 3[f(x)]^{2/3} \text{ for all } x \in \mathbf{R} \\ f(0) &= 0. \end{aligned}$$

Show that the function $f\colon \mathbf{R} \to \mathbf{R}$ that is identically 0 is a solution and the function $f\colon \mathbf{R} \to \mathbf{R}$ defined by $f(x) = 0$, if $x < 0$ and $f(x) = x^3$, if $x \ge 0$, is also a solution. Does this contradict the Picard Existence Theorem?

27. For a positive number ϵ, consider the differential equation

$$\begin{aligned} f'(x) &= (1/\epsilon)[1 + (f(x))^2] \text{ for all } x \in \mathbf{R} \\ f(0) &= 0. \end{aligned}$$

Show that on the interval $I = (-\epsilon(\pi/2), \epsilon(\pi/2))$ there is a unique solution of this differential equation that is defined by $f(x) = \tan(x/\epsilon)$ and there is no solution in an interval strictly containing I.

28. Let I be an open interval in $\mathbf{R}$ and suppose that the function $h\colon I \to \mathbf{R}$ has the property that there are points $x_1 < x_2$ in I and a number c such that $h(x_1) < c < h(x_2)$ but c does not belong to $h(I)$. Prove that there is no solution to the differential equation (5) by arguing that if $f\colon I \to \mathbf{R}$ is a solution, then the continuous function $f(x) - cx$ fails to attain a minimum value on the interval $[x_1, x_2]$.

29. (Darboux) Use the preceding exercise to prove that if I is an open interval in $\mathbf{R}$ and the function $f\colon I \to \mathbf{R}$ is differentiable, then the image of the derivative $f'\colon I \to \mathbf{R}$ is an interval.

30. State and prove a form of the Picard Existence Theorem for systems of differential equations in the following context: $\mathcal{O}$ is an open subset of $\mathbf{R} \times \mathbf{R}^n$, $\mathbf{g}\colon \mathcal{O} \to \mathbf{R}^n$ is continuous, the point $(x_0, \mathbf{y}_0)$ is in $\mathcal{O}$, and the system of differential equations is

$$\begin{aligned} \mathbf{f}'(x) &= \mathbf{g}(x, \mathbf{f}(x)) \text{ for all } x \in I \\ \mathbf{f}(x_0) &= \mathbf{y}_0. \end{aligned}$$

(Suggestion: Approximate $\mathbf{g}$ by a Lipschitz mapping and then use the Arzelà-Ascoli Theorem.)

14.3 THE BAIRE CATEGORY THEOREM

Let E be a subset of a metric space X. A point $x \in E$ is called an **interior point** of E provided that there is an open ball centered at x that is contained in E: the collection of interior points of E is called the **interior** of E and denoted by $\operatorname{int} E$. A point $x \in X \sim E$ is called an **exterior point** of E provided that there is an open ball centered at x that is contained in $X \sim E$: the collection of exterior points of E is called the **exterior** of E and denoted by $\operatorname{ext} E$. If a point $x \in X$ has the property that every ball centered at x contains points in E and points in $X \sim E$, it is called a **boundary point** of E: the collection of boundary points of E is called the **boundary** of E and denoted by $\operatorname{bd} E$. It is clear that, for any subset E of X,

$$X = \operatorname{int} E \cup \operatorname{ext} E \cup \operatorname{bd} E \text{ and the union is disjoint.} \tag{8}$$

A subset of a metric space is said to be **hollow** provided that it has empty interior[2]. Observe that for a subset E of a metric space X,

$$E \text{ is hollow in } X \text{ if and only if its complement, } X \sim E, \text{ is dense in } X. \tag{9}$$

For a metric space X, a point $x \in X$ and $0 < r_1 < r_2$, by the continuity of the metric, there is the inclusion $\overline{B(x,\, r_1)} \subseteq B(x,\, r_2)$. Therefore, $\overline{B(x,\, r_1)}$ is a closed set for which

$$B(x,\, r_1) \subseteq \overline{B(x,\, r_1)} \subseteq B(x,\, r_2).$$

Consequently, if $\mathcal{O}$ is an open subset of a metric space X, for each point $x \in \mathcal{O}$, there is an open ball centered at x whose closure is contained in $\mathcal{O}$.

The Baire Category Theorem *Let X be a complete metric space.*

(i) If $\{\mathcal{O}_n\}_{n=1}^{\infty}$ is a countable collection of open, dense subsets of X, then the intersection $\bigcap_{n=1}^{\infty} \mathcal{O}_n$ also is dense.

(ii) If $\{F_n\}_{n=1}^{\infty}$ is a countable collection of closed, hollow subsets of X, then the union $\bigcup_{n=1}^{\infty} F_n$ also is hollow.

Proof A set is dense if and only if its complement is hollow; a set is open if and only if its complement is closed. Consequently, by De Morgan's Identities, (i) and (ii) are equivalent. We establish (i). Let $x_0 \in X$ and $r_0 > 0$. It must be shown that $B(x_0,\, r_0)$ contains a point of $\bigcap_{n=1}^{\infty} \mathcal{O}_n$. The strategy of the proof is to inductively define a contracting sequence of closed sets $\{\overline{B}_n\}$ for which, for all n,

$$\overline{B}_n \subseteq \mathcal{O}_n \text{ and } \overline{B}_1 \subseteq B(x_0,\, r_0). \tag{10}$$

Assume such a sequence has been defined. The metric space X is complete. According to the Cantor Intersection Theorem, the intersection $\bigcap_{n=1}^{\infty} \overline{B}_n$ is non-empty. Let $x_* \in \bigcap_{n=1}^{\infty} \overline{B}_n$. Observe that $x_* \in \overline{B}_1 \subseteq B(x_0, r_0)$ also. This completes the proof of (i), provided that we choose such a contracting sequence of closed sets.

The set $B(x_0,\, r_0) \cap \mathcal{O}_1$ is open, and is non-empty, since $\mathcal{O}_1$ is open and dense. Let $x_1 \in B(x_0,\, r_0) \cap \mathcal{O}_1$. Choose $r_1, 0 < r_1 < 1$, for which, if $B_1 = B(x_1,\, r_1)$, then

$$\overline{B}_1 \subseteq B(x_0,\, r_0) \cap \mathcal{O}_1.$$

[2]The adjective "hollow" was suggested by Adam Ross.

Suppose n is a natural number and the descending collection of open balls $\{B_k\}_{k=1}^n$ has been chosen with the property that for $1 \leq k \leq n$, B_k has radius less than $1/k$ and $\overline{B}_k \subseteq \mathcal{O}_k$. The set $B_n \cap \mathcal{O}_{n+1}$ is non-empty, since $\mathcal{O}_{n+1}$ is open and dense. Let $x_{n+1} \in B(x_n, r_n) \cap \mathcal{O}_{n+1}$. Choose $r_{n+1}, 0 < r_{n+1} < 1/(n+1)$, for which, if we define $B_{n+1} = B(x_{n+1}, r_{n+1})$, $\overline{B}_{n+1} \subseteq B_n \cap \mathcal{O}_{n+1}$. This inductively defines a contracting sequence of closed sets for which (10) holds. □

Corollary 4 *Let X be a complete metric space and $\{F_n\}_{n=1}^\infty$ a countable collection of closed subsets of X. If $\bigcup_{n=1}^\infty F_n$ has non-empty interior, then at least one of the F_n's has non-empty interior. In particular, if $X = \bigcup_{n=1}^\infty F_n$, then at least one of the F_n's has non-empty interior.*

Corollary 5 *Let X be a complete metric space and $\{F_n\}_{n=1}^\infty$ a countable collection of closed subsets of X. Then $\bigcup_{n=1}^\infty \operatorname{bd} F_n$ is hollow.*

Proof It is left as an exercise to show that for any closed subset E of X, the boundary of $E, \operatorname{bd} E$, is hollow. The boundary of any subset of X is closed. Therefore, for each n, $\operatorname{bd} F_n$ is closed and hollow. According to the Baire Category Theorem, $\bigcup_{n=1}^\infty \operatorname{bd} F_n$ is hollow. □

Theorem 6 (A Non-linear Uniform Boundedness Theorem) *Let $\mathcal{F}$ be a collection of continuous real-valued functions on a complete metric space X that is pointwise bounded. Then there is a ball in X on which $\mathcal{F}$ is uniformly pointwise bounded.*

Proof For each index n, define $E_n = \{x \in X \mid |f(x)| \leq n \text{ for all } f \in \mathcal{F}\}$. Then E_n is closed, since each function in $\mathcal{F}$ is continuous and the intersection of a collection of closed sets is closed. Since $\mathcal{F}$ is pointwise bounded, for each $x \in X$, there is an index n such that $|f(x)| \leq n$ for all $f \in \mathcal{F}$; that is, $x \in E_n$. Therefore, $X = \bigcup\limits_{n=1}^{\infty} E_n$. Since X is a complete metric space, it follows from Corollary 4 that there is an n for which E_n contains an open ball $B(x, r)$. □

In Chapter 8, using a very different proof, we established a special case of the above theorem for continuous functions that are linear. We proved that if X is a complete normed linear space and $\{\Phi_n\}$ is a sequence of continuous, linear functionals in its dual space X^* that is pointwise bounded, then it is bounded with respect to the norm in X^*; which, because of linearity, is equivalent to the assertion that it is uniformly pointwise bounded on some ball.

If a sequence of continuous real-valued functions converges uniformly, then the limit function is continuous, but this may not hold for pointwise convergence. Nonetheless, under pointwise convergence to a real-valued function of a sequence of continuous real-valued functions on a complete metric space, the limit function is continuous at each point in a dense subset of its domain.

Theorem 7 *Let X be a complete metric space and $\{f_n\colon X \to \mathbf{R}\}$ be a sequence of continuous functions that converges pointwise to $f\colon X \to \mathbf{R}$. Then there is a dense subset D of X for which $\{f_n\colon X \to \mathbf{R}\}$ is equicontinuous and $f\colon X \to \mathbf{R}$ is continuous at each $x \in D$.*

Proof For natural numbers m and n, define

$$E(m, n) = \{x \in X \mid \ |f_j(x) - f_k(x)| \leq 1/m \text{ for all } j, k \geq n\}.$$

Since each function $f_j - f_k$ is continuous, the set $E(m, n)$, being the intersection of a collection of closed sets, is closed. According to Corollary 5,

$$D = X \sim \Big[\bigcup_{n,m \in \mathbf{N}} \operatorname{bd} E_{m,n} \Big]$$

is dense in X. Observe that if $x \in D \cap E(m, n)$, then x belongs to the interior of $E(m, n)$. We claim that $\{f_n\}$ is equicontinuous at each point of D. Indeed, let $x_0 \in D$. Let $\epsilon > 0$. Choose an m for which $1/m < \epsilon/4$. Since $\{f_n(x_0)\}$ converges to a real number, $\{f_n(x_0)\}$ is Cauchy. Choose an N for which

$$|f_j(x_0) - f_k(x_0)| \leq 1/m \text{ for all } j, k \geq N. \tag{11}$$

Therefore, x_0 belongs to $E(m, N)$. As observed above, x_0 belongs to the interior of $E(m, N)$. Choose $r > 0$ such that $B(x_0, r) \subseteq E(m, N)$, that is,

$$|f_j(x) - f_k(x)| \leq 1/m \text{ for all } j, k \geq N \text{ and all } x \in B(x_0, r). \tag{12}$$

The function f_N is continuous at x_0. Therefore, there is a $\delta, 0 < \delta < r$, for which

$$|f_N(x) - f_N(x_0)| < 1/m \text{ for all } x \in B(x_0, \delta). \tag{13}$$

Observe that for every point $x \in X$ and natural number j,

$$f_j(x) - f_j(x_0) = [f_j(x) - f_N(x)] + [f_N(x) - f_N(x_0)] + [f_N(x_0) - f_j(x_0)].$$

It follows from (11), (12), (13), and the triangle inequality that

$$|f_j(x) - f_j(x_0)| \leq 3/m < [3/4]\epsilon \text{ for all } j \geq N \text{ and all } x \in B(x_0, \delta). \tag{14}$$

The finite collection of continuous functions $\{f_j\}_{j=1}^{N-1}$ is clearly equicontinuous at x_0. It follows from (14) that $\{f_n\}$ is equicontinuous at x_0. But this implies the continuity of f at x_0. Indeed, take the limit as $j \to \infty$ in (14) to obtain

$$|f(x) - f(x_0)| < \epsilon \text{ for all } x \in B(x_0, \delta). \qquad \square$$

There is standard, but not particularly suggestive, terminology associated with the ideas of this section. A subset E of a metric space X is called **nowhere dense** provided that its closure $\overline{E}$ is hollow. The Baire Category Theorem has the following equivalent formulation: In a complete metric space, the union of a countable collection of nowhere dense sets is hollow. A subset E of a metric space X is said to be of the **first category** (or *meager*) if E is the union of a countable collection of nowhere dense subsets of X. A set that is not of the first category is said to be of the **second category** (or *non-meager*), and the complement of a set of first category is called **residual** (or *co-meager*). The Baire Category Theorem may also be rephrased as follows: a non-empty open subset of a complete metric space is of the second category. This is the reason why the adjective **category** appears in the title of the theorem.

The consequences of the Baire Category Theorem are surprisingly varied. In Chapter 17, Theorem 6 is employed to prove the Open Mapping Theorem and the Uniform Boundedness Principle for continuous linear operators, two cornerstones for the study of linear functionals and operators. In the next section, Theorem 7 is used to prove the Vitali-Hahn-Saks Theorem regarding the convergence of measures. In Problems 36 and 37, two interesting properties of continuous and differentiable functions are deduced from the Baire Category Theorem.

PROBLEMS

31. Let E be a subset of a metric space X. Show that $\operatorname{bd} E$ is closed. Also show that if E is closed, then the interior of $\operatorname{bd} E$ is empty.

32. In a metric space X, show that a subset E is nowhere dense if and only if for each open subset $\mathcal{O}$ of X, $E \cap \mathcal{O}$ is not dense in $\mathcal{O}$.

33. In a complete metric space X, is the union of a countable collection of nowhere dense sets also nowhere dense?

34. Let $\mathcal{O}$ be an open subset and F be a closed subset of a metric space X. Show that both $\overline{\mathcal{O}} \sim \mathcal{O}$ and $F \sim \operatorname{int} F$ are closed and hollow.

35. In a complete metric space, is the union of a countable collection of sets of the first category also of the first category?

36. Let F_n be the subset of $C[0, 1]$ consisting of functions for which there is a point x_0 in $[0, 1]$ such that $|f(x) - f(x_0)| \leq n|x - x_0|$ for all $x \in [0, 1]$. Show that F_n is closed. Show that F_n is hollow by observing that for $f \in C[0, 1]$ and $r > 0$, there is a piecewise linear function $g \in C[0, 1]$ for which $\rho_{\max}(f, g) < r$ and the left-hand and right-hand derivatives of g on $[0, 1]$ are greater than $n + 1$. Conclude that $C[0, 1] \neq \bigcup_{n=1}^{\infty} F_n$ and show that each $h \in C[0, 1] \sim \bigcup_{n=1}^{\infty} F_n$ fails to be differentiable at any point in $(0, 1)$.

37. Let f be a real-valued function on a metric space X. Show that the set of points at which f is continuous is the intersection of a countable collection of open sets. Conclude that there is not a real-valued function on $\mathbf{R}$ that is continuous just at the rational numbers.

38. For each n, show that in $[0, 1]$ there is a nowhere dense closed set that has Lebesgue measure $1 - 1/n$. Use this to construct a set of the first category in $[0, 1]$ that has measure 1.

39. A point x in a metric space X is called isolated provided that the singleton set $\{x\}$ is open in X.
    (i) Prove that a complete metric space without isolated points has an uncountable number of points.
    (ii) Use part (i) to prove that $[0, 1]$ is uncountable. Compare this with the proof that $[0, 1]$ is uncountable because it has positive Lebesgue measure.
    (iii) Show that if X is a complete metric space without isolated points and $\{F_n\}_{n=1}^{\infty}$ is a countable collection of closed hollow sets, then $X \sim \bigcup_{n=1}^{\infty} F_n$ is dense and uncountable.

40. Let E be a subset of a complete metric space X. Verify the following assertions.
    (i) If $X \sim E$ is dense and F is a closed set contained in E, then F is nowhere dense.
    (ii) If E and $X \sim E$ are both dense, then at most one of them is the union of a countable collection of closed sets.
    (iii) The set of rational numbers in $[0, 1]$ is not the intersection of a countable collection of open sets.

41. Show that under the hypotheses of Theorem 6 there is a dense open set $\mathcal{O} \subseteq X$ such that each $x \in \mathcal{O}$ has a neighborhood U on which $\mathcal{F}$ is uniformly bounded.

42. By Hölder's Inequality, $L^2[a, b] \subseteq L^1[a, b]$. Show that the set $L^2[a, b]$, considered as a subset of the complete metric space $L^1[a, b]$, is of the first category.

43. Let f be a continuous real-valued function on $\mathbf{R}$ with the property that for each real number x, $\lim_{n \to \infty} f(nx) = 0$. Show that $\lim_{x \to \infty} f(x) = 0$.

44. Let f be a continuous real-valued function on $\mathbf{R}$ that has derivatives of all orders. Suppose that for each real number x, there is an index $n = n(x)$ for which $f^{(n)}(x) = 0$. Show that f is a polynomial. (Suggestion: Apply the Baire Category Theorem twice.)

14.4 THE NIKODYM METRIC SPACE: THE VITALI-HAHN-SAKS THEOREM AND THE DUNFORD-PETTIS THEOREM

Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Recall that the symmetric difference of two measurable sets A and B is the measurable set $A \Delta B$ defined to be the following disjoint union:

$$A \Delta B = [A \sim [A \cap B]] \cup [B \sim [A \cap B]]$$

Observe that if $A, B \in \mathcal{M}$, then $|\chi_A - \chi_B| = \chi_{A \Delta B}$ and therefore

$$\int_X |\chi_A - \chi_B| \, d\mu = \mu(A \Delta B). \tag{15}$$

Introduce the relation $\simeq$ on $\mathcal{M}$ by defining $A \simeq B$ provided that $\mu(A \Delta B) = 0$. This is an equivalence relation. Indeed, clearly this relation is reflexive and symmetric, and, from the inclusion $A \Delta B \subseteq (A \Delta C) \cup (C \Delta B)$, it follows that it is transitive. This relation induces a decomposition of $\mathcal{M}$ into equivalence classes, which we denote $\mathcal{M}/_{\simeq}$. For $A \in \mathcal{M}$, denote the equivalence class of A by $[A]$. On $\mathcal{M}/_{\simeq}$, define the Nikodym metric ρ_μ by

$$\rho_\mu([A], [B]) = \mu(A \Delta B) \text{ for all } A, B \in \mathcal{M}.$$

It follows from (15) that this is independent of the choice of representative in the equivalence class, and the triangle inequality follows from the triangle inequality in $L^1(X, \mu)$; the remaining two properties of a metric are evident. The space $(\mathcal{M}/_{\simeq}, \rho_\mu)$ is called the **Nikodym metric space** associated with the measure space $(X, \mathcal{M}, \mu)$. As was done with functions in $L^1(X, \mu)$, for simplicity and convenience, we identify measurable sets with their equivalence classes and write $(\mathcal{M}, \rho_\mu)$ for $(\mathcal{M}/_{\simeq}, \rho_\mu)$.

Lemma 8 *Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Then the Nikodym metric space $(\mathcal{M}, \rho_\mu)$ is complete. Moreover, if $L^1(X, \mu)$ is separable, so is $(\mathcal{M}, \rho_\mu)$.*

Proof Define the operator $T\colon \mathcal{M} \to L^1(X, \mu)$ by $T(E) = \chi_E$. Then (15) is the assertion that the operator T is an isometry, that is,

$$\rho_\mu(A, B) = \|T(A) - T(B)\|_1 \text{ for all } A, B \in \mathcal{M}. \tag{16}$$

Let $\{A_n\}$ be a Cauchy sequence in $(\mathcal{M}, \rho_\mu)$. Then $\{T(A_n)\}$ is a Cauchy sequence in $L^1(X, \mu)$. According to the Riesz-Fischer Theorem, there is a function $f \in L^1(X, \mu)$ such that $\{T(A_n)\} \to f$ in $L^1(X, \mu)$ and a subsequence of $\{T(A_n)\}$ that converges pointwise to f almost everywhere on X. Since each $T(A_n)$ takes the values 0 and 1, if we define A_0 to be the points in X at which the pointwise convergent subsequence converges to 1, then $f = \chi_{A_0}$ almost everywhere on X. Therefore, by (15), $\{A_n\} \to A_0$ in $(\mathcal{M}, \rho_\mu)$. Consequently, $(X, \mathcal{M}, \mu)$ is complete. Since a subspace of a separable metric space is separable, in view of the isometry T, if $L^1(X, \mu)$ is separable, so is $(\mathcal{M}, \rho_\mu)$. $\square$

Recall that a measure $\nu\colon \mathcal{M} \to [0, \infty]$ is said to be absolutely continuous with respect to μ provided that $\nu(E) = 0$ whenever $\mu(E) = 0$. According to Proposition 16 of Chapter 11, if $\nu(X) < \infty$, this is equivalent to the assertion that for each $\epsilon > 0$, there is a $\delta > 0$ such that if $\mu(E) < \delta$ then $\nu(E) < \epsilon$.

Lemma 9 *Let $(X, \mathcal{M}, \mu)$ be a finite measure space and let the measure $\nu\colon \mathcal{M} \to [0, \infty)$ be absolutely continuous with respect to μ. Then ν is a continuous function on the Nikodym metric space $(\mathcal{M}, \rho_\mu)$.*

Proof It is first necessary to verify that ν is properly defined with respect to the equivalence relation, that is, if $\mu(A \Delta B) = 0$, then $\nu(A) = \nu(B)$. However, if $\mu(A \Delta B) = 0$, then, by absolute continuity, $\nu(A \Delta B) = 0$, so that $\nu(A) = \nu(A \cap B) = \nu(B)$. Let $\epsilon > 0$. Since $\nu(X) < \infty$, by the ϵ-δ criterion for the absolute continuity of measures, there is a $\delta > 0$ such that if $\mu(E) < \delta$, then $\nu(E) < \epsilon/2$. Let $A, B \in \mathcal{M}$ and $\mu(A \Delta B) < \delta$. We claim that $|\nu(A) - \nu(B)| < \epsilon$, and therefore ν is a uniformly continuous function on the Nikodym metric space. Indeed, $\mu(A \sim [A \cap B]) < \delta$, and therefore, by the excision property of ν and the choice of δ, $0 \le \nu(A) - \nu(A \cap B) < \epsilon/2$. Similarly, $0 \le \nu(B) - \nu(A \cap B) < \epsilon/2$. Consequently, by the triangle inequality,

$$|\nu(A) - \nu(B)| \le |\nu(A) - \nu(A \cap B)| + |\nu(B) - \nu(A \cap B)| < \epsilon. \qquad \square$$

Definition *Let $(X, \mathcal{M}, \mu)$ be a finite measure space. A sequence of measures $\{\nu_n\colon \mathcal{M} \to [0, \infty)\}$ is said to be* **uniformly absolutely continuous**[3] *with respect to μ provided that for each $\epsilon > 0$, there is a $\delta > 0$ such that for any measurable set E and any n,*

$$\textit{if } \mu(E) < \delta\,, \ \textit{then } \nu_n(E) < \epsilon.$$

Lemma 10 *Let $(X, \mathcal{M}, \mu)$ be a finite measure space and $\{\nu_n\colon \mathcal{M} \to [0, \infty)\}$ be a sequence of measures, each of which is absolutely continuous with respect to μ. Assume that there is an $E_0 \in \mathcal{M}$ at which the sequence of continuous functions $\nu_n\colon (\mathcal{M}, \rho_\mu) \to [0, \infty)\}$ on the Nikodym metric space $(\mathcal{M}, \rho_\mu)$ is equicontinuous. Then $\{\nu_n\colon \mathcal{M} \to [0, \infty)\}$ is uniformly absolutely continuous withe respect to μ.*

Proof Let $\epsilon > 0$. Choose $\delta > 0$ such that for all n, if $\rho_\mu(A, E_0) < \delta$, then$|\nu_n(A) - \nu_n(E_0)| < \epsilon/2$. We claim that if $E \in \mathcal{M}$ and $\mu(E) < \delta$, then, for all n, $\nu_n(E) < \epsilon$. Indeed, let $E \in \mathcal{M}$ with $\mu(E) < \delta$. Define

$$E' = E_0 \sim [E \cap E_0] \text{ and } E'' = E_0 \cup [E \sim E_0].$$

Then

$$\rho_\mu(E_0, E') < \delta \text{ and } \rho_\mu(E_0, E'') < \delta.$$

Therefore, by the choice of δ and the excision property of measure, for all n,

$$\nu_n(E \cap E_0) = |\nu_n(E_0) - \nu_n(E')| < \epsilon/2 \text{ and } \nu_n(E \sim E_0) = |\nu_n(E_0) - \nu_n(E'')| < \epsilon/2.$$

It follows that $\nu_n(E) = \nu_n(E') + \nu(E'') < \epsilon$. $\square$

[3] What is here called "uniformly absolutely continuous" might also be called equi absolutely continuous. There is no standard terminology.

A sequence of set functions $\{\nu_n: \mathcal{M} \to [0, \infty)\}$ is said to **converge setwise** to the set-function $\nu: \mathcal{M} \to [0, \infty)$ provided that

$$\nu(E) = \lim_{n\to\infty} \nu_n(E) \text{ for all } E \in \mathcal{M}.$$

Lemma 11 *Let $(X, \mathcal{M}, \mu)$ be a finite measure space and $\{\nu_n: \mathcal{M} \to [0, \infty)\}$ be a sequence of measures that is uniformly absolutely continuous with respect to μ. If $\{\nu_n\}$ converges setwise on $\mathcal{M}$ to $\nu: \mathcal{M} \to [0, \infty)$, then ν is a measure that is absolutely continuous with respect to μ.*

Proof The setwise limit of finitely additive set-functions is finitely additive and therefore ν is finitely additive. To verify that ν is countably additive, let $\{E_k\}_{k=1}^{\infty}$ be a disjoint collection of measurable sets. By the finite additivity of ν, for each n,

$$\nu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{n} \nu(E_k) + \nu\left(\bigcup_{k=n+1}^{\infty} E_k\right). \tag{17}$$

Let $\epsilon > 0$. By the uniform absolute continuity of the sequence $\{\nu_n\}$ with respect to μ, there is a $\delta > 0$ such that for E measurable and each m,

$$\text{if } \mu(E) < \delta, \text{ then } \nu_m(E) < \epsilon/2.$$

Since μ is countably additivity, there is an N for which

$$\mu\left(\bigcup_{k=N+1}^{\infty} E_k\right) < \delta \text{ and so, for all } m,\ \nu_m\left(\bigcup_{k=N+1}^{\infty} E_k\right) < \epsilon/2.$$

Take the limit as $m \to \infty$ to conclude that $\nu\left(\bigcup_{k=N+1}^{\infty} E_k\right) < \epsilon$. Therefore, in view of (17), ν is countably additive. The absolute continuity of ν with respect to μ follows from uniform absolute continuity of $\{\nu_n\}$ with respect to μ. $\square$

The following theorem asserts that, in the preceding lemma, one can dispense with the assumption that the sequence is uniformly absolutely continuous: uniform absolute continuity is a consequence of setwise convergence.

The Vitali-Hahn-Saks Theorem *Let $(X, \mathcal{M}, \mu)$ be a finite measure space and let $\{\nu_n: \mathcal{M} \to [0, \infty)\}$ be a sequence of measures, each of which is absolutely continuous with respect to μ. Suppose that $\{\nu_n\}$ converges setwise* to $\nu: \mathcal{M} \to [0, \infty)$. *Then ν is a measure that is absolutely continuous with respect to μ.*

Proof According to Lemma 8, the Nikodym metric space is complete, and $\{\nu_n\}$ induces a sequence of continuous functions on this metric space that converges pointwise (that is, setwise) to the real-valued function ν. It follows from Theorem 7 that there is a set $E_0 \in \mathcal{M}$ at which the sequence of functions $\{\nu_n: \mathcal{M} \to \mathbf{R}\}$ is equicontinuous. By Lemma 10, $\{\nu_n\}$ is uniformly absolutely continuous with respect to μ. An appeal to the preceding lemma completes the proof. $\square$

Theorem 12 (Nikodym) *Assume that $\{\nu_n: \mathcal{M} \to [0, \infty)\}$ is a sequence of measures that converges setwise on $\mathcal{M}$ to the set-function $\nu: \mathcal{M} \to [0, \infty)\}$, and $\{\nu_n(X)\}$ is bounded. Then ν is a measure on $\mathcal{M}$.*

Proof For a measurable set E, define

$$\mu(E) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \nu_n(E). \tag{18}$$

The verification that μ is a finite measure on $\mathcal{M}$ is left as an exercise. It is clear that each ν_n is absolutely continuous with respect to μ. The conclusion now follows from the Vitali-Hahn-Saks Theorem. □

In Section 11.4, it was shown that for a σ-finite measure space $(X, \mathcal{M}, \mu)$ and $1 < p < \infty$, if $L^q(X, \mu)$ is separable, where q is the conjugate of p, then a bounded sequences in $L^p(X, \mu)$ has a weakly convergent subsequence, while, in general, there are bounded sequences in $L^1(X, \mu)$ that fail to have weakly convergent subsequences. It therefore is interesting to identify sufficient conditions for a bounded sequence in $L^1(X, \mu)$ to possess a weakly convergent subsequence.

Theorem 13 (The Dunford-Pettis Theorem) *Let $(X, \mathcal{M}, \mu)$ be a finite measure space, $L^1(X, \mu)$ be separable, and the sequence of functions $\{f_n: X \to \mathbf{R}\}$ be bounded in $L^1(X, \mu)$. If $\{f_n\}$ is uniformly integrable, then $\{f_n\}$ has a subsequence that converges weakly in $L^1(X, \mu)$.*

Proof By examining the positive and negative parts of the f_n's, it may be assumed that each $f_n \geq 0$. For each n, define $\nu_n: \mathcal{M} \to [0, \infty)$ by

$$\nu_n(E) = \int_E f_n \, d\mu \text{ for all } E \in \mathcal{M}.$$

Then $\{\nu_n\}$ is a sequence of measures on $\mathcal{M}$, each of which is absolutely continuous with respect to μ. Clearly, the uniform integrability of $\{f_n\}$ is equivalent to the uniformly absolutely continuous with respect to μ of $\{\nu_n\}$. Arguing exactly as we did in the proof of Lemma 9, we see that the sequence $\{\nu_n\}$ is equicontinuous on $(\mathcal{M}, \rho_\mu)$. According to Lemma 8, the Nikodym metric space is complete and separable. As a consequence, by the Arzelà-Ascoli Lemma, a subsequence $\{\nu_{n_k}: \mathcal{M} \to [0, \infty)\}$ converges setwise to a real-valued function $\nu: \mathcal{M} \to [0, \infty)$. For notational convenience, assume the entire sequence converges. According to the Vitali-Hahn-Saks Theorem, $\nu: \mathcal{M} \to [0, \infty)$ is absolutely continuous with respect to μ, and therefore, by the Radon-Nikodym Theorem, there is a μ-integrable function $f: X \to [0, \infty)$ such that

$$\nu(E) = \int_E f \; dm \text{ for all } E \in \mathcal{M}.$$

Therefore

$$\lim_{n\to\infty} \int_E f_n \, d\mu = \int_E f \, d\mu \text{ for all } E \in \mathcal{M},$$

and so, for each simple function $\varphi\colon X \to \mathbf{R}$,

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

Let $g \in L^\infty(X, \mu)$. By the Simple Approximation Lemma, a sequence of simple functions converges in $L^\infty(X, \mu)$ to g. Consequently, since $\{f_n\colon X \to \mathbf{R}\}$ is bounded in $L^1(X, \mu)$,

$$\lim_{n\to\infty} \int_X f_n \cdot g \, d\mu = \int_E f \cdot g \, d\mu \text{ for all } g \in L^\infty(X, \mu).$$

According to the Riesz Representation Theorem, every bounded linear functional on $L^1(X, \mu)$ is represented by a function in $L^\infty(X, \mu)$. Therefore, $\{f_{n_k}\} \to f$ weakly in $L^1(X, \mu)$. $\square$

Corollary 14 *Assume that $E \subseteq \mathbf{R}^n$ has finite Lebesgue measure. Let $\{f_k\colon E \to \mathbf{R}\}$ be a sequence in $L^1(E, \mu_n)$ that is dominated by a function $g \in L^1(E, \mu_n)$, in the sense that*

$$|f_k| \leq g \text{ on } E \text{ for all } n. \tag{19}$$

Then $\{f_k\}$ has a subsequence that converges weakly in $L^1(E, \mu_n)$.

Proof According to Theorem 18 of Chapter 12, $L^1(E, \mu_n)$ is separable. It follows from (19) that the sequence of $\{f_k\}$ is bounded in $L^1(X, \mu)$ and uniformly integrable. Apply the Dunford-Pettis Theorem. $\square$

PROBLEMS

45. For measurable sets A and B in a finite measure space $(X, \mathcal{M}, \mu)$, show that

$$\rho_\mu(A, B) = \mu(A) + \mu(B) - 2 \cdot \mu(A \cap B).$$

46. Let $\{A_n\}$ be a sequence of measurable sets that converges to the measurable set A_0 with respect to the Nikodym metric. Show that $A_0 \simeq \cup_{n=1}^\infty [\cap_{k=n}^\infty A_k]$.
47. Let $(X, \mathcal{M}, \mu)$ be a finite measure space and $\nu\colon \mathcal{M} \to [0, \infty)$ a finitely additive set-function with the property that for each $\epsilon > 0$, there is a $\delta > 0$ such that for a measurable set E, if $\mu(E) < \delta$, then $\nu(E) < \epsilon$. Show that ν is a measure on $\mathcal{M}$.
48. Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Show that the normed linear space $L^1(X, \mu)$ is separable if and only if the associated Nikodym metric space is separable.
49. Give an example of a decreasing sequence $\{\mu_n\}$ of measures on $\mathcal{M}$ such that the set-function μ defined by $\mu(E) = \lim \mu_n(E)$ is not a measure.
50. Let $\{\mu_n\}$ a sequence of measures on $\mathcal{M}$ such that for each $E \in \mathcal{M}$, $\mu_{n+1}(E) \geq \mu_n(E)$. For each $E \in \mathcal{M}$, define $\mu(E) = \lim \mu_n(E)$. Show that μ is a measure on $\mathcal{M}$ if $\mu(X) < \infty$.
51. Verify the assertion, in the proof of the Dunford-Pettis Theorem, that it suffices to consider non-negative functions.
52. Find a bounded sequence in $L^1([a, b], m)$, where m is Lebesgue measure, which fails to have a weakly convergent subsequence.

53. Find a measure space $(X, \mathcal{M}, \mu)$ for which every bounded sequence in $L^1(X, \mu)$ has a weakly convergent subsequence.

54. Let $(X, \mathcal{M}, \mu)$ be a finite measure space and $\{\nu_n\}$ a sequence of finite measures of $\mathcal{M}$, each of which is absolutely continuous with respect to μ. Prove that the following are equivalent:

    (i) The sequence of measures $\{\nu_n\colon \mathcal{M} \to [0, \infty)\}$ is uniformly absolutely continuous with respect to the measure μ.

    (ii) The sequence of functions $\{\nu_n\colon \mathcal{M} \to \mathbf{R}\}$ is equicontinuous with respect to the metric ρ_μ.

    (iii) The sequence of Radon-Nikodym derivatives $\{\frac{d\nu_n}{d\mu}\}$ is uniformly integrable over X with respect to the measure μ.

CHAPTER 15

Topological Spaces: General Properties

Contents

15.1 Open Sets, Closed Sets, Bases, and Subbases 305
15.2 The Separation Properties 309
15.3 Countability and Separability 311
15.4 Continuous Mappings, Weak Topologies, and Metrizability 313
15.5 Compact Topological Spaces 317
15.6 Connected Topological Spaces 320

In the preceding two chapters, metric spaces were considered. In these spaces, the metric was used to define an open ball, and an open set has been defined to be the union of open balls. A number of concepts, for instance, continuity of mappings, were expressible solely in terms of the open sets associated with the metric. We here consider spaces in which the notion of an open set is fundamental. Such spaces are called topological spaces. They are more general than metric spaces. On a normed linear space X, there are important concepts, such as weak sequential convergence (we studied this in Chapter 8), which we prove cannot be formulated in the framework of a metric, but can be considered with respect to the weak topology. Also, pointwise convergence of functions can be considered with respect to the product topology, but is not a metric concept. Beyond generality, new concepts, such as the compactness of an arbitrary product of compact metric spaces, reveal properties of normed linear spaces which are not evident within the metric framework.

15.1 OPEN SETS, CLOSED SETS, BASES, AND SUBBASES

Definition *Let X be a non-empty set. A topology $\mathcal{T}$ for X is a collection of subsets of X, called* **open** *sets, possessing the following properties:*

(i) The entire set X and the empty-set $\emptyset$ are open;

(ii) The intersection of any finite collection of open sets is open;

(iii) The union of any collection of open sets is open.

A non-empty set X, together with a topology on X, is called a **topological space**. *For a point x in X, an open set that contains x is called a* **neighborhood** *of x.*

We sometimes denote a topological space by $(X, \mathcal{T})$. Often we are interested in only one topology for a given set of points, and in such cases we sometimes use the symbol X to denote both the set of points and the topological space $(X, \mathcal{T})$. When greater precision is needed, we make explicit the topology.

Proposition 1 *A subset E of a topological space X is open if and only if for each point x in E there is a neighborhood of x that is contained in E.*

Proof This follows immediately from the definition of neighborhood and the property of a topology that the union of a collection of open sets is again open. □

Metric Topology Consider a metric space (X, ρ). Define a subset $\mathcal{O}$ of X to be open provided that for each point $x \in \mathcal{O}$ there is an open ball centered at x that is contained in $\mathcal{O}$. Thus, the open sets are unions of collections of open balls. Proposition 1 of Chapter 13 is the assertion that this collection of open sets is a topology for X. We call it the **metric topology** induced by the metric ρ. As a particular case of a metric topology on a set we have the topology we call the Euclidean topology induced on $\mathbf{R}^n$ by the Euclidean metric[1].

The Discrete Topology Let X be any non-empty set. Define $\mathcal{T}$ to be the collection of all subsets of X. Then $\mathcal{T}$ is a topology for X called the discrete topology. For the discrete topology, every set containing a point is a neighborhood of that point. The discrete topology is induced by the discrete metric.

The Trivial Topology Let X be any non-empty set. Define $\mathcal{T}$ to be the collection of subsets of X consisting of $\emptyset$ and X. Then $\mathcal{T}$ is a topology for X called the trivial topology. For the trivial topology, the only neighborhood of a point is the whole set X.

Topological Subspaces Given a topological space $(X, \mathcal{T})$ and a non-empty subset E of X, we define the inherited topology $\mathcal{S}$ for E to consist of all sets of the form $E \cap \mathcal{O}$ where $\mathcal{O}$ belongs to $\mathcal{T}$. We call the topological space $(E, \mathcal{S})$ a **subspace** of $(X, \mathcal{T})$. If a subset of E is open as a subset of the topological space $(X, \mathcal{T})$, then it is open as a subset of the topological space $(E, \mathcal{S})$. The converse assertion holds for all subsets of E if and only if E is an open subset of $(X, \mathcal{T})$.

In elementary analysis, we define what it means for a subset of $\mathbf{R}$ to be open even if we have no need to use the word "topology." In Chapter 1, we proved that the topological space $\mathbf{R}$ has the property that every open set is the union of a countable disjoint collection of open intervals. In a metric space, every open set is the union of a collection of open balls. In a general topological space, it is often useful to distinguish a collection of open sets called a **base** for the topology: they are building blocks for the topology.

Definition *For a topological space $(X, \mathcal{T})$ and a point x in X, a collection of neighborhoods of x, $\mathcal{B}_x$, is called a* **base for the topology at** x *provided that for any neighborhood $\mathcal{U}$ of x, there is a set B in the collection $\mathcal{B}_x$ for which $B \subseteq \mathcal{U}$. A collection of open sets $\mathcal{B}$ is called a* **base for the topology** $\mathcal{T}$ *provided that it contains a base for the topology at each point.*

Observe that a subcollection B of a topology is a base for the topology if and only if every non-empty open set is the union of a subcollection of $\mathcal{B}$. Once a base for a topology is prescribed, the topology is completely defined: it consists of $\emptyset$ and unions of sets belonging

[1] Unless otherwise stated, by the topological space $\mathbf{R}^n$ we mean the set $\mathbf{R}^n$ with the Euclidean topology. In the problems we introduce more exotic topologies on $\mathbf{R}$ and $\mathbf{R}^2$ (see Problems 7 for the Sorgenfrey Line and 8 for the Moore Plane).

to the base. For this reason a topology is often defined by specifying a base. The following proposition describes the properties that a collection of subsets of X must possess in order for it to be a base for a topology.

Proposition 2 *For a non-empty set X, let $\mathcal{B}$ be a collection of subsets of X. Then $\mathcal{B}$ is a base for a topology for X if and only if*

(i) $\mathcal{B}$ covers X, that is, $X = \bigcup_{B \in \mathcal{B}} B$.
(ii) if B_1 and B_2 are in $\mathcal{B}$ and $x \in B_1 \cap B_2$, then there is a set B in $\mathcal{B}$ for which $x \in B \subseteq B_1 \cap B_2$.

The topology that has $\mathcal{B}$ as its base consists of $\emptyset$ and unions of subcollections of $\mathcal{B}$.

Proof Assume that $\mathcal{B}$ possesses properties (i) and (ii). Define $\mathcal{T}$ to be the collection of unions of subcollections of $\mathcal{B}$ together with $\emptyset$. We claim that $\mathcal{T}$ is a topology for X. Indeed, it follows from (i) that the set X is the union of all the sets in $\mathcal{B}$ and therefore it belongs to $\mathcal{T}$. Moreover, it is also clear that the union of a subcollection of $\mathcal{T}$ is also a union of a subcollection of $\mathcal{B}$ and therefore belongs to $\mathcal{T}$. It remains to show that if $\mathcal{O}_1$ and $\mathcal{O}_2$ belong to $\mathcal{T}$, then their intersection $\mathcal{O}_1 \cap \mathcal{O}_2$ belongs to $\mathcal{T}$. Indeed, let x belong to $\mathcal{O}_1 \cap \mathcal{O}_2$. Then there are sets B_1 and B_2 in $\mathcal{B}$ such that $x \in B_1 \subseteq \mathcal{O}_1$ and $x \in B_2 \subseteq \mathcal{O}_2$. Using (ii), choose B_x in $\mathcal{B}$ with $x \in B_x \subseteq B_1 \cap B_2$. Then, $\mathcal{O}_1 \cap \mathcal{O}_2 = \bigcup_{x \in \mathcal{O}} B_x$, the union of a subcollection of $\mathcal{B}$. Thus, $\mathcal{T}$ is a topology for which $\mathcal{B}$ is a base. We leave the proof that any base for a topology has properties (i) and (ii) as an exercise. $\square$

A base determines a unique topology. However, in general, a topology has many bases. For instance, the collection of open intervals is a base for the Euclidean topology on $\mathbf{R}$, while the collection of open, bounded intervals with rational end-points also is a base for this topology.

Example Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces. In the Cartesian product $X \times Y$, consider the collection of sets $\mathcal{B}$ consisting of products $\mathcal{O}_1 \times \mathcal{O}_2$, where $\mathcal{O}_1$ is open in X and $\mathcal{O}_2$ is open in Y. It is left as an exercise to show that $\mathcal{B}$ is a base for a topology on $X \times Y$. The topology is called the **product topology** on $X \times Y$.

Let $\mathcal{S}$ be a collection of subsets of X that covers X. Consider the collection of all intersections of finite subcollections $\mathcal{S}$. It is clear from Proposition 2 that this is a base for a topology $\mathcal{T}$ on X. We leave it as an exercise to show that this is the smallest topology that contains $\mathcal{S}$, in the sense that $\mathcal{S} \subseteq \mathcal{T}$, and $\mathcal{T} \subseteq \mathcal{T}'$, whenever $\mathcal{T}'$ is a topology on X for which $\mathcal{S} \subseteq \mathcal{T}'$. The collection $\mathcal{S}$ is said to be a **subbase** for $\mathcal{T}$.

Example Consider a closed, bounded interval $[a, b]$ as a topological space with the topology it inherits from $\mathbf{R}$. This space has a subbase consisting of intervals of the type $[a, c)$ or $(c, b]$ for $a < c < b$.

Definition *For a subset E of a topological space X, a point $x \in X$ is called a* **point of closure** *of E provided that every neighborhood of x contains a point in E. The collection of points of closure of E is called the* **closure** *of E and denoted by $\overline{E}$.*

It is clear that we always have $E \subseteq \overline{E}$. If E contains all of its points of closure, that is, $E = \overline{E}$, the set E is said to be **closed.**

Proposition 3 *For E a subset of a topological space X, its closure $\overline{E}$ is closed. Moreover, $\overline{E}$ is the smallest closed subset of X containing E, in the sense that if F is closed and $E \subseteq F$, then $\overline{E} \subseteq F$.*

Proof The set $\overline{E}$ is closed provided that it contains all its points of closure. Let x be a point of closure of $\overline{E}$. Consider a neighborhood $\mathcal{U}_x$ of x. There is a point $x' \in \overline{E} \cap \mathcal{U}_x$. Since x' is a point of closure of E and $\mathcal{U}_x$ is a neighborhood of x', there is a point $x'' \in E \cap \mathcal{U}_x$. Therefore, every neighborhood of x contains a point of E and hence $x \in \overline{E}$. So, the set $\overline{E}$ is closed. It is clear that if $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$, so that if F is closed and contains E, then $\overline{E} \subseteq \overline{F} = F$. □

Proposition 4 *A subset of a topological space X is open if and only if its complement in X is closed.*

Proof First suppose that E is open in X. Let x be a point of closure of $X \sim E$. Then x cannot belong to E because otherwise there would be a neighborhood x that is contained in E and therefore does not intersect $X \sim E$. Thus, x belongs to $X \sim E$ and hence $X \sim E$ is closed. Now suppose that $X \sim E$ is closed. Let x belong to E. Then there must be a neighborhood of x that is contained in E, for otherwise every neighborhood of x would contain points in $X \sim E$ and therefore x would be a point of closure of $X \sim E$. Since $X \sim E$ is closed, x would belong to $X \sim E$. This is a contradiction. □

Since $X \sim [X \sim E] = E$, it follows from the preceding proposition that a subset of a topological space X is closed if and only if its complement in X is open. Therefore, by De Morgan's Identities, the collection of closed subsets of a topological space possesses the following properties.

Proposition 5 *Let X be a topological space. The empty-set $\emptyset$ and the whole set X are closed; the union of any finite collection of closed subsets of X is closed; and the intersection of any collection of closed subsets of X is closed.*

PROBLEMS

1. Show that the discrete topology for a non-empty set X is a metric topology.
2. Show that the trivial topology on a set has a unique base.
3. Regarding Proposition 2, show that if $\mathcal{B}$ is a base for a topology, then properties (i) and (ii) hold.
4. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be topologies for a non-empty set X. Show that $\mathcal{T}_1 = \mathcal{T}_2$ if and only if there are bases $\mathcal{B}_1$ for $\mathcal{T}_1$ and $\mathcal{B}_2$ for $\mathcal{T}_2$ that are related as follows at each point x in X: for each neighborhood $\mathcal{N}_1$ of x belonging to $\mathcal{B}_1$, there is a neighborhood $\mathcal{N}_2$ of x belonging to $\mathcal{B}_2$ for which $\mathcal{N}_2 \subseteq \mathcal{N}_1$ and for each neighborhood $\mathcal{N}_2$ of x belonging to $\mathcal{B}_2$, there is a neighborhood $\mathcal{N}_1$ of x belonging to $\mathcal{B}_1$ for which $\mathcal{N}_1 \subseteq \mathcal{N}_2$.

5. Let E be a subset of a topological space X.
   (i) A point $x \in X$ is called an interior point of E provided that there is a neighborhood of x that is contained in E: the collection of interior points of E is called the interior of E and denoted by $\text{int}\, E$. Show that $\text{int}\, E$ is always open and E is open if and only if $E = \text{int}\, E$.
   (ii) A point $x \in X$ is called an exterior point of E provided that there is a neighborhood of x that is contained in $X \sim E$: the collection of exterior points of E is called the exterior of E and denoted by $\text{ext}\, E$. Show that $\text{ext}\, E$ is always open and E is closed if and only if $\overline{E} \sim E \subseteq \text{ext}\, E$.
   (iii) A point $x \in X$ is called a boundary point of E provided that every neighborhood of x contains points in E and points in $X \sim E$: the collection of boundary points of E is called the boundary of E and denoted by $\text{bd}\, E$. Show that (i) $\text{bd}\, E$ is always closed, (ii) E is open if and only if $E \cap \text{bd}\, E = \emptyset$, and (iii) E is closed if and only if $\text{bd}\, E \subseteq E$.
6. Let $\mathcal{O}$ be an open subset of a topological space X. For a subset E of X, show that $\mathcal{O}$ is disjoint from E if and only if it is disjoint from $\overline{E}$.
7. (The Sorgenfrey Line) Show that the collection of intervals of the form $[a, b)$, where $a < b$, is a base for a topology for the set of real numbers $\mathbf{R}$. The set of real numbers $\mathbf{R}$ with this topology is called the *Sorgenfrey Line*.
8. (The Moore Plane) Consider the upper half-plane, $\mathbf{R}^{2,+} = \{(x, y) \in \mathbf{R}^2 \mid y \geq 0\}$. For points (x, y) with $y > 0$, take as a basic open neighborhood a usual Euclidean open ball centered at (x, y) and contained in the upper half-plane. As a basic open neighborhood of a point $(x, 0)$ take the set consisting of the point itself and all the points in an open Euclidean ball in the upper half-plane that is tangent to the real line at $(x, 0)$. Show that this collection of sets is a base. The set $\mathbf{R}^{2,+}$ with this topology is called the *Moore Plane*.

15.2 THE SEPARATION PROPERTIES

In order to establish interesting results for topological spaces and continuous mappings between such spaces, it is necessary to enrich the rudimentary topological structure. In this section, we consider so-called separation properties for a topology on a set X, which ensure that the topology discriminates between certain disjoint pairs of sets and, as a consequence, ensure that there is a robust collection of continuous real-valued functions on X.

A neighborhood of a point in a topological space has been defined. For a subset K of a topological space X, by a **neighborhood of K** is meant an open set that contains K. We say that two disjoint subsets A and B of X can be **separated by disjoint neighborhoods** provided that there are neighborhoods of A and B, respectively, that are disjoint. For a topological space X, consider the following four separation properties:

The Tychonoff Separation Property For each two points u and v in X, there is a neighborhood of u that does not contain v and a neighborhood of v that does not contain u.

The Hausdorff Separation Property Any two points in X can be separated by disjoint neighborhoods.

The Regular Separation Property The Tychonoff separation property holds and, moreover, each closed set and point not in the set can be separated by disjoint neighborhoods.

The Normal Separation Property The Tychonoff separation property holds and, moreover, any two disjoint closed sets can be separated by disjoint neighborhoods.

We naturally call a topological space Tychonoff, Hausdorff, regular, or normal provided that it satisfies the respective separation property.

Proposition 6 *A topological space X is a Tychonoff space if and only if every set consisting of a single point is closed.*

Proof Let x be in X. The set $\{x\}$ is closed if and only if $X \sim \{x\}$ is open. Now $X \sim \{x\}$ is open if and only if for each point y in $X \sim \{x\}$ there is a neighborhood of y that is contained in $X \sim \{x\}$, that is, there is a neighborhood of y that does not contain x. $\square$

Proposition 7 *Every metric space is normal.*

Proof Let (X, ρ) be a metric space. For F a closed subset of X, in Chapter 13, we defined the continuous function $\text{dist}_F \colon X \to [0, \infty)$, which has the property that $\text{dist}_F(x) = 0$ only if $x \in F$. Let F_1 and F_2 be closed, disjoint subsets of X. Define

$$\mathcal{O}_1 = \{x \text{ in } X \mid \ \text{dist}_{F_1}(x) < \text{dist}_{F_2}(x)\} \text{ and } \mathcal{O}_2 = \{x \text{ in } X \mid \ \text{dist}_{F_2}(x) < \text{dist}_{F_1}(x)\}\,.$$

Therefore, $F_1 \subseteq \mathcal{O}_1$, $F_2 \subseteq \mathcal{O}_2$, and $\mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset$. Since the distance functions are continuous, the sets $\mathcal{O}_1$ and $\mathcal{O}_2$ are open. $\square$

Using obvious notation, the preceding two propositions provide the following string of inclusions between families of topologies on a set X:

$$\mathcal{T}_{\text{metric}} \subseteq \mathcal{T}_{\text{normal}} \subseteq \mathcal{T}_{\text{regular}} \subseteq \mathcal{T}_{\text{Hausdorff}} \subseteq \mathcal{T}_{\text{Tychonoff}}.$$

Proposition 8 *Let X be a Tychonoff topological space. Then X is normal if and only if whenever $\mathcal{U}$ is a neighborhood of a closed subset F of X, there is an open set $\mathcal{O}$ for which*

$$F \subseteq \mathcal{O} \subseteq \overline{\mathcal{O}} \subseteq \mathcal{U}.$$

Proof First assume that X is normal. Since F and $X \sim \mathcal{U}$ are disjoint closed sets, there are disjoint open sets $\mathcal{O}$ and $\mathcal{V}$ for which $F \subseteq \mathcal{O}$ and $X \sim \mathcal{U} \subseteq \mathcal{V}$. Thus, $\mathcal{O} \subseteq X \sim \mathcal{V} \subseteq \mathcal{U}$. Since $\mathcal{O} \subseteq X \sim \mathcal{V}$ and $X \sim \mathcal{V}$ is closed, $\overline{\mathcal{O}} \subseteq X \sim \mathcal{V} \subseteq \mathcal{U}$.

To prove the converse, suppose that the nested neighborhood property holds. Let A and B be disjoint closed subsets of X. Then $A \subseteq X \sim B$ and $X \sim B$ is open. Thus, there is an open set $\mathcal{O}$ for which $A \subseteq \mathcal{O} \subseteq \overline{\mathcal{O}} \subseteq X \sim B$. Therefore, $\mathcal{O}$ and $X \sim \overline{\mathcal{O}}$ are disjoint neighborhoods of A and B, respectively. $\square$

PROBLEMS

9. Show that if F is a closed subset of a normal space X, then the subspace F is normal. Is it necessary to assume that F is closed?
10. Let X be a topological space. Show that X is Hausdorff if and only if the diagonal $D = \{(x_1, x_2) \in X \times X \mid x_1 = x_2\}$ is a closed subset of $X \times X$.
11. Consider the set of real numbers with the topology that has as a subbase and sets of the form $(-\infty, c), c \in \mathbf{R}$. Show that this space is not Tychonoff.
12. (Zariski Topology) In $\mathbf{R}^n$ let $\mathcal{B}$ be the family of sets $\{x \in \mathbf{R}^n \mid p(x) \neq 0\}$, where p is a polynomial in n variables. Let $\mathcal{T}$ be the topology on X that has $\mathcal{B}$ as a subbase. Show that $\mathcal{T}$ is a topology for $\mathbf{R}^n$ that is Tychonoff but not Hausdorff.
13. Show the Sorgenfrey Line and the Moore Plane are Hausdorff (see Problems 7 and 8).

15.3 COUNTABILITY AND SEPARABILITY

We have defined what it means for a sequence in a metric space to converge. The following is the natural generalization of sequential convergence to topological spaces.

Definition *A sequence $\{x_n\}$ in a topological space X is said to* **converge** *to the point $x \in X$ provided that for each neighborhood $\mathcal{U}$ of x, there is an index N such that if $n \geq N$, then x_n belongs to $\mathcal{U}$. The point x is called a* **limit** *of the sequence.*

In a metric space, a sequence cannot converge to two different points, and so we refer to *the limit* of a sequence. In a general topological space, a sequence can converge to different points. For instance, for the trivial topology on a set, every sequence converges to every point. However, in a Hausdorff space, a sequence has a unique limit.

Definition *A topological space X is said to be* **first countable** *provided that there is a countable base at each point. The space X is said to be* **second countable** *provided that there is a countable base for the topology.*

Example Every metric space X is first countable since for $x \in X$, the countable collection of open balls $\{B(x, 1/n)\}_{n=1}^{\infty}$ is a base at x for the topology induced by the metric.

The proof of the following proposition is left as an exercise.

Proposition 9 *Let X be a first countable topological space. For a subset E of X, a point $x \in X$ is a point of closure of E if and only if x is a limit of a sequence in E. Therefore, a subset E of X is closed if and only if whenever a sequence in E converges to $x \in X$, then x belongs to E.*

In a topological space that is not first countable, it is possible for a point to be a point of closure of a set and yet no sequence in the set converges to the point (see Problem 19).

Definition *A subset E of topological space X is said to be* **dense** *in X provided that every open set in X contains a point of E. We call X* **separable** *provided that it has a countable dense subset.*

In Chapter 13, it was proven that a metric space is second countable if and only if it is separable. In a general topological space, a second countable space is separable, but a separable space, even one that is first countable, may fail to be second countable (see Problem 18).

A topological space is said to be **metrizable** provided that the topology is induced by a metric. Not every topology is induced by a metric. Indeed, we proved that a metric space is normal, so certainly the trivial topology on a set with more than one point is not metrizable. In the next section, we introduce the weak topology on a normed linear space X, and show that it is not metrizable if X has infinite dimension. It is natural to ask if it is possible to identify those topological spaces that are metrizable, that is, state criteria solely in terms of the open sets of the topology that are necessary and sufficient in order that the topology be induced by a metric. There are such criteria[2]. In a second countable topological space, there is the following simple necessary and sufficient criterion for metrizability.

The Urysohn Metrization Theorem *A topological space that has a countable base is metrizable if and only if it is normal.*

We already have shown that a metric space is normal. We postpone until the next chapter the proof, for spaces with a countable base, of the converse.

PROBLEMS

14. A topological space is said to be a *Lindelöf* space or to have the *Lindelöf* property provided that each open cover of X has a countable subcover. Show that if X is second countable, then it is *Lindelöf*.
15. Let X be an uncountable set of points, and let $\mathcal{T}$ consist of $\emptyset$ and all subsets of X that have finite complements. Show that $\mathcal{T}$ is a topology for X and that the space $(X, \mathcal{T})$ is not first countable.
16. Show that a second countable space is separable and every subspace of a second countable space is second countable.
17. Show that the Moore Plane is separable (see Problem 8). Show that the subspace $\mathbf{R} \times \{0\}$ of the Moore Plane is not separable. Conclude that the Moore Plane is not metrizable and not second countable.
18. Show that the Sorgenfrey Line is first countable but not second countable and yet the rationals are dense (see Problem 7). Conclude that the Sorgenfrey Line is not metrizable.
19. Let $X_1 = N \times N$, where N denotes the set of natural numbers and take $X = X_1 \cup \{\omega\}$, where ω does not belong to X_1. For each sequence $s = \{m_k\}$ of natural numbers and natural number n, define
$$B_{s,n} = \{\omega\} \cup \{(j,\, k) | j \geq m_k \text{ all } k \geq n\}.$$
    (i) Show that the sets $B_{s,n}$ together with the singleton sets $\{(j,\, k)\}$ form a base for a topology on X.
    (ii) Show that ω is a point of closure of X_1 even though no sequence $\{x_n\}$ from X_1 converges to ω.
    (iii) Show that the space X is separable but is not first countable and so is not second countable.
    (iv) Is X a Lindelöf space?

[2]The Nagata-Smirnov-Bing Metrization Theorem is such a result; see John Kelley's *General Topology* [Kel55].

15.4 CONTINUOUS MAPPINGS, WEAK TOPOLOGIES, AND METRIZABILITY

We defined continuity for mappings $f: X \to Y$ between metric spaces, in terms of convergent sequences: The mapping f is continuous at a point $x \in X$ provided that whenever a sequence in X converges to x the image sequence converges to $f(x)$. We then showed that this was equivalent to the ϵ-δ criterion expressed in terms of open balls. The concept of continuity extends naturally to mappings between topological spaces, although the sequential criterion is not primary.

Definition *For topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{S})$, a mapping $f: X \to Y$ is said to be* **continuous** *at the point x_0 in X provided that for any neighborhood $\mathcal{O}$ of $f(x_0)$, there is a neighborhood $\mathcal{U}$ of x_0 for which $f(\mathcal{U}) \subseteq \mathcal{O}$. A mapping f is said to be continuous provided that it is continuous at each point in X.*

Proposition 10 *A mapping $f: X \to Y$ between topological spaces X and Y is continuous if and only if for any open subset $\mathcal{O}$ in Y, its inverse image under f, $f^{-1}(\mathcal{O})$, is an open subset of X.*

Proof First suppose that f is continuous. Let $\mathcal{O}$ be open in Y. According to Proposition 1, to show that $f^{-1}(\mathcal{O})$ is open it suffices to show that each point in $f^{-1}(\mathcal{O})$ has a neighborhood that is contained in $f^{-1}(\mathcal{O})$. Let x belong to $f^{-1}(\mathcal{O})$. Then by the continuity of f at x there is a neighborhood of x that is mapped into $\mathcal{O}$ and therefore is contained in $f^{-1}(\mathcal{O})$. Conversely, if f^{-1} maps open sets to open sets, then it is immediate that f is continuous on all of X. $\square$

For a continuous mapping f of a topological space X to a topological space Y, by the definition of the subspace topology, the restriction of f to a subspace of X also is continuous. The proof of the next proposition is left as an exercise.

Proposition 11 *The composition of continuous mappings between topological spaces is continuous.*

Definition *Given two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ for a set X, if $\mathcal{T}_2 \subseteq \mathcal{T}_1$, $\mathcal{T}_2$ is said to be* **weaker** *than $\mathcal{T}_1$ and that $\mathcal{T}_1$ is* **stronger** *than $\mathcal{T}_2$.*

Proposition 12 *Let X be a non-empty set and $\mathcal{F} = \{f_\lambda: X \to X_\lambda\}_{\lambda \in \Lambda}$ be a collection of mappings parametrized by a set Λ, where each X_λ is a topological space. The topology $\mathcal{T}$ for X that has the collection of sets*

$$\mathcal{C} = \left\{ f_\alpha^{-1}(\mathcal{O}_\alpha) \;\middle|\; \lambda \in \Lambda,\, \mathcal{O}_\alpha \text{ open in } X_\alpha \right\} \tag{1}$$

as a subbase is the weakest among the topologies on X for which each mapping $f_\lambda: X \to X_\lambda$ is continuous. We call $\mathcal{T}$ the weak topology induced by $\mathcal{F}$.

Proof According to Proposition 10, for each λ in Λ, $f_\lambda: X \to X_\lambda$ is continuous if and only if the inverse image under f_λ of each open set in X_λ is open in X. And, we already observed that the weakest topology on X that contains the collection $\mathcal{C}$ defined by (1) is the topology that has $\mathcal{C}$ as a subbase. $\square$

Definition *A continuous mapping from a topological space X to a topological space Y is said to be a* **homeomorphism** *provided that it is one-to-one, maps X onto Y, and has a continuous inverse from Y to X.*

It is clear that the inverse of a homeomorphism is a homeomorphism and that the composition of homeomorphisms, when defined, is a homeomorphism. Two topological spaces X and Y are said to be **homeomorphic** if there is a homeomorphism between them. This is an equivalence relation among topological spaces, that is, it is reflexive, symmetric, and transitive. From a topological point of view two homeomorphic topological spaces are indistinguishable since, according to Proposition 10, for a homeomorphism f of X onto Y, a set E is open in X if and only if its image $f(E)$ is open in Y. The concept of homeomorphism plays the same role for topological spaces that isometry plays for metric spaces. As will be apparent from the following two examples, from the viewpoint either of metric spaces or normed linear spaces, two homeomorphic topological spaces may look quite different.

Example By composing a translation with a dilation, we see that any two open, bounded intervals of real numbers are homeomorphic, and since the tangent function defines a homeomorphism between $(-\pi/2, \pi/2)$ and $\mathbf{R}$, every open, bounded interval is homeomorphic to $\mathbf{R}$. As a metric space, $\mathbf{R}$ is complete, so the Baire Category Theorem holds for $\mathbf{R}$. However, the assumptions and conclusions of this theorem are purely topological, in that both are statements about open sets. Consequently, this theorem also holds for any open, bounded interval.

Example (**Mazur**) Let E be a Lebesgue measurable set of real numbers. For f in $L^1(E)$, define the function $\Phi(f)$ on E by $\Phi(f)(x) = \operatorname{sgn}(f(x))|f(x)|^{1/2}$. Then $\Phi(f)$ belongs to $L^2(E)$. We leave it as an exercise to show that for any two numbers a and b,

$$\left|\operatorname{sgn}(a) \cdot |a|^{1/2} - \operatorname{sgn}(b) \cdot |b|^{1/2}\right|^2 \leq 2 \cdot |a-b|,$$

and therefore

$$\|\Phi(f) - \Phi(g)\|_2^2 \leq 2 \cdot \|f-g\|_1 \text{ for all } f, g \text{ in } L^1(E).$$

From this we conclude that Φ is a continuous one-to-one mapping of $L^1(E)$ into $L^2(E)$. It also maps $L^1(E)$ onto $L^2(E)$ and its inverse Φ^{-1} is defined by $\Phi^{-1}(f)(x) = \operatorname{sgn}(f(x))|f(x)|^2$ for f in $L^2(E)$. Use Problem 34 to conclude that the inverse mapping Φ^{-1} is a continuous mapping from $L^2(E)$ to $L^1(E)$. Therefore, $L^1(E)$ is homeomorphic to $L^2(E)$, where each of these spaces is equipped with the topology induced by its L^p norm.

For a normed linear space X, we call the topology induced on X by the norm the **strong topology.** We considered the dual space of X, X^*, of bounded linear functionals on X, that is, the space of linear functionals on X that are continuous with respect to the strong topology.

Definition *Let X be a normed linear space. The topology induced on X by the dual space X^* is called the* **weak topology** *on X.*

Using linearity, it is nor difficult to see that a base at $x \in X$ for the weak topology on X comprises sets of the form

$$\mathcal{N}_{\epsilon,\, \psi_1,\, \ldots,\, \psi_n}(x) = \{x' \in X \mid \; |\psi_k(x' - x) < \epsilon \text{ for } 1 \leq k \leq n\}\,, \tag{2}$$

where $\epsilon > 0$ and $\{\psi_k\}_{k=1}^n$ is a finite subcollection of X^*. For topological concepts with respect to the weak topology, we use the adjective "weakly": so we have weakly compact sets, weakly open sets, etc. Therefore, consistent with our earlier definition of weak convergence, a sequence $\{x_n\}$ in X converges weakly to $x \in X$ if and only if

$$\lim_{n \to \infty} \psi(x_n) = \psi(x) \text{ for all } \psi \in X^*. \tag{3}$$

The following theorem provides a very explicit reason why an analyst should not stick just with metric spaces.

Theorem 13 *The weak topology on an infinite dimensional normed linear space is not metrizable.*

Proof Let X be an infinite dimensional normed linear space. We argue by contradiction. Suppose that there is a metric ρ on X that induces the weak topology. Fix a natural number n and consider $\mathcal{O}_n = \{x \in X \mid \rho(x, 0) < 1/n\}$. As was just observed, there are m functionals, $\psi_1, \ldots, \psi_m$, in X^* and an $\epsilon > 0$ for which

$$\mathcal{N}_{\epsilon,\, \psi_1,\, \ldots,\, \psi_m}(0) \subseteq \mathcal{O}_n.$$

Since X has infinite dimension, there is a finite dimensional subspace X_m of X with dimension larger than m. It follows from linear algebra, say, by identifying X_m with Euclidean space and restricting these functionals to X_m, that there is a point $x_n \neq 0$ in X_m for which $\psi_k(x_n) = 0$ for $1 \leq k \leq m$. Then $x_n \in \mathcal{O}_n$. We may assume $\|x_n\| = n$. The sequence $\{x_n\}$ has the property that

$$\rho(x_n, 0) < 1/n \text{ for all } n,$$

so that it converges with respect to the metric, and therefore weakly, to the origin. However, by the Uniform Boundedness Theorem in Chapter 8, a weakly convergent sequence is bounded with respect to the norm. This sequence is not bounded. We conclude that the weak topology on X is not metrizable. $\square$

In Chapter 18, we will prove that if a normed linear space has a separable dual, then the weak topology on a bounded subset is metrizable.

PROBLEMS

20. Let f be a mapping of the topological space X to the topological space Y and $\mathcal{S}$ be a subbase for the topology on Y. Show that f is continuous if and only if the inverse image under f of every set in $\mathcal{S}$ is open in X.
21. Let X be a topological space.
    (i) If X has the trivial topology, find all continuous mappings of X into $\mathbf{R}$.
    (ii) If X has the discrete topology, find all continuous mappings of X into $\mathbf{R}$.

(iii) Find all continuous one-to-one mappings from $\mathbf{R}$ to X if X has the discrete topology.

(iv) Find all continuous one-to-one mappings from $\mathbf{R}$ to X if X has the trivial topology.

22. For topological spaces X and Y, let f map X to Y. Which of the following assertions are equivalent to the continuity of f? Verify your answers.
    (i) The inverse image under f of every closed subset of Y is closed in X.
    (ii) If $\mathcal{O}$ is open in X, then $f(\mathcal{O})$ is open in Y.
    (iii) If F is closed in X, then $f(F)$ is closed in Y.
    (iv) For each subset A of X, $f(\overline{A}) \subseteq \overline{f(A)}$.
23. Prove Proposition 11.
24. Prove Proposition 12.
25. Prove that the sum and product of two real-valued continuous functions defined on a topological space X are themselves continuous.
26. Let $\mathcal{F}$ be a collection of real-valued functions on a set X. Find necessary and sufficient conditions on $\mathcal{F}$ in order that X, considered as a topological space with the weak topology induced by $\mathcal{F}$, is Tychonoff.
27. For topological spaces X and Y, let the mapping $f\colon X \to Y$ be one-to-one and onto. Show that the following assertions are equivalent.
    (i) f is a homeomorphism of X onto Y.
    (ii) A subset E of X is open in X if and only if $f(E)$ is open in Y.
    (iii) A subset E of X is closed in X if and only if $f(E)$ is closed in Y.
    (iv) The image of the closure of a set is the closure of the image, that is, for each subset A of X, $f(\overline{A}) = \overline{f(A)}$.
28. For topological spaces X and Y, let f be a continuous mapping from X onto Y. If X is Hausdorff, is Y Hausdorff? If X is normal, is Y normal?
29. Show that the inverse of a homeomorphism is a homeomorphism and the composition of two homeomorphisms, when defined, is again a homeomorphism.
30. Suppose that a topological space X has the property that every continuous real-valued function on X takes a minimum value. Show that any topological space that is homeomorphic to X also possesses this property.
31. Suppose that a topological space X has the property that every continuous real-valued function on X has an interval as its image. Show that any topological space that is homeomorphic to X also possesses this property.
32. Show that $\mathbf{R}$ is homeomorphic to the open bounded interval $(0, 1)$, but is not homeomorphic to the closed bounded interval $[0, 1]$.
33. Let X and Y be topological spaces and consider a mapping f from X to Y. Suppose $X = X_1 \cup X_2$ and the restrictions of f to the topological subspaces X_1 and to X_2 are continuous. Show that f need not be continuous at any point in X. Show that f is continuous on X if X_1 and X_2 are open. Compare this with the case of measurable functions and the inheritance of measurability from the measurability of restrictions.
34. Show that for any two numbers a and b,
$$\left| \operatorname{sgn}(a) \cdot |a|^2 - \operatorname{sgn}(b) \cdot |b|^2 \right| \leq 2 \cdot |a - b|(|a| + |b|).$$

15.5 COMPACT TOPOLOGICAL SPACES

Compactness has been considered for metric spaces. Several characterizations of compactness were considered, together with properties of continuous mappings and continuous real-valued functions defined on compact metric spaces. The concept of compactness is extended to topological spaces as follows.

