# Real Analysis

THIRD EDITION

H. L. ROYDEN

# Real Analysis

# Real Analysis

THIRD EDITION

**H. L. ROYDEN**
Stanford University

**Macmillan Publishing Company**
New York

**Collier Macmillan Publishing**
London

Printed in the United States of America



Macmillan Publishing Company
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Inc.

**Library of Congress Cataloging-in-Publication Data**

Royden, H. L.
Real analysis.

Bibliography: p.
Includes indexes.
1. Functions of real variables. 2. Functional analysis. 3. Measure theory. I. Title.
QA331.5.R6 1988 515.8 86-33216
ISBN 0-02-404151-3 (Hardcover Edition)
ISBN 0-02-946620-2 (International Edition)

PRINTING 9 YEAR 4 5

ISBN 0-02-404151-3

*To*
*John Slavens*

# Preface to the Third Edition

In the twenty years since this work was last revised it has contributed to the education of several generations of students, despite flaws in its treatment of Baire measure and the omission of invariant measures. I am therefore pleased by the opportunity of presenting a new edition ameliorating these shortcomings. Because of the difficulty of recapturing a point of view from the time of original composition, when the material was new to me, I have confined significant revisions to the theory of locally compact spaces and the study of measures on topological spaces. Elsewhere the alterations consist largely of minor improvements and the addition of new problems.

Part One is almost unchanged, the most notable change being the treatment of the Minkowski and Hölder inequalities. This treatment seems to me more natural, and it immediately gives the reversed inequalities for $0 < p < 1$. There are also relatively few changes in Chapters 11 and 12, the basic chapters on measure and integration. The principal additions consist of a section on integral operators and a section on Hausdorff measure added to Chapter 12.

Part Two sees somewhat more reorganization and extension: The sections on compact metric spaces and the Ascoli theorem have been moved from the chapter on compact spaces to the chapter on metric spaces, making these topics independent of the general theory of topological spaces. The material on Baire Category has been expanded with an indication of the principles used in applying this theory to proofs.

Chapter 8 on topological spaces is virtually unchanged, but Chapter 9 has been largely rewritten to expand the treatment accorded to locally compact spaces. Properties of locally compact spaces needed for measure theory are developed, and the concepts of paracompactness, exhaustion, and $\sigma$-compactness are discussed at length in the context of locally compact Hausdorff spaces. There is also a section on manifolds and the significance of paracompactness for them.

Chapter 10 is again little changed, with the exception of some material on convexity.

The material on Baire and Borel measures in locally compact spaces has been entirely rewritten. The treatment in the 2nd edition was seriously flawed. I do not think there were any actual misstatements of fact in the theorems and propositions, but the text was misleading, and a number of the statements in the problems were false. The difficulties, as in some other published treatments of measures in spaces that are not $\sigma$-compact, arose from problems of regularity. They were caused in my case by a misguided attempt to avoid talking about regularity directly. The current treatment meets these problems face-to-face and shows that one can have Baire (or Borel) measures that are inner regular or that are quasi regular but not always ones that are both. Included with this material is a direct proof of the Riesz–Markoff Theorem on the structure of positive linear functionals on $C_0(X)$. This proof is independent of the Daniell integral, allowing the chapter on the Daniell integral to be relegated to the end of the book.

Chapter 15 on automorphisms of measure spaces has been largely rewritten so that it now gives an extended treatment of Borel measures on complete separable metric spaces. I have tried to please my friend George Mackey by stressing the equivalence of these spaces with certain standard measure spaces, essentially Lebesgue measure on an interval of **R**.

The present edition contains a new chapter on invariant measures in Part Three. This topic was omitted from earlier editions because I was unsatisfied with the usual development of the theory. I thought the standard presentations of Haar measure awkward in the manner of their use of the Axiom of Choice to assure additivity, and I wanted to use instead a suitable generalization of the notion of limit along the lines used by Banach in the separable metric case. I also believed the proper context for invariant measures to be that of a transitive group of homeomorphisms on a locally compact space $X$. Thus the topology should be on the homeogeneous space $X$, with the group of homeomorphisms an abstract group without topology.

Of course, the group must satisfy some conditions in order that there should be a Baire measure on $X$ invariant under the group. I introduce a property, called topological equicontinuity, and show that it suffices for the existence of an invariant measure. The unicity of such measures is considered in a number of particular cases, including that of locally compact topological groups. We also consider groups of diffeomorphisms and introduce the Hurwitz invariant integral when it exists. This integral has the advantage that one can give specific formulas for the integrand in many cases.

When this book was originally planned and written, the theory of Lebesgue integration was generally considered to be graduate level material, and the book was designed to be covered in a year-long course for first-year graduate students. Since that time the undergraduate curriculum has tended to include material on Lebesgue integration for advanced students, and this book has found increasing use at this level. The material presented here is of varying difficulty and sophistication. I have tried to arrange the chapters with considerable independence so the book will be useful for a variety of courses. One possibility for a short course is to cover Part One and Chapters 11 and 12. This gives a thorough treatment of integration and differentiation on $\mathbf{R}$ together with the fundamentals of abstract measure and integration. This could be supplemented by Chapter 7, covering metric spaces, and some topics on Banach spaces from Chapter 10. For students who are already familiar with basic measure and integration theory as well as the elements of metric spaces, one could construct a short course on measure and integration in topological spaces covering Chapters 9, 13, 14, and 15 with Chapters 7 and 8 as background material.

Were I writing the chapter on set theory today, I would give it a different tone, emphasize the various philosophical points of view about the foundations of mathematics and warn against endowing sets with reality and significance apart from the formal system in which they are embedded. The temptation to rewrite the chapter along these lines has been resisted, but I hope the readers of this book will ultimately read some of the many books on the foundations of mathematics before coming to a fixed opinion on the nature of infinite sets.

I wish to thank all of the diligent readers who have given me corrections and improvements over the last twenty years. My special thanks go to Jay Jorgenson and Hala Khuri for proofreading and checking the work in galley proof and to Elizabeth Arrington and

Elizabeth Harvey for turning large amounts of handwritten corrections and material into copy suitable for the printer.

H. L. R.

*Stanford, California*
*July 1987*

# Preface to the Second Edition

This book is the outgrowth of a course at Stanford entitled "Theory of Functions of a Real Variable," which I have given from time to time during the last ten years. This course was designed for first-year graduate students in mathematics and statistics. It presupposes a general background in undergraduate mathematics and specific acquaintance with the material in an undergraduate course on the fundamental concepts of analysis. I have attempted to cover the basic material that every graduate student should know in the classical theory of functions of a real variable and in measure and integration theory, as well as some of the more important and elementary topics in general topology and normed linear space theory. The treatment of material given here is quite standard in graduate courses of this sort, although Lebesgue measure and Lebesgue integration are treated in this book before the general theory of measure and integration. I have found this a happy pedagogical practice, since the student first becomes familiar with an important concrete case and then sees that much of what he has learned can be applied in very general situations.

There is considerable independence among chapters, and the chart on page 4 gives the essential dependencies. The instructor thus has considerable freedom in arranging the material here into a course according to his taste. Sections that are peripheral to the principal line of argument have been starred (*). The Prologue to the Student lists some of the notations and conventions and makes some suggestions.

The material in this book belongs to the common culture of mathematics and reflects the craftsmanship of many mathematicians. My

treatment of it is particularly indebted to the published works of Constantine Carathéodory, Paul Halmos, and Stanislaw Saks and to the lectures and conversations of Andrew Gleason, John Herriot, and Lynn Loomis. Chapter 15 is the result of intensive discussion with John Lamperti.

I also wish to acknowledge my indebtedness for helpful suggestions and criticism from numerous students and colleagues. Of the former I should like to mention in particular Peter Loeb, who read the manuscript of the original edition and whose helpful suggestions improved the clarity of a number of arguments, and Charles Stanton, who read the manuscript for this revision, correcting a number of fallacious statements and problems. Among my colleagues, particular thanks are due to Paul Berg, who pointed out Littlewood's "three principles" to me, to Herman Rubin, who provided counterexamples to many of the theorems the first time I taught the course, and to John Kelley, who read the manuscript, giving helpful advice and making me omit my polemical remarks. (A few have reappeared as footnotes, however.) Finally, my thanks go to Margaret Cline for her patience and skill in transforming illegible copy into a finished typescript for the original edition, to William Glassmire for reading proofs of the revised edition, to Valerie Yuchartz for typing the material for this edition, and to the editors at Macmillan for their forbearance and encouragement during the years in which this book was written.

H. L. R.

*Stanford, California*
*September 1967*

# Contents

# Part Three

GENERAL MEASURE AND INTEGRATION THEORY

# Real Analysis

# Prologue to the Student

This book covers a portion of the material that every graduate student in mathematics must know. For want of a better name the material here is denoted by real analysis, by which I mean those parts of modern mathematics which have their roots in the classical theory of functions of a real variable. These include the classical theory of functions of a real variable itself, measure and integration, point-set topology, and the theory of normed linear spaces. This book is accordingly divided into three parts. The first part contains the classical theory of functions, including the classical Banach spaces. The second is devoted to general topology and to the theory of general Banach spaces, and the third to abstract treatment of measure and integration.

*Prerequisites.* It is assumed that the reader already has some acquaintance with the principal theorems on continuous functions of a real variable and with Riemann integration. No formal use of this knowledge is made here, and Chapter 2 provides (formally) all of the basic theorems required. The material in Chapter 2 is, however, presented in a rather brief fashion and is intended for review and as introduction to the succeeding chapters. The reader to whom this material is not already familiar may find it difficult to follow the presentation here. We also presuppose some acquaintance with the elements of modern algebra as taught in the usual undergraduate course. The definitions and elementary properties of groups and rings are used in some of the peripheral sections, and the basic notations of linear vector spaces are used in Chapter 10. The theory of sets underlies all of the material in this book, and I have sketched some of the basic facts from set theory in Chapter 1. Since the remainder of the book is full of applications of set theory, the students should become adept at set-theoretic arguments in progressing through the book. I recommend that the student first read Chapter 1 lightly and then refer back to it as needed. The books by Halmos [6][1] and Suppes [14] contain a more thorough treatment of set theory and can be profitably read by the student while reading this book.

*Logical notation.* It is convenient to use some abbreviations for logical expressions. We use '&' to mean 'and' so that '$A$ & $B$' means '$A$ and $B$'; '$\vee$' means 'or' so that '$A \vee B$' means '$A$ or $B$ (or both)'; $\neg$ means 'not' or 'it is not the case that', so that '$\neg A$' means 'it is not the case that $A$'. Another important notation is the one that we express by the symbol '$\Rightarrow$'. It has a number of synonyms in English, so that the statement '$A \Rightarrow B$' can be expressed by saying 'if $A$, then $B$', '$A$ implies $B$', '$A$ only if $B$', '$A$ is sufficient for $B$', or '$B$ is necessary for $A$'. The statement '$A \Rightarrow B$' is equivalent to each of the statements '$(\neg A) \vee B$' and '$\neg(A \;\&\; (\neg B))$'. We also use the notation '$A \Leftrightarrow B$' to mean '$(A \Rightarrow B) \;\&\; (B \Rightarrow A)$'. English synonyms for '$A \Leftrightarrow B$' are '$A$ if and only if $B$', '$A$ iff $B$', '$A$ is equivalent to $B$', and '$A$ is necessary and sufficient for $B$'.

In addition to the preceding symbols we use two further abbreviations: '$(x)$' to mean 'for all $x$' or 'for every $x$', and '$(\exists x)$' to mean 'there is an $x$' or 'for some $x$'. Thus the statement $(x)(\exists y)(x < y)$ says that for every $x$ there is a $y$ which is larger than $x$. Similarly, $(\exists y)(x)(x < y)$ says that there is a $y$ which is larger than every $x$. Note that these two statements are different: As applied to real numbers, the first is true and the second is false.

Since saying that there is an $x$ such that $A(x)$ means that it is not the case that for every $x$ we have $\neg A(x)$, we see that $(\exists x)A(x) \Leftrightarrow \neg(x)\neg A(x)$. Similarly, $(x)A(x) \Leftrightarrow \neg(\exists x)\neg A(x)$. This rule is often convenient when we wish to express the negative of a complex statement. Thus

$$\begin{aligned}\neg\{(x)(\exists y)(x < y)\} &\Leftrightarrow \neg(x)\neg(y)\neg(x < y)\\ &\Leftrightarrow (\exists x)(y)\neg(x < y)\\ &\quad (\exists x)(y)(y \leq x),\end{aligned}$$

[1] Numbers in brackets refer to the Bibliography, p. 435.

where we have used properties of the real numbers to infer that $\neg(x < y) \Leftrightarrow (y \leq x)$.

We sometimes modify the standard logical notation slightly and write $(\epsilon > 0)$ (...), $(\exists\delta > 0)$ (...), and $(\exists x \in A)$ (...) to mean 'for every $\epsilon$ greater than 0 (...)', 'there is a $\delta$ greater than 0 such that (...)', and 'there is an $x$ in the set $A$ such that (...)'. This modification shortens our expressions. For example, $(\epsilon > 0)$ (...) would be written in standard notation $(\epsilon)$ $\{(\epsilon > 0) \Rightarrow (\ldots)\}$.