Recall that a collection of sets $\{E_\lambda\}_{\lambda\in\Lambda}$ is said to be a **cover** of a set E provided that $E \subseteq \bigcup_{\lambda\in\Lambda} E_\lambda$. If each E_λ is contained in a topological space, a cover is said to be open provided that each set in the cover is open.

Definition *A topological space X is said to be* **compact** *provided that every open cover of X has a finite subcover. A subset K of X is called compact provided that K, considered as a topological space with the subspace topology inherited from X, is compact.*

In view of the definition of the subspace topology, a subset K of X is compact provided that every covering of K by a collection of open subsets of X has a finite subcover. Certain results regarding compactness in a topological space carry over directly from the metric space setting; for example, the image of a compact topological space under a continuous mapping also is compact. Other properties of compact metric spaces, for example, the equivalence of compactness and sequential compactness, carry over to the topological setting only for spaces that possess some additional topological structure. Other properties of compact metric spaces, such as total boundedness, have no simple correspondent in the topological setting.

Recall that a collection of sets is said to have the **finite intersection property** provided that every finite subcollection has non-empty intersection. Since a subset of a topological space X is closed if and only if its complement in X is open, we have, by De Morgan's Identities, the following extension to topological spaces of a result we previously established for metric spaces.

Proposition 14 *A topological space X is compact if and only if every collection of closed subsets of X that possesses the finite intersection property has non-empty intersection.*

Proposition 15 *A closed subset K of a compact topological space X is compact.*

Proof Let $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$ be an open cover for K by open subsets of X. Since $X \sim K$ is an open subset of X, $[X \sim F] \cup \{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$ is an open cover of X. By the compactness of X this cover has a finite subcover, and, by possibly removing the set $X \sim K$ from this finite subcover, the remaining collection is a finite subcollection of $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$ that covers K. Thus K is compact. □

We proved that a compact subspace K of a metric space X must be a closed subset of X. This is also true for topological spaces that are Hausdorff.

Proposition 16 *A compact subspace K of a Hausdorff topological space X is a closed subset of X.*

Proof We will show that $X \sim K$ is open so that K must be closed. Let y belong to $X \sim K$. Since X is Hausdorff, for each $x \in K$ there are disjoint neighborhoods $\mathcal{O}_x$ and $\mathcal{U}_x$ of x and y, respectively. Then $\{\mathcal{O}_x\}_{x\in K}$ is an open cover of K, and so, since K is compact, there is a finite subcover $\{\mathcal{O}_{x_1}, \mathcal{O}_{x_2}, \ldots, \mathcal{O}_{x_n}\}$. Define $\mathcal{N} = \bigcap_{i=1}^{n} \mathcal{U}_{x_i}$. Then $\mathcal{N}$ is a neighborhood of y which is disjoint from each $\mathcal{O}_{x_i}$ and hence is contained in $X \sim K$. Therefore $X \sim K$ is open. □

Definition *A topological space X is said to be* **sequentially compact** *provided that each sequence in X has a subsequence that converges to a point of X.*

We have shown that a metric space is compact if and only if it is sequentially compact. The same holds for topological spaces that are second countable.

Proposition 17 *A topological space that has a countable base is compact if and only if it is sequentially compact.*

Proof Assume that X has a countable base. First assume that X is compact. Let $\{x_n\}$ be a sequence in X. For each n, let F_n be the closure of the non-empty set $\{x_k \mid k \geq n\}$. Then $\{F_n\}$ is a descending sequence of non-empty closed sets. Since $\{F_n\}$ has the finite intersection property, by Proposition 14, $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$, choose a point x_0 in this intersection. Since X has a countable base, there is a countable base at x_0. Let $\{B_n\}_{n=1}^{\infty}$ be a base for the topology at the point x_0. By taking appropriate intersections, we may assume that each $B_{n+1} \subseteq B_n$. Since x_0 belongs to the closure of $\{x_k \mid k \geq n\}$, for each n, the neighborhood B_n has non-empty intersection with $\{x_k \mid k \geq n\}$. Therefore, we may inductively select a strictly increasing sequence of indices $\{n_k\}$ such that for each index k, $x_{n_k} \in B_k$. Since for each neighborhood $\mathcal{O}$ of x_0, there is an index N for which $B_n \subseteq \mathcal{O}$ for $n \geq N$, the subsequence $\{x_{n_k}\}$ converges to x_0. Thus, X is sequentially compact.

Now suppose that X is sequentially compact. Since X is second countable, every open cover has a countable subcover. Therefore, to show that X is compact it suffices to show that every countable open cover of X has a finite subcover. Let $\{\mathcal{O}_n\}_{n=1}^{\infty}$ be such a cover. We argue by contradiction. Assume that there is no finite subcover. Then for each index n, there is an index $m(n) > n$ for which $\mathcal{O}_{m(n)} \sim \bigcup_{i=1}^{n} \mathcal{O}_i \neq \emptyset$. For each natural number n, choose $x_n \in \mathcal{O}_{m(n)} \sim \bigcup_{i=1}^{n} \mathcal{O}_i$. Then, since X is sequentially compact, a subsequence of $\{x_n\}$ converges to $x_0 \in X$. But $\{\mathcal{O}_n\}_{n=1}^{\infty}$ is an open cover of X, so there is some $\mathcal{O}_N$ that is a neighborhood of x_0. Therefore, there are infinitely many indices n for which x_n belongs to $\mathcal{O}_N$. This is not possible since $x_n \notin \mathcal{O}_N$ for $n > N$. □

Theorem 18 *A compact, Hausdorff space is normal.*

Proof Let X be compact and Hausdorff. We first show it is regular, that is, each closed set and point not in the set can be separated by disjoint neighborhoods. Let F be a closed subset of X and x belong to $X \sim F$. Since X is Hausdorff, for each $y \in F$ there are disjoint neighborhoods $\mathcal{O}_y$ and $\mathcal{U}_y$ of x and y, respectively. Then $\{\mathcal{U}_y\}_{y \in F}$ is an open cover of F. But F is compact. Thus, here is a finite subcover $\{\mathcal{U}_{y_1}, \mathcal{U}_{y_2}, \ldots, \mathcal{U}_{y_n}\}$. Define $\mathcal{N} = \bigcap_{i=1}^{n} \mathcal{O}_{y_i}$. Then $\mathcal{N}$ is a neighborhood of y which is disjoint from $\bigcup_{i=1}^{n} \mathcal{U}_{y_i}$, a neighborhood of F. Thus X is regular. A repeat of this argument, now using regularity, shows that X is normal. □

Proposition 19 *The continuous image of a compact topological space is compact.*

Proof Let f be a continuous mapping of compact topological space X to a topological space Y. Let $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$ be a covering of $f(X)$ by open subsets of Y. Then, by the continuity of f, $\{f^{-1}(\mathcal{O}_\lambda)\}_{\lambda\in\Lambda}$ is an open cover of X. By the compactness of X, there is a finite subcollection $\{f^{-1}(\mathcal{O}_{\lambda_1}), \ldots, f^{-1}(\mathcal{O}_{\lambda_n})\}$ that also covers X. The finite collection $\{\mathcal{O}_{\lambda_1}, \ldots, \mathcal{O}_{\lambda_n}\}$ covers $f(X)$. □

Proposition 20 *A continuous one-to-one mapping f of a compact space X onto a Hausdorff space Y is a homeomorphism.*

Proof In order to show that f is a homeomorphism, it is necessary to show that f maps open sets to open sets or equivalently closed sets to closed sets. Let F be a closed subset of X. Then F is compact, since X is compact. Therefore, by Proposition 19, $f(F)$ is compact. Hence, by Proposition 16, since Y is Hausdorff, $f(F)$ is closed. □

Corollary 21 *A continuous real-valued function on a compact topological space takes a maximum and minimum functional value.*

Proof Let X be compact and $f\colon X \to \mathbf{R}$ be continuous. By the preceding proposition, $f(X)$ is a compact set of real numbers. Thus $f(X)$ is closed and bounded. But a closed and bounded set of real numbers contains a smallest and largest member. □

PROBLEMS

35. A topological space is said to be countably compact provided that every countable open cover has a finite subcover. For a second countable space X, show that X is compact if and only if it is countably compact.

36. (Fréchet Intersection Theorem) Let X be a topological space. Prove that X is countably compact if and only if whenever $\{F_n\}$ is a descending sequence of non-empty closed subsets of X, the intersection $\bigcap_{n=1}^{\infty} F_n$ is non-empty.

37. Let X be compact, Hausdorff and $\{F_n\}_{n=1}^{\infty}$ be a descending collection of closed subsets of X. Let $\mathcal{O}$ be a neighborhood of the intersection $\bigcap_{n=1}^{\infty} F_n$. Show there is an index N such that $F_n \subseteq \mathcal{O}$ for $n \geq N$.

38. Show that it is not possible to express a closed, bounded interval of real numbers as the pairwise disjoint union of a countable collection (having more than one member) of closed, bounded intervals.

39. Let f be a continuous mapping of the compact space X onto the Hausdorff space Y. Show that any mapping g of Y into Z for which $g \circ f$ is continuous must itself be continuous.

40. Let $(X, \mathcal{T})$ be a topological space.
    (i) Prove that if $(X, \mathcal{T})$ is compact, then $(X, \mathcal{T}_1)$ is compact for any topology $\mathcal{T}_1$ weaker than $\mathcal{T}$.
    (ii) Show that if $(X, \mathcal{T})$ is Hausdorff, then $(X, \mathcal{T}_2)$ is Hausdorff for any topology $\mathcal{T}_2$ stronger than $\mathcal{T}$.
    (iii) Show that if $(X, \mathcal{T})$ is compact and Hausdorff, then any strictly weaker topology is not Hausdorff and any strictly stronger topology is not compact.

41. (The Compact-Open Topology) Let X and Y be Hausdorff topological spaces and Y^X the collection of maps from X into Y. On Y^X we define a topology, called the compact-open topology, by taking as a subbase sets of the form $\mathcal{U}_{K,\mathcal{O}} = \{f\colon X \to Y \mid f(K) \subseteq \mathcal{O}\}$, where K is a compact subset of X and $\mathcal{O}$ is an open subset of Y. Thus the compact-open topology is the weakest topology on Y^X such that the sets $\mathcal{U}_{K,\mathcal{O}}$ are open.
    (i) Let $\{f_n\}$ be a sequence in Y^X that converges with respect to the compact-open topology to $f \in Y^X$. Show that $\{f_n\}$ converges pointwise to f on X.
    (ii) Now assume that Y is a metric space. Show that a sequence $\{f_n\}$ in Y^X converges with respect to the compact-open topology to a continuous function $f \in Y^X$ if and only if $\{f_n\}$ converges to f uniformly on each compact subset K of X.
42. (Dini's Theorem) Let $\{f_n\}$ be a sequence of continuous real-valued functions on a countably compact space X. Suppose that for each $x \in X$, the sequence $\{f_n(x)\}$ decreases monotonically to zero. Show that $\{f_n\}$ converges to zero uniformly.

15.6 CONNECTED TOPOLOGICAL SPACES

Two non-empty open subsets of a topological space X are said to **separate** X if they are disjoint and their union is X. A topological space which cannot be separated by such a pair is said to be **connected**. Since the complement of an open set is closed, each of the open sets in a separation of a space is also closed. Therefore, a topological space is connected if and only if the only subsets that are both open and closed are the whole space and the empty-set.

A subset E of X is said to be connected provided that it is a connected topological subspace. Thus a subset E of X is connected if there do not exist open subsets $\mathcal{O}_1$ and $\mathcal{O}_2$ of X for which

$$\mathcal{O}_1 \cap E \neq \emptyset,\ \mathcal{O}_2 \cap E \neq \emptyset,\ E \subseteq \mathcal{O}_1 \cup \mathcal{O}_2 \text{ and } E \cap \mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset.$$

Proposition 22 *The continuous image of a connected topological space is connected.*

Proof Let X be connected, $f\colon X \to Y$ be continuous, and $f(X)$ have the subspace topology inherited from Y. We argue by contradiction. Suppose $f(X)$ is not connected. Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be a separation of $f(X)$. Now, for $k = 1, 2, \mathcal{O}_k = \mathcal{O}'_k \cap f(X)$ with each $\mathcal{O}'_k$ open in Y. Then $f^{-1}(\mathcal{O}_1)$ and $f^{-1}(\mathcal{O}_2)$ are disjoint non-empty open sets in X whose union is X. Therefore, this pair is a separation of X, in contradiction to the connectedness of X. $\square$

We leave it as an exercise to show that for a set C of real numbers, the following are equivalent:

$$(i)\ C \text{ is an interval; } (ii)\ C \text{ is convex; } (iii)\ C \text{ is connected.} \tag{4}$$

Definition *A topological space X is said to have the intermediate value property provided that the image of any continuous real-valued function on X is an interval.*

In Chapter 13, we proved that a metric space has the extreme value property if and only if it is compact. The following is a companion criterion for topological spaces pertaining to connectedness.

Proposition 23 *A topological space has the intermediate value property if and only if it is connected.*

Proof According to (4), a connected set of real numbers is an interval. It follows from Proposition 22 that a connected topological space has the intermediate value property. To prove the converse, we suppose that X is a topological space that is not connected and conclude that it fails to have the intermediate value property. Indeed, since X is not connected, there is a pair of non-empty open subsets of X, $\mathcal{O}_1$ and $\mathcal{O}_2$, for which $X = \mathcal{O}_1 \cup \mathcal{O}_2$. Define the function f on X to take the value 0 on $\mathcal{O}_1$ and 1 on $\mathcal{O}_2$. Then f is continuous, since $f^{-1}(A)$ is an open subset of X for every subset A of $\mathbf{R}$ and therefore f is continuous. On the other hand, f fails to have the intermediate value property. □

If a topological space is not connected, then for any separation of the space, a subspace that has non-empty intersection with each of the sets in the separation also fails to be connected. Moreover, the image under a continuous map of an interval of real numbers is connected. Therefore, a topological space X is connected if for each pair of points $u, v \in X$, there is a continuous map $f\colon [0,1] \to X$ for which $f(0) = u$ and $f(1) = v$. A topological space possessing this property is said to be **arcwise connected.** While an arcwise-connected topological space is connected, there are connected spaces that fail to be arcwise connected (see Problem 45). However, for an open subset of a Euclidean space $\mathbf{R}^n$, connectedness is equivalent to arcwise connectedness (see Problem 46).

PROBLEMS

43. Let $\{C_\lambda\}_{\lambda \in \Lambda}$ be a collection of connected subsets of a topological space X and suppose that any two of them have a point in common. Show that the union of $\{C_\lambda\}_{\lambda \in \Lambda}$ also is connected.
44. Let A be a connected subset of a topological space X, and suppose that $A \subseteq B \subseteq \overline{A}$. Show that B is connected.
45. Show that the following subset of the plane is connected but not arcwise connected.
$$X = \{(x, y) \mid \; x = 0, \; -1 \leq y \leq 1\} \cup \{(x, y) \mid \; y = \sin 1/x, \; 0 < x \leq 1\}.$$
46. Show that an arcwise connected topological space X is connected. Also show that each connected open subset $\mathcal{O}$ of a Euclidean space $\mathbf{R}^n$ is arcwise connected. (Suggestion: Let x belong to $\mathcal{O}$. Define C to be the set of points in $\mathcal{O}$ that can be connected in $\mathcal{O}$ to x by a piecewise linear arc. Show that C is both open and closed in $\mathcal{O}$.)
47. Consider the circle $C = \{(x, y) \,|\, x^2 + y^2 = 1\}$ in the plane $\mathbf{R}^2$. Show that C is connected.
48. Show that $\mathbf{R}^n$ is connected.
49. Show that a compact metric space (X, ρ) fails to be connected if and only if there are two disjoint, non-empty subsets A and B whose union is X and $\epsilon > 0$ such that $\rho(u, v) \geq \epsilon$ for all $u \in A, v \in B$. Show that this is not necessarily the case for non-compact metric spaces.

50. (i) Show that a closed, bounded interval is not homeomorphic to an open bounded interval.

    (ii) Show that $[0, 1)$ is not homeomorphic to $(0, 1)$.

51. Show that for any point (x, y) in the plane $\mathbf{R}^2$, the subspace $\mathbf{R}^2 \sim \{(x, y)\}$ is connected. Use this to show that $\mathbf{R}$ is not homeomorphic to $\mathbf{R}^2$.

52. Verify the equivalence of the three assertions in (4).

CHAPTER 16

Topological Spaces: Three Fundamental Theorems

Contents

16.1 Urysohn's Lemma and the Tietze Extension Theorem 323
16.2 The Tychonoff Product Theorem 327
16.3 The Stone-Weierstrass Theorem 331

In the preceding chapter, we considered several different topological concepts and examined relationships between these concepts. In this chapter, we focus on three theorems in topology which, beyond their intrinsic interest, are indispensable tools in several areas of analysis.

16.1 URYSOHN'S LEMMA AND THE TIETZE EXTENSION THEOREM

In Chapter 13, we defined for a closed subset F of a metric space (X, σ), the distance function $\text{dist}_F\colon X \to [0, \infty)$, which is continuous and only vanishes on F. If A and B are disjoint closed subsets of X, there is a continuous real-valued function f on X for which

$$f(X) \subseteq [0,1], f = 0 \text{ on } A \text{ and } f = 1 \text{ on } B.$$

The function f is given by

$$f = \frac{\text{dist}_A}{\text{dist}_A + \text{dist}_B} \text{ on } X.$$

This explicit construction of f depends on the metric on X. However, there are such functions on any normal topological space and, in particular, on any compact, Hausdorff space.

Urysohn's Lemma *Let A and B be two non-empty, disjoint, closed subsets of a normal topological space X. Then for any closed, bounded interval of real numbers $[a, b]$, there is a continuous function $f\colon X \to [a, b]$ for which $f = a$ on A and $f = b$ on B.*

This lemma may be considered to be an extension result: Indeed, define the real-valued function f on $A \cup B$ by setting $f = a$ on A and $f = b$ on B. This is a continuous function on the closed subset $A \cup B$ of X which takes values in $[a, b]$. Urysohn's Lemma asserts that this function can be extended to a continuous function on all of X which also takes values in $[a, b]$. We note that if a Tychonoff topological space X possesses the property described in Urysohn's Lemma, then X must be normal. Indeed, for A and B non-empty, disjoint closed, subsets of X and a continuous real-valued function f on X that takes the value 0 on A and 1 on B, if I_1 and I_2 are disjoint open intervals containing 0 and 1, respectively, then $f^{-1}(I_1)$ and $f^{-1}(I_2)$ are disjoint neighborhoods of A and B, respectively.

The proof of Urysohn's Lemma becomes clearer if we introduce the following concept and then establish two preliminary results.

Definition *Let X be a topological space and Λ a set of real numbers. A collection of open subsets of X indexed by Λ, $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$, is said to be* **normally ascending** *provided that for any $\lambda_1, \lambda_2 \in \Lambda$,*

$$\overline{\mathcal{O}}_{\lambda_1} \subseteq \mathcal{O}_{\lambda_2} \text{ if } \lambda_1 < \lambda_2.$$

Example Let f be a continuous real-valued function on the topological space X. Let Λ be any set of real numbers and define, for $\lambda \in \Lambda$,

$$\mathcal{O}_\lambda = \{x \in X \mid f(x) < \lambda\}.$$

By continuity it is clear that if $\lambda_1 < \lambda_2$, then

$$\overline{\mathcal{O}}_{\lambda_1} \subseteq \{x \in X \mid f(x) \leq \lambda_1\} \subseteq \{x \in X \mid f(x) < \lambda_2\} = \mathcal{O}_{\lambda_2}$$

and therefore the collection of open sets $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$ is normally ascending.

We leave the proof of the following lemma as an exercise.

Lemma 1 *Let X be a topological space. For Λ a dense subset of the open, bounded interval of real numbers (a, b), let $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$ be a normally ascending collection of open subsets of X. Define the function $f\colon X \to \mathbf{R}$ by setting $f = b$ on $X \sim \bigcup_{\lambda\in\Lambda} \mathcal{O}_\lambda$ and otherwise setting*

$$f(x) = \inf\{\lambda \in \Lambda \mid x \in \mathcal{O}_\lambda\}. \tag{1}$$

Then $f\colon X \to [a, b]$ is continuous.

We next provide a strong generalization of Proposition 8 of the preceding chapter.

Lemma 2 *Let X be a normal topological space, F a closed subset of X, and $\mathcal{U}$ a neighborhood of F. Then for any open, bounded interval (a, b), there is a dense subset Λ of (a, b) and a normally ascending collection of open subsets of X, $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda}$, for which*

$$F \subseteq \mathcal{O}_\lambda \subseteq \overline{\mathcal{O}}_\lambda \subseteq \mathcal{U} \text{ for all } \lambda \in \Lambda. \tag{2}$$

Proof Since there is a strictly increasing continuous function of $(0, 1)$ onto (a, b) we may assume that $(a, b) = (0, 1)$. For the dense subset of $(0, 1)$ we choose the set of dyadic rationals belonging to $(0, 1)$:

$$\Lambda = \{m/2^n \mid m \text{ and } n \text{ natural numbers, } 1 \leq m \leq 2^n - 1\}.$$

For each natural number n, let Λ_n be the subset of Λ whose elements have denominator 2^n. We will inductively define a sequence of collections of normally ascending open sets $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda_n}$, where each collection is an extension of its predecessor.

By Proposition 8 of the preceding chapter, we may choose an open set $\mathcal{O}_{1/2}$ for which

$$F \subseteq \mathcal{O}_{1/2} \subseteq \overline{\mathcal{O}_{1/2}} \subseteq \mathcal{U}.$$

Thus, we have defined $\{\mathcal{O}_\lambda\}_{\lambda\in\Lambda_1}$. Now we use Proposition 8 twice more, first with F the same and $\mathcal{U} = \mathcal{O}_{1/2}$ and then with $F = \overline{\mathcal{O}}_{1/2}$ and $\mathcal{U}$ the same, to find open sets $\mathcal{O}_{1/4}$ and $\mathcal{O}_{3/4}$ for which

$$F \subseteq \mathcal{O}_{1/4} \subseteq \overline{\mathcal{O}}_{1/4} \subseteq \mathcal{O}_{1/2} \subseteq \overline{\mathcal{O}}_{1/2} \subseteq \mathcal{O}_{3/4} \subseteq \overline{\mathcal{O}}_{3/4} \subseteq \mathcal{U}.$$

Thus, we have extended the normally ascending collection $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda_1}$ to the normally ascending collection $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda_2}$. It is now clear how to proceed inductively to define for each natural number n, the normally ascending collection of open sets $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda_n}$. Observe that the union of this countable collection is a normally ascending collection of open sets parametrized by Λ, each of which is a neighborhood of F that has closure contained in $\mathcal{U}$. $\square$

Proof of Urysohn's Lemma By Lemma 2, applied with $F = A$ and $\mathcal{U} = X \sim B$, we can choose a dense subset Λ of (a, b) and a normally ascending collection of open subsets of X, $\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda}$, for which

$$A \subseteq \mathcal{O}_\lambda \subseteq X \sim B \text{ for all } \lambda \in \Lambda.$$

Define the function $f \colon X \to [a, b]$ by setting $f = b$ on $X \sim \bigcup_{\lambda \in \Lambda} \mathcal{O}_\lambda$ and otherwise setting

$$f(x) = \inf \{\lambda \in \Lambda \mid x \in \mathcal{O}_\lambda\}.$$

Then $f = a$ on A and $f = b$ on B. According to Lemma 1, f is continuous. $\square$

We mentioned above that Urysohn's Lemma may be considered to be an extension result. We now use this lemma to prove a much stronger extension theorem. In formulating Lusin's Theorem, first, in Chapter 3, we proved this theorem in the case $X = R$, and then, in Chapter 11, we used it to prove Luzin's Theorem in Euclidean space.

The Tietze Extension Theorem *If F is a closed subset of a normal topological space X, then each continuous function $f \colon F \to [a, b]$ has a continuous extension to $f \colon X \to [a, b]$.*

Proof Since the closed, bounded intervals $[a, b]$ and $[-1, 1]$ are homeomorphic, it is sufficient to consider the case $[a, b] = [-1, 1]$. We proceed by constructing a sequence $\{g_n\}$ of continuous real-valued functions on X that has the following two properties: for each index n,

$$|g_n| \leq (2/3)^n \text{ on } X \tag{3}$$

and

$$|f - [g_1 + \cdots + g_n]| \leq (2/3)^n \text{ on } F. \tag{4}$$

Indeed, suppose, for the moment, that this sequence of functions has been constructed. Define, for each index n, the real-valued function s_n on X by

$$s_n(x) = \sum_{k=1}^{n} g_n(x) \text{ for } x \text{ in } X.$$

It follows from the estimate (3) that, for each x in X, $\{s_n(x)\}$ is a Cauchy sequence of real numbers. Since $\mathbf{R}$ is complete, this sequence converges. Define

$$g(x) = \lim_{n \to \infty} s_n(x) \text{ for } x \text{ in } X.$$

Since each g_n is continuous on X, so is each s_n. It follows from the estimate (3) that $\{s_n\}$ converges to g uniformly on X, and therefore g is continuous. From the estimate (4) it is clear that $f = g$ on F. Thus, the theorem is proved provided we construct the sequence $\{g_n\}$. We do so by induction.

Claim: For each $a > 0$ and continuous function $h\colon F \to \mathbf{R}$ for which $|h| \leq a$ on F, there is a continuous function $g\colon X \to \mathbf{R}$ such that

$$|g| \leq (2/3)a \text{ on } X \text{ and } |h - g| \leq (2/3)a \text{ on } F. \tag{5}$$

Indeed, define

$$A = \{x \text{ in } F \mid h(x) \leq -(1/3)a\} \text{ and } B = \{x \text{ in } F \mid h(x) \geq (1/3)a\}.$$

Since h is continuous on F and F is a closed subset of X, A and B are disjoint closed subsets of X. Therefore, by Urysohn's Lemma, there is a continuous real-valued function g on X for which

$$|g| \leq (1/3)a \text{ on } X,\ g(X) = -a/3 \text{ for all } X \text{ in } A,\ g(X) = a/3 \text{ for all } X \text{ in } B.$$

It is clear that (5) holds for this choice of g. Apply the above approximation claim with $h = f$ and $a = 1$ to find a continuous function $g_1\colon X \to \mathbf{R}$ for which

$$|g_1| \leq (2/3) \text{ on } X \text{ and } |f - g_1| \leq (2/3) \text{ on } F.$$

Now apply the claim once more with $h = f - g_1$ and $a = 2/3$ to find a continuous function $g_2\colon X \to \mathbf{R}$ for which

$$|g_2| \leq (2/3)^2 \text{ on } X \text{ and } |f - [g_1 + g_2]| \leq (2/3)^2 \text{ on } F.$$

It is now clear how to proceed to inductively choose the sequence $\{g_n\}$ which possesses properties (3) and (4). □

The Urysohn Metrization Theorem *A topological space that has a countable base is metrizable if and only if it is normal.*

Proof We have already shown that a metric space is normal. Now let X be normal and have a countable base. Choose a countable base $\{\mathcal{U}_n\}_{n \in \mathbf{N}}$ for the topology. Let A be the subset of the product $\mathbf{N} \times \mathbf{N}$ defined by

$$A = \left\{(n, m) \text{ in } \mathbf{N} \times \mathbf{N} \;\middle|\; \overline{\mathcal{U}}_n \subseteq \mathcal{U}_m\right\}.$$

Since X is normal, according to Urysohn's Lemma, for each pair (n, m) in A, there is a continuous real-valued function $f_{n,m}\colon X \to [0, 1]$ for which

$$f_{n,m} = 0 \text{ on } \overline{\mathcal{U}}_n \text{ and } f_{n,m} = 1 \text{ on } X \sim \mathcal{U}_m.$$

For x, y in X, define

$$\rho(x, y) = \sum_{(n,\, m) \in A} \frac{1}{2^{n+m}} |f_{n,m}(x) - f_{n,m}(y)|. \tag{6}$$

The set A is countable so this sum converges. It is not difficult to see that this is a metric. We claim that the topology induced by ρ is the given topology on X. We do so by the following comparison of two bases for the topology. For each $x \in X$:

(i) If $\mathcal{U}_n$ contains x, then there is an $\epsilon > 0$ for which $B_\rho(x, \epsilon) \subseteq \mathcal{U}_n$.
(ii) For each $\epsilon > 0$, there is a $\mathcal{U}_n$ that contains x and $\mathcal{U}_n \subseteq B_\rho(x, \epsilon)$.

We leave the verification of these assertions as an exercise. □

PROBLEMS

1. Let C be a closed subset of a metric space (X, ρ). Show that the distance to C function dist_C is continuous and $\mathrm{dist}_C(x) = 0$ if and only if x belongs to C.
2. Provide an example of a continuous real-valued function on the open interval $(0, 1)$ that is not extendable to a continuous function on $\mathbf{R}$. Does this contradict the Tietze Extension Theorem?
3. Deduce Urysohn's Lemma as a consequence of the Tietze Extension Theorem.
4. State and prove a version of the Tietze Extension Theorem for functions with values in $\mathbf{R}^n$.
5. Suppose that a topological space X has the property that every continuous, bounded real-valued function on a closed subset has a continuous extension to all of X. Show that if X is Tychonoff, then it is normal.
6. Let $(X, \mathcal{T})$ be a normal topological space and $\mathcal{F}$ the collection of continuous real-valued functions on X. Show that $\mathcal{T}$ is the weak topology induced by $\mathcal{F}$.
7. Show that the function ρ defined in the proof of the Urysohn Metrization Theorem is a metric that defines the same topology as the given topology.
8. Let X be a normal topological space, F a closed subset of X, and f a continuous not necessarily bounded, real-valued function on F. Then f has a continuous extension to a real-valued function $\overline{f}$ on all of X. Prove this as follows:
   (i) Apply the Tietze Extension Theorem to obtain a continuous extension $h: X \to [-1, 1]$ of the function $f \cdot (1 + |f|)^{-1}: F \to [-1, 1]$;
   (ii) Once more, apply the Tietze Extension Theorem to obtain a function $\phi: X \to [0, 1]$ such that $\phi = 1$ on F and $\phi = 0$ on $h^{-1}(1)$ and $h^{-1}(-1)$;
   (iii) Consider the function $\overline{f} = \phi \cdot h/(1 - \phi \cdot |h|)$.
9. Show that a mapping f from a topological space X to a topological space Y is continuous if and only if there is a subbase $\mathcal{S}$ for the topology on Y such that the preimage under f of each set in $\mathcal{S}$ is open in X. Use this to show that if Y is a closed, bounded interval $[a, b]$, then f is continuous if and only if for each real number $c \in (a, b)$, the sets $\{x \in X \mid f(x) < c\}$ and $\{x \in X \mid f(x) > c\}$ are open.
10. Use the preceding problem to prove Lemma 1.

16.2 THE TYCHONOFF PRODUCT THEOREM

For a collection of sets indexed by a set Λ, $\{X_\lambda\}_{\lambda \in \Lambda}$, we defined the Cartesian product $\Pi_{\lambda \in \Lambda} X_\lambda$ to be the collection of mappings from the index set Λ to the union $\bigcup_{\lambda \in \Lambda} X_\lambda$ such that each index $\lambda \in \Lambda$ is mapped to a member of X_λ. For a member x of the Cartesian product and an index $\lambda \in \Lambda$, it is customary to denote $x(\lambda)$ by x_λ and call x_λ the λ-th component of x. For each parameter $\lambda_0 \in \Lambda$, we define the λ_0 projection mapping $\pi_{\lambda_0}: \Pi_{\lambda \in \Lambda} X_\lambda \to X_{\lambda_0}$ by

$$\pi_{\lambda_0}(x) = x_{\lambda_0} \text{ for } x \in \Pi_{\lambda \in \Lambda} X_\lambda.$$

According to Proposition 7 of Chapter 13, there is metric on the Cartesian product of a countable collection of metric spaces, with respect to which sequential convergence is componentwise convergence. There is a natural definition of a topology on the Cartesian product of a finite collection of topological spaces. Given a collection $\{(X_k, \mathcal{T}_k)\}_{k=1}^n$ of topological spaces, the collection of products

$$\mathcal{O}_1 \times \cdots \mathcal{O}_k \cdots \times \mathcal{O}_n,$$

where each $\mathcal{O}_k$ belongs to $\mathcal{T}_k$, is a base for a topology on $\Pi_{1\leq k\leq n}X_k$. The topology on the Cartesian product consisting of unions of these basic sets is called the **product topology** on $\Pi_{1\leq k\leq n}X_k$. What is novel for topological spaces is that a product topology can be defined on an arbitrary Cartesian product $\Pi_{\lambda\in\Lambda}X_\lambda$ of topological spaces. The index set is not required to be finite or even countable.

Definition *Let $\{(X_\lambda, \mathcal{T}_\lambda)\}_{\lambda\in\Lambda}$ be a collection of topological spaces indexed by a set Λ. The* **product topology** *on the Cartesian product $\Pi_{\lambda\in\Lambda}X_\lambda$ is the topology that has as a base sets of the form $\Pi_{\lambda\in\Lambda}\mathcal{O}_\lambda$, where each $\mathcal{O}_\lambda \in \mathcal{T}_\lambda$ and $\mathcal{O}_\lambda = X_\lambda$, except for finitely many λ.*

The product topology on the Cartesian product of topological spaces $\Pi_{\lambda\in\Lambda}X_\lambda$ is the weak topology associated to the collection of projections $\{\pi_\lambda\colon \Pi_{\lambda\in\Lambda}X_\lambda \to X_\lambda\}_{\lambda\in\Lambda}$, that is, the product topology is the weakest topology with respect to which all the projections are continuous.

If all the X_λ's are the same space X, it is customary to denote $\Pi_{\lambda\in\Lambda}X_\lambda$ by X^Λ. In particular, if $\mathbf{N}$ denotes the set on natural numbers, then $X^{\mathbf{N}}$ is the collection of sequences in X while $\mathbf{R}^X$ is the collection of real-valued functions that have domain X. If X is a metric space and Λ is countable, then the product topology on X^Λ is induced by a metric (see Problem 16). In general, if X is a metric space but Λ is uncountable, the product topology is not induced by a metric. For example, the product topology on $\mathbf{R}^{\mathbf{R}}$ is not induced by a metric (see Problem 17). It is left as an exercise to verify the following proposition.

Proposition 3 *Let X be a topological space. A sequence $\{f_n\colon \Lambda \to X\}$ converges to f in the product space X^Λ if and only if $\{f_n(\lambda)\}$ converges to $f(\lambda)$ for each λ in Λ. Thus, convergence of a sequence with respect to the product topology is pointwise convergence.*

The centerpiece of this section is the Tychonoff Product Theorem, according to which the product $\Pi_{\lambda\in\Lambda}X_\lambda$ of compact topological spaces is compact. There are no restrictions on the index space Λ. In preparation for the proof of this theorem, two lemmas regarding collections of sets that possess the finite intersection property are established.

Lemma 4 *Let $\mathcal{A}$ be a collection of subsets of a set X that possesses the finite intersection property. Then there is a collection $\mathcal{B}$ of subsets of X which contains $\mathcal{A}$, has the finite intersection property, and is maximal with respect to this property; that is, no collection of subsets of X that properly contains $\mathcal{B}$ possesses the finite intersection property.*

Proof Consider the family $\mathcal{F}$ of all collections of subsets of X containing $\mathcal{A}$ and possessing the finite intersection property. Order $\mathcal{F}$ by inclusion. Every linearly ordered subfamily $\mathcal{F}_0$ of $\mathcal{F}$ has an upper bound consisting of the sets belonging to any collection in $\mathcal{F}_0$. According to Zorn's Lemma, there is maximal member of $\mathcal{F}$. This maximal member is a collection of sets that has the properties described in the conclusion of the lemma. $\square$

Lemma 5 *Let $\mathcal{B}$ be a collection of subsets of a set X of a set that is maximal with respect to the finite intersection property. Then each intersection of a finite number of sets in $\mathcal{B}$ is again in $\mathcal{B}$, and each subset of X that has non-empty intersection with each set in $\mathcal{B}$ is itself in $\mathcal{B}$.*

Proof Let $\mathcal{B}'$ be the collection of all sets that are finite intersections of sets in $\mathcal{B}$. Then $\mathcal{B}'$ is a collection having the finite intersection property and containing $\mathcal{B}$. Thus $\mathcal{B}' = \mathcal{B}$ by the maximality, with respect to inclusion, of $\mathcal{B}$. Now suppose that C is a subset of X that has non-empty intersection with each member of $\mathcal{B}$. Since $\mathcal{B}$ contains each finite intersection of sets in $\mathcal{B}$, it follows that $\mathcal{B} \cup \{C\}$ has the finite intersection property. By the maximality, with respect to inclusion, of $\mathcal{B}$, $\mathcal{B} \cup \{C\} = \mathcal{B}$, and so C is a member of $\mathcal{B}$. $\square$

The Tychonoff Product Theorem *Let $\{X_\lambda\}_{\lambda \in \Lambda}$ be a collection of compact topological spaces indexed by a set Λ. Then the Cartesian product $\Pi_{\lambda \in \Lambda} X_\lambda$, with the product topology, also is compact.*

Proof Let $\mathcal{F}$ be a collection of closed subsets of $X = \Pi_{\lambda \in \Lambda} X_\lambda$ possessing the finite intersection property. It must be shown that $\mathcal{F}$ has non-empty intersection. By Lemma 4, there is a collection $\mathcal{B}$ of subsets of X that contains $\mathcal{F}$ and is maximal with respect to the finite intersection property. Fix $\lambda \in \Lambda$. Define

$$\mathcal{B}_\lambda = \{\pi_\lambda(B) \mid B \in \mathcal{B}\}.$$

Then $\mathcal{B}_\lambda$ is a collection of subsets of the set X_λ that has the finite intersection property, as does the collection of closures of members of $\mathcal{B}_\lambda$. By the compactness of X_λ there is a point $x_\lambda \in X_\lambda$ for which

$$x_\lambda \in \bigcap_{B \in \mathcal{B}} \overline{\pi_\lambda(B)}.$$

Define x to be the point in X whose λ-th coordinate is x_λ. We claim that

$$x \in \bigcap_{F \in \mathcal{F}} F. \tag{7}$$

Indeed, the point x has the property that for each index λ, x_λ is a point of closure of $\pi_\lambda(B)$ for every $B \in \mathcal{B}$. Thus

$$\text{every subbasic neighborhood } \mathcal{N}_x \text{ of } x \text{ has non-empty intersection with every set } B \text{ in } \mathcal{B}. \tag{8}$$

From the maximality of $\mathcal{B}$ and Lemma 5, it follows that every subbasic neighborhood of x belongs to $\mathcal{B}$. Once more using Lemma 5, we conclude that every basic neighborhood of x belongs to $\mathcal{B}$. But $\mathcal{B}$ has the finite intersection property and contains the collection $\mathcal{F}$. Let F be a set in $\mathcal{F}$. Then every basic neighborhood of x has non-empty intersection with F. Hence x is a point of closure of the closed set F, so that x belongs to F. Thus (7) holds. $\square$

PROBLEMS

11. Show that the product of an arbitrary collection of Tychonoff spaces, with the product topology, also is Tychonoff.
12. Show that the product of an arbitrary collection of Hausdorff spaces, with the product topology, also is Hausdorff.
13. Show that the product topology on n-fold product of $\mathbf{R}$ is the same as the metric topology on $\mathbf{R}^n$ induced by the Euclidean metric.

14. Let (X, ρ_1) and (Y, ρ_2) be metric spaces. Show that the product topology on $X \times Y$, where X and X have the topologies induced by their respective metrics, is the same as the topology induced by the product metric

$$\rho((x_1, y_1), (x_2, y_2)) = \sqrt{[\rho_1(x_1, x_2)]^2 + [\rho_2(y_1, y_2)]^2}.$$

15. Show that if X is a metric space with metric ρ, then

$$\rho^*(x, y) = \frac{\rho(x, y)}{1 + \rho(x, y)}$$

also is a metric on X and it induces the same topology as the metric ρ.

16. Consider the countable collection of metric spaces $\{(X_n, \rho_n)\}_{n=1}^{\infty}$. For the Cartesian product of these sets $X = \prod_{n=1}^{\infty} X_n$, define $\rho\colon X \times X \to \mathbf{R}$ by

$$\rho(x, y) = \sum_{n=1}^{\infty} \frac{\rho_n(x_n, y_n)}{2^n[1 + \rho_n(x_n, y_n)]}.$$

Use the preceding problem to show that ρ is a metric on $X = \prod_{n=1}^{\infty} X_n$ which induces the product topology on X, where each X_n has the topology induced by the metric ρ_n.

17. Consider the set $X = \mathbf{R}^{\mathbf{R}}$ with the product topology. Let E be the subset of X consisting of functions that take the value 0 on a countable set and elsewhere take the value 1. Let f_0 be the function that is identically zero. Then it is clear that f_0 is a point of closure of E. But there is no sequence $\{f_n\}$ in E that converges to f_0, since for any sequence $\{f_n\}$ in E there is some $x_0 \in \mathbf{R}$ such that $f_n(x_0) = 1$ for all n and so the sequence $\{f_n(x_0)\}$ does not converge to $f_0(x_0)$. This shows, in particular, that $X = \mathbf{R}^{\mathbf{R}}$ is not first countable and therefore not metrizable.

18. Let X denote the discrete topological space with two elements. Show that X^N is homeomorphic to the Cantor set.

19. Using the Tychonoff Product Theorem and the compactness of each closed, bounded interval of real numbers prove that any closed, bounded subset of $\mathbf{R}^n$ is compact.

20. Provide a direct proof of the assertion that if X is compact and I is a closed, bounded interval, then $X \times I$ is compact. (Suggestion: Let $\mathcal{U}$ be an open covering of $X \times I$, and consider the smallest value of $t \in I$ such that for each $t' < t$ the set $X \times [0, t']$ can be covered by a finite number of sets in $\mathcal{U}$. Use the compactness of X to show that $X \times [0, t]$ can also be covered by a finite number of sets in $\mathcal{U}$ and that if $t < 1$, then for some $t'' > t$, $X \times [0, t'']$ can be covered by a finite number of sets in $\mathcal{U}$.)

21. Prove that the product of a countable collection of sequentially compact topological spaces is sequentially compact.

22. A product I^A of unit intervals is called a (generalized) cube. Prove that every compact, Hausdorff space X is homeomorphic to a closed subset of some cube. (Let $\mathcal{F}$ be the family of continuous real-valued functions on X with values in $[0, 1]$. Let $Q = \Pi_{f \in \mathcal{F}} I_f$. Then, since X is normal and compact, the mapping g of X into Q that takes x into the point whose f-th coordinate is $f(x)$ is one-to-one, continuous, and has closed image.)

23. Let $Q = I^A$ be a cube, and let f be a continuous real-valued function on Q. Then, given $\epsilon > 0$, there is a continuous real-valued function g on Q for which $|f - g| < \epsilon$ and g is a function of only a finite number of coordinates. (Suggestion: Cover the range of f by a finite number of intervals of length ϵ and look at the inverse images of these intervals.)

16.3 THE STONE-WEIERSTRASS THEOREM

The following theorem is one of the jewels of classical analysis[1].

The Weierstrass Approximation Theorem *Let $f\colon [a, b] \to \mathbf{R}$ be continuous. Then for each $\epsilon > 0$, there is a polynomial p for which*

$$|f(x) - p(x)| < \epsilon \text{ for all } x \in [a, b].$$

In this section, a far-reaching extension of this theorem is proven. For a compact, Hausdorff space X, consider the linear space $C(X)$ of continuous real-valued functions on X with the maximum norm. According to the Weierstrass Approximation Theorem, the polynomials are dense in $C[a, b]$. Now $C(X)$ has a product structure not possessed by all linear spaces, namely, the product $f \cdot g$ of two functions f and g in $C(X)$ is again in $C(X)$. A linear subspace $\mathcal{A}$ of $C(X)$ is called an **algebra** provided the product of any two functions in $\mathcal{A}$ also belongs to $\mathcal{A}$. A collection $\mathcal{A}$ of real-valued functions on X is said to **separate points** in X provided for any two distinct points u and v in X, there is an f in $\mathcal{A}$ for which $f(u) \neq f(v)$. Observe that since X is compact and Hausdorff, according to Theorem 18 of the preceding chapter, it is normal and therefore it follows from Urysohn's Lemma that the whole algebra $C(X)$ separates points in X.

The Stone-Weierstrass Approximation Theorem *Let X be a compact, Hausdorff space. Assume that $\mathcal{A}$ is an algebra of continuous real-valued functions on X that separates points in X and contains the constant functions. Then $\mathcal{A}$ is dense in $C(X)$.*

Observe that this is a generalization of the Weierstrass Approximation Theorem since the closed, bounded interval $[a, b]$ is compact and Hausdorff and the collection of polynomials is an algebra that contains the constant functions and separates points.

Before proving this theorem, a few words concerning strategy are in order[2]. Since X is compact and Hausdorff, it is normal. It follows from Urysohn's Lemma that for each pair of disjoint closed subsets A and B of X and $\epsilon \in (0, 1/2)$, there is a function $f \in C(X)$ for which

$$f = \epsilon/2 \text{ on } A, f = 1 - \epsilon/2 \text{ on } B, \text{ and } \epsilon/2 \leq f \leq 1 - \epsilon/2 \text{ on } X.$$

Therefore, if $|h - f| < \epsilon/2$ on X,

$$h < \epsilon \text{ on } A, h > 1 - \epsilon \text{ on } B, \text{ and } 0 \leq h \leq 1 \text{ on } X. \tag{9}$$

The proof proceeds in two steps. First, it is shown that for each pair of disjoint closed subsets A and B of X and $\epsilon \in (0, 1/2)$, there is a function h belonging to the algebra $\mathcal{A}$ for which (9) holds. It is then shown that any function f in $C(X)$ can be uniformly approximated by arithmetic means of such h's.

---

[1] A proof of this theorem, due to Serguei Bernstein, is given in Chapter 8 of Patrick Fitzpatrick's *Advanced Calculus*.

[2] The proof we present is due to B. Brasowski and F. Deutsch, *Proceedings of the American Mathematical Society, 81* (1981). Many very different-looking proofs of the Stone-Weierstrass Theorem have been given since the first proof in 1937 by Marshall Stone.

Lemma 6 *Assume that X and $\mathcal{A}$ satisfy the assumptions of the Stone-Weierstrass Theorem. Then for each closed subset F of X and point x_0 belonging to $X \sim F$, there is a neighborhood $\mathcal{U}$ of x_0 that is disjoint from F and has the following property: for each $\epsilon > 0$, there is a function $h \in \mathcal{A}$ for which*

$$h < \epsilon \text{ on } \mathcal{U},\ h > 1 - \epsilon \text{ on } F, \text{ and } 0 \leq h \leq 1 \text{ on } X. \tag{10}$$

Proof We first claim that for each point $y \in F$, there is a function g_y in $\mathcal{A}$ for which

$$g_y(x_0) = 0,\ g_y(y) > 0, \text{ and } 0 \leq g_y \leq 1 \text{ on } X. \tag{11}$$

Indeed, since $\mathcal{A}$ separates points, there is a function $f_y \in \mathcal{A}$ for which $f(x_0) \neq f(y)$. The function

$$g_y = \left[\frac{f_y - f_y(x_0)}{||f_y - f_y(x_0)||_{\max}}\right]^2$$

belongs to $\mathcal{A}$ and satisfies (11). By the continuity of g_y, there is a neighborhood $\mathcal{N}_y$ of y on which g_y only takes positive values. However, F is a closed subset of the compact space X and therefore F itself is compact. Consequently, there is a finite collection of these neighborhoods $\{\mathcal{N}_{y_1}, \ldots, \mathcal{N}_{y_n}\}$ that covers F. Define the function $g \in \mathcal{A}$ by

$$g = \frac{1}{n}\sum_{i=1}^{n} g_{y_i}.$$

Then

$$g(x_0) = 0,\ g > 0 \text{ on } F, \text{ and } 0 \leq g \leq 1 \text{ on } X. \tag{12}$$

But a continuous function on a compact set takes a minimum value, so there is a $c > 0$ for which $g \geq c$ on F, and by rescaling if necessary it can be assumed that $c < 1$. On the other hand, g is continuous at x_0, so there is a neighborhood $\mathcal{U}$ of x_0 for which $g < c/2$ on U. Thus g belongs to the algebra $\mathcal{A}$ and

$$g < c/2 \text{ on } \mathcal{U},\ g \geq c \text{ on } F, \quad \text{and } 0 \leq g \leq 1 \text{ on } X. \tag{13}$$

We claim that (10) holds for this choice of neighborhood $\mathcal{U}$. Let $\epsilon > 0$. By the Weierstrass Approximation Theorem, there is a polynomial p such that[3]

$$p < \epsilon \text{ on } [0,\, c/2],\ p > 1 - \epsilon \text{ on } [c,\, 1], \quad \text{and } 0 \leq p \leq 1 \text{ on } [0,\, 1]. \tag{14}$$

Since p is a polynomial and g belongs to the algebra $\mathcal{A}$, the composition $h = p \circ g$ also belongs to $\mathcal{A}$. From (13) and (14) it follows that (10) holds. □

Lemma 7 *Assume that X and $\mathcal{A}$ satisfy the assumptions of the Stone-Weierstrass Theorem. Then for each pair of disjoint closed subsets A and B of X and $\epsilon > 0$, there is a function h belonging to $\mathcal{A}$ for which*

$$h < \epsilon \text{ on } A,\ h > 1 - \epsilon \text{ on } B, \text{ and } 0 \leq h \leq 1 \text{ on } X. \tag{15}$$

[3]Rather than using the Weierstrass Approximation Theorem here, one can show that (14) holds for a polynomial of the form $p(x) = 1 - (1 - x^n)^m$, where n and m are suitably chosen natural numbers.

Proof By the preceding lemma in the case $F = B$, for each point $x \in A$, there is a neighborhood $\mathcal{N}_x$ of x that is disjoint from B and has the property (10). However, A is compact since it is a closed subset of the compact space X, and therefore there is a finite collection of neighborhoods $\{\mathcal{N}_{x_1}, \ldots, \mathcal{N}_{x_n}\}$ that covers A. Choose ϵ_0 for which $0 < \epsilon_0 < \epsilon$ and $(1 - \epsilon_0/n)^n > 1 - \epsilon$. For $1 \le i \le n$, since $\mathcal{N}_{x_i}$ has property (10) with $B = F$, there is an $h_i \in \mathcal{A}$ such that

$$h_i < \epsilon_0/n \text{ on } \mathcal{N}_{x_i}, \;\; h_i > 1 - \epsilon_0/n \text{ on } B, \text{ and } 0 \le h_i \le 1 \text{ on } X.$$

Define

$$h = h_1 \cdot h_2 \cdots h_n \text{ on } X.$$

Then $h \in \mathcal{A}$. Since for each i, $0 \le h_i \le 1$ on X, $0 \le h \le 1$ on X. Also, for each i, $h_i > 1 - \epsilon_0/n$ on B, so $h \ge (1 - \epsilon_0/n)^n > 1 - \epsilon$ on B. Finally, for each point x in A there is an index i for which x belongs to $\mathcal{N}_{x_i}$. Thus $h_i(x) < \epsilon_0/n < \epsilon$ and since for the other indices j, $0 \le h_j(x) \le 1$, we conclude that $h(x) < \epsilon$. □

Proof of the Stone-Weierstrass Theorem Let f belong to $C(X)$. Set $c = ||f||_{\max}$. If we can arbitrarily closely uniformly approximate the function

$$\frac{f + c}{||f + c||_{\max}}$$

by functions in $\mathcal{A}$, we can do the same for f. Assume that $0 \le f \le 1$ on X. For each n, consider the uniform partition $\{0, 1/n, 2/n, \ldots, (n-1)/n, 1\}$ of $[0, 1]$ into n intervals, each of length $1/n$. Fix j, $1 \le j \le n$. Define

$$A_j = \{x \text{ in } X \mid \; f(x) \le (j-1)/n\} \text{ and } B_j = \{x \text{ in } X \mid \; f(x) \ge j/n\}.$$

Since f is continuous, both A_j and B_j are closed subsets of X and, of course, they are disjoint. By the preceding lemma, with $A = A_j$, $B = B_j$, and $\epsilon = 1/n$, there is a function g_j in the algebra $\mathcal{A}$ for which

$$g_j(x) < 1/n \text{ if } f(x) \le (j-1)/n, \;\; g_j(x) > 1 - 1/n \text{ if } f(x) \ge j/n \;\text{ and } 0 \le g_j \le 1 \text{ on } X. \tag{16}$$

Define

$$g = \frac{1}{n} \sum_{j=1}^{n} g_j.$$

Then g belongs to $\mathcal{A}$. We claim that

$$||f - g||_{\max} < 3/n. \tag{17}$$

Once we establish this claim the proof is complete since, given $\epsilon > 0$, we simply select n such that $3/n < \epsilon$ and therefore $||f - g||_{\max} < \epsilon$. To verify (17), we first show that

$$\text{if } 1 \le k \le n \text{ and } f(x) \le k/n, \text{ then } g(x) \le k/n + 1/n. \tag{18}$$

Indeed, for $j = k+1, \ldots, n$, since $f(x) \leq k/n$, $f(x) \leq (j-1)/n$ and therefore $g_j(x) \leq 1/n$. Thus

$$\frac{1}{n} \sum_{j=k+1}^{n} g_j \leq (n-k)/n^2 \leq 1/n.$$

Consequently, since each $g_j(x) \leq 1$, for all j,

$$g(x) = \frac{1}{n} \sum_{j=1}^{n} g_j = \frac{1}{n} \sum_{j=1}^{k} g_j + \frac{1}{n} \sum_{j=k+1}^{n} g_j \leq \frac{1}{n} \sum_{j=1}^{k} g_j + 1/n \leq k/n + 1/n.$$

Thus (18) holds. A similar argument shows that

$$\text{if } 1 \leq k \leq n \text{ and } (k-1)/n \leq f(x), \text{ then } (k-1)/n - 1/n \leq g(x). \tag{19}$$

For $x \in X$, choose k, $1 \leq k \leq n$, such that $(k-1)/n \leq f(x) \leq k/n$. It follows from (18) and (19) that $|f(x) - g(x)| < 3/n$. □

We have the following elegant consequence of the Stone-Weierstrass Theorem and Urysohn's Lemma.

Theorem 8 (**Borsuk's Theorem**) *Let X be a compact, Hausdorff space. Then the normed linear space $C(X)$ is separable if and only if the topology on X is metrizable.*

Proof First assume that X is metrizable. Let ρ be a metric that induces the topology on X. Then X, being a compact metric space, is separable. Choose a countable, dense subset $\{x_n\}$ of X. For each natural number n, define $f_n(x) = \rho(x, x_n)$ for all $x \in X$. Since ρ induces the topology, f_n is continuous. Since the set $\{x_n\}$ is dense, it separates points in X. Define $f_0 \equiv 1$ on X. Now let $\mathcal{A}$ be the collection of polynomials, with real coefficients, in a finite number of the $f_k, 0 \leq k < \infty$. Then $\mathcal{A}$ is an algebra that contains the constant functions, and it separates points in X since it contains each f_k. According to the Stone-Weierstrass Theorem, $\mathcal{A}$ is dense in $C(X)$. But the collection of functions f in $\mathcal{A}$ that are polynomials with rational coefficients is a countable set that is dense in $\mathcal{A}$. Therefore, $C(X)$ is separable.

Conversely, suppose that $C(X)$ is separable. Let $\{g_n\}$ be a countable, dense subset of $C(X)$. For each natural number n, define $\mathcal{O}_n = \{x \in X \mid g_n(x) > 1/2\}$. Then $\{\mathcal{O}_n\}_{1 \leq n < \infty}$ is a countable collection of open sets. We claim that this countable collection of open sets is a base for X. Once this claim is verified, since X, being compact and Hausdorff, is normal, it follows that the Urysohn Metrization Theorem that X, also being second countable, is metrizable. To verify the claim let $x \in X$ belong to the open set O. Since X is normal, there is an open set $\mathcal{U}$ for which $x \in \mathcal{U} \subseteq \overline{\mathcal{U}} \subseteq \mathcal{O}$. By Urysohn's Lemma, there is a g in $C(X)$ such that $g(x) = 1$ on $\mathcal{U}$ and $g = 0$ on $X \sim \mathcal{O}$. By the denseness of $\{g_n\}$ in $C(X)$, there is a natural number n for which $|g - g_n| < 1/2$ on X. Therefore, $x \in \mathcal{O}_n \subseteq \mathcal{O}$. This completes the proof. □

PROBLEMS

24. Suppose that X is a topological space for which there is a collection of continuous real-valued functions on X that separates points in X. Show that X is Hausdorff.
25. Let X be a compact, Hausdorff space and $\mathcal{A} \subseteq C(X)$ an algebra that contains the constant functions. Show that $\mathcal{A}$ is dense in $C(X)$ if and only if $\mathcal{A}$ separates points in X.

26. Let $\mathcal{A}$ be an algebra of continuous real-valued functions on a compact space X that contains the constant functions. Let $f \in C(X)$ have the property that for some constant function c and real number α, the function $\alpha(f+c)$ belongs to $\overline{\mathcal{A}}$. Show that f also belongs to $\overline{\mathcal{A}}$.

27. For $f, g \in C[a, b]$, show that $f = g$ if and only if $\int_a^b x^n f(x)\,dx = \int_a^b x^n g(x)\,dx$ for all n.

28. For $f \in C[a, b]$ and $\epsilon > 0$, show that there are real numbers $c_0, c_1, \ldots, c_n$ for which

$$|f(x) - c_0 - \sum_{k=1}^{n} c_k \cdot e^{kx}| < \epsilon \text{ for all } x \in [a, b].$$

29. For $f \in C[0, \pi]$ and $\epsilon > 0$, show that there are real numbers $c_0, c_1, \ldots, c_n$ for which

$$|f(x) - c_0 - \sum_{k=1}^{n} c_k \cdot \cos kx| < \epsilon \text{ for all } x \in [0, \pi].$$

30. Let f be a continuous real-valued function on $\mathbf{R}$ that is periodic with period 2π. For $\epsilon > 0$, show that there are real numbers, $c_0, a_1, \ldots, a_n, b_1, \ldots, b_n$, such that

$$|f(x) - c_0 - \sum_{k=1}^{n} [a_k \cos kx + b_k \sin kx]| < \epsilon \text{ for all } x \in \mathbf{R}.$$

(Suggestion: A periodic function may be identified with a continuous function on the unit circle in the plane and the unit circle is compact and Hausdorff with the topology it inherits from the plane.)

31. Let X and Y be compact, Hausdorff spaces and f belong to $C(X \times Y)$. Show that for each $\epsilon > 0$, there are functions $f_1, \ldots, f_n$ in $C(X)$ and $g_1, \ldots, g_n$ in $C(Y)$ such that

$$|f(x, y) - \sum_{k=1}^{n} f_k(x) \cdot g_k(y)| < \epsilon \text{ for all } (x, y) \in X \times Y.$$

32. Rather than use the Weierstrass Approximation Theorem in the proof of the Stone-Weierstrass Theorem, show that there are natural numbers m and n for which the polynomial $p(x) = 1 - (1 - x^n)^m$ satisfies (14). (Suggestion: Since $p(0) = 0, p(1) = 1$ and $p' > 0$ on $(0, 1)$, it suffices to choose m and n such that $p(c/2) < \epsilon$ and $p(c) > 1 - \epsilon$.)

33. Let $\mathcal{A}$ be a collection of continuous real-valued functions on a compact, Hausdorff space X that separates the points of X. Show that every continuous real-valued function on X can be uniformly approximated arbitrarily closely by a polynomial in a finite number of functions of $\mathcal{A}$.

34. Let $\mathcal{A}$ be an algebra of continuous real-valued functions on a compact, Hausdorff space X. Show that the closure of $\mathcal{A}$, $\overline{\mathcal{A}}$, also is an algebra.

35. Let $\mathcal{A}$ be an algebra of continuous real-valued functions on a compact, Hausdorff space X that separates points. Show that either $\overline{\mathcal{A}} = C(X)$ or there is a point $x_0 \in X$ for which $\overline{\mathcal{A}} = \{f \in C(X) \mid f(x_0) = 0\}$. (Suggestion: If $1 \in \overline{\mathcal{A}}$, we are done. Moreover, if for each $x \in X$ there is an $f \in \mathcal{A}$ with $f(x) \neq 0$, then there is a $g \in \mathcal{A}$ that is positive on X and this implies that $1 \in \overline{\mathcal{A}}$.)

36. Let X be a compact, Hausdorff space and $\mathcal{A}$ an algebra of continuous functions on X that separates points and contains the constant functions.
    (i) Given any two numbers a and b and points $u, v \in X$, show that there is a function f in $\mathcal{A}$ for which $f(u) = a$ and $f(v) = b$.
    (ii) Is it the case that given any two numbers a and b and disjoint closed subsets A and B of X, there is a function f in $\mathcal{A}$ for which $f = a$ on A and $f = b$ on B?

CHAPTER 17

Continuous Linear Operators Between Banach Spaces

Contents

17.1 Normed Linear Spaces 337
17.2 Linear Operators 340
17.3 Compactness Lost: Infinite Dimensional Normed Linear Spaces 343
17.4 The Open Mapping and Closed Graph Theorems 347
17.5 The Uniform Boundedness Principle 353

We have already examined important examples of normed linear spaces. The most prominent of these are (i) for a natural number n, Euclidean space $\mathbf{R}^n$; (ii) for a general measure space $(X, \mathcal{M}, \mu)$ and $1 \leq p \leq \infty$, the space $L^p(X, \mu)$; (iii) for K a compact topological space, the linear space $C(K)$ of continuous real-valued functions on K, normed by the maximum norm. In this and the following three chapters, general normed linear spaces and the continuous linear operators between such spaces are considered. The results established in the preceding four chapters for metric and topological spaces are the basic tools.

17.1 NORMED LINEAR SPACES

A linear space X is an abelian group with the group operation of addition denoted by $+$, for which, given a real number α and $u \in X$, there is defined the scalar product $\alpha \cdot u \in X$ for which the following three properties hold: for real numbers α and β and members u and v in X,

$$(\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot u,$$

$$\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v,$$

$$(\alpha\beta) \cdot u = \alpha \cdot (\beta \cdot u) \text{ and } 1 \cdot u = u.$$

A linear space is also called a vector space and, paying respect to $\mathbf{R}^n$, members of a linear space are often called vectors. The quintessential example of a linear space is the collection of real-valued functions on an arbitrary non-empty set D where, for two functions $f, g\colon D \to \mathbf{R}$ and real number λ, addition $f + g$ and scalar multiplication $\lambda \cdot f$ are defined pointwise on D by

$$(f + g)(x) = f(x) + g(x) \text{ and } (\lambda \cdot f)(x) = \lambda f(x) \text{ for all } x \in D.$$

Recall the concept of norm on a linear space X, which we first studied in Chapter 7; a non-negative real-valued function $\|\cdot\|$ defined on a linear space X is called a **norm** provided that for all $u, v \in X$ and $\alpha \in \mathbf{R}$:

$$\|u\| = 0 \text{ if and only if } u = 0;$$

$$\|u + v\| \leq \|u\| + \|v\|;$$

$$\|\alpha u\| = |\alpha|\|u\|.$$

As we observed in Chapter 13, a norm on a linear space induces a metric on the space, where the distance between u and v is defined to be $\|u - v\|$. When we refer to metric properties of a normed space, such as boundedness and completeness, we mean with respect to the metric induced by the norm. Similarly, when we refer to topological properties, such as a sequence converging or a set being open, closed, or compact, we are referring to the topology induced by the above metric[1].

Definition *Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a linear space X are said to be* **equivalent** *provided that there are constants $c_1 \geq 0$ and $c_2 \geq 0$ for which*

$$c_1 \cdot \|x\|_1 \leq \|x\|_2 \leq c_2 \cdot \|x\|_1 \textit{ for all } x \in X.$$

If a norm on a linear space is replaced by an equivalent norm, the topological and metric properties remain unchanged.

Concepts from linear algebra in finite dimensional spaces are also important for general linear spaces[2]. Given vectors $x_1, \ldots, x_n$ in a linear space X and real numbers $\lambda_1, \ldots, \lambda_n$, the vector

$$x = \sum_{k=1}^{n} \lambda_k x_k$$

is called a **linear combination** of the x_i's. A non-empty subset Y of X is called a **linear subspace**, or simply a subspace, provided that every linear combination of vectors in Y also belongs to Y.

For a non-empty subset S of X, the **span** of S is defined to be the set of all linear combinations of vectors in S: it is denoted by span$[S]$. It is left as an exercise to show that span$[S]$ is a linear subspace of X, which is the smallest subspace of X that contains S, in the sense that it is contained in any linear subspace that contains S. If $Y = \text{span}[S]$ we say that S spans Y. It will also be useful to consider the closure of the span of S, which we denote by $\overline{\text{span}}[S]$. It is left as another exercise to show that the closure of a linear subspace of X is a linear subspace. Therefore $\overline{\text{span}}[S]$ is a linear subspace of X which is the smallest closed linear subspace of X that contains S, in the sense that it is contained in any closed linear subspace that contains S. The space $\overline{\text{span}}[S]$ is called the **closed linear span** of S.

[1] In the following chapters, we consider topologies on a normed linear space X other than that induced by the norm and are explicit when we refer to topological properties with respect to these other topologies.

[2] We later refer to a few results from linear algebra but, this chapter require nothing more than knowing that any two bases of a finite dimensional linear space have the same number of vectors, so dimension is properly defined, and that any linearly independent set of vectors in a finite dimensional linear space is a subset of a basis: see Peter Lax's *Linear Algebra* [Lax97].

For any two non-empty subsets A and B of a linear space X, the **sum** of A with B, written $A+B$, is defined by

$$A+B=\{x+y \mid \; x\in A, y\in B\}\,.$$

In case B is the singleton set $\{x_0\}$, $A+\{x_0\}$ is denoted by $A+x_0$ and called a **translate** of A. For $\lambda\in\mathbf{R}$, λA is defined to be the set of all elements of the form λx with $x\in A$. Observe that if Y and Z are subspaces of X, then the sum $Y+Z$ is also a subspace of X. In the case $Y\cap Z=\{0\}$, we denote $Y+Z$ by $Y\oplus Z$ and call this subspace of X the **direct sum** of Y and Z.

For a normed linear space X, the open ball of radius 1 centered at the origin, $\{x\in X \mid \|x\|<1\}$, is called the **open unit ball** in X and $\{x\in X \mid \|x\|\leq 1\}$ is called the **closed unit ball** in X. We call a vector $x\in X$ for which $\|x\|=1$ a **unit vector**.

Many of the important theorems for metric spaces require completeness. Therefore, it is not surprising that among normed linear spaces those that are complete with respect to the metric induced by the norm will also be important. Interestingly enough, every complete metric space is isometric to a closed subset of a Banach space (see Problem 17).

Definition *A normed linear space is called a* **Banach space** *provided that it is complete as a metric space with the metric induced by the norm.*

The Riesz-Fischer Theorem tells us that for a general measure space $(X, \mathcal{M}, \mu)$ and $1\leq p\leq\infty$, $L^p(X, \mu)$ is a Banach space. We also proved that for X a compact topological space, $C(X)$, with the maximum norm, is a Banach space. Of course, it follows from the Completeness Axiom for $\mathbf{R}$ that each Euclidean space $\mathbf{R}^n$ is a Banach space.

PROBLEMS

1. Show that a non-empty subset S of a linear space X is a subspace if and only if $S+S=S$ and $\lambda\cdot S=S$ for each $\lambda\in\mathbf{R}, \lambda\neq 0$.
2. If Y and Z are subspaces of the linear space X, show that $Y+Z$ also is a subspace and $Y+Z=\text{span}[Y\cup Z]$.
3. Let S be a subset of a normed linear space X.
   (i) Show that the intersection of a collection of linear subspaces of X also is a linear subspace of X.
   (ii) Show that span$[S]$ is the intersection of all the linear subspaces of X that contain S and therefore is a linear subspace of X.
   (iii) Show that $\overline{\text{span}}\,[S]$ is the intersection of all the closed linear subspaces of X that contain S and therefore is a closed linear subspace of X.
4. For a normed linear space X, show that the function $\|\cdot\|\colon X\to\mathbf{R}$ is continuous.
5. For two normed linear spaces $(X, \|\cdot\|_1)$ and $(Y, \|\cdot\|_2)$, define a linear structure on the Cartesian product $X\times Y$ by $\lambda\cdot(x, y)=(\lambda x, \lambda y)$ and $(x_1, y_1)+(x_2, y_2)=(x_1+x_2, y_1+y_2)$. Define the product norm $\|\cdot\|$ by $\|(x, y)\|=\|x\|_1+\|y\|_2$, for $x\in X$ and $y\in Y$. Show that this is a norm with respect to which a sequence converges if and only if each of the two component sequences converges. Furthermore, show that if X and Y are Banach spaces, then so is $X\times Y$.

6. Let X be a normed linear space.

   (i) Let $\{x_n\}$ and $\{y_n\}$ be sequences in X such that $\{x_n\} \to x$ and $\{y_n\} \to y$. Show that for any real numbers α and β, $\{\alpha x_n + \beta y_n\} \to \alpha x + \beta y$.

   (ii) Use (i) to show that if Y is a subspace of X, then its closure $\overline{Y}$ also is a linear subspace of X.

   (iii) Use (i) to show that the vector sum is continuous from $X \times X$ to X and scalar multiplication is continuous from $\mathbf{R} \times X$ to X.

7. Show that the set $\mathcal{P}$ of all polynomials on $[a, b]$ is a linear space. For $\mathcal{P}$ considered as a subset of the normed linear space $C[a, b]$, show that $\mathcal{P}$ fails to be closed. For $\mathcal{P}$ considered as a subset of the normed linear space $L^1[a, b]$, show that $\mathcal{P}$ fails to be closed.

8. A non-negative real-valued function $\|\cdot\|$ defined on a vector space X is called a **pseudonorm** if $\|x + y\| \le \|x\| + \|y\|$ and $\|\alpha x\| = |\alpha|\ \|x\|$. Define $x \cong y$, provided that $\|x - y\| = 0$. Show that this is an equivalence relation. Define $X/_{\cong}$ to be the set of equivalence classes of X under $\cong$ and for $x \in X$ define $[x]$ to be the equivalence class of x. Show that $X/_{\cong}$ is a normed vector space if we define $\alpha[x] + \beta[y]$ to be the equivalence class of $\alpha x + \beta y$ and define $\|[x]\| = \|x\|$. Illustrate this procedure with $X = L^p[a, b], 1 \le p < \infty$.

17.2 LINEAR OPERATORS

Definition *Let X and Y be linear spaces. A mapping $T\colon X \to Y$ is said to be* **linear** *provided that for each $u, v \in X$, and real numbers α and β,*

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v).$$

Linear mappings are often called **linear operators** or linear transformations. In linear algebra, linear operators between finite dimensional linear spaces are considered, which, with respect to a choice of bases for the domain and range, are all given by matrix multiplication. For the $L^p(X, \mu)$ spaces, $1 \le p \le \infty$, we considered continuous linear operators from L^p to $\mathbf{R}$, which we called functionals, and, in Chapter 12, proved Representation Theorems, which characterized all such functionals.

Definition *Let X and Y be normed linear spaces. A linear operator $T\colon X \to Y$ is said to be* **bounded** *provided that there is a constant $M \ge 0$ for which*

$$\|T(u)\| \le M\|u\| \text{ for all } u \in X. \tag{1}$$

The infimum of all such M is called the **operator norm** *of T and denoted by $\|T\|$. The collection of bounded linear operators from X to Y is denoted by $\mathcal{L}(X, Y)$.*

Let X and Y be normed linear spaces and T belong to $\mathcal{L}(X, Y)$. It is easy to see that (1) holds for $M = \|T\|$. Hence, by the linearity of T,

$$\|T(u) - T(v)\| \le \|T\| \cdot \|u - v\| \text{ for all } u, v \in X. \tag{2}$$

From this we obtain the following continuity property of a bounded linear operator T:

$$\text{if } \{u_n\} \to u \text{ in } X, \text{ then } \{T(u_n)\} \to T(u) \text{ in } Y. \tag{3}$$

Indeed, there is the following basic result for linear operators, which we proved in Chapter 8 for linear functionals.

Theorem 1 *Let X and Y be normed linear spaces and the operator T: $X \to Y$ be linear. Then T is continuous if and only if it is bounded.*

Proof If T is bounded, (3) tells us that T is continuous. Now suppose that T is continuous. Since T is linear, $T(0) = 0$. Therefore, by the ϵ-δ criterion for continuity at $u = 0$, with $\epsilon = 1$, there is a $\delta > 0$ such that $\|T(u) - T(0)\| < 1$ if $\|u - 0\| < \delta$, that is, $\|T(u)\| < 1$ if $\|u\| < \delta$. For any $u \in X, u \neq 0$, set $\lambda = \delta/\|u\|$ and observe by the positive homogeneity of the norm, $\|\lambda u\| \leq \delta$. Thus $\|T(\lambda u)\| \leq 1$. Since $\|T(\lambda u)\| = \lambda\|T(u)\|$, we conclude that (1) holds for $M = 1/\delta$. □

Definition *Let X and Y be linear spaces. For T: $X \to Y$ and S: $X \to Y$ linear operators and real numbers α, β, the operator $\alpha T + \beta S$: $X \to Y$ is defined pointwise by*

$$(\alpha T + \beta S)(u) = \alpha T(u) + \beta S(u) \textit{ for all } u \in X. \tag{4}$$

Under pointwise scalar multiplication and addition the collection of linear operators between two linear spaces is a linear space.

Proposition 2 *Let X and Y be normed linear space. Then the collection of bounded linear operators from X to Y, $\mathcal{L}(X, Y)$, is a normed linear space.*

Proof Let T and S belong to $\mathcal{L}(X, Y)$. We infer from the triangle inequality for the norm on Y and (2) that

$$\|(T + S)(u)\| \leq \|T(u)\| + \|S(u)\| \leq \|T\|\|u\| + \|S\|\|u\| = (\|T\| + \|S\|)\|u\| \text{ for all } u \in X.$$

Therefore, $T + S$ is bounded and $\|T + S\| \leq \|T\| + \|S\|$. It is clear that for a real number α, αT is bounded and $\|\alpha T\| = |\alpha|\|T\|$ and $\|T\| = 0$ if and only if $T(u) = 0$ for all $u \in X$. □

Theorem 3 *Let X and Y be normed linear spaces. If Y is a Banach space, then so is $\mathcal{L}(X, Y)$.*

Proof Let $\{T_n\}$ be a Cauchy sequence in $\mathcal{L}(X, Y)$. Let u belong to X. Then, by (2), for all indices n and m,

$$\|T_n(u) - T_m(u)\| = \|(T_n - T_m)u\| \leq \|T_n - T_m\| \cdot \|u\|.$$

Thus $\{T_n(u)\}$ is a Cauchy sequence in Y. Since, by assumption, Y is complete, the sequence $\{T_n(u)\}$ converges to a member of Y, which we denote by $T(u)$. This defines a mapping T: $X \to Y$. It must be shown that T belongs to $\mathcal{L}(X, Y)$ and $\{T_n\} \to T$ in $\mathcal{L}(X, Y)$. To establish linearity, observe that for each u_1, u_2 in X, since each T_n is linear,

$$T(u_1) + T(u_2) = \lim_{n\to\infty} T_n(u_1) + \lim_{n\to\infty} T_n(u_2) = \lim_{n\to\infty} T_n(u_1 + u_2) = T(u_1 + u_2),$$

and similarly $T(\lambda u) = \lambda T(u)$.

We establish the boundedness of T and the convergence of $\{T_n\}$ to T in $\mathcal{L}(X, Y)$ simultaneously. Let $\epsilon > 0$. Choose an index N such that for all $n \geq N, k \geq 1$, $\|T_n - T_{n+k}\| < \epsilon/2$. Thus, by (2), for all $u \in X$,

$$\|T_n(u) - T_{n+k}(u)\| = \|(T_n - T_{n+k})u\| \leq \|T_n - T_{n+k}\| \cdot \|u\| < \epsilon/2\|u\|.$$

Fix $n \geq N$ and $u \in X$. Since $\lim_{k\to\infty} T_{n+k}(u) = T(u)$ and the norm is continuous we conclude that

$$\|T_n(u) - T(u)\| \leq \epsilon/2\|u\|.$$

In particular, the linear operator $T_N - T$ is bounded and therefore, since T_N also is bounded, so is T. Moreover, $\|T_n - T\| < \epsilon$ for $n \geq N$. Thus $\{T_n\} \to T$ in $\mathcal{L}(X, Y)$. □

For two normed linear spaces X and Y, an operator $T \in \mathcal{L}(X, Y)$ is called an **isomorphism** provided that it is one-to-one, onto, and has a continuous inverse. For T in $\mathcal{L}(X, Y)$, if it is one-to-one and onto, its inverse is linear. To be an isomorphism requires that the inverse be bounded, that is, the inverse belong to $\mathcal{L}(Y, X)$. Two normed linear spaces are said to be **isomorphic** provided that there is an isomorphism between them. This is an equivalence relation that plays the same role for normed linear spaces that homeomorphism plays for topological spaces. An isomorphism that also preserves the norm is called an **isometric isomorphism**: it is an isomorphism that is also an isometry of the metric structures associated with the norms.

For a linear operator $T: X \to Y$, the subspace of X, $\{x \in X \mid T(x) = 0\}$, is called the **kernel** of T and denoted by ker T. Observe that T is one-to-one if and only if ker $T = \{0\}$. We denote the **image** of T, $T(X)$, by Im T.

PROBLEMS

9. Let X and Y be normed linear spaces and $T: X \to Y$ be linear.
   (i) Show that T is continuous if and only if it is continuous at a single point u_0 in X.
   (ii) Show that T is Lipschitz if and only if it is continuous.
   (iii) Show that neither (i) nor (ii) hold in the absence of the linearity assumption on T.
10. For X and Y normed linear spaces and $T \in \mathcal{L}(X, Y)$, show that $\|T\|$ is the smallest Lipschitz constant for the mapping T, that is, the smallest number $c \geq 0$ for which
$$\|T(u) - T(v)\| \leq c \cdot \|u - v\| \text{ for all } u, v \in X.$$
11. For X and Y normed linear spaces and $T \in \mathcal{L}(X, Y)$, show that
$$\|T\| = \sup\{\|T(u)\| \mid u \in X, \|u\| \leq 1\}.$$
12. For X and Y normed linear spaces, let $\{T_n\} \to T$ in $\mathcal{L}(X, Y)$ and $\{u_n\} \to u$ in X. Show that $\{T_n(u_n)\} \to T(u)$ in Y.
13. Let X be a Banach space and $T \in \mathcal{L}(X, X)$ have $\|T\| < 1$.
   (i) Use the Contraction Mapping Principle to show that $I - T \in \mathcal{L}(X, X)$ is one-to-one and onto.
   (ii) Show that $I - T$ is an isomorphism.

14. (Neumann Series) Let X be a Banach space and $T \in \mathcal{L}(X, X)$ have $\|T\| < 1$. Define $T^0 = \text{Id}$.
   (i) Use the completeness of $\mathcal{L}(X, X)$ to show that $\sum_{n=0}^{\infty} T^n$ converges in $\mathcal{L}(X, X)$.
   (ii) Show that $(I - T)^{-1} = \sum_{n=0}^{\infty} T^n$.
15. For X and Y normed linear spaces and $T \in \mathcal{L}(X, Y)$, show that T is an isomorphism if and only if there is an operator $S \in \mathcal{L}(Y, X)$ such that for each $u \in X$ and $v \in Y$,
$$S(T(u)) = u \text{ and } T(S(v)) = v.$$
16. For X and Y normed linear spaces and $T \in \mathcal{L}(X, Y)$, show that $\ker T$ is a closed subspace of X and that T is one-to-one if and only if $\ker T = \{0\}$.
17. Let (X, ρ) be a metric space containing the point x_0. Define $\text{Lip}_0(X)$ to be the set of real-valued Lipschitz functions f on X that vanish at x_0. Show that $\text{Lip}_0(X)$ is a linear space that is normed by defining, for $f \in \text{Lip}_0(X)$,
$$\|f\| = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\rho(x, y)}.$$
Show that $\text{Lip}_0(X)$ is a Banach space. For each $x \in X$, define the linear functional F_x on $\text{Lip}_0(X)$ by setting $F_x(f) = f(x)$. Show that F_x belongs to $\mathcal{L}(\text{Lip}_0(X), \mathbf{R})$ and that for $x, y \in X$, $\|F_x - F_y\| = \rho(x, y)$. Thus, X is isometric to a subset of the Banach space $\mathcal{L}(\text{Lip}_0(X), \mathbf{R})$. Since any closed subset of a complete metric space is complete, this provides another proof of the existence of a completion for any metric space X. It also shows that any metric space is isometric to a subset of a normed linear space.
18. Use the preceding problem to show that every normed linear space is a dense subspace of a Banach space.
19. For X a normed linear space and $T, S \in \mathcal{L}(X, X)$, show that the composition $S \circ T$ also belongs to $\mathcal{L}(X, X)$ and $\|S \circ T\| \leq \|S\| \cdot \|T\|$.
20. Let X be a normed linear space and Y a closed linear subspace of X. Show that $\|x\|_1 = \inf_{y \in Y} \|x - y\|$ defines a pseudonorm on X. The normed linear space induced by the pseudonorm $\|\cdot\|_1$ (see Problem 8) is denoted by X/Y and called the **quotient space** of X modulo Y. Show that the natural map φ of X onto X/Y takes open sets into open sets.
21. Show that if X is a Banach space and Y a closed linear proper subspace of X, then the quotient X/Y also is a Banach space and the natural map $\varphi\colon X \to X/Y$ has norm 1.
22. Let X and Y be normed linear spaces, $T \in \mathcal{L}(X, Y)$ and $\ker T = Z$. Show that there is a unique bounded linear operator S from X/Z into Y such that $T = S \circ \varphi$ where $\varphi\colon X \to X/Z$ is the natural map. Moreover, show that $\|T\| = \|S\|$.

17.3 COMPACTNESS LOST: INFINITE DIMENSIONAL NORMED LINEAR SPACES

A linear space X is said to be finite dimensional provided that there is a subset $\{e_1, \ldots, e_n\}$ of X that spans X. If no proper subset also spans X, the set $\{e_1, \ldots, e_n\}$ is called a basis for X and n called the dimension of X. If X is not spanned by a finite collection of vectors, it is said to be infinite dimensional.

Theorem 4 *Any two norms on a finite dimensional linear space are equivalent.*

Proof Since equivalence of norms is an equivalence relation on the set of norms on X, it suffices to select a particular norm $\|\cdot\|_*$ on X and show that any norm on X is equivalent to $\|\cdot\|_*$. Let $\dim X = n$ and $\{e_1, \ldots, e_n\}$ be a basis for X. For any $x = x_1e_1 + \cdots + x_ne_n \in X$, define

$$\|x\|_* = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Since the Euclidean norm is a norm on $\mathbf{R}^n$, $\|\cdot\|_*$ is a norm on X.

Let $\|\cdot\|$ be any norm on X. We claim it is equivalent to $\|\cdot\|_*$. First we find a $c_1 \geq 0$ for which

$$\|x\| \leq c_1 \cdot \|x\|_* \text{ for all } x \in X. \tag{5}$$

Indeed, for $x = x_1e_1 + \cdots + x_ne_n \in X$, by the triangle inequality and positive homogeneity of the norm $\|\cdot\|$, together with the Cauchy-Schwarz inequality on $\mathbf{R}^n$,

$$\|x\| \leq \sum_{i=1}^n |x_i|\|e_i\| \leq c_1 \cdot \|x\|_*, \text{ where } c_1 = \sqrt{\sum_{i=1}^n \|e_i\|^2}.$$

We now find a $c_2 > 0$ for which

$$\|x\|_* \leq c_2 \cdot \|x\| \text{ for all } x \in X. \tag{6}$$

Define the real-valued function $f\colon \mathbf{R}^n \to \mathbf{R}$ by

$$f(x_1, \ldots, x_n) = \left\|\sum_{i=1}^n x_ie_i\right\|.$$

This function is continuous since it is Lipschitz with Lipschitz constant c_1 if $\mathbf{R}^n$ is considered as a metric space with the Euclidean metric. Since $\{e_1, \ldots, e_n\}$ is linearly independent, f takes positive values on the boundary of the unit ball, $S = \{x \in \mathbf{R}^n \mid \sum_{i=1}^n x_i^2 = 1\}$, which is compact since it is both closed and bounded. A continuous real-valued function on a compact topological space takes a minimum value. Let $m > 0$ be the minimum value of f on S. By the homogeneity of the norm $\|\cdot\|$,

$$\|x\| \geq m \cdot \|x\|_* \text{ for all } x \in X.$$

Therefore (6) holds for $c_2 = 1/m$. $\square$

Corollary 5 *Any two normed linear spaces of the same finite dimension are isomorphic.*

Proof Since being isomorphic is an equivalence relation among normed linear spaces, it suffices to show that if X is a normed linear space of dimension n, then it is isomorphic to the Euclidean space $\mathbf{R}^n$. Let $\{e_1, \ldots, e_n\}$ be a basis for X. Define the linear mapping $T\colon \mathbf{R}^n \to X$ by setting, for $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$,

$$T(x) = \sum_{i=1}^{n} x_i e_i.$$

Since $\{e_1, \ldots, e_n\}$ is a basis, T is one-to-one and onto. Clearly T is linear. It remains to show that T and its inverse are continuous. Since a linear operator is continuous if and only if it is bounded, this amounts to showing that there are constants c_1 and c_2 such that for each $x \in \mathbf{R}^n$,

$$\|T(x)\| \le c_1 \cdot \|x\|_* \text{ and } \|T(x)\| \ge c_2 \cdot \|x\|_*,$$

where $\|\cdot\|_*$ denotes the Euclidean norm on $\mathbf{R}^n$. The existence of these two constants follows from the observation that $x \mapsto \|T(x)\|$ defines a norm on $\mathbf{R}^n$, which, since all norms on $\mathbf{R}^n$ are equivalent, is equivalent to the Euclidean norm. □

Corollary 6 *Any finite dimensional normed linear space is complete, and therefore any finite dimensional subspace of a normed linear space is closed.*

Proof A finite dimensional space of dimension n is isomorphic to the Euclidean space $\mathbf{R}^n$, which is complete since $\mathbf{R}$ is complete. Since completeness is preserved under isomorphisms, every finite dimensional normed linear space is complete. For a finite dimensional subspace Y of a normed linear space X, since Y, with the metric induced by the inherited norm, is complete, Y is a closed subset of the metric space X, where X is considered as a metric space with the metric induced by the norm. □

Corollary 7 *The closed unit ball in a finite dimensional normed linear space is compact.*

Proof Let X be a normed linear space of dimension n and B be its closed unit ball. Let $T\colon X \to \mathbf{R}^n$ be an isomorphism. Since T is a linear isomorphism, $T(B)$ is a closed, bounded subset of $\mathbf{R}^n$, and therefore is compact. Consequently, so is B, the image of $T(B)$ under the continuous mappings T^{-1}. □

Riesz's Theorem *The closed unit ball of a normed linear space X is compact if and only if X is finite dimensional.*

The heart of the proof of this theorem lies in the following lemma.

Riesz's Lemma *Let Y be a closed, proper linear subspace of a normed linear space X. Then for each $\epsilon > 0$, there is a unit vector $x_0 \in X$ for which*

$$\|x_0 - y\| > 1 - \epsilon \text{ for all } y \in Y.$$

Proof We consider the case $\epsilon = 1/2$ and leave the general case as an exercise. Since Y is a proper subset of X, we may choose $x \in X \sim Y$. Since Y is a closed subset of X, its complement in X is open and therefore there is a ball centered at x that is disjoint from Y, that is,

$$\inf\{\|x - y'\| \mid y' \in Y\} = d > 0. \tag{7}$$

Choose a vector $y_1 \in Y$ for which

$$\|x - y_1\| < 2d. \tag{8}$$

Define

$$x_0 = \frac{x - y_1}{||x - y_1||}.$$

Then x_0 is a unit vector. Moreover, observe that for any $y \in Y$,

$$x_0 - y = \frac{x - y_1}{||x - y_1||} - y = \frac{1}{||x - y_1||}\{x - y_1 - ||x - y_1||\,y\} = \frac{1}{||x - y_1||}\{x - y'\},$$

where $y' = y_1 + ||x - y_1||\,y$ belongs to Y. Therefore, by (7) and (8),

$$||x_0 - y|| > \frac{1}{2d}||x - y'|| \geq \frac{1}{2} \text{ for all } y \in Y. \qquad \square$$

Proof of Riesz's Theorem We have already shown that the closed unit ball B in a finite dimensional normed linear space is compact. It remains to show that B fails to be compact if X is infinite dimensional. Assume that X is infinite dimensional. We will inductively choose a sequence $\{x_n\}$ in B such that $\|x_n - x_m\| > 1/2$ for $n \neq m$. This sequence has no Cauchy subsequence and therefore no convergent subsequence. Thus B is not sequentially compact, and therefore, since B is a metric space, not compact.

It remains to choose this sequence. Choose any vector $x_1 \in B$. For a natural number n, suppose that we have chosen n vectors in B, $\{x_1, \ldots, x_n\}$, each pair of which are more than a distance $1/2$ apart. Let X_n be the linear space spanned by these n vectors. Then X_n is a finite dimensional subspace of X and so it is closed. Moreover, X_n is a proper subspace of X since $\dim X = \infty$. By the preceding lemma, there is an x_{n+1} in B for which $\|x_i - x_{n+1}\| > 1/2$ for $1 \leq i \leq n$. Therefore, we have inductively chosen a sequence in B any two terms of which are more than a distance $1/2$ apart. $\square$

PROBLEMS

23. Show that a subset of a finite dimensional normed linear space X is compact if and only if it is closed and bounded.
24. Complete the proof of Riesz's Lemma for $\epsilon \neq 1/2$.
25. Exhibit an open cover of the closed unit ball of $X = \ell^2$ that has no finite subcover. Then do the same for $X = C[0, 1]$ and $X = L^2[0, 1]$.
26. For normed linear spaces X and Y, let $T \colon X \to Y$ be linear. If X is finite dimensional, show that T is continuous. If Y is finite dimensional, show that T is continuous if and only if $\ker T$ is closed.

27. (Another proof of Riesz's Theorem) Let X be an infinite dimensional normed linear space, B the closed unit ball in X, and B_0 the unit open ball in X. Suppose B is compact. Then the open cover $\{x + (1/3)B_0\}_{x \in B}$ of B has a finite subcover $\{x_i + (1/3)B_0\}_{1 \le i \le n}$. Use Riesz's Lemma with $Y = \text{span}[\{x_1, \ldots, x_n\}]$ to derive a contradiction.
28. Let X be a normed linear space. Show that X is separable if and only if there is a compact subset K of X for which $\overline{\text{span}}[K] = X$.

17.4 THE OPEN MAPPING AND CLOSED GRAPH THEOREMS

In this section, the Baire Category Theorem is used to establish two essential tools, the Open Mapping Theorem and the Closed Graph Theorem, for the analysis of linear operators between infinite dimensional Banach spaces. We first use it to prove the following theorem.

Theorem 8 *Let X and Y be Banach spaces and the linear operator $T\colon X \to Y$ be continuous. Then $T(X)$ is a closed subspace of Y if and only if there is a constant $M > 0$ for which given any $y \in T(X)$, there is an $x \in X$ such that*

$$T(x) = y \text{ and } \|x\| \le M\|y\|. \tag{9}$$

Proof First suppose that there is a constant $M > 0$ for which (9) holds. Let $\{y_n\}$ be a sequence in $T(X)$ that converges to $y_* \in Y$. It must be shown that y_* belongs to $T(X)$. By selecting a subsequence if necessary, it may be assumed that

$$\|y_n - y_{n-1}\| \le 1/2^n \text{ for all } n \ge 2.$$

By the choice of M, for each natural number $n \ge 2$, there is a vector $u_n \in X$ for which

$$T(u_n) = y_n - y_{n-1} \text{ and } \|u_n\| \le M/2^n.$$

Therefore, for $n \ge 2$, if we define $x_n = \sum_{j=2}^n u_j$, then

$$T(x_n) = y_n - y_1 \tag{10}$$

and

$$\|x_{n+k} - x_n\| \le M \cdot \sum_{j=n}^{\infty} 1/2^j \text{ for all } k \ge 1. \tag{11}$$

But X is a Banach space and therefore the Cauchy sequence $\{x_n\}$ converges to a vector $x_* \in X$. Take the limit as $n \to \infty$ in (10) and use the continuity of T to infer that $y_* = T(x_*) + y_1$. Since y_1 belongs to $T(X)$ so does y_*. Therefore, $T(X)$ is closed.

To prove the converse, assume that $T(X)$ is a closed subspace of Y. For notational convenience, assume that $Y = T(X)$. Let B_X and B_Y denote the open unit balls in X and Y, respectively. Since $T(X) = Y$,

$$Y = \bigcup_{n=1}^{\infty} n \cdot T(B_X) = \bigcup_{n=1}^{\infty} n \cdot \overline{T(B_X)}.$$

The Banach space Y has non-empty interior and therefore it follows from the Baire Category Theorem that there is a natural number n such that the closed set $n \cdot \overline{T(B_X)}$ contains an open ball, which we write as $y_0 + [r_1 \cdot B_Y]$. Thus

$$r_1 B_Y \subseteq n\overline{T(B_X)} - y_0 \subseteq 2n\overline{T(B_X)}.$$

Hence, if we set $r = 2n/r_1$, since $\overline{T(B_X)}$ is closed, we obtain $\overline{B_Y} \subseteq r \cdot \overline{T(B_X)}$. Therefore, since $\overline{B}_Y$ is the closed unit ball in Y, for each $y \in Y$ and $\epsilon > 0$, there is an $x \in X$ for which

$$\|y - T(x)\| < \epsilon \text{ and } \|x\| \leq r \cdot \|y\|. \tag{12}$$

We claim that (9) holds for $M = 2r$. Indeed, let y_* belong to Y, $y_* \neq 0$. According to (12) with $\epsilon = 1/2 \cdot \|y_*\|$ and $y = y_*$, there is a vector $u_1 \in X$ for which

$$\|y_* - T(u_1)\| < 1/2 \cdot \|y_*\| \text{ and } \|u_1\| \leq r \cdot \|y_*\|.$$

Now use (12) again, this time with $\epsilon = 1/2^2 \cdot \|y_*\|$ and $y = y_* - T(u_1)$. There is a vector u_2 in X for which

$$\|y_* - T(u_1) - T(u_2)\| < 1/2^2 \cdot \|y_*\| \text{ and } \|u_2\| \leq r/2 \cdot \|y_*\|.$$

We continue this selection process and inductively choose a sequence $\{u_k\}$ in X such that for each k,

$$\|y_* - T(u_1) - T(u_2) - \cdots - T(u_k)\| < 1/2^k \cdot \|y_*\| \text{ and } \|u_k\| \leq r/2^{k-1} \cdot \|y_*\|.$$

For each natural number n, define $x_n = \sum_{k=1}^{n} u_k$. Then, by the linearity of T, for each n,

$$\|y_* - T(x_n)\| \leq 1/2^n \cdot \|y_*\|,$$

$$\|x_{n+k} - x_n\| \leq r \cdot \|y_*\| \cdot \sum_{j=n}^{\infty} 1/2^j \text{ and } \|x_n\| \leq 2 \cdot r \cdot \|y_*\|.$$

By assumption, X is complete. Therefore, the Cauchy sequence $\{x_n\}$ converges to a vector x_* in X. Since T is continuous and the norm is continuous,

$$T(x_*) = y_* \text{ and } \|x_*\| \leq 2 \cdot r \cdot \|y_*\|.$$

Thus (9) holds for $M = 2 \cdot r$. The proof is complete. □

A mapping f: $X \to Y$ from the topological space X to the topological space Y is said to be **open** provided that the image of each open set in X is open in the topological space $f(X)$, where $f(X)$ has the subspace topology inherited from Y. Therefore, a continuous one-to-one mapping f of X into Y is open if and only if f is a topological homeomorphism between X and $f(X)$.

The Open Mapping Theorem *Let X and Y be Banach spaces and the linear operator T: $X \to Y$ be continuous. Then its image $T(X)$ is a closed subspace of Y if and only if the operator T is open.*

Proof The preceding theorem tells us that it suffices to show that T is open if and only if there is a constant $M > 0$ for which (9) holds. Let B_X and B_Y denote the open unit balls in X and Y, respectively. We infer from the homogeneity of T and of the norms that (9) is equivalent to the inclusion

$$\overline{B_Y} \cap T(X) \subseteq M \cdot T(\overline{B_X}).$$

By homogeneity, this inclusion is equivalent to the existence of a constant M' for which $B_Y \cap T(X) \subseteq M' \cdot T(B_X)$. Therefore, it must be shown that T is open if and only if there is an $r > 0$ for which

$$[r \cdot B_Y] \cap T(X) \subseteq T(B_X). \tag{13}$$

First assume that the operator T is open. Then $T(B_X) \cap T(X)$ is an open subset of $T(X)$ which contains 0. Thus there is an $r > 0$ for which $r \cdot B_Y \cap T(X) \subseteq T(B_X) \cap T(X) \subseteq T(B_X)$. Therefore, (13) holds for this choice of r. To prove the converse, assume that (13) holds. Let $\mathcal{O}$ be an open subset of X and x_0 belong to $\mathcal{O}$. We must show that $T(x_0)$ is an interior point of $T(\mathcal{O})$. Since x_0 is an interior point of $\mathcal{O}$, there is an $R > 0$ for which $x_0 + R \cdot B_X \subseteq \mathcal{O}$. We infer from (13) that the open ball of radius $r \cdot R$ about $T(x_0)$ in $T(X)$ is contained in $T(\mathcal{O})$. Thus $T(x_0)$ is an interior point of $T(\mathcal{O})$. □

Corollary 9 *Let X and Y be Banach spaces and $T \in \mathcal{L}(X, Y)$ be one-to-one and onto. Then T^{-1} is continuous.*

Proof The operator T^{-1} is continuous if and only if the operator T is open. □

Corollary 10 *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on a linear space X for which both $(X, \|\cdot\|_1)$ and $(X, \|\cdot\|_2)$ are Banach spaces. Suppose there is a $c \geq 0$ for which*

$$\|\cdot\|_2 \leq c \cdot \|\cdot\|_1 \text{ on } X.$$

Then these two norms are equivalent.

Proof Define the identity map $\mathrm{Id}\colon X \to X$ by $\mathrm{Id}(x) = x$ for all $x \in X$. By assumption,

$$\mathrm{Id}\colon (X, \|\cdot\|_1) \to (X, \|\cdot\|_2)$$

is a bounded, and therefore continuous, operator between Banach spaces and, of course, it is both one-to-one and onto. By the Open Mapping Theorem, the inverse of the identity, $\mathrm{Id}\colon (X, \|\cdot\|_2) \to (X, \|\cdot\|_1)$ also is continuous, that is, it is bounded: there is an $M \geq 0$ for which

$$\|\cdot\|_1 \leq M \cdot \|\cdot\|_2 \text{ on } X.$$

Therefore, the two norms are equivalent. □

Definition *A linear operator $T\colon X \to Y$ between normed linear spaces X and Y is said to be* **closed** *provided that whenever $\{x_n\}$ is a sequence in X*

$$\text{if } \{x_n\} \to x_0 \text{ and } \{T(x_n)\} \to y_0, \text{ then } T(x_0) = y_0.$$

The **graph** of a mapping $T: X \to Y$ is the set $\{(x, T(x)) \in X \times Y \mid x \in X\}$. Therefore, an operator is closed if and only if its graph is a closed subspace of the product space $X \times Y$.

The Closed Graph Theorem *Let $T: X \to Y$ be a linear operator between the Banach spaces X and Y. Then T is continuous if and only if it is closed.*

Proof It is clear that T is closed if it is continuous. To prove the converse, assume that T is closed. Introduce a new norm $\|\cdot\|_*$ on X by

$$\|x\|_* = \|x\| + \|T(x)\| \text{ for all } x \in X.$$

The closedness of the operator T is equivalent to the completeness of the normed linear space $(X, \|\cdot\|_*)$. On the other hand, clearly

$$\|\cdot\| \leq \|\cdot\|_* \text{ on } X.$$

Since both $(X, \|\cdot\|_*)$ and $(X, \|\cdot\|)$ are Banach spaces it follows from the preceding corollary that there is a $c \geq 0$ for which

$$\|\cdot\|_* \leq c \cdot \|\cdot\| \text{ on } X.$$

Thus for all $x \in X$,

$$\|T(x)\| \leq \|x\| + \|T(x)\| \leq c\|x\|.$$

Therefore, T is bounded and hence is continuous. □

Let X and Y be Banach spaces and the operator $T: X \to Y$ be linear. To directly establish the continuity of T it is, of course, necessary to show that $\{T(x_n)\} \to T(x_0)$ in Y if $\{x_n\} \to x_0$ in X. The Closed Graph Theorem provides a drastic simplification in this criterion. It tells us that to establish the continuity of T it suffices to check that $\{T(x_n)\} \to T(x_0)$ in Y for sequences $\{x_n\}$ such that $\{x_n\} \to x_0$ in X and $\{T(x_n)\}$ is Cauchy in Y. The usefulness of this simplification will be evident in the proof of Theorem 11.

Let V be a linear subspace of a linear space X. An argument using Zorn's Lemma (see Problem 35) shows that there is a subspace W of X for which there is the direct sum decomposition

$$X = V \oplus W. \tag{14}$$

We call W a **linear complement** of V in X. If a subspace of X has a finite dimensional linear complement in X, then it is said to have **finite codimension** in X. For $x \in X$ and the decomposition (14), let $x = v + w$, for $v \in V$ and $w \in W$. Define $P(x) = v$. We leave it as an algebraic exercise to show that $P: X \to X$ is linear,

$$P^2 = P \text{ on } X,\ P(X) = V \text{ and } (\mathrm{Id} - P)(X) = W. \tag{15}$$

We call P the **projection** of X onto V along W. We leave it as a second algebraic exercise to show that if $P: X \to X$ is any linear operator for which $P^2 = P$, then

$$X = P(X) \oplus (\mathrm{Id} - P)(X). \tag{16}$$

We therefore call a linear operator $P: X \to X$ for which $P^2 = P$ a **projection**. If P is a projection, then $(\mathrm{Id} - P)^2 = \mathrm{Id} - P$ and therefore $\mathrm{Id} - P$ also is a projection.

Now assume that the linear space X is normed. A closed subspace W of X for which (14) holds is called a **closed linear complement** of V in X. Whereas every linear subspace has a linear complement, in general, it is very difficult to determine if a linear subspace has a closed linear complement. According to Corollary 8 of the next chapter, every finite dimensional subspace of a normed linear space has a closed linear complement. By the forthcoming Theorem 3 of Chapter 20, every closed subspace of a Hilbert space has a closed linear complement. For now, there is the following criterion, in terms of the continuity of projections, for the existence of closed linear complements.

Theorem 11 *Let V be a closed subspace of a Banach space X. Then V has a closed linear complement in X if and only if there is a continuous projection of X onto V.*

Proof Suppose P is a continuous projection defined on X, and let $\{x_n\}$ be a sequence in $P(X)$ converging to x. Then $x_n = P(x_n) \to P(x)$, so $x = P(x)$. This shows that $P(X)$ is closed. Now assume that there is a continuous projection P of X onto V. There is the direct sum decomposition $X = V \oplus (\mathrm{Id} - P)(X)$. Since $\mathrm{Id} - P$ is a continuous projection, $(\mathrm{Id} - P)(X)$ is closed. To prove the converse, assume that there is a closed subspace W of X for which there is the direct sum decomposition (14). Define P to be the projection of X onto V along W. We claim that the operator P is continuous. According to the Closed Graph Theorem, to verify this claim it is sufficient to show that P is closed. Let $\{x_n\}$ be a sequence in X for which $\{x_n\} \to x_0$ and $\{P(x_n)\} \to y_0$. Since $\{P(x_n)\}$ is a sequence in the closed set V that converges to y_0, the vector y_0 belongs to V. Since $(\mathrm{Id} - P)(x_n)\}$ is a sequence in the closed set W that converges to $x_0 - y_0$, the vector $x_0 - y_0$ belongs to W. Therefore $P(y_0) = y_0$ and $P(x_0 - y_0) = 0$. Hence $y_0 = P(x_0)$, so the operator P is closed. $\square$

In view of Theorem 8 and its corollary, the Open Mapping Theorem, it is interesting to provide criteria to determine when the image of a continuous linear operator is closed. The following theorem provides one such criterion.

Theorem 12 *Let X and Y be Banach spaces and the linear operator $T: X \to Y$ be continuous. If $T(X)$ has a closed linear complement in Y, then $T(X)$ is closed in Y. In particular, if $T(X)$ has finite codimension in Y, then $T(X)$ is closed in Y.*

Proof Let Y_0 be a closed subspace of Y for which

$$T(X) \oplus Y_0 = Y. \tag{17}$$

Since Y is a Banach space, so is Y_0. Consider the Banach space $X \times Y_0$, where the linear structure on the Cartesian product is defined componentwise and the norm is defined by

$$\|(x, y)\| = \|x\| + \|y\| \text{ for all } (x, y) \in X \times Y_0.$$

Then $X \times Y_0$ is a Banach space. Define the linear operator $S: X \times Y_0 \to Y$ by

$$S(x, y) = T(x) + y \text{ for all } (x, y) \in X \times Y_0.$$

Then S is a continuous linear mapping of the Banach space $X \times Y_0$ onto the Banach space Y. By Theorem 8, there is an $M > 0$ such that for each $y \in T(x)$ there is an $(x, y') \in X \times Y_0$ for which

$$T(x) + y' = y \text{ and } \|x\| + \|y'\| \leq M \cdot \|y\|.$$

But since $T(X) \cap Y_0 = \{0\}$, $y' = 0$. Therefore,

$$T(x) = y \text{ and } \|x\| \le M \cdot \|y\|.$$

Once more we use Theorem 8 to conclude that $T(X)$ is a closed subspace of Y. Finally, since every finite dimensional subspace of a normed linear space is closed, if $T(X)$ has finite codimension, it is closed. □

All linear operators on a finite dimensional normed linear space are continuous, open, and have closed images. The results in the section are only significant for linear operators defined on infinite dimensional Banach spaces, in which case continuity of the operator does not imply that the image is closed. We leave it as an exercise to verify that the operator $T\colon \ell^2 \to \ell^2$ defined by

$$T(\{x_n\}) = \{x_n/n\} \text{ for } \{x_n\} \in \ell^2$$

is continuous but does not have closed image and is not open.

PROBLEMS

29. Let X be a finite dimensional normed linear space and Y a normed linear space. Show that every linear operator $T\colon X \to Y$ is continuous and open.
30. Let X be a Banach space and $P \in \mathcal{L}(X, X)$ be a projection. Show that P is open.
31. Let $T\colon X \to Y$ be a continuous linear operator between the Banach spaces X and Y. Show that T is open if the image under T of the open unit ball in X is dense in a neighborhood of the origin in Y.
32. Let $\{u_n\}$ be a sequence in a Banach space X. Suppose that $\sum_{k=1}^{\infty} \|u_k\| < \infty$. Show that there is an $x \in X$ for which
$$\lim_{n\to\infty} \sum_{k=1}^{n} u_k = x.$$
33. Let T be a linear operator from a normed linear space X to a finite-dimensional normed linear space Y. Show that T is continuous if and only if ker T is a closed subspace of X.
34. Let X be a Banach space, the operator $T \in \mathcal{L}(X, X)$ be open and X_0 be a closed subspace of X. The restriction T_0 of T to X_0 is continuous. Is T_0 necessarily open?
35. Let V be a linear subspace of a linear space X. Argue as follows to show that V has a linear complement in X.
    (i) If $\dim X < \infty$, let $\{e_i\}_{i=1}^{n}$ be a basis for V. Extend this basis for V to a basis $\{e_i\}_{i=1}^{n+k}$ for X. Then define $W = \text{span}[\{e_{n+1}, \ldots, e_{n+k}\}]$.
    (ii) If $\dim X = \infty$, apply Zorn's Lemma to the collection $\mathcal{F}$ of all subspaces Z of X for which $V \cap Z = \{0\}$, ordered by set inclusion.
36. Verify (15) and (16).
37. Let Y be a normed linear space. Show that Y is a Banach space if and only if there is a Banach space X and a continuous, linear, open mapping of X onto Y.

17.5 THE UNIFORM BOUNDEDNESS PRINCIPLE

We have the following Uniform Boundedness Principle for continuous linear operators, which, as was the Non-linear Uniform Boundedness Principle in Chapter 15, is a consequence of the Baire Category Theorem.

The Uniform Boundedness Principle *For X a Banach space and Y a normed linear space, consider a collection $\mathcal{F} \subseteq \mathcal{L}(X, Y)$. Assume that the family $\mathcal{F}$ is pointwise bounded, in the sense that for each x in X there is a constant $M_x \geq 0$ for which*

$$\|T(x)\| \leq M_x \text{ for all } T \in \mathcal{F}.$$

Then the family $\mathcal{F}$ is bounded in $\mathcal{L}(X, Y)$, in the sense that there is a constant $M \geq 0$ for which $\|T\| \leq M$ for all T in $\mathcal{F}$.

Proof For each n, define $F_n = \{x \in X \mid ||T(x)|| \leq n, \text{for all}\, T \in \mathcal{F}\}$. Since the operators in $\mathcal{F}$ are continuous, each $\mathcal{F}_n$ is closed, and since $\mathcal{F}$ is pointwise bounded, $X = \cup_{n=1}^{\infty} F_n$. But X is a complete metric space and therefore, by the Baire Category Theorem, there is an open ball $B(x_0, r)$ in X and a constant $C \geq 0$ for which

$$\|T(x)\| \leq C \text{ for all } x \in B(x_0, r) \text{ and } T \in \mathcal{F}.$$

Thus, for each $T \in \mathcal{F}$,

$$\|T(x)\| = \|T([x + x_0] - x_0)\| \leq \|T(x + x_0)\| + \|T(x_0)\| \leq C + M_{x_0} \text{ for all } x \in B(0, r).$$

Therefore, setting $M = (1/r)(C + M_{x_0})$, we have $\|T\| \leq M$ for all T in $\mathcal{F}$. $\square$

The Banach-Saks-Steinhaus Theorem *Let X be a Banach space, Y a normed linear space, and $\{T_n \colon X \to Y\}$ a sequence of continuous linear operators. Assume that for each $x \in X$,*

$$\lim_{n \to \infty} T_n(x) \text{ exists in } Y. \tag{18}$$

Then the sequence of operators $\{T_n \colon X \to Y\}$ is uniformly bounded. Furthermore, the operator $T \colon X \to Y$ defined by

$$T(x) = \lim_{n \to \infty} T_n(x) \text{ for all } x \in X$$

is linear, continuous, and

$$\|T\| \leq \liminf \|T_n\|.$$

Proof The pointwise limit of a sequence of linear operators is linear. Thus T is linear. We infer from the Uniform Boundedness Principle that the sequence $\{T_n\}$ is uniformly bounded. Therefore, $\liminf \|T_n\|$ is finite. Let x belong to X. By the continuity of the norm on Y, $\lim_{n \to \infty} \|T_n(x)\| \to \|T(x)\|$. Since, for all n, $\|T_n(x)\| \leq \|T_n\| \cdot \|x\|$, we also have $\|T(x)\| \leq \liminf \|T_n\| \cdot \|x\|$. Therefore, T is bounded and $\|T\| \leq \liminf \|T_n\|$. A bounded linear operator is continuous. $\square$

PROBLEMS

38. As a consequence of the Baire Category Theorem, it was proven that a real-valued mapping that is the pointwise limit of a sequence of continuous mappings on a complete metric space must be continuous at all points of a dense subset of its domain. Adapt that proof so that it applies to mappings into any metric space. Use this to prove that the pointwise limit of a sequence of continuous linear operators on a Banach space has a limit that is continuous at some point and hence, by linearity, is continuous.

39. Let $\{f_n\}$ be a sequence in $L^\infty[a, b]$. Suppose that for each $g \in L^1[a, b]$, $\lim_{n\to\infty} \int_a^b g \cdot f_n \mathrm{dm}$ exists. Show that there is a function $f \in L^\infty[a, b]$ such that $\lim_{n\to\infty} \int_a^b g \cdot f_n \mathrm{dm} = \int_a^b g \cdot f \mathrm{dm}$ for all $g \in L^1[a, b]$.

40. Let X be the linear space of all polynomials defined on $\mathbf{R}$. For $p \in X$, define $\|p\|$ to be the sum of the absolute values of the coefficients of p. Show that this is a norm on X. For each n, define $\psi_n \colon X \to \mathbf{R}$ by $\psi_n(p) = p^{(n)}(0)$. Use the properties of the sequence $\{\psi_n\}$ in $\mathcal{L}(X, \mathbf{R})$ to show that X is not a Banach space.

CHAPTER 18

Duality for Normed Linear Spaces

Contents

18.1 Linear Functionals, Bounded Linear Functionals, and Weak Topologies . . . 355
18.2 The Hahn-Banach Theorem 361
18.3 Reflexive Banach Spaces and Weak Sequential Convergence 367
18.4 Locally Convex Topological Vector Spaces 370
18.5 The Separation of Convex Sets and Mazur's Theorem 375
18.6 The Krein-Milman Theorem 379

For a normed linear space X, we denoted the normed linear space of continuous linear real-valued functions of X by X^* and called it the **dual space** of X. In this and the following chapter, we explore properties of the mapping from $X \times X^*$ to $\mathbf{R}$ defined by

$$(x, \psi) \mapsto \psi(x) \text{ for all } x \in X, \psi \in X^*$$

to uncover analytic, geometric, and topological properties of Banach spaces. The departure point for this exploration is the Hahn-Banach Theorem. This is a theorem regarding the extension of certain linear functionals on subspaces of an unnormed linear space to linear functionals on the whole space. The elementary nature of this theorem provides it with such flexibility that in this chapter we deduce from it the following three properties of linear functionals: (i) for a normed linear space X, any bounded linear functional on a subspace of X may be extended to a bounded linear functional on all of X, without increasing its norm; (ii) for a locally convex topological vector space X, any closed, convex subset of X and point outside this subset may be separated by a closed hyperplane; and (iii) for a reflexive Banach space X, any bounded sequence in X has a weakly convergent subsequence.

18.1 LINEAR FUNCTIONALS, BOUNDED LINEAR FUNCTIONALS, AND WEAK TOPOLOGIES

In this chapter, our principle interest is in continuous linear functionals on a normed linear space. As preparation, in this first section, we begin by considering results that depend solely on linearity. Let X be a linear space. We denote by $X^\sharp$ the linear space of linear real-valued functions on X. For $\psi \in X^\sharp$, $\psi \neq 0$, and $x_0 \in X$ for which $\psi(x_0) \neq 0$, we claim that X may be expressed as the direct sum

$$X = [\ker\, \psi] \oplus \operatorname{span}[x_0], \tag{1}$$

where the **kernel** of ψ, $\ker\, \psi$, is the subspace $\{x \in X \mid \psi(x) = 0\}$. Indeed, clearly $[\ker\, \psi] \cap \operatorname{span}[x_0] = \{0\}$. On the other hand, we may write each $x \in X$ as

$$x = \left[x - \frac{\psi(x)}{\psi(x_0)} \cdot x_0\right] + \frac{\psi(x)}{\psi(x_0)} \cdot x_0 \text{ and } \psi\left(x - \frac{\psi(x)}{\psi(x_0)} \cdot x_0\right) = 0.$$

Observe that for a real number c, if x_0 belongs to X and $\psi(x_0) = c$, then

$$\psi^{-1}(c) = \{x \in X \mid \psi(x) = c\} = \ker \psi + x_0.$$

Therefore, by (1), if X is finite dimensional of dimension n and ψ is non-zero, then for each $c \in \mathbf{R}$, the level set $\psi^{-1}(c)$ is the translate of an $(n-1)$ dimensional subspace of X.

If a linear subspace X_0 of X has the property that there is some $x_0 \in X$, $x_0 \neq 0$ for which $X = X_0 \oplus \text{span}\,[x_0]$, then X_0 is said to be of **codimension** 1 in X. A translate of a subspace of codimension 1 is called a **hyperplane**.

Proposition 1 *A linear subspace X_0 of a linear space X is of codimension 1 if and only if $X_0 = \ker \psi$, for some non-zero $\psi \in X^\sharp$.*

Proof We already observed that the kernel of a non-zero linear functional is of codimension 1. Conversely, suppose X_0 is a subspace of codimension 1. Then there is a vector $x_0 \neq 0$ for which $X = X_0 \oplus \text{span}\,[x_0]$. For $\lambda \in \mathbf{R}$ and $x \in X_0$, define $\psi(x + \lambda x_0) = \lambda$. Then $\psi \neq 0$, ψ is linear and $\ker \psi = X_0$. $\square$

The following proposition tells us that the linear functionals on a linear space are plentiful.

Proposition 2 *Let Y be a linear subspace of a linear space X. Then each linear functional on Y has an extension to a linear functional on all of X. In particular, for each $x \in X, x \neq 0$, there is a $\psi \in X^\sharp$ for which $\psi(x) \neq 0$.*

Proof As we observed in the preceding chapter (see Problem 36 of Chapter 17), Y has a linear complement in X, that is, there is a linear subspace X_0 of X for which there is the direct sum decomposition

$$X = Y \oplus X_0.$$

Let η belong to $Y^\sharp$. For $x \in X$, we have $x = y + x_0$, where $y \in Y$ and $x_0 \in X_0$. Define $\eta(x) = \eta(y)$. This defines an extension of η to a linear functional on all of X.

Now let $x \neq 0$ belong to X. Define $\eta\colon \text{span}\,[x] \to \mathbf{R}$ by $\eta(\lambda x) = \lambda$. By the first part of the proof, with span $[X]$ playing the role of Y, the linear functional η has an extension to a linear functional on all of X. $\square$

We are particularly interested in linear spaces X that are normed and subspaces of $X^\sharp$ that are contained in the dual space of X, X^*, that is, linear spaces of linear functionals that are continuous with respect to the topology induced by the norm. If X is a finite dimensional normed linear space, then every linear functional on X belongs to X^* (see Problem 4). This property characterizes finite dimensional normed linear spaces.

A subset $\mathcal{B}$ of a linear space X is called a **Hamel basis** for X provided that each vector in X is expressible as a unique finite linear combination of vectors in $\mathcal{B}$. We leave it as an exercise to deduce from Zorn's Lemma that every linear space possesses a Hamel basis (see Problem 17).

Proposition 3 *Let X be a normed linear space. Then X is finite dimensional if and only if $X^\sharp = X^*$.*

Proof We leave it as an exercise to show that since all norms on a finite dimensional linear space are equivalent, all linear functionals on such spaces are bounded. Assume that X is infinite dimensional. Let $\mathcal{B}$ be a Hamel basis for X. Without loss of generality we assume the vectors in $\mathcal{B}$ are unit vectors. Since X is infinite dimensional, we may choose a countably infinite subset of $\mathcal{B}$, which we enumerate as $\{x_k\}_{k=1}^{\infty}$. For each natural number k and vector $x \in X$, define $\psi_k(x)$ to be k times the coefficient of x_k with respect to the expansion of x in the Hamel basis $\mathcal{B}$. Then each ψ_k belongs to $X^\sharp$. Therefore, the functional ψ: $X \to \mathbf{R}$ is properly defined by

$$\psi(x) = \sum_{k=1}^{\infty} k \cdot \psi_k(x) \text{ for all } x \in X$$

and belongs to $X^\sharp$. This linear functional is not bounded since each x_k is a unit vector for which $\psi(x_k) = k \cdot \psi_k(x_k) = k$. □

The following algebraic property of linear functionals is useful in establishing properties of weak topologies induced by linear functionals on a linear space.

Proposition 4 *Let X be a linear space, the functional ψ belong to $X^\sharp$, and $\{\psi_i\}_{i=1}^n$ be contained in $X^\sharp$. Then ψ is a linear combination of $\{\psi_i\}_{i=1}^n$ if and only if*

$$\bigcap_{i=1}^{n} \text{ker } \psi_i \subseteq \text{ker } \psi. \tag{2}$$

Proof It is clear that if ψ is a linear combination of $\{\psi_i\}_{i=1}^n$, then the inclusion (2) holds. We argue inductively to prove the converse. For $n = 1$, suppose (2) holds. We assume $\psi \neq 0$, for otherwise there is nothing to prove. Choose $x_0 \neq 0$ for which $\psi(x_0) = 1$. Then $\psi_1(x_0) \neq 0$ also since ker $\psi_1 \subseteq$ ker ψ. However, $X =$ ker $\psi_1 \oplus \text{span}\,[x_0]$. Therefore, if we define $\lambda_1 = 1/\psi_1(x_0)$ we see that $\psi = \lambda_1 \psi_1$. Now assume that for linear functionals on any linear space, if (2) holds for $n = k-1$, then ψ is a linear combination of $\psi_1, \ldots, \psi_{k-1}$. Suppose (2) holds for $n = k$. If $\psi_k = 0$, there is nothing to prove. So choose $x_0 \in X$ with $\psi_k(x_0) = 1$. Then $X = Y \oplus \text{span}\,[x_0]$, where $Y =$ ker ψ_k, and therefore

$$\bigcap_{i=1}^{k-1} [\text{ker } \psi_i \cap Y] \subseteq \text{ker } \psi \cap Y.$$

By the induction assumption, there are real numbers $\lambda_1, \ldots, \lambda_{k-1}$ for which

$$\psi = \sum_{i=1}^{k-1} \lambda_i \cdot \psi_i \text{ on } Y.$$

A direct substitution shows that if we define $\lambda_k = \psi(x_0) - \sum_{i=1}^{k-1} \lambda_i \cdot \psi_i(x_0)$, then

$$\psi = \sum_{i=1}^{k} \lambda_i \cdot \psi_i \text{ on } X.$$

□

Recall that for two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on a set X, we say that $\mathcal{T}_1$ is weaker than $\mathcal{T}_2$, or $\mathcal{T}_2$ is stronger than $\mathcal{T}_1$, provided that $\mathcal{T}_1 \subseteq \mathcal{T}_2$. Observe that a function on X that is continuous with respect to a topology on X, then it is also continuous with respect to any stronger topology on X but may not be continuous with respect to a weaker topology. If $\mathcal{F}$ is any collection of real-valued functions on a set X, the weak topology on X induced by $\mathcal{F}$, or the $\mathcal{F}$-weak topology on X, has been defined to be the weakest topology on X (that is, the topology with the fewest number of sets) for which each function in $\mathcal{F}$ is continuous. A base at $x \in X$ for the $\mathcal{F}$-weak topology on X comprises sets of the form

$$\mathcal{N}_{\epsilon, f_1, \ldots, f_n}(x) = \{x' \in X \mid |f_k(x') - f_k(x)| < \epsilon \text{ for } 1 \leq k \leq n\}, \tag{3}$$

where $\epsilon > 0$ and $\{f_k\}_{k=1}^n$ is a finite subcollection of $\mathcal{F}$. For a topology on a set, we know what it means for a sequence in the set to converge, with respect to the topology, to a point in the set. It is easy to see that a sequence $\{x_n\}$ in X converges to $x \in X$ with respect to the $\mathcal{F}$-weak topology if and only if

$$\lim_{n\to\infty} f(x_n) = f(x) \text{ for all } f \in \mathcal{F}. \tag{4}$$

A function on X that is continuous with respect to the $\mathcal{F}$-weak topology is called $\mathcal{F}$-weakly continuous. Similarly, we have $\mathcal{F}$-weakly open sets, $\mathcal{F}$-weakly closed sets, and $\mathcal{F}$-weakly compact sets.

For a linear space X, it is natural and very useful to consider weak topologies induced on X by linear subspaces W of $X^\sharp$.

Proposition 5 *Let X be a linear space and W a subspace of $X^\sharp$. Then a linear functional $\psi\colon X \to \mathbf{R}$ is W-weakly continuous if and only if it belongs to W.*

Proof By the definition of the W-weak topology, each linear functional in W is W-weakly continuous. It remains to prove the converse. Suppose the linear functional $\psi\colon X \to \mathbf{R}$ is W-weakly continuous. By the continuity of ψ at 0, there is a neighborhood $\mathcal{N}$ of 0 for which $|\psi(x)| = |\psi(x) - \psi(0)| < 1$ if $x \in \mathcal{N}$. There is a neighborhood in the base for the W-topology at 0 that is contained in $\mathcal{N}$. Choose $\epsilon > 0$ and $\psi_1, \ldots, \psi_n$ in W for which $\mathcal{N}_{\epsilon, \psi_1, \ldots, \psi_n} \subseteq \mathcal{N}$. Thus

$$|\psi(x)| < 1 \text{ if } |\psi_k(x)| < \epsilon \text{ for all } 1 \leq k \leq n.$$

By the linearity of ψ and the ψ_k's, we have the inclusion $\cap_{k=1}^n \ker \psi_k \subseteq \ker \psi$. According to Proposition 4, ψ is a linear combination of $\psi_1, \ldots, \psi_n$. Therefore, since W is a linear space, ψ belongs to W. □

The above proposition establishes a one-to-one correspondence between linear subspaces of $X^\sharp$ and weak topologies on X induced by such subspaces.

Definition *Let X be a normed linear space. The weak topology induced on X by the dual space X^* is called the* **weak topology** *on X.*

It follows from linearity that a base at $x \in X$ for the weak topology on X is comprised of sets of the form

$$\mathcal{N}_{\epsilon, \psi_1, \ldots, \psi_n}(x) = \{x' \in X \mid |\psi_k(x' - x)| < \epsilon \text{ for } 1 \leq k \leq n\}, \tag{5}$$

where $\epsilon > 0$ and $\{\psi_k\}_{k=1}^n$ is a finite subcollection of X^*. For topological concepts with respect to the weak topology, we use the adjective "weakly": so we have weakly compact sets, weakly open sets, etc. Thus a sequence $\{x_n\}$ in X converges weakly to $x \in X$ if and only if

$$\lim_{n\to\infty} \psi(x_n) = \psi(x) \text{ for all } \psi \in X^*. \tag{6}$$

Frequently, for a normed linear space, we call the topology induced by the norm the **strong topology** on X. For X a normed linear space and W a subspace of X^*, there is the following inclusion among topologies on X:

$$W\text{-weak topology on } X \subseteq \text{weak topology on } X \subseteq \text{strong topology on } X.$$

It follows from Proposition 5 that the W-weak topology coincides with the weak topology if and only if $W = X^*$. Furthermore, the weak topology coincides with the strong topology if and only if X is finite dimensional (see Problem 6). If no adjective is attached to a topological concept associated with a normed linear space, it is implicitly assumed that the reference topology is the strong topology.

For normed linear spaces that are dual spaces, there is a third important topology on the space besides the weak and the strong topologies. Indeed, for a normed linear space X and $x \in X$ we define the functional $J(x)\colon X^* \to \mathbf{R}$ by

$$J(x)[\psi] = \psi(x) \text{ for all } \psi \in X^*.$$

It is clear that the **evaluation functional** $J(x)$ is linear and is bounded on X^* with $\|J(x)\| \leq \|x\|$. Moreover, the operator $J\colon X \to (X^*)^*$ is linear and therefore $J(X)$ is a linear subspace of $(X^*)^*$.

Definition *Let X be a normed linear space. The weak topology on X^* induced by $J(X) \subseteq (X^*)^*$ is called the* **weak-$*$ topology** *on X^*.*

It follows from linearity that a base at $\psi \in X^*$ for the weak-$*$ topology on X^* comprises sets of the form

$$\mathcal{N}_{\epsilon,\, x_1,\, \ldots,\, x_n}(\psi) = \{\psi' \in X^* \mid |(\psi' - \psi)(x_k)| < \epsilon \text{ for } 1 \leq k \leq n\}, \tag{7}$$

where $\epsilon > 0$ and $\{x_k\}_{k=1}^n$ is a finite subset of X. A subset of X^* that is open with respect to the weak-$*$ topology is said to be weak-$*$ open. We use similarly terminology for other topological concepts. Observe that a sequence $\{\psi_n\}$ in X^* is weak-$*$ convergent to $\psi \in X^*$ if and only if

$$\lim_{n\to\infty} \psi_n(x) = \psi(x) \text{ for all } x \in X. \tag{8}$$

Therefore, weak-$*$ convergence in X^* is exactly pointwise convergence. For a normed linear space X, the strong, weak, and weak-$*$ topologies on X^* are related by the following inclusions:

$$\text{weak-}*\text{ topology on } X^* \subseteq \text{weak topology on } X^* \subseteq \text{strong topology on } X^*.$$

Definition *Let X be a normed linear space. The linear operator $J\colon X \to (X^*)^*$ defined by*

$$J(x)[\psi] = \psi(x) \text{ for all } x \in X, \psi \in X^*$$

is called the **natural embedding** *of X into $(X^*)^*$. The space X is said to be* **reflexive** *provided that $J(X) = (X^*)^*$.*

It is customary to denote $(X^*)^*$ by X^{**} and call X^{**} the **bidual** of X.

Proposition 6 *A normed linear space X is reflexive if and only if the weak and weak-$*$ topologies on X^* are the same.*

Proof Clearly if X is reflexive, then the weak and weak-$*$ topologies on X^* are the same. Conversely, suppose these two topologies are the same. Let $\Psi\colon X^* \to \mathbf{R}$ be a continuous linear functional. By definition of the weak topology, Ψ is continuous with respect to the weak topology on X^*. Therefore, it is continuous with respect to the weak-$*$ topology. It follows from Proposition 5 that Ψ belongs to $J(X)$. Therefore $J(X) = X^{**}$. $\square$

At present, we are not justified in calling $J\colon X \to X^{**}$ an "embedding" since we have not shown that J is one-to-one. In fact, we have not even shown that on a general normed linear space X there are any non-zero bounded linear functionals. We need a variation of Proposition 2 for linear functionals that are bounded. The forthcoming Hahn-Banach Theorem will provide this variation and, moreover, show that J is an isometry. Of course, we have already studied the dual spaces of some particular normed linear spaces. For instance, if $(X, \mathcal{M}, \mu)$ is a σ-finite general measure space and $1 \leq p < \infty$, the Riesz Representation Theorem, which we proved in Chapter 12, characterizes the dual of $L^p(X, \mu)$, and the Kanterovich Representation Theorem characterizes the dual of $L^\infty(X, \mu)$.

PROBLEMS

1. (i) Use Zorn's Lemma to show that every linear space has a Hamel basis.
   (ii) Show that any Hamel basis for an infinite dimensional Banach space must be uncountable.
   (iii) Let X be the linear space of all polynomials defined on $\mathbf{R}$. Show that there is not a norm on X with respect to which X is a Banach space.
2. Verify the two direct substitution assertions in the proof of Proposition 4.
3. Let X_0 be a codimension 1 subspace of a normed linear space X. Show that X_0 is closed with respect to the strong topology if and only if $X_0 = \ker \psi$ for some $\psi \in X^*$.
4. Show that if X is a finite dimensional normed linear space, then every linear functional on X is continuous.
5. Let X be a finite dimensional normed linear space of dimension n. Let $\{e_1, \ldots, e_n\}$ be a basis for X. For $1 \leq i \leq n$, define $\psi_i \in X^*$ by $\psi_i(x) = x_i$ for $x = x_i e_i + \cdots + x_n e_n \in X$. Show that $\{\psi_1, \ldots, \psi_n\}$ is a basis for X^*. Thus $\dim X^* = n$.
6. Let X be a finite dimensional linear space. Show that the weak and strong topologies on X are the same.

7. Show that every non-empty weakly open subset of an infinite dimensional normed linear space is unbounded with respect to the norm.
8. Let X be a finite dimensional space. Show that the natural embedding $J: X \to X^{**}$ is one-to-one. Then use Problem 5 to show that $J: X \to X^{**}$ is onto, so X is reflexive.
9. For a vector $v \neq 0$ in Euclidean space $\mathbf{R}^n$, explicitly exhibit a linear functional $\psi: \mathbf{R}^n \to \mathbf{R}$ for which $\psi(v) = 1$.
10. For a sequence $\{x_n\} \neq 0$ in ℓ^2, explicitly exhibit a continuous linear functional $\psi : \ell^2 \to \mathbf{R}$ for which $\psi(\{x_n\}) = 1$.
11. For a function $f \neq 0$ in $L^p[a, b], 1 \leq p \leq \infty$, explicitly exhibit a continuous linear functional $\psi: L^p[a, b] \to \mathbf{R}$ for which $\psi(f) = 1$.
12. Consider $C[a, b]$ with the maximum norm. For a function $f \neq 0$ in $C[a, b]$, explicitly exhibit a continuous linear functional $\psi: C[a, b] \to \mathbf{R}$ for which $\psi(f) = 1$.
13. For $1 < p < \infty$, let Y be a closed subspace of $L^p[a, b]$ of codimension 1. Show that there is function $g \in L^q[a, b]$, where q is the conjugate of p, for which
$$Y = \left\{ f \in L^p[a, b] \;\middle|\; \int_{[a, b]} f \cdot g \, dm = 0 \right\}.$$
14. Let X be a normed linear space and ψ belong to $X^\sharp \sim X^*$. Show that $\ker \psi$ is dense, with respect to the strong topology, in X.
15. Let X be the normed linear space of polynomials restricted to $[a, b]$. For $p \in X$, define $\psi(p)$ to be the sum of the coefficients of p. Show that ψ is linear. Is ψ continuous if X has the topology induced by the maximum norm?
16. Let X be the normed linear space of sequences of real numbers that have only a finite number of non-zero terms. For $x = \{x_n\} \in X$, define $\psi(x) = \sum_{n=1}^{\infty} x_n$. Show that ψ is linear. Is ψ continuous if X has the topology induced by the ℓ^∞ norm?
17. Let X be a linear space. A subset E of X is said to be linearly independent provided that each $x \in E$ fails to be a finite linear combination of points in $E \sim \{x\}$. Define $\mathcal{F}$ to be the collection of non-empty subsets of X that are linearly independent. Order $\mathcal{F}$ by set inclusion. Apply Zorn's Lemma to conclude that X has a Hamel basis.
18. Provide an example of a discontinuous linear operator T from a normed linear space X to a normed linear space Y for which T has a closed graph. (Suggestion: Let ψ be a discontinuous linear functional on a normed linear space X and $Y = \{y \in X \times \mathbf{R} \,|\, y = \langle x, \psi(x)\rangle\}$, the graph of ψ.) Does this contradict the Closed Graph Theorem?

18.2 THE HAHN-BANACH THEOREM

Definition *A non-negative functional $p: X \to [0, \infty)$ on a linear space X is said to be* **positively homogeneous** *provided that*
$$p(\lambda x) = \lambda p(x) \text{ for all } x \in X, \lambda > 0,$$
and said to be **subadditive** *provided that*
$$p(x + y) \leq p(x) + p(y) \text{ for all } x, y \in X.$$

Any norm on a linear space is both subadditive (the triangle inequality) and positively homogeneous.

The Hahn-Banach Lemma *Let p be a positively homogeneous, subadditive functional on the linear space X and Y a subspace of X on which there is defined a linear functional ψ for which*

$$\psi \leq p \text{ on } Y.$$

Let z be a point of X that lies outside of Y. Then ψ can be extended to a linear functional ψ on span $[Y+z]$ *for which*

$$\psi \leq p \text{ on } \operatorname{span}[Y+z].$$

Proof Since every vector in span $[Y+z]$ may be written uniquely as $y+\lambda z$, for $y \in Y$ and $\lambda \in \mathbf{R}$, it is sufficient to find a number $\psi(z)$ with the property that

$$\psi(y)+\lambda\psi(z) \leq p(y+\lambda z) \text{ for all } y \in Y \text{ and } \lambda \in \mathbf{R}. \tag{9}$$

Indeed, for such a number $\psi(z)$, define $\psi(y+\lambda z) = \psi(y)+\lambda\psi(z)$ for all y in Y and $\lambda \in \mathbf{R}$ to obtain the required extension.

For any vectors $y_1, y_2 \in Y$, since ψ is linear, $\psi \leq p$ on Y and p is subadditive,

$$\psi(y_1)+\psi(y_2) = \psi(y_1+y_2) \leq p(y_1+y_2) = p((y_1-z)+(y_2+z)) \leq p(y_1-z)+p(y_2+z),$$

and therefore

$$\psi(y_1)-p(y_1-z) \leq -\psi(y_2)+p(y_2+z).$$

As we vary y_1 and y_2 among all vectors in Y, any number on the left-hand side of this inequality is no greater than any number on the right. By the completeness of $\mathbf{R}$, if we define $\psi(z)$ to be the supremum of the numbers on the left-hand side of this inequality, then $\psi(z) \in \mathbf{R}$. Furthermore, for any $y \in Y$, $\psi(y)-p(y-z) \leq \psi(z)$, by the choice of $\psi(z)$ as an upper bound and $\psi(z) \leq -\psi(y)+p(y+z)$ by the choice of $\psi(z)$ as the least upper bound. Therefore

$$\psi(y)-p(y-z) \leq \psi(z) \leq -\psi(y)+p(y+z) \text{ for all } y \in Y. \tag{10}$$

Let y belong to Y. For $\lambda > 0$, in the inequality $\psi(z) \leq -\psi(y)+p(y+z)$, replace y by y/λ, multiply each side by λ, and use the positive homogeneity of both p and ψ to obtain the desired inequality (9). For $\lambda < 0$, in the inequality $\psi(y)-p(y-z) \leq \psi(z)$, replace y by $-y/\lambda$, multiply each side by $-\lambda$, and once more use positive homogeneity to obtain the desired inequality (9). Therefore, (9) holds if the number $\psi(z)$ is chosen so that (10) holds. $\square$

The Hahn-Banach Theorem *Let p be a positively homogeneous, subadditive functional on a linear space X and Y a subspace of X on which there is defined a linear functional ψ for which*

$$\psi \leq p \text{ on } Y.$$

Then ψ may be extended to a linear functional ψ on all of X for which $\psi \leq p$ on all of X.

Proof Consider the family $\mathcal{F}$ of all linear functionals η defined on a subspace Y_η of X for which $Y \subseteq Y_\eta$, $\eta = \psi$ on Y, and $\eta \leq p$ on Y_η. This particular family $\mathcal{F}$ of extensions of ψ is partially ordered by defining $\eta_1 \prec \eta_2$ provided that $Y_{\eta_1} \subseteq Y_{\eta_2}$ and $\eta_1 = \eta_2$ on Y_{η_1}.

Let $\mathcal{F}_0$ be a totally ordered subfamily of $\mathcal{F}$. Define Z to be the union of the domains of the functionals in $\mathcal{F}_0$. Since $\mathcal{F}_0$ is totally ordered, any two such domains are contained in just one of them and therefore, since each domain is a linear subspace of X, so is Z. For $z \in Z$, choose $\eta \in \mathcal{F}_0$ such that $z \in Y_\eta$: define $\eta^*(z) = \eta(z)$. Then, again by the total ordering of $\mathcal{F}_0$, η^* is a properly defined linear functional on Z. Observe that $\eta^* \leq p$ on Z, $Y \subseteq Z$ and $\eta^* = \psi$ on Y, since each functional in $\mathcal{F}_0$ has these three properties. Thus, $\eta \prec \eta^*$ for all $\eta \in \mathcal{F}_0$. Therefore, every totally ordered subfamily of $\mathcal{F}$ has an upper bound. Hence, by Zorn's Lemma, $\mathcal{F}$ has a maximal member ψ_0. Let the domain of ψ_0 be Y_0. By definition, $Y \subseteq Y_0$ and $\psi_0 \leq p$ on Y_0. It follows from the Hahn-Banach Lemma that this maximal extension ψ_0 is defined on all of X. □

Theorem 7 *Let X_0 be a linear subspace of a normed linear space X. Then each bounded linear functional ψ on X_0 has an extension to a bounded linear functional on all of X that has the same norm as ψ. In particular, for each $x \in X$, $x \neq 0$, there is a $\psi \in X^*$ for which*

$$\psi(x) = ||x|| \text{ and } ||\psi||_* = 1. \tag{11}$$

Proof Let $\psi\colon X_0 \to \mathbf{R}$ be linear and bounded. Define

$$M = \sup\left\{|\psi(x)| \mid x \in X_0, ||x|| \leq 1\right\}.$$

Define $p\colon X \to \mathbf{R}$ by

$$p(x) = M \cdot \|x\| \text{ for all } x \in X.$$

The functional p is subadditive and positively homogeneous. By the definition of M,

$$\psi \leq p \text{ on } X_0.$$

By the Hahn-Banach Theorem, ψ may be extended to a continuous linear functional ψ on all of X and $\psi(x) \leq p(x) = M||x||$ for all $x \in X$. Replacing x by $-x$, it follows that $|\psi(x)| \leq p(x) = M||x||$ for all $x \in X$ and therefore the extension of ψ to all of X has the same norm as $\psi\colon X_0 \to \mathbf{R}$.

Now let $x \neq 0$ belong to X. Define $\eta\colon \text{span}\,[x] \to \mathbf{R}$ by $\eta(\lambda x) = \lambda \cdot \|x\|$. Observe that $\|\eta\|_* = 1$. Note that $\eta(x) \neq 0$. By the first part of the proof, the functional η has an extension to a bounded linear functional on all of X that also has norm 1. □

Example Let x_0 belong to the closed, bounded interval $[a, b]$. Define

$$\psi(f) = f(x_0) \text{ for all } f \in C[a, b].$$

We consider $C[a, b]$ as a subspace of $L^\infty[a, b]$ (see Problem 29). It follows from the preceding theorem that ψ has an extension to a bounded linear functional $\psi\colon L^\infty[a, b] \to \mathbf{R}$. No such extension has ever been explicitly presented.

Example Define the positively homogeneous, subadditive functional p on ℓ^∞ by

$$p(\{x_n\}) = \limsup\{x_n\} \text{ for all } \{x_n\} \in \ell^\infty.$$

Let $c \subseteq \ell^\infty$ be the subspace of convergent sequences. Define L on c by

$$L(\{x_n\}) = \lim_{n\to\infty} x_n \text{ for all } \{x_n\} \in c.$$

Since L is linear and $L \leq p$ on c, L has an extension to a linear functional L on ℓ^∞ for which $L \leq p$ on ℓ^∞. Any such extension is called a **Banach limit**.

In the preceding chapter, we considered whether a subspace X_0 of a Banach space X has a closed linear complement in X. The following corollary tells us it does if X_0 is finite dimensional.