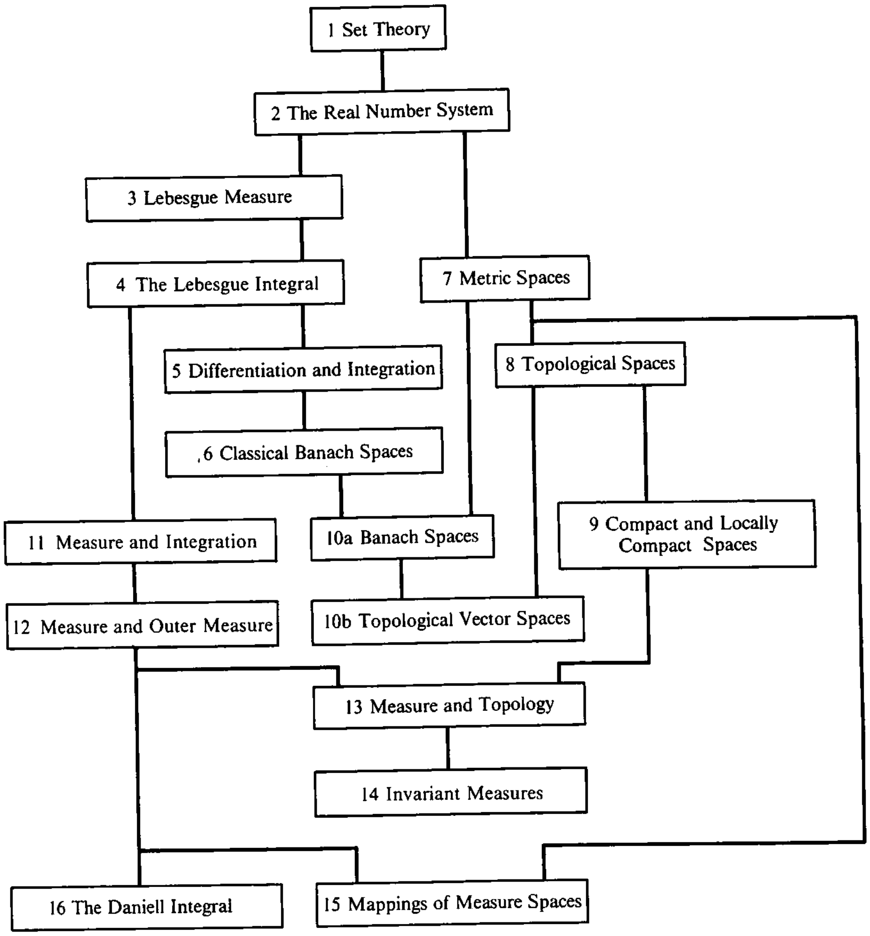
For a thorough discussion of the formal use of logical symbolism, the student should refer to Suppes [14].

*Statements and their proofs.* Most of the principal statements (theorems, propositions, etc.) in mathematics have the standard form 'if $A$, then $B$' or in symbols '$A \Rightarrow B$'. The *contrapositive* of $A \Rightarrow B$ is the statement $(\neg B) \Rightarrow (\neg A)$. It is readily seen that a statement and its contrapositive are equivalent; that is, if one is true, then so is the other. The direct method of proving a theorem of the form '$A \Rightarrow B$' is to start with $A$, deduce various consequences from it, and end with $B$. It is sometimes easier to prove a theorem by contraposition, that is, by starting with $\neg B$ and deriving $\neg A$. A third method of proof is proof by contradiction or *reductio ad absurdum:* We begin with $A$ and $\neg B$ and derive a contradiction. All students are enjoined in the strongest possible terms to eschew proofs by contradiction! There are two reasons for this prohibition: First, such proofs are very often fallacious, the contradiction on the final page arising from an erroneous deduction on an earlier page, rather than from the incompatibility of $A$ with $\neg B$. Second, even when correct, such a proof gives little insight into the connection between $A$ and $B$, whereas both the direct proof and the proof by contraposition construct a chain of argument connecting $A$ with $B$. One reason that mistakes are so much more likely in proofs by contradiction than in direct proofs or proofs by contraposition is that in a direct proof (assuming the hypothesis is not always false) all deductions from the hypothesis are true in those cases where the hypothesis holds, and similarly for proofs by contraposition (if the conclusion is not always true) the deductions from the negation of the conclusion are true in those cases where the conclusion is false. Either way, one is dealing with true statements, and one's intuition and knowledge about what is true help to keep one from making erroneous statements. In proofs by contradiction, however, you are (assuming the theorem true) in the unreal world where any statement can be derived, and so the falsity of a statement is no indication of an erroneous deduction.

The principal statements in this book are numbered consecutively in each chapter and are variously labeled lemma, proposition, theorem, or corollary. A theorem is a statement of such importance that it should be remembered since it will be used frequently. A proposition is a statement of some interest in its own right but which has less frequent application. A lemma is usually used only for proving propositions and theorems in the same section. References to statements in the same chapter are made by giving the statement number, as Theorem 17. References to statements in another chapter take the form Proposition 3.21, meaning Proposition 21 of Chapter 3. A similar convention is followed with respect to problems. I have tried to restrict the essential use of interchapter references to named theorems such as 'the Lebesgue Convergence Theorem'; the references to numbered statements are mostly auxiliary references which the student should find it unnecessary to consult.

The proof of a theorem, proposition, etc., in this book begins with the word '***proof***' and ends with the symbol '▮', which has the meaning of 'this completes the proof'. If a theorem has the form '$A \Leftrightarrow B$', the proof is usually divided into two parts, one, the 'only if' part, proving $A \Rightarrow B$, the other, the 'if' part, proving $B \Rightarrow A$.

*Interdependence of the chapters.* The dependence of one chapter on the preceding ones is indicated by the chart on page 5 (except for a few peripheral references). Chapter 10a indicates Sections 1 through 4 and 8 of Chapter 10; Chapter 10b denotes Sections 5 through 7 of that chapter.


1 Set Theory
2 The Real Number System
3 Lebesgue Measure
4 The Lebesgue Integral
5 Differentiation and Integration
6 Classical Banach Spaces
7 Metric Spaces
8 Topological Spaces
9 Compact and Locally Compact Spaces
10a Banach Spaces
10b Topological Vector Spaces
11 Measure and Integration
12 Measure and Outer Measure
13 Measure and Topology
14 Invariant Measures
15 Mappings of Measure Spaces
16 The Daniell Integral

# 1 Set Theory

## 1 Introduction

One of the most important tools in modern mathematics is the theory of sets. The study of sets and their use in the foundations of mathematics was begun just before the turn of the century by Cantor, Frege, Russell, and others, and it appeared that all mathematics could be based on set theory alone. It is in fact possible to base most of mathematics on set theory, but unfortunately this set theory is not quite as simple and natural as Frege and Russell supposed; for it was soon discovered that a free and uncritical use of set theory leads to contradictions and that set theory had to have a careful development with various devices used to exclude the contradictions. Roughly speaking, the contradictions appear when one uses sets that are "too big," such as trying to speak of a set which contains everything. In this book we shall avoid these contradictions by having some set or space $X$ fixed for a given discussion and considering only sets whose elements are elements of $X$, or sets (collections) whose elements are subsets of $X$, or sets (families) whose elements are collections of subsets of $X$, and so forth. In the first few chapters $X$ will usually be the set of real numbers.

In the present chapter we describe some of the notions from set theory which will be useful later. Our purpose is descriptive and the arguments given are directed toward plausibility and (hopefully) understanding rather than rigorous proof in some fixed basis for set theory. The descriptions and notations given here are for the most part consistent with the set theory described by Halmos in his book *Naive Set Theory* [5], although we assume as known, or primitive, several notions such as the natural numbers, the rational numbers, functions, and so forth, which can be defined (as in Halmos) in terms of the notions of set theory.

For an axiomatic treatment, I recommend Suppes' book *Axiomatic Set Theory* [14] or the appendix to Kelley's book, *General Topology* [9].

The natural numbers (positive integers) play such an important role in this book that we introduce the special symbol **N** for the set of natural numbers. We also shall take for granted the principle of mathematical induction and the well-ordering principle. The principle of mathematical induction states that if $P(n)$ is a proposition defined for each $n$ in **N**, then $\{P(1) \ \& \ [P(n) \Rightarrow P(n+1)]\} \Rightarrow (n)P(n)$. The well-ordering principle asserts that each nonempty subset of **N** has a smallest element.

The basic notions of set theory are those of set and the idea of membership in a set. We express this latter notion by $\varepsilon$, and write '$x \,\varepsilon\, A$' for the statement '$x$ is an element (or member) of $A$'. A set is completely determined by its members; that is, if two sets $A$ and $B$ have the property that $x \,\varepsilon\, A$ if and only if $x \,\varepsilon\, B$, then $A = B$. Suppose that each $x$ in a set $A$ is in the set $B$, that is, $x \,\varepsilon\, A \Rightarrow x \,\varepsilon\, B$; then we say that $A$ is a subset of $B$ or that $A$ is contained in $B$ and write $A \subset B$. Thus we always have $A \subset A$, and if $A \subset B$ and $B \subset A$, then $B = A$. It is perhaps unfortunate that the English phrase "contained in" is often used to represent both the notions $\varepsilon$ and $\subset$, but we shall use it only in the latter context. We write '$x \notin A$' to mean 'not ($x \,\varepsilon\, A$)', that is, that $x$ is not an element of $A$.

Since a set is determined by its elements, one of the most usual ways of determining a set is by specifying its elements, as in the definition: The set $A$ is the set of all elements $x$ in $X$ which have the property $P$. We abbreviate this by writing

$$A = \{x \,\varepsilon\, X : P(x)\}.$$

Thus $x \,\varepsilon\, A \Leftrightarrow [x \,\varepsilon\, X \ \& \ P(x)]$. Where the set $X$ is understood, we sometimes write[1]

$$A = \{x : P(x)\}.$$

[1] The presence (explicit or implicit) of the qualifying set $X$ is essential. Otherwise, we are confronted with the Russell paradox (cf. Suppes [14], p. 6).

We usually think of a set as having some members, but it turns out to be convenient to consider also a set which has no members. Since a set is determined by its elements, there is only one such set, and we call it the **empty set** and denote it by $\emptyset$. If $A$ is any set, then each member of $\emptyset$ (there are none) is a member of $A$, and so $\emptyset \subset A$. Thus the empty set is a subset of every set.

If $x$, $y$, $z$ are elements of $X$, we define the set $\{x\}$ to be the set whose only element is $x$; the set $\{x, y\}$ to be the set whose elements are exactly $x$ and $y$; the set $\{x, y, z\}$ to be the set whose elements are $x$, $y$, and $z$; and so forth. The set $\{x\}$ is called a **unit set,** or the **singleton** of $x$. One should distinguish carefully between $x$ and $\{x\}$. For example, we always have $x \in \{x\}$, while we seldom have $x \in x$.

In $\{x, y\}$ there is no preference given to $x$ over $y$; that is, $\{x, y\} = \{y, x\}$. For this reason we call $\{x, y\}$ an **unordered pair.** It is often useful to consider also the **ordered pair** $\langle x, y\rangle$, where we distinguish between the first element $x$ and the second element $y$. Thus $\langle x, y\rangle = \langle a, b\rangle$ if and only if $x = a$ and $y = b$, so that $\langle x, y\rangle \neq \langle y, x\rangle$ if $x \neq y$. Similarly, we shall consider ordered triplets $\langle x, y, z\rangle$, quadruplets $\langle x, y, z, w\rangle$, and so forth, where we distinguish between the first, second, third, ..., elements. Although it is possible to define ordered pairs, triplets, and so forth, in terms of unordered pairs (cf. Halmos), we shall not do so here.

If $X$ and $Y$ are two sets, we define the **Cartesian,** or **direct, product** $X \times Y$ to be the set $\{\langle x, y\rangle\}$ of all ordered pairs whose first element belongs to $X$ and whose second element belongs to $Y$. Similarly, $X \times Y \times Z$ is the set $\{\langle x, y, z\rangle\}$ of all ordered triples such that $x \in X$, $y \in Y$, and $z \in Z$. If $X$ is the set of real numbers, then $X \times X$ is the set of ordered pairs of real numbers and, as we know from analytic geometry, is equivalent to the set of points in the plane. We sometimes write $X^2$ for $X \times X$, $X^3$ for $X \times X \times X$, and so forth.

## Problems

**1.** Show that $\{x\colon x \neq x\} = \emptyset$.

**2.** Show that if $x \in \emptyset$, then $x$ is a green-eyed lion.

**3.** Show that in general the sets $X \times (Y \times Z)$ and $(X \times Y) \times Z$ are different but that there is a natural correspondence between each of them and $X \times Y \times Z$.

**4.** Show that the well-ordering principle implies the principle of mathematical induction. [Consider the set $\{n \in \mathbf{N}\colon P(n) \text{ false}\}$.]

5. Use mathematical induction to establish the well-ordering principle. [Given a set $S$ of positive integers, let $P(n)$ be the proposition 'If $n \varepsilon S$, then $S$ has a least element'.]

## 2 Functions

By a function $f$ from (or on) a set $X$ to (or into) a set $Y$ we mean a rule that assigns to each $x$ in $X$ a unique element $f(x)$ in $Y$. The collection $G$ of pairs of the form $\langle x, f(x)\rangle$ in $X \times Y$ is called the **graph** of the function $f$. A subset $G$ of $X \times Y$ is the graph of a function on $X$ if and only if for each $x \varepsilon X$ there is a unique pair in $G$ whose first element is $x$. Since a function is determined by its graph, many people like to define a function to be its graph. It is irrelevant for our purposes whether we do this or consider the notion of function as primitive.[2]

The word 'mapping' is often used as a synonym for 'function'. We express the fact that $f$ is a function of $X$ into $Y$ by writing

$$f: X \to Y.$$

The set $X$ is called the *domain* (or domain of definition) of $f$. The set of values taken by $f$, that is, the set $\{y \varepsilon Y: (\exists x)[y = f(x)]\}$, is called the *range* of $f$. The range of a function $f$ will generally be smaller than $Y$. If the range of $f$ is $Y$, then we say that $f$ is a function **onto** $Y$. (Another terminology for this case: $f$ is *surjective.*)

If $A$ is a subset of $X$, we define the *image* under $f$ of $A$ to be the set of elements in $Y$ such that $y = f(x)$ for some $x$ in $A$. We denote this image by $f[A]$. Thus

$$f[A] = \{y \varepsilon Y: (\exists x)\,[x \varepsilon A \quad \text{and} \quad y = f(x)]\}.$$

Thus the range of $f$ is $f[X]$, and $f$ is onto $Y$ if and only if $Y = f[X]$.

More important than the notion of image of set under $f$ is the notion of an inverse image. If $B$ is a subset of $Y$, we define the **inverse image** $f^{-1}[B]$ of $B$ to be the set of those $x$ in $X$ for which $f(x)$ is in $B$;

[2] The equivalence between functions and their graphs is valid only for functions from a given set $X$ to a set $Y$. There are difficulties, for example, in defining the graph of the identity function $i$ defined by $i(x) = x$ for all $x$. In a formal treatment which takes function as a primitive notion, we must have axioms describing the properties of functions, such as $(f = g) \Leftrightarrow (x)[f(x) = g(x)]$, as well as axioms enabling us to construct functions.

that is,

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

It should be noted that $f$ is onto $Y$ if and only if the inverse image of each nonempty subset of $Y$ is nonempty.