Corollary 8 *Let X be a normed linear space. If X_0 is a finite dimensional subspace of X, then there is a closed linear subspace X_1 of X for which $X = X_0 \oplus X_1$.*

Proof Let $\{e_k\}_{k=1}^n$ be a basis for X_0. Define $\psi_k\colon X_0 \to \mathbf{R}$ by $\psi_k(\sum_{i=1}^n \lambda_i \cdot e_i) = \lambda_k$ for $1 \leq k \leq n$. Since X_0 is finite dimensional, the ψ_k's are continuous. According to Theorem 7, each ψ_k has an extension to a continuous functional ψ'_k on all of X. Therefore each ψ'_k has a closed kernel so that the subspace $X_1 = \cap_{k=1}^n \ker \psi'_k$ also is closed. It is easy to check that $X = X_0 \oplus X_1$. □

Corollary 9 *Let X be a normed linear space. Then the natural embedding $J\colon X \to X^{**}$ is an isometry.*

Proof Let x belong to X. Observe that by the definition of the norm on the dual space

$$|\psi(x)| \leq \|\psi\|_* \cdot \|x\| \text{ for all } \psi \in X^*.$$

Thus

$$|J(x)(\psi)| \leq \|x\| \cdot \|\psi\|_* \text{ for all } \psi \in X^*.$$

Therefore, $J(x)$ is bounded and $\|J(x)\|_* \leq \|x\|$. On the other hand, according to Theorem 7, there is a $\psi \in X^*$ for which $\|\psi\|_* = 1$ and $J(x)(\psi) = \|x\|$. Therefore $\|x\| \leq \|J(x)\|_*$. We conclude that J is an isometry. □

Theorem 10 *Let X_0 be a subspace of the normed linear space X. Then a point x in X belongs to the closure of X_0 if and only if whenever a continuous functional $\psi \in X^*$ vanishes on X_0, it also vanishes at x.*

Proof It is clear by continuity that if a continuous functional vanishes on X_0 it also vanishes on the closure of X_0. To prove the converse, let x_0 belong to $X \sim \overline{X_0}$. We must show that there is a $\psi \in X^*$ that vanishes on X_0 but $\psi(x_0) \neq 0$. Define $Z = \overline{X_0} \oplus [x_0]$ and $\psi\colon Z \to \mathbf{R}$ by

$$\psi(x + \lambda x_0) = \lambda \text{ for all } x \in \overline{X_0} \text{ and } \lambda \in \mathbf{R}.$$

We claim that ψ is bounded. Indeed, since $\overline{X_0}$ is closed, its complement is open. Thus, there is an $r > 0$ for which $\|u - x_0\| \geq r$ for all $u \in \overline{X_0}$. Thus, for $x \in \overline{X_0}$ and $\lambda \in \mathbf{R}$,

$$\|x + \lambda x_0\| = |\lambda| \|(-1/\lambda \cdot x) - x_0\| \geq |\lambda| \cdot r.$$

From this it follows that $\psi\colon Z \to \mathbf{R}$ is bounded with $\|\psi\|_* \leq 1/r$. Theorem 7 tells us that ψ has an extension to a bounded linear functional on all of X. This extension belongs to X^*, vanishes on X_0, and yet $\psi(x_0) \neq 0$. □

We leave the proof of the following corollary as an exercise.

Corollary 11 *Let $\mathcal{S}$ be a subset of the normed linear space X. Then the linear span of $\mathcal{S}$ is dense in X if and only if whenever $\psi \in X^*$ vanishes on $\mathcal{S}$, then $\psi = 0$.*

Theorem 12 *Let X be a normed linear space. Then every weakly convergent sequence in X is bounded. Moreover, if $\{x_n\} \to x$ weakly in X, then*

$$\|x\| \leq \liminf \|x_n\|. \tag{12}$$

Proof Let $\{x_n\} \to x$ weakly in X. Then $\{J(x_n)\colon X^* \to \mathbf{R}\}$ is a sequence of bounded functionals that converges pointwise to $J(x)\colon X^* \to \mathbf{R}$. In Chapter 8, we proved, using a Uniform Boundedness Theorem, that a weakly convergent sequence is bounded. Now

$$|\psi(x_n)| \leq \|\psi\| \cdot \|x_n\| = \|x_n\| \text{ for all } n.$$

Moreover, $|\{\psi(x_n)\}|$ converges to $|\psi(x)| = \|x\|$. Therefore

$$\|x\| = \lim_{n\to\infty} |\psi(x_n)| \leq \liminf \|x_n\|.$$

□

The Hahn-Banach Theorem has a rather humble nature. The only mathematical concepts needed for its statement are linear spaces and linear, subadditive, and positively homogeneous functionals. Besides Zorn's Lemma, its proof relies on nothing more than the rudimentary properties of the real numbers. Nevertheless, often by making a clever choice of the functional p, this theorem permits us to create basic analytical, geometric, and topological tools for functional analysis. We established Theorem 7 by applying the Hahn-Banach Theorem with the functional p chosen to be a multiple of the norm. In Section 5 of this chapter, we use the Hahn-Banach Theorem with p the so-called gauge functional associated with a convex set to separate disjoint, convex subsets of a linear space by a hyperplane. In the next chapter, we use the natural embedding J of a normed linear space into its bidual to prove that the closed unit ball of a Banach space X is weakly sequentially compact if and only if X is reflexive[1].

PROBLEMS

19. Let X be a linear subspace of $C[0, 1]$ that is closed as a subset of $L^2[0, 1]$. Verify the following assertions to show that X has finite dimension. The sequence $\{f_n\}$ belongs to X.
    (i) Show that X is a closed subspace of $C[0, 1]$.
    (ii) Show that there is a constant $M \geq 0$ such that for all $f \in X$ we have $\|f\|_2 \leq \|f\|_\infty$ and $\|f\|_\infty \leq M \cdot \|f\|_2$.
    (iii) Show that for each $y \in [0, 1]$, there is a function k_y in L^2 such that for each $f \in X$ we have $f(y) = \int_0^1 k_y(x) f(x)\, dx$.
    (iv) Show that if $\{f_n\} \to f$ weakly in L^2, then $\{f_n\} \to f$ pointwise on $[0,1]$.

[1]More applications of the Hahn-Banach Theorem may be found in Peter Lax's *Functional Analysis* [Lax02].

(v) Show $\{f_n\} \to f$ weakly in L^2, then $\{f_n\}$ is bounded (in what sense?), and hence $\{f_n\} \to f$ strongly in L^2 by the Dominated Convergence Theorem.

(vi) Conclude that X, when normed by $\|\cdot\|_2$, has a compact closed unit ball and therefore, by Riesz's Theorem, is finite dimensional.

20. Let X be a normed linear space, ψ belong to X^*, and $\{\psi_n\}$ be in X^*. Show that if $\{\psi_n\}$ converges weak-$*$ to ψ, then
$$\|\psi\|_* \leq \liminf \|\psi_n\|_*.$$

21. Let $X = \mathbf{R}^n$ be normed with the Euclidean norm, Y a subspace of X, and $\psi\colon Y \to \mathbf{R}$ a linear functional. Define $Y^\perp$ to be the linear subspace of $\mathbf{R}^n$ consisting of vectors orthogonal to Y. Then $\mathbf{R}^n = Y \oplus Y^\perp$. For $x = y + y'$, $y \in Y$, $y' \in Y^\perp$, define $\psi(x) = \psi(y)$. Show that this properly defines $\psi \in (\mathbf{R}^n)^*$, is an extension of ψ on Y, and has the same norm as $\psi|_Y$.

22. Let $X = L^p = L^p[a, b], 1 < p < \infty$ and m be Lebesgue measure. For $f \neq 0$ in L^p, define
$$\psi(h) = \frac{1}{\|f\|_p^{p-1}} \int_{[a,\, b]} \operatorname{sgn}(f) \cdot |f|^{p-1} \cdot h\, dm \text{ for all } h \in L^p.$$
Use Hölder's Inequality to show that $\psi \in (L^p)^*, ||\psi||_* = 1$ and $\psi(f) = \|f\|_p$.

23. For each point x in a normed linear space X, show that
$$\|x\| = \sup\left\{\psi(x) \;\middle|\; \psi \in X^*, \|\psi\|_* \leq 1\right\}.$$

24. Let X be a normed linear space and Y a closed subspace of X. Show that for each $x_0 \in X \sim Y$, there is a $\psi \in X^*$ such that
$$||\psi||_* = 1, \psi = 0 \text{ on } Y \text{ and } \psi(x_0) = d, \text{ where } d = \operatorname{dist}(x_0, Y) = \inf\{\|x_0 - y\| \mid y \in Y\}.$$

25. Let Y be a linear subspace of a normed linear space X and z be a vector in X. Show that
$$\operatorname{dist}(z, Y) = \sup\{\psi(z) \mid \|\psi\|_* \leq 1, \psi = 0 \text{ on } Y\}.$$

26. Let X be a vector space. A subset C of X is called a **cone** provided that $x + y \in C$ and $\lambda x \in C$ whenever x, y belong to C and $\lambda > 0$. Define a partial order in X by defining $x \prec y$ to mean $y - x \in C$. A linear functional f on X is said to be positive (with respect to the cone C) provided that $f \geq 0$ on C. Let Y be any subspace of X with the property that for each $x \in X$ there is a $y \in Y$ with $x \prec y$. Suppose f is a linear functional on Y that is positive with respect to the cone $C \cap Y$. Show that f may be extended to a linear functional on X that is positive with respect to C. (Suggestion: Adapt the Hahn-Banach Lemma and use Zorn's Lemma to find a maximal extension.)

27. Let X_0 be a subset of a metric space X. Use the Tietze Extension Theorem to show that every continuous real-valued function on X_0 has a continuous extension to all of X if and only if X_0 is closed. Does this contradict Theorem 7?

28. Let F be closed subset of a metric space (X, ρ). Show that a point $x \in X$ belongs to F if and only if every continuous real-valued function on X that vanishes on F also vanishes at x. Can this be used to prove Theorem 10?

29. Let $[a, b]$ be a closed, bounded interval of real numbers and consider $L^\infty[a, b]$, now formally considered as the collection of equivalence classes for the relation of pointwise equality almost everywhere among essentially bounded functions. Let X be the subspace of $L^\infty[a, b]$ comprising those equivalence classes that contain a continuous function. Show that such an equivalence class contains exactly one continuous function. Thus X is linearly isomorphic to $C[a, b]$ and therefore, modulo this identification, we may consider $C[a, b]$ to be a linear subspace $L^\infty[a, b]$. Show that $C[a, b]$ is a closed subspace of the Banach space $L^\infty[a, b]$.
30. Define ψ: $C[a, b] \to \mathbf{R}$ by $\psi(f) = f(a)$ for all $f \in C[a, b]$. Use Theorem 7 to extend ψ to a continuous linear functional on all of $L^\infty[a, b]$ (see the preceding problem). Show that there is no function $h \in L^1[a, b]$ for which

$$\psi(f) = \int_a^b h \cdot f \, dm \text{ for all } f \in L^\infty[a, b].$$

18.3 REFLEXIVE BANACH SPACES AND WEAK SEQUENTIAL CONVERGENCE

Theorem 13 *If X is a normed linear space that has a separable dual space X^*, then X also is separable.*

Proof Since X^* is separable, so is its closed unit sphere $S^* = \{\psi \in X^* \mid \|\psi\| = 1\}$. Let $\{\psi_n\}_{n=1}^\infty$ be a countable dense subset of S^*. For each index n, choose $x_n \in X$ for which

$$\|x_n\| = 1 \text{ and } \psi_n(x_n) > 1/2.$$

Define X_0 to be the closed linear span of the set $\{x_n \mid 1 \leq n < \infty\}$. Then X_0 is separable since finite linear combinations, with rational coefficients, of the x_n's is a countable set that is dense in X_0. We claim $X_0 = X$. Indeed, otherwise, by Theorem 10, we may choose $\psi^* \in X^*$ for which

$$\|\psi^*\| = 1 \text{ and } \psi^* = 0 \text{ on } X_0.$$

Since $\{\psi_n\}_{n=1}^\infty$ is dense in S^*, there is a natural number n_0 for which $\|\psi^* - \psi_{n_0}\| < 1/2$. Therefore

$$|(\psi_{n_0} - \psi^*)(x_{n_0})| \leq \|\psi_{n_0} - \psi^*\| \cdot \|x_{n_0}\| < 1/2 \text{ and yet } (\psi_{n_0} - \psi^*)(x_{n_0}) = \psi_{n_0}(x_{n_0}) > 1/2.$$

From this contradiction we conclude that X is separable. □

Corollary 14 *A reflexive Banach space is separable if and only if its dual is separable.*

Proof Let X be a Banach space. The preceding theorem tells us that if X^* is separable so is X, irrespective of any reflexivity assumption. Now assume that X is reflexive and separable. Thus $J(X) = X^{**} = (X^*)^*$ is separable since J is an isometry. According to the preceding theorem, with X replaced by X^*, X^* is separable. □

Proposition 15 *A closed subspace of a reflexive Banach space is reflexive.*

Proof Let X_0 be a closed subspace of reflexive Banach space X. Define J to be the natural embedding of X in its bidual X^{**}. Let J_0 be the natural embedding of X_0 in its bidual X_0^{**}. To show that J_0 is onto, let S belong to X_0^{**}. Define $S' \in X^{**}$ by

$$S'(\psi) = S(\psi|_{X_0}) \text{ for all } \psi \in X^*.$$

Then $S'\colon X_0^* \to \mathbf{R}$ is linear and it is bounded with $\|S'\| \le \|S\|$. By the reflexivity of X, there is an $x_0 \in X$ for which $S' = J(x_0)$. But if $\psi \in X^*$ vanishes on X_0, then $S'(\psi) = 0$, so that

$$\psi(x_0) = J(x_0)[\psi] = S'(\psi) = 0.$$

Theorem 10 tells us that x_0 belongs to X_0. Therefore $S = J_0(x_0)$. □

We record again Helly's Theorem, which we proved in Chapter 8.

Theorem 16 (Helly's Theorem) *If X is a separable normed linear space and $\{\psi_n\}$ is a bounded sequence in X^*, then a subsequence of $\{\psi_n\}$ converges pointwise on X to $\psi \in X^*$, that is, $\{\psi_n\}$ has a subsequence that converges to ψ with respect to the weak-$*$ topology.*

Theorem 17 *Let X be a reflexive Banach space. Then every bounded sequence in X has a weakly convergent subsequence.*

Proof Let $\{x_n\}$ be a bounded sequence in X. Define X_0 to be the closure of the linear span of the set $\{x_n \,|\, n \in \mathbf{N}\}$. Then X_0 is separable since finite linear combinations of the x_n's, with rational coefficients, is a countable set that is dense in X_0. Of course X_0 is closed. Proposition 15 tells us that X_0 is reflexive. Let J_0 be the natural embedding of X_0 in its bidual X_0^{**}. It follows from Corollary 14 that X_0^* also is separable. Then $\{J_0(x_n)\}$ is a bounded sequence of bounded linear functionals on a separable Banach space X_0^*. According to Helly's Theorem, a subsequence $\{J_0(x_{n_k})\}$ converges weak-$*$ to $S \in (X_0^*)^*$. Since X_0 is reflexive, there is some $x_0 \in X_0$ for which $S = J_0(x_0)$. Since every functional in X^* restricts to a functional in X_0^*, the weak-$*$ convergence of $\{J_0(x_{n_k})\}$ to $J_0(x_0)$ means precisely that $\{x_{n_k}\}$ converges weakly to x_0. □

Corollary 18 *Let X be a reflexive Banach space. Then every $\psi \in X^*$ takes a maximum and a minimum value on B, the closed unit ball of X.*

Proof Let ψ belong to X^*. It suffices to show that ψ takes a maximum value on B. The supremum of the functional values of ψ on B is $\|\psi\|$. Choose a sequence $\{x_n\}$ in B for which $\lim_{n\to\infty} \psi(x_n) = \|\psi\|$. In view of Theorem 17, we may assume that $\{x_n\}$ converges weakly to x_0. According to (12), x_0 belongs to B. Since

$$\lim_{n\to\infty} \psi(x_n) = \psi(x_0),$$

ψ takes a maximum value on B at x_0. □

Theorem 17 makes it interesting to identify which of the classical Banach spaces are reflexive. It suffices to show that ψ takes a maximum value.

Proposition 19 *Let $[a, b]$ be a closed, bounded interval of real numbers. Then $C[a, b]$, normed with the maximum norm, is not reflexive.*

Proof Assume that $[a, b] = [0, 1]$. For $x \in [0, 1]$, define the evaluation functional ψ_x: $C[0, 1] \to \mathbf{R}$ by $\psi_x(f) = f(x)$. Then ψ_x is a bounded linear functional on $C[0, 1]$. Therefore, if $\{f_n\}$ converges weakly to f in $C[0, 1]$, then $\{f_n\} \to f$ pointwise on $[0, 1]$. For a natural number n, define $f_n(x) = x^n$ for $x \in [0, 1]$. Then $\{f_n\}$ converges pointwise to a function f that is not continuous. Therefore, no subsequence can converge pointwise to a continuous function and hence no subsequence can converge weakly to a function in $C[0, 1]$. It follows from Theorem 17 that $C[0, 1]$ fails to be reflexive. $\square$

Proposition 20 *If $(X, \mathcal{M}, \mu)$ is a σ-finite measure space and $1 < p < \infty$, then $L^p(X, \mu)$ is reflexive.*

Proof Let q be the conjugate to p. According to the Riesz Representation Theorem, the Riesz representation operator $\mathcal{R}$: $(L^p(X, \mu))^* \to L^q(X, \mu)$ is one-to one and onto. Consider the inverse $\mathcal{S}$: $L^q(X, \mu) \to (L^p(X, \mu))^*$ of $\mathcal{R}$. Then

$$\mathcal{S}(g)[f] = \int_W g \cdot f \, d\mu \text{ for all } g \in L^q(X, \mu), f \in L^p(X, \mu).$$

Let T be a bounded linear functional on $(L^p(X, \mu))^*$. We must show that there is a function $f \in L^p(X, \mu)$ for which $T = J(f)$, that is, since $\mathcal{S}$ is onto,

$$T(\mathcal{S}(g)) = J(f)[\mathcal{S}(g)] = \mathcal{S}(g)[f] = \int_X g \cdot f \, d\mu \text{ for all } g \in L^q(X, \mu). \tag{13}$$

However, the composition $T \circ \mathcal{S}$ is a bounded linear functional on $L^q(X, \mu)$. The Riesz Representation Theorem, with p and q interchanged, tells us that there is a function $f \in L^p(X, \mu)$ which represents the bounded linear functional $T \circ \mathcal{S}$, that is, (13) holds. $\square$

In general, $L^1(X, \mu)$ is not reflexive. Consider Lebesgue measure on $[0, 1]$. Observe that $L^1[0, 1]$ is separable, while $(L^1[0, 1])^*$ is not separable since it is isomorphic to $L^\infty[0, 1]$, which is not separable. It follows from Corollary 14 that $L^1[0, 1]$ is not reflexive. Observe that $C[0, 1]$ is a closed subspace of $L^\infty[0, 1]$ (see Problem 29). By Proposition 19, $C[0, 1]$ is not reflexive and therefore Proposition 15 tells us that neither is $L^\infty[0, 1]$.

The contrast between reflexivity for spaces of integrable functions and spaces of continuous functions is striking. According to the preceding theorem, for $1 < p < \infty$, the L^p spaces for a σ-finite measure spaces are reflexive. On the other hand, if K is any compact Hausdorff space and $C(K)$ is normed by the maximum norm, then $C(K)$ is reflexive if and only if K is a finite set (see Problem 11 of Chapter 19).

PROBLEMS

31. Show that a collection of bounded linear functionals is equicontinuous if and only if it is uniformly bounded.
32. Let X be a separable normed linear space. Show that its closed unit sphere $S = \{x \in X \mid \|x\| = 1\}$ also is separable.
33. Find a compact metric space X for which $C(X)$, normed by the maximum norm, is reflexive.

34. Let c_0 be the subspace of ℓ^∞ consisting of sequences that converge to 0. Show that c_0 is a closed subspace of ℓ^∞ whose dual space is isomorphic to ℓ^1. Conclude that c_0 is not reflexive and therefore neither is ℓ^∞.
35. For $1 \leq p \leq \infty$, show that the sequence space ℓ^p is reflexive if and only if $1 < p < \infty$. (For $p = \infty$, see the preceding problem.)
36. Consider the functional $\psi \in (C[-1, 1])^*$ defined by

$$\psi(h) = \int_{-1}^{0} h(x)\,dx - \int_{0}^{1} h(x)\,dx \text{ for all } h \in C[-1, 1].$$

Show that ψ fails to take a maximum on the closed unit ball of $C[-1, 1]$. Use this to provide another proof that $C[-1, 1]$ fails to be reflexive.
37. For $1 < p < \infty$, show that a bounded sequence in ℓ^p converges weakly if and only if it converges componentwise.
38. For $1 \leq p < \infty$ and $[a, b]$ a closed, bounded interval of real numbers, show that a bounded sequence $\{f_n\}$ in $L^p[a, b]$ converges weakly to f if and only if $\{\int_E f_n\,dm\} \to \int_E f\,dm$ for every Lebesgue measurable subset E of $[a, b]$.
39. For $[a, b]$ a closed, bounded interval of real numbers, show that if a sequence $\{f_n\}$ in $C[a, b]$ converges weakly, then it converges pointwise.
40. For X and Y normed linear spaces and an operator $S \in \mathcal{L}(X, Y)$, define the adjoint of S, $S^* \in \mathcal{L}(Y^*, X^*)$ by

$$[S^*(\psi)](x) = \psi(S(x)) \text{ for all } \psi \in Y^*, x \in X.$$

(i) Show that $\|S^*\| = \|S\|$ and that S^* is an isomorphism if S is an isomorphism.

(ii) For $1 < p < \infty$ and $X = L^p(E)$, where E is a measurable set of real numbers, show that the natural embedding $J\colon X \to X^{**}$ may be expressed as the composition

$$J = [\mathcal{S}_q^*]^{-1} \circ \mathcal{S}_p,$$

where $\mathcal{R}_p$ and $\mathcal{R}_q$ are the Riesz representing operators.
41. Let X be a reflexive Banach space and $T\colon X \to X$ a linear operator. Show that T belongs to $\mathcal{L}(X, X)$ if and only if whenever $\{x_n\}$ converges weakly to x, $\{T(x_n)\}$ converges weakly to $T(x)$.

18.4 LOCALLY CONVEX TOPOLOGICAL VECTOR SPACES

There is a very nice class of topologies on a vector space X, topologies for which X is said to be a locally convex topological vector space, which, for our purposes, has two virtues: This class is large enough so that, for a normed linear space X, it includes both the strong topology on X induced by a norm and the weak topologies on X induced by any subspace W of X^* that separates points. On the other hand, this class of topologies is small enough so that for linear spaces with these topologies, if K is a closed, convex set that does not contain the point x_0, there is a closed hyperplane passing through x_0 that contains no point of K.

For two vectors u, v in a linear space X, a vector x that can be expressed as

$$x = \lambda u + (1 - \lambda)v \text{ for } 0 \leq \lambda \leq 1$$

is called a **convex combination** of u and v. A subset K of X is said to be **convex** provided that it contains all convex combinations of vectors in K. Every linear subspace of a linear space is convex, and the open and closed balls in a normed space also are convex.

Definition *A* **locally convex topological vector space** *is a linear space X together with a Hausdorff topology that possesses the following properties:*

(i) Vector addition is continuous, that is, the map $(x, y) \mapsto x + y$ is continuous from $X \times X$ to X;

(ii) Scalar multiplication is continuous, that is, the map $(\lambda, x) \mapsto \lambda \cdot x$ is continuous from $\mathbf{R} \times X$ to X;

(iii) There is a base at the origin for the topology consisting of convex sets.

For a normed linear space X, a subspace W of X^* is said to **separate points** in X provided that for each $u, v \in X$, there is a $\psi \in W$ for which $\psi(u) \neq \psi(v)$. Recall that for a subspace W of X^* and point $x \in X$, a neighborhood base of x with respect to the W-weak topology comprises sets of the form

$$\mathcal{N}_{\epsilon, \psi_1, \ldots, \psi_n}(x) = \{x' \in X \mid \ |\psi_i(x - x')| < \epsilon \text{ for } 1 \leq i \leq n\},$$

where $\epsilon > 0$ and each ψ_i belongs to W.

Proposition 21 *Let X be a normed linear space. Then the linear space X is a locally convex topological vector space with respect to the topology induced by the norm and also with respect to the W-weak topology induced by any subspace W of X^* that separates points in X.*

Proof First consider X with the topology induced by the norm. Since the topology is induced by a metric, it is Hausdorff. From the subadditivity and homogeneity of the norm, it follows that vector addition and scalar multiplication are continuous. Finally, each open ball centered at the origin is convex and the collection of such balls is a base at the origin for the topology induced by the norm.

Now let W be a subspace of X^* that separates points. To show that the W-weak topology is Hausdorff, let u and v be distinct vectors in X. Since W separates points there is a $\psi \in W$ such that $|\psi(u) - \psi(v)| = r > 0$. Then $\{x \in X \mid |\psi(u) - \psi(x)| < r/2\}$ and $\{x \in X \mid |\psi(v) - \psi(x)| < r/2\}$ are disjoint W-weak neighborhoods of u and v, respectively. To show that vector addition is continuous, let x_1 and x_2 belong to X. Consider a W-weak neighborhood $\mathcal{N}_{\epsilon, \psi_1, \ldots, \psi_n}(x_1 + x_2)$ of $x_1 + x_2$. Then the W-weak neighborhoods $\mathcal{N}_{\epsilon/2, \psi_1, \ldots, \psi_n}(x_1)$ and $\mathcal{N}_{\epsilon/2, \psi_1, \ldots, \psi_n}(x_2)$ of x_1 and x_2, respectively, have the property that

$$\text{if } (u, v) \in \mathcal{N}_{\epsilon/2, \psi_1, \ldots, \psi_n}(x_1) \times \mathcal{N}_{\epsilon/2, \psi_1, \ldots, \psi_n}(x_2), \text{ then } u + v \in \mathcal{N}_{\epsilon, \psi_1, \ldots, \psi_n}(x_1 + x_2).$$

Thus vector addition is continuous at $(x_1, x_2) \in X \times X$. A similar argument shows that scalar multiplication is continuous. Finally, a basic neighborhood of the origin is of the form $\mathcal{N}_{\epsilon, \psi_1, \ldots, \psi_n}(0)$ and this set is convex since each ψ_k is linear. $\square$

Definition *Let E be a subset of a linear space X. A point $x_0 \in E$ is said to be an* **internal point** *of E provided that for each $x \in X$, there is some $\lambda_0 > 0$ for which $x_0 + \lambda \cdot x$ belongs to E if $|\lambda| \leq \lambda_0$.*

Proposition 22 *Let X be a locally convex topological vector space.*

(i) A subset $\mathcal{N}$ of X is open if and only if for each $x_0 \in X$ and $\lambda \neq 0$, $x_0 + \mathcal{N}$ and $\lambda \cdot \mathcal{N}$ are open.

(ii) The closure of a convex subset of X is convex.

(iii) Every point in an open subset $\mathcal{O}$ of X is an internal point of $\mathcal{O}$.

Proof We first verify (i). For $x_0 \in X$, define the translation map $T_{x_0}: X \to X$ by $T_{x_0}(x) = x + x_0$. Then T_{x_0} is continuous since vector addition is continuous. The map T_{-x_0} also is continuous and is the inverse of T_{x_0}. Therefore, T_{x_0} is a homeomorphism of X onto X. Thus $\mathcal{N}$ is open if and only if $\mathcal{N} + x_0$ is open. The proof of invariance of the topology under non-zero scalar multiplication is similar.

To verify (ii), let K be a convex subset of X. Fix $\lambda \in [0, 1]$. Define the mapping $\Psi: X \times X \to X$ by

$$\Psi(u, v) = \lambda u + (1 - \lambda)v \text{ for all } u, v \in X.$$

Since scalar multiplication and vector addition are continuous, $\Psi: X \times X \to X$ is continuous, where $X \times X$ has the product topology. A continuous mapping maps the closure of a set into the closure of the image of the set. Thus $\Psi(\overline{K \times K}) \subseteq \overline{\Psi(K \times K)}$. However, $\overline{K \times K} = \overline{K} \times \overline{K}$. Moreover, since K is convex, $\Psi(K \times K) \subseteq K$. Therefore, $\Psi(\overline{K} \times \overline{K}) \subseteq \overline{K}$. Since this holds for all $\lambda \in [0, 1]$, the closure of K is convex.

To verify (iii), let x_0 belong to $\mathcal{O}$. Fix $x \in X$. Define the function $g: R \to X$ by $g(\lambda) = \lambda \cdot x + x_0$. Since addition and scalar multiplication are continuous, so is g, and consequently, because $g(0) = x_0$ and $\mathcal{O}$ is a neighborhood of x_0, there is a $\lambda > 0$ such that $g(\lambda) \in \mathcal{O}$ if $|\lambda| < \lambda_0$. So $\lambda \cdot x + x_0 \in \mathcal{O}$ if $|\lambda| < \lambda_0$. Therefore, x_0 is an internal point of $\mathcal{O}$. □

Proposition 23 *Let X be a locally convex topological vector space and $\psi: X \to \mathbf{R}$ be linear. Then ψ is continuous if and only if there is a neighborhood of the origin on which $|\psi|$ is bounded, that is, there is a neighborhood of the origin, $\mathcal{N}_0$, and an $M > 0$ for which*

$$|\psi| \leq M \text{ on } \mathcal{N}_0. \tag{14}$$

Proof First suppose ψ is continuous. Then it is continuous at $x = 0$ and so, since $\psi(0) = 0$, there is a neighborhood $\mathcal{N}_0$ of 0 such that $|\psi(x)| = |\psi(x) - \psi(0)| < 1$ for $x \in \mathcal{N}_0$. Thus $|\psi|$ is bounded on $\mathcal{N}_0$. To prove the converse, let $\mathcal{N}_0$ be a neighborhood of 0 and $M > 0$ be such that (14) holds. For each $\lambda > 0$, $\lambda \cdot \mathcal{N}_0$ is also a neighborhood of 0 and $|\psi| \leq \lambda \cdot M$ on $\lambda \cdot \mathcal{N}_0$. To verify continuity of $\psi: X \to \mathbf{R}$, let x_0 belong to X and $\epsilon > 0$. Choose λ so that $\lambda \cdot M < \epsilon$. Then $x_0 + \lambda \cdot \mathcal{N}_0$ is a neighborhood of x_0 and if x belongs to $x_0 + \lambda \cdot \mathcal{N}_0$, then $x - x_0$ belongs to $\lambda \cdot \mathcal{N}_0$ so that

$$|\psi(x) - \psi(x_0)| = |\psi(x - x_0)| \leq \lambda \cdot M < \epsilon.$$ □

According to Theorem 14 of Chapter 15, the weak topology on an infinite dimensional normed linear space is not induced by a metric. The following example suggests another way of proving this.

Example (von Neumann) For each natural number n, let e_n denote the sequence in ℓ^2 whose nth component is 1 and other components vanish. Define

$$E = \{e_n + n \cdot e_m \mid \ n \text{ and } m \text{ any natural numbers, } m > n\}.$$

We leave it as an exercise to show that 0 is a point of closure, with respect to the weak topology, of E but there is no sequence in E that converges weakly to 0. Therefore, the weak topology on ℓ^2 is not metrizable.

Two metrics on a set X induce the same topology if and only if a sequence that is convergent with respect to one of the metrics is convergent with respect to the other. Things are quite different for locally convex topological vector spaces. There are linear spaces X that have distinct locally convex topologies with respect to which the convergence of sequences is the same. A classic example of this is the sequence space $X = \ell^1$. The space X is a locally convex topological vector space with respect to the strong topology and with respect to the weak topology and these topologies are distinct. However, a lemma of Schur asserts that a sequence converges weakly in ℓ^1 if and only if it converges strongly in ℓ^1.[2]

A topological vector space is defined to be a linear space with a Hausdorff topology with respect to which vector addition and scalar multiplication are continuous. In the absence of local convexity, such spaces can be rather pathological. For instance, if $0 < p < 1$, let X be the linear space of all Lebesgue measurable extended real-valued functions on $[0, 1]$ for which $|f|^p$ is integrable over $[0, 1]$ with respect to Lebesgue measure. Define

$$\rho(f, g) = \int_{[0, 1]} |f - g|^p \, dm \text{ for all } f, g \in X.$$

Then, after identifying functions that are equal almost everywhere, ρ is a metric that induces a Hausdorff topology on X with respect to which vector addition and scalar multiplication are continuous. But there are no continuous linear functionals on X besides the zero functional (see Problem 52). In the next section, we show that there are lots of continuous linear functionals on a topological vector space that is locally convex.

PROBLEMS

42. Let X be a normed linear space and W be a subspace of X^*. Show that the W-weak topology on X is Hausdorff if and only if W separates points in X.
43. Let X be a normed linear space and $\psi\colon X \to \mathbf{R}$ be linear. Show that ψ is continuous with respect to the weak topology if and only if it is continuous with respect to the strong topology.
44. Let X be a locally convex topological vector space and $\psi\colon X \to \mathbf{R}$ be linear. Show that ψ is continuous if and only if it is continuous at the origin.
45. Let X be a locally convex topological vector space and $\psi\colon X \to \mathbf{R}$ be linear. Show that ψ is continuous if and only if there is a neighborhood $\mathcal{O}$ of the origin for which $f(\mathcal{O}) \neq \mathbf{R}$.
46. Let X be a normed linear space and W a subspace of X^* that separates points. For any topological space Z, show that a mapping $f\colon Z \to X$ is continuous, where X has the W-weak topology, if and only if $\psi \circ f\colon Z \to \mathbf{R}$ is continuous for all $\psi \in W$.

[2] See Robert E. Megginson's *An Introduction to Banach Space Theory* [Meg98].

47. Show that the topology on a finite dimensional locally convex topological vector space is induced by a norm.

48. Let X be a locally convex topological space. Show that the linear space X' of all linear continuous functionals ψ: $X \to \mathbf{R}$ also has a topology with respect to which it is a locally convex topological space on which, for each $x \in X$, the linear functional $\psi \mapsto \psi(x)$ is continuous.

49. Let X and Y be locally convex topological vector spaces and T: $X \to Y$ be linear, one-to-one, and onto. Show that T is a topological homeomorphism if and only if it maps a base at the origin for the topology on X to a base at the origin for the topology on Y.

50. Let X be a linear space and the function σ: $X \to [0, \infty)$ have the following properties: for all $u, v \in X$, (i) $\sigma(u+v) \leq \sigma(u)+\sigma(v)$; (ii) $\sigma(u) = 0$ if and only if $u = 0$; (iii) $\sigma(u) = \sigma(-u)$. Define $\rho(u, v) = \sigma(u - v)$. Show that ρ is a metric on X.

51. (Nikodym) Let X be the linear space of all Lebesgue measurable functions on $[0, 1]$. Define
$$\sigma(f) = \int_{[0,\,1]} \frac{|f|}{1+|f|}\, dm \text{ for all } f \in X.$$
   (i) After identifing functions that are equal almost everywhere, use Problem 50 to show that $\rho(u, v) = \sigma(u - v)$ defines a metric on X.
   (ii) Show that $\{f_n\} \to f$ with respect to the metric ρ if and only if $\{f_n\} \to f$ in measure.
   (iii) Show that (X, ρ) is a complete metric space.
   (iv) Show that the mapping $(f, g) \mapsto f + g$ is a continuous mapping of $X \times X$ into X.
   (v) Show that the mapping $(\lambda, f) \mapsto \lambda \cdot f$ is a continuous mapping of $\mathbf{R} \times X$ into X.
   (vi) Show that there are no non-zero continuous linear functionals ψ on X. (Suggestion: Let ψ: $X \to \mathbf{R}$ be linear and continuous. Show that there is an n such that $\psi(f) = 0$ whenever f is the characteristic function of an interval of length less than $1/n$. Hence $\psi(f) = 0$ for all step-functions f.)

52. (Day) For $0 < p < 1$, let X be the linear space of all measurable (with respect to the Lebesgue measure m) real-valued functions on $[0, 1]$ for which $|f|^p$ is integrable. Define
$$\sigma(f) = \int_{[0,\,1]} |f|^p\, dm \text{ for all } f \in X.$$
   (i) Use Problem 50 to show that $\rho(u, v) = \sigma(u - v)$ defines a metric on X.
   (ii) Show that the linear space X, with the topology determined by ρ, is a topological vector space.
   (iii) For a non-negative function f in X and natural number n, show that there is a partition $0 = x_0 < x_1 < \ldots x_n = 1$ of $[0, 1]$ for which $\int_{x_{k-1}}^{x_k} f^p\, dm = 1/n \cdot \int_0^1 f^p\, dm$, for all $1 \leq k \leq n$.
   (iv) For a non-negative function f in X and natural number n, show that there are functions $f_1, \ldots, f_n$ for which $\rho(f_k, 0) < 1/n^{1-p} \int_0^1 f^p dm$ for $1 \leq k \leq n$ and
$$f = \sum_{k=1}^{n} 1/n \cdot f_k.$$
   (v) Show that there are no continuous non-zero linear functionals on X.

53. Let $\mathcal{S}$ be the space of all sequences of real numbers, and define

$$\sigma(x) = \sum \frac{|x_n|}{2^n[1+|x_n|]} \text{ for all } x = \{x_n\} \in \mathcal{S}.$$

Show that σ induces a metric on $\mathcal{S}$ for which $\mathcal{S}$ is a topological vector space with the metric topology. What is the most general continuous linear functional on $\mathcal{S}$?

18.5 THE SEPARATION OF CONVEX SETS AND MAZUR'S THEOREM

In the first section of this chapter, we showed that a hyperplane in a linear space X is the level set of a non-zero linear functional on X. We therefore say that two non-empty subsets A and B of X may be **separated by a hyperplane** provided that there is a linear functional $\psi\colon X \to \mathbf{R}$ and $c \in \mathbf{R}$ for which

$$\psi < c \text{ on } A \text{ and } \psi > c \text{ on } B.$$

Observe that if A is the singleton set $\{x_0\}$, then this means precisely that

$$\psi(x_0) < \inf_{x \in B} \psi(x).$$

Definition *Let K be a convex subset of a linear space X for which the origin is an internal point. The* **gauge functional**[3] *for K, $p_K\colon X \to [0, \infty)$, is defined by*

$$p_K(x) = \inf\{\lambda > 0 \mid x \in \lambda \cdot K\}$$

Note it is precisely because the origin is an internal point of the convex set K that its gauge functional is properly defined. Also note that the gauge functional associated with the unit ball of a normed linear space is the norm itself.

Proposition 24 *Let K be a convex subset of a linear space X that contains the origin as an internal point and p_K the gauge functional for K. Then p_K is subadditive and positively homogeneous.*

Proof We establish subadditivity and leave the proof of positive homogeneity as an exercise. Let $u, v \in X$ and suppose, for $\lambda > 0$ and $\mu > 0$, that $x \in \lambda K$ and $y \in \mu K$. Then, since K is convex,

$$\frac{1}{\lambda+\mu}\cdot(x+y) = \frac{\lambda}{\lambda+\mu}\cdot\frac{x}{\lambda} + \frac{\mu}{\lambda+\mu}\cdot\frac{y}{\mu} \in K.$$

Therefore, $x+y \in (\lambda+\mu)K$ so that $p_K(x+y) \le \lambda+\mu$. Taking infima, first over all such λ and then over all such μ, we conclude that $p_K(x+y) \le p_K(x)+p_K(y)$. □

The Hyperplane Separation Lemma *Let K_1 and K_2 be two non-empty, disjoint, convex subsets of a linear space X, one of which has an internal point. Then there is a non-zero linear functional $\psi\colon X \to \mathbf{R}$ for which*

$$\sup_{x \in K_1} \psi(x) \le \inf_{x \in K_2} \psi(x). \tag{15}$$

[3] A gauge functional is often called a Minkowski functional.

Proof Suppose x_1 is an internal point of K_1 and x_2 is any point of K_2. Define

$$z = x_2 - x_1 \text{ and } K = K_1 + [-K_2] + z.$$

Then K is a convex set that contains the origin as an internal point and does not contain z. Let $p = p_K\colon X \to \mathbf{R}$ be the gauge functional for K. Define $Y = \operatorname{span}[z]$ and the linear functional $\psi\colon Y \to \mathbf{R}$ by $\psi(\lambda z) = \lambda$. Thus $\psi(z) = 1$, and since $1 \leq p(z)$ because $z \notin K$, we conclude that $\psi \leq p$ on Y. According to the preceding proposition, p is subadditive and positively homogeneous. Thus the Hahn-Banach Theorem tells us that ψ may be extended to a linear functional on all of X so that $\psi \leq p$ on all of X. Let $x \in K_1$ and $y \in K_2$. Then $x - y + z \in K$ so that $p(x - y + z) \leq 1$ and thus, since ψ is linear and $\psi \leq p$ on all of X,

$$\psi(x) - \psi(y) + \psi(z) = \psi(x - y + z) \leq p(x - y + z) \leq 1.$$

Since $\psi(z) = 1$, we have $\psi(x) \leq \psi(y)$. This holds for each $x \in K_1$ and y in K_2, so

$$\sup_{x \in K_1} \psi(x) \leq \inf_{y \in K_2} \psi(y).$$

Of course, $\psi \neq 0$ since $\psi(z) = 1$. If K_2 has the internal point, we apply the same argument, but at the end replace ψ by $-\psi$. □

The Hyperplane Separation Theorem *Let X be a locally convex topological vector space, K a non-empty, closed, convex subset of X, and x_0 a point in X that lies outside of K. Then K and x_0 may be separated by a closed hyperplane, that is, there is a continuous linear functional $\psi\colon X \to \mathbf{R}$ for which*

$$\psi(x_0) < \inf_{x \in K} \psi(x). \tag{16}$$

Proof Since K is closed, $X \sim K$ is open. Choose a convex neighborhood $\mathcal{N}_0$ of 0 for which

$$K \cap [\mathcal{N}_0 + x_0] = \emptyset.$$

We may, possibly by replacing $\mathcal{N}_0$ with $\mathcal{N}_0 \cap [-\mathcal{N}_0]$, suppose $\mathcal{N}_0$ is symmetric with respect to the origin, that is, $\mathcal{N}_0 = -\mathcal{N}_0$. By the Hyperplane Separation Lemma, there is a non-zero linear functional $\psi\colon X \to \mathbf{R}$ for which

$$\sup_{x \in \mathcal{N}_0 + x_0} \psi(x) \leq \inf_{x \in K} \psi(x). \tag{17}$$

Since $\psi \neq 0$, we may choose $z \in X$ such that $\psi(z) > 0$. According to Proposition 22, an interior point of a set is an internal point. Choose $\lambda > 0$ such that $\lambda \cdot z \in \mathcal{N}_0$. Since $\lambda\psi(z) > 0$ and $\lambda z + x_0 \in \mathcal{N}_0 + x_0$, it follows from the linearity of ψ and inequality (17) that

$$\psi(x_0) < \lambda\psi(z) + \psi(x_0) = \psi(\lambda z + x_0) \leq \sup_{x \in \mathcal{N}_0 + x_0} \psi(x) \leq \inf_{x \in K} \psi(x).$$

It remains to show that ψ is continuous. Define $M = [\inf_{x \in K} \psi(x)] - \psi(x_0)$. It follows from (17) that $\psi \leq M$ on $\mathcal{N}_0$. Since $\mathcal{N}_0$ is symmetric, $|\psi| \leq M$ on $\mathcal{N}_0$. By Proposition 23, ψ is continuous. □

Corollary 25 *Let X be a normed linear space, K a non-empty, strongly closed, convex subset of X, and x_0 a point in X that lies outside of K. Then there is a functional $\psi \in X^*$ for which*

$$\psi(x_0) < \inf_{x \in K} \psi(x). \tag{18}$$

Proof According to Proposition 21, the linear space X is a locally convex topological vector space with respect to the strong topology. The conclusion now follows from the Hyperplane Separation Theorem. $\square$

Corollary 26 *Let X be a normed linear space and W a subspace of its dual space X^* that separates points in X. Furthermore, let K be a non-empty W-weakly closed, convex subset of X and x_0 a point in X that lies outside of K. Then there is a functional $\psi \in W$ for which*

$$\psi(x_0) < \inf_{x \in K} \psi(x). \tag{19}$$

Proof According to Proposition 21, the linear space X is a locally convex topological vector space with respect to the W-weak topology. Proposition 5 tells us that the W-weakly continuous linear functionals on X belong to W. The conclusion now follows from the Hyperplane Separation Theorem. $\square$

Mazur's Theorem *Let K be a convex subset of a normed linear space X. Then K is strongly closed if and only if it is weakly closed.*

Proof Since each $\psi \in X^*$ is continuous with respect to the strong topology, each weakly open set is strongly open and therefore each weakly closed set is strongly closed, irrespective of any convexity assumption. Now suppose K is non-empty, strongly closed, and convex. Let x_0 belong to $X \sim K$. By Corollary 25, there is a $\psi \in X^*$ for which

$$\psi(x_0) < \alpha = \inf_{x \in K} \psi(x).$$

Then $\{x \in X \mid \psi(x) < \alpha\}$ is a weak neighborhood of x_0 that is disjoint from K. Thus $X \sim K$ is weakly open and therefore its complement in X, K, is weakly closed. $\square$

Corollary 27 *Let K be a strongly closed, convex subset of a normed linear space X. Suppose $\{x_n\}$ is a sequence in K that converges weakly to $x \in X$. Then x belongs to K.*

Proof The weak limit of a sequence in K is a point of closure of K with respect to the weak topology. Therefore x belongs to the weak closure of K. But Mazur's Theorem tells us that the weak closure of K is K itself. $\square$

Theorem 28 *Let X be a reflexive Banach space. Then each strongly closed, bounded, convex subset of X is weakly sequentially compact.*

Proof Theorem 17 tells us that every bounded sequence in X has a weakly convergent subsequence. Therefore, by the preceding corollary, every sequence in K has a subsequence that converges weakly to a point in K. $\square$

The following is a variation of the Banach-Saks Theorem which we proved in Section 8.4 for $X = L^2$; the conclusion is weaker, but it holds for general normed linear spaces.

Theorem 29 *Let X be a normed linear space and $\{x_n\}$ a sequence X that converges weakly to $x \in X$. Then there is a sequence $\{z_n\}$ that converges strongly to x and each z_n is a convex combination of $\{x_n, x_{n+1}, \ldots\}$.*

Proof We argue by contradiction. If the conclusion is false, then there is a natural number n and an $\epsilon > 0$ for which, if we define K_0 to be the set of all convex combinations of $\{x_n, x_{n+1}, \ldots\}$, then

$$\|x - z\| \geq \epsilon \text{ for all } z \in K_0.$$

Define K to be the strong closure of K_0. Then x does not belong to K. The strong closure of a convex set is convex. Therefore, by Mazur's Theorem, K is weakly closed. Since $\{x_n\}$ converges to x with respect to the weak topology, x is a point of closure of K with respect to the weak topology. But a point of closure of a closed set belongs to the set. This contradiction concludes the proof. $\square$

The following theorem generalizes a minimization theorem in Chapter 8; namely, that for $1 \leq p < \infty$, a continuous, convex function defined on a closed, bounded, convex subset of L^p takes a minimum value.

Theorem 30 *Let K be a strongly closed, bounded, convex subset of a reflexive Banach space X. Let the function $f\colon K \to \mathbf{R}$ be continuous with respect to the strong topology on K and convex, in the sense that for $u, v \in K$ and $0 \leq \lambda \leq 1$,*

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v).$$

Then f takes a minimum value on K.

Proof Define m to be the infimum of $f(K)$. Choose a sequence $\{x_n\}$ in K such that $\{f(x_n)\}$ converges to m. According to Theorem 28, K is weakly sequentially compact. We may assume that $\{x_n\}$ converges weakly to $x \in K$. First assume m is finite. Let $\epsilon > 0$. Choose a natural number N such that

$$m \leq f(x_k) < m + \epsilon \text{ for all } k \geq N. \tag{20}$$

Theorem 29 tells us that there is a sequence $\{z_n\}$ that converges strongly to x and each z_n is a convex combination of $\{x_n, x_{n+1}, \ldots\}$. By the continuity of f with respect to the strong topology on K, $\{f(z_n)\} \to f(x)$. On the other hand, from the convexity of f and (20),

$$m \leq f(z_n) < m + \epsilon \text{ for all } n \geq N.$$

Therefore, $m \leq f(x) \leq m + \epsilon$. This holds for all $\epsilon > 0$ and hence f takes its minimum value on K at the point x. The same argument shows that, in fact, m is always finite. $\square$

PROBLEMS

54. For each natural number n, let e_n denote the sequence in ℓ^2 whose nth component is 1 and other components vanish. Define $E = \{e_n + n \cdot e_m \mid n \text{ and } m \text{ any natural numbers}, m > n\}$. We showed that 0 is a point of closure of E but no sequence in E converges weakly to 0.

Consider the topological space $X = E \cup \{0\}$ with the weak topology. Find a function $f\colon X \to \mathbf{R}$ that fails to be continuous at 0 and yet has the property that whenever a sequence $\{x_n\}$ in E converges weakly to 0, its image sequence $\{f(x_n)\}$ converges to $f(0)$.

55. Find a subset of the plane $\mathbf{R}^2$ for which the origin is an internal point but not an interior point.

56. Let X be a locally convex topological vector space and V a convex, symmetric with respect to the origin (that is, $V = -V$) neighborhood of the origin. If p_V is the gauge functional for V and ψ is a linear real-valued functional on X such that $\psi \leq p_V$ on X, show that ψ is continuous.

57. Let X be a locally convex topological vector space, Y a closed subspace of X, and x_0 belong to $X \sim Y$. Show that there is a continuous functional $\psi\colon X \to \mathbf{R}$ such that

$$\psi(x_0) \neq 0 \text{ and } \psi = 0 \text{ on } Y.$$

58. Let X be a normed linear space and W a proper subspace of X^* that separates points. Let ψ belong to $X^* \sim W$. Show that ker ψ is strongly closed and convex but not W-weakly closed. (Suggestion: Otherwise, apply Corollary 26 with $K = \ker \psi$.)

59. Let X be a normed linear space. Show that the closed unit ball B^* of X^* is weak-$*$ closed.

60. Show that the Hyperplane Separation Theorem may be amended as follows: the point x_0 may be replaced by a convex set K_0 that is disjoint from K and the conclusion is that K and K_0 can be separated by a closed hyperplane if K_0 is compact. If K_0 is open, there is a continuous linear functional ψ on X for which $\psi(x_0) < \inf_{x \in K} \psi(x)$ for all $x_0 \in K_0$.

61. Show that the weak topology on an infinite dimensional normed linear space is not first countable.

62. Show that every weakly compact subset of a normed linear space is bounded with respect to the norm.

63. Let Y be a closed subspace of a reflexive Banach space X. For $x_0 \in X \sim Y$, show that there is a point in Y that is closest to x_0.

64. Let X be normed linear space, W a finite dimensional subspace of X^* and ψ a functional in $X^* \sim W$. Show that there is a vector $x \in X$ such that $\psi(x) \neq 0$ while $\varphi(x) = 0$ for all φ in W. (Suggestion: First show this is true if X is finite dimensional.)

65. Let X be a normed linear space. Show that any dense subset of $B^* = \{\psi \in X^* \mid \|\psi\| \leq 1\}$ separates points in X.

66. Complete the final part of the proof of Proposition 24.

67. Find an example of a bounded subset A of a normed linear space X, $\mathcal{F}$ a set of functionals in X^* containing $\mathcal{F}_0$ as a dense subset of $\mathcal{F}$ (dense in the sense of the norm topology on X^*) such that $\mathcal{F}$ and $\mathcal{F}_0$ generate different weak topologies for X, but the same weak topology for A.

18.6 THE KREIN-MILMAN THEOREM

Definition *Let K be a non-empty, convex subset of a locally convex topological vector space X. A non-empty subset E of K is called an* **extreme subset**[4] *of K provided that it is both convex and closed, and whenever $x \in E$ and $x = \lambda u + (1-\lambda)v$ for $0 < \lambda < 1$ and*

[4]An extreme subset of K is also often called a supporting set for K.

$u, v \in K$, then $u, v \in E$. A point $x \in K$ is called an **extreme point** *of K provided that the singleton set $\{x\}$ is an extreme subset of K.*

We leave it as two exercises to show that if the intersection of a collection of extreme subsets of K is non-empty, then the intersection is an extreme subset of K and, moreover, if A is an extreme subset of B and B is an extreme subset of K, then A is an extreme subset of K.

Lemma 31 *Let K be a non-empty, compact, convex subset of a locally convex topological vector space X and $\psi: X \to \mathbf{R}$ be linear and continuous. Then the set of points in K at which ψ takes its maximum value on K is an extreme subset of K.*

Proof Since K is compact and ψ is continuous, ψ takes a maximum value, m, on K. The subset M of K on which ψ takes its maximum value is closed, since ψ is continuous, and is convex since ψ is linear. Let $x \in M$ and suppose there are vectors u, v in K and $0 < \lambda < 1$ for which $x = \lambda u + (1 - \lambda)v$. Since

$$\psi(u) \leq m, \psi(v) \leq m \text{ and } m = \psi(x) = \lambda\psi(u) + (1 - \lambda)\psi(v),$$

we must have $\psi(u) = \psi(v) = m$, that is, $u, v \in M$. □

The Krein-Milman Lemma *Let K be a non-empty, compact, convex subset of a locally convex topological vector space X. Then K has an extreme point.*

Proof The strategy of the proof is first to apply Zorn's Lemma to find an extreme subset E of K that contains no proper subset which also is an extreme subset of K. It follows from the Hyperplane Separation Lemma and the preceding lemma that E is a singleton set.

Consider the collection $\mathcal{F}$ of extreme subsets of K. Then $\mathcal{F}$ is non-empty since it contains K. Order $\mathcal{F}$ by containment. Let $\mathcal{F}_0 \subseteq \mathcal{F}$ be totally ordered. Then $\mathcal{F}_0$ has the finite intersection property, since for any finite subcollection of $\mathcal{F}_0$, because $\mathcal{F}_0$ is totally ordered, one of the subsets of this finite subcollection is contained in all the others and therefore the intersection is non-empty. Thus $\mathcal{F}_0$ is a collection of non-empty, closed subsets of the compact set K, which has the finite intersection property. Hence if we let E_0 be the intersection of the sets in $\mathcal{F}_0$, E_0 is non-empty. As already observed, E_0 is an extreme subset of K, since it is the non-empty intersection of such sets. Therefore E_0 is a lower bound for $\mathcal{F}_0$. Consequently, every totally ordered subcollection of $\mathcal{F}$ has a lower bound and so, by Zorn's Lemma, $\mathcal{F}$ has a minimal member, that is, there is an extreme subset E of K which contains no proper extreme subset.

We claim that E is a singleton set. Indeed, otherwise, there are two points u and v in E. It follows from the Hyperplane Separation Lemma that there is a $\psi \in X^*$ for which $\psi(u) < \psi(v)$. According to Lemma 31, the subset M of E on which ψ takes its maximum value on E is an extreme subset of E. Since E is an extreme subset of K, M is also an extreme subset of K. Clearly $u \notin M$, and therefore, M is a proper subset of E. This contradicts the minimality of E. Thus E is a singleton set and therefore K has an extreme point. □

Definition *Let K be a subset of a locally convex topological vector space X. Then the* **closed, convex hull** *of K is defined to be the intersection of all closed, convex subsets of X that contain K.*

It follows from Mazur's Theorem that in a normed linear space the weakly closed, convex hull of a set equals its strongly closed, convex hull. It is clear that the closed, convex hull of a set K is a closed convex set that contains K and that is contained in any other closed convex set that contains K.

The Krein-Milman Theorem *Let K be a non-empty, compact, convex subset of a locally convex topological vector space X. Then K is the closed, convex hull of its extreme points.*

Proof By the Krein-Milman Lemma, the set E of extreme points of K is non-empty. Let C be the closed, convex hull of E. If $K \neq C$, choose $x_0 \in K \sim C$. By the Hyperplane Separation Theorem, since C is convex and closed, there is a continuous linear functional $\psi: X \to \mathbf{R}$ such that

$$\psi(x_0) > \max_{x \in C} \psi(x) \geq \max_{x \in E} \psi(x). \tag{21}$$

By Lemma 31, if m is the maximum value taken by ψ on K, then $M = \{x \in K \mid \psi(x) = m\}$ is an extreme subset of K. By the Krein-Milman Lemma, applied now with K replaced by the non-empty, compact, convex set M, there is a point $z \in M$ that is an extreme point of M. As we already observed, an extreme point of an extreme subset of K is also an extreme point of K. It follows from (21) that $\psi(z) \geq \psi(x_0) > \psi(z)$. This contradiction shows that $K = C$. $\square$

There are many interesting applications of the Krein-Milman Theorem[5]. In Chapter 22, this theorem is used to prove the existence of ergodic measure preserving transformations.

PROBLEMS

68. Find the extreme points of each of the following subsets of the plane $\mathbf{R}^2$:
(i) $\{(x, y) \mid x^2 + y^2 \leq 1\}$; (ii) $\{(x, y) \mid |x| + |y| \leq 1\}$; (iii) $\{(x, y) \mid \max\{x, y\} \leq 1\}$.
69. In each of the following, B denotes the closed unit ball of a normed linear space X.
   (i) If X contains more than one point, show that the only possible extreme points of B have norm 1.
   (ii) If $X = L^p[a, b], 1 < p < \infty$, show that every unit vector in B is an extreme point of B.
   (iii) If $X = L^\infty[a, b]$, show that the extreme points of B are those functions $f \in B$ such that $|f| = 1$ almost everywhere on $[a, b]$.
   (iv) If $X = L^1[a, b]$, show that B fails to have any extreme points.
   (v) If $X = l^p, 1 \leq p \leq \infty$, what are the extreme points of B?
   (vi) If $X = C(K)$, where K is a compact Hausdorff topological space and X is normed by the maximum norm, what are the extreme points of B?
70. A norm on a linear space is said to be **strictly convex** provided that whenever u and v are distinct unit vectors and $0 < \lambda < 1$, then $\|\lambda u + (1 - \lambda)v\| < 1$. Show that the Euclidean norm on $\mathbf{R}^n$ and the usual norm on $L^p[a, b], 1 < p < \infty$ are strictly convex.

[5] See Peter Lax's *Functional Analysis* [Lax97].

CHAPTER 19

Compactness Regained: The Weak Topology

Contents

19.1 Alaoglu's Extension of Helly's Theorem 382
19.2 Reflexivity and Weak Compactness: Kakutani's Theorem 384
19.3 Compactness and Weak Sequential Compactness: The Eberlein-Šmulian Theorem 386
19.4 Metrizability of Weak Topologies 389

The closed unit ball of an infinite dimensional normed linear space fails to be compact with respect to the strong topology induced by the norm. In this chapter, a precise theorem is proven regarding the manner in which, for an infinite dimensional Banach space, compactness of the closed unit ball is regained with respect to the weak topology. We prove that if B is the closed unit ball of a Banach space X, then the following are equivalent:

(i) X is reflexive;

(ii) B is weakly compact;

(iii) B is weakly sequentially compact.

The first compactness result established is Alaoglu's Theorem, an extension of Helly's Theorem to non-separable spaces, which tells us that for a normed linear space X, the closed unit ball of the dual space X^* is compact with respect to the weak-$*$ topology. This immediate consequence of the Tychonoff Product Theorem enables us to prove the equivalences (i)–(ii). What is rather surprising is that, for the weak topology on B, sequential compactness is equivalent to compactness despite the fact that in general the weak topology on B is not metrizable.

19.1 ALAOGLU'S EXTENSION OF HELLY'S THEOREM

Let X be a normed linear space, B its closed unit ball, and B^* the closed unit ball of its dual space X^*. Assume that X is separable. Choose $\{x_n\}$ to be a dense subset of B and define

$$\rho(\psi, \eta) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot |(\psi - \eta)(x_n)| \text{ for all } \psi, \eta \in B^*.$$

Then ρ is a metric that induces the weak-$*$ topology on B^* (see Corollary 11). For a metric space, compactness is the same as sequential compactness. Therefore Helly's Theorem may be restated as follows: If X is a separable normed linear space, then the closed unit ball B^* of its dual space X^* is weak-$*$ compact. We now use the Tychonoff Product Theorem to show that the separability assumption is not needed.

Recall a special case of the Tychonoff Product Theorem: Let Λ be any set. Consider the collection $\mathcal{F}(\Lambda)$ of all real-valued functions on Λ that take values in the closed, bounded interval $[-1, 1]$. Consider $\mathcal{F}(\Lambda)$ as a topological space with the product topology. A base at $f \in \mathcal{F}(\Lambda)$ for the product topology on $\mathcal{F}(\Lambda)$ comprise sets of the form

$$\mathcal{N}_{\epsilon, \lambda_1, \ldots, \lambda_n}(f) = \{f' \in \mathcal{F}(\Lambda) \mid |f'(\lambda_k) - f(\lambda_k)| < \epsilon \text{ for } 1 \leq k \leq n\},$$

where $\epsilon > 0$ and the λ_k's belong to Λ. The Tychonoff Product Theorem implies that the topological space consisting of $\mathcal{F}(\Lambda)$ with the product topology is compact. Therefore, every closed subset $\mathcal{F}_0(\Lambda)$ of $\mathcal{F}(\Lambda)$, with the topology induced by the product topology, also is compact[1].

Alaoglu's Theorem *Let X be a normed linear space. Then the closed unit ball B^* of its dual space X^* is compact with respect to the weak-$*$ topology.*

Proof Denote the closed unit balls in X and X^* by B and B^*, respectively. By the preceding discussion, the topological space $\mathcal{F}(B)$ consisting of functions from B to $[-1, 1]$, with the product topology, is compact.

Define the restriction map $R: B^* \to \mathcal{F}(B)$ by $R(\psi) = \psi|_B$ for $\psi \in B^*$. We claim that (i) $R(B^*)$ is a closed subset of $\mathcal{F}(B)$ and (ii) the restriction map R is a topological homeomorphism from B^*, with the weak-$*$ topology, onto $R(B^*)$, with the product topology. Suppose, for the moment, that (i) and (ii) have been established. By the preceding discussion, the Tychonoff Product Theorem tells us that $R(B^*)$ is compact. Therefore any space topologically homeomorphic to $R(B^*)$ is compact. In particular, by (ii), B^* is weak-$*$ compact.

It remains to verify (i) and (ii). First observe that R is one-to-one since for $\psi, \eta \in B^*$, with $\psi \neq \eta$, there is some $x \in B$ for which $\psi(x) \neq \eta(x)$ and thus $R(\psi) \neq R(\eta)$. A direct comparison of the basic open sets in the weak-$*$ topology with basic open sets in the product topology reveals that R is a homeomorphism of B^* onto $R(B^*)$. It remains to show that $R(B^*)$ is closed with respect to the product topology. Let $f: B \to [-1, 1]$ be a point of closure, with respect to the product topology, of $R(B^*)$. To show that $f \in R(B^*)$ it suffices (see Problem 1) to show that for all $u, v \in B$ and $\lambda \in \mathbf{R}$ for which $u+v$ and λu also belong to B,

$$f(u+v) = f(u) + f(v) \text{ and } f(\lambda u) = \lambda f(u). \tag{1}$$

However, for any $\epsilon > 0$, the weak-$*$ neighborhood of f, $\mathcal{N}_{\epsilon,u,v,u+v}(f)$, contains some $R(\psi_\epsilon)$ and since $\psi_\in$ is linear, we have $|f(u+v) - f(u) - f(v)| < 3\epsilon$. Therefore, the first equality in (1) holds. The proof of the second is similar. $\square$

Corollary 1 *Let X be a normed linear space. Then there is a compact, Hausdorff space K for which X is isomorphic to a linear subspace of $C(K)$, normed by the maximum norm.*

Proof Let K be the closed unit ball of the dual space, with the weak-$*$ topology. Alaoglu's Theorem tells us that K is compact, and it clearly it is Hausdorff. Consider the linear space $C(K)$, normed with the maximum norm. For $x \in X$, defines $\Phi(x)$ to be the restriction of

[1] Actually, Tychonoff proved his theorem just for this special case. Several years later, it was observed that his proof worked for general Cartesian products of compact spaces.

$J(x)$ to K. It is clear that $\Phi(x)\colon K \to R$ is continuous and, according to Theorem 7 of the preceding chapter, there is a $\psi \in X^*$ such that $\psi(x) = \|x\|$ and $\|\psi\|_* = 1$. Therefore, the maximum value of $|\Phi(x)|$ on K is $\|x\|$. Consequently, Φ is an isomorphism of X onto its image, a subspace of $C(K)$. □

Corollary 2 *Let X be a normed linear space. Then the closed unit ball B^* of its dual space X^* possesses an extreme point.*

Proof We consider X^* as a locally convex topological space with its weak-$*$ topology. According to Alaoglu's Theorem, B^* is convex and compact. The Krein-Milman Lemma tells us that B^* possesses an extreme point. □

Alaoglu's Theorem does not tell us that the closed unit ball of the dual of a normed linear space is sequentially compact with respect to the weak-$*$ topology. For instance, for $X = \ell^\infty$, B^*, the closed unit ball of X^*, is not weak-$*$ sequentially compact. Indeed, the sequence $\{\psi_n\} \subseteq B^*$ defined for each n by

$$\psi_n(\{x_k\}) = x_n \text{ for all } \{x_k\} \in \ell^\infty,$$

fails to have a weak-$*$ convergent subsequence. Alaoglu's Theorem is a generalization of Helly's Theorem from the viewpoint of compactness, not sequential compactness. By Helly's Theorem, B^* is weak-$*$ sequentially compact if X is separable, and the forthcoming Corollary 5 tells us that B^* also is weak-$*$ sequentially compact if X is reflexive.

PROBLEMS

1. For X a normed linear space with closed unit ball B, suppose that the function $f\colon B \to [-1, 1]$ has the property that whenever $u, v, u+v$, and λu belong to B, $f(u+v) = f(u) + f(v)$ and $f(\lambda u) = \lambda f(u)$. Show that f is the restriction to B of a linear functional on all of X which belongs to the closed unit ball of X^*.
2. Let X be a normed linear space and K be a bounded convex weak-$*$ closed subset of X^*. Show that K possesses an extreme point.
3. Show that any non-empty weakly open set in an infinite dimensional normed linear space is unbounded with respect to the norm.
4. Use the Baire Category Theorem and the preceding problem to show that the weak topology on an infinite dimensional Banach space is not metrizable by a complete metric.
5. Is every Banach space isomorphic to the dual of a Banach space?

19.2 REFLEXIVITY AND WEAK COMPACTNESS: KAKUTANI'S THEOREM

Proposition 3 *Let X be a normed linear space. Then the natural embedding $J\colon X \to X^{**}$ is a topological homeomorphism between the locally convex topological vector spaces X and $J(X)$, where X has the weak topology and $J(X)$ has the weak-$*$ topology.*

Proof Let x_0 belong to X. A neighborhood base for the weak topology at $x_0 \in X$ is defined by sets of the form, for $\epsilon > 0$ and $\psi_1, \ldots, \psi_n \in X^*$,

$$\mathcal{N}_{\epsilon, \psi_1, \ldots, \psi_n}(x_0) = \{x \in X \mid \ |\psi_i(x - x_0)| < \epsilon \text{ for } 1 \leq i \leq n\}.$$

Now $[J(x) - J(x_0)]\psi_i = \psi_i(x - x_0)$ for each $x \in X$ and $1 \le i \le n$, and therefore

$$J(\mathcal{N}_{\epsilon,\psi_1,\ldots,\psi_n}(x_0)) = J(X) \cap \mathcal{N}_{\epsilon,\psi_1,\ldots,\psi_n}(J(x_0)).$$

Therefore, J maps a base for the weak topology at the origin in X onto a base for the weak-$*$ topology at the origin in $J(X)$. Thus J is a homeomorphism from X, with the weak topology, onto $J(X)$, with the weak-$*$ topology. $\square$

Kakutani's Theorem *A Banach space is reflexive if and only if its closed unit ball is weakly compact.*

Proof Let X be a Banach space. Denote the closed unit balls in X and X^{**} by B and B^{**}, respectively. Assume that X is reflexive. The natural embedding is an isomorphism and therefore J is a one-to-one map of B onto B^{**}. On the other hand, according to Proposition 3, J is a homeomorphism from B, with the weak topology, onto B^{**}, with the weak-$*$ topology. But by Alaoglu's Theorem, applied with X replaced by X^*, B^{**} is weak-$*$ compact, so any topological space homeomorphic to it also is compact. In particular, B is weakly compact.

Now assume that that B is weakly compact. The continuous image of compact topological spaces is compact. It follows from Proposition 3 that $J(B)$ is compact with respect to the weak-$*$ topology. Hence $J(B)$ is closed since it is clear that the weak-$*$ topology is Hausdorff. Of course, $J(B)$ is convex. To establish the reflexivity of X, we argue by contradiction. Assume that X is not reflexive. Let T belong to $B^{**} \sim J(B)$. Apply Corollary 26 of the Hyperplane Separation Theorem in the case that X is replaced by X^{**} and $W = J(X)$. Thus there is a functional $\psi \in X^*$ for which $\|\psi\|_* = 1$ and

$$T(\psi) < \inf_{S \in J(B)} S(\psi) = \inf_{x \in B} \psi(x).$$

The right-hand infimum equals -1, since $\|\psi\|_* = 1$. Therefore $T(\psi) < -1$. This is a contradiction since $\|T\| \le 1$ and $\|\psi\|_* = 1$. Therefore, X is reflexive. $\square$

Corollary 4 *Every closed, bounded, convex subset of a reflexive Banach space is weakly compact.*

Proof Let X be a Banach space. According to Kakutani's Theorem, the closed unit ball of X is weakly compact. Hence so is any closed ball. According to Mazur's Theorem, every closed, convex subset of X is weakly closed. Therefore any closed, convex, bounded subset of X is a weakly closed subset of a weakly compact set and hence must be weakly compact. $\square$

Corollary 5 *Let X be a reflexive Banach space. Then the closed unit ball of its dual space, B^*, is sequentially compact with respect to the weak-$*$ topology.*

Proof Since X is reflexive, the weak topology on B^* is the same as the weak-$*$ topology. Therefore, by Alaoglu's Theorem, B^* is weakly compact. It follows from Kakutani's Theorem that X^* is reflexive. Therefore, by Theorem 17 of the preceding chapter, every bounded sequence in X^* has a weak-$*$ convergent subsequence. But B^* is weak-$*$ closed. Therefore, B^* is sequentially compact with respect to the weak-$*$ topology. $\square$

PROBLEMS

6. Show that every weakly compact subset of a normed linear space is bounded with respect to the norm.
7. Show that the closed unit ball B^* of the dual X^* of a Banach space X has an extreme point.
8. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two compact, Hausdorff topologies on a set $\mathcal{S}$ for which $\mathcal{T}_1 \subseteq \mathcal{T}_2$. Show that $\mathcal{T}_1 = \mathcal{T}_2$.
9. Let X be a normed linear space containing the subspace Y. For $A \subseteq Y$, show that the weak topology on A induced by Y^* is the same as the topology A inherits as a subspace of X with its weak topology.
10. Argue as follows to show that a Banach space X is reflexive if and only if its dual space X^* is reflexive.
    (i) If X is reflexive, show that the weak and weak-$*$ topologies on B^* are the same, and infer from this that X^* is reflexive.
    (ii) If X^* is reflexive, use part (i) and Proposition 15 of Chapter 18 to show that X is reflexive.
11. For X a Banach space, by the preceding problem, if X is reflexive, then so in X^*. Conclude that X is not reflexive if there is a closed subspace of X^* that is not reflexive. Let K be an infinite, compact, Hausdorff space and $\{x_n\}$ an enumeration of a countably infinite subset of K. Define the operator $T\colon l^1 \to [C(K)]^*$ by

$$[T(\{\eta_k\})](f) = \sum_{k=1}^{\infty} \eta_k \cdot f(x_k) \text{ for all } \{\eta_k\} \in l^1 \text{ and } f \in C(k).$$

    Show that T is an isomorphism and therefore, since l^1 is not reflexive, neither is $T(l^1)$ and therefore neither is $C(K)$. Use a dimension counting argument to show that $C(K)$ is reflexive if K is a finite set.
12. If Y is a linear subspace of a Banach space X, we define the *annihilator* $Y^\perp$ to be the subspace of X^* consisting of those $\psi \in X^*$ for which $\psi = 0$ on Y. If Y is a subspace of X^*, we define Y^0 to be the subspace of vectors in X for which $\psi(x) = 0$ for all $\psi \in Y$.
    (i) Show that $Y^\perp$ is a closed linear subspace of X^*.
    (ii) Show that $(Y^\perp)^0 = \overline{Y}$.
    (iii) If X is reflexive and Y is subspace of X^*, show that $Y^\perp = J(Y^0)$.