A function $f\colon X \to Y$ is called **one-to-one** (or *univalent*, or *injective*) if $f(x_1) = f(x_2)$ only when $x_1 = x_2$. Functions which are one-to-one from $X$ onto $Y$ are often called one-to-one correspondences between $X$ and $Y$. (They are also called *bijective*.) In this case there is a function $g\colon Y \to X$ such that for all $x$ and $y$ we have $g(f(x)) = x$ and $f(g(y)) = y$. The function $g$ is called the inverse of $f$ and is sometimes denoted by $f^{-1}$.

It should be noted that if we denote $g$ by $f^{-1}$, then $f^{-1}[E]$ can be considered to be the inverse image of $E$ under $f$ or the image of $E$ under $f^{-1}$. Fortunately for our notation, these are the same set.

If $f\colon X \to Y$ and $g\colon Y \to Z$, we define a new function $h\colon X \to Z$ by setting $h(x) = g(f(x))$. The function $h$ is called the *composition* of $g$ with $f$ and denoted by $g \circ f$. If $f\colon X \to Y$ and $A$ is a subset of $X$, we can define a new function $g\colon A \to Y$ by defining $g(x) = f(x)$ for $x \in A$. This new function $g$ is called the *restriction* of $f$ to $A$ and is sometimes written $f \mid A$. In many cases it is important to distinguish carefully between the functions $g$ and $f$. They have different ranges, and the inverse images under $g$ are different from those under $f$.

Having mentioned the possibility of defining functions by means of ordered pairs, it is only fair to point out that, conversely, ordered pairs may be defined in terms of the notion of a function. An ordered pair is a function whose domain is the set $\{1, 2\}$. Similarly, a **finite sequence,** or $n$-tuple, is a function whose domain is the first $n$ natural numbers, that is, the set $\{i \in \mathbf{N} : i \leq n\}$. (We call such a set a *segment* of $\mathbf{N}$.) Similarly, an **infinite sequence** is a function whose domain is the set $\mathbf{N}$ of natural numbers. We use the term "sequence" to mean a finite or infinite sequence. If the range of a sequence is in a set $X$, we speak of a sequence from (or in) $X$ or of a sequence of elements of $X$. It is customary to depart somewhat from the usual functional notation when dealing with sequences and denote the value of the function at $i$ by $x_i$ and to call this value the $i$-th element of the sequence. We shall often denote ordered $n$-tuples by $\langle x_i \rangle_{i=1}^{n}$ and infinite sequences by $\langle x_i \rangle_{i=1}^{\infty}$. When no misunderstanding is likely to arise we often write simply $\langle x_i \rangle$. The range of the sequence $\langle x_i \rangle$ will be denoted by $\{x_i\}$. Thus the range of an ordered $n$-tuple $\langle x_i \rangle_{i=1}^{n}$ is the unordered $n$-tuple $\{x_i\}_{i=1}^{n}$. This notation is reasonably consistent with our earlier notation concerning ordered and unordered pairs, triplets, and so forth.

A set $A$ is called **countable** if it is the range of some sequence and **finite** if it is the range of some finite sequence. A set that is not finite is called infinite. (Many authors restrict the use of the word 'countable' to sets that are infinite and countable, but our definition includes the finite sets among the countable sets.) We use the term 'countably infinite' for infinite countable sets. We shall return to this notion again in Section 6.

A useful way of defining an infinite sequence is given by the following principle:

**Principle of Recursive Definition:** *Let $f$ be a function from a set $X$ to itself, and let $a$ be an element of $X$. Then there is a unique infinite sequence $\langle x_i \rangle$ from $X$ such that $x_1 = a$ and $x_{i+1} = f(x_i)$ for each $i$.*

The existence of such a sequence is intuitively clear: We define $x_1 = a$, $x_2 = f(a)$, $x_3 = f(f(a))$, and so on. A more formal proof can be given as follows: We first prove by induction on $n$ that for each natural number $n$ there is a unique finite sequence

$$x_1^{(n)}, x_2^{(n)}, \ldots, x_n^{(n)},$$

such that

$$x_1^{(n)} = a \quad \text{and} \quad x_{i+1}^{(n)} = f(x_i^{(n)}) \qquad \text{for } 1 \le i < n.$$

From the uniqueness it follows that $x_i^{(n)} = x_i^{(m)}$ for $i \le n \le m$. Thus if we define $x_n$ to be $x_n^{(n)}$, we have $x_i^{(n)} = x_i$ for $i \le n$, and we see that the sequence $\langle x_i \rangle$ satisfies the requirements of our principle.

A slight extension of this principle is the following: For each natural number $n$ let $f_n$ be a function on $X^n$ to $X$ and let $a \in X$. Then there is a unique sequence $\langle x_i \rangle$ from $X$ such that $x_1 = a$ and $x_{i+1} = f_i(x_1, \ldots, x_i)$.

An important notion connected with that of a sequence is the notion of a subsequence. We say that a mapping $g$ of $\mathbf{N}$ into $\mathbf{N}$ is **monotone** if $(i > j) \Rightarrow (g(i) > g(j))$. If $f$ is an infinite sequence (that is, a function whose domain is $\mathbf{N}$), we say that $h$ is an infinite **subsequence** of $f$ if there is a monotone mapping $g$ of $\mathbf{N}$ into $\mathbf{N}$ such that $h = f \circ g$. If we write $f$ as $\langle f_i \rangle$ and $g$ as $\langle g_i \rangle$, then we usually denote $f \circ g$ by $\langle f_{g_i} \rangle$.

## Problems

**6.** Let $f: X \to Y$ be a mapping of a nonempty space $X$ into $Y$. Show that $f$ is one-to-one if and only if there is a mapping $g: Y \to X$ such that $g \circ f$ is the identity map on $X$, that is, such that $g(f(x)) = x$ for all $x \in X$.

**7.** Let $f: X \to Y$ be a mapping of $X$ into $Y$. Show that $f$ is onto if there is a mapping $g: Y \to X$ such that $f \circ g$ is the identity map on $Y$, that is, such that $f(g(y)) = y$ for all $y \in Y$. (For the converse, see Problem 21.)

**8.** Use mathematical induction to prove the generalized principle of recursive definition stated in the text.

## 3 Unions, Intersections, and Complements

Let us fix a given set $X$ and consider the set $\mathcal{P}(X)$ consisting of all subsets of $X$. There are certain set-theoretic operations that we can perform on subsets of $X$. If $A$ and $B$ are subsets of $X$, we define their intersection $A \cap B$ to be the set of all elements that belong to both $A$ and $B$. Thus

$$A \cap B = \{x : x \in A \;\&\; x \in B\}.$$

We note that the definition is symmetric in $A$ and $B$; that is, $A \cap B = B \cap A$. Also, $A \cap B \subset A$ and $A \cap B = A \Leftrightarrow A \subset B$. We have $(A \cap B) \cap C = A \cap (B \cap C)$ and write this set as $A \cap B \cap C$. It is the set of all elements that belong to each of the sets $A$, $B$, and $C$.

We define the union $A \cup B$ of two sets $A$ and $B$ to be the set of elements that are in either $A$ or $B$. Thus

$$A \cup B = \{x : x \in A \vee x \in B\}.$$

We have

$$A \cup B = B \cup A$$

$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$$

$$A \subset A \cup B$$

$$A = A \cup B \Leftrightarrow B \subset A.$$

We also have relations between unions and intersections which are called distributive laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

The empty set $\emptyset$ and the space $X$ play special roles:

$$A \cup \emptyset = A, \qquad A \cap \emptyset = \emptyset$$
$$A \cup X = X, \qquad A \cap X = A.$$

If $A$ is a subset of $X$, we define the complement $\tilde{A}$ of $A$ (relative to $X$) as the set of elements not in $A$. Thus

$$\tilde{A} = \{x \in X : x \notin A\}.$$

We sometimes write $\sim A$ instead of $\tilde{A}$. We have

$$\tilde{\emptyset} = X, \qquad \tilde{X} = \emptyset$$
$$\tilde{\tilde{A}} = A, \qquad A \cup \tilde{A} = X, \qquad A \cap \tilde{A} = \emptyset$$
$$A \subset B \Leftrightarrow \tilde{B} \subset \tilde{A}.$$

Two special laws relating complements to unions and intersections are De Morgan's laws:

$$\sim(A \cup B) = \tilde{A} \cap \tilde{B}$$
$$\sim(A \cap B) = \tilde{A} \cup \tilde{B}.$$

If $A$ and $B$ are two subsets of $X$, we define the *difference* $B \sim A$, or *relative complement* of $A$ in $B$, as the set of elements in $B$ that are not in $A$. Thus

$$B \sim A = \{x : x \in B \ \&\ x \notin A\}.$$

We have $B \sim A = B \cap \tilde{A}$.

We shall also use the notation $A \,\Delta\, B$ for the *symmetric difference* of two sets defined by

$$A \,\Delta\, B = (A \sim B) \cup (B \sim A).$$

The symmetric difference of two sets consists of all those points that belong to one or the other of the two sets but not to both.

If the intersection of two sets is empty, we say the sets are **disjoint.** A collection $\mathcal{C}$ of sets is said to be a **disjoint collection** of sets or a collection of pairwise disjoint sets if any two sets in $\mathcal{C}$ are disjoint.

The process of taking unions (or intersections) of two sets can be extended by repetition to give unions (or intersections) of any finite collection of sets. However, we can give a definition of intersection for an arbitrary collection $\mathcal{C}$ of sets: The intersection of the collection $\mathcal{C}$ is the set of those elements of $X$ that belong to each member of $\mathcal{C}$.

We denote this intersection by $\bigcap\limits_{A \in \mathcal{C}} A$ or $\bigcap \{A : A \in \mathcal{C}\}$. Thus

$$\bigcap_{A \in \mathcal{C}} A = \{x \in X : (A)(A \in \mathcal{C} \Rightarrow x \in A)\}.$$

Similarly, we define the union of an arbitrary collection of sets by

$$\bigcup_{A \in \mathcal{C}} A = \{x \in X : (\exists A)(A \in \mathcal{C} \;\&\; x \in A)\}.$$

De Morgan's laws hold for arbitrary unions and intersections:

$$\sim\left[\bigcup_{A \in \mathcal{C}} A\right] = \bigcap_{A \in \mathcal{C}} \tilde{A}$$

$$\sim\left[\bigcap_{A \in \mathcal{C}} A\right] = \bigcup_{A \in \mathcal{C}} \tilde{A}.$$

We also have the distributive laws:

$$B \cap \left[\bigcup_{A \in \mathcal{C}} A\right] = \bigcup_{A \in \mathcal{C}} (B \cap A)$$

$$B \cup \left[\bigcap_{A \in \mathcal{C}} A\right] = \bigcap_{A \in \mathcal{C}} (B \cup A).$$

It follows from our definition that the union of an empty collection of sets is empty and that the intersection of the empty collection of sets is $X$.

By a sequence of subsets of $X$ we mean a sequence from $\mathcal{P}(X)$, that is, a mapping of $\mathbf{N}$ (or a segment of $\mathbf{N}$) into $\mathcal{P}(X)$. If $\langle A_i \rangle$ is an infinite sequence of subsets of $X$, we write $\bigcup\limits_{i=1}^{\infty} A_i$ for the union of the range of the sequence. Thus

$$\bigcup_{i=1}^{\infty} A_i = \{x : (\exists i)(x \in A_i)\}.$$

Similarly, if $\langle B_i \rangle_{i=1}^{n}$ is a finite sequence of subsets of $X$, we write $\bigcap\limits_{i=1}^{n} B_i$ for the intersection of the range of the sequence, so that

$$\bigcap_{i=1}^{n} B_i = B_1 \cap B_2 \cap \cdots \cap B_n.$$

This notation for sequences of sets is so convenient that we often generalize it to arbitrary collections of sets by using the notion of an indexed collection: An **indexed subset** of $X$ (or collection of subsets of $X$) is a function on an index set $\Lambda$ to $X$ (or the set of subsets of $X$). If $\Lambda$ is the set of natural numbers, then the notion of an indexed set coincides with the notion of a sequence.

In keeping with the notation for sequences, we usually write $x_\lambda$ instead of $x(\lambda)$, and denote the indexed set itself by $\{x_\lambda\}$ or $\{x_\lambda: \lambda \varepsilon \Lambda\}$. We say that $\{x_\lambda\}$ is indexed by $\Lambda$. We define the union and intersection of an indexed set to be the union and intersection of the range of the function defining the indexed set. Thus

$$\bigcup_{\lambda \varepsilon \Lambda} A_\lambda = \{x \varepsilon X: (\exists\lambda)(\lambda \varepsilon \Lambda \ \&\ x \varepsilon A_\lambda)\}$$

and

$$\bigcap_{\lambda \varepsilon \Lambda} A_\lambda = \{x \varepsilon X: (\lambda)(\lambda \varepsilon \Lambda \Rightarrow x \varepsilon A_\lambda)\}.$$

In the case when $\Lambda$ is the set $\mathbf{N}$ of natural numbers we have

$$\bigcap_{i \varepsilon N} A_i = \bigcap_{i=1}^{\infty} A_i,$$

and similarly for unions.

If $f$ maps $X$ into $Y$ and $\{A_\lambda\}$ is a collection of subsets of $X$, then

$$f\left[\bigcup_\lambda A_\lambda\right] = \bigcup_\lambda f[A_\lambda],$$

but we can conclude only that

$$f\left[\bigcap_\lambda A_\lambda\right] \subset \bigcap_\lambda f[A_\lambda].$$

For inverse images we have, for a collection $\{B_\lambda\}$ of subsets of $Y$,

$$f^{-1}\left[\bigcup_\lambda B_\lambda\right] = \bigcup_\lambda f^{-1}[B_\lambda]$$

$$f^{-1}\left[\bigcap_\lambda B_\lambda\right] = \bigcap_\lambda f^{-1}[B_\lambda]$$

and

$$f^{-1}[\tilde{B}] = \sim f^{-1}[B]$$

for $B \subset Y$. Also,

$$f[f^{-1}[B]] \subset B$$

and

$$f^{-1}[f[A]] \supset A$$

for $A \subset X$ and $B \subset Y$.

## Problems

**9.** Show that $A \subset B \Leftrightarrow A \cap B = A \Leftrightarrow A \cup B = B$.

**10.** Prove the distributive laws.

**11.** Show that $A \subset B \Leftrightarrow \tilde{B} \subset \tilde{A}$.

**12.** Show that:

**a.** $A \,\Delta\, B = B \,\Delta\, A$ and $A \,\Delta\, (B \,\Delta\, C) = (A \,\Delta\, B) \,\Delta\, C$.

**b.** $A \,\Delta\, B = \emptyset \Leftrightarrow A = B$.

**c.** $A \,\Delta\, B = X \Leftrightarrow A = \tilde{B}$.

**d.** $A \,\Delta\, \emptyset = A$ and $A \,\Delta\, X = \tilde{A}$.

**e.** $(A \,\Delta\, B) \cap E = (A \cap E) \,\Delta\, (B \cap E)$.

**13.** Prove De Morgan's laws (for arbitrary unions and intersections).

**14.** Show that

$$B \cap \left[\bigcup_{A \in \mathcal{C}} A\right] = \bigcup_{A \in \mathcal{C}} (B \cap A).$$

**15.** Show that if $\mathfrak{A}$ and $\mathfrak{B}$ are two collections of sets, then

$$[\bigcup \{A\colon A \in \mathfrak{A}\}] \cap [\bigcup \{B\colon B \in \mathfrak{B}\}] = \bigcup \{A \cap B\colon \langle A, B\rangle \in \mathfrak{A} \times \mathfrak{B}\}.$$

**16. a.** Show that $f[\bigcup A_\lambda] = \bigcup f[A_\lambda]$.

**b.** Show that $f[\bigcap A_\lambda] \subset \bigcap f[A_\lambda]$.

**c.** Give an example where

$$f[\bigcap A_\lambda] \neq \bigcap f[A_\lambda].$$

**17.** Show that

**a.** $f^{-1}[\bigcup B_\lambda] = \bigcup f^{-1}[B_\lambda]$.

**b.** $f^{-1}[\bigcap B_\lambda] = \bigcap f^{-1}[B_\lambda]$.

**c.** $f^{-1}[\tilde{B}] = \sim f^{-1}[B]$ for $B \subset Y$.

**18. a.** Show that if $f$ maps $X$ into $Y$ and $A \subset X$, $B \subset Y$, then

$$f[f^{-1}[B]] \subset B$$

and

$$f^{-1}[f[A]] \supset A.$$

**b.** Give examples to show that we need not have equality.

**c.** Show that if $f$ maps $X$ onto $Y$ and $B \subset Y$, then

$$f[f^{-1}[B]] = B.$$

## 4 Algebras of Sets

A collection $\mathfrak{A}$ of subsets of $X$ is called an **algebra** of sets or a **Boolean algebra** if (i) $A \cup B$ is in $\mathfrak{A}$ whenever $A$ and $B$ are, and (ii) $\tilde{A}$ is in $\mathfrak{A}$ whenever $A$ is. It follows from De Morgan's laws that (iii) $A \cap B$ is in $\mathfrak{A}$ whenever $A$ and $B$ are. If a collection $\mathfrak{A}$ of subsets of $X$ satisfies (ii) and (iii), then by De Morgan's laws it also satisfies (i) and is therefore a Boolean algebra. By taking unions two at a time, we see that if $A_1, \ldots, A_n$ are sets in $\mathfrak{A}$, then $A_1 \cup A_2 \cup \cdots \cup A_n$ is again in $\mathfrak{A}$. Similarly, $A_1 \cap A_2 \cap \cdots \cap A_n$ is in $\mathfrak{A}$.

We shall find several propositions concerning algebras of sets useful. The first is the following:

**1. Proposition:** *Given any collection $\mathcal{C}$ of subsets of $X$, there is a smallest algebra $\mathfrak{A}$ which contains $\mathcal{C}$; that is, there is an algebra $\mathfrak{A}$ containing $\mathcal{C}$ and such that if $\mathfrak{B}$ is any algebra containing $\mathcal{C}$, then $\mathfrak{B}$ contains $\mathfrak{A}$.*

***Proof:*** Let $\mathcal{F}$ be the family of all algebras (of subsets of $X$) that contain $\mathcal{C}$. Let $\mathfrak{A} = \bigcap \{\mathfrak{B} : \mathfrak{B} \in \mathcal{F}\}$. Then $\mathcal{C}$ is a subcollection of $\mathfrak{A}$, since each $\mathfrak{B}$ in $\mathcal{F}$ contains $\mathcal{C}$. Moreover, $\mathfrak{A}$ is an algebra. For if $A$ and $B$ are in $\mathfrak{A}$, then for each $\mathfrak{B} \in \mathcal{F}$ we have $A \in \mathfrak{B}$ and $B \in \mathfrak{B}$. Since $\mathfrak{B}$ is an algebra, $A \cup B$ belongs to $\mathfrak{B}$. Since this is true for every $\mathfrak{B} \in \mathcal{F}$, we have $A \cup B$ in $\bigcap \{\mathfrak{B} : \mathfrak{B} \in \mathcal{F})$. Similarly, we see that if $A \in \mathfrak{A}$, then $\tilde{A} \in \mathfrak{A}$. From the definition of $\mathfrak{A}$, it follows that if $\mathfrak{B}$ is an algebra containing $\mathcal{C}$, then $\mathfrak{B} \supset \mathfrak{A}$. ▮

The smallest algebra containing $\mathcal{C}$ is called the algebra generated by $\mathcal{C}$.

**2. Proposition:** *Let $\mathfrak{A}$ be an algebra of subsets and $\langle A_i \rangle$ a sequence of sets in $\mathfrak{A}$. Then there is a sequence $\langle B_i \rangle$ of sets in $\mathfrak{A}$ such that $B_n \cap B_m = \emptyset$ for $n \neq m$ and*

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

***Proof:*** Since the proposition is trivial when $\langle A_i \rangle$ is finite, we assume $\langle A_i \rangle$ to be an infinite sequence. Set $B_1 = A_1$, and for each natural number $n > 1$ define

$$\begin{aligned} B_n &= A_n \sim [A_1 \cup A_2 \cup \cdots \cup A_{n-1}] \\ &= A_n \cap \tilde{A}_1 \cap \tilde{A}_2 \cap \cdots \cap \tilde{A}_{n-1}. \end{aligned}$$

Since the complements and intersections of sets in $\mathfrak{A}$ are in $\mathfrak{A}$, we have each $B_n \in \mathfrak{A}$. We also have $B_n \subset A_n$. Let $B_n$ and $B_m$ be two such sets, and suppose $m < n$. Then $B_m \subset A_m$, and so

$$\begin{aligned} B_m \cap B_n &\subset A_m \cap B_n \\ &= A_m \cap A_n \cap \cdots \cap \tilde{A}_m \cap \cdots \\ &= (A_m \cap \tilde{A}_m) \cap \cdots \\ &= \emptyset \cap \cdots \\ &= \emptyset. \end{aligned}$$

Since $B_i \subset A_i$, we have

$$\bigcup_{i=1}^{\infty} B_i \subset \bigcup_{i=1}^{\infty} A_i.$$

Let $x \in \bigcup_{i=1}^{\infty} A_i$. Then $x$ must belong to at least one of the $A_i$'s. Let $n$ be the smallest value of $i$ such that $x \in A_i$. Then $x \in B_n$, and so $x \in \bigcup_{n=1}^{\infty} B_n$. Thus

$$\bigcup_{n=1}^{\infty} B_n \supset \bigcup_{n=1}^{\infty} A_n,$$

and we have

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n. \quad \blacksquare$$

An algebra $\mathfrak{A}$ of sets is called a **σ-algebra,** or a **Borel field,** if every union of a countable collection of sets in $\mathfrak{A}$ is again in $\mathfrak{A}$. That is, if $\langle A_i \rangle$ is a sequence of sets, then $\bigcup_{i=1}^{\infty} A_i$ must again be in $\mathfrak{A}$. From

De Morgan's laws it follows that the intersection of a countable collection of sets in $\mathcal{A}$ is again in $\mathcal{A}$. A slight modification of the proof of Proposition 1 gives us the following proposition:

**3. Proposition:** *Given any collection $\mathcal{C}$ of subsets of $X$, there is a smallest $\sigma$-algebra that contains $\mathcal{C}$; that is, there is a $\sigma$-algebra $\mathcal{A}$ containing $\mathcal{C}$ such that if $\mathcal{B}$ is any $\sigma$-algebra containing $\mathcal{C}$, then $\mathcal{A} \subset \mathcal{B}$.*

The smallest $\sigma$-algebra containing $\mathcal{C}$ is called the $\sigma$-algebra generated by $\mathcal{C}$.

## Problems

**19. a.** Prove Proposition 3.

**b.** If $\mathcal{A}$ is the algebra generated by $\mathcal{C}$, then $\mathcal{A}$ and $\mathcal{C}$ generate the same $\sigma$-algebra.

**20.** Let $\mathcal{C}$ be a collection of sets and $E$ an element in the $\sigma$-algebra generated by $\mathcal{C}$. Then there is a countable subcollection $\mathcal{C}_0 \subset \mathcal{C}$ such that $E$ is an element of the $\sigma$-algebra $\mathcal{A}_0$ generated by $\mathcal{C}_0$. [Hint: Let $\mathcal{A}'$ be the union of all $\sigma$-algebras generated by countable subsets of $\mathcal{C}$. Then $\mathcal{A}' \supset \mathcal{A}$.]

# 5 The Axiom of Choice and Infinite Direct Products

An important axiom in set theory is the so-called axiom of choice. It is somewhat less elementary than the other axioms used in axiomatic set theory and is known to be independent of them. Many mathematicians like to be very explicit about the use of the axiom of choice and its consequences, but we shall be rather informal about its use. The axiom is the following:

**Axiom of Choice:** *Let $\mathcal{C}$ be any collection of nonempty sets. Then there is a function $F$ defined on $\mathcal{C}$ which assigns to each set $A \in \mathcal{C}$ an element $F(A)$ in $A$.*

The function $F$ is called a choice function, and its existence may be thought of as the result of choosing for each of the sets $A$ in $\mathcal{C}$ an element in $A$. There is, of course, no difficulty in doing this if there are only a finite number of sets in $\mathcal{C}$, but we need the axiom of choice in case the collection $\mathcal{C}$ is infinite. If the sets in $\mathcal{C}$ are disjoint, we may think of the axiom of choice as asserting the possibility of selecting a "parliament" consisting of one member from each of the sets in $\mathcal{C}$.

Let $\mathcal{C} = \{X_\lambda\}$ be a collection of sets indexed by an index set $\Lambda$. We define the **direct product**

$$\bigtimes_\lambda X_\lambda$$

to be the collection of all sets $\{x_\lambda\}$ indexed by $\Lambda$ and having the property that $x_\lambda \in X_\lambda$. If $\Lambda = \{1, 2\}$, we have our earlier definition of the direct product $X_1 \times X_2$ of the two sets $X_1$ and $X_2$. If $z = \{x_\lambda\}$ is an element of $\bigtimes_\lambda X_\lambda$, we call $x_\lambda$ the $\lambda$-th coordinate of $z$.

If one of the $X_\lambda$ is empty, then $\bigtimes_\lambda X_\lambda$ is also empty. The axiom of choice is equivalent to the converse statement: If none of the $X_\lambda$ are empty, then $\bigtimes_\lambda X_\lambda$ is not empty. For this reason Bertrand Russell prefers to call the axiom of choice the multiplicative axiom.

## Problem

**21.** Let $f: X \to Y$ be a mapping *onto* $Y$. Then there is a mapping $g: Y \to X$ such that $f \circ g$ is the identity map on $Y$. [Apply the axiom of choice to the collection $\{A: (\exists y \in Y) \text{ with } A = f^{-1}[\{y\}]\}$.]

# 6 Countable Sets

In Section 2 we defined a set to be countable if it was the range of some sequence. If it is the range of a finite sequence, we have a finite set, but the range of an infinite sequence may also be finite. In fact, every nonempty finite set is the range of an infinite sequence. For example, the finite set $\{x_1, \ldots, x_n\}$ is the range of the infinite sequence defined by setting $x_i = x_n$ for $i > n$. Thus a set is countably infinite if it is the range of some infinite sequence but not the range of any finite sequence. The set $\mathbf{N}$ of natural numbers is an example of a countably infinite set.

Before proceeding further, we had better come to terms with the empty set. The empty set is not the range of any sequence (unless we admit sequences with zero terms). It is convenient, however, to define finite and countable sets so that the empty set is both finite and countable. Hence:

**Definition:** *A set is called **finite** if it is either empty or the range of a finite sequence. A set is called **countable** (or denumerable) if it is either empty or the range of a sequence.*

It follows at once from this definition that the image of any countable set is countable, that is, that the range of any function with a countable domain is itself countable, and similarly for finite sets.