19.3 COMPACTNESS AND WEAK SEQUENTIAL COMPACTNESS: THE EBERLEIN-ŠMULIAN THEOREM

Theorem 6 (Goldstine's Theorem) *Let X be a normed linear space, B the closed unit ball of X, and B^{**} the closed unit ball of X^{**}. Then the weak-$*$ closure of $J(B)$ is B^{**}.*

Proof According to Corollary 9 of the preceding chapter, J is an isometry. Thus $J(B) \subseteq B^{**}$. Let C be the weak-$*$ closure of $J(B)$. We leave it as an exercise to show that B^{**} is weak-$*$ closed. Therefore, $C \subseteq B^{**}$. Since B is convex and J is linear, $J(B)$ is convex. Proposition 22 of the preceding chapter tells us that, in a locally convex topological vector space, the closure of a convex set is convex. Thus C is a convex set that is closed with respect to the weak-$*$ topology. Suppose $C \neq B^{**}$. Let T belong to $B^{**} \sim C$. We now invoke the Hyperplane Separation Theorem in the case that X is replaced by $(X^*)^*$ and $(X^*)^*$ is considered as a locally convex topological vector space with the weak-$*$ topology; see Corollary 26 of the preceding chapter. Thus there is some $\psi \in X^*$ for which $\|\psi\| = 1$ and

$$T(\psi) < \inf_{S \in C} S(\psi). \tag{2}$$

Observe that since C contains $J(B)$,

$$\inf_{S \in C} S(\psi) \leq \inf_{x \in B} \ \psi(x) = -1.$$

Therefore, $T(\psi) < -1$. This is a contradiction since $\|T\| \leq 1$ and $\|\psi\|_* = 1$. Consequently, $C = J(B)$ and the proof is complete. □

Lemma 7 *Let X be a normed linear space and W a finite dimensional subspace of X^*. Then there is a finite subset F of X for which*

$$\textit{for all } \psi \in W \ \ \|\psi\|_*/2 \leq \max_{x \in F} \psi(x). \tag{3}$$

Proof Since W is finite dimensional, its closed unit sphere $S^* = \{\psi \in W \mid \|\psi\| = 1\}$ is compact and therefore is totally bounded. Choose a finite subset $\{\psi_1, \ldots, \psi_n\}$ of S^* for which $S^* \subseteq \cup_{k=1}^n B(\psi_k, 1/4)$. For $1 \leq k \leq n$, choose a unit vector x_k in X for which $\psi_k(x_k) > 3/4$. Let ψ belong to S^*. Observe that

$$\psi(x_k) = \psi_k(x_k) + [\psi - \psi_k]x_k \geq 3/4 + [\psi - \psi_k]x_k \text{ for } 1 \leq k \leq n.$$

If we choose k such that $\|\psi - \psi_k\| < 1/4$, then since $\|x_k\| = 1$, $\psi(x_k) \geq 1/2 = 1/2\|\psi\|_*$. Thus (3) holds if $F = \{x_1, \ldots, x_k\}$ and $\psi \in W$ has $\|\psi\|_* = 1$. It therefore holds for all $\psi \in W$. □

Theorem 8 (The Eberlein-Šmulian Theorem) *The closed unit ball B of a Banach space X is weakly compact if and only if it is weakly sequentially compact.*

Proof We first assume that B is weakly compact. Kakutani's Theorem tells us that X is reflexive. According to Theorem 17 of the preceding chapter, every bounded sequence in X has a weakly convergent subsequence. Since B is weakly closed, B is weakly sequentially compact.

To prove the converse, assume that B is weakly sequentially compact. To show that B is compact it suffices, by Kakutani's Theorem, to show that X is reflexive[2]. Let T belong to B^{**}. Goldstine's Theorem tells us that T belongs to the weak-$*$ closure of $J(B)$. We will use the preceding lemma to show that T belongs to $J(B)$.

Choose $\psi_1 \in B^*$. Since T belongs to the weak-$*$ closure of $J(B)$, we may choose $x_1 \in B$ for which $J(x_1)$ belongs to $\mathcal{N}_{1,\psi_1}(T)$. Define $N(1) = 1$ and $W_1 = \text{span}[\{T, J(x_1)\}] \subseteq X^{**}$. Let n be a natural number for which there has been defined a natural number $N(n)$,

[2]This elegant proof that sequential compactness of the closed unit ball implies reflexivity is due to R. J. Whitley, "An elementary proof of the Eberlein-Šmulian Theorem," *Mathematische Annalen*, 1967.

a subset $\{x_k\}_{1\le k\le n}$ of B, a subset $\{\psi_k\}_{1\le k\le N(n)}$ of X^* and we have defined $W_n = \operatorname{span}[\{T,\, J(x_1),\, \ldots,\, J(x_n)\}]$. Since T belongs to the weak-$*$ closure of $J(B)$, we may choose $x_{n+1}\in B$ for which

$$J(x_{n+1}) \in \mathcal{N}_{1/(n+1),\psi_1,\ldots,\psi_{N(n)}}(T). \tag{4}$$

Define

$$W_{n+1} = \operatorname{span}[\{T,\, J(x_1),\, \ldots,\, J(x_{n+1})\}]. \tag{5}$$

It follows from the preceding lemma, in the case X is replaced by X^*, that there is a natural number $N(n+1) > N(n)$ and a finite subset $\{\psi_k\}_{N(n)<k\le N(n+1)}$ of X^* for which

$$\|S\|/2 < \max_{N(n)<k\le N(n+1)} S(\psi_k) \text{ for all } S\in W_{n+1}. \tag{6}$$

We have therefore inductively defined a strictly increasing sequence of natural numbers $\{N(n)\}$, a sequence $\{x_n\}$ in B, a sequence $\{\psi_n\}$ in X^*, and a sequence $\{W_n\}$ of subspaces of X^{**} for which (4) and (6) hold. Since $\{W_n\}$ is an ascending sequence for which (6) holds for every index n,

$$\|S\|/2 \le \sup_{1\le k<\infty} S(\psi_k) \text{ for all } S\in W \equiv \overline{\operatorname{span}}[\{T,\, J(x_1),\, \ldots,\, J(x_n),\ldots\}]. \tag{7}$$

Since (4) holds for all n,

$$|(T - J(x_m))[\psi_n]| < 1/m \text{ if } n \le N(m-1). \tag{8}$$

Since B is sequentially compact, there is a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ that converges weakly to $x\in B$. Mazur's Theorem tells us that a sequence of convex combinations of the terms of the sequence $\{x_{n_k}\}$ converges strongly to x. The image under J of this sequence of convex combinations converges strongly to $J(x)$ in X^{**}. Thus $J(x)$ belongs to W. But T also belongs to W. Therefore, $T - J(x)$ belongs to W. We claim that $T = J(x)$. In view of (7) to verify this claim it is necessary and sufficient to show that

$$(T - J(x))[\psi_n] = 0 \text{ for all } n. \tag{9}$$

Fix a natural number n. Observe that for each index k,

$$(T - J(x))[\psi_n] = (T - J(x_{n_k}))[\psi_n] + (J(x_{n_k}) - J(x))[\psi_n].$$

It follows from (8) that if $N(n_k - 1) > n$, then $|(T - J(x_{n_k}))[\psi_n]| < 1/n_k$. On the other hand,

$$(J(x_{n_k}) - J(x))[\psi_n] = \psi_n(x_{n_k} - x) \text{ for all } k$$

and $\{x_{n_k}\}$ converges weakly to x. Thus

$$(T - J(x))[\psi_n] = \lim_{k\to\infty}(T - J(x_{n_k}))[\psi_n] + \lim_{k\to\infty}(J(x_{n_k}) - J(x))[\psi_n] = 0. \qquad \square$$

We gather Kakutani's Theorem and the Eberlein-Šmulian Theorem into the following statement.

Characterization of Weak Compactness *Let B be the closed unit ball of a Banach space X. Then the following three assertions are equivalent:*

(i) X is reflexive;

(ii) B is weakly compact;

(iii) B is weakly sequentially compact.

PROBLEMS

13. In a general topological space that is not metrizable a sequence may converge to more than one point. Show that this cannot occur for the W-weak topology on a normed linear space X, where W is a subspace of X^* that separates points in X.
14. Show that there is a bounded sequence in $L^\infty[0,1]$ that fails to have a weakly convergent subsequence. Show that the closed unit ball of $C[a,b]$ is not weakly compact.
15. Let K be a compact metric space with infinitely many points. Show that there is a bounded sequence in $C(K)$ that fails to have a weakly convergent subsequence (see Problem 11), but every bounded sequence of continuous linear functionals on $C(K)$ has a subsequence that converges pointwise to a continuous linear functional on $C(K)$.

19.4 METRIZABILITY OF WEAK TOPOLOGIES

If the weak topology on the closed unit ball of a Banach space is metrizable, then the Eberlein-Šmulian Theorem is an immediate consequence of the equivalence of compactness and sequential compactness for a metric space. To better appreciate this theorem, we now establish some metrizability properties of weak topologies.

Theorem 9 *Let X be an infinite dimensional normed linear space. Then neither the weak topology on X nor the weak-$*$ topology on X^* is metrizable.*

Proof According to Theorem 14 of Chapter 15, the weak topology on X is not induced by a metric. We argue by contradiction to show that the weak-$*$ topology on X^* is not induced by a metric. Otherwise, there is a metric $\rho^*\colon X^* \times X^* \to [0, \infty)$ that induces the weak-$*$ topology on X^*. Fix a natural number n. Consider the weak-$*$ neighborhood $\{\psi \in X^* \mid \rho^*(\psi, 0) < 1/n\}$ of 0. We may choose a finite subset A_n of X and $\epsilon_n > 0$ for which

$$\{\psi \in X^* \mid \ |\psi(x)| < \epsilon_n \text{ for all } x \in A_n\} \subseteq \{\psi \in X^* \mid \ \rho^*(\psi, 0) < 1/n\}.$$

Define X_n to be the linear span of A_n. Then

$$\{\psi \in X^* \mid \ \psi(x) = 0 \text{ for all } x \in X_n\} \subseteq \{\psi \in X^* \mid \ \rho^*(\psi, 0) < 1/n\}. \tag{10}$$

Since X_n is finite dimensional, it is closed and is a proper subspace of X since X is infinite dimensional. It follows from Corollary 11 of the preceding chapter that there is a non-zero functional $\psi_n \in X^*$ which vanishes on X_n. Define $\varphi_n = n \cdot \psi_n / \|\psi_n\|$. Observe that $\|\varphi_n\| = n$ and, by (10), that $\rho^*(\varphi_n, 0) < 1/n$. Therefore, $\{\varphi_n\}$ is an unbounded sequence in X^* that converges pointwise to 0. This contradicts the Uniform Boundedness Theorem. Thus, the weak-$*$ topology on X^* is not metrizable. □

Theorem 10 *Let X be a normed linear space and W a separable subspace of X^* that separates points in X. Then the W-weak topology on the closed unit ball B of X is metrizable.*

Proof Since W is separable, $B^* \cap W$ also is separable, where B^* is the closed unit ball of X^*. Choose a countable dense subset $\{\psi_k\}_{k=1}^\infty$ of $B^* \cap W$. Define $\rho\colon B \times B \to \mathbf{R}$ by

$$\rho(u, v) = \sum_{k=1}^\infty \frac{1}{2^k} \cdot |\psi_k(u - v)| \text{ for all } u, v \in B.$$

This is properly defined since each ψ_k belongs to B^*. We first claim that ρ is a metric on B. The symmetry and triangle inequality are inherited by ρ from the linearity of ψ_k's. On the other hand, since W separates points in X, any dense subset of $S^* \cap W$ also separates points in X. Therefore, for $u, v \in B$ with $u \neq v$, there is a natural number k for which $\psi_k(u - v) \neq 0$ and therefore $\rho(u, v) > 0$. Thus ρ is a metric on B. Observe that for each natural number n, since each ψ_k belongs to B^*,

$$\frac{1}{2^n}\left[\sum_{k=1}^{n} |\psi_k(x)|\right] \leq \rho(x, 0) \leq \sum_{k=1}^{n} |\psi_k(x)| + 1/2^n \text{ for all } x \in B. \tag{11}$$

We leave it as an exercise to conclude from these inequalities and the denseness of $\{\psi_k\}_{k=1}^{\infty}$ in $B^* \cap W$ that $\{x \in B \mid \rho(x, 0) < 1/n\}_{n=1}^{\infty}$ is a base at the origin for the W-weak topology on B. Therefore, the topology induced by the metric ρ is the W-weak topology on B. $\square$

Corollary 11 *Let X be a normed linear space.*

(i) The weak topology on the closed unit ball B of X is metrizable if X^* is separable.

(ii) The weak-$*$ topology on the closed unit ball B^* of X^* is metrizable if X is separable.

Theorem 12 *Let X be a reflexive Banach space. Then the weak topology on the closed unit ball B is metrizable if and only if X is separable.*

Proof Since X is reflexive, Theorem 14 of the preceding chapter tells us that if X is separable, so is X^*. Therefore, by the preceding corollary, if X is separable, then the weak topology on B is metrizable. Conversely, suppose that the weak topology on B is metrizable. Let $\rho\colon B \times B \to [0, \infty)$ be a metric that induces the weak topology on B. Let n be a natural number. We may choose a finite subset F_n of X^* and $\epsilon_n > 0$ for which

$$\{x \in B \mid \ |\psi(x)| < \epsilon_n \text{ for all } \psi \in F_n\} \subseteq \{x \in B \mid \ \rho(x, 0) < 1/n\}.$$

Therefore,

$$\left[\bigcap_{\psi \in F_n} \ker \psi\right] \cap B \subseteq \{x \in B \mid \ \rho(x, 0) < 1/n\}. \tag{12}$$

Define Z to be the closed linear span of $\bigcup_{n=1}^{\infty} F_n$. Then Z is separable since finite linear combinations, with rational coefficients, of functionals in $\bigcup_{n=1}^{\infty} F_n$ is a countable dense subset of Z. We claim that $Z = X^*$. Otherwise, Corollary 11 of the preceding chapter tells us that there is a non-zero $S \in (X^*)^*$, which vanishes on Z. Since X is reflexive, there is some $x_0 \in X$ for which $S = J(x_0)$. Thus $x_0 \neq 0$ and $\psi(x_0) = 0$ for all $\psi \in Z$. According to (12), $\rho(x_0, 0) < 1/n$ for all n. Hence $x_0 \neq 0$ but $\rho(x_0, 0) = 0$. This is a contradiction. Therefore, X^* is separable. Theorem 13 of the preceding chapter tells us that X also is separable. $\square$

PROBLEMS

16. Show that the dual of an infinite dimensional normed linear space also is infinite dimensional.
17. Complete the last step of the proof of Theorem 10 by showing that the inequalities (11) imply that the metric ρ induces the W-weak topology.
18. Let X be a Banach space, W a closed subspace of its dual X^*, and ψ_0 belong to $X^* \sim W$. Show that if either W or X is reflexive, then there is a vector x_0 in X for which $\psi_0(x_0) \neq 0$ but $\psi(x_0) = 0$ for all $\psi \in W$. Exhibit an example of an infinite dimensional closed subspace W of X^* for which this separation property fails.

CHAPTER 20

Continuous Linear Operators on Hilbert Spaces

Contents

20.1 The Inner-Product and Orthogonality 393
20.2 Separability, Bessel's Inequality, and Orthonormal Bases 397
20.3 The Dual Space and Weak Sequential Compactness 399
20.4 Symmetric Operators 404
20.5 Compact Operators 409
20.6 The Hilbert-Schmidt Theorem 411
20.7 The Riesz-Schauder Theorem: Characterization of Fredholm Operators . . . 414

The inner-product $\langle u, v\rangle$ of two vectors $u = (u_1, \ldots, u_n)$ and $v \in (v_1, \ldots, v_n)$ in Euclidean space $\mathbf{R}^n$ is defined by

$$\langle u, v\rangle = \sum_{k=1}^{n} u_k v_k.$$

This is called the Euclidean inner-product, and the Euclidean norm $\|\cdot\|$ is defined by

$$\|u\| = \sqrt{\langle u, u\rangle} \text{ for all } u \in \mathbf{R}^n.$$

With respect to the Euclidean inner-product, there is the important notion of orthogonality, which brings a geometric viewpoint to the study of problems in finite dimensional spaces: subspaces have orthogonal complements and the solvability of systems of equations can be determined by orthogonality relations. The inner-product also brings to light interesting classes of linear operators that have quite special structure: prominent among these are the symmetric operators for which there is a beautiful eigenvector representation. In this chapter, we consider Banach spaces H that have an inner-product that is related to the norm as it is in the Euclidean spaces. These spaces are called Hilbert spaces. We prove that if V is a closed subspace of a Hilbert space H, then H is the direct sum of V and its orthogonal complement. Based on this structural property, we prove the Riesz-Fréchet Representation Theorem, which characterizes the dual space of a Hilbert space. It follows from this, using Helly's Theorem, that every bounded sequence in a Hilbert space has a weakly convergent subsequence. We prove Bessel's Inequality from which it follows that a Hilbert space is separable if and only if it has an orthonormal basis. The chapter concludes with an examination of bounded, symmetric operators and compact operators on a Hilbert space, in preparation for the proof of two theorems: the Hilbert-Schmidt Theorem regarding an eigenvalue expansion for compact, symmetric operators and the Riesz-Schauder Theorem regarding the Fredholm properties of compact perturbations of the identity operator.

20.1 THE INNER-PRODUCT AND ORTHOGONALITY

Definition *Let H be a linear space. A function $\langle \cdot, \cdot \rangle : H \times H \to \mathbf{R}$ is called an* **inner-product** *on H provided that for all x_1, x_2, x and $y \in H$ and real numbers α and β,*

(i) $\langle \alpha x_1 + \beta x_2, y \rangle = \alpha \langle x_1, y \rangle + \beta \langle x_2, y \rangle$,

(ii) $\langle x, y \rangle = \langle y, x \rangle$,

(iii) $\langle x, x \rangle > 0$ *if* $x \neq 0$.

A linear space H together with an inner-product on H is called an **inner-product space**.

Property (ii) is called **symmetry**. From (i) and (ii) it follows that $\langle x, \alpha y_1 + \beta y_2 \rangle = \alpha \langle x, y_1 \rangle + \beta \langle x, y_2 \rangle$: this property, together with (i), is called **bilinearity**.

Among infinite dimensional linear spaces, two examples of inner-product spaces come to mind. For two sequences $x = \{x_k\}$ and $y = \{y_k\} \in \ell^2$, the ℓ^2 inner-product, $\langle x, y \rangle$, is defined by

$$\langle x, y \rangle = \sum_{k=1}^{\infty} x_k y_k.$$

For a general measure space $(X, \mathcal{M}, \mu)$, and f and $g \in L^2(X, \mu)$, the L^2 inner-product, $\langle f, g \rangle$, is defined by

$$\langle f, g \rangle = \int_X f \cdot g \, d\mu.$$

For a vector h in an inner-product space H, define

$$\|h\| = \sqrt{\langle h, h \rangle}.$$

Then $\|\cdot\|$ is a norm on H called the norm induced by the inner-product $\langle \cdot, \cdot \rangle$.

The Cauchy-Schwarz Inequality *For any two vectors u, v in an inner-product space H,*

$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\|.$$

To verify this, observe that

$$0 \leq \|u + tv\|^2 = \|u\|^2 + 2t\langle u, v \rangle + t^2\|v\|^2 \text{ for all } t \in \mathbf{R}.$$

The quadratic polynomial in t defined by the right-hand side fails to have distinct real roots and therefore its discriminant is not positive, that is, the Cauchy-Schwarz Inequality holds.

We claim that $\|\cdot\|$ is a norm. Indeed, The only property of a norm that is not evident is the triangle inequality. This, however, is a consequence of the Cauchy-Schwarz Inequality since, for two vectors $u, v \in H$,

$$\|u+v\|^2 = \langle u+v, u+v \rangle = \|u\|^2 + 2\langle u, v \rangle + \|v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2.$$

The following identity characterizes norms that are induced by an inner-product; see Problem 7.

The Parallelogram Identity *For any two vectors u, v in an inner-product space H,*

$$\|u - v\|^2 + \|u + v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

To verify this identity, just add the following two equalities:

$$\|u-v\|^2 = \|u\|^2 - 2\langle u, v\rangle + \|v\|^2;$$
$$\|u+v\|^2 = \|u\|^2 + 2\langle u, v\rangle + \|v\|^2.$$

Definition *An inner-product space H is called a* **Hilbert space** *provided that it is complete with respect to the norm induced by the inner-product.*

For a general measure space $(X, \mathcal{M}, \lambda)$, according to the Riesz-Fischer Theorem $(X, \mathcal{M}, \lambda)$, the inner-product space $L^2(X, \mu)$ is complete and, as a consequence, so is ℓ^2.

Proposition 1 *Let K be a non-empty, closed, convex subset of a Hilbert space H and h_0 be a vector in H that lies outside of K. Then there is exactly one vector $h_* \in K$ that is closest to h_0, in the sense that*

$$\|h_0 - h_*\| = \text{dist}_K(h_0) = \inf_{h\in K} \|h_0 - h\|.$$

Proof By replacing K by $K - h_0$, we may assume that $h_0 = 0$. Let $\{h_n\}$ be a sequence in K for which

$$\lim_{n\to\infty} \|h_n\| = \inf_{h\in K} \|h\|. \tag{1}$$

It follows from the parallelogram identity and the convexity of K that for each m and n,

$$\|h_n\|^2 + \|h_m\|^2 = 2\left\|\frac{h_n+h_m}{2}\right\|^2 + 2\left\|\frac{h_n-h_m}{2}\right\|^2 \geq 2\cdot \inf_{h\in K}\|h\|^2 + 2\cdot\left\|\frac{h_n-h_m}{2}\right\|^2. \tag{2}$$

From (1) and (2) we conclude that $\{h_n\}$ is a Cauchy sequence. Since H is complete and K is closed, $\{h_n\}$ converges strongly to $h_* \in K$. By the continuity of the norm, $\|h_*\| = \inf_{h\in K}\|h\|$. This point in K that is closest to the origin is unique. Indeed, if h^* is another vector in K that is closest to the origin, then, if we substitute h^* for h_n and h_* for h_m in inequality (2), we have

$$0 = \|h^*\|^2 + \|h_*\|^2 - 2\cdot \inf_{h\in K}\|h\|^2 \geq 2\cdot \left\|\frac{h^*-h_*}{2}\right\|^2.$$

Thus $h^* = h_*$. □

Definition *Two vectors u, v in the inner-product space H are said to be* **orthogonal** *provided that $\langle u, v\rangle = 0$. A vector u is said to be* **orthogonal** *to a subset S of H provided that it is orthogonal to each vector in S. We denote by $S^\perp$ the collection of vectors in H that are orthogonal to S.*

We leave it as an exercise to conclude from the Cauchy-Schwarz Inequality that if S is a subset of an inner-product space H, then $S^\perp$ is a closed subspace of H. The following theorem is fundamental.

Theorem 2 *Let V be a closed subspace of a Hilbert space H. Then H has the orthogonal direct sum decomposition*

$$H = V \oplus V^\perp. \tag{3}$$

Proof Let h_0 be a vector in H that lies outside of V. According to the preceding proposition, there is a unique vector $h^* \in V$ that is closest to h_0. Let h be any vector in V. For a real number t, since V is a linear space, the vector $h^* - th$ belongs to V and therefore

$$\begin{aligned}\langle h_0 - h^*, h_0 - h^* \rangle &= \|h_0 - h^*\|^2 \\ &\leq \|h_0 - (h^* - th)\|^2 \\ &= \langle h_0 - h^*, h_0 - h^* \rangle + 2t \cdot \langle h_0 - h^*, h \rangle + t^2 \langle h, h \rangle.\end{aligned}$$

Hence

$$0 \leq 2t \cdot \langle h_0 - h^*, h \rangle + t^2 \langle h, h \rangle \text{ for all } t \in R,$$

and therefore $\langle h_0 - h^*, h \rangle = 0$. Thus, the vector $h_0 - h^*$ is orthogonal to V. Observe that $h_0 = h^* + [h_0 - h^*]$. We conclude that $H = V + V^\perp$ and since $V \cap V^\perp = \{0\}$, $H = V \oplus V^\perp$. $\square$

We denote $\mathcal{L}(H, H)$ by $\mathcal{L}(H)$. In view of (3), for a closed subspace V of H, we call $V^\perp$ the **orthogonal complement** of V in H and refer to (3) as an **orthogonal decomposition** of H. For $u \in H$ we have

$$u = v + w, \text{ where } v \in V, w \in V^\perp,$$

and define $v = P(u)$, so that $w = (\mathrm{Id} - P)u$. The operator $P \in \mathcal{L}(H)$ is called the **orthogonal projection** of H onto V. Observe that $\mathrm{Id} - P$ is the orthogonal projection of H onto $V^\perp$.

Proposition 3 *Let P be the orthogonal projection of a Hilbert space H onto a non-zero, closed subspace V of H. Then $\|P\| = 1$ and*

$$\langle Pu, v \rangle = \langle u, Pv \rangle \quad \textit{for all } u, v \in H. \tag{4}$$

Proof Let the vector u belong to H. Then

$$\|u\|^2 = \langle P(u) + (\mathrm{Id} - P)(u), P(u) + (\mathrm{Id} - P)(u) \rangle = \|P(u)\|^2 + \|(\mathrm{Id} - P)(u)\|^2 \geq \|P(u)\|^2,$$

and hence $\|P(u)\| \leq \|u\|$. We therefore have $\|P\| \leq 1$ and conclude that $\|P\| = 1$ since $P(v) = v$ for each non-zero vector in V. If the vector v also belongs to H, then

$$\langle P(u), (\mathrm{Id} - P)(v) \rangle = \langle (\mathrm{Id} - P)(u), P(v) \rangle = 0,$$

so that

$$\langle P(u), v \rangle = \langle P(u), P(v) \rangle = \langle u, P(v) \rangle.$$ $\square$

A Banach space X is said to be **complemented** provided that every closed subspace of X has a closed linear complement. A Banach space X is said to be **Hilbertable** provided that there is an inner-product on X for which the induced norm is equivalent to the given norm. It follows from Theorem 2 that a Hilbertable Banach space is complemented. A theorem of Joram Lindenstrauss and Lior Tzafriri asserts that the converse is true: If a Banach space is complemented, then it is Hilbertable.

The proofs of many results that were established for general Banach spaces are much simpler for the special case of Hilbert spaces, as is seen in Problems 11–15.

PROBLEMS

In the following problems, H is a Hilbert space.

1. Let $[a,\, b]$ be a closed, bounded interval of real numbers. Show that the $L^2[a,\, b]$ inner-product is also an inner-product on $C[a,\, b]$. Is $C[a,\, b]$, considered as an inner-product space with the $L^2[a,\, b]$ inner-product, a Hilbert space?
2. Show that the maximum norm on $C[a,\, b]$ is not induced by an inner-product and neither is the usual norm on ℓ^1.
3. Let H_1 and H_2 be Hilbert spaces. Show that the Cartesian product $H_1 \times H_2$ also is a Hilbert space with an inner-product with respect to which $H_1 \times \{0\} = [\{0\} \times H_2]^{\perp}$.
4. Show that if S is a subset of an inner-product space H, then $S^{\perp}$ is a closed subspace of H.
5. Let S be a subset of H. Show that $S = (S^{\perp})^{\perp}$ if and only if S is a closed subspace of H.
6. (Polarization Identity) Show that for any two vectors $u, v \in H$,
$$\langle u,\, v\rangle = \frac{1}{4}[\|u+v\|^2 - \|u-v\|^2].$$
7. (Jordan-von Neumann) Let X be a linear space normed by $\|\cdot\|$. Use the polarization identity to show that a norm $\|\cdot\|$ is induced by an inner-product if and only if the parallelogram identity holds.
8. Let V be a closed subspace of H and P a projection of H onto V. Show that P is the orthogonal projection of H onto V if and only if (4) holds.
9. Let T belong to $\mathcal{L}(H)$. Show that T is an isometry if and only if
$$\langle T(u),\, T(v)\rangle = \langle u,\, v\rangle \text{ for all } u, v \in H.$$
10. Let V be a finite dimensional subspace of H and $\varphi_1, \ldots, \varphi_n$ a basis for V consisting of unit vectors, each pair of which is orthogonal. Show that the orthogonal projection P of H onto V is given by
$$P(h) = \sum_{k=1}^{n} \langle h,\, \varphi_k\rangle \varphi_k \text{ for all } h \in V.$$
11. For h a vector in H, show that the function $u \mapsto \langle h,\, u\rangle$ belongs to H^*.
12. For any vector $h \in H$, show that there is a bounded linear functional $\psi \in H^*$ for which
$$\|\psi\| = 1 \text{ and } \psi(h) = \|h\|.$$
13. Let V be a closed subspace of H and P the orthogonal projection of H onto V. For any normed linear space X and $T \in \mathcal{L}(V,\, X)$, show that $T \circ P$ belongs to $\mathcal{L}(H,\, X)$, and is an extension of $T\colon V \to X$ for which $\|T \circ P\| = \|T\|$.
14. Prove the Hyperplane Separation Theorem for H, considered as a locally convex topological vector space with respect to the strong topology, by directly using Proposition 1.
15. Use Proposition 1 to prove the Krein-Milman Lemma in a Hilbert space.

20.2 SEPARABILITY, BESSEL'S INEQUALITY, AND ORTHONORMAL BASES

Throughout this section, H is a Hilbert space.

Definition *A subset S of H is said to be* **orthogonal** *provided that every two vectors in S are orthogonal. If such a set has the further property that each vector in S is a unit vector, then S is said to be* **orthonormal.**

The General Pythagorean Identity *If $u_1, u_2, \ldots, u_n$ are n orthonormal vectors in H, and $\alpha_1, \ldots, \alpha_n$ are real numbers, then*

$$\|\alpha_1 u_1 + \cdots + \alpha_n u_n\|^2 = |\alpha_1|^2 + \cdots + |\alpha_n|^2.$$

This identity follows from an expansion of the right-hand side of the following identity:

$$\|\alpha_1 u_1 + \cdots + \alpha_n u_n\|^2 = \langle \alpha_1 u_1 + \cdots + \alpha_n u_n,\, \alpha_1 u_1 + \cdots + \alpha_n u_n \rangle.$$

Bessel's Inequality *For $\{\varphi_k\}$ an orthonormal sequence in H and h a vector in H,*

$$\sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle^2 \leq \|h\|^2.$$

To verify this inequality, fix an n and define $h_n = \sum_{k=1}^{n} \langle \varphi_k,\, h \rangle \varphi_k$. Then, by the General Pythagorean Identity,

$$\begin{aligned} 0 \leq \|h - h_n\|^2 &= \|h\|^2 - 2\langle h,\, h_n \rangle + \|h_n\|^2 \\ &= \|h\|^2 - 2\sum_{k=1}^{n} \langle h,\, \varphi_k \rangle \langle h,\, \varphi_k \rangle + \sum_{k=1}^{n} \langle h,\, \varphi_k \rangle^2 \\ &= \|h\|^2 - \sum_{k=1}^{n} \langle h,\, \varphi_k \rangle^2. \end{aligned}$$

Therefore,

$$\sum_{k=1}^{n} \langle \varphi_k,\, h \rangle^2 \leq \|h\|^2.$$

Take the limit as $n \to \infty$ to obtain Bessel's Inequality.

Proposition 4 *Let $\{\varphi_k\}$ be an orthonormal sequence in a Hilbert space H and the vector h belong to H. Then the series $\sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k$ converges strongly in H and the vector $h - \sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k$ is orthogonal to each φ_k.*

Proof For each n, define $h_n = \sum_{k=1}^{n} \langle \varphi_k,\, h \rangle \varphi_k$. By the General Pythagorean Identity, for each n and k,

$$\|h_{n+k} - h_n\|^2 = \sum_{i=n+1}^{n+k} \langle \varphi_i,\, h \rangle^2.$$

However, by Bessel's Inequality, the series $\sum_{i=1}^{\infty} \langle \varphi_i,\, h \rangle^2$ converges and hence $\{h_n\}$ is a Cauchy sequence in H. Since H is complete, $\sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k$ converges strongly to a vector $h_* \in H$. Fix a natural number m. Observe that if $n > m$, then $h - h_n$ is orthogonal to φ_m. By the continuity of the inner-product, $h - h_*$ is orthogonal to φ_m. $\square$

Definition *An orthonormal sequence $\{\varphi_k\}$ in a Hilbert space H is said to be* **complete** *provided that the only vector $h \in H$ that is orthogonal to every φ_k is $h = 0$*[1]*, and said to be an* **orthonormal basis** *provided that*

$$h = \sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k \text{ for all } h \in H. \tag{5}$$

Proposition 5 *An orthonormal sequence $\{\varphi_k\}$ in a Hilbert space H is complete if and only if it is an orthonormal basis.*

Proof First assume that $\{\varphi_k\}$ is complete. By the preceding proposition, $h - \sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k$ is orthogonal to each φ_k. Therefore, by the completeness of $\{\varphi_k\}$, (5) holds. Conversely, suppose (5) holds. Then if $h \in H$ is orthogonal to all φ_k, then

$$h = \sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k = \sum_{k=1}^{\infty} 0 \cdot \varphi_k = 0.$$

□

Example The countable collection of functions in $L^2[0,\, 2\pi]$ consisting of the constant function that takes the value $1/\sqrt{2\pi}$ and the functions $\{1/\sqrt{\pi} \cdot \sin kt,\, 1/\sqrt{\pi} \cdot \cos kt\}_{k=1}^{\infty}$ are a complete orthonormal sequence for the Hilbert space $L^2[0,\, 2\pi]$. Indeed, by the elementary trigonometric identities, this sequence is orthonormal. It follows from the Stone-Weierstrass Theorem that the linear span of this sequence is dense, with respect to the maximum norm, in the Banach space $C[a,\, b]$. Therefore, by the density of $C[a,\, b]$ in $L^2[0,\, 2\pi]$, the linear span of this sequence is dense in $L^2[0,\, 2\pi]$.

If a Hilbert space H possesses an orthonormal basis $\{\varphi_k\}$, then, since finite rational linear combinations of the φ_k's are a countable dense subset of H, H must be separable. It turns out that separability is also a sufficient condition for a Hilbert space to possess an orthonormal basis.

Theorem 6 *A Hilbert space is separable if and only if it possesses an orthonormal basis.*

Proof Let H be a separable Hilbert space. Let $\mathcal{F}$ be the collection of subsets of H that are orthonormal. Order $\mathcal{F}$ by inclusion. For every linearly ordered subcollection of $\mathcal{F}$, the union of the sets in the subcollection is an upper bound for the subcollection. By Zorn's Lemma, there is a maximal subset S_0 of $\mathcal{F}$. Since H is separable, S_0 is countable (see problem 16). Let $\{\varphi_k\}_{k=1}^{\infty}$ be an enumeration of S_0. If $h \in H, h \neq 0$, then, by Proposition 4, $h - \sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k$ is orthogonal to each φ_k. Therefore, $h - \sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k = 0$, for otherwise the union of S_0 and the normalization of $h - \sum_{k=1}^{\infty} \langle \varphi_k,\, h \rangle \varphi_k$ would be an orthonormal set that properly contains S_0. Therefore, $\{\varphi_k\}_{k=1}^{\infty}$ is an orthonormal basis for H. □

Recall that we proved that if $E \subseteq \mathbf{R}^n$ is Lebesgue measurable, then the Hilbert space $L^2(E, \mu_n)$ is separable, and consequently it has an orthonormal basis.

[1] Unfortunately, this is the third use of the adjective "complete"; we already have complete measure spaces and complete metric spaces.

PROBLEMS

16. Show that an orthonormal subset of a separable Hilbert space H must be countable. (Suggestion: Argue by contradiction. Otherwise, there is an uncountable set, each two vectors of which are a distance $\sqrt{2}$ apart.)
17. Let $\{\varphi_k\}$ be an orthonormal sequence in a Hilbert space H. Show that $\{\varphi_k\}$ converges weakly to 0 in H.
18. Let $\{\varphi_k\}$ be an orthonormal basis for the separable Hilbert space H. Show that $\{u_n\} \to u$ weakly if and only if $\{u_n\}$ is bounded and, for each k, $\lim_{n\to\infty}\langle u_n, \varphi_k\rangle = \langle u, \varphi_k\rangle$.
19. Show that any two infinite dimensional separable Hilbert spaces are isometrically isomorphic and that any such isomorphism preserves the inner-product.
20. Let V a closed, separable subspace of H for which $\{\varphi_k\}$ is an orthonormal basis. Show that the orthogonal projection of H onto V, P, is given by

$$P(h) = \sum_{k=1}^{\infty} \langle \varphi_k, h\rangle \varphi_k \text{ for all } h \in H.$$

21. (Parseval's Identities) Let $\{\varphi_k\}$ be an orthonormal basis for a Hilbert space H. Verify that

$$\|h\|^2 = \sum_{k=1}^{\infty} \langle \varphi_k, h\rangle^2 \text{ for all } h \in H.$$

Also verify that

$$\langle u, v\rangle = \sum_{k=1}^{\infty} a_k \cdot b_k \text{ for all } u, v \in H,$$

where, for each natural number k, $a_k = \langle u, \varphi_k\rangle$ and $b_k = \langle v, \varphi_k\rangle$.
22. Verify the assertions in the example of the orthonormal basis for $L^2[0, 2\pi]$.
23. Use Proposition 5 and the Stone-Weierstrass Theorem to show that for each $f \in L^2[-\pi, \pi]$,

$$f(x) = a_0/2 + \sum_{k=1}^{\infty} [a_k \cdot \cos kx + b_k \cdot \sin kx],$$

where the convergence is in $L^2[-\pi, \pi]$ and each

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\, dx \text{ and } b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\, dx.$$

20.3 THE DUAL SPACE AND WEAK SEQUENTIAL COMPACTNESS

For $1 \le p < \infty$, q the conjugate of p, and $(X, \mathcal{M}, \mu)$ a σ-finite measure space, the Riesz Representation Theorem explicitly describes a linear isometry of $[L^p(X, \mu)]^*$ onto $L^q(X, \mu)$. The case $p = 2$ extends to general Hilbert spaces.

The Riesz-Fréchet Representation Theorem *Let H be a Hilbert space. Define the operator $T\colon H \to H^*$ by assigning to each $h \in H$ the linear functional $T(h)\colon H \to \mathbf{R}$ defined by*

$$T(h)[u] = \langle h,\, u\rangle \text{ for all } u \in H. \tag{6}$$

Then T is a linear isometry of H onto H^*.

Proof Let h belong to H. Then $T(h)$ is linear. It follows from the Cauchy-Schwarz Inequality that the functional $T(h)\colon H \to \mathbf{R}$ is bounded and $\|T(h)\| \leq \|h\|$. But if $h \neq 0$, then $T(h)[h/\|h\|] = \|h\|$. Therefore, $\|T(h)\| = \|h\|$. Thus, T is an isometry. It is clear that T is linear. It remains to show that $T(H) = H^*$. Let $\psi_0 \neq 0$ belong to H^*. Since ψ_0 is continuous, its kernel is a closed, proper subspace of H. By Theorem 2, since $\ker \psi_0 \neq H$, we may choose a unit vector $h_* \in H$ that is orthogonal to $\ker\, \psi_0$. Define $h_0 = \psi_0(h_*)h_*$. We claim that $T(h_0) = \psi_0$. Indeed, for $h \in H$,

$$h - \frac{\psi_0(h)}{\psi_0(h_*)}h_* \in \ker\, \psi_0, \text{ so that } \langle h - \frac{\psi_0(h)}{\psi_0(h_*)}h_*,\, h_*\rangle = 0$$

and therefore $\psi_0(h) = \langle h_0,\, h\rangle = T(h_0)[h]$. Hence $T(H) = H^*$. □

In Problem 26, we outline von Neumann's use of this representation theorem to provided an elegant, simple proof of the Radon-Nikodym Theorem. In view of the Riesz-Fréchet Representation Theorem,

$$\{u_n\} \to u \text{ weakly in } H \text{ if and only if } \lim_{n\to\infty} \langle h,\, u_n\rangle = \langle h,\, u\rangle \text{ for all } h \in H.$$

Theorem 7 *Every bounded sequence in a Hilbert space has a weakly convergent subsequence.*

Proof Let $\{h_n\}$ be a bounded sequence in H. Define H_0 to be the closed linear span of $\{h_n\}$. Then H_0 is separable. For each n, define $\psi_n \in [H_0]^*$ by

$$\psi_n(h) = \langle h_n,\, h\rangle \text{ for all } h \in H_0.$$

Since $\{h_n\}$ is bounded, it follows from the Cauchy-Schwarz Inequality that $\{\psi_n\}$ also is bounded. Then $\{\psi_n\}$ is a bounded sequence of bounded linear functionals on the separable normed linear space H_0. According to Helly's Theorem, there is a subsequence $\{\psi_{n_k}\}$ of $\{\psi_n\}$ that converges pointwise to $\psi_0 \in [H_0]^*$. According to the Riesz-Fréchet Representation Theorem, there is a vector $h_0 \in H_0$ for which $\psi_0 = T(h_0)$. Thus

$$\lim_{k\to\infty} \langle h_{n_k},\, h\rangle = \langle h_0,\, h\rangle \text{ for all } h \in H_0.$$

Let P be the orthogonal projection on H onto H_0. For each index k, since $(\mathrm{Id} - P)[H] = P(H)^\perp$,

$$\langle h_{n_k},\, (\mathrm{Id} - P)[h]\rangle = \langle h_0,\, (\mathrm{Id} - P)[h]\rangle = 0 \text{ for all } h \in H.$$

Therefore,

$$\lim_{k\to\infty} \langle h_{n_k},\, h\rangle = \langle h_0,\, h\rangle \text{ for all } h \in H,$$

and so $\{h_{n_k}\} \to h$ weakly to h_0. □

We gather in the following proposition some properties regarding weakly convergent sequences which we established earlier for general Banach spaces.

Proposition 8 *Let $\{u_n\} \to u$ weakly in the Hilbert space H. Then $\{u_n\}$ is bounded and*

$$||u|| \leq \liminf ||u_n||.$$

Moreover, if $\{v_n\} \to v$ strongly in H, then

$$\lim_{n\to\infty} \langle u_n, v_n \rangle = \langle u, v \rangle. \tag{7}$$

In Chapter 8, we proved the following theorem for the Hilbert space L^2.

The Banach-Saks Theorem *Let $\{u_n\} \to u$ weakly in the Hilbert space H. Then there is a subsequence $\{u_{n_k}\}$ of $\{u_n\}$ for which*

$$\lim_{k\to\infty} \frac{u_{n_1} + \cdots + u_{n_k}}{k} = u \textit{ strongly in } H. \tag{8}$$

Proof Replacing each u_n with $u_n - u$ we may suppose that $u = 0$. Since a weakly convergent sequence is bounded, we may choose $M > 0$ such that

$$\|u_n\|^2 \leq M \text{ for all } n.$$

We will inductively choose a subsequence $\{u_{n_k}\}$ of $\{u_n\}$ with the property that for all k,

$$\|u_{n_1} + \cdots + u_{n_k}\|^2 \leq (2 + M)k. \tag{9}$$

For such a sequence,

$$\left\| \frac{u_{n_1} + \cdots + u_{n_k}}{k} \right\|^2 \leq \frac{(2+M)}{k} \text{ for all } k \tag{10}$$

and the proof is complete, provided we choose a sequence for which (9) holds.

Define $n_1 = 1$. Since $\{u_n\} \to u$ weakly and u_{n_1} belongs to H, we can choose an index n_2 such that $|\langle u_{n_1}, u_{n_2} \rangle| < 1$. Suppose we have chosen natural numbers $n_1 < n_2 < \cdots < n_k$ such that

$$\|u_{n_1} + \cdots + u_{n_j}\|^2 \leq (2 + M)j \text{ for } j = 1, \ldots, k.$$

Since $\{u_n\} \to u$ weakly and $u_{n_1} + \cdots + u_{n_k}$ belongs to H, we may choose $n_{k+1} > n_k$ such that

$$|\langle u_{n_1} + \cdots + u_{n_k}, u_{n_{k+1}} \rangle| \leq 1.$$

However,

$$\|u_{n_1} + \cdots + u_{n_k} + u_{n_{k+1}}\|^2 = \|u_{n_1} + \cdots + u_{n_k}\|^2 + 2\langle u_{n_1} + \cdots + u_{n_k}, u_{n_{k+1}} \rangle + \|u_{n_{k+1}}\|^2$$

Therefore,

$$\|u_{n_1} + \cdots + u_{n_{k+1}}\|^2 \leq (2 + M)k + 2 + M = (2 + M)(k + 1).$$

We have chosen a subsequence so that (9) holds. $\square$

Corollary 9 *Every strongly closed, bounded, convex subset K of a Hilbert space H is weakly sequentially compact.*

Proof This is an immediate consequence of Theorem 7 and the Banach-Saks Theorem. □

Theorem 10 *Every Hilbert space H is reflexive.*

Proof To establish reflexivity it is necessary to show that the natural embedding $J\colon H \to [H^*]^*$ is onto. Let $\Psi\colon H^* \to \mathbf{R}$ be a bounded linear functional. Let $T\colon H \to [H]^*$ be the isomorphism described by the Riesz-Fréchet Representation Theorem. Then $\Psi \circ T\colon H \to \mathbf{R}$, being the composition of bounded linear operators, is bounded. According to the Riesz-Fréchet Representation Theorem, there is a vector $h_0 \in H$ for which $\Psi \circ T = T(h_0)$. Therefore,

$$\Psi(T(h)) = T(h_0)[h] = T(h)[h_0] = J(h_0)[T(h)] \text{ for all } h \in H.$$

Since $T(H) = H^*$, $\Psi = J(h_0)$. Thus H is reflexive. □

According to Theorem 17 in Chapter 18, a bounded sequence in a reflexive Banach space has a weakly convergent subsequence; according to Theorem 28 of Chapter 18, every strongly closed, bounded, convex subset of a reflexive Banach space is weakly closed. Consequently, both Theorem 8 and Corollary 10 are a consequence of reflexivity, which we have just established for a Hilbert space. Here, we have provided independent proofs of these two results because of their simplicity in the Hilbert space setting.

An operator $T \in \mathcal{L}(H)$ is said to be **positive-definite** provided that there is a $c > 0$ for which

$$\langle T(u),\, u\rangle \geq c \cdot \|u\|^2 \text{ for all } u \in H.$$

Lemma 11 *If the operator $T \in \mathcal{L}(H)$ is positive-definite, then it is invertible.*

Proof Clearly, T is one-to-one. By linearity, the Cauchy-Schwarz Inequality, and the definition of $\|T\|$, if $c_0 = c/\|T\|$,

$$\|T(u) - T(v)\| \geq c_0 \cdot \|u - v\| \text{ for all } u, v \in H.$$

It follows that if $\{T(u_n)\}$ is Cauchy, then so is $\{u_n\}$. Consequently, in view of the completeness of H, $\operatorname{Im} T$ is closed. On the other hand, $[\operatorname{Im} T]^\perp = \{0\}$, since if $v \in [\operatorname{Im} T]^\perp$, then $\langle T(v),\, v\rangle = 0$, so $v = 0$. According to Theorem 3, $\operatorname{Im} T = H$. Therefore, T is invertible. □

The following Lax-Milgram Lemma has many important applications in the study of partial differential equations. It is a generalization of the representation theorem for the dual of a Hilbert space, in which, in contrast to the inner-product, the bilinear form is not assumed to be symmetric.

Theorem 12 (The Lax-Milgram Lemma) *Let H be a Hilbert space and assume that the function $B\colon H \times H \to \mathbf{R}$ has the following three properties:*

(i) For each $u \in H$, the following two functionals are linear on H;

$$v \mapsto B(u,\, v) \text{ and } v \mapsto B(v,\, u).$$

(ii) There is a $c_1 > 0$ *for which*

$$|B(u,\, v)| \leq c_1 \cdot \|u\| \cdot \|v\| \text{ for all } u, v \in H.$$

(iii) There is a $c_2 > 0$ *for which*

$$B(h,\, h) \geq c_2 \cdot \|h\|^2 \text{ for all } h \in H.$$

Then for each $\psi \in H^*$*, there is a unique* $h \in H$ *for which*

$$\psi(u) = B(h,\, u) \text{ for all } u \in H.$$

Proof Let $T\colon H \to H^*$ be the isomorphism defined by the Riesz-Fréchet Representation Theorem, that is, for each $h \in H$,

$$T(h)[u] = \langle h,\, u\rangle \text{ for all } u \in H. \tag{11}$$

For each $h \in H$, define the functional $S(h)\colon H \to \mathbf{R}$ by

$$S(h)[u] = B(h,\, u) \text{ for all } u \in H. \tag{12}$$

It follows from assumptions (i) and (ii) that each $S(h)$ is a bounded linear functional on H and that the operator $S\colon H \to H^*$ is linear and continuous. Since T is an isomorphism of H onto H^*, to show that S is an isomorphism of H onto H^* is equivalent to showing that the operator $T^{-1} \circ S \in \mathcal{L}(H)$ is invertible. However, by assumption (iii),

$$\langle (T^{-1} \circ S)(h),\, h\rangle = S(h)[h] = B(h,\, h) \geq c_2 \cdot \|h\|^2 \text{ for all } h \in H.$$

According to the preceding lemma, the operator $T^{-1} \circ S \in \mathcal{L}(H)$ is invertible. □

PROBLEMS

24. Show that neither ℓ^1, ℓ^∞, $L^1[a,\, b]$ nor $L^\infty[a,\, b]$ is Hilbertable.
25. Let H be an inner-product space. Show that since H is a dense subset of a Banach space X whose norm restricts to the norm induced by the inner-product on H, the inner-product on H extends to X and induces the norm on X. Therefore, inner-product spaces have Hilbert space completions.
26. (von Neumann's proof of the Radon-Nikodym Theorem) The proof is based on the Riesz-Fréchet Representation Theorem. Let $(X,\, \mathcal{M},\, \mu)$ be a finite measure space, and $\nu\colon \mathcal{M} \to [0,\, \infty)$ be a finite measure that is absolutely continuous with respect to μ. Fill in the following outline of the proof that there is a non-negative measurable function $h\colon X \to [0,\, \infty)$ for which $\nu(E) = \int_E h\, d\mu$ for all $E \in \mathcal{M}$.
    (i) Define the measure $\lambda = \mu + \nu$ on the σ-algebra $\mathcal{M}$ and the functional ψ on $L^2(X,\, \lambda)$ by
    $$\psi(f) = \int_X f\; d\mu \text{ for all } f \in L^2(X,\, \lambda).$$
    Show that, since $\lambda(X) < \infty$, ψ is a bounded linear functional on $L^2(X,\, \lambda)$.

(ii) By the Riesz-Fréchet Representation Theorem for the Hilbert space $L^2(X, \lambda)$, there is a function $g \in L^2(X, \lambda)$ such that

$$\int_X f \, d\mu = \int_X f \cdot g \, d\lambda \text{ for all } f \in L^2(X, \lambda),$$

and therefore

$$\int_X f \, d\mu = \int_X f \cdot g \, d\mu + \int_X f \cdot g \, d\nu \text{ for all } f \in L^2(X, \lambda) \tag{*}$$

(iii) Conclude that, since $\lambda(X) < \infty$,

$$\mu(E) = \int_E g \, d\mu + \int_E g \, d\nu \text{ for all } E \in \mathcal{M}.$$

Therefore, $\lambda\{x \in X \,|\, g(x) > 1\} = 0$, and, using the absolute continuity of ν with respect to μ, we also have $\lambda\{x \in X \,|\, g(x) = 0\} = 0$. Consequently, we may assume that $0 < g \leq 1$ on X.

(iv) For a natural number n and $E \in \mathcal{M}$, substitute $f = \chi_E \cdot \frac{1}{g+1/n}$ in (*) and, using the Monotone Convergence Theorem, take the limit as $n \to \infty$ to conclude that

$$\nu(E) = \int_E h \, d\mu \text{ for all } E \in \mathcal{M}, \text{ where } h = [1/g - 1],$$

and h is non-negative and measurable.

(v) Show that, by taking a countable measurable partition of X such that μ and ν are finite on each set in the partition, the above special case of the Radon-Nikodym Theorem extends to the case that the measures μ and ν are σ-finite.

20.4 SYMMETRIC OPERATORS

For $T \in \mathcal{L}(H)$ and a fixed vector v in H, the mapping

$$u \mapsto \langle T(u), v \rangle \text{ for } u \in H,$$

belongs to H^* since it is linear and, by the Cauchy-Schwarz Inequality, $|\langle T(u), v \rangle| \leq c \cdot \|u\|$ for all $u \in H$, where $c = \|T\| \cdot \|v\|$. According to the Riesz-Fréchet Representation Theorem, there is a unique vector $h \in H$ such that $\langle T(u), v \rangle = \langle h, u \rangle = \langle u, h \rangle$ for all $u \in H$. We denote this vector h by $T^*(v)$. This defines a mapping $T^*\colon H \to H$ that is determined by the relation

$$\langle T(u), v \rangle = \langle u, T^*(v) \rangle \text{ for all } u, v \in H. \tag{13}$$

We call T^* the **adjoint** of T.

Proposition 13 *If $T \in \mathcal{L}(H)$, then $T^* \in \mathcal{L}(H)$ and $\|T\| = \|T^*\|$.*

Proof Clearly T^* is linear. Let h be a unit vector in H. Then, by the Cauchy-Schwarz Inequality,

$$\|T^*(h)\|^2 = \langle T^*(h), T^*(h) \rangle = \langle T(T^*(h)), h \rangle \leq \|T\| \|T^*(h)\|.$$

Thus T^* belongs to $\mathcal{L}(H)$ and $\|T^*\| \leq \|T\|$. But also observe that

$$\|T(h)\|^2 = \langle T(h),\, T(h)\rangle = \langle T^*(T(h)),\, h\rangle \leq \|T^*\|\|T(h)\|.$$

Therefore, $\|T\| \leq \|T^*\|$. □

We leave it as an exercise to verify the following structural properties of adjoints: for $T, S \in \mathcal{L}(H)$,

$$(T^*)^* = T,\ (T+S)^* = T^* + S^* \text{ and } (T \circ S)^* = S^* \circ T^*. \tag{14}$$

Proposition 14 *If $T \in \mathcal{L}(H)$ has a closed image, then*

$$\operatorname{Im} T \oplus \ker\, T^* = H. \tag{15}$$

Proof Since $\operatorname{Im} T$ is closed, it suffices, by Theorem 2, to show that $\ker\, T^* = [\operatorname{Im} T]^\perp$. But this is an immediate consequence of the relation (13). □

Definition *An operator $T \in \mathcal{L}(H)$ is said to be* **symmetric** *or* **self-adjoint** *provided that $T = T^*$, that is,*

$$\langle T(u),\, v\rangle = \langle u,\, T(v)\rangle \text{ for all } u, v \in H.$$

Example Let $\{\varphi_k\}$ be an orthonormal basis for the separable Hilbert space H and T belong to $\mathcal{L}(H)$. Then, by the continuity of the inner-product, T is symmetric if and only if

$$\langle T(\varphi_i),\, \varphi_j\rangle = \langle T(\varphi_j),\, \varphi_i\rangle \text{ for all } 1 \leq i, j < \infty.$$

In particular, if H is Euclidean space $\mathbf{R}^n$, then T is symmetric if and only if the $n \times n$ matrix that represents T with respect to the Euclidean basis is a symmetric matrix.

A symmetric operator $T \in \mathcal{L}(H)$ is said to be **non-negative**, written $T \geq 0$, provided that $\langle T(h),\, h\rangle \geq 0$ for all $h \in H$. Moreover, for two symmetric operators $A, B \in \mathcal{L}(H)$, we write $A \geq B$ provided that $A - B \geq 0$. The sum of non-negative, symmetric operators is non-negative and symmetric. Moreover,

$$\text{if } T \in \mathcal{L}(H) \text{ is symmetric and non-negative, then so is } S^*TS \text{ for any } S \in \mathcal{L}(H), \tag{16}$$

since for each $h \in H$, $\langle S^*TS(h),\, h\rangle = \langle T(S(h)),\, S(h)\rangle \geq 0$. In Problems 37–43, we explore a few of the many interesting consequences of this order relation among symmetric operators.

The Polarization Identity *For a symmetric operator $T \in \mathcal{L}(H)$,*

$$\langle T(u),\, v\rangle = \frac{1}{4}\left[\langle T(u+v),\, u+v\rangle - \langle T(u-v),\, u-v\rangle\right] \text{ for all } u, v \in H. \tag{17}$$

To verify this identity, simply expand the two inner-products on the right-hand side. It is useful to associate with a symmetric operator $T \in \mathcal{L}(H)$ the **quadratic form** $Q_T\colon H \to \mathbf{R}$ defined by

$$Q_T(u) = \langle T(u),\, u\rangle \text{ for all } u \in H.$$

By the Polarization Identity, T is completely determined by Q_T. In particular, $T = 0$ on H if and only if $Q_T = 0$ on H. In fact, the following much sharper result holds.

Proposition 15 *Let $T \in \mathcal{L}(H)$ be a symmetric operator. Then*

$$\|T\| = \sup_{\|u\|=1} |\langle T(u),\, u\rangle|. \tag{18}$$

Proof Define $\eta = \sup_{\|u\|=1} |\langle T(u),\, u\rangle|$. If $\eta = 0$, it follows from the Polarization Identity that $T = 0$. Assume $\eta > 0$. Observe that, by the Cauchy-Schwarz Inequality, for a unit vector $u \in H$,

$$|\langle T(u),\, u\rangle| \leq \|T(u)\|\|u\| \leq \|T\|.$$

Therefore, $\eta \leq \|T\|$. To prove the inequality in the opposite direction, observe that the two symmetric operators $\eta \cdot \mathrm{Id} - T$ and $\eta \cdot \mathrm{Id} + T$ are non-negative and therefore, by (16), the two operators

$$(\eta \cdot \mathrm{Id} + T)^*(\eta \cdot \mathrm{Id} - T)(\eta \cdot \mathrm{Id} + T) = (\eta \cdot \mathrm{Id} + T)(\eta \cdot \mathrm{Id} - T)(\eta \cdot \mathrm{Id} + T)$$

and

$$(\eta \cdot \mathrm{Id} - T)^*(\eta \cdot \mathrm{Id} + T)(\eta \cdot \mathrm{Id} - T) = (\eta \cdot \mathrm{Id} - T)(\eta \cdot \mathrm{Id} + T)(\eta \cdot \mathrm{Id} - T)$$

also are non-negative, and hence so is their sum $2\eta \cdot \mathrm{Id}(\eta^2 \cdot \mathrm{Id} - T^2)$. Since $2\eta \cdot \mathrm{Id} > 0$, $\eta^2 \cdot \mathrm{Id} - T^2$ is non-negative, that is,

$$\|T(u)\|^2 = \langle T(u),\, T(u)\rangle = \langle T^2(u),\, u\rangle \leq \eta^2 \cdot \mathrm{Id}\langle u,\, u\rangle = \eta^2 \cdot \mathrm{Id}\, \|u\|^2 \text{ for all } u \in H.$$

Hence $\|T\| \leq \eta \cdot \mathrm{Id}$. □

For $T \in \mathcal{L}(H)$, define

$$m(T) = \inf_{h \in H, h \neq 0} \frac{\langle T(h),\, h\rangle}{\langle h,\, h\rangle} \text{ and } M(T) = \sup_{h \in H, h \neq 0} \frac{\langle T(h),\, h\rangle}{\langle h,\, h\rangle},$$

and call $m = m(T)$ and $M = M(T)$ the **spectral bounds** for T. Clearly

$$m\langle h,\, h\rangle \leq \langle T(h),\, h\rangle \leq M\langle h,\, h\rangle \text{ for all } h \text{ in } H.$$

For a non-negative, symmetric operator T, there is the following useful inequality:

$$\|T(h)\|^4 \leq (\|T\|^3\|h\|^2)\langle T(h),\, h\rangle \text{ for all } h \text{ in } H. \tag{19}$$

To see this, for u, v in H, observe that, since $T \geq 0$, $p(t) \equiv \langle T(u + tv),\, u + tv\rangle \geq 0$ for all t. Therefore, the discriminant of the quadratic polynomial p is non-positive, that is,

$$\langle T(u),\, v\rangle^2 \leq \langle T(u),\, u\rangle\langle T(v),\, v\rangle.$$

Substitute $u = h$ and $v = T(h)$ in this inequality and use the Cauchy-Schwarz Inequality to deduce (19).

For $T \in \mathcal{L}(H)$, we define the **spectrum** of T, $\sigma(T)$, by

$$\sigma(T) = \{\lambda \in \mathbf{R} \mid \ \lambda \cdot Id - T \text{ is not invertible}\}.$$

The complement in $\mathbf{R}$ of $\sigma(T)$ is called the **resolvent** of T and denoted by $\rho(T)$.

Theorem 16 (The Spectral Boundary Theorem) *If the operator $T \in \mathcal{L}(H)$ is symmetric and m and M are its spectral bounds, then $\sigma(T) \subseteq [m, M]$ and both m and M belong to $\sigma(T)$.*

Proof According to Lemma 13, a positive-definite operator is invertible, and therefore $\sigma(T) \subseteq [m, M]$. To show that M belongs to $\sigma(T)$, observe that since $M\,\mathrm{Id} - T$ is a nonnegative, symmetric operator, according to (19),

$$\|(M\,\mathrm{Id} - T)(h)\|^4 \leq \|M\,\mathrm{Id} - T\|^3 \|h\|^2 \langle (M\,\mathrm{Id} - T)(h),\, h\rangle \text{ for all } \in H.$$

By the definition of M, there is a sequence $\{h_n\}$ of unit vectors such that the sequence $\{\langle (M\,\mathrm{Id} - T)(h_n),\, h_n\rangle\}$ converges to 0, and so $\{(M - T)(h_n)\}$ converges to 0. Therefore, $M\,\mathrm{Id} - T$ cannot be invertible, since otherwise, by the Open Mapping Theorem, its inverse would be continuous. Hence $M \in \sigma(T)$. Replacing $M\,\mathrm{Id} - T$ by $T - m\,\mathrm{Id}$, the same argument show that $m \in \sigma(T)$. $\square$

PROBLEMS

27. Verify (14).
28. Let $T, S \in \mathcal{L}(H)$ be symmetric. Show that $T = S$ if and only if $Q_T = Q_S$.
29. Show the bounded, symmetric operators are a closed subspace of $\mathcal{L}(H)$. Also show that if T and S are symmetric, then so is the composition $S \circ T$ if and only if T commutes with S with respect to composition, that is, $S \circ T = T \circ S$.
30. (Hellinger-Toplitz) Let H be a Hilbert space and the linear operator $T \colon H \to H$ have the property that $\langle T(u),\, v\rangle = \langle u,\, T(v)\rangle$ for all $u, v \in H$. Show that T belongs to $\mathcal{L}(H)$, that is, T is bounded.
31. Exhibit an operator $T \in \mathcal{L}(\mathbf{R}^2)$ for which $\|T\| > \sup_{\|u\|=1} |\langle T(u),\, u\rangle|$.
32. Let $S, T \in \mathcal{L}(H)$ be symmetric. Assume that $S \geq T$ and $T \geq S$. Prove that $T = S$.
33. Let V be a closed, nontrivial subspace of a Hilbert space H and P the orthogonal projection of H onto V. Show that $P = P^*$, $P \geq 0$, and $\|P\| = 1$.
34. Let $P \in \mathcal{L}(H)$ be a projection. Show that P is the orthogonal projection of H onto $P(H)$ if and only if $P = P^*$.
35. Let $\{\varphi_k\}$ be an orthonormal basis for a Hilbert space H and for each natural number n, define P_n to be the orthogonal projection of H onto the linear span of $\{\varphi_1, \ldots, \varphi_n\}$. Show that P_n is symmetric and
$$0 \leq P_n \leq P_{n+1} \leq \mathrm{Id} \text{ for all } n.$$
Show that $\{P_n\}$ converges pointwise on H to Id but does not converge uniformly on the unit ball.
36. Show that if $T \in \mathcal{L}(H)$ is invertible, so is $T^* \circ T$ and therefore so is T^*.

37. (General Cauchy-Schwarz Inequality) Let $T \in \mathcal{L}(H)$ be symmetric and non-negative. Show that for all $u, v \in H$,
$$|\langle T(u),\, v\rangle|^2 \leq \langle T(u),\, u\rangle \cdot \langle T(v),\, v\rangle.$$

38. Use the preceding problem to show that if $S, T \in \mathcal{L}(H)$ are symmetric and $S \geq T$, then for each $u \in H$,
$$\|S(u) - T(u)\|^4 = \langle (S-T)(u),\, (S-T)(u)\rangle^2 \leq |\langle (S-T)(u),\, u\rangle||\langle (S-T)^2(u),\, (S-T)(u)\rangle|$$
and thereby conclude that
$$\|S(u) - T(u)\|^4 \leq |\langle S(u),\, u\rangle - \langle T(u),\, u\rangle| \cdot \|S - T\|^3 \cdot \|u\|^2.$$

39. (Monotone Convergence Theorem for Symmetric Operators) A sequence $\{T_n\}$ of symmetric operators in $\mathcal{L}(H)$ is said to be monotone increasing provided that $T_{n+1} \geq T_n$ for each n, and said to be bounded above provided that there is a symmetric operator S in $\mathcal{L}(H)$ such that $T_n \leq S$ for all n.
   (i) Use the preceding problem to show that a monotone increasing sequence $\{T_n\}$ of symmetric operators in $\mathcal{L}(H)$ converges pointwise to a symmetric operator in $\mathcal{L}(H)$ if and only if it is bounded above.
   (ii) Show that a monotone increasing sequence $\{T_n\}$ of symmetric operators in $\mathcal{L}(H)$ is bounded above if and only if it is pointwise bounded, that is, for each $h \in H$, the sequence $\{T_n(h)\}$ is bounded.

40. Let $S \in \mathcal{L}(H)$ be a symmetric operator for which $0 \leq S \leq \mathrm{Id}$. Define a sequence $\{T_n\}$ in $\mathcal{L}(H)$ by letting $T_1 = 1/2(\mathrm{Id} - S)$ and if n is a natural number for which $T_n \in \mathcal{L}(H)$ has been defined, defining $T_{n+1} = 1/2(\mathrm{Id} - S + T_n^2)$.
   (i) Show that for each natural number n, T_n and $T_{n+1} - T_n$ are polynomials in $\mathrm{Id} - S$ with non-negative coefficients.
   (ii) Show that $\{T_n\}$ is a monotone increasing sequence of symmetric operators that is bounded above by Id.
   (iii) Use the preceding problem to show that $\{T_n\}$ converges pointwise to a symmetric operator T for which $0 \leq T \leq \mathrm{Id}$ and $T = 1/2(\mathrm{Id} - S + T^2)$.
   (iv) Define $A = (\mathrm{Id} - T)$. Show that $A^2 = S$.

41. (Square Roots of Non-negative Symmetric Operators) Let $T \in \mathcal{L}(H)$ be a non-negative, symmetric operator. A non-negative, symmetric operator $A \in \mathcal{L}(H)$ is called a square root of T provided that $A^2 = T$. Use the inductive construction in the preceding problem to show that T has a square root A which commutes with each operator in $\mathcal{L}(H)$ that commutes with T. Show that the square root is unique: it is denoted by $\sqrt{T}$. Finally, show that T is invertible if and only if $\sqrt{T}$ is invertible.

42. An invertible operator $T \in \mathcal{L}(H)$ is said to be **orthogonal** provided that $T^{-1} = T^*$. Show that an invertible operator is orthogonal if and only if it is an isometry.

43. (Polar Decompositions) Let $T \in \mathcal{L}(H)$ be invertible. Show that there is an orthogonal invertible operator $A \in \mathcal{L}(H)$ and a non-negative, symmetric invertible operator $B \in \mathcal{L}(H)$ such that $T = B \circ A$. (Hint: Show that TT^* is invertible and symmetric and let $B = \sqrt{T \circ T^*}$.)

20.5 COMPACT OPERATORS

Definition *An operator $T \in \mathcal{L}(H)$ is said to be* **compact** *provided that $T(B)$ has compact closure, with respect to the strong topology, where B is the closed unit ball in H.*

Any operator $T \in \mathcal{L}(H)$ maps bounded sets to bounded sets. An operator $T \in \mathcal{L}(H)$ is said to be of **finite rank** provided that its image is finite dimensional. Since a bounded subset of a finite dimensional space has compact closure, every operator of finite rank is compact. In particular, if H is finite dimensional, then every operator in $\mathcal{L}(H)$ is compact. On the other hand, according to Riesz's Theorem, the identity operator $\mathrm{Id}\colon H \to H$ fails to be compact if H is infinite dimensional. For the same reason, an invertible operator in $\mathcal{L}(H)$ fails to be compact if H is infinite dimensional.

In any metric space, compactness of a set is the same as sequential compactness. Furthermore, since a metric space is compact if and only if it is complete and totally bounded, a subset of a complete metric space has compact closure if and only if it is totally bounded. We therefore have the following useful characterizations of compactness for a bounded linear operator.

Proposition 17 *For $T \in \mathcal{L}(H)$, the following three properties are equivalent:*

(i) T is compact;

(ii) $T(B)$ is totally bounded, where B is the closed unit ball in H;

(iii) If $\{h_n\}$ is a bounded sequence in H, then $\{T(h_n)\}$ has a strongly convergent subsequence.

Example Let $\{\varphi_k\}$ be an orthonormal basis for the separable Hilbert space H and $\{\lambda_k\}$ a sequence of real numbers that converges to 0. Define

$$T(h) = \sum_{k=1}^{\infty} \lambda_k \langle h,\, \varphi_k\rangle \varphi_k \text{ for } h \in H.$$

It follows from Bessel's Inequality and the boundedness of $\{\lambda_k\}$ that T belongs to $\mathcal{L}(H)$ and we claim that T is compact. According to the preceding proposition, to show that T is compact it suffices to show that $T(B)$ is totally bounded. Let $\epsilon > 0$. Choose N such that $|\lambda_k| < \epsilon/2$ for $k \geq N$. Define $T_N \in \mathcal{L}(H)$ by

$$T_N(h) = \sum_{k=1}^{N} \lambda_k \langle h,\, \varphi_k\rangle \varphi_k \text{ for } h \in H.$$

Then, by Bessel's Inequality, $\|T(h) - T_N(h)\| < \epsilon/2\,\|h\|$ for $h \in H$. But $T_N(B)$ is a bounded subset of a finite dimensional space, so it is totally bounded. Let $\epsilon > 0$. There is a finite $\epsilon/2$-net for $T_N(B)$ and by doubling the radius of each of the balls in this net we get a finite ϵ-net for $T(B)$.

Example For a closed, bounded interval $I = [a,\, b]$ consider a continuous function $k\colon I \times I \to \mathbf{R}$. Consider $K\colon L^2[a,\, b] \to L^2[a,\, b]$ defined by

$$K(u)(s) = \int_a^b k(s,t)u(t)\,dt \text{ for } s \in I.$$

Then $K\colon L^2[a, b] \to L^2[a, b]$ is compact. This follows from the Arzelà-Ascoli Theorem and the uniform continuity of k.

A linear operator $T\colon H \to H$ belongs to $\mathcal{L}(H)$ if and only if it maps weakly convergent sequences to weakly convergent sequences (see Problem 47).