It is usual in mathematics to give a slightly different but equivalent definition based on the notion of a one-to-one correspondence. We first note that any set that can be put in one-to-one correspondence with a finite set is finite and that any set that can be put in one-to-one correspondence with a countable set must be countable. Since the set **N** of natural numbers is countable but not finite, any set which can be put in one-to-one correspondence with **N** is countably infinite. It is customary to use this property to define the notion of countably infinite. Thus to show that our definition is equivalent to the customary one, we must show that if an infinite set $E$ is the range of a sequence $\langle x_n \rangle$, then $E$ can be put in one-to-one correspondence with **N**. To do this we define a function $\varphi$ from **N** into **N** by recursion as follows: Let $\varphi(1) = 1$, and define $\varphi(n+1)$ to be the smallest value of $m$ such that $x_m \neq x_i$ for all $i \leq \varphi(n)$. Since $E$ is infinite, such an $m$ always exists, and by the well-ordering principle for **N** there is always a least such $m$. The correspondence $n \to x_{\varphi(n)}$ is a one-to-one correspondence between **N** and $E$. Thus we have shown that a set is countably infinite if and only if it can be put into one-to-one correspondence with **N**. We are now in a position to prove some simple propositions about countable sets:

**4. Proposition:** *Every subset of a countable set is countable.*

***Proof:*** Let $E = \{x_n\}$ be a countable set, and let $A$ be a subset of $E$. If $A$ is empty, $A$ is countable by definition. If $A$ is not empty, choose $x$ in $A$. Define a new sequence $\langle y_n \rangle$ by setting $y_n = x_n$ if $x_n \in A$ and $y_n = x$ if $x_n \notin A$. Then $A$ is the range of $\langle y_n \rangle$ and is therefore countable. ▮

**5. Proposition:** *Let $A$ be a countable set. Then the set of all finite sequences from $A$ is also countable.*

***Proof:*** Since $A$ is countable, it can be put into one-to-one correspondence with a subset of the set $\mathbf{N}$ of natural numbers. Thus it suffices to prove that the set $S$ of all finite sequences of natural numbers is countable. Let $\langle 2, 3, 5, 7, 11, \ldots, p_k, \ldots\rangle$ be the sequence of prime numbers. Then each $n$ in $\mathbf{N}$ has a unique factorization of the form[3] $n = 2^{x_1} 3^{x_2} \cdots p_k^{x_k}$, where $x_i \in \mathbf{N}_0 = \mathbf{N} \cup \{0\}$ and $x_k > 0$. Let $f$ be the function on $\mathbf{N}$ that assigns to the natural number $n$ the finite sequence $\langle x_1, \ldots, x_k\rangle$ from $\mathbf{N}_0$. Then $S$ is a subset of the range of $f$. Hence $S$ is countable by Proposition 4. ▮

**6. Proposition:** *The set of all rational numbers is countable.*

**7. Proposition:** *The union of a countable collection of countable sets is countable.*

***Proof:*** Let $\mathcal{C}$ be a countable collection of countable sets. If all the sets in $\mathcal{C}$ are empty, the union is empty and thus countable. Thus we may as well assume that $\mathcal{C}$ contains nonempty sets, and since the empty set contributes nothing to the union of $\mathcal{C}$, we can assume that the sets in $\mathcal{C}$ are nonempty. Thus $\mathcal{C}$ is the range of an infinite sequence $\langle A_n\rangle_{n=1}^{\infty}$, and each $A_n$ is the range of an infinite sequence $\langle x_{nm}\rangle_{m=1}^{\infty}$. But the mapping of $\langle n, m\rangle$ to $x_{nm}$ is a mapping of the set of ordered pairs of natural numbers onto the union of $\mathcal{C}$. Since the set of pairs of natural numbers is countable, the union of the collection $\mathcal{C}$ must also be countable. ▮

## Problems

**22.** Show that every subset of a finite set is finite.

**23.** Prove Proposition 6 by using Propositions 4 and 5. [Hint: The mapping

$$\langle p, q, 1\rangle \to p/q$$
$$\langle p, q, 2\rangle \to -p/q$$
$$\langle 1, 1, 3\rangle \to 0$$

[3] Except 1; we agree to write $1 = 2^0$.

is a function whose range is the set of rational numbers and whose domain is a subset of the set of finite sequences from N.]

**24.** Show that the set $E$ of infinite sequences from $\{0, 1\}$ is not countable. [Hint: Let $f$ be a function from N to $E$. Then $f(\nu)$ is a sequence $\langle a_{\nu n}\rangle_{n=1}^{\infty}$. Let $b_\nu = 1 - a_{\nu\nu}$. Then $\langle b_n\rangle$ is again a sequence from $\{0, 1\}$, and for each $\nu \in \mathbf{N}$ we have $\langle b_n\rangle \neq \langle a_{\nu n}\rangle$. This method of proof is known as the Cantor diagonal process.]

**25.** Let $f$ be a function from a set $X$ to the collection $\mathcal{P}(X)$ of subsets of $X$. Then there is a set $E \subset X$ that is not in the range of $f$. [Hint: Let $E = \{x: x \notin f(x)\}$.]

**26.** Use the axiom of choice and the generalized principle of recursive definition to show that each infinite set $X$ contains a countably infinite subset.

## 7 Relations and Equivalences

Two given entities $x$ and $y$ may be "related" to each other in many ways, as in $x = y$, $x \in y$, $x \subset y$, or for numbers $x < y$. In general we say that R denotes a relation if, given $x$ and $y$, either $x$ stands in the relation R to $y$ (written $x \mathrel{R} y$) or $x$ does not stand in the relation R to $y$. A relation R is said to be a relation on a set $X$ if $x \mathrel{R} y$ implies $x \in X$ and $y \in X$. If R is a relation on a set $X$, we define the graph of R to be the set $\{\langle x, y\rangle: x \mathrel{R} y\}$. Since we consider two relations R and S to be the same if $(x \mathrel{R} y) \Leftrightarrow (x \mathrel{S} y)$, each relation on a set $X$ is uniquely determined by its graph, and conversely each subset of $X \times X$ is the graph of some relation on $X$. Thus we may if we like identify a relation on $X$ with its graph and define a relation to be a subset of $X \times X$. In many formalized treatments of set theory a relation is in general defined simply as a set of ordered pairs.[4]

A relation R is said to be *transitive* on a set $X$ if $x \mathrel{R} y$ and $y \mathrel{R} z$ imply $x \mathrel{R} z$ for all $x$, $y$, and $z$ in $X$. Thus $=$ and $<$ are transitive relations on the set of real numbers. A relation R is said to be symmetric on $X$ if $x \mathrel{R} y$ implies $y \mathrel{R} x$ for all $x$ and $y$ in $X$. It is said to be *reflexive* on $X$ if for all $x \in X$ we have $x \mathrel{R} x$.

A relation that is transitive, reflexive, and symmetric on $X$ is said to be an *equivalence* relation on $X$ or simply an equivalence on $X$.

[4] Cf. Suppes [14], p. 57, or Halmos [5], p. 26. It should be pointed out, however, that there is a difficulty in this approach in that $=$, $\in$, and $\subset$ are no longer relations. For this reason I prefer some treatment such as the one in Kelley [9], p. 260, where relations are not necessarily *sets* of order pairs.

Suppose that $\equiv$ is an equivalence relation on a set $X$. For a given $x \in X$, let $E_x$ be the set of elements equivalent to $x$, that is, $E_x = \{y: y \equiv x\}$. If $y$ and $z$ are both in $E_x$, then $y \equiv x$ and $z \equiv x$, and by symmetry and transitivity we have $z \equiv y$. Thus any two elements of $E_x$ are equivalent. If $y \in E_x$ and $z \equiv y$, then $z \equiv y$ and $y \equiv x$, whence $z \equiv x$, and so $z \in E_x$. Thus any element of $X$ equivalent to an element of $E_x$ is itself an element of $E_x$. Consequently, for any two elements $x$ and $y$ of $X$, the sets $E_x$ and $E_y$ are either identical (if $x \equiv y$) or disjoint (if $x \not\equiv y$). The sets in the collection $\{E_x: x \in X\}$ are called equivalence sets or classes of $X$ under $\equiv$. Thus $X$ is the disjoint union of the equivalence classes under $\equiv$. Note that $x \in E_x$, and so no equivalence class is empty.

The collection of equivalence classes under an equivalence $\equiv$ is called the *quotient* of $X$ with respect to $\equiv$, and is sometimes denoted by $X/\equiv$. The mapping $x \to E_x$ is called the natural mapping of $X$ onto $X/\equiv$.

A *binary operation* on set $X$ is a mapping from $X \times X$ to $X$. We say that an equivalence relation $\equiv$ is *compatible* with a binary operation $+$ if $x \equiv x'$ and $y \equiv y'$ imply that $(x + y) \equiv (x' + y')$. In this case $+$ defines an operation on the quotient $Q = X/\equiv$ as follows: If $E$ and $F$ belong to $Q$, choose $x \in E$, $y \in F$ and define $E + F$ to be $E_{(x+y)}$. Since $\equiv$ is an equivalence, $E + F$ is seen to depend only on $E$ and $F$ and not on the choice of $x$ and $y$.

For further details readers may consult Birkhoff and MacLane [2], pp. 145ff.

## Problems

**27.** Prove that $F + G$ as defined above depends only on $F$ and $G$.

**28.** Let $X$ be an Abelian group under $+$. Then $\equiv$ is compatible with $+$ if and only if $x \equiv x'$ implies $x + y \equiv x' + y$. The induced operation then makes the quotient space into a group.

## 8 Partial Orderings and the Maximal Principle

A relation R is said to be *antisymmetric* on a set $X$ if $x \mathrel{R} y$ and $y \mathrel{R} x$ imply $x = y$ for all $x$ and $y$ in $X$. A relation $\prec$ is said to be a **partial ordering** of a set $X$ (or to partially order $X$) if it is transitive and antisymmetric on $X$. Thus $\leq$ is a partial ordering on the real numbers and $\subset$ is a partial ordering on $\mathcal{P}(X)$. A partial ordering $\prec$ on a set $X$ is said to be a **linear ordering** (or simple ordering) of $X$ if for any two elements $x$ and $y$ of $X$ we have either $x \prec y$ or $y \prec x$. Thus $\leq$ linearly orders the set of real numbers, while $\subset$ is not a linear ordering on $\mathcal{P}(X)$.

If $\prec$ is a partial order on $X$ and if $a \prec b$, we often say that $a$ *precedes* $b$ or that $b$ *follows* $a$. Sometimes we say that $a$ is less than $b$ or $b$ greater than $a$. If $E \subset X$, we say that an element $a \in E$ is the *first element* in $E$ or the smallest element in $E$ if, whenever $x \in E$ and $x \neq a$, we have $a \prec x$. Similarly for last (or largest) elements. An element $a \in E$ is called a *minimal element* of $E$ if there is no $x \in E$ with $x \neq a$ and $x \prec a$, and similarly for maximal elements. It should be observed that if a set has a smallest element, then that element is a minimal element. If $\prec$ is a linear ordering, a minimal element is a least element, but in general it is possible to have minimal elements which are not least elements.

Our definition of partial order makes no assertion about the possibility or necessity of $x \prec x$. If we have $x \prec x$ for all $x$, we call $\prec$ a *reflexive* partial order. If we never have $x \prec x$, then $\prec$ is called a *strict* partial order. Thus $<$ is a strict partial order for the real numbers and $\leq$ is a reflexive partial order. To any partial order $\prec$ there is associated a unique strict partial order and a unique reflexive partial order that agree with $\prec$ for all $\langle x, y\rangle$ with $x \neq y$. If $<$ is any partial order, we use $\leq$ for the associated reflexive partial order.

The following principle is equivalent to the axiom of choice and is often more convenient to apply. For a proof of this equivalence and a discussion of related principles, see Suppes [14], Chapter 8, or Kelley [9], pp. 31–36.

**Hausdorff Maximal Principle:** *Let $\prec$ be a partial ordering on a set $X$. Then there is a maximal linearly ordered subset $S$ of $X$, that is, a subset $S$ of $X$ which is linearly ordered by $\prec$ and has the property that if $S \subset T \subset X$ and $T$ is linearly ordered by $\prec$, then $S = T$.*

## Problems

**29.** Let $\prec$ be a partial order on $X$. Then there is a unique strict partial order $<$ and a unique reflexive partial order $\leq$ on $X$ such that for $x \neq y$ we have $x \prec y \Leftrightarrow x < y \Leftrightarrow x \leq y$.

**30.** Give an example of a partially ordered set that has a unique minimal element but no smallest element.

## 9 Well Ordering and the Countable Ordinals

A strict linear ordering $<$ on a set $X$ is called a **well ordering** for $X$ or is said to well order $X$ if every nonempty subset of $X$ contains a first element. Thus, if we take $X = \mathbf{N}$ and $<$ to mean less than, then $\mathbf{N}$ is well ordered by $<$. On the other hand, the set $\mathbf{R}$ of all real numbers is not well ordered by the relation "less than." The following principle clearly implies the axiom of choice and can be shown equivalent to it (see Suppes [14], Chapter 8, or Kelley [9], pp. 31–36).

**Well-Ordering Principle:** *Every set $X$ can be well ordered; that is, there is a relation $<$ that well orders $X$.*

**8. Proposition:** *There is an uncountable set $X$ that is well ordered by a relation $<$ in such a way that:*

i. *There is a last element $\Omega$ in $X$.*
ii. *If $x \in X$ and $x \neq \Omega$, then the set $\{y \in X : y < x\}$ is countable.*

***Proof:*** Let $Y$ be any uncountable set, say the one given by Problem 23. By the well-ordering principle, there is a well ordering $<$ for $Y$. If $Y$ does not have a last element, take an element $\alpha \notin Y$, replace $Y$ by $Y \cup \{\alpha\}$, and extend the order $<$ by setting $y < \alpha$ for all $y \in Y$. This new $Y$ has a last element and is well ordered by $<$. The set of $y$ in $Y$ for which the set $\{x \in Y : x < y\}$ is uncountable is a nonempty set, since it contains the last element of $Y$. Let $\Omega$ be the smallest element in this set, and let $X = \{x \in Y : x < \Omega \text{ or } x = \Omega\}$. Then $X$ is the required set. ▮

The well-ordered set $X$ given in the proposition will be very useful for constructing examples. It can be shown to be unique in the sense that if $Y$ is any other well-ordered set with the same properties, then there is a one-to-one order preserving correspondence between $X$ and $Y$. The last element $\Omega$ in $X$ is called the *first uncountable ordinal* and $X$ is called the set of ordinals less than or equal to the first uncountable ordinal. The elements $x < \Omega$ are called *countable ordinals*. If $\{y : y < x\}$ is finite, we call $x$ a *finite ordinal*. If $\omega$ is the first *nonfinite ordinal*, then $\{x: x < \omega\}$ is the set of finite ordinals and is equivalent, as an ordered set, to the set $\mathbf{N}$ of positive integers.

## Problems

**31. a.** Show that any subset of a well-ordered set is well ordered.

**b.** If $<$ is a partial order on $X$ with the property that every nonempty subset of $X$ has a least element, then $<$ is a linear ordering and consequently a well ordering.

**32.** Let $Y$ be the set of ordinals less than the first uncountable ordinal; i.e., $Y = \{x \in X: x < \Omega\}$. Show that every countable subset $E$ of $Y$ has an upper bound in $Y$ and hence a least upper bound. (An element $b$ is an upper bound for $E$ if $x \leq b$ for each $x \in E$; it is a least upper bound if $b \leq b^*$ for each upper bound $b^*$.)

**33.** A subset $S$ of a well-ordered set $X$ is called a *segment* if

$$S = \{x \in X: x < y\}$$

for some $y \in X$ or if $S = X$. Show that a union of segments is again a segment.

**34.** Let $X$ and $Y$ be two well-ordered sets. A function $f$ from $X$ to $Y$ is called successor preserving if for each $x \in X$ the element $f(x)$ is the first element of $Y$ not in $f[\{z: z < x\}]$.

**a.** Show that there is at most one successor-preserving map of $X$ into $Y$.

**b.** Show that the range of a successor-preserving map is a segment.

**c.** Show that if $f$ is successor preserving, then $f$ is a one-to-one order-preserving map and $f^{-1}$ is also order preserving.

**d.** If $f$ is a successor-preserving map of $X$ into $Y$, then the restriction of $f$ to a segment is also successor preserving.

**e.** If $X$ and $Y$ are well-ordered sets, then there is a successor-preserving map from one of them onto a segment of the other; that is, either there is a successor-preserving map of $X$ onto a segment of $Y$ or a successor-preserving map of $Y$ onto a segment of $X$. [Hint: Consider the collection of all those segments of $X$ on which there is a successor-preserving map into $Y$. Show that there is a successor-preserving map $f$ of the union $S$ of this collection into $Y$ and that either $S = X$ or $f[S] = Y$.]

**f.** Show that the well-ordered set $X$ in Proposition 8 is unique up to isomorphism.

# Part One

## THEORY OF FUNCTIONS OF A REAL VARIABLE

# 2 The Real Number System

## 1 Axioms for the Real Numbers

We assume that the reader has a familiarity with the set **R** of real numbers and those basic properties of real numbers which are usually treated in an undergraduate course in analysis. The present chapter is devoted to a review and systematization of those results which will be useful later.

One approach to the subject of real numbers is to define them as Dedekind cuts of rational numbers, the rational numbers in turn being defined in terms of the natural numbers. Such a program gives an elegant construction of the real numbers out of more primitive concepts and set theory. We shall not concern ourselves here with the construction of the real numbers but will think of them as already given, and list a set of axioms for them. All the properties we need are consequences of these axioms, and in fact these axioms completely characterize the real numbers.

We thus assume as given the set **R** of real numbers, the set $P$ of positive real numbers, and the functions ' + ' and '·' on $\mathbf{R} \times \mathbf{R}$ to **R** and assume that these satisfy the following axioms, which we list in three groups. The first group describes the algebraic properties and the second the order properties. The third comprises the least upper bound axiom.

**A. The Field Axioms:** *For all real numbers $x$, $y$, and $z$ we have:*

A1. $x + y = y + x$.
A2. $(x + y) + z = x + (y + z)$.
A3. $\exists 0 \in \mathbf{R}$ *such that* $x + 0 = x$ *for all* $x \in \mathbf{R}$.

A4. *For each* $x \varepsilon \mathbf{R}$ *there is a* $w \varepsilon \mathbf{R}$ *such that* $x + w = 0$.
A5. $xy = yx$.
A6. $(xy)z = x(yz)$.
A7. $\exists 1 \varepsilon \mathbf{R}$ *such that* $1 \neq 0$ *and* $x \cdot 1 = x$ *for all* $x \varepsilon \mathbf{R}$.
A8. *For each* $x$ *in* $\mathbf{R}$ *different from* $0$ *there is* $w \varepsilon \mathbf{R}$ *such that* $xw = 1$.
A9. $x(y + z) = xy + xz$.

Any set that satisfies these axioms is called a field (under $+$ and $\cdot$). It follows from A1 that the 0 in A3 is unique, a fact which we have assumed in the formulation of A4, A7, and A8. The $w$ in A4 is unique and denoted by '$-x$'. We define subtraction $x - y$ as $x + (-y)$. The 1 in A7 is unique. The $w$ in A8 can be shown to be unique and is denoted by '$x^{-1}$'. If we have a field, that is, any system satisfying A1 through A9, we can perform all the operations of elementary algebra, including the solution of simultaneous linear equations. We shall use the various consequences of these axioms without explicit mention.

The second class of properties possessed by the real numbers have to do with the fact that the real numbers are ordered. We could axiomatize the notion of $a$ less than $b$, but it is somewhat more convenient to use the notation of a positive real number as the primitive one. When we do this our second group of axioms takes the following form:

**B. Axioms of Order:** *The subset* $P$ *of positive real numbers satisfies the following:*

B1. $(x, y \varepsilon P) \Rightarrow x + y \varepsilon P$.
B2. $(x, y \varepsilon P) \Rightarrow xy \varepsilon P$.
B3. $(x \varepsilon P) \Rightarrow -x \notin P$.
B4. $(x \varepsilon \mathbf{R}) \Rightarrow (x = 0)$ *or* $(x \varepsilon P)$ *or* $(-x \varepsilon P)$.

Any system satisfying the axioms of groups A and B is called an **ordered field.** Thus the real numbers are an ordered field. The rational numbers give another example of an ordered field.

In an ordered field we define the notion $x < y$ to mean $y - x \varepsilon P$. We write '$x \leq y$' for '$x < y$ or $x = y$'. In terms of $<$ Axiom B1 is equivalent to

$$(x < y \ \& \ z < w) \Rightarrow x + z < y + w,$$

and B2 is equivalent to

$$(0 < x < y \ \& \ 0 < z < w) \Rightarrow xz < yw.$$

Axiom B3 asserts that a number cannot be both greater than and less than another, while B4 states that of any two different numbers one must be the larger. Since Axiom B1 implies that the relation $<$ is transitive, we see that the real numbers are linearly ordered by $<$. Except for the discussion in the beginning of the next section we take all the consequences of these two axiom groups for granted and use them without explicit mention. For further properties of ordered fields, the reader should see Birkhoff and MacLane [2].

The third axiom group consists of a single axiom, and it is this axiom that distinguishes the real numbers from other ordered fields. In contrast to our cavalier policy about the consequences of the first two axiom groups, we shall be explicit about the use of this last axiom. Before stating this final axiom, let us introduce some terminology: If $S$ is a set of real numbers, we say that $b$ is an **upper bound** for $S$ if for each $x \in S$ we have $x \leq b$. We sometimes express this by writing $S \leq b$. A number $c$ is called a **least upper bound** for $S$ if it is an upper bound for $S$ and if $c \leq b$ for each upper bound $b$ of $S$. Clearly, the least upper bound of a set $S$ is unique if it exists. Our final axiom for real numbers simply guarantees its existence for sets with an upper bound.

**C. Completeness Axiom:** *Every nonempty set $S$ of real numbers which has an upper bound has a least upper bound.*

As a consequence of Axiom C we have the following proposition:

**1. Proposition:** *Let $L$ and $U$ be nonempty subsets of $\mathbf{R}$ with $\mathbf{R} = L \cup U$ and such that for each $l$ in $L$ and each $u$ in $U$ we have $l < u$. Then either $L$ has a greatest element or $U$ has a least element.*

We shall often denote the least upper bound of $S$ by $\sup S$ or by $\sup\limits_{x \in S} x$ and occasionally by $\sup \{x : x \in S\}$. We can define lower bounds and greatest lower bounds in a similar fashion, and it follows from Axiom C that every set of real numbers with a lower bound has a greatest lower bound. We denote the greatest lower bound of a set $S$ by $\inf S$ or by $\inf\limits_{x \in S} x$. Note that $\inf\limits_{x \in S} x = -\sup\limits_{x \in S} -x$.

## Problems

1. Show that $1 \varepsilon P$.

2. Use Axiom C to show that every nonempty set of real numbers with a lower bound has a greatest lower bound.

3. Prove Proposition 1 using Axiom C.

4. If $x$ and $y$ are two real numbers, we define max $(x, y)$ to be $x$ if $x \geq y$ and $y$ if $y \geq x$. We often denote max $(x, y)$ by '$x \vee y$'. Similarly, we define min $(x, y)$ to be the smaller of $x$ and $y$, and denote it by '$x \wedge y$'. Show that

   a. $(x \wedge y) \wedge z = x \wedge (y \wedge z)$.

   b. $x \wedge y + x \vee y = x + y$.

   c. $(-x) \wedge (-y) = -(x \vee y)$.

   d. $x \vee y + z = (x + z) \vee (y + z)$.

   e. $z(x \vee y) = (zx) \vee (zy)$ if $z \geq 0$.

5. We define $|x|$ to be $x$ if $x \geq 0$ and to be $-x$ if $x < 0$. Show that

   a. $|xy| = |x|\,|y|$.

   b. $|x + y| \leq |x| + |y|$.

   c. $|x| = x \vee (-x)$.

   d. $x \vee y = \frac{1}{2}(x + y + |x - y|)$.

   e. if $-y \leq x \leq y$, then $|x| \leq y$.

## 2 The Natural and Rational Numbers as Subsets of R

We have adopted the procedure of taking the natural numbers for granted and of using them as counting numbers. Yet we all regard such a number as 3 not only as a natural number but also as a real number. In fact, we have used the symbol 1 not only to denote the first natural number but also the special real number given by Axiom A7. One is tempted to define the real number 3 as $1 + 1 + 1$, and, in a "similar" fashion, we can define real numbers corresponding to any natural number. We may use the tools at hand to do this in a more precise fashion.

By the principle of recursive definition there is a function $\varphi$ from the natural numbers to the real numbers defined by $\varphi(1) = 1$ and $\varphi(n + 1) = \varphi(n) + 1$. (Here 1 denotes a real number on the right side and a natural number on the left.) We shall show that the mapping $\varphi$ is a one-to-one mapping of **N** into **R**. Let $p$ and $q$ be two different natural numbers, say $p < q$. Then $q = p + n$, and we shall show that $\varphi(p) < \varphi(q)$ by induction on $n$. For $n = 1$ we have $q = p + 1$ and $\varphi(q) = \varphi(p) + 1 > \varphi(p)$. For general $n$ we have $\varphi(p + n + 1) = \varphi(p + n) + 1 > \varphi(p + n)$, and so $\varphi(p + n) > \varphi(p)$ implies $\varphi(p + n + 1) > \varphi(p)$. Thus by induction $\varphi(p + n) > \varphi(p)$, and we see that the mapping $\varphi$ is one-to-one. We can also prove by induction that $\varphi(p + q) = \varphi(p) + \varphi(q)$ and $\varphi(pq) = \varphi(p)\varphi(q)$. Thus $\varphi$ gives a one-to-one correspondence between the natural numbers and a subset of **R**, and $\varphi$ preserves sums, products, and the relation $<$. Strictly speaking, we should distinguish between the natural number $n$ and its image $\varphi(n)$ under $\varphi$, but we shall not make the distinction here; we shall consider the set **N** of natural numbers to be a subset of **R**. By taking differences of natural numbers, we obtain the integers as a subset of **R**, and taking quotients of integers gives us the rationals. Since Axiom C was not used in this discussion, the same results hold for any ordered field. Thus we have shown the following:

**2. Proposition:** *Every ordered field contains (sets isomorphic to) the natural numbers, the integers, and the rational numbers.*

If we make use of Axiom C, we can prove several further facts about the integers and rational numbers as subsets of the reals. One of the most important is the following theorem, which for historical reasons[1] is called the Axiom of Archimedes:

**3. Axiom of Archimedes:** *Given any real number $x$, there is an integer $n$ such that $x < n$.*

***Proof:*** If $x < 0$, take $n = 0$. Otherwise the set of integers $k$ such that $k \leq x$ is nonempty. Since $S$ has the upper bound $x$, it has a least upper bound $y$ by Axiom C. Since $y$ is the least upper bound for $S$, $y - \frac{1}{2}$ cannot be an upper bound for $S$, and so there is a $k \in S$ such that $k > y - \frac{1}{2}$. But $k + 1 > y + \frac{1}{2} > y$, and so $(k + 1) \notin S$. Since $k + 1$ is an integer not in $S$, we must have $k + 1$ greater than $x$ by the definition of $S$. ∎

**4. Corollary:** *Between any two real numbers is a rational; that is, if $x < y$, then there is a rational $r$ with $x < r < y$.*

[1] Or rather, unhistorical: Archimedes explicitly attributes it to Eudoxus.

***Proof:*** Let us first suppose $0 \leq x$. By the Axiom of Archimedes there is an integer $q > (y - x)^{-1}$. Then $(1/q) < y - x$. The set of integers $n$ such that $y \leq (n/q)$ is a nonempty (by the Axiom of Archimedes) set of positive integers, and so has a least element $p$. Then $(p - 1)/q < y \leq (p/q)$, and $x = y - (y - x) < (p/q) - (1/q) = (p - 1)/q$. Thus $r = (p - 1)/q$ lies between $x$ and $y$. If $x < 0$, we can find an integer $n$ such that $n > -x$. Then $n + x > 0$, and there is a rational $r$ with $n + x < r < n + y$, and $r - n$ is a rational between $x$ and $y$. ▮

## 3 The Extended Real Numbers

It is often convenient to extend the system of real numbers by the addition of two elements, $+\infty$ and $-\infty$. This enlarged set is called the set of **extended real numbers.** We extend the definition of $<$ to the extended real numbers by postulating $-\infty < x < \infty$, for each real number $x$. We define

$$x + \infty = \infty, \qquad x - \infty = -\infty$$

$$x \cdot \infty = \infty \qquad \text{if } x > 0$$

$$x \cdot -\infty = -\infty \qquad \text{if } x > 0$$

for all real numbers $x$, and set

$$\infty + \infty = \infty, \quad -\infty - \infty = -\infty$$

$$\infty \cdot (\pm\infty) = \pm\infty, \quad -\infty \cdot (\pm\infty) = \mp\infty.$$

The operation $\infty - \infty$ is left undefined, but we shall adopt the arbitrary *convention* that $0 \cdot \infty = 0$.

One use of extended real numbers is in the expression "sup $S$." If $S$ is a nonempty set of real numbers with an upper bound, we define sup $S$ to be the least upper bound of $S$. If $S$ has no upper bound, we write sup $S = \infty$. Then sup $S$ is defined for all nonempty sets $S$. If we define sup $\emptyset$ to be $-\infty$, then in all cases sup $E$ is the smallest extended real number which is greater than or equal to each element of $E$. Similar conventions are adopted with respect to inf $S$.