Proposition 18 *An operator T in $\mathcal{L}(H)$ is compact if and only if it maps weakly convergent sequences to strongly convergent sequences, that is,*

$$\text{if } \{h_n\} \to h \text{ weakly, then } \{T(h_n)\} \to T(h) \text{ strongly.}$$

Proof According to the preceding proposition, an operator is compact if and only if it maps bounded sequences to sequences that have a strongly convergent subsequence. First assume that T is compact. Observe that for any operator $T \in \mathcal{L}(H)$, if $\{u_n\} \to u$ weakly, then $\{T(u_n)\} \to T(u)$ weakly, since for each $v \in H$,

$$\lim_{k\to\infty} \langle T(u_k),\, v\rangle = \lim_{k\to\infty} \langle u_k,\, T^*(v)\rangle = \langle u,\, T^*(v)\rangle = \langle T(u),\, v\rangle.$$

Let $\{u_k\} \to u$ weakly. By the compactness of T, every subsequence of $\{T(h_n)\}$ has a further subsequence that converges strongly and, by the preceding observation, its strong limit must be $T(h)$. Therefore, the entire sequence $\{T(h_n)\}$ converges strongly to $T(h)$. To prove the converse, assume that T maps weakly convergent sequences to strongly convergent subsequences. Let $\{h_n\}$ be a bounded sequence. According to Theorem 7, $\{h_n\}$ has a weakly convergent subsequence. The image of this weakly convergent subsequence converges strongly. □

Schauder's Theorem *A compact linear operator on a Hilbert space has a compact adjoint.*

Proof Let $K \in \mathcal{L}(H)$ be compact. According to the preceding proposition, it suffices to show that K^* maps weakly convergent sequences to strongly convergent sequences. Let $\{h_n\} \to h$ weakly in H. For each n,

$$\|K^*(h_n) - K^*(h)\|^2 = \langle KK^*(h_n) - KK^*(h),\, h_n - h\rangle. \tag{20}$$

Since K^* is continuous, $\{K^*(h_n)\}$ converges weakly to $K^*(h)$. According to the preceding proposition, $\{KK^*(h_n)\} \to KK^*(h)$ strongly in H. Therefore, by Proposition 8,

$$\lim_{k\to\infty} \langle KK^*(h_n) - KK^*(h),\, h_n - h\rangle = 0.$$

It follows from (20) that $\{K^*(h_n)\}$ converges strongly to $K^*(h)$. □

PROBLEMS

44. Show that if H is infinite dimensional and $T \in \mathcal{L}(H)$ is invertible, then T is not compact.
45. Prove Proposition 17.
46. Let $\mathcal{K}(H)$ denote the set of compact operators in $\mathcal{L}(H)$. Show that $\mathcal{K}(H)$ is a linear subspace of $\mathcal{L}(H)$. Moreover, show that for $K \in \mathcal{K}(H)$ and $T \in \mathcal{L}(H)$, both $K \circ T$ and $T \circ K$ belong to $\mathcal{K}(H)$.

47. Show that a linear operator $T\colon H \to H$ is continuous if and only if it maps weakly convergent sequences to weakly convergent sequences.
48. Show that $K \in \mathcal{L}(H)$ is compact if and only if whenever if $\{u_n\} \to u$ weakly and $\{v_n\} \to v$ weakly, then $\langle K(u_n), v_n\rangle \to \langle K(u), v\rangle$.
49. Let $\{P_n\}$ be a sequence of orthogonal projections in $\mathcal{L}(H)$ with the property that for natural numbers n and m, $P_n(H)$ and $P_m(H)$ are orthogonal finite dimensional subspaces of H. Let $\{\lambda_n\}$ be a bounded sequence of real numbers. Show that

$$K = \sum_{n=1}^{\infty} \lambda_n \cdot P_n$$

is a properly defined symmetric operator in $\mathcal{L}(H)$ that is compact if and only if $\{\lambda_n\}$ converges to 0.
50. For X a Banach space, define an operator $T \in \mathcal{L}(X)$ be compact provided that $T(B)$ has compact closure. Show that Proposition 17 holds for a general Banach space and Proposition 18 holds for a reflexive Banach space.

20.6 THE HILBERT-SCHMIDT THEOREM

A non-zero vector $u \in H$ is said to be an **eigenvector** of the operator $T \in \mathcal{L}(H)$ provided that there is some $\lambda \in \mathbf{R}$ for which $T(u) = \lambda u$. We call λ the **eigenvalue** of T associated with the eigenvector u. One of the centerpieces of linear algebra is the following assertion: If H is a finite dimensional Hilbert space and $T \in \mathcal{L}(H)$ is symmetric, then there is an orthonormal basis for H consisting of eigenvectors of T, that is, if H has dimension n, there is an orthonormal basis $\{\varphi_1, \ldots, \varphi_n\}$ for H and numbers $\{\lambda_1, \ldots, \lambda_n\}$ such that $T(\varphi_k) = \lambda_k \varphi_k$ for $1 \leq k \leq n$. Thus

$$T(h) = \sum_{k=1}^{n} \lambda_k \langle h, \varphi_k\rangle \varphi_k \text{ for all } h \in H. \tag{21}$$

Of course, in the absence of symmetry, a bounded linear operator, even on a finite dimensional space, may fail to have any eigenvectors. It should be emphasized that we are considering real-linear operators on real vector spaces. Spectral theory for complex-linear operators on complex linear spaces is rather different. As the following example shows, even a symmetric operator on an infinite dimensional Hilbert space may fail to have any eigenvectors.

Example Define $T \in \mathcal{L}(L^2[a, b])$ by $[T(f)](x) = xf(x)$ for $f \in L^2[a, b]$. For $u, v \in L^2[a, b]$,

$$\langle T(u), v\rangle = \int_a^b xu(x)v(x)\,dx = \langle u, T(v)\rangle.$$

Then T is symmetric and one easily checks that it has no eigenvectors.

It is useful to define the **Rayleigh quotient** for T, $R_T\colon H \sim \{0\} \to \mathbf{R}$, by

$$R_T(h) = \frac{\langle T(h), h\rangle}{\langle h, h\rangle} \text{ for all } h \in H \sim \{0\}.$$

Observe that a maximizer h_* for the quadratic form Q_T on the unit sphere $S = \{h \in H \mid \|h\| = 1\}$ is a maximizer for the Rayleigh quotient R_T on $H \sim \{0\}$.

The Hilbert-Schmidt Lemma *Let $T \in \mathcal{L}(H)$ be compact and symmetric. Then T has an eigenvalue λ for which*

$$|\lambda| = \|T\| = \sup_{\|h\|=1} |\langle T(h),\, h\rangle|. \tag{22}$$

Proof If $T = 0$ on H, then every non-zero vector in H is an eigenvector of T with corresponding eigenvalue $\lambda = 0$. So consider the case $T \neq 0$. According to Proposition 15,

$$\|T\| = \sup_{\|h\|=1} |\langle T(h),\, h\rangle|.$$

By possibly replacing T by $-T$ we may suppose that $\|T\| = \sup_{\|h\|=1}\langle T(h),\, h\rangle$. Denote $\sup_{\|h\|=1}\langle T(h),\, h\rangle$ by η. Let $S = \{h \in H \mid \|h\| = 1\}$ be the unit sphere in H.

Let $\{h_k\}$ be a sequence of unit vectors for which $\lim_{k\to\infty}\langle T(h_k),\, h_k\rangle = \eta$. By Theorem 7, by possibly passing to a subsequence, we may suppose that $\{h_k\}$ converges weakly to h_*. We have $\|h_*\| \le \liminf \|h_n\| = 1$. According to Proposition 18, since T is compact, $\{T(h_n)\}$ converges strongly to $T(h_*)$. Therefore, by Proposition 8,

$$\lim_{k\to\infty} \langle T(h_k),\, h_k\rangle = \langle T(h_*),\, h_*\rangle.$$

Thus $\eta = \langle T(h_*),\, h_*\rangle$. Now $h_* \neq 0$ since $\eta \neq 0$. Moreover, h_* must be a unit vector. Indeed, otherwise $0 < \|h_*\| < 1$, in which case the quadratic form Q_T takes a value greater than η at $h_*/\|h_*\| \in S$, contradicting the choice of η as being an upper bound for Q_T on S. Thus $h_* \in S$ and $Q_T(h) \le Q_T(h_*)$ for all $h \in S$. Therefore, for the Rayleigh Quotient for T, R_T, we have

$$R_T(h) \le R_T(h_*) \text{ for all } h \in H \sim \{0\}.$$

Let h_0 be any vector in H. The function $f\colon \mathbf{R} \to \mathbf{R}$, defined by $f(t) = R_T(h_* + th_0)$ for $t \in R$, has a maximum value at $t = 0$ and therefore $f'(0) = 0$. Since $h_* \neq 0$, f is differentiable at $t = 0$ and a direct calculation gives

$$0 = f'(0) = \frac{\langle T(h_0),\, h_*\rangle + \langle T(h_*),\, h_0\rangle}{\|h_*\|^2} - \langle T(h_*),\, h_*\rangle \frac{\langle h_*,\, h_0\rangle + \langle h_0,\, h_*\rangle}{\|h_*\|^4}.$$

But T is symmetric, h_* is a unit vector, and $\eta = \langle T(h_*),\, h_*\rangle$ so that

$$\langle T(h_*) - \eta h_*,\, h_0\rangle = 0.$$

Since this holds for all $h_0 \in H$, $T(h_*) = \eta h_*$. □

A general strategy in the study of a linear operator $T \in \mathcal{L}(H)$ is to express H as a direct sum $H_1 \oplus H_2$ for which $T(H_1) \subseteq H_1$ and $T(H_2) \subseteq H_2$, in which case it is said that the decomposition $H = H_1 \oplus H_2$ **reduces** the operator T. In general, if $T(H_1) \subseteq H_1$, it cannot be concluded that $T(H_2) \subseteq H_2$. However, for symmetric operators on H and an orthogonal direct sum decomposition of H, there is the following simple but very useful result.

Proposition 19 *Assume that the operator $T \in \mathcal{L}(H)$ is symmetric and V is a subspace of H for which $T(V) \subseteq V$. Then $T(V^\perp) \subseteq V^\perp$.*

Proof Let u belong to $V^\perp$. Then for any $v \in V$, $\langle T(u), v\rangle = \langle u, T(v)\rangle$ and $\langle u, T(v)\rangle = 0$ since $T(V) \subseteq V$ and $u \in V^\perp$. Therefore, $T(u) \in V^\perp$. □

The Hilbert-Schmidt Theorem Let $K \in \mathcal{L}(H)$ *be a compact, symmetric operator on a Hilbert space H that is not of finite rank. Then there is an orthonormal basis $\{\psi_k\}$ for* $[\ker K]^\perp$ *together with a sequence of non-zero numbers $\{\lambda_k\}$ such that* $\lim_{k\to\infty} \lambda_k = 0$ *and $K(\psi_k) = \lambda_k \psi_k$ for each k. Therefore,*

$$K(h) = \sum_{k=1}^{\infty} \lambda_k \langle h, \psi_k\rangle \psi_k \text{ for all } h \in H. \tag{23}$$

Proof Let S be the unit sphere in H. According to the Hilbert-Schmidt Lemma, we may choose a vector $\psi_1 \in S$ and $\mu_1 \in \mathbf{R}$ for which

$$K(\psi_1) = \mu_1 \psi_1 \text{ and } |\mu_1| = \sup_{h\in S} |\langle K(h), h\rangle|.$$

Since $K \neq 0$, it follows from Proposition 15 that $\mu_1 \neq 0$. Define $H_1 = [\text{span}\,\{\psi_1\}]^\perp$. Since $K(\text{span}\,\{\psi_1\}) \subseteq \text{span}\,\{\psi_1\}$, it follows from Proposition 19 that $K(H_1) \subseteq H_1$. Thus, if we define K_1 to be the restriction of K to H_1, then $K_1 \in \mathcal{L}(H_1)$ is compact and symmetric. We again apply the Hilbert-Schmidt Lemma to choose a vector $\psi_2 \in S \cap H_1$ and $\mu_2 \in \mathbf{R}$ for which

$$K(\psi_2) = \mu_2 \psi_2 \text{ and } |\mu_2| = \sup\,\{|\langle K(h), h\rangle| \mid h \in S \cap H_1\}.$$

Observe that $|\mu_2| \leq |\mu_1|$. Moreover, since K does not have finite rank, we again use Proposition 15 to conclude that $\mu_2 \neq 0$. We argue inductively to choose an orthogonal sequence of unit vectors in H, $\{\psi_k\}$, and a sequence of non-zero real numbers $\{\mu_k\}$ such that for each index k,

$$K(\psi_k) = \mu_k \psi_k \text{ and } |\mu_k| = \sup\left\{|\langle K(h), h\rangle| \;\middle|\; h \in S \cap [\text{span}\,\{\psi_1, \ldots, \psi_{k-1}\}]^\perp\right\}. \tag{24}$$

Observe that $\{|\mu_k|\}$ is decreasing. We claim that $\{\mu_k\} \to 0$. Indeed, otherwise, since this sequence is decreasing, there is some $\epsilon > 0$ such that $|\mu_k| \geq \epsilon$ for all k. Therefore, for natural numbers m and n, since ψ_n is orthogonal to ψ_m,

$$\|K(\psi_n) - K(\psi_m)\|^2 = \mu_n^2 \|\psi_n\|^2 + \mu_m^2 \|\psi_m\|^2 \geq 2\epsilon^2.$$

Thus $\{K(\psi_k)\}$ has no strongly convergent subsequence and this contradicts the compactness of the operator K. Therefore, $\{\mu_k\} \to 0$. Define H_0 to be the closed linear span of $\{\psi_k\}_{k=1}^\infty$. Then, by Proposition 5, $\{\psi_k\}_{k=1}^\infty$ is an orthonormal basis for H_0. Since $K(H_0) \subseteq H_0$, it follows from Proposition 19 that $K(H_0^\perp) \subseteq H_0^\perp$. But observe that if $h \in H_0^\perp$ is a unit vector, then, for each k, $h \in S \cap [\text{span}\,\{\psi_1, \ldots, \psi_{k-1}\}]^\perp$ and therefore $|\langle K(h), h\rangle| \leq |\mu_k|$. Since $\{\mu_k\} \to 0$, $\langle K(h), h\rangle = 0$. Thus $Q_T = 0$ on $H_0^\perp$ and hence, by the polarization identity, $\ker K = H_0^\perp$. Thus $H_0 = [\ker K]^\perp$. □

In case a symmetric operator $T \in \mathcal{L}(H)$ has finite rank, define H_0 to be the image of T. Then $\ker T = H_0^\perp$. The above argument obtains a finite orthonormal basis for H_0 consisting of eigenvectors of T, thereby establishing a basic result of linear algebra that was mentioned at the beginning of this section.

PROBLEMS

51. Let $T \in \mathcal{L}(H)$ be compact and symmetric. Define

$$\alpha = \inf_{\|h\|=1} \langle T(h),\, h\rangle \text{ and } \beta = \sup_{\|h\|=1} \langle T(h),\, h\rangle.$$

Show that if $\alpha < 0$, then α is an eigenvalue of T and if $\beta > 0$, then β is an eigenvalue of T. Exhibit an example where $\alpha = 0$ and yet α is not an eigenvalue of T, that is, T is one-to-one.

52. Let $K \in \mathcal{L}(H)$ be compact and symmetric. Suppose

$$\sup_{\|h\|=1} \langle K(h),\, h\rangle = \beta > 0.$$

Let $\{h_n\}$ be a sequence of unit vectors for which $\lim_{n\to\infty}\langle K(h_n),\, h_n\rangle = \beta$. Show that a subsequence of $\{h_n\}$ converges strongly to an eigenvector of T with corresponding eigenvalue β.

20.7 THE RIESZ-SCHAUDER THEOREM: CHARACTERIZATION OF FREDHOLM OPERATORS

A subspace X_0 of the linear space X is said to be of **finite codimension** in X provided that X_0 has a finite dimensional linear complement in X, that is, there is a finite dimensional subspace X_1 of X for which $X = X_0 \oplus X_1$. The codimension of X_0, denoted by codim X_0, is properly defined to be the dimension of a linear complement of X_0; all linear complements have the same dimension (see Problem 66). A cornerstone of linear algebra is the assertion that if X is a finite dimensional linear space and $T\colon X \to X$ is linear, then the sum of the rank of T and the nullity of T equals the dimension of X, that is, if $\dim X = n$,

$$\dim \operatorname{Im} T + \dim \ker T = n,$$

and therefore, since $\operatorname{codim} \operatorname{Im} T = n - \dim \operatorname{Im} T$,

$$\dim \ker T = \operatorname{codim} \operatorname{Im} T. \tag{25}$$

Our principal goal in this section is to prove that if H is a Hilbert space and the operator $T \in \mathcal{L}(H)$ is a compact perturbation of the identity operator, then T has a finite dimensional kernel and a finite codimensional image for which (25) holds.

Proposition 20 *If $K \in \mathcal{L}(H)$ is a compact operator, then* $\mathrm{Id} + K$ *has a finite dimensional kernel and a closed image.*

Proof We argue by contradiction. Otherwise, $\ker(\mathrm{Id}+K)$ is infinite dimensional. By appealing to Riesz's Lemma of Chapter 17, and then following the proof of the succeeding Riesz's Theorem, there is a sequence of unit vectors $\{u_k\}$ contained in $\ker(\mathrm{Id}+K)$ for which $\|u_n - u_m\| \geq 1/2$. if $m \neq n$. Now observe that each $\|K(u_n) - K(u_m)\| = \|u_n - u_m\|$, so that the sequence $\{K(u_n)\}$ has no convergent subsequence. This contradicts the compactness of the operator K. Thus $\dim[\ker(\mathrm{Id}+K)] < \infty$. Let $H_0 = [\ker(\mathrm{Id}+K)]^{\perp}$. We claim that there is a $c > 0$ for which

$$\|u + K(u)\| \geq c\|u\| \text{ for all } u \in H_0. \tag{26}$$

Indeed, if there is no such c, then we can choose a sequence $\{u_n\}$ of unit vectors in H_0 such that $\{u_n + K(u_n)\} \to 0$ strongly in H. Since K is compact, by passing to a subsequence if necessary, we may suppose that $\{K(u_n)\} \to h_0$ strongly. Therefore $\{u_n\} \to -h_0$ strongly. By the continuity of K, $h_0 + K(h_0) = 0$. Thus, since H_0 is closed, h_0 is a unit vector that belongs to both $[\ker(\mathrm{Id} + K)]^{\perp}$ and $\ker(\mathrm{Id} + K)$. This contradiction confirms the existence of a $c > 0$ for which (26) holds. Using Theorem 8 of Chapter 17, we conclude from (26) and the completeness of H_0 that $(\mathrm{Id} + K)(H_0)$ is closed. Since $(\mathrm{Id} + K)(H_0) = (\mathrm{Id} + K)(H)$, $\mathrm{Im}(\mathrm{Id} + K)$ is closed. □

Definition *Let $\{\varphi_k\}$ be an orthonormal basis for the separable Hilbert space H. For each n, define $P_n \in \mathcal{L}(H)$ by*

$$P_n(h) = \sum_{k=1}^{n} \langle \varphi_k, h \rangle \varphi_k \text{ for all } h \in H.$$

We call $\{P_n\}$ the **orthogonal projection sequence** *induced by $\{\varphi_k\}$.*

For an orthogonal projection sequence $\{P_n\}$ induced by an orthonormal basis $\{\varphi_n\}$, each P_n is the orthogonal projection of H into $\mathrm{span}\{\varphi_1, \ldots, \varphi_n\}$ and therefore $\|P_n\| = 1$. Moreover, by the very definition of an orthonormal basis, $\{P_n\} \to \mathrm{Id}$ pointwise on H. Therefore, for any $T \in \mathcal{L}(H)$, $\{P_n \circ T\}$ is a sequence of operators of finite rank that converges pointwise to T on H.

Proposition 21 *Let $\{P_n\}$ be the orthogonal projection sequence induced by an orthonormal basis $\{\varphi_n\}$ for the separable Hilbert space H. Then an operator $T \in \mathcal{L}(H)$ is compact if and only if $\{P_n \circ T\} \to T$ in $\mathcal{L}(H)$.*

Proof First assume that $\{P_n \circ T\} \to T$ in $\mathcal{L}(H)$. For each natural number n, $P_n \circ T$ has a finite dimensional range and therefore $(P_n \circ T)(B)$ is totally bounded, where B is the unit ball in H. Since $\{P_n \circ T\} \to T \in \mathcal{L}(H)$, $\{P_n \circ T\}$ converges uniformly on B to T. Therefore, $T(B)$ also is totally bounded. It follows from Proposition 17 that the operator T is compact. To prove the converse, assume that T is compact. Then the set $\overline{T(B)}$ is compact with respect to the strong topology. For each natural number n, define $\psi_n \colon \overline{T(B)} \to \mathbf{R}$ by

$$\psi_n(h) = \|P_n(h) - h\| \text{ for all } h \in \overline{T(B)}.$$

Since each P_n has norm 1, the sequence of real-valued functions $\{\psi_n \colon \overline{T(B)} \to \mathbf{R}\}$ is equicontinuous, bounded, and converges pointwise to 0 on the compact set $\overline{T(B)}$. Consequently, by the Arzelà-Ascoli Theorem, $\{\psi_n \colon \overline{T(B)} \to \mathbf{R}\}$ converges uniformly to 0. This means precisely that $\{P_n \circ T\} \to T$ in $\mathcal{L}(H)$. □

Proposition 22 *Let $K \in \mathcal{L}(H)$ be a compact operator. If* $\mathrm{Id} + K$ *is one-to-one, then it is onto.*

Proof We leave it as an exercise (Problem 53) to show that there is a closed, separable subspace H_0 of H for which $K(H_0) \subseteq H_0$ and $K = 0$ on $H_0^{\perp}$. Therefore, by replacing H by H_0 we may suppose H is separable. Since $\mathrm{Id} + K$ is one-to-one, we see that $\ker(\mathrm{Id} + K)^{\perp} = H$.

Using this fact, we argue as we did in the proof of Proposition 19 to show that there is a $c > 0$ for which

$$\|h + K(h)\| \geq c\|h\| \text{ for all } h \in H. \tag{27}$$

According to Theorem 6, H has an orthonormal basis $\{\varphi_n\}$. Let $\{P_n\}$ be the orthogonal projection sequence induced by $\{\varphi_n\}$. For each natural number n, let H_n be the linear span of $\{\varphi_1, \ldots, \varphi_n\}$. Since the operator K is compact, according to the preceding proposition, $\{P_n \circ K\} \to K$ in $\mathcal{L}(H)$. Choose a natural number N for which $\|P_n \circ K - K\| < c/2$ for $n \geq N$. We conclude from (27) that

$$\|u + P_n \circ K(u)\| \geq c/2\|u\| \text{ for all } u \in H \text{ and all } n \geq N. \tag{28}$$

To show that $(\mathrm{Id} + K)(H) = H$, let h_* belong to H. Let $n \geq N$. It follows from (28) that the restriction to H_n of $\mathrm{Id} + P_n \circ K$ is a one-to-one linear operator that maps the finite dimensional space H_n into itself. A one-to-one linear operator on a finite dimensional space is onto. Therefore this restriction maps H_n onto H_n. Thus, there is a vector $u_n \in H_n$ for which

$$u_n + (P_n \circ K)(u_n) = P_n(h_*). \tag{29}$$

Take the inner-product of each side of this equality with $v \in H$ and use the symmetry of the projection P_n to conclude that

$$\langle u_n + K(u_n),\, P_n(v)\rangle = \langle h_*,\, P_n(v)\rangle \text{ for all } n \geq N, v \in H. \tag{30}$$

It follows from (29) and the estimate (28) that

$$\|h_*\| \geq \|P_n(h_*)\| = \|u_n + (P_n \circ K)u_n\| \geq c/2\|u_n\| \text{ for all } n \geq N.$$

Therefore, the sequence $\{u_n\}$ is bounded. According to Theorem 7, there is a subsequence $\{u_{n_k}\}$ that converges weakly to $u \in H$. Therefore, $\{u_{n_k} + K(u_{n_k})\}$ converges weakly to $u + K(u)$. Take the limit as $k \to \infty$ in (30) with $n = n_k$ to conclude, by Proposition 8, that

$$\langle u + K(u),\, v\rangle = \langle h_*,\, v\rangle \text{ for all } v \in H.$$

Therefore, $u + K(u) = h_*$. Thus $(\mathrm{Id} + K)(H) = H$. □

The Riesz-Schauder Theorem *Let $K \in \mathcal{L}(H)$ be a compact operator. Then* $\mathrm{Im}(\mathrm{Id} + K)$ *is closed and*

$$\dim \ker(\mathrm{Id} + K) = \operatorname{codim} \mathrm{Im}\ (\mathrm{Id} + K) < \infty. \tag{31}$$

In particular, $\mathrm{Id} + K$ *is one-to-one if and only if it is onto.*

Proof According to Proposition 20, a compact perturbation of the identity has finite dimensional kernel and a closed image. Therefore, by Proposition 15, codim $\mathrm{Im}(Id + K) = \dim(Id + K^*)$. We will show that

$$\dim \ker(\mathrm{Id} + K) \geq \dim \ker(\mathrm{Id} + K^*). \tag{32}$$

Once this is established, we replace K by K^* and use the observation that $(K^*)^* = K$, together with Schauder's Theorem regarding the compactness of K^*, to obtain the inequality in the opposite direction. We argue by contradiction to verify (32). Otherwise, $\dim\ker(\mathrm{Id}+K) < \dim\ker(\mathrm{Id}+K^*)$. Let P be the orthogonal projection of H onto $\ker(\mathrm{Id}+K)$ and A a linear mapping of $\ker(\mathrm{Id}+K)$ into $\ker(\mathrm{Id}+K^*)$ that is one-to-one but not onto. Define $K' = K + A \circ P$. Since $\mathrm{Id}+K$ has closed image, according to Proposition 14,

$$H = \mathrm{Im}(\mathrm{Id}+K) \oplus \ker(\mathrm{Id}+K^*)$$

and therefore $\mathrm{Id}+K'$ is one-to-one but not onto. On the other hand, since $A \circ P$ is of finite rank, it is compact and therefore so is K'. These two assertions contradict the preceding proposition. Therefore, (31) is established. Since $\mathrm{Id}+K$ has closed image, it follows from (14) and (31) that $\mathrm{Id}+K$ is one-to-one if and only if it is onto. □

Corollary 23 (The Fredholm Alternative) *Let $K \in \mathcal{L}(H)$ be a compact operator, and $\mu \neq 0$. Then exactly one of the following holds:*

(i) There is a non-zero solution of the equation

$$\mu h - K(h) = 0, h \in H.$$

(ii) For every $h_0 \in H$, there is a unique solution of the equation

$$\mu h - K(h) = h_0, h \in H.$$

Definition *An operator $T \in \mathcal{L}(H)$ is said to be **Fredholm** provided that the kernel of T is finite dimensional and the image of T has finite codimension. For such an operator, its index, ind T, is defined by*

$$\mathrm{ind}\ T = \dim\ \ker T - \mathrm{codim}\ \mathrm{Im}\, T.$$

Theorem 24 *An operator $T \in \mathcal{L}(H)$ is Fredholm of index 0 if and only if $T = S + K$, where $S \in \mathcal{L}(H)$ is invertible and $K \in \mathcal{L}(H)$ is compact.*

Proof First assume that T is Fredholm of index 0: According to Theorem 12 of Chapter 17, the image of T is closed since it is of finite codimension in H, and consequently, by Proposition 15,

$$H = \mathrm{Im}\, T \oplus \ker\ T^*. \tag{33}$$

Since $\dim\ker\ T = \dim\ker\ T^* < \infty$, we may choose a one-to-one linear operator A of $\ker\ T$ onto $\ker\ T^*$. Let P be the orthogonal projection of H onto $\ker\ T$. Define $K = A \circ P \in \mathcal{L}(H)$ and $S = T - K$. Then $T = S + K$. The operator K is compact since it is of finite rank, while the operator S is invertible by (33) and the choice of P and A. Hence T is a compact perturbation of a invertible operator.

To prove the converse, suppose $T = S+K$, where $S \in \mathcal{L}(H)$ is invertible and $K \in \mathcal{L}(H)$ is compact. Observe that

$$T = S \circ [\mathrm{Id}+S^{-1} \circ K]. \tag{34}$$

Since S^{-1} is continuous and K is compact, $S^{-1} \circ K$ is compact. According to the Riesz-Schauder Theorem, $\mathrm{Id}+S^{-1} \circ K$ is Fredholm of index 0. The composition of a Fredholm operator with an invertible operator is also Fredholm of index 0 (see Problem 55). It follows from (34) that T is Fredholm of index 0. $\square$

We leave it as an exercise to establish the following corollary.

Corollary 25 *If the operators T and S in $\mathcal{L}(H)$ are Fredholm of index 0, then the composition $S \circ T$ also is Fredholm of index 0.*

Theorem 24 holds for operators on a general Banach space, where a compact linear operator is defined to be an operator that maps the unit ball to a set having strongly compact closure[2]. However, the general method of proof must be quite different. First of all, the adjoint of a linear operator on a general Banach space is an operator on the dual space. Also, an essential ingredient in the proof of Proposition 22 is the approximation in $\mathcal{L}(H)$ of a linear compact operator by a linear operator of finite rank. Per Enflo has shown that there are linear compact operators on a separable Banach space that cannot be approximated in $\mathcal{L}(H)$ by linear operators of finite rank.

PROBLEMS

53. Let $K \in \mathcal{L}(H)$ be compact. Show that $T = K^*K$ is compact and symmetric. Then use the Hilbert-Schmidt Theorem to show that there is an orthonormal sequence $\{\varphi_k\}$ of H such that $T(\varphi_k) = \lambda_k \varphi_k$ for all k and $T(h) = 0$ if h is orthogonal to $\{\varphi_k\}_{k=0}^{\infty}$. Conclude that if h is orthogonal to $\{\varphi_k\}_{k=0}^{\infty}$, then
$$\|K(h)\|^2 = \langle K(h),\, K(h)\rangle = \langle T(h),\, h\rangle = 0.$$
Define H_0 to be the closed linear span of $\{K^m(\varphi_k) \mid m \geq 0, k \geq 1\}$. Show that H_0 is closed and separable, $K(H_0) \subseteq H_0$ and $K = 0$ on $H_0^{\perp}$.
54. Let $\mathcal{K}(H)$ denote the set of compact operators in $\mathcal{L}(H)$. Show that $\mathcal{K}(H)$ is a closed subspace of $\mathcal{L}(H)$ that has the set of operators of finite rank as a dense subspace. Is $\mathcal{K}(H)$ an open subset of $\mathcal{L}(H)$?
55. Show that the composition in either order of a Fredholm operator of index 0 with an invertible operator is also Fredholm of index 0.
56. Show that the composition of two Fredholm operators of index 0 is also Fredholm of index 0.
57. Show that an operator $T \in \mathcal{L}(H)$ is Fredholm of index 0 if and only if it is the perturbation of an invertible operator by an operator of finite rank.
58. Argue as follows to show that the collection of invertible operators in $\mathcal{L}(H)$ is an open subset of $\mathcal{L}(H)$.
    (i) For $A \in \mathcal{L}(H)$ with $\|A\| < 1$, use the completeness of $\mathcal{L}(H)$ to show that the so-called Neumann series $\sum_{n=0}^{\infty} A^n$ converges to an operator in $\mathcal{L}(H)$ that is the inverse of $\mathrm{Id} - A$.
    (ii) For an invertible operator $S \in \mathcal{L}(H)$ show that for any $T \in \mathcal{L}(H)$, $T = S[\mathrm{Id} + S^{-1}(T - S)]$.
    (iii) Use (i) and (ii) to show that if $S \in \mathcal{L}(H)$ is invertible then so is any $T \in \mathcal{L}(H)$ for which $\|S - T\| < 1/\|S^{-1}\|$.

[2] See Chapter 21 of Peter Lax's *Functional Analysis* ([Lax02]).

59. Show that the set of operators in $\mathcal{L}(H)$ that are Fredholm of index 0 is an open subset of $\mathcal{L}(H)$.
60. By following the orthogonal approximation sequence method used in the proof of Proposition 22, provide another proof of Lemma 14 in case H is separable.
61. For $T \in \mathcal{L}(H)$, suppose that $\langle T(h), h\rangle \geq \|h\|^2$ for all $h \in H$. Assume that $K \in \mathcal{L}(H)$ is compact and $T+K$ is one-to-one. Show that $T+K$ is onto.
62. Let $K \in \mathcal{L}(H)$ be compact and $\mu \in \mathbf{R}$ have $|\mu| > \|K\|$. Show that $\mu - K$ is invertible.
63. Let $S \in \mathcal{L}(H)$ have $\|S\| < 1$, $K \in \mathcal{L}(H)$ be compact and $(\mathrm{Id} + S + K)(H) = H$. Show that $\mathrm{Id} + S + K$ is one-to-one.
64. Let $\mathcal{G}L(H)$ denote the set of invertible operators in $\mathcal{L}(H)$.
    (i) Show that under the operation of composition of operators, $\mathcal{G}L(H)$ is a group: it is called the general linear group of H.
    (ii) An operator T in $\mathcal{G}L(H)$ is said be orthogonal, provided that $T^* = T^{-1}$. Show that the set of orthogonal operators is a subgroup of $\mathcal{G}L(H)$: it is called the orthogonal group.
65. Let $T \in \mathcal{L}(H)$ be Fredholm of index zero, and $K \in \mathcal{L}(H)$ be compact. Show that $T+K$ is Fredholm of index zero.
66. Let X_0 be a finite codimensional subspace of a linear space X. Show that all finite dimensional linear complements of X_0 in X have the same dimension.

This page intentionally left blank

PART FOUR

MEASURE AND TOPOLOGY: INVARIANT MEASURES

This page intentionally left blank

CHAPTER 21

Measure and Topology

Contents

21.1 Locally Compact Topological Spaces 424
21.2 Separating Sets and Extending Functions 428
21.3 The Construction of Radon Measures 430
21.4 The Representation of Positive Linear Functionals on $C_c(X)$: The Riesz-Markov Theorem 434
21.5 The Representation of the Dual of $C(X)$: The Riesz-Kakutani Theorem . . 439
21.6 Regularity Properties of Baire Measures 446

In the study of Lebesgue measure, μ_n, and Lebesgue integration on the Euclidean spaces $\mathbf{R}^n$ and, in particular, on the real line, we explored connections between Lebesgue measure and the Euclidean topology and between the measurable functions and continuous ones. The Borel σ-algebra $\mathcal{B}(\mathbf{R}^n)$ is contained in the σ-algebra of Lebesgue measurable sets. Therefore, if we define $C_c(\mathbf{R}^n)$ to be the linear space of compactly supported, continuous functions $f\colon R^n \to R$, the functional

$$f \mapsto \int_{\mathbf{R}^n} f\, d\mu_n \text{ for all } f \in C_c(\mathbf{R}^n)$$

is properly defined and positive[1]. Moreover, for K a closed, bounded subset of $\mathbf{R}^n$, the operator

$$f \mapsto \int_K f\, d\mu_n \text{ for all } f \in C(K)$$

is properly defined, positive, and is a bounded linear operator if $C(K)$ has the maximum norm. In this chapter, we consider a general locally compact topological space $(X, \mathcal{T})$, the Borel σ-algebra $\mathcal{B}(X)$ comprising the smallest σ-algebra containing the topology $\mathcal{T}$, and integration with respect to a Borel measure $\mu\colon \mathcal{B}(X) \to [0, \infty)$. The chapter has two centerpieces. The first is the Riesz-Markov Theorem, which tells us that every positive linear function on $C_c(X)$ is given by integration against a Borel measure on $\mathcal{B}(X)$. The Riesz-Markov Theorem enables us to prove the Riesz-Kakutani Representation Theorem, which tells us that, for X a compact, Hausdorff topological space, every bounded linear functional on the linear space $C(X)$, normed with the maximum norm, is given by integration against a signed Borel measure. The Riesz-Kakutani Representation Theorem provides the opportunity, which we pursue in the final chapter, for the application of the Helly, Alaoglu, and Krein-Milman Theorems to collections of Borel measures.

[1] A linear functional L on a space of real-valued functions on a set X is called positive, provided $L(f) \geq 0$ if $f \geq 0$ on X. But, for a linear functional, positivity means $L(h) \geq L(g)$ if $h \geq g$ on X. So in view of our perpetual dependence on the monotonicity property of integration, the adjective "monotone" is certainly better than "positive." However, we will respect convention and use of the adjective "positive."

The proofs of these two representation theorems require an examination of the relationship between the topology on a set and the measures on the Borel sets associated with the topology. The technique by which we construct Borel measures that represent functionals is the same one we used to construct Lebesgue measure on Euclidean space: We study the Carathéodory extension of premeasures defined on a collection $\mathcal{S}$ of subsets of X, now taking $\mathcal{S} = \mathcal{T}$, the topology on X. We begin the chapter with a preliminary section on locally compact topological spaces. In the second section, we gather all the properties of such spaces that we need into a single theorem and provide a separate very simple proof of this theorem for X a locally compact metric space.

21.1 LOCALLY COMPACT TOPOLOGICAL SPACES

A topological space X is called **locally compact** provided each point in X has a neighborhood that has compact closure. Every compact space is locally compact, while the Euclidean spaces $\mathbf{R}^n$ are examples of spaces that are locally compact but not compact. Riesz's Theorem tells us that an infinite dimensional normed linear space, with the topology induced by the norm, is not locally compact. In this section, we establish properties of locally compact spaces, which will be the basis of our subsequent study of measure and topology.

Variations on Urysohn's Lemma

Recall that we broadened the meaning of the word *neighborhood* and, for a subset K of a topological space X call an open set that contains K a neighborhood of K.

Lemma 1 *Let x be a point in a locally compact, Hausdorff space X and $\mathcal{O}$ a neighborhood of x. Then there is a neighborhood $\mathcal{V}$ of x that has compact closure contained in $\mathcal{O}$, that is,*

$$x \in \mathcal{V} \subseteq \overline{\mathcal{V}} \subseteq \mathcal{O} \text{ and } \overline{\mathcal{V}} \text{ is compact.}$$

Proof Let $\mathcal{U}$ be a neighborhood of x that has compact closure. Then the topological space $\overline{\mathcal{U}}$ is compact and Hausdorff and therefore is normal. The set $\mathcal{O} \cap \mathcal{U}$ is a neighborhood, with respect to the $\overline{\mathcal{U}}$ topology, of x. Therefore, by the normality of $\overline{\mathcal{U}}$, there is a neighborhood $\mathcal{V}$ of x that has compact closure contained in $\mathcal{O} \cap \mathcal{U}$: Here both neighborhood and closure mean with respect to the $\overline{\mathcal{U}}$ topology. Since $\mathcal{O}$ and $\mathcal{U}$ are open in X, it follows from the definition of the subspace topology that $\mathcal{V}$ is open in X and $\overline{\mathcal{V}} \subseteq \mathcal{O}$ where the closure now is with respect to the X topology. □

Proposition 2 *Let K be a compact subset of a locally compact, Hausdorff space X and $\mathcal{O}$ a neighborhood of K. Then there is a neighborhood $\mathcal{V}$ of K that has compact closure contained in $\mathcal{O}$, that is,*

$$K \subseteq \mathcal{V} \subseteq \overline{\mathcal{V}} \subseteq \mathcal{O} \text{ and } \overline{\mathcal{V}} \text{ is compact.}$$

Proof By the preceding lemma, each point $x \in K$ has a neighborhood $\mathcal{N}_x$ that has compact closure contained in $\mathcal{O}$. Then $\{\mathcal{N}_x\}_{x \in K}$ is an open cover of the compact set K. Choose a finite subcover $\{\mathcal{N}_{x_i}\}_{i=1}^n$ of K. The set $\mathcal{V} = \bigcup_{i=1}^n \mathcal{N}_{x_i}$ is a neighborhood of K and

$$\overline{\mathcal{V}} \subseteq \bigcup_{i=1}^n \overline{\mathcal{N}}_{x_i} \subseteq \mathcal{O}.$$

The set $\bigcup_{i=1}^{n} \overline{\mathcal{N}}_{x_i}$, being the union of a finite collection of compact sets, is compact and hence so is $\overline{\mathcal{V}}$ since it is a closed subset of a compact space. □

For a real-valued function f on a topological space X, the **support** of f, which we denote by $\operatorname{supp} f$, is defined to be the closure of the set $\{x \in X \mid f(x) \neq 0\}$, that is,

$$\operatorname{supp} f = \overline{\{x \in X \mid \ f(x) \neq 0\}}.$$

We denote the collection of continuous functions $f \colon X \to \mathbf{R}$ that have compact support by $C_c(X)$. Thus, a function belongs to $C_c(X)$ if and only if it is continuous and vanishes outside of a compact set.

Proposition 3 *Let K be a compact subset of a locally compact, Hausdorff space X and $\mathcal{O}$ a neighborhood of K. Then there is a function $f \in C_c(X)$ for which*

$$f = 1 \text{ on } K, f = 0 \text{ on } X \sim \mathcal{O} \text{ and } 0 \leq f \leq 1 \text{ on } X. \tag{1}$$

Proof By the preceding proposition, there is a neighborhood $\mathcal{V}$ of K that has compact closure contained in $\mathcal{O}$. Since $\overline{\mathcal{V}}$ is compact and Hausdorff, it is normal. Moreover, K and $\overline{\mathcal{V}} \sim \mathcal{V}$ are disjoint closed subsets of $\overline{\mathcal{V}}$. According to Urysohn's Lemma, there is a continuous real-valued function f on $\overline{\mathcal{V}}$ for which

$$f = 1 \text{ on } K, \ f = 0 \text{ on } \overline{\mathcal{V}} \sim \mathcal{V} \text{ and } 0 \leq f \leq 1 \text{ on } \overline{\mathcal{V}}.$$

Extend f to all of X by setting $f = 0$ on $X \sim \overline{\mathcal{V}}$. Then $f \in C_c(X)$ and has the properties described in (1). □

Recall that a subset of a topological space is called a G_δ set provided it is the intersection of a countable collection of open sets.

Corollary 4 *Let K be a compact G_δ subset of a locally compact, Hausdorff space X. Then there is a function $f \in C_c(X)$ for which*

$$K = \{x \in X \mid \ f(x) = 1\}.$$

Proof According to Proposition 2, there is a neighborhood $\mathcal{U}$ of K that has compact closure. Since K is a G_δ set, there is a countable collection $\{\mathcal{O}_k\}_{k=1}^{\infty}$ of open sets whose intersection is K. We may assume that $\mathcal{O}_k \subseteq \mathcal{U}$ for all k. By the preceding proposition, for each k there is a continuous real-valued function f_k on X for which

$$f_k = 1 \text{ on } K, f_k = 0 \text{ on } X \sim \overline{\mathcal{O}}_k \text{ and } 0 \leq f_k \leq 1 \text{ on } X.$$

Since the sequence of functions $\{f_k\}$ is uniformly pointwise bounded and a geometric series converges, we see that the function f defined by

$$f = \sum_{k=1}^{\infty} 2^{-k} f_k \text{ on } X$$

has the desired property. □

Partitions of Unity

Definition *Let K be a subset of a topological space X and $\{\mathcal{O}_k\}_{k=1}^n$ be a finite cover of K by open sets. A collection of continuous real-valued functions on X, $\{\varphi_k\}_{k=1}^n$, is called a* **partition of unity** *for K subordinate to $\{\mathcal{O}_k\}_{k=1}^n$ provided*

$$\operatorname{supp}\varphi_i \subseteq \mathcal{O}_i,\ 0 \leq \varphi_i \leq 1 \text{ on } X \text{ for } 1 \leq i \leq n$$

and

$$\varphi_1 + \varphi_2 + \cdots + \varphi_n = 1 \text{ on } K.$$

Proposition 5 *Let K be a compact subset of a locally compact, Hausdorff space X and $\{\mathcal{O}_k\}_{k=1}^n$ be a finite cover of K by open sets. Then there is a partition of unity $\{\varphi_k\}_{k=1}^n$ for K subordinate to this finite cover and each φ_k has compact support.*

Proof We first claim that there is an open cover $\{\mathcal{U}_k\}_{k=1}^n$ of K such that for each k, $\overline{\mathcal{U}_k}$ is a compact subset of $\mathcal{O}_k$. Indeed, invoking Proposition 2 n times, for each $x \in K$, there is a neighborhood $\mathcal{N}_x$ of x that has compact closure and such that if $1 \leq j \leq n$ and x belongs to $\mathcal{O}_j$, then $\overline{\mathcal{N}_x} \subseteq \mathcal{O}_j$. The collection of open sets $\{\mathcal{N}_x\}_{x \in K}$ is a cover of K and K is compact. Therefore, there is a finite set of points $\{x_k\}_{k=1}^m$ in K for which $\{\mathcal{N}_{x_k}\}_{1 \leq k \leq m}$ also covers K. For $1 \leq k \leq n$, let $\mathcal{U}_k$ be the unions of those $\mathcal{N}_{x_j}$'s that are contained in $\mathcal{O}_k$. Then $\{\mathcal{U}_1, \ldots, \mathcal{U}_n\}$ is an open cover of K and for each k, $\overline{\mathcal{U}}_k$ is a compact subset of $\mathcal{O}_k$ since it is the finite union of such sets. We infer from Proposition 3 that for each $k, 1 \leq k \leq n$, there is a function $f_k \in C_c(X)$ for which $f_k = 1$ on $\overline{\mathcal{U}_k}$ and $f = 0$ on $X \sim \mathcal{O}_k$. The same proposition tells us that there is a function $h \in C(X)$ for which $h = 1$ on K and $h = 0$ on $X \sim \bigcup_{k=1}^n \mathcal{U}_k$. Define

$$f = \sum_{k=1}^n f_k \text{ on } X.$$

Observe that $f + [1 - h] > 0$ on X and $h = 0$ on K. Therefore, if we define

$$\varphi_k = \frac{f_k}{f + [1-h]} \text{ on } X \text{ for } 1 \leq k \leq n,$$

$\{\varphi_k\}_{k=1}^n$ is a partition of unity for K subordinate to $\{\mathcal{O}_k\}_{k=1}^n$ and each φ_k has compact support. □

The Alexandroff One-Point Compactification

If X is a locally compact, Hausdorff space, then a new space X^* may be defined by adjoining to X a single point ω not in X and defining a set in X^* to be open provided it is either an open subset of X or the complement of a compact subset in X. Then X^* is a compact, Hausdorff space, and the identity mapping of X into X^* is a homeomorphism of X and $X^* \sim \{\omega\}$. The space X^* is called the **Alexandroff one-point compactification** of X, and ω is often referred to as **the point at infinity** in X^*.

The proof of the following variant, for locally compact, Hausdorff spaces, of the Tietze Extension Theorem nicely illustrates the usefulness of the Alexandroff compactification.

Theorem 6 *Let K be a compact subset of a locally compact, Hausdorff space X. Each continuous function $f\colon K \to \mathbf{R}$ has a continuous extension $f\colon X \to \mathbf{R}$.*

Proof The Alexandroff compactification of X, X^*, is a compact, Hausdorff space. Moreover, K is a closed subset of X^*, since its complement in X^* is open. A compact, Hausdorff space is normal. It follows from the Tietze Extension Theorem that f may be extended to a continuous real-valued function on all of X^*. The restriction to X of this extension is a continuous extension of f to all of X. □

PROBLEMS

1. Let X be a locally compact, Hausdorff space, and F a set that has closed intersection with each compact subset of X. Show that F is closed.
2. Regarding the proof of Proposition 3:
   (i) Show that F and $\overline{\mathcal{V}} \sim \mathcal{V}$ are closed subsets of $\overline{\mathcal{V}}$.
   (ii) Show that the function f is continuous.
3. Let X be a locally compact, Hausdorff space and X^* the Alexandroff one-point compactification of X:
   (i) Prove that the subsets of X^* that are either open subsets of X or the complements of compact subsets of X are a topology for X^*.
   (ii) Show that the identity mapping from X to the subspace $X^* \sim \{\omega\}$ is a homeomorphism.
   (iii) Show that X^* is compact and Hausdorff.
4. Show that the Alexandroff one-point compactification of $\mathbf{R}^n$ is homeomorphic to the n-sphere $S^n = \{x \in \mathbf{R}^{n+1} \mid ||x|| = 1\}$.
5. Show that an open subset of a locally compact, Hausdorff space, with its subspace topology, is locally compact.
6. Show that a closed subset of a locally compact space, with its subspace topology, is locally compact.
7. Show that a locally compact, Hausdorff space X is compact if and only if the set consisting of the point at infinity is an open subset of the Alexandroff one-point compactification X^* of X.
8. Let X be a locally compact, Hausdorff space. Show that the Alexandroff one-point compactification X^* is separable if and only if X is separable.
9. Consider the topological space X consisting of the set of real numbers with the topology that has complements of countable sets as a base. Show that X is not locally compact.
10. Provide a proof of Proposition 3 by applying Urysohn's Lemma to the Alexandroff one-point compactification of X.
11. Let f continuously map the locally compact, Hausdorff space X onto the topological space Y. Is Y necessarily locally compact?
12. Let X be a topological space and f a continuous function of X that has compact support. Define $K = \{x \in X \mid f(x) = 1\}$. Show that K is a compact G_δ set.
13. Let $\mathcal{O}$ be an open subset of a compact, Hausdorff space X. Show that the mapping of X to the Alexandroff one-point compactification of $\mathcal{O}$ that is the identity on $\mathcal{O}$ and takes each point in $X \sim \mathcal{O}$ into ω is continuous.

14. Let X and Y be locally compact, Hausdorff spaces, and f a continuous mapping of X into Y. Let X^* and Y^* be the Alexandroff one-point compactifications of X and Y, and f^* the mapping of X^* into Y^* whose restriction to X is f and that takes the point at infinity in X^* into the point at infinity in Y^*. Show that f^* is continuous if and only if $f^{-1}(K)$ is compact whenever $K \subseteq Y$ is compact. A mapping f with this property is said to be proper.
15. Let X be a locally compact, Hausdorff space. Show that a subset F of X is closed if and only if $F \cap K$ is closed for each compact subset K of X. Moreover, show that the same equivalence holds if instead of being locally compact the space X is first countable.
16. Let $\mathcal{F}$ be a family of real-valued continuous functions on a locally compact, Hausdorff space X which has the following properties:
    (i) If $f \in \mathcal{F}$ and $g \in \mathcal{F}$, then $f + g \in \mathcal{F}$.
    (ii) If $f \in \mathcal{F}$ and $g \in \mathcal{F}$, then $f/g \in \mathcal{F}$, provided that $\operatorname{supp} f \subseteq \{x \in X \mid g(x) \neq 0\}$.
    (iii) Given a neighborhood $\mathcal{O}$ of a point $x_0 \in X$, there is a $f \in \mathcal{F}$ with $f(x_0) = 1$, $0 \leq f \leq 1$ and $\operatorname{supp} f \subseteq \mathcal{O}$.

    Show that Proposition 5 is still true if we require that the functions in the partition of unity belong to $\mathcal{F}$.
17. Let K be a compact G_δ subset of a locally compact, Hausdorff space X. Show that there is a decreasing sequence of continuous non-negative real-valued functions on X that converges pointwise on X to the characteristic function of K.
18. The Baire Category Theorem asserts that in a complete metric space the intersection of a countable collection of open dense sets is dense. At the heart of its proof lies the Cantor Intersection Theorem. Show that the Fréchet Intersection Theorem is a sufficiently strong substitute for the Cantor Intersection Theorem to provide a proof of the following assertion, by first proving it in the case in which X is compact: Let X be a locally compact, Hausdorff space.
    (i) If $\{F_n\}_{n=1}^\infty$ is a countable collection of closed subsets of X for which each F_n has empty interior, then the union $\bigcup_{n=1}^\infty F_n$ also has empty interior.
    (ii) If $\{\mathcal{O}_n\}_{n=1}^\infty$ is a countable collection of open dense subsets of X, then the intersection $\bigcap_{n=1}^\infty \mathcal{O}_n$ also is dense.
19. Use the preceding problem to prove the following: Let X be a locally compact, Hausdorff space. If $\mathcal{O}$ is an open subset of X that is contained in a countable union $\bigcup_{n=1}^\infty F_n$ of closed subsets of X, then the union of their interiors, $\bigcup_{n=1}^\infty \operatorname{int} F_n$, is an open dense subset of $\mathcal{O}$.
20. For a map $f\colon X \to Y$ and a collection $\mathcal{C}$ of subsets of Y we define $f^*\mathcal{C}$ to be the collection of subsets of X given by
$$f^*\mathcal{C} = \left\{E \;\middle|\; E = f^{-1}[C] \text{ for some } C \in \mathcal{C}\right\}.$$
    Show that if $\mathcal{A}$ is the σ-algebra generated by $\mathcal{C}$, then $f^*\mathcal{A}$ is the σ-algebra generated by $f^*\mathcal{C}$.
21. For a map $f\colon X \to Y$ and a collection $\mathcal{C}$ of subsets of X, let $\mathcal{A}$ be the σ-algebra generated by $\mathcal{C}$. If $f^{-1}[f[C]] = C$ for each $C \in \mathcal{C}$, show that $f^{-1}[f[A]] = A$ for each $A \in \mathcal{A}$.

21.2 SEPARATING SETS AND EXTENDING FUNCTIONS

We gather in the statement of the following theorem the three properties of locally compact, Hausdorff spaces which we will employ in the proofs of our forthcoming representation theorems.

Theorem 7 *Let $(X, \mathcal{T})$ be a Hausdorff space. Then the following four properties are equivalent:*

(i) $(X, \mathcal{T})$ *is locally compact.*

(ii) If $\mathcal{O}$ is a neighborhood of a compact subset K of X, then there is a neighborhood $\mathcal{U}$ of K that has compact closure contained in $\mathcal{O}$.

(iii) If $\mathcal{O}$ is a neighborhood of a compact subset K of X, then the constant function on K that takes the value 1 may be extended to a function f in $C_c(X)$ for which $0 \leq f \leq 1$ on X and f vanishes outside of $\mathcal{O}$.

(iv) For K a compact subset of X and $\mathcal{F}$ a finite open cover of K, there is a partition of unity subordinate to $\mathcal{F}$ consisting of functions of compact support.

Proof We first establish the equivalence of (i) and (ii). Assume that (ii) holds. Let x be a point in X. Then X is a neighborhood of the compact set $\{x\}$. By property (ii) there is a neighborhood of $\{x\}$ that has compact closure. Thus X is locally compact. Now assume that X is locally compact. Proposition 2 tells us that (ii) holds.

Next we establish the equivalence of (i) and (iii). Assume that (iii) holds. Let x be a point in X. Then X is a neighborhood of the compact set $\{x\}$. By property (iii) there is a function f in $C_c(X)$ to take the value 1 at x. Then $\mathcal{O} = f^{-1}(1/2, 3/2)$ is a neighborhood of x and it has compact closure since f has compact support and $\overline{\mathcal{O}} \subseteq f^{-1}[1/2, 3/2]$. Thus X is locally compact. Now assume that X is locally compact. Proposition 3 tells us that (iii) holds.

Finally, we establish the equivalence of (i) and (iv). Assume that property (iv) holds. Let x be a point in X. Then X is a neighborhood of the compact set $\{x\}$. By property (iv) there is a single function f that is a partition of unity subordinate to the covering of the compact set $\{x\}$ by single open set X. Then $\mathcal{O} = f^{-1}(1/2, 3/2)$ is a neighborhood of x and it has compact closure. Thus X is locally compact. Now assume that X is locally compact. Proposition 5 tells us that (iv) holds. □

The substantial implications in the above theorem are that a locally compact, Hausdorff space possesses properties (ii), (iii), and (iv). Their proofs, which we presented in the preceding section, depend on Urysohn's Lemma. It is interesting to note, however, that if X is a locally compact metric space, then very direct proofs show that X possesses properties (ii), (iii), and (iv). Indeed, suppose there is a metric $\rho: X \times X \to \mathbf{R}$ that induces the topology $\mathcal{T}$ and X is locally compact.

Proof of Property (ii) For each $x \in K$, since $\mathcal{O}$ is open and X is locally compact, there is an open ball $B(x, r_x)$ of compact closure that is contained in $\mathcal{O}$. Then $\{B(x, r_x/2)\}_{x \in K}$ is a cover of K by open sets. The set K is compact. Therefore, there are a finite set of points $x_1, \ldots, x_n$ in K for which $\{B(x, r_{x_k}/2)\}_{1 \leq k \leq n}$ cover K. Then $\mathcal{U} = \bigcup_{1 \leq k \leq n} B(x, r_{x_k}/2)$ is a neighborhood of K that, since $\overline{\mathcal{U}} \subseteq \bigcup_{1 \leq k \leq n} \overline{B}(x, r_{x_k}/2)$, has closure contained in $\mathcal{O}$.

Proof of Property (iii) For a subset A of X, define the function called the distance to A and denoted by $\mathrm{dist}_A: X \to [0, \infty)$ by

$$\mathrm{dist}_A(x) = \inf_{y \in A} \rho(x, y) \text{ for } x \in X.$$

The function dist_A is continuous; indeed, it is Lipschitz with Lipschitz constant 1 (see Problem 25). Moreover, if A is closed subset of X, then $\text{dist}_A(x) = 0$ if and only if $x \in A$. For $\mathcal{O}$ a neighborhood of a compact set K, by part (i) choose $\mathcal{U}$ to be a neighborhood of K that has compact closure contained in $\mathcal{O}$. Define

$$f = \frac{\text{dist}_{X \sim \mathcal{U}}}{\text{dist}_{X \sim \mathcal{U}} + \text{dist}_K} \text{ on } X.$$

Then $f \in C_c(X)$ takes values in $[0, 1]$, $f = 1$ on K, and $f = 0$ on $X \sim \mathcal{O}$.

Proof of Property (iv) This follows from properties (ii) and (i) as it did in the case in which X is Hausdorff but not necessarily metrizable; see the proof of Proposition 5.

We see that property (ii) is equivalent to the assertion that two disjoint closed subsets of X, one of which is compact, may be separated by disjoint neighborhoods. We therefore refer to property (ii) as **the locally compact separation property**. It is convenient to call (iii) **the locally compact extension property**.

PROBLEMS

22. Show that Euclidean space $\mathbf{R}^n$ is locally compact.
23. Show that ℓ^p, for $1 \le p \le \infty$, fails to be locally compact.
24. Show that $C([0, 1])$, with the topology induced by the maximum norm, is not locally compact.
25. Let $\rho\colon X \times X \to \mathbf{R}$ be a metric on a set X. For $A \subseteq X$, consider the distance function

$$\text{dist}_A\colon X \to [0, \infty).$$

   (i) Show that the function dist_A is continuous.
   (ii) If $A \subseteq X$ is closed and x is a point in X, show that $\text{dist}_A(x) = 0$ if and only if $x \in A$.
   (iii) If $A \subseteq X$ is closed and $x \in X$, show that there may not exist a point x_0 in A for which $\text{dist}_A(x) = \rho(x, x_0)$, but there is such a point x_0 if K is compact.

26. Show that property (ii) in the statement of Theorem 7 is equivalent to the assertion that two disjoint closed subsets of X, one of which is compact, may be separated by disjoint neighborhoods.

21.3 THE CONSTRUCTION OF RADON MEASURES

Let $(X, \mathcal{T})$ be a topological space. The purpose of this section is to construct measures on the Borel σ-algebra, $\mathcal{B}(X)$, comprising the smallest σ-algebra that contains the topology $\mathcal{T}$. A natural place to begin is to consider premeasures $\mu\colon \mathcal{T} \to [0, \infty]$ defined on the topology $\mathcal{T}$ and consider the Carathéodory measure induced by μ. If we can establish that each open set is measurable with respect to μ^*, then, by the minimality with respect to inclusion of the Borel σ-algebra among all σ-algebras containing the open sets, each Borel set will be μ^*-measurable and the restriction of μ^* to $\mathcal{B}(X)$ will be an extension of μ. We ask the following question: What properties of $\mu\colon \mathcal{T} \to [0, \infty]$ are sufficient in order that every open set be measurable with respect to μ^*, the outer-measure induced by μ. It is not useful to invoke the Carathéodory-Hahn Theorem here. A topology, in general, is not a semi-ring.

Indeed, it is not difficult to see that a Hausdorff topology $\mathcal{T}$ is a semi-ring if and only if $\mathcal{T}$ is the discrete topology, that is, every subset of X is open (see Problem 27).

Lemma 8 *Let $(X, \mathcal{T})$ be a topological space, $\mu\colon \mathcal{T} \to [0, \infty]$ a premeasure, and μ^* the outer-measure induced by μ. Then for any subset E of X,*

$$\mu^*(E) = \inf \{\mu(\mathcal{U}) \mid \mathcal{U} \text{ a neighborhood of } E\}. \tag{2}$$

Furthermore, E is μ^*-measurable if and only if

$$\mu(\mathcal{O}) \geq \mu^*(\mathcal{O} \cap E) + \mu^*(\mathcal{O} \sim E) \text{ for each open set } \mathcal{O} \text{ for which } \mu(\mathcal{O}) < \infty. \tag{3}$$

Proof Since the union of any collection of open sets is open, (2) follows from the countable monotonicity of μ. Let E be a subset of X for which (3) holds. To show that E is μ^*-measurable, let A be a subset of X for which $\mu^*(A) < \infty$ and let $\epsilon > 0$. We must show that

$$\mu^*(A) + \epsilon \geq \mu^*(A \cap E) + \mu^*(A \sim E). \tag{4}$$

By the above characterization (2) of outer-measure, there is an open set $\mathcal{O}$ for which

$$A \subseteq \mathcal{O} \text{ and } \mu^*(A) + \epsilon \geq \mu^*(\mathcal{O}). \tag{5}$$

On the other hand, by (3) and the monotonicity of μ^*,

$$\mu^*(\mathcal{O}) \geq \mu^*(\mathcal{O} \cap E) + \mu^*(\mathcal{O} \sim E) \geq \mu^*(A \cap E) + \mu^*(A \sim E). \tag{6}$$

Inequality (4) follows from the inequalities (5) and (6). □

Proposition 9 *Let $(X, \mathcal{T})$ be a topological space and $\mu\colon \mathcal{T} \to [0, \infty]$ a premeasure. Assume that for each open set $\mathcal{O}$ for which $\mu(\mathcal{O}) < \infty$,*

$$\mu(\mathcal{O}) = \sup \left\{\mu(\mathcal{U}) \;\middle|\; \mathcal{U} \text{ open and } \overline{\mathcal{U}} \subseteq \mathcal{O}\right\}. \tag{7}$$

Then every open set is μ^*-measurable and the measure $\mu^*\colon \mathcal{B}(X) \to [0, \infty]$ is an extension of μ.

Proof A premeasure is countable monotone and hence, for each open set $\mathcal{V}$, $\mu^*(\mathcal{V}) = \mu(\mathcal{V})$. Therefore, by the minimality property of $\mathcal{B}(X)$, to complete the proof it suffices to show that each open set is μ^*-measurable.

Let $\mathcal{V}$ be open. To verify the μ^*-measurability of $\mathcal{V}$ it suffices, by the preceding lemma, to let $\mathcal{O}$ be open with $\mu(\mathcal{O}) < \infty$, let $\epsilon > 0$ and show that

$$\mu(\mathcal{O}) + \epsilon \geq \mu(\mathcal{O} \cap \mathcal{V}) + \mu^*(\mathcal{O} \sim \mathcal{V}). \tag{8}$$

However, $\mathcal{O} \cap \mathcal{V}$ is open and, by the monotonicity of μ, $\mu(\mathcal{O} \cap \mathcal{V}) < \infty$. By assumption (7) there is an open set $\mathcal{U}$ for which $\overline{\mathcal{U}} \subseteq \mathcal{O} \cap \mathcal{V}$ and

$$\mu(\mathcal{U}) > \mu(\mathcal{O} \cap \mathcal{V}) - \epsilon.$$

The pair of sets $\mathcal{U}$ and $\mathcal{O} \sim \overline{\mathcal{U}}$ are disjoint open subsets of $\mathcal{O}$. Therefore, by the monotonicity and finite additivity of the premeasure μ,

$$\mu(\mathcal{O}) \geq \mu(\mathcal{U} \cup [\mathcal{O} \sim \overline{\mathcal{U}}]) = \mu(\mathcal{U}) + \mu(\mathcal{O} \sim \overline{\mathcal{U}}).$$

On the other hand, since $\overline{\mathcal{U}} \subseteq \mathcal{V} \cap \mathcal{O}$,

$$\mathcal{O} \sim \mathcal{V} = \mathcal{O} \sim [\mathcal{O} \cap \mathcal{V}] \subseteq \mathcal{O} \sim \overline{\mathcal{U}}.$$

Hence, by the monotonicity of outer-measure,

$$\mu(\mathcal{O} \sim \overline{\mathcal{U}}) \geq \mu^*(\mathcal{O} \sim \mathcal{V}).$$

Therefore,

$$\begin{aligned} \mu(\mathcal{O}) &\geq \mu(\mathcal{U}) + \mu(\mathcal{O} \sim \overline{\mathcal{U}}) \\ &\geq \mu(\mathcal{O} \cap \mathcal{V}) - \epsilon + \mu(\mathcal{O} \sim \overline{\mathcal{U}}) \\ &\geq \mu(\mathcal{O} \cap \mathcal{V}) - \epsilon + \mu^*(\mathcal{O} \sim \mathcal{V}). \end{aligned}$$

We have established (8). The proof is complete. □

Definition *Let $(X, \mathcal{T})$ be a topological space. We call a measure μ on the Borel σ-algebra $\mathcal{B}(X)$ a* **Borel measure** *provided every compact subset of X has finite measure. A Borel measure μ is called a* **Radon measure** *provided*

(i) (Outer Regularity) for each Borel subset E of X,

$$\mu(E) = \inf\{\mu(\mathcal{U}) \mid \mathcal{U} \text{ a neighborhood of } E\};$$

(ii) (Inner Regularity) for each open subset $\mathcal{O}$ of X,

$$\mu(\mathcal{O}) = \sup\{\mu(K) \mid K \text{ a compact subset of } \mathcal{O}\}.$$

We proved that the restriction to the Borel sets of Lebesgue measure on a Euclidean space $\mathbf{R}^n$ is a Radon measure. A Dirac delta measure on a topological space is a Radon measure.

While property (7) is sufficient in order for a premeasure $\mu: \mathcal{T} \to [0, \infty]$ to be extended by the measure $\mu^*: \mathcal{B} \to [0, \infty]$, in order that this extension be a Radon measure it is necessary, if X is locally compact, Hausdorff space, that μ be what we now name a Radon premeasure (see Problem 35).

Definition *A premeasure $\mu: \mathcal{T} \to [0, \infty]$ is called a* **Radon premeasure** *provided that it possesses the following two properties:*

(i) for each open set $\mathcal{U}$ that has compact closure, $\mu(\mathcal{U}) < \infty$;

(ii) for each open set $\mathcal{O}$,

$$\mu(\mathcal{O}) = \sup\left\{\mu(\mathcal{U}) \;\middle|\; \mathcal{U} \text{ open and } \overline{\mathcal{U}} \text{ a compact subset of } \mathcal{O}\right\}.$$

In earlier chapters, we called a Borel measure regular provided that it possesses properties (i) and (ii). In Chapter 10, we proved that every finite Borel measure on R^n is regular, while, in Chapter 13, we proved Ulam's Regularity Theorem, according to which a finite Borel measure on a complete, separable metric space is regular.

Theorem 10 *Let $(X, \mathcal{T})$ be a locally compact, Hausdorff space and $\mu: \mathcal{T} \to [0, \infty]$ a Radon premeasure. Then the restriction to the Borel σ-algebra $\mathcal{B}(X)$ of the Carathéodory outer-measure μ^* induced by μ is a Radon measure that extends μ.*

Proof A compact subset of the Hausdorff space X is closed, and hence assumption (ii) implies property (7). According to Proposition 9, the set-function $\mu^*: \mathcal{B}(X) \to [0, \infty]$ is a measure that extends μ. Assumption (i) and the locally compact separation property possessed by X imply that if K is compact, then $\mu^*(K) < \infty$. Therefore $\mu^*: \mathcal{B}(X) \to [0, \infty]$ is a Borel measure. Since μ is a premeasure, Lemma 8 tells us that every subset of X and, in particular, every Borel subset of X, is outer regular with respect to μ^*. It remains only to establish the inner regularity of every open set with respect to μ^*. However, this follows from assumption (ii) and the monotonicity of μ^*. $\square$

The natural functions on a topological space are the continuous ones. Of course, every continuous function on a topological space X is measurable with respect to the Borel σ-algebra $\mathcal{B}(X)$. For Lebesgue measure on $\mathbf{R}^n$, we proved Lusin's Theorem, which made precise J. E. Littlewood's second principle: a measurable function is "nearly continuous." We leave it as an exercise to use inner regularity (see Problem 36), together with Theorem 6, as was done in the case of Lebesgue measure, in order to prove the following general version of Lusin's Theorem.

Lusin's Theorem *Let X be a locally compact, Hausdorff space, $\mu: \mathcal{B}(X) \to [0, \infty)$ a Radon measure, and $f: X \to \mathbf{R}$ a Borel measurable function that vanishes outside of a set of finite measure. Then for each $\epsilon > 0$, there is a Borel subset X_0 of X and a function $g \in C_c(X)$ for which*

$$f = g \text{ on } X_0 \text{ and } \mu(X \sim X_0) < \epsilon.$$

PROBLEMS

27. Let $(X, \mathcal{T})$ be a Hausdorff topological space. Show that $\mathcal{T}$ is a semi-ring if and only if $\mathcal{T}$ is the discrete topology.
28. (Tyagi) Let $(X, \mathcal{T})$ be a topological space and $\mu: \mathcal{T} \to [0, \infty]$ a premeasure. Assume that if $\mathcal{O}$ is open and $\mu(\mathcal{O}) < \infty$, then $\mu(\text{bd}\, \mathcal{O}) = 0$. Show that every open set is μ^*-measurable.
29. Show that the restriction of Lebesgue measure on the real line to the Borel σ-algebra is a Radon measure.
30. Show that the restriction of Lebesgue measure on the Euclidean space $\mathbf{R}^n$ to the Borel σ-algebra is a Radon measure.
31. Show that a Dirac delta measure on a topological space is a Radon measure.
32. Let X be an uncountable set with the discrete topology and $\{x_k\}_{1 \le k < \infty}$ a countable subset of X. For $E \subseteq X$, define
$$\mu(E) = \sum_{\{n \mid x_n \in E\}} 2^{-n}.$$
Show that $2^X = \mathcal{B}(X)$ and $\mu: \mathcal{B}(X) \to$ is a Radon measure.

33. Show that the sum of two Radon measures also is Radon.
34. Let μ and ν be Borel measures on $\mathcal{B}(X)$, where X is a compact topological space, and suppose that μ is absolutely continuous with respect to ν. If ν is Radon show that μ also is Radon.
35. Let $(X, \mathcal{T})$ be a locally compact, Hausdorff space and μ: $\mathcal{T} \to [0, \infty]$ a premeasure for which the restriction to $\mathcal{B}(X)$ of μ^* is a Radon measure. Show that μ is a Radon premeasure.
36. Let X be a locally compact, Hausdorff space and μ: $\mathcal{B}(X) \to [0, \infty]$ a Radon measure. Show that any Borel set E of finite measure is inner regular in the sense that

$$\mu(E) = \sup \{\mu(K) \mid K \subseteq E, K \text{ compact}\}.$$

Conclude that if μ is σ-finite, then every Borel set is inner regular.
37. Let X be a topological space, μ: $\mathcal{B}(X) \to [0, \infty]$ a σ-finite Radon measure, and $E \subseteq X$ a Borel set. Show that there is a G_δ subset A of X and an F_σ subset B of X for which

$$A \subseteq E \subseteq B \text{ and } \mu(B \sim E) = \mu(E \sim A) = 0.$$

38. For a metric space X, show that $\mathcal{B}(X)$ is the smallest σ-algebra with respect to which all of the continuous real-valued functions on X are measurable.

21.4 THE REPRESENTATION OF POSITIVE LINEAR FUNCTIONALS ON $C_c(X)$: THE RIESZ-MARKOV THEOREM

Let X be a topological space. A real-valued functional ψ on $C(X)$ is said to be **monotone** provided $\psi(g) \geq \psi(h)$ if $g \geq h$ on X, and said to be **positive** provided $\psi(f) \geq 0$ if $f \geq 0$ on X. If ψ is linear, $\psi(g - h) = \psi(g) - \psi(h)$ and, of course, if $f = g - h$, then $f \geq 0$ on X if and only if $g \geq h$ on X. Therefore, for a linear functional, positivity is the same as monotonicity.