A function whose values are in the set of extended real numbers is called an *extended real-valued function.*

## Problem

6. Show that $\inf E < \sup E$ if and only $E \neq \emptyset$.

## 4 Sequences of Real Numbers

By a sequence[2] $\langle x_n \rangle$ of real numbers we mean a function that maps each natural number $n$ into the real number $x_n$. We say that a real number $l$ is a *limit* of the sequence $\langle x_n \rangle$ if for each positive $\epsilon$ there is an $N$ such that for all $n \geq N$ we have $|x_n - l| < \epsilon$. It is easily verified that a sequence can have at most one limit, and we denote this limit by $\lim x_n$ when it exists. In symbols $l = \lim x_n$ if

$$(\epsilon > 0)(\exists N)(n \geq N)(|x_n - l| < \epsilon).$$

A sequence $\langle x_n \rangle$ of real numbers is called a *Cauchy sequence* if given $\epsilon > 0$, there is an $N$ such that for all $n \geq N$ and all $m \geq N$ we have $|x_n - x_m| < \epsilon$. The *Cauchy Criterion* states that a sequence of real numbers converges if and only if it is a Cauchy sequence (see Problem 11).

We extend this notion of limit of a sequence to include the value $\infty$ as follows: $\lim x_n = \infty$ if given $\Delta$ there is an $N$ such that for all $n \geq N$ we have $x_n > \Delta$. A sequence is called *convergent* if it has a limit. This definition is ambiguous; it depends on whether or not we mean a limit which is a real number or an extended real number. In most of analysis it is more usual to use the restricted definition for convergence, which requires a limit to be a real number, but we shall find it convenient in the next few chapters to allow $\pm\infty$ as limits in good standing. In those cases in which it is important to distinguish between the two concepts of a limit we shall try to be explicit by the use of such phrases as 'converges to a real number' or 'converges in the set of extended real numbers.'

If $l = \lim x_n$, we often write $x_n \to l$. If, in addition, $\langle x_n \rangle$ is monotone, that is, $x_n \leq x_{n+1}$, we write $x_n \uparrow l$.

In the case of a real number we can paraphrase the definition of limit as follows: $l$ is the limit of $\langle x_n \rangle$ if given $\epsilon > 0$, all but a finite number of terms of the sequence $\langle x_n \rangle$ are within $\epsilon$ of $l$. A weaker requirement is to have infinitely many terms of the sequence within $\epsilon$

[2] These are *infinite* sequences in the terminology of Chapter 1. Since we shall be mostly interested in infinite sequences in the remainder of this book, we drop the adjective "infinite" and assume all sequences infinite unless otherwise specified.

of $l$. In this case we say that $l$ is a **cluster point** of the sequence $\langle x_n \rangle$. Thus $l$ is a cluster point of $\langle x_n \rangle$ if, given $\epsilon > 0$ and given $N$, $\exists n \geq N$ such that $|x_n - l| < \epsilon$. We extend this definition to the case $l = \infty$ by saying that $\infty$ is a cluster point of $\langle x_n \rangle$ if, given $\Delta$ and given $N$, $\exists n \geq N$ such that $x_n \geq \Delta$. An obvious modification applies to $-\infty$. Thus, if a sequence has a limit $l$, then $l$ is a cluster point, but the converse is not usually true. For example, the sequence $\langle x_n \rangle$ defined by $x_n = (-1)^n$ has $+1$ and $-1$ as cluster points but has no limit.

If $\langle x_n \rangle$ is a sequence, we define its limit superior by

$$\overline{\lim}\ x_n = \inf_n \sup_{k \geq n} x_k .$$

The symbols $\overline{\lim}$ and lim sup are both used for the limit superior. A real number $l$ is the limit superior of the sequence $\langle x_n \rangle$ if and only if (i) given $\epsilon > 0$, $\exists n$ such that $x_k < l + \epsilon$ for all $k \geq n$, and (ii) given $\epsilon > 0$ and given $n$, $\exists k \geq n$ such that $x_k > l - \epsilon$. The extended real number $\infty$ is the limit superior of $\langle x_n \rangle$ if and only if given $\Delta$ and $n$, there is a $k \geq n$ such that $x_k > \Delta$. The extended real number $-\infty$ is the limit superior of $\langle x_n \rangle$ if and only if $-\infty = \lim x_n$.

We define the limit inferior by

$$\underline{\lim}\ x_n = \sup_n \inf_{k \geq n} x_k .$$

We have $\overline{\lim}\ (-x_n) = -\underline{\lim}\ x_n$, and $\underline{\lim}\ x_n \leq \overline{\lim}\ x_n$. The sequence $\langle x_n \rangle$ converges to an extended real number $l$ if and only if $l = \underline{\lim}\ x_n = \overline{\lim}\ x_n$. If $\langle x_n \rangle$ and $\langle y_n \rangle$ are two sequences, we have

$$\begin{aligned} \underline{\lim}\ x_n + \underline{\lim}\ y_n \leq \underline{\lim}\ (x_n + y_n) &\leq \overline{\lim}\ x_n + \underline{\lim}\ y_n \\ &\leq \overline{\lim}\ (x_n + y_n) \leq \overline{\lim}\ x_n + \overline{\lim}\ y_n , \end{aligned}$$

provided no sum is of the form $\infty - \infty$.

## Problems

**7.** Show that a sequence can have at most one limit.

**8.** Show that $l$ is a cluster point of $\langle x_n \rangle$ if and only if there is a subsequence $\langle x_{n_j} \rangle_{j=1}^{\infty}$ that converges to $l$.

**9. a.** Show that $\overline{\lim}\ x_n$ and $\underline{\lim}\ x_n$ are the largest and smallest cluster points of the sequence $\langle x_n \rangle$.

**b.** Show that every bounded infinite sequence has a subsequence that converges to a real number.

**10.** Show that a sequence $\langle x_n \rangle$ is convergent if and only if there is exactly one extended real number that is a cluster point of the sequence. Is this statement true if we omit the word "extended"?

**11. a.** Show that a sequence $\langle x_n \rangle$ which converges to a real number $l$ is a Cauchy sequence.

**b.** Show that each Cauchy sequence is bounded.

**c.** Show that if a Cauchy sequence has a subsequence that converges to $l$, then the original sequence converges to $l$.

**d.** Establish the Cauchy Criterion: There is a real number $l$ to which the sequence $\langle x_n \rangle$ converges if and only if $\langle x_n \rangle$ is a Cauchy sequence.

**12.** Show that $x = \lim x_n$ if and only if every subsequence of $\langle x_n \rangle$ has in turn a subsequence that converges to $x$.

**13.** Show that the real number $l$ is the limit superior of the sequence $\langle x_n \rangle$ if and only if (i) given $\epsilon > 0$, $\exists n$ such that $x_k < l + \epsilon$ for all $k \geq n$, and (ii) given $\epsilon > 0$ and $n$, $\exists k \geq n$ such that $x_k > l - \epsilon$.

**14.** Show that $\overline{\lim}\, x_n = \infty$ if and only if given $\Delta$ and $n$, $\exists k \geq n$ with $x_k > \Delta$.

**15.** Show that $\underline{\lim}\, x_n \leq \overline{\lim}\, x_n$ and that $\underline{\lim}\, x_n = \overline{\lim}\, x_n = l$ if and only if $l = \lim x_n$.

**16.** Prove that

$$\overline{\lim}\, x_n + \underline{\lim}\, y_n \leq \overline{\lim}\, (x_n + y_n) \leq \overline{\lim}\, x_n + \overline{\lim}\, y_n,$$

provided the right and left sides are not of the form $\infty - \infty$.

**17.** Prove that if $x_n > 0$ and $y_n \geq 0$, then

$$\overline{\lim}\, (x_n y_n) \leq (\overline{\lim}\, x_n)(\overline{\lim}\, y_n),$$

provided the product on the right is not of the form $0 \cdot \infty$.

**18.** We shall say that a sequence (or series) $\langle x_n \rangle$ is **summable** to the real number $s$ or has a sum $s$ if the sequence $\langle s_n \rangle$ defined by $s_n = \sum_{\nu=1}^{n} x_\nu$ has $s$ as a limit. In this case we write $s = \sum_{\nu=1}^{\infty} x_\nu$. Show that if each $x_\nu \geq 0$, there is always an extended real number $s$ such that

$$s = \sum_{\nu=1}^{\infty} x_\nu .$$

**19.** Show that the series $\langle x_\nu \rangle$ has a sum if

$$\sum_{\nu=1}^{\infty} |x_\nu| < \infty.$$

**20.** Let $\langle x_n \rangle$ be a sequence of real numbers. Show that $x = \lim_{n \to \infty} x_n$ if and only if

$$x = x_1 + \sum_{\nu=1}^{\infty} (x_{\nu+1} - x_\nu).$$

**21.** Let $E$ be a set of positive real numbers. We define $\sum_{x \varepsilon E} x$ to be $\sup_{F \varepsilon \mathcal{F}} s_F$, where $\mathcal{F}$ is the collection of finite subsets of $E$ and $s_F$ is the (finite) sum of the elements of $F$.

**a.** Show that $\sum_{x \varepsilon E} x < \infty$ only if $E$ is countable.

**b.** Show that if $E$ is countable and $\langle x_n \rangle$ is a one-to-one mapping of $\mathbf{N}$ onto $E$, then $\sum_{x \varepsilon E} x = \sum_{n=1}^{\infty} x_n$.

**22.** Let $p$ be an integer greater than 1, and $x$ a real number, $0 < x < 1$. Show that there is a sequence $\langle a_n \rangle$ of integers with $0 \le a_n < p$ 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 $q/p^n$, in which case there are exactly two such sequences. Show that, conversely, if $\langle a_n \rangle$ is any sequence of integers with $0 \le a_n < p$, the series

$$\sum_{n=1}^{\infty} \frac{a_n}{p^n}$$

converges to a real number $x$ with $0 \le x \le 1$. 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.

**23.** Show that $\mathbf{R}$ is uncountable. [Use Problem 1.24. Another proof will be given by Corollary 3.4.]

## 5 Open and Closed Sets of Real Numbers

The simplest sets of real numbers are the intervals. We define the open interval $(a, b)$ to be the set $\{x: a < x < b\}$. We always take $a < b$, but we consider also the infinite intervals $(a, \infty) = \{x: a < x\}$ and $(-\infty, b) = \{x: x < b\}$. Sometimes we write $(-\infty, \infty)$ for the set of all real numbers. We define the closed interval $[a, b]$ to be the set $\{x: a \le x \le b\}$. For closed intervals we take $a$ and $b$ finite but always assume that $a < b$. The half-open interval $(a, b]$ is defined to be

$\{x: a < x \leq b\}$, and $[a, b) = \{x: a \leq x < b\}$. A generalization of the notion of an open interval is given by that of an open set:

**Definition:** *A set $O$ of real numbers is called* ***open*** *if for each $x \in O$ there is a $\delta > 0$ such that each $y$ with $|x - y| < \delta$ belongs to $O$.*

Another way of phrasing this definition is to say that a set $O$ is open if for every $x$ in $O$ there is an open interval $I$ such that $x \in I \subset O$. The open intervals are examples of open sets, and both the empty set $\emptyset$ and the set $\mathbf{R}$ of real numbers are open. We establish some properties of open sets.

**5. Proposition:** *The intersection $O_1 \cap O_2$ of two open sets $O_1$ and $O_2$ is open.*

***Proof:*** Let $x \in O_1 \cap O_2$. Since $x \in O_1$ and $O_1$ is open, there is a $\delta_1 > 0$ such that all $y$ with $|x - y| < \delta_1$ belong to $O_1$. Similarly, there is a $\delta_2 > 0$ such that all $y$ with $|x - y| < \delta_2$ belong to $O_2$. Take $\delta$ to be the smaller of $\delta_1$ and $\delta_2$. Then $\delta > 0$, and if $|x - y| < \delta$, then $y$ belongs to both $O_1$ and $O_2$, i.e., to $O_1 \cap O_2$. ▮

**6. Corollary:** *The intersection of any finite collection of open sets is open.*

**7. Proposition:** *The union of any collection $\mathcal{C}$ of open sets is open.*

***Proof:*** Let $U$ be the union of the collection $\mathcal{C}$, and $x \in U$. Then there is an $O \in \mathcal{C}$ with $x \in O$. Since $O$ is open, there is an $\epsilon > 0$ such that all $y$ with $|x - y| < \epsilon$ belong to $O$ and hence to $U$, since $O \subset U$. Thus $U$ is open. ▮

It follows from Proposition 5 that the intersection of any *finite* collection of open sets is open. It is not true, however, that the intersection of any collection of open sets is open. Take, for example, $O_n$ to be the open interval $(-1/n, 1/n)$. Then $\bigcap_{n=1}^{\infty} O_n = \{0\}$, and $\{0\}$ is not an open set.