Proposition 11 *Let X be a locally compact, Hausdorff space and μ_1, μ_2 be Radon measures on $\mathcal{B}(X)$ for which*

$$\int_X f \, d\mu_1 = \int_X f \, d\mu_2 \text{ for all } f \in C_c(X).$$

Then $\mu_1 = \mu_2$.

Proof By the outer regularity of every Borel set, these measures are equal if and only if they agree on open sets and therefore, by the inner regularity of every open set, if and only if they agree on compact sets. Let K be a compact subset of X. We will show that

$$\mu_1(K) = \mu_2(K).$$

Let $\epsilon > 0$. By the outer regularity of both μ_1 and μ_1 and the excision and monotonicity properties of measure, there is a neighborhood $\mathcal{O}$ of K for which

$$\mu_1(\mathcal{O} \sim K) < \epsilon/2 \text{ and } \mu_2(\mathcal{O} \sim K) < \epsilon/2. \tag{9}$$

Since X is locally compact and Hausdorff, it has the locally compact extension property. Hence there is a function $f \in C_c(X)$ for which $0 \le f \le 1$ on X, $f = 0$ on $X \sim \mathcal{O}$, and $f = 1$ on K. For $i = 1, 2$,

$$\int_X f\, d\mu_i = \int_{\mathcal{O}} f\, d\mu_i = \int_{\mathcal{O} \sim K} f\, d\mu_i + \int_K f\, d\mu_i = \int_{X \sim \mathcal{O}} f\, d\mu_i + \mu_i(K).$$

By assumption,

$$\int_X f\, d\mu_1 = \int_X f\, d\mu_2.$$

Therefore,

$$\mu_1(K) - \mu_2(K) = \int_{\mathcal{O} \sim K} f\, d\mu_2 - \int_{\mathcal{O} \sim K} f\, d\mu_1.$$

But $0 \le f \le 1$ on X and we have the measure estimates (9). Hence, by the monotonicity of integration,

$$|\mu_1(K) - \mu_2(K)| \le \int_{\mathcal{O} \sim K} f\, d\mu_2 + \int_{\mathcal{O} \sim K} f\, d\mu_1 < \epsilon.$$

Therefore, $\mu_1(K) = \mu_2(K)$. The proof is complete. $\square$

The Riesz-Markov Theorem *Let X be a locally compact, Hausdorff space and I a positive linear functional on $C_c(X)$. Then there is a unique Radon measure $\widehat{\mu}$ on $\mathcal{B}(X)$, the Borel σ-algebra associated with the topology on X, for which*

$$I(f) = \int_X f\, d\widehat{\mu} \text{ for all } f \in C_c(X). \tag{10}$$

Proof[2] Define $\mu(\emptyset) = 0$. For each non-empty open subset $\mathcal{O}$ of X, define

$$\mu(\mathcal{O}) = \sup \left\{ I(f) \mid f \in C_c(X),\ 0 \le f \le 1, \operatorname{supp} f \subseteq \mathcal{O} \right\}.$$

Our strategy is to first show that μ is a Radon premeasure. Hence, by Theorem 10, if we denote by $\widehat{\mu}$ the restriction to the Borel sets of the outer-measure induced by μ, then $\widehat{\mu}$ is a Radon measure that extends μ. We then show that integration with respect to $\widehat{\mu}$ represents the functional I. The uniqueness assertion is a consequence of the preceding proposition.

Since I is positive, μ takes values in $[0, \infty]$. We begin by showing that μ is a premeasure. To establish countable monotonicity, let $\{\mathcal{O}_k\}_{k=1}^{\infty}$ be a collection of open subsets of X that covers the open set $\mathcal{O}$. Let f be a function in $C_c(X)$ with $0 \le f \le 1$ and $\operatorname{supp} f \subseteq \mathcal{O}$. Define $K = \operatorname{supp} f$. By the compactness of K, there is a finite collection $\{\mathcal{O}_k\}_{k=1}^{n}$ that also

[2]To prove the theorem we need to determine the measure of a set by knowing the values of the "integrals" of certain functions. It is an instructive exercise to show that if μ is Lebesgue measure of $\mathbf{R}$ and $I = (a, b)$, an open, bounded interval, then

$$\mu(I) = b - a = \sup \left\{ \int_{\mathbf{R}} f\, d\mu \;\middle|\; f \in C_c(\mathbf{R}),\ 0 \le f \le 1, \operatorname{supp} f \subseteq I \right\}.$$

covers K. According to Proposition 5, there is a partition of unity subordinate to this finite cover, that is, there are functions $\varphi_1, \ldots, \varphi_n$ in $C_c(X)$ such that

$$\sum_{i=1}^{n} \varphi_i = 1 \text{ on } K \text{ and, for } 1 \leq k \leq n, 0 \leq \varphi_k \leq 1 \text{ on } X \text{ and } \operatorname{supp} \varphi_k \subseteq \mathcal{O}_k.$$

Then, since $f = 0$ on $X \sim K$,

$$f = \sum_{k=1}^{n} \varphi_k \cdot f \text{ on } X \text{ and, for } 1 \leq k \leq n,\ 0 \leq f \cdot \varphi_k \leq 1 \text{ and } \operatorname{supp}(\varphi_k \cdot f) \subseteq \mathcal{O}_k.$$

By the linearity of the functional I and the definition of μ,

$$I(f) = I\left(\sum_{k=1}^{n} \varphi_k \cdot f\right) = \sum_{k=1}^{n} I(\varphi_k \cdot f) \leq \sum_{k=1}^{n} \mu(\mathcal{O}_k) \leq \sum_{k=1}^{\infty} \mu(\mathcal{O}_k).$$

Take the supremum over all such f to conclude that

$$\mu(\mathcal{O}) \leq \sum_{k=1}^{\infty} \mu(\mathcal{O}_k).$$

Therefore, μ is countably monotone.

Since μ is countably monotone and, by definition, $\mu(\emptyset) = 0$, μ is finitely monotone. Therefore, to show that μ is finitely additive it suffices, using an induction argument, to let $\mathcal{O} = \mathcal{O}_1 \cup \mathcal{O}_2$ be the disjoint union of two open sets and show that

$$\mu(\mathcal{O}) \geq \mu(\mathcal{O}_1) + \mu(\mathcal{O}_2). \tag{11}$$

Let the functions f_1, f_2 belong to $C_c(X)$ and have the property that for $1 \leq k \leq 2$,

$$0 \leq f_k \leq 1 \text{ and } \operatorname{supp} f_k \subseteq \mathcal{O}_k.$$

Then the function $f = f_1 + f_2$ has support contained in $\mathcal{O}$, and, since $\mathcal{O}_1$ and $\mathcal{O}_2$ are disjoint, $0 \leq f \leq 1$. Again using the linearity of I and the definition of μ, we have

$$I(f_1) + I(f_2) = I(f) \leq \mu(\mathcal{O}).$$

If we first take the supremum over all such f_1 and then over all such f_2 we have

$$\mu(\mathcal{O}_1) + \mu(\mathcal{O}_2) \leq \mu(\mathcal{O}).$$

Hence we have established (11) and thus the finite additivity of μ. Therefore μ is a premeasure.

We next establish the inner regularity property of a Radon premeasure. Let $\mathcal{O}$ be open. Suppose $\mu(\mathcal{O}) < \infty$. We leave the case $\mu(\mathcal{O}) = \infty$ as an exercise (see Problem 42). Let $\epsilon > 0$. We have to establish the existence of an open set $\mathcal{U}$ that has compact closure contained in

$\mathcal{O}$ and $\mu(\mathcal{U}) > \mu(\mathcal{O}) - \epsilon$. Indeed, by the definition of μ, there is a function $f_\epsilon \in C_c(X)$ that has support contained in $\mathcal{O}$ and for which $I(f_\epsilon) > \mu(\mathcal{O}) - \epsilon$. Let $K = \operatorname{supp} f$. But X is locally compact and Hausdorff and therefore has the locally compact separation property. Choose $\mathcal{U}$ to be a neighborhood of K that has compact closure contained in $\mathcal{O}$. Then

$$\mu(\mathcal{U}) \geq I(f_\epsilon) > \mu(\mathcal{O}) - \epsilon.$$

It remains only to show that if $\mathcal{O}$ is an open set of compact closure, then $\mu(\overline{\mathcal{O}}) < \infty$. But X is locally compact and Hausdorff and therefore has the locally compact extension property. Choose a function in $C_c(X)$ that takes the constant value 1 on $\overline{\mathcal{O}}$. Thus, since I is positive, $\mu(\mathcal{O}) \leq I(f) < \infty$. This concludes the proof that μ is a Radon premeasure.

Theorem 10 tells us that the Carathéodory measure induced by μ restricts to a Radon measure $\widehat{\mu}$ on $\mathcal{B}(X)$ that extends μ. We claim that (10) holds for $\widehat{\mu}$. The first observation is that a continuous function is measurable with respect to any Borel measure and that a continuous function of compact support is integrable with respect to such a measure, since compact sets have finite measure and continuous functions on compact sets are bounded. By the linearity of I and of integration with respect to a given measure and the representation of each $f \in C_c(X)$ as the difference of non-negative functions in $C_c(X)$, to establish (10) it suffices to verify that

$$I(f) = \int_X f \, d\widehat{\mu} \text{ for all } f \in C_c(X) \text{ for which } 0 \leq f \leq 1. \tag{12}$$

Let $f \in C_c(X)$ with $0 \leq f \leq 1$. Fix a natural number n. For $1 \leq k \leq n$, define the function $\varphi_k \colon X \to [0, 1]$ as follows:

$$\varphi_k(x) = \begin{cases} 1 & \text{if } f(x) > \frac{k}{n} \\ nf(x) - (k-1) & \text{if } \frac{k-1}{n} < f(x) \leq \frac{k}{n} \\ 0 & \text{if } f(x) \leq \frac{k-1}{n}. \end{cases}$$

The function φ_k is continuous. We claim that

$$f = \frac{1}{n} \sum_{k=1}^{n} \varphi_k \text{ on } X. \tag{13}$$

To verify this claim, let $x \in X$. If $f(x) = 0$, then $\varphi_k(x) = 0$ for $1 \leq k \leq n$, and therefore (13) holds. Otherwise, choose k_0 such that $1 \leq k_0 \leq n$ and $\frac{k_0-1}{n} < f(x) \leq \frac{k_0}{n}$. Then

$$\varphi_k(x) = \begin{cases} 1 & \text{if } 1 \leq k \leq k_0 - 1 \\ nf(x) - (k_0 - 1) & \text{if } k = k_0 \\ 0 & \text{if } k_0 < k \leq n. \end{cases}$$

Thus (13) holds.

Since X is locally compact and Hausdorff, it has the locally compact separation property. Therefore, since $\operatorname{supp} f$ is compact, we may choose an open set $\mathcal{O}$ of compact closure for which $\operatorname{supp} f \subseteq \mathcal{O}$. Define $\mathcal{O}_0 = \mathcal{O}, \mathcal{O}_{n+1} = \emptyset$ and, for $1 \leq k \leq n$, define

$$\mathcal{O}_k = \left\{ x \in \mathcal{O} \;\middle|\; f(x) > \frac{k-1}{n} \right\}.$$

By construction,

$$\operatorname{supp}\varphi_k \subseteq \overline{\mathcal{O}}_k \subseteq \mathcal{O}_{k-1} \text{ and } \varphi_k = 1 \text{ on } \mathcal{O}_{k+1}.$$

Therefore, by the monotonicity of I and of integration with respect to $\widehat{\mu}$, and the definition of μ,

$$\mu(\mathcal{O}_{k+1}) \leq I(\varphi_k) \leq \mu(\mathcal{O}_{k-1}) = \mu(\mathcal{O}_k) + [\mu(\mathcal{O}_{k-1}) - \mu(\mathcal{O}_k)]$$

and

$$\mu(\mathcal{O}_{k+1}) \leq \int_X \varphi_k \, d\widehat{\mu} \leq \mu(\mathcal{O}_{k-1}) = \mu(\mathcal{O}_k) + [\mu(\mathcal{O}_{k-1}) - \mu(\mathcal{O}_k)].$$

However,

$$\mu(\mathcal{O}) = \mu(\mathcal{O}_0) \geq \mu(\mathcal{O}_1) \geq \cdots \geq \mu(\mathcal{O}_{n-1}) \geq \cdots \geq \mu(\mathcal{O}_n) = 0.$$

Therefore, since the compactness of $\overline{\mathcal{O}}$ implies the finiteness of $\mu(\mathcal{O})$, we have

$$-\mu(\mathcal{O}) - \mu(\mathcal{O}) \leq \sum_{k=1}^{n} \left[I(\varphi_k) - \int_X \varphi_k \, d\widehat{\mu} \right] \leq \mu(\mathcal{O}) + \mu(\mathcal{O}).$$

Divide this inequality by n, use the linearity of I and of integration, together with (13) to obtain

$$\left| I(f) - \int_X f d\widehat{\mu} \right| \leq \frac{4}{n} \mu(\mathcal{O}).$$

This holds for all natural numbers n and $\mu(\mathcal{O}) < \infty$. Hence (10) holds. □

PROBLEMS

39. Let X be a locally compact, Hausdorff space and $C_0(X)$ the space of all uniform limits of functions in $C_c(X)$.
    (i) Show that a continuous real-valued function f on X belongs to $C_0(X)$ if and only if for each $\alpha > 0$ the set $\{x \in X \mid |f(x)| \geq \alpha\}$ is compact.
    (ii) Let X^* be the one-point compactification of X. Show that $C_0(X)$ consists precisely of the restrictions to X of those functions in $C(X^*)$ that vanish at the point at infinity.
40. Let X be an uncountable set with the discrete topology.
    (i) What is $C_c(X)$?
    (ii) What are the Borel subsets of X?
    (iii) Let X^* be the one-point compactification of X. What is $C(X^*)$?
    (iv) What are the Borel subsets of X^*?
    (v) Show that there is a Borel measure μ on X^* such that $\mu(X^*) = 1$ and $\int_X f \, d\mu = 0$ for each f in $C_c(X)$.
41. Let X and Y be two locally compact, Hausdorff spaces.
    (i) Show that each $f \in C_c(X \times Y)$ is the limit of sums of the form

$$\sum_{i=1}^{n} \varphi_i(x)\psi_i(y)$$

where $\varphi_i \in C_c(X)$ and $\psi_i = C_c(Y)$. (The Stone-Weierstrass Theorem is useful.)

(ii) Show that $\mathcal{B}(X \times Y) \subseteq \mathcal{B}(X) \times \mathcal{B}(Y)$.

(iii) Show that $\mathcal{B}(X \times Y) = \mathcal{B}(X) \times \mathcal{B}(Y)$ if and only if X or Y is the union of a countable collection of compact subsets.

42. In the proof of the Riesz-Markov Theorem, establish inner regularity in the case in which $\mu(\mathcal{O}) = \infty$.

43. Let $k(x, y)$ be a bounded Borel measurable function on $X \times Y$, and let μ and ν be Radon measures on X and Y.

(i) Show that

$$\begin{aligned} \iint_{X \times Y} \varphi(x) k(x, y) \psi(y) d(\mu \times \nu) &= \int_Y \left[\int_X \varphi(x) k(x, y) d\mu \right] \psi(y) d\nu \\ &= \int_X \varphi(x) \left[\int_Y k(x, y) \psi(y) d\nu \right] d\mu \end{aligned}$$

for all $\varphi \in C_c(X)$ and $\psi \in C_c(Y)$.

(ii) If the integral in (i) is zero for all φ and ψ in $C_c(X)$ and $C_c(Y)$, show that then $k = 0$ almost everywhere $[\mu \times \nu]$.

44. Let X be a compact, Hausdorff space and μ a Borel measure on $\mathcal{B}(X)$. Show that there is a constant $c > 0$ such that

$$\left| \int_X f \, d\mu \right| \leq c \|f\|_{\max} \text{ for all } f \in C(X).$$

21.5 THE REPRESENTATION OF THE DUAL OF $C(X)$: THE RIESZ-KAKUTANI THEOREM

In the preceding section, we considered the linear space $C_c(X)$, without placing a norm on this space, and examined positive linear functionals on this space, in the absence of continuity. Let X be a compact, Hausdorff space and $C(X) = C_c(X)$ the space of real-valued continuous functions on X. We now consider $C(X)$ as a normed linear space with the maximum norm and characterize the continuous linear functionals on $C(X)$. First observe that each positive linear functional is continuous, that is, is bounded. Indeed, if L is a positive linear functional on $C(X)$ and $f \in C(X)$ with $\|f\| \leq 1$, then $-1 \leq f \leq 1$ on X and hence, by the homogeneity and positivity of $L, -L(1) \leq L(f) \leq L(1)$, that is, $|L(f)| \leq L(1)$. Therefore, L is bounded and the norm of the functional L equals the value of L at the constant function with value 1, that is,

$$\|L\| = L(1).$$

According to the Jordan Decomposition Theorem, a signed measure may be expressed as the difference of two mutually singular measures. Therefore, integration with respect to a signed measure may be expressed as the difference of positive linear functionals. The following proposition is a variation, for general continuous linear functionals on $C(X)$, of this decomposition property.

Proposition 12 *Let X be a compact, Hausdorff space and $C(X)$ the linear space of continuous real-valued functions on X, normed by the maximum norm. Then for each continuous linear functional L on $C(X)$, there are two positive linear functionals L_+ and L_- on $C(X)$ for which*

$$L = L_+ - L_- \text{ and } \|L\| = L_+(1) + L_-(1).$$

Proof For $f \in C(X)$ such that $f \geq 0$, define

$$L_+(f) = \sup_{0 \leq \psi \leq f} L(\psi).$$

Since the functional L is bounded, $L_+(f)$ is a real number. We begin by showing that for $f \geq 0, g \geq 0,$ and $c \geq 0$,

$$L_+(cf) = cL_+(f) \text{ and } L_+(f+g) = L_+(f) + L_+(g).$$

Indeed, by the positive homogeneity of L, $L_+(cf) = cL_+(f)$ for $c \geq 0$. Let f and g be two non-negative functions in $C(X)$. If $0 \leq \varphi \leq f$ and $0 \leq \psi \leq g$, then $0 \leq \varphi + \psi \leq f + g$ and so

$$L(\varphi) + L(\psi) = L(\varphi + \psi) \leq L_+(f+g).$$

Taking suprema, first over all such φ and then over all such ψ, we obtain

$$L_+(f) + L_+(g) \leq L_+(f+g).$$

On the other hand, if $0 \leq \psi \leq f+g$, then $0 \leq \min\{\psi, f\} \leq f$ and thus $0 \leq \psi - \min\{\psi, f\} \leq g$, and therefore

$$\begin{aligned} L(\psi) &= L(\min\{\psi, f\}) + L(\psi - [\min\{\psi, f\}]) \\ &\leq L_+(f) + L_+(g). \end{aligned}$$

Taking the supremum over all such ψ, we get

$$L_+(f+g) \leq L_+(f) + L_+(g).$$

Therefore,

$$L_+(f+g) = L_+(f) + L_+(g).$$

Let f be an arbitrary function in $C(X)$, and let M and N be two non-negative constants for which $f + M$ and $f + N$ are non-negative. Then

$$L_+(f + M + N) = L_+(f + M) + L_+(N) = L_+(f + N) + L_+(M).$$

Hence

$$L_+(f + M) - L_+(M) = L_+(f + N) - L_+(N).$$

Thus the value of $L_+(f + M) - L_+(M)$ is independent of the choice of M, and we define $L_+(f)$ to be this value.

Clearly, $L_+\colon C(X) \to \mathbf{R}$ is positive and we claim that it is linear. Indeed, it is clear that $L_+(f+g) = L_+(f) + L_+(g)$. We also have $L_+(cf) = cL_+(f)$ for $c \geq 0$. On the other hand,

$L_+(-f)+L_+(f) = L_+(0) = 0$, so that we have $L_+(-f) = -L_+(f)$. Thus $L_+(cf) = cL_+(f)$ for all c. Therefore L_+ is linear.

Define $L_- = L_+ - L$. Then L_- is a linear functional on $C(X)$ and it is positive since, by the definition of L_+, $L(f) \leq L_+(f)$ for $f \geq 0$. We have expressed L as the difference, $L_+ - L_-$, of two positive linear functionals on $C(X)$.

We always have $\|L\| \leq \|L_+\| + \|L_-\| = L_+(1) + L_-(1)$. To establish the inequality in the opposite direction, let φ be any function in $C(X)$ for which $0 \leq \varphi \leq 1$. Then $\|2\varphi - 1\| \leq 1$ and hence

$$\|L\| \geq L(2\varphi - 1) = 2L(\varphi) - L(1).$$

Taking the supremum over all such φ, we have

$$\|L\| \geq 2L_+(1) - L(1) = L_+(1) + L_-(1).$$

Hence $\|L\| = L_+(1) + L_-(1)$. □

For a compact topological space X, we call a signed measure on $\mathcal{B}(X)$ a **signed Radon measure** provided it is the difference of Radon measures. We denote by $\mathcal{R}adon(X)$ the normed linear space of signed Radon measures on X with the norm of $\nu \in \mathcal{R}adon(X)$ given by its total variation $||\nu||_{\text{var}}$, which, we recall, may be expressed as

$$||\nu||_{\text{var}} = \nu^+(X) + \nu^-(X),$$

where $\nu = \nu^+ - \nu^-$ is the Jordan decomposition of ν. We leave it as an exercise to show that $|| \cdot ||_{\text{var}}$ is a norm on the linear space of signed Radon measures.

The Riesz-Kakutani Representation Theorem *Let X be a compact, Hausdorff space and $C(X)$ the linear space of continuous real-valued functions on X, normed by the maximum norm. Define the operator T: $\mathcal{R}adon(X) \to [C(X)]^*$ by setting, for $\nu \in \mathcal{R}adon(X)$,*

$$T_\nu(f) = \int_X f\, d\nu \text{ for all } f \text{ in } C(X).$$

Then T is a linear isometric isomorphism of $\mathcal{R}adon(X)$ onto $[C(X)]^*$.

Proof The operator T is linear and we infer from Proposition 11 that it is one-to-one. To verify that T is onto, let L be a bounded linear functional on $C(X)$. By the preceding proposition, choose positive linear functionals L_1 and L_2 on $C(X)$ for which $L = L_1 - L_2$ and $\|L\| = L_+(1) + L_-(1)$. According to the Riesz-Markov Theorem, there are Radon measures on X, μ_1 and μ_2, for which

$$L_1(f) = \int_X f\, d\mu_1 \text{ and } L_2(f) = \int_X f\, d\mu_2 \text{ for all } f \in C(X).$$

Define $\mu = \mu_1 - \mu_2$. Thus μ is a signed Radon measure for which $L = T_\mu$. Hence T is onto. Moreover, since

$$\|L\| = L_1(1) + L_2(1) = \mu_1(X) + \mu_2(X) = |\mu|(X),$$

$\|L\| = \|\mu\|_{\text{var}}$. Therefore, T is an isometric isomorphism. □

Corollary 13 *Let X be a compact, Hausdorff space and K^* a bounded subset of $\mathcal{R}adon(X)$ that is weak-$*$ closed. Then K^* is weak-$*$ compact. If, furthermore, K^* is convex, then K^* is the weak-$*$ closed convex hull of its extreme points.*

Proof Alaoglu's Theorem tells us that each closed ball in $[C(X)]^*$ is weak-$*$ compact. A closed subset of a compact topological space is compact. Thus K^* is weak-$*$ compact. We infer from the Krein-Milman Theorem, applied to the locally convex topological space comprising $[C(X)]^*$ with its weak-$*$ topology, that, if K^* is convex, then K^* is the weak-$*$ closed convex hull of its extreme points. □

Theorem 14 *Let X be a compact metric space and $\{\mu_n\colon \mathcal{B}(X) \to [0, \infty)\}$ a sequence of Borel measures for which the sequence $\{\mu_n(X)\}$ is bounded. Then there is a subsequence $\{\mu_{n_k}\}$ and a Borel measure μ for which*

$$\lim_{k\to\infty} \int_X f\, d\mu_{n_k} = \int_X f\, d\mu \text{ for all } f \in C(X).$$

Proof A compact metric space is both complete and separable. Therefore, by Ulam's Regularity Theorem, every finite measure on such a space is a Radon measure. Consequently, by the Riesz-Kakutani Representation Theorem, all the bounded linear functionals on $C(X)$ are given by integration against finite signed Borel measures. Now, X being a compact metric space, it follows from Borsuk's Theorem that the space $C(X)$, normed by the maximum norm, is separable. Therefore, the weak-$*$ sequential compactness conclusion now follows from Helly's Theorem. □

In 1909, Frigyes Riesz proved the representation theorem for the dual of $C[a, b]$, in the following form: For each bounded linear functional L on $C[a, b]$, there is a function $g\colon [a, b] \to \mathbf{R}$ of bounded variation for which

$$L(f) = \int_a^b f(x)\, dg(x) \text{ for all } f \in C[a, b]:$$

the integral is in the sense of Stieltjes. This integral is defined in the case that f is continuous and g is of bounded variation. Moreover, for a fixed function of bounded variation g, as a functional of $f \in C[a,b]$, the Stieltjes integral is a bounded linear functional on $C[a, b]$[3]. According to Jordan's Theorem, any function of bounded variation is the difference of increasing functions. Therefore, it is interesting, given an increasing function g on $[a, b]$, to identify, with respect to the properties of g, the unique Borel measure μ for which

$$\int_a^b f(x)\, dg(x) = \int_{[a,\, b]} f\, d\mu \text{ for all } f \in C[a, b]. \tag{14}$$

For a closed, bounded interval $[a, b]$, let $\mathcal{S}$ be the semi-ring of subsets of $[a, b]$ comprising the singleton set $\{a\}$ together with subintervals of the form $(c, d]$. Then $\mathcal{B}[a, b]$ is the smallest σ-algebra containing $\mathcal{S}$. We infer from the uniqueness assertion in the

[3]Chapter 3 of Richard Wheeden and Antoni Zygmund's book Measure and Integral is devoted to the Stieltjes integral, which is sometimes called the Lebesgue-Stieltjes integral.

Carathéodory-Hahn Theorem that a Borel measure on $\mathcal{B}[a, b]$ is uniquely determined by its values on $\mathcal{S}$. Therefore the following proposition characterizes the Borel measure that represents Stieltjes integration against a given increasing function. For an increasing real-valued function on the closed, bounded interval $[a, b]$ we define, for $a < c < b$,

$$f(c^+) = \inf_{c < x \leq b} f(x) \text{ and } f(c^-) = \sup_{a \leq x < c} f(x).$$

Define $f(a^+)$ and $f(b^-)$ in the obvious manner, and set $f(a^-) = f(a)$, $f(b^+) = f(b)$. The function f is said to be continuous on the right at $x \in [a, b)$ provided $f(x) = f(x^+)$.

Proposition 15 *Let g be an increasing function on the closed, bounded interval $[a, b]$ and μ the unique Borel measure for which (14) holds. Then $\mu\{a\} = g(a^+) - g(a)$ and*

$$\mu(c, d] = g(d^+) - g(c^+) \text{ for all } (c, d] \subseteq (a, b]. \tag{15}$$

Proof We first verify that

$$\mu[c, b] = g(b) - g(c^-) \text{ for all } c \in (a, b]. \tag{16}$$

Fix a natural number n. The increasing function g is continuous except at a countable number of points in $[a, b]$. Choose a point $c_n \in (a, c)$ at which g is continuous and $c - c_n < 1/n$. Now choose a point $c_n' \in (a, c_n)$ at which g is continuous and $g(c_n) - g(c_n') < 1/n$. Construct a continuous function f_n on $[a, b]$ for which $0 \leq f_n \leq 1$ on $[a, b]$, $f_n = 1$ on $[c_n, b]$ and $f_n = 0$ on $[a, c_n']$. By the additivity over intervals property of the Stieltjes integral,

$$\int_a^b f_n(x)\, dg(x) = \int_{c_n'}^{c_n} f_n(x)\, dg(x) + [g(b) - g(c_n)].$$

By the additivity of integration with respect to μ over finite disjoint unions of Borel sets,

$$\int_{[a, b]} f_n\, d\mu = \int_{(c_n', c_n)} f_n\, d\mu + \mu[c_n, b].$$

Substitute $f = f_n$ in (14) to conclude that

$$\int_{c_n'}^{c_n} f_n(x)\, dg(x) + [g(b) - g(c_n)] = \int_{(c_n', c_n)} f_n\, d\mu + \mu[c_n, b]. \tag{17}$$

However, since $0 \leq f_n \leq 1$ on $[a, b]$,

$$\left| \int_{c_n'}^{c_n} f_n(x)\, dg(x) \right| \leq g(c_n) - g(c_n') < 1/n$$

and

$$\left| \int_{(c_n', c_n)} f_n\, d\mu \right| \leq \mu(c_n', c_n) \leq \mu(c_n', c).$$

Take the limit as $n \to \infty$ in (17) and use the continuity of measure to conclude that (16) holds. A similar argument shows that $\mu\{a\} = g(a^+) - g(a)$ and

$$\mu\{c\} = g(c^+) - g(c^-) \text{ for all } c \in (a, b). \tag{18}$$

Finally, we infer from (16), (18), and the finite additivity of μ that for $a < c < d \leq b$,

$$\mu(c, d] = \mu[c, b] - \mu[d, b] - \mu\{c\} + \mu\{d\} = g(d^+) - g(c^+).$$

The proof is complete. □

We have the following, slightly amended, version of Riesz's original representation theorem from 1909.

Theorem 16 (Riesz) *Let $[a, b]$ be a closed, bounded interval and $\mathcal{F}$ be the collection of real-valued functions on $[a, b]$ that are of bounded variation on $[a, b]$, continuous on the right on (a, b), and vanish at a. Then for each bounded linear functional ψ on $C[a, b]$, there is a unique function $g \in \mathcal{F}$ for which*

$$\psi(f) = \int_a^b f(x)\, dg(x) \textit{ for all } f \in C[a, b]. \tag{19}$$

Proof To establish existence, it suffices, by the Riesz-Markov Theorem, to do so for ψ a positive bounded linear functional on $C[a, b]$. For such a ψ, the Riesz-Kakutani Representation Theorem tells us there is a Borel measure μ for which

$$\psi(f) = \int_a^b f\, d\mu \text{ for all } f \in C[a, b].$$

Consider the increasing real-valued function g defined on $[a, b]$ by $g(a) = 0$ and $g(x) = \mu(a, x] + \mu\{a\}$ for $x \in (a, b]$. The functon g inherits continuity on the right at each point in (a, b) from the continuity of the measure μ. Therefore, $g \in \mathcal{F}$. It follows from Proposition 15 that

$$\int_a^b f\, d\widehat{\mu} = \int_a^b f(x)\, dg(x) \text{ for all } f \in C[a, b],$$

where $\widehat{\mu}$ is the unique Borel measure on $\mathcal{B}[a, b]$ for which

$$\widehat{\mu}(c, b] = g(b) - g(c^+) = g(b) - g(c) \text{ for all } c \in (a, b) \text{ and } \widehat{\mu}\{a\} = g(a^+) - g(a).$$

However, the measure μ has these properties. This completes the proof of existence.

To establish uniqueness, by Jordan's Theorem regarding the expression of a function of bounded variation as the difference of increasing functions, it suffices to let $g_1, g_2 \in \mathcal{F}$ be increasing functions which have the property that

$$\psi(f) = \int_a^b f(x)\, dg_1(x) = \int_a^b f(x)\, dg_2(x) \text{ for all } f \in C[a, b],$$

and show that $g_1 = g_2$. Take $f \equiv 1$ in this integral equality to conclude that

$$g_1(b) = g_1(b) - g_1(a) = g_2(b) - g_2(a) = g_2(b).$$

Let ψ be represented by integration against the Borel measure μ. We infer from Proposition 15 and the right continuity of g_1 and g_2 at each point in (a, b) that if x belongs to (a, b), then

$$g_1(b) - g_1(x) = \mu(x, b] = g_2(b) - g_2(x),$$

and hence $g_1(x) = g_2(x)$. Therefore $g_1 = g_2$ on $[a, b]$. $\square$

PROBLEMS

45. Let x_0 be a point in the compact, Hausdorff space X. Define $L(f) = f(x_0)$ for each $f \in C(X)$. Show that L is a bounded linear functional on $C(X)$. Find the signed Radon measure that represents L.

46. Let X be a compact, Hausdorff space and μ a Borel measure on $\mathcal{B}(X)$. Show that there is a Radon measure μ_0 for which

$$\int_X f\, d\mu = \int_X f\, d\mu_0 \text{ for all } f \text{ in } C(X).$$

47. Let g_1 and g_2 be two increasing functions on the closed, bounded interval $[a, b]$ that agree at the end-points. Show that

$$\int_a^b f(x)\, dg_1(x) = \int_a^b f(x)\, dg_2(x) \text{ for all } f \in C[a, b]$$

if and only if $g_1(x^+) = g_2(x^+)$ for all $a \leq x < b$.

48. Let X be a compact, Hausdorff space. Show that the Jordan Decomposition Theorem for signed Borel measures on $\mathcal{B}(X)$ follows from the Riesz-Kakutani Representation Theorem for the dual of $C(X)$ and Proposition 12.

49. What are the extreme points of the unit ball of the linear space of signed Radon measures $\mathcal{R}adon(X)$, where X is a compact, Hausdorff space?

50. Let X be a compact metric space. On the linear space of functions $\mathcal{F}$ defined in the statement of Theorem 16, define the norm of a function to be its total variation. Show that with this norm $\mathcal{F}$ is a Banach space.

51. (Alternate proof of the Stone-Weierstrass Theorem (de Branges)) Let $\mathcal{A}$ be an algebra of real-valued continuous functions on a compact space X that separates points and contains the constants. Let $\mathcal{A}^\perp$ be the set of signed Radon measures on X such that $|\mu|(X) \leq 1$ and $\int_X f d\mu = 0$ for all $f \in \mathcal{A}$.
    (i) Use the Hahn-Banach Theorem and the Riesz-Kakutani Representation Theorem to show that if $\mathcal{A}^\perp$ contains only the zero measure, then $\overline{\mathcal{A}} = C(X)$.
    (ii) Use the Krein-Milman Theorem and the weak-$*$ compactness of the unit ball in $\mathcal{R}adon(X)$ to show that if the identically zero measure is the only extreme point of $\mathcal{A}^\perp$, then $\mathcal{A}^\perp$ contains only the identically zero measure.
    (iii) Let μ be an extreme point of $\mathcal{A}^\perp$. Let $f \in \mathcal{A}$, with $0 \leq f \leq 1$. Define measures μ_1 and μ_2 by

$$\mu_1(E) = \int_E f\, d\mu \text{ and } \mu_2(E) = \int_E (1 - f)\, d\mu \text{ for } E \in \mathcal{B}(X).$$

    Show that μ_1 and μ_2 belong to $\mathcal{A}^\perp$ and, moreover, $\|\mu_1\| + \|\mu_2\| = \|\mu\|$, and $\mu_1 + \mu_2 = \mu$. Since μ is an extreme point, conclude that $\mu_1 = c\mu$ for some constant c.

(iv) Show that $f = c$ on the support of μ.

(v) Since $\mathcal{A}$ separates points, show that the support of μ can contain at most one point. Since $\int_X 1 d\mu = 0$, conclude that the support of μ is empty and hence μ is the zero measure.

21.6 REGULARITY PROPERTIES OF BAIRE MEASURES

Definition *Let X be a topological space. The* **Baire** *σ-algebra, which is denoted by $\mathcal{B}a(X)$, is defined to be the smallest σ-algebra of subsets of X for which the functions in $C_c(X)$ are measurable.*

Evidently

$$\mathcal{B}a(X) \subseteq \mathcal{B}(X).$$

There are compact, Hausdorff spaces for which this inclusion is strict (see Problem 56). The forthcoming Theorem 19 tells us that these two σ-algebras are equal if X is a compact metric space. A measure on $\mathcal{B}a(X)$ is called a **Baire measure** provided it is finite on compact sets. Given a Borel measure μ on the Borel σ-algebra $\mathcal{B}(X)$, we define μ_0 to be the restriction of μ to the Baire σ-algebra $\mathcal{B}a(X)$. Then μ_0 is a Baire measure. Moreover, each function $f \in C_c(X)$ is integrable over X with respect to μ_0 since it is measurable with respect to $\mathcal{B}a(X)$, bounded, and vanishes outside a set of finite measure. Since $\mathcal{B}a(X) \subseteq \mathcal{B}(X)$,

$$\int_X f\, d\mu = \int_X f\, d\mu_0 \text{ for all } f \in C_c(X). \tag{20}$$

We will establish regularity properties for Baire measures from which we obtain finer uniqueness properties for Baire representations than are possible for Borel representations in the Riesz-Markov and Riesz-Kakutani Representation Theorems.

Proposition 17 *Let X be a locally compact, Hausdorff space and μ_1 and μ_2 be two regular Baire measures on $\mathcal{B}a(X)$. Suppose*

$$\int_X f\, d\mu_1 = \int_X f\, d\mu_2 \textit{ for all } f \in C_c(X).$$

Then $\mu_1 = \mu_2$.

Proof The proof is exactly the same as the corresponding uniqueness result for integration with respect to Radon measures. □

Proposition 18 *Let X be a compact, Hausdorff space, $\mathcal{S}$ a σ-algebra of subsets of X, and $\mu: \mathcal{S} \to [0, \infty)$ a finite measure. Then the collection of sets in $\mathcal{S}$ that are regular with respect to μ is a σ-algebra.*

Proof Define $\mathcal{F}$ to be the collection of sets in $\mathcal{S}$ that are regular with respect to μ. Since X is compact and Hausdorff, a subset of X is open if and only if its complement in X is compact. Thus, since μ is finite, by the excision property of measure, a set belongs to $\mathcal{F}$ if and only if its complement in X belongs to $\mathcal{F}$. We leave it as an exercise to show that the union of two regular sets is regular. Therefore, the regular sets are closed with respect to the formation of finite unions, finite intersections, and relative complements. It remains to show that $\mathcal{F}$ is closed with respect to the formation of countable unions. Let $E = \bigcup_{n=1}^{\infty} E_n$, where each E_n is a regular set. By replacing each E_n by $E_n \sim \bigcup_{i=1}^{n-1} E_i$, we may suppose that the E_n's are disjoint. Let $\epsilon > 0$. For each n, by the outer regularity of E_n, we may choose a neighborhood $\mathcal{O}_n$ of E_n, which belongs to $\mathcal{S}$ and $\mu(\mathcal{O}_n) < \mu(E_n) + \epsilon/2^n$. Define $\mathcal{O} = \bigcup_{n=1}^{\infty} \mathcal{O}_n$. Then $\mathcal{O}$ is a neighborhood of E, $\mathcal{O}$ belongs to $\mathcal{S}$, and, since

$$\mathcal{O} \sim E \subseteq \bigcup_{n=1}^{\infty} [\mathcal{O}_n \sim K_n],$$

by the excision and countable monotonicity properties of the measure μ,

$$\mu(\mathcal{O}) - \mu(E) = \mu(\mathcal{O} \sim E) \leq \sum_{n=1}^{\infty} \mu(\mathcal{O}_n \sim E_n) < \epsilon.$$

Thus E is outer regular. A similar argument established inner regularity of E. This completes the proof. □

Theorem 19 *Let X be a compact, Hausdorff space in which every closed set is a G_δ set. Then the Borel σ-algebra equals the Baire σ-algebra and every Borel measure is regular. In particular, if X is a compact metric space, then the Borel σ-algebra equals the Baire σ-algebra and every Borel measure is regular.*

Proof To show that the Baire σ-algebra equals the Borel σ-algebra, it is necessary and sufficient to show that every closed set is a Baire set. Let K be a closed subset of X. Then K is compact and, by assumption, is a G_δ set. According to Proposition 4, there is a function $f \in C_c(X)$ for which $K = \{x \in X \mid f(x) = 1\}$. Since $f \in C_c(X)$, the set $\{x \in X \mid f(x) = 1\}$ is a Baire set.

Let μ be a Borel measure on $\mathcal{B}(X)$. The preceding proposition tells us that the collection of regular Borel sets is a σ-algebra. Therefore, to establish the regularity of $\mathcal{B}(X)$ it is necessary and sufficient to show that every closed set is regular with respect to the Borel σ-algebra. Let K be a closed subset of X. Then K is compact since X is compact and thus K is inner regular. Since K is a G_δ set and $\mu(X) < \infty$, we infer from the continuity of measure that K is outer regular with respect to the Borel σ-algebra.

To conclude the proof, assume that X is a compact metric space. Let K be a closed subset of X. We will show that K is a G_δ set. Let n be a natural number. Define $\mathcal{O}_n = \cup_{x \in K} B(x, 1/n)$. Then $\mathcal{O}_n$ is a neighborhood of the compact set K. According to the locally compact extension property, the function that takes the value 1 on K may be extended to a function $f_n \in C_c(X)$ that has support contained in $\mathcal{O}_n$. Define $\mathcal{U}_n = f_n^{-1}(-1/n, 1/n)$. Then $\mathcal{U}_n$ is an open Baire set. By the compactness of K, $K = \cap_{n=1}^{\infty} \mathcal{U}_n$. We infer from the continuity of measure that K is outer regular. □

In the preceding section, the Riesz-Markov Theorem was employed to show that if X is a compact metric space, then every Borel measure on $\mathcal{B}(X)$ is a Radon measure.

Proposition 20 *Let X be locally compact, Hausdorff space. The Baire σ-algebra $\mathcal{B}a(X)$ is the smallest σ-algebra that contains all the compact G_δ subsets of X, in the sense that it is contained in every σ-algebra that contains all such subsets.*

Proof Define $\mathcal{F}$ to be the smallest σ-algebra that contains all the compact G_δ sets. Let K be a compact G_δ set. According to Proposition 4 there is a function $f \in C_c(X)$ for which $K = \{x \in X \mid f(x) = 1\}$. Therefore K belongs to $\mathcal{B}a(X)$. Thus $\mathcal{F} \subseteq \mathcal{B}a(X)$. To establish the inclusion in the opposite direction we let f belong to $C_c(X)$ and show it is measurable with respect to the Baire σ-algebra. For a closed, bounded interval $[a, b]$ that does not contain 0, $f^{-1}[a, b]$ is compact and equal to $\cap_{n=1}^{\infty}(a - 1/n, b + 1/n)$ since f is continuous and has compact support. Since $\mathcal{F}$ is closed with respect to the formation of countable unions, $f^{-1}(I)$ also belongs to $\mathcal{F}$ if I is any interval that does not contain 0. Finally, since

$$f^{-1}\{0\} = X \sim [f^{-1}(-\infty, 0) \cup f^{-1}(0, \infty)],$$

it follows that the inverse image under f of any non-empty interval belongs to $\mathcal{F}$ and therefore f is measurable with respect to the Baire σ-algebra. □

Proposition 21 *Let X be a compact, Hausdorff space. Then every Baire measure on $\mathcal{B}a(X)$ is regular.*

Proof Let μ be a Baire measure on $\mathcal{B}a(X)$. Proposition 18 tells us that the collection of subsets of $\mathcal{B}a(X)$ that are regular with respect to μ is a σ-algebra. We infer from Proposition 20 that to prove the proposition it is sufficient to show that each compact G_δ subset K of X is regular. Let K be such a set. Clearly K is inner regular. Since $\mu(X) < \infty$ and K is a G_δ set, by the continuity of measure, K is outer regular. □

We have the following small improvement regarding uniqueness of the Riesz-Kakutani Representation Theorem.

Theorem 22 *Let X be a compact, Hausdorff space and $I\colon C(X) \to \mathbf{R}$ a bounded linear functional. Then there is a unique signed Baire measure μ for which*

$$I(f) = \int_X f\, d\mu \textit{ for all } f \in C_c(X).$$

Proof The Riesz-Kakutani Representation Theorem tells us that I is given by integration against a signed Radon measure μ' on the Borel subsets of X. Let μ be the restriction of μ' to the Baire σ-algebra. Then, arguing as we did in establishing (20), integration against μ represents I. The uniqueness assertion follows from Proposition 17 and the preceding regularity result. □

Definition *A topological space X is said to be* **σ-compact** *provided it is the countable union of compact subsets.*

Each Euclidean space $\mathbf{R}^n$ is σ-compact. The discrete topology on an uncountable space is not σ-compact. Our final goal of this chapter is to prove regularity for Baire measures on a locally compact, σ-compact, Hausdorff space. To that end we need the following three lemmas whose proof we leave as exercises.

Lemma 23 *Let X be a locally compact, Hausdorff space and $F \subseteq X$ a closed Baire set. Then for $A \subseteq F$,*

$$A \in \mathcal{B}a(X) \text{ if and only if } A \in \mathcal{B}a(F).$$

Lemma 24 *Let X be a locally compact, Hausdorff space and $E \subseteq X$ a Baire set that has compact closure. Then E is regular with respect to any Baire measure μ on $\mathcal{B}a(X)$.*

Lemma 25 *Let X be a locally compact, σ-compact Hausdorff space and $A \subseteq X$ a Baire set. Then A is the union of a countable collection of Baire sets, each of which has compact closure.*

Theorem 26 *Let X be a locally compact, σ-compact Hausdorff space. Then every Baire measure on $\mathcal{B}a(X)$ is regular.*

Proof Since X is locally compact, Lemma 24 tells us that any Baire set of compact closure is regular. Moreover, by the preceding lemma, since X is σ-compact, every Baire set is the union of a countable collection of Baire sets each of which has compact closure. Therefore, to complete the proof it is sufficient to show that the countable union of Baire sets, each of which has compact closure, is regular.

Let $E = \bigcup_{k=1}^{\infty} E_k$, where each E_k is a Baire set of compact closure. Since the Baire sets are an algebra, we may suppose that the E_k's are disjoint. Let $\epsilon > 0$. For each k, by the regularity of E_k, we may choose Baire sets K_k and $\mathcal{O}_k$, with K_k compact and $\mathcal{O}_k$ open, for which

$$K_k \subseteq E_k \subseteq \mathcal{O}_k$$

and

$$\mu(E_k) - \epsilon/2^k < \mu(K_k) \leq \mu(\mathcal{O}_k) < \mu(E_k) + \epsilon/2^k.$$

If $\mu(E) = \infty$ then, of course, E is outer regular. Moreover, since

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \mu(E) = \infty \text{ and } \mu\left(\bigcup_{k=1}^{\infty} K_k\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} K_k\right)$$

E contains compact Baire sets of the form $\bigcup_{k=1}^{N} K_k$, which have arbitrarily large measure and therefore E is inner regular.

Now suppose that $\mu(E) < \infty$. Then $\mathcal{O} = \bigcup_{k=1}^{\infty} \mathcal{O}_k$ is again an open Baire set and since

$$\mathcal{O} \sim E \subseteq \bigcup_{k=1}^{\infty} [\mathcal{O}_k \sim E_k],$$

by the countable monotonicity and excision properties of measure,

$$\mu(\mathcal{O}) - \mu(E) = \mu(\mathcal{O} \sim E) \leq \sum_{k=1}^{\infty} \mu(\mathcal{O}_k \sim E_k) < \epsilon.$$

Thus E is outer regular. To establish inner regularity observe that

$$\mu(E) = \lim_{N\to\infty} \sum_{k=1}^{N} \mu(E_k) \leq \lim_{N\to\infty} \sum_{k=1}^{N} \mu(K_k) + \epsilon$$

Thus E contains compact Baire subsets of the form $\bigcup_{k=1}^{N} K_k$, which have measure arbitrarily close to the measure of E. Therefore E is inner regular. □

We have the following small improvement regarding uniqueness of the Riesz-Markov Theorem for σ-compact spaces.

Theorem 27 *Let X be a locally compact, σ-compact Hausdorff space and I: $C_c(X) \to \mathbf{R}$ a positive linear functional. Then there is a unique Baire measure μ for which*

$$I(f) = \int_X f \, d\mu \text{ for all } f \in C_c(X).$$

The reader should be warned that standard terminology regarding sets and measures that are either Baire or Borel has not been established. Neither has the terminology regarding Radon measures. Some authors take the class of Baire sets to be the smallest σ-algebra for which all continuous real-valued functions on X are measurable. Others do not assume that every Borel or Baire measure is finite on every compact set. Others restrict the class of Borel sets to be the smallest σ-algebra that contains the compact sets. Authors (such as Halmos [Hal50]) who do measure theory on σ-rings rather than σ-algebras often take the Baire sets to be the smallest σ-ring containing the compact G_δ's and the Borel sets to be the smallest σ-ring containing the compact sets. In reading works dealing with Baire and Borel sets or measures and Radon measures, it is imperative to check carefully the author's definitions. A given statement may be true for one usage and false for another.

PROBLEMS

52. Let X be a separable, compact, Hausdorff space. Show that every closed set is a G_δ set.
53. Let X be a Hausdorff space and μ: $\mathcal{B}(X) \to [0, \infty]$ a σ-finite Borel measure. Show that μ is Radon if and only if it is regular.
54. Show that a Hausdorff space X is both locally compact and σ-compact if and only if there is an ascending countable collection $\{\mathcal{O}_k\}_{k=1}^{\infty}$ of open subsets of X that covers X and for each k,
$$\overline{\mathcal{O}}_k \text{ is a compact subset of } \mathcal{O}_{k+1}.$$
55. Let x_0 be a point in the locally compact, Hausdorff space X. Is the Dirac delta measure concentrated at x_0, δ_{x_0}, a regular Baire measure?
56. Let X be an uncountable set with the discrete topology and X^* its Alexandroff compactification with x^* the point at infinity. Show that the singleton set $\{x^*\}$ is a Borel set that is not a Baire set.
57. Let X be a locally compact, Hausdorff space. Show that a Borel measure μ: $\mathcal{B}(X) \to [0, \infty]$ is Radon if and only if every Borel set is measurable with respect to the Carathéodory measure induced by the premeasure μ: $\mathcal{B}(X) \to [0, \infty]$.

58. Prove Lemmas 23, 24, and 25.

59. Let X be a compact, Hausdorff space and $f_1, \ldots, f_n$ continuous real-valued functions on X. Let ν be a signed Radon measure on X with $|\nu|(X) \leq 1$ and let $c_i = \int_X f_i d\nu,$ for $1 \leq i \leq n$.
    (i) Show that there is a signed Radon measure μ on X with $|\mu|(X) \leq 1$ for which
    $$\int_X f_i \, d\mu = c_i$$
    and
    $$\int_X g \, d\mu \leq \int_X g \, d\lambda \text{ for all } g \in C(X)$$
    for any signed Radon measure λ with $|\lambda|(X) \leq 1$ and such that $\int_X f_i \, d\lambda = c_i$ for each i, $1 \leq i \leq n$.
    (ii) Suppose that there is a Radon measure ν on X with $\nu(X) = 1$ and $\int_X f_i d\nu = c_i$ for each $i, 1 \leq i \leq n$. Show that there is a Radon measure μ on X with $\mu(X) = 1$ and $\int_X f_i \, d\mu = c_i,$ for $1 \leq i \leq n$, which minimizes $\int_X g d\mu$ among all Radon measures that satisfy these conditions.

CHAPTER 22

Invariant Measures

Contents

22.1 Topological Groups: The General Linear Group 452
22.2 Kakutani's Fixed-Point Theorem 455
22.3 Invariant Borel Measures on Compact Groups: von Neumann's Theorem . . 460
22.4 Measure Preserving Transformations and Ergodicity 463

A topological group is a group $\mathcal{G}$ together with a Hausdorff topology on $\mathcal{G}$ for which the group operation and inversion are continuous. We prove a seminal theorem of John von Neumann which tells us that on any compact topological group $\mathcal{G}$ there is a unique measure μ on the Borel σ-algebra $\mathcal{B}(\mathcal{G})$, called Haar measure, that is invariant under the left action of the group, that is,

$$\mu(g \cdot E) = \mu(E) \text{ for all } g \in \mathcal{G}, E \in \mathcal{B}(\mathcal{G}).$$

Uniqueness follows from Fubini's Theorem; existence is a consequence of a fixed-point theorem of Shizuo Kakutani which asserts that for a compact group $\mathcal{G}$, there is a functional $\psi \in [C(\mathcal{G})]^*$ for which

$$\psi[f \equiv 1] = 1 \text{ and } \psi[x \mapsto f(x)] = \psi[x \mapsto f(g \cdot x)] \text{ for all } g \in \mathcal{G}, f \in C(\mathcal{G}).$$

Alaoglu's Theorem is crucial in the proof of this fixed-point theorem. Details of the proof of the existence of Haar measure are framed in the context of a group homomorphism of $\mathcal{G}$ into the general linear group of $[C(\mathcal{G})]^*$. We also consider mappings f of a compact metric space X into itself and finite measures on $\mathcal{B}(X)$. Based on Helly's Theorem, we prove the Bogoliubov-Krilov Theorem which tells us that if f is a continuous mapping on a compact metric space X, then there is a measure μ on $\mathcal{B}(X)$ for which

$$\mu(X) = 1 \text{ and } \int_X \varphi \circ f \, d\mu = \int_X \varphi \, d\mu \text{ for all } \varphi \in L^1(X, \mu).$$

Based on the Krein-Milman Theorem, we prove that the above μ may be chosen so that f is ergodic with respect to μ, that is, if A belongs to $\mathcal{B}(X)$ and $\mu([A \sim f(A)] \cup [f(A) \sim A]) = 0$, then $\mu(A) = 0$ or $\mu(A) = 1$.

22.1 TOPOLOGICAL GROUPS: THE GENERAL LINEAR GROUP

Consider a group $\mathcal{G}$ together with a Hausdorff topology on $\mathcal{G}$. For two members g_1, g_2 of $\mathcal{G}$, denote the group operation by $g_1 \cdot g_2$, denote the inverse of a member g of the group by g^{-1}, and let e be the identity of the group. We say that $\mathcal{G}$ is a **topological group** provided the mapping $(g_1, g_2) \mapsto g_1 \cdot g_2$ is continuous from $\mathcal{G} \times \mathcal{G}$ to $\mathcal{G}$, where $\mathcal{G} \times \mathcal{G}$ has the product topology, and the mapping $g \mapsto g^{-1}$ is continuous from $\mathcal{G}$ to $\mathcal{G}$. By a **compact group** we mean a topological group that is compact as a topological space. For subsets $\mathcal{G}_1$ and $\mathcal{G}_2$ of $\mathcal{G}$, we define $\mathcal{G}_1 \cdot \mathcal{G}_2 = \{g_1 \cdot g_2 \mid g_1 \in \mathcal{G}_1,\, g_2 \in \mathcal{G}_2\}$ and $\mathcal{G}_1^{-1} = \{g^{-1} \mid g \in \mathcal{G}_1\}$. If $\mathcal{G}_1$ has just one member g, we denote $\{g\} \cdot \mathcal{G}_2$ by $g \cdot \mathcal{G}_2$.

Let E be a Banach space and $\mathcal{L}(E)$ the Banach space of continuous linear operators on E. The composition of two operators in $\mathcal{L}(E)$ also belongs to $\mathcal{L}(E)$ and clearly, for operators $T,\, S \in \mathcal{L}(E)$,

$$\|S \circ T\| \leq \|S\| \cdot \|T\|. \tag{1}$$

Define $\mathrm{GL}(E)$ to be the collection of invertible operators in $\mathcal{L}(E)$. By definition, an operator in $\mathcal{L}(E)$ is invertible if and only if it is one-to-one and onto; the inverse is continuous by the Open Mapping Theorem. Observe that for $T,\, S \in \mathrm{GL}(E)$, $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$. Therefore, under the operation of composition, $\mathrm{GL}(E)$ is a group called the **general linear group** of E. We denote its identity element by Id. It also is a topological space with the topology induced by the operator norm.

Lemma 1 *Let E be a Banach space and the operator $C \in \mathcal{L}(E)$ have $\|C\| < 1$. Then $\mathrm{Id} - C$ is invertible and*

$$\|(\mathrm{Id} - C)^{-1}\| \leq (1 - \|C\|)^{-1}. \tag{2}$$

Proof We infer from (1) that for each natural number k, $\|C^k\| \leq \|C\|^k$. Hence, since $\|C\| < 1$, the series of real numbers $\sum_{k=0}^{\infty} \|C^k\|$ converges. The normed linear space $\mathcal{L}(E)$ is complete. Therefore, the series[1] of operators $\sum_{k=0}^{\infty} C^k$ converges in $\mathcal{L}(E)$ to a continuous linear operator. But observe that

$$(\mathrm{Id} - C) \circ \Big(\sum_{k=0}^{n} C^k\Big) = \Big(\sum_{k=0}^{n} C^k\Big) \circ (\mathrm{Id} - C) = \mathrm{Id} - C^{n+1} \text{ for all } n.$$

Therefore the series $\sum_{k=0}^{\infty} C^k$ converges to the inverse of $\mathrm{Id} - C$. The estimate (2) follows from this series representation of the inverse of $\mathrm{Id} - C$. □

Theorem 2 *Let E be a Banach space. Then the general linear group of E, $\mathrm{GL}(E)$, is a topological group with respect to the group operation of composition and the topology induced by the operator norm on $\mathcal{L}(E)$.*

Proof For operators T, T', S, S' in $\mathrm{GL}(E)$, observe that

$$T \circ S - T' \circ S' = T \circ (S - S') + (T - T') \circ S'.$$

Therefore, by the triangle inequality for the operator norm and inequality (1),

$$\|T \circ S - T' \circ S'\| \leq \|T\| \cdot \|S - S'\| + \|T - T'\| \cdot \|S'\|.$$

The continuity of composition follows from this inequality.

If S belongs to $\mathrm{GL}(E)$ and $\|S - \mathrm{Id}\| < 1$, then from the identity

$$S^{-1} - \mathrm{Id} = (\mathrm{Id} - S)S^{-1} = (\mathrm{Id} - S)[\mathrm{Id} - (I - S)]^{-1},$$

[1] The series $\sum_{k=0}^{\infty} C^k$ is called the Neumann series for the inverse of $I - C$.

together with the inequalities (1) and (2), we infer that

$$\|S^{-1}-\mathrm{Id}\,\|\leq\frac{\|S-\mathrm{Id}\,\|}{1-\|S-\mathrm{Id}\,\|}. \tag{3}$$

Therefore inversion is continuous at the identity. Now let T and S belong to $\mathrm{GL}(E)$ and $\|S-T\|<\|T^{-1}\|^{-1}$. Then

$$\|T^{-1}S-\mathrm{Id}\,\|=\|T^{-1}(S-T)\|\leq\|T^{-1}\|\cdot\|S-T\|<1.$$

Thus, if we substitute $T^{-1}S$ for S in (3) we have

$$\|S^{-1}T-\mathrm{Id}\,\|\leq\frac{\|T^{-1}S-\mathrm{Id}\,\|}{1-\|T^{-1}S-\mathrm{Id}\,\|}.$$

From this inequality and the identities

$$S^{-1}-T^{-1}=(S^{-1}T-\mathrm{Id})T^{-1}\text{ and }T^{-1}S-\mathrm{Id}=T^{-1}(S-T)$$

we infer that

$$\|S^{-1}-T^{-1}\|\leq\frac{\|T^{-1}\|^2\cdot\|T-S\|}{1-\|T^{-1}\|\cdot\|T-S\|}.$$

The continuity of inversion at T follows from this inequality. □

In the case E is the Euclidean space $\mathbf{R}^n$, $\mathrm{GL}(E)$ is denoted by $\mathrm{GL}(n, \mathbf{R})$. If a choice of basis is made for $\mathbf{R}^n$, then the topology on $\mathrm{GL}(n, \mathbf{R})$ is the topology imposed by the requirement that each of the $n\times n$ entries of the matrix representing the operator with respect to this basis is a continuous function.

A subgroup of a topological group with the subspace topology is also a topological group. For example, if H is a Hilbert space, then the subset of $\mathrm{GL}(H)$ consisting of those operators that leave invariant the inner-product is a topological group that is called the orthogonal linear group of H and denoted by $\mathcal{O}(H)$.

PROBLEMS

In the following exercises, $\mathcal{G}$ is a topological group with unit element e and E is a Banach space.

1. If $\mathcal{T}_e$ is a base for the topology at e, show that $\{g\cdot\mathcal{O}\,|\,\mathcal{O}\in\mathcal{T}_e\}$ is a base for the topology at $g\in\mathcal{G}$.
2. Show that $K_1\cdot K_2$ is compact if K_1 and K_2 are compact subsets of $\mathcal{G}$.
3. Let $\mathcal{O}$ be a neighborhood of e. Show that there is also a neighborhood $\mathcal{U}$ of e for which $\mathcal{U}=\mathcal{U}^{-1}$ and $\mathcal{U}\cdot\mathcal{U}\subseteq\mathcal{O}$.
4. Show that the closure $\overline{H}$ of a subgroup H is a subgroup of $\mathcal{G}$.
5. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be topological groups and $h\colon\mathcal{G}_1\to\mathcal{G}_2$ a group homomorphism. Show that h is continuous if and only if it is continuous at the identity element of $\mathcal{G}_1$.
6. Use the Banach Contraction Principle to prove Lemma 1.

7. Use the completeness of $\mathcal{L}(E)$ to show that if $C \in \mathcal{L}(E)$ and $\|C\| < 1$, then $\sum_{k=0}^{\infty} C^k$ converges in $\mathcal{L}(E)$.
8. Show that the set of $n \times n$ invertible real matrices with determinant 1 is a topological group if the group operation is matrix multiplication and the topology is entrywise continuity. This topological group is called the special linear group and denoted by $\mathrm{SL}(n, \mathbf{R})$.
9. Let H be a Hilbert space. Show that an operator in $\mathrm{GL}(H)$ preserves the norm if and only if it preserves the inner-product.
10. Consider $\mathbf{R}^n$ with the Euclidean inner-product and norm. Characterize those $n \times n$ matrices that represent orthogonal operators with respect to an orthonormal basis.
11. Show that $\mathrm{GL}(E)$ is open in $\mathcal{L}(E)$.
12. Show that the set of operators in $\mathrm{GL}(E)$ comprising operators that are linear compact perturbations of the identity is a subgroup of $\mathrm{GL}(E)$. It is denoted by $\mathrm{GL_C}(E)$.

22.2 KAKUTANI'S FIXED-POINT THEOREM

For two groups $\mathcal{G}$ and $\mathcal{H}$, a mapping $\varphi\colon \mathcal{G} \to \mathcal{H}$ is called a **group homomorphism** provided for each pair of elements g_1, g_2 in $\mathcal{G}$, $\varphi(g_1 \cdot g_2) = \varphi(g_1) \cdot \varphi(g_2)$.

Definition *Let $\mathcal{G}$ be a topological group and E a Banach space. A group homomorphism $\pi\colon \mathcal{G} \to GL(E)$ is called a* **representation**[2] *of $\mathcal{G}$ on E.*

As usual, for a Banach space E, its dual space, the Banach space of bounded linear functionals on E, is denoted by E^*. We recall that the weak-$*$ topology on E^* is the topology with the fewest number of sets among the topologies on E^* such that, for each $x \in E$, the functional on E^* defined by $\psi \mapsto \psi(x)$ is continuous. Alaoglu's Theorem tells us that the closed unit ball of E^* is compact with respect to the weak-$*$ topology.

Definition *Let $\mathcal{G}$ be a topological group, E a Banach space, and $\pi\colon \mathcal{G} \to GL(E)$ a representation of $\mathcal{G}$ on E. The* **adjoint representation** *$\pi^*\colon \mathcal{G} \to GL(E^*)$ is a representation of $\mathcal{G}$ on E^* defined for $g \in \mathcal{G}$ by*

$$\pi^*(g)\psi = \psi \circ \pi(g^{-1}) \textit{ for all } \psi \in E^*. \tag{4}$$

We leave it as an exercise to verify that π^* is a group homomorphism.

A gauge or Minkowski functional on a vector space V is a positively homogeneous, subadditive functional $p\colon V \to \mathbf{R}$. Such functionals determine a base at the origin for the topology of a locally convex topological vector space V. In the presence of a representation π of a compact group $\mathcal{G}$ on a Banach space E, the following lemma establishes the existence of a family, parametrized by $\mathcal{G}$, of positively homogeneous, subadditive functionals on E^*, each of which is invariant under π^* and, when restricted to bounded subsets of E^*, is continuous with respect to the weak-$*$ topology.

Lemma 3 *Let $\mathcal{G}$ be a compact group, E a Banach space, and $\pi\colon \mathcal{G} \to GL(E)$ a representation of $\mathcal{G}$ on E. Let x_0 belong to E and assume that the mapping $g \mapsto \pi(g)x_0$ is continuous from $\mathcal{G}$ to E, where E has the norm topology. Define $p\colon E^* \to \mathbf{R}$ by*

$$p(\psi) = \sup_{g \in \mathcal{G}} |\psi(\pi(g)x_0)| \textit{ for } \psi \in E^*.$$

[2] Observe that no continuity assumption is made regarding a representation. It is convenient to view it as a purely algebraic object and impose continuity assumptions as they are required in a particular context.

Then p is a positively homogeneous, subadditive functional on E^*. It is invariant under π^*, that is,

$$p(\pi^*(g_0)\psi) = p(\psi) \text{ for all } \psi \in E^* \text{ and } g_0 \in \mathcal{G}.$$

Furthermore, the restriction of p to any bounded subset of E^* is continuous with respect to the weak-$*$ topology on E^*.

Proof Since $\mathcal{G}$ is compact and, for $\psi \in \mathcal{G}$, the functional $g \mapsto \psi(\pi(g)x_0)$ is continuous on $\mathcal{G}$, $p: E^* \to \mathbf{R}$ is properly defined. It is clear that p is positively homogeneous, subadditive, and invariant under π^*. Let B^* be a bounded subset of E^*. To establish the weak-$*$ continuity at $p: B^* \to \mathbf{R}$, it suffices to show that for each $\psi_0 \in B^*$ and $\epsilon > 0$, there is a weak-$*$ neighborhood $\mathcal{N}(\psi_0)$ of ψ_0 for which

$$|\psi(\pi(g)x_0) - \psi_0(\pi(g)x_0)| < \epsilon \text{ for all } \psi \in \mathcal{N}(\psi_0) \cap B^* \text{ and } g \in \mathcal{G}. \tag{5}$$

Let ψ_0 belong to B^* and $\epsilon > 0$. Choose $M > 0$ such that $\|\psi\| \leq M$ for all $\psi \in B^*$. The mapping $g \mapsto \pi(g)x_0$ is continuous and $\mathcal{G}$ is compact. Therefore there are a finite number of points $\{g_1, \ldots, g_n\}$ in $\mathcal{G}$ and for each $k, 1 \leq k \leq n$, a neighborhood $\mathcal{O}_k$ of g_k such that $\{\mathcal{O}_{g_k}\}_{k=1}^n$ covers $\mathcal{G}$ and, for $1 \leq k \leq n$,

$$\|\pi(g)x_0 - \pi(g_k)x_0\| < \epsilon/4M \text{ for all } g \in \mathcal{O}_{g_k}. \tag{6}$$

Define the weak-$*$ neighborhood $\mathcal{N}(\psi_0)$ of ψ_0 by

$$\mathcal{N}(\psi_0) = \{\psi \in E^* \mid |(\psi - \psi_0)(\pi(g_k)x_0)| < \epsilon/2 \text{ for } 1 \leq k \leq n\}.$$

Observe that for any $g \in \mathcal{G}$, $\psi \in E^*$ and $1 \leq k \leq n$,

$$\psi(\pi(g)x_0) - \psi_0(\pi(g)x_0) = (\psi - \psi_0)[\pi(g_k)x_0] + (\psi - \psi_0)[\pi(g)x_0 - \pi(g_k)x_0]. \tag{7}$$

To verify (5), let g belong to $\mathcal{G}$ and ψ belong to $\mathcal{N}(\psi_0) \cap B^*$. Choose $k, 1 \leq k \leq n$, for which g belongs to $\mathcal{O}_k$. Then $|(\psi - \psi_0)[\pi(g_k)x_0]| < \epsilon/2$ since ψ belongs to $\mathcal{N}(\psi_0)$. On the other hand, since $\|\psi - \psi_0\| \leq 2M$, we infer from (6) that $|(\psi - \psi_0)[\pi(g)x_0 - \pi(g_k)x_0]| < \epsilon/2$. Therefore, by (7), (5) holds for $\mathcal{N}(\psi_0)$. $\square$

Definition *Let $\mathcal{G}$ be a topological group, E a Banach space, and $\pi: \mathcal{G} \to GL(E)$ a representation of $\mathcal{G}$ on E. A subset K of E is said to be* **invariant** *under π provided $\pi(g)K \subseteq K$ for all $g \in \mathcal{G}$. A point $x \in E$ is said to be* **fixed** *under π provided $\pi(g)x = x$ for all $g \in \mathcal{G}$.*

Theorem 4 *Let $\mathcal{G}$ be a compact group, E a Banach space, and $\pi: \mathcal{G} \to GL(E)$ a representation of $\mathcal{G}$ on E. Assume that for each $x \in E$, the mapping $g \mapsto \pi(g)x$ is continuous from $\mathcal{G}$ to E, where E has the norm topology. Assume that there is a non-empty, convex, weak-$*$ compact subset K^* of E^* that is invariant under π^*. Then there is a functional ψ in K^* that is fixed under π^*.*

Proof Let $\mathcal{F}$ be the collection of all non-empty, convex, weak-$*$ closed subsets of K^* that are invariant under π^*. The collection $\mathcal{F}$ is non-empty since K^* belongs to $\mathcal{F}$. Order $\mathcal{F}$ by set inclusion. This defines a partial ordering on $\mathcal{F}$. Every totally ordered subcollection of $\mathcal{F}$ has the finite intersection property. But for any compact topological space, a collection of non-empty closed subsets that has the finite intersection property has non-empty intersection. The intersection of any collection of convex sets is convex and the intersection of any collection of π^* invariant sets is π^* invariant. Therefore every totally ordered subcollection of $\mathcal{F}$ has its non-empty intersection as a lower bound. We infer from Zorn's Lemma that there is a set K_0^* in $\mathcal{F}$ that is minimal with respect to containment, that is, no proper subset of K_0^* belongs to $\mathcal{F}$. This minimal subset is weak-$*$ closed and therefore weak-$*$ compact. We relabel and assume K^* itself is this minimal subset.

We claim that K^* consists of a single functional. Otherwise, choose two distinct functionals ψ_1 and ψ_2 in K^*. Choose $x_0 \in E$ such that $\psi_1(x_0) \neq \psi_2(x_0)$. Define the functional $p\colon K^* \to \mathbf{R}$ by

$$p(\psi) = \sup_{g \in \mathcal{G}} |\psi(\pi(g)x_0)| \text{ for } \psi \in K^*.$$

Since K^* is weak-$*$ compact, the Uniform Boundedness Principle tells us that K^* is bounded. According to the preceding lemma, p is continuous with respect to the weak-$*$ topology. Therefore, if, for $r > 0$ and $\eta \in K^*$, we define

$$B_0(\eta, r) = \{\psi \in K^* \mid p(\psi - \eta) < r\} \text{ and } \overline{B}_0(\eta, r) = \{\psi \in K^* \mid p(\psi - \eta) \leq r\}, \tag{8}$$

then $B_0(\eta, r)$ is open with respect to the weak-$*$ topology on K^* and $\overline{B}_0(\eta, r)$ is closed with respect to the same topology. Each of these sets is convex since, again by the preceding lemma, p is positively homogeneous and subadditive.

Define $d = \sup\{p(\psi - \varphi) \mid \psi, \varphi \in K^*\}$. Then d is finite since p is continuous on the weak-$*$ compact set K^*, and $d > 0$ since $p(\psi_1 - \psi_2) > 0$. Since K^* is weak-$*$ compact and each $B_0(\eta, r)$ is weak-$*$ open, we may choose a finite subset $\{\psi_k\}_{k=1}^n$ of K^* for which

$$K^* = \bigcup_{k=1}^n B_0(\psi_k, d/2).$$

Define

$$\psi^* = \frac{\psi_1 + \cdots + \psi_k + \cdots + \psi_n}{n}.$$

The functional ψ^* belongs to K^* since K^* is convex. Let ψ be any functional in K^*. By the definition of d, $p(\psi - \psi_k) \leq d$ for $1 \leq k \leq n$. Since $\{B_0(\psi_k, d/2)\}_{k=1}^n$ covers K^*, ψ belongs to some $B_0(\psi_{k_0}, d/2)$ for some k_0. Thus, by the positive homogeneity and subadditivity of p,

$$p(\psi - \psi^*) \leq d' \text{ where } d' = \frac{n-1}{n} \cdot d + \frac{d}{4} < d.$$

Define

$$K' = \bigcap_{\psi \in K^*} \overline{B}_0(\psi, d').$$

Then K' is a weak-$*$ closed, and hence weak-$*$ compact, convex subset of K^*. It is non-empty since it contains the functional ψ^*. We claim that K' is invariant under π^*. To verify this, for $\eta \in K'$, $\psi \in K^*$ and $g \in \mathcal{G}$, we must show that $p(\pi^*(g)\eta - \psi) \leq d'$. Since p is π^* invariant and $p(\eta - \pi^*(g^{-1})\psi) \leq d'$,

$$p(\pi^*(g)\eta - \psi) = p(\eta - \pi^*(g^{-1})\psi) \leq d'.$$

By the minimality of K^*, $K^* = K'$. This is a contradiction since, by the definition of d, there are functionals ψ' and ψ'' in K^* for which $p(\psi' - \psi'') > d'$ and hence ψ'' does not belong to $\overline{B}_0(\psi', d')$. We infer from this contradiction that K^* consists of a single functional. The proof is complete. $\square$

Definition *Let $\mathcal{G}$ be a compact group and $C(\mathcal{G})$ the Banach space of continuous real-valued functions on $\mathcal{G}$, normed by the maximum norm. By the* **regular representation** *of $\mathcal{G}$ on $C(\mathcal{G})$ we mean the representation $\pi: \mathcal{G} \to GL(C(\mathcal{G}))$ defined by*

$$[\pi(g)f](x) = f(g^{-1} \cdot x) \text{ for all } f \in C(\mathcal{G}), x \in \mathcal{G} \text{ and } g \in \mathcal{G}.$$

We leave it as an exercise to show that the regular representation is indeed a representation. The following lemma shows that the regular representation of a compact group $\mathcal{G}$ on $C(\mathcal{G})$ possesses the continuity property imposed in Theorem 4.

Lemma 5 *Let $\mathcal{G}$ be a compact group and $\pi: \mathcal{G} \to GL(C(\mathcal{G}))$ the regular representation of $\mathcal{G}$ on $C(\mathcal{G})$. Then for each $f \in C(\mathcal{G})$, the mapping $g \mapsto \pi(g)f$ is continuous from $\mathcal{G}$ to $C(\mathcal{G})$, where $C(\mathcal{G})$ has the topology induced by the maximum norm.*

Proof Let f belong to $C(\mathcal{G})$. It suffices to check that the mapping $g \mapsto \pi(g)f$ is continuous at the identity $e \in \mathcal{G}$. Let $\epsilon > 0$. We claim that there is a neighborhood of the identity, $\mathcal{U}$, for which

$$|f(g \cdot x) - f(x)| < \epsilon \text{ for all } g \in \mathcal{U}, x \in \mathcal{G}. \tag{9}$$

Let x belong to $\mathcal{G}$. Choose a neighborhood of x, $\mathcal{O}_x$, for which

$$|f(x') - f(x)| < \epsilon/2 \text{ for all } x' \in \mathcal{O}_x.$$

Thus

$$|f(x') - f(x'')| < \epsilon \text{ for all } x', x'' \in \mathcal{O}_x. \tag{10}$$

By the continuity of the group operation, we may choose a neighborhood of the identity, $\mathcal{U}_x$, and a neighborhood $x, \mathcal{V}_x$, for which $\mathcal{V}_x \subseteq \mathcal{O}_x$ and $\mathcal{U}_x \cdot \mathcal{V}_x \subseteq \mathcal{O}_x$. By the compactness of $\mathcal{G}$, there is a finite collection $\{\mathcal{V}_{x_k}\}_{k=1}^n$ that covers $\mathcal{G}$. Define $\mathcal{U} = \bigcap_{k=1}^n \mathcal{U}_{x_k}$. Then $\mathcal{U}$ is a neighborhood of the identity in $\mathcal{G}$. We claim that (9) holds for this choice of $\mathcal{U}$. Indeed, let g belong in $\mathcal{U}$ and x belong to $\mathcal{G}$. Then x belongs to some $\mathcal{V}_{x_k}$. Hence

$$x \in \mathcal{V}_{x_k} \subseteq \mathcal{O}_{x_k} \text{ and } g \cdot x \in \mathcal{U} \times \mathcal{V}_{x_k} \subseteq \mathcal{U}_{x_k} \times \mathcal{V}_{x_k} \subseteq \mathcal{O}_{x_k}.$$

Therefore, both x and $g \cdot x$ belong to $\mathcal{O}_{x_k}$ so that, by (10), $|f(g \cdot x) - f(x)| < \epsilon$. Thus (9) is established. Replace $\mathcal{U}$ by $\mathcal{U} \cap \mathcal{U}^{-1}$. Therefore,

$$|f(g^{-1} \cdot x) - f(x)| < \epsilon \text{ for all } g \in \mathcal{U}, x \in \mathcal{G},$$

that is,

$$\|\pi(g)f - \pi(e)f\|_{\max} < \epsilon \text{ for all } g \in \mathcal{U}.$$

This establishes the required continuity. $\square$

For $\mathcal{G}$ a compact group, we call a functional $\psi \in [C(\mathcal{G})]^*$ a **probability functional** provided it takes the value 1 at the constant function $f = 1$ and is **positive** in the sense that for $f \in C(\mathcal{G})$, if $f \geq 0$ on $\mathcal{G}$, then $\psi(f) \geq 0$.

Theorem 6 (Kakutani) *Let $\mathcal{G}$ be a compact group and $\pi\colon \mathcal{G} \to GL(C(\mathcal{G}))$ be the regular representation of $\mathcal{G}$ on $C(\mathcal{G})$. Then there is a probability functional $\psi \in [C(\mathcal{G})]^*$ that is fixed under the adjoint action π^*, that is,*

$$\psi(f) = \psi(\pi^*(g)f) \text{ for all } f \in C(\mathcal{G}) \text{ and } g \in \mathcal{G}. \tag{11}$$

Proof According to Alaoglu's Theorem, the closed unit ball of $[C(\mathcal{G})]^*$ is weak-$*$ compact. Let K^* be the collection of positive probability functionals on $C(\mathcal{G})$. Observe that if ψ is a probability functional and f belongs to $C(\mathcal{G})$ with $\|f\|_{\max} \leq 1$, then, by the positivity and linearity of ψ, since $-1 \leq f \leq 1$,

$$-1 = \psi(-1) \leq \psi(f) \leq \psi(1) = 1.$$

Thus $|\psi(f)| \leq 1$ and hence $\|\psi\| \leq 1$. Therefore, K^* is a convex subset of the closed unit ball of E^*. We claim that K^* is weak-$*$ closed. Indeed, for each non-negative function $f \in C(\mathcal{G})$, the set $\{\psi \in [C(\mathcal{G})]^* \mid \psi(f) \geq 0\}$ is weak-$*$ closed, as is the set of functionals ψ that take the value 1 at the constant function $f \equiv 1$. The set K^* is therefore the intersection of weak-$*$ closed sets and so it is weak-$*$ closed. As a closed subset of a compact set, K^* is weak-$*$ compact. Finally, the set K^* is non-empty since if x_0 is any point in $\mathcal{G}$, the Dirac functional that takes the value $f(x_0)$ at each $f \in C(\mathcal{G})$ belongs to K^*.

It is clear that K^* is invariant under π^*. The preceding lemma tells us that the regular representation possesses the continuity required to apply Theorem 4. According to that theorem, there is a functional in $\psi \in K^*$ that is fixed under π^*, that is, (11) holds. $\square$

PROBLEMS

13. Show that the adjoint of a representation also is a representation.
14. Show that a probability functional has norm 1.
15. Let E be a reflexive Banach space and K^* a convex subset of E^* that is closed with respect to the metric induced by the norm. Show that K^* is weak-$*$ closed. On the other hand, show that if E is not reflexive, then the image of the closed unit ball of E under the natural embedding of E in $(E^*)^* = E^{**}$ is a subset of E^{**} that is convex, closed and bounded with respect to the metric induced by the norm but is not weak-$*$ closed.
16. Let $\mathcal{G}$ be a compact group, E a reflexive Banach space, and $\pi\colon \mathcal{G} \to \mathrm{GL}(E)$ a representation. Suppose that for each $x \in E$, the mapping $g \mapsto \pi(g)x$ is continuous. Assume that there is a non-empty strongly closed, bounded, convex subset K of E that is invariant with respect to π. Show that K contains a point that is fixed by π.
17. Let $\mathcal{G}$ be a topological group, E be a Banach space, and $\pi\colon \mathcal{G} \to \mathrm{GL}(E)$ a representation. For $x \in E$, show that the mapping $g \mapsto \pi(g)x$ is continuous if and only if it is continuous at e.
18. Suppose $\mathcal{G}$ is a topological group, X a topological space, and $\varphi\colon \mathcal{G} \times X \to X$ a mapping. For $g \in \mathcal{G}$, define the mapping $\pi(g)\colon X \to X$ by $\pi(g)x = \varphi(g, x)$ for all $x \in X$. What properties must φ possess in order for π to be a representation on $\mathcal{G}$ on $C(X)$? What further properties must φ possess in order that for each $x \in X$, the mapping $g \mapsto \pi(g)x$ is continuous?

22.3 INVARIANT BOREL MEASURES ON COMPACT GROUPS: VON NEUMANN'S THEOREM

Recall that a Borel measure on a compact topological space X is a finite measure on $\mathcal{B}(X)$, the smallest σ-algebra that contains the topology on X. We now consider Borel measures on compact groups and their relation to the group operation.

Lemma 7 *Let $\mathcal{G}$ be a compact group and μ a Borel measure on $\mathcal{B}(\mathcal{G})$. For $g \in \mathcal{G}$, define the set-function $\mu_g \colon \mathcal{B}(\mathcal{G}) \to [0, \infty)$ by*

$$\mu_g(A) = \mu(g \cdot A) \text{ for all } A \in \mathcal{B}(\mathcal{G}).$$

Then μ_g is a Borel measure. If μ is Radon, so is μ_g. Furthermore, if π is the regular representation of $\mathcal{G}$ on $C(\mathcal{G})$,[3] then

$$\int_{\mathcal{G}} \pi(g) f \, d\mu = \int_{\mathcal{G}} f \, d\mu_g \text{ for all } f \in C(\mathcal{G}). \tag{12}$$

Proof Let g belong to $\mathcal{G}$. Observe that multiplication on the left by g defines a topological homeomorphism of G onto G. From this we infer that A is a Borel set if and only if $g \cdot A$ is a Borel set. Therefore the set-function μ_g is properly defined on $\mathcal{B}(\mathcal{G})$. Clearly, μ_g inherits countable additivity from μ and hence, since $\mu_g(\mathcal{G}) = \mu(\mathcal{G}) < \infty$, μ_g is a Borel measure. Now suppose μ is a Radon measure. To establish the inner regularity of μ_g, let $\mathcal{O}$ be open in $\mathcal{G}$ and $\epsilon > 0$. Since μ is inner regular and $g \cdot \mathcal{O}$ is open, there is a compact set K contained in $g \cdot \mathcal{O}$ for which $\mu(g \cdot \mathcal{O} \sim K) < \epsilon$. Hence $K' = g^{-1} \cdot K$ is compact, contained in $\mathcal{O}$, and $\mu_g(\mathcal{O} \sim K') < \epsilon$. Thus μ_g is inner regular. A similar argument shows μ_g is outer regular. Therefore μ_g is a Radon measure.

We now verify (12). Integration is linear. Therefore, if (12) holds for characteristic functions of Borel sets it also holds for simple Borel functions. We infer from the Simple Approximation Theorem and the Bounded Convergence Theorem that (12) holds for all $f \in C(\mathcal{G})$ if it holds for simple Borel functions. It therefore suffices to verify (12) in the case $f = \chi_A$, the characteristic function of the Borel set A. However, for such a function,

$$\int_{\mathcal{G}} \pi(g) f \, d\mu = \mu(g \cdot A) = \int_{\mathcal{G}} f \, d\mu_g.$$

□

Definition *Let $\mathcal{G}$ be a compact group. A Borel measure $\mu \colon \mathcal{B}(\mathcal{G}) \to [0, \infty)$ is said to be* **left-invariant** *provided*

$$\mu(A) = \mu(g \cdot A) \text{ for all } g \in \mathcal{G} \text{ and } A \in \mathcal{B}(\mathcal{G}). \tag{13}$$

It is said to be a **probability measure** *provided $\mu(\mathcal{G}) = 1$.*

[3] A continuous function on a topological space is measurable with respect to the Borel σ-algebra on the space and, if the space is compact and the measure is Borel, it is integrable with respect to this measure. Therefore, each side of the following formula is properly defined because, for each $f \in C(\mathcal{G})$ and $g \in \mathcal{G}$, both f and $\pi(g)f$ are continuous functions on the compact topological space $\mathcal{G}$ and both μ and μ_g are Borel measures.

A right-invariant measure is defined similarly. If we consider $\mathbf{R}^n$ as a topological group under the operation of addition, we showed that the restriction of Lebesgue measure μ_n on $\mathbf{R}^n$ to $\mathcal{B}(\mathbf{R}^n)$ is left-invariant with respect to addition, that is, $\mu_n(E+x) = \mu_n(E)$ for each Borel subset E of $\mathbf{R}^n$ and each point $x \in \mathbf{R}^n$. Of course, this also holds for any Lebesgue measurable subset E of $\mathbf{R}^n$.

Proposition 8 *On each compact group $\mathcal{G}$ there is a Radon probability measure on $\mathcal{B}(\mathcal{G})$ that is left-invariant and also one that is right-invariant.*

Proof Theorem 6 tells us that there is a probability functional $\psi \in [C(\mathcal{G})]^*$ that is fixed under the adjoint of the regular representation on $\mathcal{G}$ on $C(\mathcal{G})$. This means that $\psi(1) = 1$ and

$$\psi(f) = \psi(\pi(g^{-1})f) \text{ for all } f \in C(\mathcal{G}) \text{ and } g \in \mathcal{G}. \tag{14}$$

On the other hand, according to the Riesz-Markov Theorem, there is a unique Radon measure μ on $\mathcal{B}(\mathcal{G})$ that represents ψ in the sense that

$$\psi(f) = \int_{\mathcal{G}} f\, d\mu \text{ for all } f \in C(\mathcal{G}). \tag{15}$$

Therefore, by (14),

$$\psi(f) = \psi(\pi(g^{-1})f) = \int_{\mathcal{G}} \pi(g^{-1})f\, d\mu \text{ for all } f \in C(\mathcal{G}) \text{ and } g \in \mathcal{G}. \tag{16}$$

Hence, by Lemma 7,

$$\psi(f) = \int_{\mathcal{G}} f\, d\mu_{g^{-1}} \text{ for all } f \in C(\mathcal{G}) \text{ and } g \in \mathcal{G}.$$

By the same lemma, $\mu_{g^{-1}}$ is a Radon measure. We infer from the uniqueness of the representation of the functional ψ that

$$\mu = \mu_{g^{-1}} \text{ for all } g \in \mathcal{G}.$$

Thus μ is a left-invariant Radon measure. It is a probability measure because ψ is a probability functional and thus

$$1 = \psi(1) = \int_{\mathcal{G}} d\mu = \mu(\mathcal{G}).$$

A dual argument (see Problem 25) establishes the existence of a right-invariant Radon probability measure. □

Definition *Let $\mathcal{G}$ be a topological group. A Radon measure on $\mathcal{B}(\mathcal{G})$ is said to be a* **Haar** *measure provided it is a left-invariant probability measure.*

Theorem 9 (von Neumann) *Let $\mathcal{G}$ be a compact group. Then there is a unique Haar measure μ on $\mathcal{B}(\mathcal{G})$. The measure μ is also right-invariant.*

Proof According to the preceding proposition, there is a left-invariant Radon probability measure μ on $\mathcal{B}(\mathcal{G})$ and a right-invariant Radon probability measure ν on $\mathcal{B}(\mathcal{G})$. We claim that

$$\int_{\mathcal{G}} f\,d\mu = \int_{\mathcal{G}} f\,d\nu \text{ for all } f \in C(\mathcal{G}). \tag{17}$$

Once this is verified, we infer from the uniqueness of representations of bounded linear functionals on $C(\mathcal{G})$ by integration against Radon measures that $\mu = \nu$. Therefore every left-invariant Radon measure equals ν. Hence there is only one left-invariant Radon measure and it is right-invariant.

To verify (17), let f belong to $C(\mathcal{G})$. Define $h\colon \mathcal{G} \times \mathcal{G} \to \mathbf{R}$ by $h(x, y) = f(x \cdot y)$ for $(x, y) \in \mathcal{G} \times \mathcal{G}$. Then h is a continuous function on $\mathcal{G} \times \mathcal{G}$. Moreover the product measure $\nu \times \mu$ is defined on a σ-algebra of subsets of $\mathcal{G} \times \mathcal{G}$ containing $\mathcal{B}(\mathcal{G} \times \mathcal{G})$. Therefore, since h is measurable and bounded on a set $\mathcal{G} \times \mathcal{G}$ of finite $\nu \times \mu$ measure, it is integrable with respect to the product measure $\nu \times \mu$ over $\mathcal{G} \times \mathcal{G}$. To verify (17) it suffices to show that

$$\int_{\mathcal{G}\times\mathcal{G}} h\,d[\nu \times \mu] = \int_{\mathcal{G}} f\,d\mu \;\text{ and }\; \int_{\mathcal{G}\times\mathcal{G}} h\,d[\mu \times \nu] = \int_{\mathcal{G}} f\,d\nu. \tag{18}$$

However, by Fubini's Theorem,

$$\int_{\mathcal{G}\times\mathcal{G}} h\,d[\nu \times \mu] = \int_{\mathcal{G}} \left[\int_{\mathcal{G}} h(x, \cdot)\,d\mu(y)\right] d\nu(x).$$

By the left-invariance of μ and (12),

$$\int_{\mathcal{G}} h(x, \cdot)\,d\mu(y) = \int_{\mathcal{G}} f\,d\mu \text{ for all } x \in \mathcal{G}.$$

Thus, since $\nu(\mathcal{G}) = 1$,

$$\int_{\mathcal{G}\times\mathcal{G}} h\,d[\mu \times \nu] = \int_{\mathcal{G}} f\,d\mu \cdot \nu(\mathcal{G}) = \int_{\mathcal{G}} f\,d\mu.$$

A similar argument establishes the right-hand equality in (18) and thereby completes the proof. $\square$

The methods studied here may be extended to show that there is a left-invariant Haar measure on any locally compact group $\mathcal{G}$, although it may not be right-invariant. Here we investigated one way in which the topology on a topological group determines its measure theoretic properties. Of course, it is also interesting to investigate the influence of measure on topology. For further study of this interesting circle of ideas see John von Neumann's classic lecture notes *Invariant Measures* [vN91].

PROBLEMS

19. Let μ be a Borel probability measure on a compact group $\mathcal{G}$. Show that μ is Haar measure if and only if

$$\int_{\mathcal{G}} f \circ \varphi_g\,d\mu = \int_{\mathcal{G}} f\,d\mu \text{ for all } g \in \mathcal{G}, f \in C(\mathcal{G}),$$

where $\varphi_g(g') = g \cdot g'$ for all $g' \in \mathcal{G}$.

20. Let μ be Haar measure on a compact group $\mathcal{G}$. Show that $\mu \times \mu$ is Haar measure on $\mathcal{G} \times \mathcal{G}$.
21. Let $\mathcal{G}$ be a compact group whose topology is given by a metric. Show that there is a $\mathcal{G}$-invariant metric. (Hint: Use the preceding two problems and average the metric over the group $\mathcal{G} \times \mathcal{G}$.)
22. Let μ be Haar measure on a compact group $\mathcal{G}$. If $\mathcal{G}$ has infinitely many members, show that $\mu(\{g\}) = 0$ for each $g \in \mathcal{G}$. If $\mathcal{G}$ is finite, explicitly describe μ.
23. Show that if μ is Haar measure on a compact group, then $\mu(\mathcal{O}) > 0$ for every open subset $\mathcal{O}$ of $\mathcal{G}$.
24. Let $S^1 = \{z = e^{i\theta} \mid \theta \in \mathbf{R}\}$ be the circle with the group operation of complex multiplication and the topology it inherits from the Euclidean plane.
    (i) Show that S^1 is a topological group.
    (ii) Define $\Lambda = \{(\alpha, \beta) \mid \alpha, \beta \in \mathbf{R}, 0 < \beta - \alpha < 2\pi\}$. For $\lambda = (\alpha, \beta) \in \Lambda$, define $I_\alpha = \{e^{i\theta} \mid \alpha < \theta < \beta\}$. Show that every proper open subset of S^1 is the countable disjoint union of sets of the form I_λ, $\lambda \in \Lambda$.
    (iii) For $\lambda = (\alpha, \beta) \in \Lambda$, define $\mu(I_\alpha) = (\beta - \alpha)/2\pi$. Define $\mu(S^1) = 1$. Use part (ii) to extend μ to set-function defined on the topology $\mathcal{T}$ of S^1. Then verify that, by Proposition 9 from the preceding chapter, μ may be extended to a Borel measure μ on $\mathcal{B}(S^1)$.
    (iv) Show that the measure defined in part (ii) is Haar measure on S^1.
    (v) The torus T^n is the topological group consisting of the Cartesian product of n copies of S^1 with the product topology and group structure. What is Haar measure on T^n?
25. Let μ be a Borel measure on a topological group $\mathcal{G}$. For a Borel set E, define $\mu'(E) = \mu(E^{-1})$, where $E^{-1} = \{g^{-1} \mid g \in E\}$. Show that μ' also is a Borel measure. Moreover, show that μ is left-invariant if and only if μ' is right-invariant.

22.4 MEASURE PRESERVING TRANSFORMATIONS AND ERGODICITY

For a measurable space $(X, \mathcal{M})$, a mapping $T: X \to X$ is said to be a **measurable transformation** provided for each measurable set E, $T^{-1}(E)$ also is measurable. Observe that for a mapping $T: X \to X$,

$$T \text{ is measurable if and only if } g \circ T \text{ is measurable whenever the function } g \text{ is measurable.} \tag{19}$$

Recall that, in Chapter 6, we proved the von Neumann Composition Theorem for Lebesgue measure on R, according to which a continuous, strictly increasing function $f: [a, b] \to R$ has the property that $g \circ f \circ [a, b] \to R$ is measurable whenever $g: R \to R$ is measurable if and only if its inverse f^{-1} is absolutely continuous.

For a measure space $(X, \mathcal{M}, \mu)$, a measurable transformation $T: X \to X$ is said to be **measure preserving** provided

$$\mu(T^{-1}(A)) = \mu(A) \text{ for all } A \in \mathcal{M}.$$

In Chapter 10, we proved that an invertible linear operator on R^n is Lebesgue measure preserving if and only if it is orthogonal, and for such an operator proved the following change of variable theorem in R^n.

Proposition 10 *Let $(X, \mathcal{M}, \mu)$ be a finite measure space and $T: X \to X$ a measurable transformation. Then T is measure preserving if and only if $g \circ T$ is integrable over X whenever g is, and*

$$\int_X g \circ T \, d\mu = \int_X g \, d\mu \text{ for all } g \in L^1(X, \mu). \tag{20}$$

Proof First assume (20) holds. For $A \in \mathcal{M}$, since $\mu(X) < \infty$, the function $g = \chi_A$ belongs to $L^1(X, \mu)$ and $g \circ T = \chi_{T^{-1}(A)}$. We infer from (20) that $\mu(T^{-1}(A)) = \mu(A)$.

Conversely, assume T is measure preserving. Let g be integrable over X. If g^+ is the positive part of g, then $(g \circ T)^+ = g^+ \circ T$. Similarly for the negative part. We may therefore assume that g is non-negative. For a simple function $g = \sum_{k=1}^n c_k \cdot \chi_{A_k}$, since T is measure preserving,

$$\int_X g \circ T \, d\mu = \int_X \left[\sum_{k=1}^n c_k \cdot \chi_{A_k} \circ T \right] d\mu = \int_X \left[\sum_{k=1}^n c_k \cdot \chi_{T^{-1}(A_k)} \right] d\mu = \sum_{k=1}^n c_k \cdot \mu(A_k) = \int_X g \, d\mu.$$

Therefore, (20) holds for simple functions g. According to the Simple Approximation Theorem, there is an increasing sequence $\{g_n\}$ of simple functions on X that converge pointwise on X to g. Hence $\{g_n \circ T\}$ is an increasing of simple functions on X that converge pointwise on X to $g \circ T$. Using the Monotone Convergence Theorem twice and the validity of (20) for simple functions, we have

$$\int_X g \circ T \, d\mu = \lim_{n \to \infty} \left[\int_X g_n \circ T \, d\mu \right] = \lim_{n \to \infty} \left[\int_X g_n \, d\mu \right] = \int_X g \, d\mu.$$

□

For a measure space $(X, \mathcal{M}, \mu)$ and measurable transformation $T: X \to X$, a measurable set A is said to be **invariant** under T (with respect to μ) provided

$$\mu(A \sim T^{-1}(A)) = \mu(T^{-1}(A) \sim A) = 0,$$

that is, modulo sets of measure 0, $T^{-1}(A) = A$. It is clear that

$$A \text{ is invariant under } T \text{ if and only if } \chi_A \circ T = \chi_A \text{ almost everywhere on } X. \tag{21}$$

If $(X, \mathcal{M}, \mu)$ also is a probability space, that is, $\mu(X) = 1$, a measure preserving transformation T is said to be **ergodic** provided any set A that is invariant under T with respect to μ has $\mu(A) = 0$ or $\mu(A) = 1$.

Proposition 11 *Let $(X, \mathcal{M}, \mu)$ be a probability space and $T: X \to X$ a measure preserving transformation. Then, among real-valued measurable functions g on X,*

T is ergodic if and only if whenever $g \circ T = g$ almost everywhere on X, then g is constant almost everywhere on X. (22)

Proof First assume that whenever $g \circ T = g$ almost everywhere on X, then g is constant almost everywhere on X. Let $A \in \mathcal{M}$ be invariant under T. Then $g = \chi_A$, the characteristic function of A, is measurable and $\chi_A \circ T = \chi_A$ almost everywhere on X. Thus χ_A is constant almost everywhere, that is, $\mu(A) = 0$ or $\mu(A) = 1$.

Conversely, assume T is ergodic. Let g be a real-valued measurable function on X for which $g \circ T = g$ almost everywhere on X. Let k be an integer. Define $X_k = \{x \in X \mid k \leq g(x) < k+1\}$. Then X_k is a measurable set that is invariant under T. By the ergodicity of T, either $\mu(X_k) = 0$ or $\mu(X_k) = 1$. The countable collection $\{X_k\}_{k \in \mathbf{Z}}$ is disjoint and its union is X. Since $\mu(X) = 1$ and μ is countably additive, $\mu(X_k) = 0$, except for exactly one integer k'. Define $I_1 = [k', k'+1]$. Then $\mu\{x \in X \mid g(x) \in I_1\} = 1$ and the length of I_1, $\ell(I_1)$, is 1.

Let n be a natural number for which the descending finite collection $\{I_k\}_{k=1}^n$ of closed, bounded intervals have been defined for which

$$\ell(I_k) = 1/2^{k-1} \text{ and } \mu\{x \in X \mid \ g(x) \in I_k\} = 1 \text{ for } 1 \leq k \leq n.$$

Let $I_n = [a_n, b_n]$, define $c_n = (b_n - a_n)/2$,

$$A_n = \{x \in X \mid \ a_n \leq g(x) < c_n\} \text{ and } B_n = \{x \in X \mid \ c_n \leq g(x) \leq b_n\}.$$

Then A_n and B_n are disjoint measurable sets whose union is $\mu\{x \in X \mid g(x) \in I_n\}$, a set of measure 1. Since both A_n and B_n are invariant under T, we infer from the ergodicity of T that exactly one of these sets has measure 1. If $\mu(A_n) = 1$, define $I_{n+1} = [a_n, c_n]$. Otherwise, define $I_{n+1} = [c_n, b_n]$. Then $\ell(I_{n+1}) = 1/2^n$ and $\mu\{x \in X \mid g(x) \in I_{n+1}\} = 1$. We have inductively defined a descending countable collection $\{I_n\}_{n=1}^\infty$ of closed, bounded intervals such that

$$\ell(I_n) = 1/2^{n-1} \text{ and } \mu\{x \in X \mid \ g(x) \in I_n\} = 1 \text{ for all } n.$$

By the Nested Set Theorem for the real numbers, there is a number c that belongs to every I_n. We claim that $g = c$ almost everywhere on X. Indeed, observe that if $g(x)$ belongs to I_n, then $|g(x) - c| \leq 1/2^{n-1}$ and therefore

$$1 = \mu\{x \in X \mid \ g(x) \in I_n\} \leq \mu\left\{x \in X \ \middle| \ |g(x) - c| \leq 1/2^{n-1}\right\} \leq 1.$$

Since

$$\{x \in X \mid \ g(x) = c\} = \bigcap_{n=1}^{\infty} \left\{x \in X \ \middle| \ |g(x) - c| \leq 1/2^{n-1}\right\},$$

we infer from the continuity of measure that

$$\mu\{x \in X \mid \ g(x) = c\} = \lim_{n \to \infty} \mu\left\{x \in X \ \middle| \ |g(x) - c| \leq 1/2^{n-1}\right\} = 1.$$

$\square$

Theorem 12 (Bogoliubov-Krilov) *Let X be a compact metric space and the mapping $f\colon X \to X$ be continuous. Then there is a probability measure μ on the Borel σ-algebra $\mathcal{B}(X)$ with respect to which f is measure preserving.*

Proof Consider the Banach space $C(X)$ of continuous real-valued functions on X with the maximum norm. Since X is a compact metric space, Borsuk's Theorem tells us that $C(X)$ is separable. Let η be any Borel probability measure on $\mathcal{B}(X)$. Define the sequence $\{\psi_n\}$ of linear functionals on $C(X)$ by

$$\psi_n(g) = \int_X \left[\frac{1}{n} \sum_{k=0}^{n-1} g \circ f^k \right] d\eta \text{ for all } n \in \mathbf{N} \text{ and } g \in C(X). \tag{23}$$

Observe that

$$|\psi_n(g)| \leq \|g\|_{\max} \text{ for all } n \in \mathbf{N} \text{ and } g \in C(X).$$

Thus $\{\psi_n\}$ is a bounded sequence in $[C(X)]^*$. Since the Banach space $C(X)$ is separable, we infer from Helly's Theorem that there is a subsequence $\{\psi_{n_k}\}$ of $\{\psi_n\}$ that converges, with respect to the weak-$*$ topology, to a bounded functional $\psi \in [C(X)]^*$, that is,

$$\lim_{k\to\infty} \psi_{n_k}(g) = \psi(g) \text{ for all } g \in C(X).$$

Therefore,

$$\lim_{k\to\infty} \psi_{n_k}(g \circ f) = \psi(g \circ f) \text{ for all } g \in C(X).$$

However, for each k and $g \in C(X)$,

$$\psi_{n_k}(g \circ f) - \psi_{n_k}(g) = \frac{1}{n_k} \left[\int_X [g \circ f^{n_k+1} - g] \, d\eta \right].$$

Take the limit as $k \to \infty$ and conclude that

$$\psi(g \circ f) = \psi(g) \text{ for all } g \in C(X). \tag{24}$$

Since each ψ_n is a positive functional, the limit functional ψ also is positive. The Riesz-Markov Theorem tells us that there is a Borel measure μ for which

$$\psi(g) = \int_X g \, d\mu \text{ for all } g \in C(X).$$

We infer from (24) that

$$\int_X g \circ f \, d\mu = \int_X g \, d\mu \text{ for all } g \in C(X).$$

According to Proposition 10, f is measure preserving with respect to μ. Finally, for the constant function $g = 1$, $\psi_n(g) = 1$ for all n. Therefore, $\psi(g) = 1$, that is, μ is a probability measure. $\square$

Proposition 13 *Let $f \colon X \to X$ be a continuous mapping on a compact metric space X. Define $\mathcal{M}_f$ to be the set of probability measures on $\mathcal{B}(X)$ with respect to which f is measure preserving. Then a measure μ in $\mathcal{M}_f$ is an extreme point of $\mathcal{M}_f$ if and only if f is ergodic with respect to μ.*

Proof First suppose that μ is an extreme point of $\mathcal{M}_f$. To prove that f is ergodic, we assume otherwise. Then there is a Borel subset A of X that is invariant under f with respect to μ and yet $0 < \mu(A) < 1$. Define

$$\nu(E) = \mu(E \cap A)/\mu(A) \text{ and } \eta(E) = \mu(E \cap [X \sim A])/\mu(X \sim A) \text{ for all } E \in \mathcal{B}(X).$$

Then, since $\mu(X) = 1$,

$$\mu = \lambda \cdot \nu + (1 - \lambda) \cdot \eta \text{ where } \lambda = \mu(A).$$

Both ν and η are Borel probability measures on $\mathcal{B}(X)$. We claim that f is measure preserving with respect to each of these measures. Indeed, since f is measure preserving with respect to μ and A is invariant under f with respect to μ, for each $E \in \mathcal{B}(X)$,

$$\mu(E \cap A) = \mu(f^{-1}(E \cap A)) = \mu(f^{-1}(E) \cap f^{-1}(A)) = \mu(f^{-1}(E) \cap A).$$

Therefore, f is invariant with respect to ν. By a similar argument, it is also invariant with respect to η. Therefore ν and η belong to $\mathcal{M}_f$ and hence μ is not an extreme point of $\mathcal{M}_f$. Therefore f is ergodic.

Now suppose f is ergodic with respect to $\mu \in \mathcal{M}_f$. To show that μ is an extreme point of $\mathcal{M}_f$, let $\lambda \in (0, 1)$ and $\nu, \eta \in \mathcal{M}_f$ be such that

$$\mu = \lambda\nu + (1 - \lambda)\eta. \tag{25}$$

The measure ν is absolutely continuous with respect to μ. Since $\mu(X) < \infty$, the Radon-Nikodym Theorem tells us that there is a function $h \in L^1(X, \mu)$ for which

$$\nu(A) = \int_A h \, d\mu \text{ for all } A \in \mathcal{B}(X).$$

It follows from the Simple Approximation Theorem and the Bounded Convergence Theorem that

$$\int_X g \, d\nu = \int_X g \cdot h \, d\mu \text{ for all } g \in L^\infty(X, \mu). \tag{26}$$

Fix $\epsilon > 0$, and define $X_\epsilon = \{x \in X \mid h(x) \geq 1/\lambda + \epsilon\}$. We infer from (25) that

$$\mu(X_\epsilon) \geq \lambda \cdot \int_{X_\epsilon} h \, d\mu \geq (1 + \lambda \cdot \epsilon) \cdot \mu(X_\epsilon).$$

Hence $\mu(X_\epsilon) = 0$. Therefore, h and $h \circ f$ are essentially bounded on X with respect to μ. Hence, using (26), first with $g = h \circ f$ and then with $g = h$, and the invariance of f with respect to ν, we have

$$\int_X [h \circ f \cdot h] \, d\mu = \int_X h \circ f \, d\nu = \int_X h \, d\nu = \int_X h^2 \, d\mu.$$

We infer from this equality and the invariance of f with respect to μ that

$$\begin{aligned}\int_X [h \circ f - h]^2 \, d\mu &= \int_X [h \circ f]^2 \, d\mu - 2 \cdot \int_X h \circ f \cdot h \, d\mu + \int_X h^2 \, d\mu \\ &= 2 \cdot \int_X h^2 \, d\mu - 2 \cdot \int_X h \circ f \cdot h \, d\mu \\ &= 2 \cdot \int_X h^2 \, d\mu - 2 \cdot \int_X h^2 \, d\mu = 0.\end{aligned}$$

Therefore, $h \circ f = h$ almost everywhere $[\mu]$ on X. By the ergodicity of f and Proposition 11, there is a constant c for which $h = c$ almost everywhere $[\mu]$ on X. But μ and ν are probability measures and hence

$$1 = \nu(X) = \int_X h \, d\mu = c \cdot \mu(X) = c.$$

Hence $\mu = \nu$ and thus $\mu = \eta$. Therefore μ is an extreme point of $\mathcal{M}_f$. □

Theorem 14 *Let $f \colon X \to X$ be a continuous mapping on a compact metric space X. Then there is a probability measure μ on the Borel σ-algebra $\mathcal{B}(X)$ with respect to which f is ergodic.*

Proof Let $\mathcal{R}adon(X)$ be the Banach space of signed Radon measures on $\mathcal{B}(X)$ and the linear operator $\Phi \colon \mathcal{R}adon(X) \to [C(X)]^*$ be defined by

$$\Phi(\mu)(g) = \int_X g \, d\mu \text{ for all } \mu \in \mathcal{R}adon(X) \text{ and } g \in C(X).$$

The Riesz-Kakutani Representation Theorem for the dual of $C(X)$ tells us that Φ is a linear isomorphism of $\mathcal{R}adon(X)$ onto $[C(X)]^*$. Define $\mathcal{M}_f$ to be the set of probability measures on $\mathcal{B}(X)$ with respect to which f is measure preserving. Then the measure μ is an extreme point of $\mathcal{M}_f$ if and only if $\Phi(\mu)$ is an extreme point of $\Phi(\mathcal{M}_f)$. Therefore, by the preceding proposition, to prove the theorem we must show that the set $\Phi(\mathcal{M}_f)$ possesses an extreme point. According to the Bogoliubov-Krilov Theorem, $\mathcal{M}_f$ is non-empty. A consequence of the Krein-Milman Theorem, Corollary 13 of the preceding chapter, tells us that $\Phi(\mathcal{M}_f)$ possesses an extreme point provided it is bounded, convex, and closed with respect to the weak-$*$ topology. The Riesz-Markov Theorem tells us that Φ defines an isomorphism of Radon measures onto positive functionals. The positive functionals are certainly weak-$*$ closed, as are the functionals that take the value 1 at the constant function 1. According to Proposition 11, a functional $\psi \in [C(X)]^*$ is the image under Φ of a measure that is invariant with respect to f if and only if

$$\psi(g \circ f) - \psi(g) = 0 \text{ for all } g \in C(X).$$

Fix $g \in C(X)$. Evaluation at the function $g \circ f - g$ is a linear functional on $[C(X)]^*$ that is continuous with respect to the weak-$*$ topology and therefore its kernel is weak-$*$ closed. Hence the intersection

$$\bigcap_{g \in C(X)} \{\psi \in [C(X)]^* \mid \psi(g \circ f) = \psi(g)\}$$

also is a weak-$*$ closed set. This completes the proof of the weak-$*$ closedness of $\Phi(\mathcal{M}_f)$ and also the proof of the theorem. $\square$

Asymptotic averaging phenomena were originally introduced in the analysis of the dynamics of gases. One indication of the significance of ergodicity in the study of such phenomena is revealed in the statement of the following theorem. Observe that the right-hand side of (27) is independent of the point $x \in X$.

Theorem 15 *Let T be a measure preserving transformation on the probability space $(X, \mathcal{M}, \mu)$. Then T is ergodic if and only if for every $g \in L^1(X, \mu)$,*

$$\lim_{n\to\infty}\left[\frac{1}{n}\sum_{k=0}^{n-1} g(T^k(x))\right] = \frac{1}{\mu(X)}\int_X g\,d\mu \text{ for almost all } x \in X. \tag{27}$$

A proof of this theorem may be found in the books *Introduction to Dynamical Systems* [BS02] by Michael Brin and Garrett Stuck and *Lectures on Ergodic Theory* [Hal06] by Paul Halmos. These books also contain varied examples of measure preserving and ergodic transformations.

PROBLEMS

26. Does the proof of the Bogoliubov-Krilov Theorem also provide a proof in the case X is compact Hausdorff but not necessarily metrizable?
27. Let $(X, \mathcal{M}, \mu)$ be a finite measure space and $T: X \to X$ a measurable transformation. For a measurable function g on X, define the measurable function $U_T(g)$ by $U_T(g)(x) = g(T(x))$. Show that T is measure preserving if and only if for every $1 \leq p < \infty$, U_T maps $L^p(X, \mu)$ into itself and is an isometry.
28. Suppose that $T: \mathbf{R}^n \to \mathbf{R}^n$ is linear. Establish necessary and sufficient conditions for T to be measure preserving with respect to Lebesgue measure on $\mathbf{R}^n$.
29. Let $S^1 = \{z = e^{i\theta} \mid \theta \in \mathbf{R}\}$ be the circle with the group operation of complex multiplication and μ be Haar measure on this group (see Problem 24). Define $T: S^1 \to S^1$ by $T(z) = z^2$. Show that T preserves μ.
30. (Poincaré Recurrence) Let T be a measure preserving transformation on a finite measure space $(X, \mathcal{M}, \mu)$ and the set A be measurable. Show that for almost all $x \in X$, there are infinitely many natural numbers n for which $T^n(x)$ belongs to A.

Bibliography

[Bir73] Garrett Birkhoff, *A Source Book in Classical Analysis*, Harvard University Press, 1973.

[BM97] Garrett Birkhoff and Saunders MacLane, *A Survey of Modern Algebra*, AK Peters, Ltd., 1997.

[BS02] Michael Brin and Garrett Stuck, *Introduction to Dynamical Systems*, Cambridge University Press, 2002.

[B07] V.I. Bogachev, *Measure Theory*, Vol 1, Springer-Verlag, 2007.

[BS18] Theo Bühler and Dietmar Salamon, *Functional Analysis*, Graduate Studies in Mathematics, Vol 191, American Mathematical Society, 2018.

[CC74] Micha Cotlar and Roberto Cignoli, *An Introduction to Functional Analysis*, Elsevier, 1974.

[DS71] Nelson Dunford and Jacob Schwartz, *Linear Operators Volumes I and II*, Interscience, 1958–71.

[Eva98] Lawrence C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, Vol 19, American Mathematical Society, 1998.

[EG15] Lawrence C. Evans and Ronald G. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, 2015.

[Fit09] Patrick M. Fitzpatrick, *Advanced Calculus*, Pure and Applied Undergraduate Texts, American Mathematical Society, 2009.

[Hal50] Paul R. Halmos, *Measure Theory*, Van Nostrand, 1950.

[Hal06] ______, *Lectures on Ergodic Theory*, American Mathematical Society, 2006.

[Haw01] Thomas Hawkins, *Lebesgue's Theory of Integration: Its Origins and Development*, American Mathematical Society, 2001.

[Lax97] Peter D. Lax, *Linear Algebra*, John Wiley and Sons, 1997.

[Lax02] ______, *Functional Analysis*, Wiley-Interscience, 2002.

[Lit41] J.E. Littlewood, *Lectures on the Theory of Functions*, Oxford University Press, 1941.

[Meg98] Robert E. Megginson, *An Introduction to Banach Space Theory*, Graduate Texts in Mathematics, Springer, 1998.

[Pie07] Albrecht Pietsch, *History of Banach Spaces and Linear Operators*, Birkhaüser, 2007.

[RSN90] Frigyes Riesz and Béla Sz.-Nagy, *Functional Analysis*, Dover, 1990.

[Sak64] Stanislaw Saks, *Theory of the Integral*, Dover, 1964.

[Sim03] George F. Simmons, *Introduction to Topology and Analysis*, International Series in Pure and Applied Mathematics, McGraw-Hill, 2003.

[Sim15] ______, *Gems of Calculus: Brief Lives and Memorable Moments*, AMS/MAA (Spectrum), Vol 95, 2015.

[Tao14] Terence Tao, *Analysis 1*, *Texts and Readings in Mathematics,* Hindustan Book Agency, 2014.

[vN50] John von Neumann, *Functional Operators, Volume I: Measures and Integrals*, Annals of Mathematics Studies, Princeton University Press, 1950.

[vN91] ______, *Invariant Measures*, American Mathematical Society, 1991.

[WZ77] Richard L. Wheeden and Antoni Zygmund, *Measure and Integral*, Marcel Dekker, 1977.

Index

Absolute continuity
of a function, 103
Fundamental Theorem of Calculus, 107
of an indefinite integral, 108
of a measure, 220
Radon-Nikndym Theorem, 220
von Neumann Composition Theorem, 114
Adjoint, 185, 404
Adjoint representation, 455
Alaoglu's Theorem, 383–384, 452, 455
Alexandroff one-point compactification, 426–428
Algebra
of functions, 331
of sets, 36
Annihilator, 386
Approximation Theorems
continuous approximation of measurable functions, 61, 206
simple approximation of measurable functions, 56–57, 203–204
smooth approximation in $L^p(\mathbf{R}^n), 1 \leq p < \infty$, 243–247
Stone-Weierstrass Theorem, 331
Weierstrass Approximation Theorem, 331
Archimedean property of $\mathbf{R}$, 11–12
Arcwise connected topological spaces, 321
Arzelà Bounded Convergence Theorem, 63, 69
Arzelà-Ascoli Theorem, 155, 285–289, 410, 415
Ascending sequence of sets, 18
Ascoli-Young Lemma, 163
Average value function, 97
Axiom of Choice, 5–6

Baire Category Theorem, 295
Baire measures, 446–451
Baire σ-algebra, 446, 448
Banach Contraction Principle, 290–291
Banach limit, 364
Banach space, 130, 339, 349–352
Banach-Saks-Steinhaus Theorem, 353
Closed Graph Theorem, 350–352
closed unit ball, 382, 388
Hilbertable, 395
Open Mapping Theorem, 349–350
reflexive, 367–370
reflexivity and weak compactness, 385
Banach-Saks Theorem, 151, 401
Banach-Saks-Steinhaus Theorem, 353
Base for a topology, 306
Beppo Levi's Lemma, 153
Beppo Levi's Theorem, 72, 211
Bessel's Inequality, 397–399, 409
Bidual of a normed linear space, 360
Bogoliubov-Krilov Theorem, 465–466
Bolzano-Weierstrass Theorem, 21, 277
Borel measurable, 55
Borel measure, 160, 163, 190–193, 281–282, 423, 432, 447, 460
Borel measure space, 178
Borel σ-algebra, 39, 180, 192, 423, 468
Borel set, 18–20, 163
Borel-Cantelli Lemma, 44, 162
Borsuk's Theorem, 334, 442, 466
Boundary point, 295
Bounded Convergence Theorem, 68–69, 460
Bounded linear functionals, 137
Bounded linear operator, 340–341, 353, 402, 409, 411, 423

Canonical representation of a simple function, 56, 203
Cantor Intersection Theorem, 272
Cantor sets, 46–49
Cantor-Lebesgue function, 46–49, 163
Carathéodory extension, 168, 177, 226
Carathéodory measure, 166
Carathéodory outer-measures, 193–197, 433
Carathéodory strategy, 164
Carathéodory-Hahn Theorem, 168–173
Cauchy Convergence Criterion, 21–22
Cauchy sequence, 394
in metric space, 271, 278
in normed linear space, 242
in $\mathbf{R}$, 21, 287
rapidly Cauchy sequence, 130–131, 245
Cauchy-Schwarz Inequality, 127, 393
Cavalieri's Principle, 233–237
Change of Variables Theorem, 110–112
Characteristic function, 56
Characterization of Compactness Theorem, 276–277
Chebychev's Inequality, 72, 208
Chordal Slope Lemma, 119
Closed
ball, 266
linear complement, 350
linear span, 338
set in a topological space, 307–308
unit ball, 266, 288, 339, 346, 382, 388
Closed Graph Theorem, 350–352

Closure of a set
in the real numbers, 17
in a topological space, 307
Codimension, 356
Compact group, 452–453
Compact metric spaces, 274–280
Characterization of Compactness Theorem, 276–277
Extreme Value Theorem, 26, 277–278
finite intersection property, 274
Lebesgue Covering Lemma, 278–279
open cover, 274
sequentially, 275
totally bounded, 275
Compact open topology, 320
Compact operators, 409–411
Compact topological spaces, 317–320
finite intersection property, 317
sequentially compact, 318
Compactly supported functions, 175–176, 244, 246, 248–249, 251
Compactness Criteria
Arzelà-Ascoli Theorem, 297
Bolzano-Weierstrass Theorem, 21, 277
Heine-Borel Theorem, 18, 277
for subsets of a metric space, 275
Tychonoff Product Theorem, 329
for weak compactness in a Hilbert space, 400
for weak compactness in a reexive Banach space, 382
for weak compactness in $L^p, 1 \leq p < \infty$, 148, 259
for weak-* compactness, Alouglu's Theorem, 383
Compactness Theorem
Alaoglu's Theorem, 383–384
Arzelà-Ascoli Theorem, 287
Bolzano-Weierstrass Theorem, 21, 277
Dunford-Pettis Theorem, 302–303
Eberlein-Šmulian Theorem, 387–388
Heine-Borel Theorem, 18, 277
Helly's Theorem, 140
Kakutani's Theorem, 384–386
Schauder's Theorem, 410
Tychonoff Product Theorem, 329
Completenes of a measure space, 163
of every Carathéodory extension, 171
of every product measure, 227
Completeness Axiom, 9, 21
Completeness of a metric space
Axiom in **R**, 9
Cantor Intersection Theorem, 272
of $L^p, 1 \leq p \leq \infty$, 133, 246
Compositions, 51–55, 114
Conjugate
of $f \in L^p, 1 \leq p < \infty$, 126
of $p, 1 \leq p \leq \infty$, 125
Connected topological spaces, 320–322
Continuity of measure, 43–44, 159
Continuous functions, 25–29
Continuous mappings, 269–271, 313–315
Contracting sequence, 272
Convergence Theorem
Arzelâ Bounded, 63, 69
Beppo Levi Convergence Theorem, 72, 211
Bounded, 68–69
Dominated, 76–77, 215
Monotone, 72, 210
Vitali, 83, 84, 217, 244
Convex functions, 117–122
Banach-Saks Theorem, 151–153
Beppo Levi's Lemma, 153
Chordal Slope Lemma, 119
Jensen's Inequality, 120–121
left-hand and right-hand derivatives, 119
minimization, 150–155
non-negative second derivative, 118
supporting line, 120
Convex set, 375–379
Hyperplane Separation Lemma, 375–376
Hyperplane Separation Theorem, 376–377
Convolution, 243–244, 247
Countable additivity
of integration, 78–79, 214
of measure, 43, 159
Countable compact topological spaces, 319
Countably monotone set-function
measure, 161
outer-measure, 33, 164
premeasure, 168
Counting measure, 30, 160
Cover, 17
finite, 18
open, 17
Cumulative distribution function, 192

Darboux sums, 62, 89
De Morgan's Identities, 4
Decomposition Theorems
Hahn, 199, 254
Jordan, 200, 258, 259
Dense set
in metric space, 297
in **R**, 125
in a topological space, 311
Derivative
Dini, 119
left-hand and right-hand, 120

Derivative (*Continued*)
Radon-Nikodym, 222
upper and lower, 95
Descending sequence of sets, 18
Diagonal linear operator, 185–186
Differentiability, of monotone function, 94–99
Dini derivative, 95
Dini's Theorem, 58
Dirac measure, 160
Direct sum, 184, 339
Dirichlet function, 63
Discrete
metric, 264
topology, 306
Divided difference function, 97
Dominated Convergence Theorem, 76–77, 215
Dunford-Pettis Theorem, 155, 302–303

Eberlein-Šmulian Theorem, 387–388
ϵ-dilation, 31, 160, 176, 314
Egoroff's Theorem, 59–60, 204–205
Eiegenvalue, 411
ϵ-net, 275
Equicontinuous, 286
Equisummable, 288
Equivalence class, 5
Equivalence relation, 5–6
Equivalent norms, 338
Ergodicity, 463–469
Essential upper bound, 124, 238
Essentially bounded function, 124, 238–239
Euclidean basis for $\mathbf{R}^n$, 183
Euclidean inner-product, 174, 392
Euclidean metric, 344
Euclidean norm, 174, 392
Euclidean space, 174–181, 233–237, 275, 454
Excision property, 35, 161
Extension Theorems
Hahn-Banach Theorem, 361–367
Locally compact extension property, 430
Tietze Extension Theorem, 205
Exterior point, 295
Extreme point, 380
Extreme subset, 379
Extreme Value Theorem, 26, 277–278

Fatou's Lemma, 71–72, 209–210
Finite codimension, 350
Finite dimensional normed linear spaces, 343–346
Finite intersection property, 274, 317
Finite measure, 13, 159
Finite rank, 409
Finitely supported function, 57, 202
First category, set of the, 297
First countable space, 311
Fixed-point of mapping, 289
Fredholm Alternative, 417
Fredholm operators, 414–419
Fubini's Theorem, 228
Fubini-Tonelli Lemma, 228
Function
absolutely continuous, 103
bounded variation, 100
convex, 118
integrable, 75, 213
Lipschitz, 25
measurable, 52, 202
monotone, 92
simple, 56
total variation, 101

Gauge functional, 365, 375
General Dominated Convergence Theorem, 77
General linear group, 452–455, 453
General Pythagorean Identity, 397
General Vitali Convergence Theorem, 84–85
Goldstine's Theorem, 386–387
Greatest lower bound, 9

Haar measure, 452, 461
Hahn Decomposition Theorem, 199, 254
Hahn-Banach Lemma, 362
Hahn-Banach Theorem, 361–367
Hahn's Decomposition Theorem, 200
Hahn's Lemma, 198–199
Hamel basis, 356–357
Hausdorff dimension, 196
Hausdorff space, 382–383, 424–425, 447
locally compact, 449
σ-compact, 450
Hausdorff α-dimensional measure, 193–197
Hausdorff measures, 193–197
Hausdorff separation property, 309
Heine-Borel Theorem, 18, 177
Hellinger-Toplitz, 407
Helly's Theorem, 141–142
Hilbert space, 222, 392, 394
orthonormal sequence, 398
reflexivity, 402
separability, 398
Hilbertable Banach space, 397
Hilbert-Schmidt Lemma, 412
Hilbert-Schmidt Theorem, 392, 411–414, 413
Hölder's Inequality, 126, 242
Hollow set, 295
Homeomorphism, 314
Hyperplane, 356, 375
Hyperplane Separation Lemma, 375–376
Hyperplane Separation Theorem, 376–377, 385
Indefinite integral, 107–114

Inequalities
- Cauchy-Schwarz Inequality, 127, 393
- Chebychev's Inequality, 72, 208
- Hölder's Inequality, 126, 242
- Jensen's Inequality, 121
- Minkowski's Inequality, 127, 242
- Minkowski's Integral Inequality, 247
- triangle inequality, 9, 123, 263
- Young's Convolution Inequality, 250
- Young's Inequality, 126

Infinite dimensional normed linear spaces, 343–347
Inner regular set, 469
Inner regularity, 432
Inner-product, 393
Inner-product space, 393
Integration
- indenite, 107–114
- Lebesgue, over subsets of **R**, 62–79
- Over general measure spaces, 207–219
- Riemann, 57, 62–64, 92, 227
- Stieltjes, 442

Integral Comparison Test, 214
Integral count, 31, 180
Interior point, 20, 268, 295
Intermediate value property, 320–321
Intermediate Value Theorem, 26–27
Invariant measures, 452–469
- Bogoliubov-Krilov Theorem, 465–466
- ergodicity, 463–469
- general linear group, 452–455
- Kakutani's Fixed-Point Theorem, 455–459
- measurable transformations, 463–469
- measure preserving transformations, 463
- Radon probability measure, 461
- von Neumann's Theorem, 460–463

Invariant subset, 456
Invertible operator, 417
Isometric isomorphism, 352
Isometric metric spaces, 264
Isometry, 264
Isomorphism
- isometric, 342
- between linear spaces, 341

Jensen's Inequality, 120–121
Jordan content, 34
Jordan Decomposition, 258, 259
Jordan Decomposition Theorem, 200
Jordan's Lemma, 101
Jordan's Theorem, 99–102, 442, 444
Jump discontinuity, 27

Kakutani's Fixed-Point Theorem, 455–459
Kantorovitch Representation Theorem, 257–260
Krein-Milman Lemma, 380
Krein-Milman Theorem, 379–382, 441, 452

Lax-Milgram Lemma, 402–404
Least upper bound, 9
Lebesgue Covering Lemma, 278–279, 280
Lebesgue Decomposition Theorem, 222–223
Lebesgue Differentiation Theorem, 96–97, 108, 115
Lebesgue integrable, 65
Lebesgue integral, 65
- Fundamental Theorem of Calculus, 107, 110

Lebesgue integration, 62–80
- Arzelà Bounded Convergence Theorem, 69
- Beppo Levi's Theorem, 72
- Bounded Convergence Theorem, 68–69
- bounded function, 65
- continuity of integration, 78–80
- convergence in mean and measure, 85–88
- countable addivirty, 78–80
- Dominated Convergence Theorem, 76–77
- Fatou's Lemma, 71–72
- General Dominated Convergence Theorem, 77
- general Lebesgue integral, 74–78
- General Vitali Convergence Theorem, 84–85
- indefinite integral, 108
- integral comparison test, 75
- integral of a non-negative measurable function, 70–74
- linearity, 65, 68
- linearity of integration, 75–76
- Monotone Convergence Theorem, 72
- monotonicity, 65, 68
- Riemann integrability, 88–91
- Riemann integral, 62–64
- Riesz's Theorem, 85–88
- Simple Approximation Lemma, 66–67
- uniform integrability, 81–85
- Vitali Convergence Theorem, 83–84

Lebesgue measurability, 182–190
Lebesgue measurable, 29–30, 179–180, 246
Lebesgue measure, 29–50, 39, 233–237
- continuity, 43–44
- countable additivity, 43
- definition, 29, 30
- Euclidean space, 174–181, 178–179
- outer-measure, 30, 31–34
- on **R**, 160, 178
- regularity, 39

Lebesgue measure zero, 184
Lebesgue number, 278
Lebesgue outer-measure, 178
Lebesgue's Differentiation Theorem, 96
Limit inferior, 23

Limit superior, 23
Lindelöf property, 312
Lindenstrauss-Tzafriri Theorem, 395
Linear complement, 350
Linear functional, 137, 355–361
  bounded, 137–143
  positive, 434–439
Linear operator, 185
  bounded, 340–341, 353, 402, 409, 411, 423
  bounded linear, 423
  diagonal, 185–186
  orthogonal, 185
Linear sequence spaces, 125
Lipschitz constant, 182, 344
Lipschitz function, 25–29, 120
Lipschitz mapping, 182
Littlewood's three principles, 59
Locally compact extension property, 430
Locally compact separation property, 430
Locally compact topological spaces, 424–428
  Alexandroff one-point compactification, 426–428
  Hausdorff space, 424–425
  partitions of unity, 426
  point of infinity, 426
  Urysohn's Lemma, 424
Locally convex topological vector space, 370–375
Lower Darboux sums, 62, 89
Lower derivative, 95
Lusin's Theorem, 61, 205–206, 244

Maximal member, 6, 329, 363
Maximum norm, 125
Mazur's Theorem, 377, 388
Meager set, 297
Mean Value Theorem, 95, 118, 175, 184, 249
Measurable functions
  composing measurable with continuous, 54, 114, 202
  sequential pointwise limits, 56, 203
  sequential simple approximation, 57, 203
  sums and products, 52, 202
Measurable partition, 37, 159
Measurable sets, 34, 164
  characterization of Lebesgue measurability, 40, 41, 167
  characterization of product measurability, 230
  measurable partition, 37, 159
Measurable transformations, 463–469
Measure, 163
  absolutely continuous, 220
  Borel, 160
  Carathéodory, 166
  continuity, 43, 161, 161–162
  countable additivity, 42, 159
  countable monotonicity, 161
  counting, 160
  excision, 35, 161
  finite, 159
  Hausdorrff, 193–197
  Lebesgue, 160
  outer-measure, 164–167
  premeasure, 168
  product, 233
  σ-finite, 159
  signed, 197
Measure preserving transformation, 463
Metric spaces, 263
  Arzelà-Ascoli Theorem, 285–289
  Baire Category Theorem, 295–299
  Banach Contraction Principle, 290–291
  compact, 274–280
  complete, 271–274
  continuons mapping between, 269–271
  convergent sequences, 266–267
  discrete metric, 264
  examples of, 263–265
  first category, 297
  ϵ-δ criterion for continuity, 269
  isometric, 264
  Nikodym, 299–304
  Non-linear Uniform Boundedness Theorem, 296–297
  nowhere dense, 297
  Picard Local Existence Theorem, 292–293
  point of closure in, 266
  second category, 297
  separable, 280–284
  triangle inequality, 263–264
  uniform metric, 285
Metric topology, 306
Metrizability
  Borsuk's Theorem, 334
  Urysohn's Metrization Theorem, 312
  weak topology on a ball B, 389–390
  weak topology on X, 315, 389–390
  weak-* topology on a ball B^*, 389–390
  weak-* topology on X, 389–390
Metrizable topological spaces, 312, 389–391
Minimization Theorems
  for bounded linear functionals, 368
  for continuous functions on a compact metric space, 266
  for continuous, convex functions, 378
  for the distance function in a Hilbert space, 394
  for the eigenvalues of symmetric operators, 412
Minkowski functional, 375
Minkowski's Inequality, 123, 240

Minkowski's Integral Inequality, 247
Monotone Convergence Criterion, 21
Monotone Convergence Theorem, 72, 210
Monotone functions, 94–96
absolutely continuous, 109
continuity property, 92–94
differentiability property, 94–99
Lebesgue Differentiation Theorem, 96
Moore Plane, 309, 312
Mutually singular, 199, 222

Natural embedding, 360, 368, 384–385
n-dimensional Euclidean space, 175
Negative part of a function, 213
Negative set, 198, 199
Neighborhood of a point
in metric space, 265
in a topological space, 309
Nested Set Theorem, 18–19, 26, 47
Nikodym metric space, 264, 299–304
Non-degenerate interval, 15
Non-measurable set, 45, 182
Non-negative symmetric operator, 405
Norm, 123, 264, 338
Normal separation property, 310
Normally ascending collection, 324–325
Normed linear space, 124
Nowhere dense set, 297

Open ball, 175, 265
Open cover, 274
Open mapping, 347–350
Open Mapping Theorem, 297, 348–350, 407, 453
Open sets, 16–17, 265, 305
Open unit ball, 339
Operator
bounded linear, 340–341, 353, 402, 409, 411
diagonal linear, 185–186
Fredholm, 414–419
invertible, 417
linear, 340–343
non-negative symmetric, 405
orthogonal linear, 185
positive-definite, 402
self-adjoint, 405
symmetric, 404–408
translation, 245–246
Orthogonal, 394
complement, 395
linear operator, 185
operators, 408
projection, 395
projection sequence, 415
Orthonomal bases, 397–399

Outer regularity, 432
Outer-measure, 30, 31–34, 164–167
Carathéodory, 193–197
countably monotone, 33
excision property, 35
of intervals, 32
monotonicity, 33
translation invariant, 32–33

Parallelepiped set, 189
Parallelogram Identity, 393–394
Parallelogram Law, 181
Parseval's Identities, 399
Partitions of unity, 426
Picard Local Existence Theorem, 292–293
Pigeonhole principle, 13
Point at infinity, 426
Point of closure
in metric space, 266
in the real numbers, 17
in a topological space, 307
Pointwise bounded sequence, 139, 286
Pointwise convergent sequence, 139
Polar decompositions, 408
Polarization Identity, 405–407
Positive homogeneity, 123
Positive part of a function, 213
Positive-definite operator, 402
Positively homogenous functional, 361
Premeasure, 168
Prime factorization theorem, 12
Principle of Mathematical Induction, 11
Probability measure, 460
Product measures
Cavalieri's Principle, 233–237
completeness, 227
Fubini's Theorem, 228
Fubini-Tonelli Lemma, 228–231
Tonelli's Theorem, 227–229
Product topology, 307, 328
Projection, 350
Proper mapping, 280
Pseudonorm, 340

Quadratic form, 406

Radamacher functions, 151
Radon measure, 430, 460
Radon premeasure, 432
Radon probability measure, 461
Radon-Nikodym derivative, 222
Radon-Nikodym Theorem, 219–223
Radon-Riesz Theorem, 151
Rapidly Cauchy sequence, 131–132, 245

Rayleigh quotient, 411
Real numbers, 16–18
  absolute value, 9
  Borel set, 18–20
  closed set, 18
  completeness axioms, 9
  extended, 9–10
  filed axioms, 7–8
  irrational numbers, 12
  natural numbers, 11–12
  open set, 16–17, 16–18
  positivity axioms, 8–9
  rational numbers, 11–12
  sequences, 20–24
  triangle inequality, 9
Real-valued functions
  continuous, 25–29
  Extreme Value Theorem, 26
  Intermediate Value Theorem, 26–27
Reflexive Banach space
  $C(K)$, only if K finite, 389
  Hilbert space, 402
  Kakutani's Theorem, 385
  $L^p, 1 \leq p < \infty$, 369
  weak compactness, 382
  weak sequential compactness, 382
Regular representation, 458–459, 460, 461
Regular separation property, 310
Regularity of Lebesgue Measure, 39, 180
Representation, 455
Representation Theorems for the Dual
  Kantorovich Representation Theorem, L^∞, 257–260
  Riesz Representation Theorem, $C[a, b]$, 444
  Riesz Representation Theorem, $L^p, 1 \leq p < \infty$, 147, 253
  Riesz-Fréchet Representation Theorem, H, 400
  Riesz-Kakutani Representation Theorem, $C(K)$, 441
Riemann integral, 62
  Arzelà's Bounded Convergence Theorem, 69
  Discrete formulation of Fundamental Theorem, 92
  Riemann integrability criterion, 89
Riemann Rearrangement Theorem, 24
Riemann-Lebesgue Lemma, 149–150
Riesz-Fischer Theorem, 133, 246
Riesz-Fréchet Representation Theorem, 400
Riesz-Kakutani Representation Theorem, 441
Riesz-Markov Theorem, 435
Riesz's Lemma, 345–346
Riesz's Theorem, 85–88, 288, 345–346, 409
Riesz-Schauder Theorem, 392, 414–419
Scalar product, 337
Schauder's Theorem, 410
Second category, set of the, 297
Second countable space, 311
Self-adjoint operators, 405
Semi-ring, 169
Separability, 397–399
  Hilbert space, 397–398
  in normed linear spaces, 135–136
  topological spaces, 311–312
Separation properties, 309–311
Sequential compactness
  in metric space, 275
  in topological space, 318
Sequential Convergence Modes
  in the mean, 85
  in measure, 86
  in a metric space, 266
  in a normed linear space, 124
  pointwise (Lusin's Theorem), 61, 205
  in a topological space, 311
  uniformly (Egoroff's Theorem), 59, 204
  weakly, 147, 382
Shifted product, 250–251
σ-algebra, 38–39
σ-compact topological spaces, 448
σ-finite measure, 159
Signed measure, 125, 197–200
  Hahn Decomposition Theorem, 199
  Hahn's Lemma, 198–199
  Jordan Decomposition Theorem, 200
Simple Approximation Lemma, 56, 203
Simple Approximation Theorem, 57, 204
Smooth approximation of the identity, 250
Sorgenfrey Line, 309
Space
  arcwise connected topological, 321
  Banach, 131, 339, 349–352, 367–370
  Borel measure, 178
  compact metric, 274–280
  compact topological, 317–320
  connected topological, 320–322
  countable compact topological, 319
  dual, 139, 355, 399–404
  Euclidean, 174–181, 233–237, 275, 454
  Hausdorff, 382–383, 424–425, 447, 449, 450
  Hilbert, 222, 392, 394, 398, 402
  Hilbertable Banach, 397
  infinite dimensional normed linear, 343–347
  inner-product, 393
  Lindelöf, 312
  locally compact topological, 424–428
  measure, 159–164
  metric, 263
  metrizable topological, 312, 389–391
  Nikodym metric, 264, 299–304

normed linear, 124–126, 238–243, 264, 337–340
topological, 305–322
Spectral Boundary Theorem, 407
Spectral bounds, 406
Step-function, 63, 66, 89, 91, 146, 151
Stieltjes integral, 442
Stone-Weierstrass Theorem, 331–336
Strong topology, 359
Subadditive functional, 361
Subbase, 307
Supporting line, 120
Supremum, 9
Symmetric difference of sets, 41
Symmetric mapping, 174
Symmetric operators, 404–408
Monotone Convergence Theorem, 408
Polarization Identity, 405–406
Spectral Boundary Theorem, 407

Taylor series, 249
Tietze Extension Theorem, 205
Tightness, 81–85
Tonelli's Theorem, 228
Topological group, 452
Topological spaces, 305–322
arcwise connected, 321
base for a topology, 306
Borsuk's Theorem, 334
closed sets, 307–308
compact, 317–320
connected, 320–322
continuous mappings, 313–316
countable compact, 319
defined, 305
dense set, 311
discrete topology, 306
first countable, 311
Hausdorff, 317–318
Hausdorrff separation property, 310
homeomorphism, 314
locally compact, 424–428
metric topology, 306
metrizability, 313–316
normal separation property, 310
open mapping, 347
open sets, 305
point of closure, 307
product topology, 307
regular separation property, 310
second countable, 311
separability, 311–312
separation properties, 309–311
σ-compact, 448
subbase for a topology, 307
subspace, 306
Tietze Extension Theorem, 325–326
trivial topology, 306
Tychonoff Product Theorem, 327–330
Tychonoff separation property, 309
Urysohn Metrization Theorem, 312, 326–327
Urysohn's Lemma, 323–325
weak topology, 314
Topological vector space, 373
Total variation, 99–104, 200, 257, 441
Total variation function, 101
Totally bounded metric spaces, 275
Totally ordered set, 6
Translation invariant measure, 32, 185
Translation operator, 245–246
Triangle inequality, 9, 123, 131, 263–264
Trivial topology, 306
Tychonoff Product Theorem, 327–330, 382–383
Tychonoff Separation Property, 309
Tychonoff space, 310

Ulam's Regularity Theorem, 282–283
Uniform absolute continuity, 300
Uniform Boundedness Theorems
for sequences non-linear continuous functions, 296
for sequences of bounded linear functionals, 143–144
for sequences of bounded linear operators, 353
Uniform integrability, 81–85, 104, 105, 107, 217, 302
Uniform metric, 285
Uniformly absolutely continuous sequence, 300
Uniformly bounded sequence, 64, 285, 289
Uniformly continuous function, 28, 300
Uniformly pointwise bounded sequence, 286
Upper Darboux sums, 62, 89
Upper derivative, 95
Urysohn Metrization Theorem, 312, 326–327
Urysohn's Lemma, 175–176, 251, 323–325, 424

Vitali Convergence Theorems, 83, 84, 217, 244
Vitali Covering Lemma, 94–96, 116
Vitali non-measurable set, 42, 182
Vitali Partition Theorem, 187–188
Vitali-Hahn-Saks Theorem, 301
Volume, 176
von Neumann
Composition Theorem, 114
integral count, 31, 180
measure on a compact group, 460
proof of the Radon-Nikodym Theorem, 403

Weak compactnesss, 388
Weak sequential compactness, 255–257, 386–389, 399–404
Weak sequential convergence, 147–150, 367–370
Weak topology, 314, 358–359, 382–391
Alaoglu's Theorem, 383–384
compactness and weak sequential compactness, 386–389
Eberlein-Šmulian Theorem, 387–388
Goldstine's Theorem, 386–387
metrizability, 389–391
reflexivity and weak compactness, 385
Weaker topology, 319, 358
Weakly convergent sequence, 149, 365, 401, 410–411
Weierstrass Approximation Theorem, 134, 141, 280, 331

Young's Convolution Inequality, 247
Young's Inequality, 126

Zermelo-Fraenkel Axioms, 5
Zorn's Lemma, 6, 329, 356, 380

This page intentionally left blank

This page intentionally left blank