Every union of open intervals is an open set by Proposition 7. A strong form of the converse of this is also true:

**8. Proposition:** *Every open set of real numbers is the union of a countable collection of disjoint open intervals.*

***Proof:*** Since $O$ is open, for each $x \varepsilon O$ there is a $y > x$ such that $(x, y) \subset O$. Let $b = \sup \{y: (x, y) \subset O\}$. Let $a = \inf \{z: (z, x) \subset O\}$. Then $a < x < b$, and $I_x = (a, b)$ is an open interval containing $x$. Now $I_x \subset O$, for if $w \varepsilon I_x$, say $x < w < b$, we have by the definition of $b$ a number $y > w$ such that $(x, y) \subset O$, and so $w \varepsilon O$. Moreover, $b \notin O$, for if $b \varepsilon O$, then for some $\epsilon > 0$ we have $(b - \epsilon, b + \epsilon) \subset O$, whence $(x, b + \epsilon) \subset O$, contradicting the definition of $b$. Similarly, $a \notin O$. Consider the collection of open intervals $\{I_x\}$, $x \varepsilon O$. Since each $x$ in $O$ is contained in $I_x$, and each $I_x$ is contained in $O$, we have $O = \bigcup I_x$. Let $(a, b)$ and $(c, d)$ be two intervals in this collection with a point in common. Then we must have $c < b$ and $a < d$. Since $c$ does not belong to $O$, it does not belong to $(a, b)$ and we have $c \leq a$. Since $a$ does not belong to $O$ and hence not to $(c, d)$, we have $a \leq c$. Thus $a = c$. Similarly, $b = d$, and $(a, b) = (c, d)$. Thus two *different* intervals in the collection $\{I_x\}$ must be disjoint. Thus $O$ is the union of the disjoint collection $\{I_x\}$ of open intervals, and it remains only to show that this collection is countable. But each open interval contains a rational number by the corollary to the Axiom of Archimedes. Since we have a collection of disjoint open intervals, each open interval contains a different rational number, and the collection can be put in one-to-one correspondence with a subset of the rationals. Thus it is a countable collection. ∎

**9. Proposition (Lindelöf):** *Let $\mathcal{C}$ be a collection of open sets of real numbers. Then there is a countable subcollection $\{O_i\}$ of $\mathcal{C}$ such that*

$$\bigcup_{O \varepsilon \mathcal{C}} O = \bigcup_{i=1}^{\infty} O_i.$$

***Proof:*** Let $U = \bigcup \{O: O \varepsilon \mathcal{C}\}$, and $x \varepsilon U$. Then there is an $O \varepsilon \mathcal{C}$ with $x \varepsilon O$. Since $O$ is open, there is an open interval $I_x$ such that $x \varepsilon I_x \subset O$. It follows from Corollary 4 that we can find an open interval $J_x$ with rational endpoints such that $x \varepsilon J_x \subset I_x$. Since the collection of all open intervals with rational endpoints is countable, the collection $\{J_x\}$, $x \varepsilon U$, is countable, and $U = \bigcup_{x \varepsilon U} J_x$. For each interval in $\{J_x\}$ choose a set $O$ in $C$ that contains it. This gives a countable subcollection $\{O_i\}_{i=1}^{\infty}$ of $C$, and $U = \bigcup_{i=1}^{\infty} O_i$. ▮

We shall also study the notion of a closed set, which generalizes the notion of a closed interval. We begin by defining a point of closure:

**Definition:** *A real number $x$ is called a **point of closure** of a set $E$ if for every $\delta > 0$ there is a $y$ in $E$ such that $|x - y| < \delta$.*

This is equivalent to saying that $x$ is a point of closure of $E$ if every open interval containing $x$ also contains a point of $E$. Each point of $E$ is trivially a point of closure of $E$. We denote the set of points of closure of $E$ by $\bar{E}$. Thus $E \subset \bar{E}$.

**10. Proposition:** *If $A \subset B$, then $\bar{A} \subset \bar{B}$. Also, $\overline{(A \cup B)} = \bar{A} \cup \bar{B}$.*

***Proof:*** The first part follows immediately from the definition of points of closure. Since $A \subset A \cup B$, we have $\bar{A} \subset \overline{(A \cup B)}$. Similarly, $\bar{B} \subset \overline{(A \cup B)}$. Thus $\bar{A} \cup \bar{B} \subset \overline{(A \cup B)}$. Suppose that $x \notin \bar{A} \cup \bar{B}$. Then there is a $\delta_1 > 0$ such that there is no $y \in A$ with $|x - y| < \delta_1$, and there is a $\delta_2 > 0$ such that there is no $y \in B$ with $|x - y| < \delta_2$. Thus if $\delta = \min(\delta_1, \delta_2)$, there is no $y \in A \cup B$ with $|x - y| < \delta$. Consequently, $x \notin \overline{(A \cup B)}$, and we have $\overline{(A \cup B)} \subset \bar{A} \cup \bar{B}$. ▮

**Definition:** *A set $F$ is called **closed** if $F = \bar{F}$.*

Since we always have $F \subset \bar{F}$, a set $F$ is closed if $\bar{F} \subset F$, that is, if $F$ contains all of its points of closure. The empty set $\emptyset$ and the set $\mathbf{R}$ of all real numbers are closed. The closed intervals $[a, b]$ and $[a, \infty)$ are closed. It is customary to use the letter $F$ to denote closed sets (French, *fermé*).

**11. Proposition:** *For any set $E$ the set $\bar{E}$ is closed; that is, $\bar{\bar{E}} = \bar{E}$.*

***Proof:*** Let $x$ be a point of closure of $\bar{E}$. Then, given $\delta > 0$, there is a point $y \in \bar{E}$ with $|x - y| < \delta/2$. Since $y \in \bar{E}$, there is a $z \in E$ with $|y - z| < \delta/2$. Thus $|x - z| < \delta$, and we see that $x$ is a point of closure of $E$. ▮

**12. Proposition:** *The union $F_1 \cup F_2$ of two closed sets $F_1$ and $F_2$ is closed.*

***Proof:*** By Proposition 10 we have

$$\overline{(F_1 \cup F_2)} = \bar{F}_1 \cup \bar{F}_2 = F_1 \cup F_2 . \quad \blacksquare$$

**13. Proposition:** *The intersection of any collection $\mathcal{C}$ of closed sets is closed.*

***Proof:*** Let $x$ be a point of closure of $\bigcap \{F: F \in \mathcal{C}\}$. Then, given $\delta > 0$, there is a $y \in \bigcap \{F: F \in \mathcal{C}\}$ such that $|x - y| < \delta$. Since such a $y$ belongs to each $F \in \mathcal{C}$, we see that $x$ is a point of closure of each $F \in \mathcal{C}$. Since each $F$ is closed, we have $x \in F$ for each $F$ in $\mathcal{C}$. Hence $x \in \bigcap \{F: F \in \mathcal{C}\}$. $\blacksquare$

**14. Proposition:** *The complement of an open set is closed and the complement of a closed set is open.*

***Proof:*** Let $O$ be open. If $x \in O$, there is a $\delta > 0$ such that if $|x - y| < \delta$, then $y \in O$. Hence $x$ cannot be a point of closure of $\tilde{O}$, since there is no $y \in \tilde{O}$ with $|x - y| < \delta$. Thus $\tilde{O}$ contains all of its points of closure and is therefore closed.

On the other hand, let $F$ be closed and $x \in \tilde{F}$. Then, since $x$ is not a point of closure of $F$, there is a $\delta > 0$ such that there is no $y \in F$ with $|x - y| < \delta$. Hence, if $|x - y| < \delta$, then $y \in \tilde{F}$. Thus $\tilde{F}$ is open. $\blacksquare$

We say that a collection $\mathcal{C}$ of sets **covers** a set $F$ if $F \subset \bigcup \{O: O \in \mathcal{C}\}$. In this case the collection $\mathcal{C}$ is called a **covering** of $F$. If each $O \in \mathcal{C}$ is open, we call $\mathcal{C}$ an **open** covering of $F$. If $\mathcal{C}$ contains only a finite number of sets, we call $\mathcal{C}$ a **finite** covering. This terminology is inconsistent: In 'open covering' the adjective 'open' refers to the sets in the covering; in 'finite covering' the adjective 'finite' refers to the collection and does not imply that the sets in the collection are finite sets. Thus the term 'open covering' is an abuse of language and should properly be 'covering by open sets'. Unfortunately, the former terminology is well established in mathematics. With this terminology we state the following theorem:

**15. Theorem (Heine–Borel):** *Let $F$ be a closed and bounded set of real numbers. Then each open covering of $F$ has a finite subcovering.*

*That is, if* $\mathcal{C}$ *is a collection of open sets such that* $F \subset \bigcup \{O : O \in \mathcal{C}\}$, *then there is a finite collection* $\{O_1, \ldots, O_n\}$ *of sets in* $\mathcal{C}$ *such that*

$$F \subset \bigcup_{i=1}^{n} O_i.$$

***Proof:*** Let us first consider the case when $F$ is the closed interval $[a, b]$, where $-\infty < a < b < \infty$. Let $E$ be the set of numbers $x \leq b$ with the property that the interval $[a, x]$ can be covered by a finite number of the sets of $\mathcal{C}$. Since $a \in E$, $E$ is nonempty. Since it is bounded by $b$, it has at least upper bound $c$. Since $c \in [a, b]$, there is an $O \in \mathcal{C}$ that contains $c$. Since $O$ is open, there is an $\epsilon > 0$ such that the interval $(c - \epsilon, c + \epsilon)$ is contained in $O$. 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 $\{O_1, \ldots, O_k\}$ of sets in $\mathcal{C}$ which covers $[a, x]$. Consequently, the finite collection $\{O_1, \ldots, O_k, O\}$ covers $[a, c + \epsilon)$. Thus each point of $[c, c + \epsilon)$ would be in $E$ if it were less than or equal to $b$. Since no point of $[c, c + \epsilon)$ except $c$ can belong to $E$, we must have $c = b$ and $b \in E$. Thus $[a, b]$ can be covered by a finite number of sets from $\mathcal{C}$, proving our special case.

Now let $F$ be any closed and bounded set and $\mathcal{C}$ an open covering of $F$. Since $F$ is bounded, it is contained in some closed bounded interval $[a, b]$. Let $\mathcal{C}^*$ be the collection obtained by adding $\tilde{F}$ to $\mathcal{C}$; that is, $\mathcal{C}^* = \mathcal{C} \cup \{\tilde{F}\}$. Since $F$ is closed, $\tilde{F}$ is open, and $\mathcal{C}^*$ is a collection of open sets. By hypothesis $F \subset \bigcup \{O : O \in \mathcal{C}\}$, and so $\mathbf{R} = \tilde{F} \cup F \subset \tilde{F} \cup \bigcup \{O : O \in \mathcal{C}\} = \bigcup \{O : O \in \mathcal{C}^*\}$. Thus $\mathcal{C}^*$ is an open covering of $\mathbf{R}$ and therefore of $[a, b]$. By our previous case there is a finite subcollection of $\mathcal{C}^*$ that covers $[a, b]$ and hence $F$. If this finite subcollection does not contain $\tilde{F}$, it is a subcollection of $\mathcal{C}$ and the conclusion of our theorem holds. If the subcollection contains $\tilde{F}$, denote it by $\{O_1, \ldots, O_n, \tilde{F}\}$. Then $F \subset \tilde{F} \cup O_1 \cup \cdots \cup O_n$. Since no point of $F$ is contained in $\tilde{F}$, we have $F \subset O_1 \cup \cdots \cup O_n$, and the collection $\{O_1, \ldots, O_n\}$ is a finite subcollection of $\mathcal{C}$ which covers $F$. ▌

**16. Proposition:** *Let* $\mathcal{C}$ *be a collection of closed sets (of real numbers) with the property that every finite subcollection of* $\mathcal{C}$ *has a nonempty intersection, and suppose that one of the sets in* $\mathcal{C}$ *is bounded. Then*

$$\bigcap_{F \in \mathcal{C}} F \neq \emptyset.$$

## Problems

**24.** Is the set of rational numbers open or closed?

**25.** What are the sets of real numbers that are both open and closed?

**26.** Find two sets $A$ and $B$ such that $A \cap B = \emptyset$ and $\bar{A} \cap \bar{B} \neq \emptyset$.

**27.** Show that $x$ is a point of closure of $E$ if and only if there is a sequence $\langle y_n \rangle$ with $y_n \in E$ and $x = \lim y_n$.

**28.** A number $x$ is called an **accumulation point** of a set $E$ if it is a point of closure of $E \sim \{x\}$. Show that the set $E'$ of accumulation points of $E$ is a closed set.

**29.** Show that $\bar{E} = E \cup E'$.

**30.** A set is called **isolated** if $E \cap E' = \emptyset$. Show that every isolated set of real numbers is countable.

**31.** A set $D$ is called **dense** in **R** if $\bar{D} = \mathbf{R}$. Show that the set of rational numbers is dense in **R**.

**32.** Prove Propositions 12 and 13 using Propositions 5, 7, and 14.

**33.** Prove Propositions 5 and 7 using Propositions 12, 13, and 14.

**34.** A point $x$ is called an **interior point** of a set $A$ if there is a $\delta > 0$ such that the interval $(x - \delta, x + \delta)$ is contained in $A$. The set of interior points of $A$ is denoted by $A^\circ$. Show that

**a.** $A$ is open if and only if $A = A^\circ$.

**b.** $A^\circ = \sim(\bar{\tilde{A}})$.

**35.** Derive Proposition 16 from the Heine–Borel Theorem using De Morgan's laws.

**36.** Let $\langle F_n \rangle$ be a sequence of nonempty closed sets of real numbers with $F_{n+1} \subset F_n$. Show that if one of the sets $F_n$ is bounded, then $\bigcap_{i=1}^{\infty} F_i \neq \emptyset$. Give an example to show that this conclusion may be false if we do not require one of the sets to be bounded.

**37.** The **Cantor ternary set** $C$ consists of all those real numbers in $[0, 1]$ that have ternary expansion (cf. Problem 22) $\langle a_n \rangle$ for which $a_n$ is never 1. (If $x$ has two ternary expansions, we put $x$ in the Cantor set if *one* of the expansions has no term equal to 1.) Show that $C$ is a closed set, and that $C$ is obtained by first removing the middle third $(\frac{1}{3}, \frac{2}{3})$ from $[0, 1]$, then removing the middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ of the remaining intervals, and so on.

**38.** Show that the Cantor set can be put into a one-to-one correspondence with the interval $[0, 1]$.

**39.** Show that the set of accumulation points of the Cantor set is the Cantor set itself.

## 6 Continuous Functions

Let $f$ be a real-valued function whose domain of definition is a set $E$ of real numbers. We say that $f$ is **continuous at the point** $x$ in $E$ if given $\epsilon > 0$, there is a $\delta > 0$ such that for all $y$ in $E$ with $|x - y| < \delta$ we have $|f(x) - f(y)| < \epsilon$. The function $f$ is said to be **continuous** on a subset $A$ of $E$ if it is continuous at each point of $A$. If we merely say that $f$ is continuous, we mean that $f$ is continuous on its domain.

**17. Proposition:** *Let $f$ be a real-valued function defined and continuous on a closed and bounded set $F$. Then $f$ is bounded on $F$ and assumes its maximum and minimum on $F$; that is, there are points $x_1$ and $x_2$ in $F$ such that $f(x_1) \leq f(x) \leq f(x_2)$ for all $x$ in $F$.*

***Proof:*** We shall first show that $f$ is bounded on $F$. Since $f$ is continuous on $F$, for each $x \in F$ there is an open interval $I_x$ containing $x$ such that $|f(y) - f(x)| < 1$ for $y \in I_x \cap F$. Thus for $y \in I_x \cap F$, we have $|f(y)| \leq |f(x)| + 1$ and so $f$ is bounded in $I_x$. The collection $\{I_x : x \in F\}$ is a collection of open intervals which covers $F$, and by the Heine–Borel Theorem there is a finite subcollection $\{I_{x_1}, \ldots, I_{x_n}\}$ that covers $F$. Let $M = 1 + \max\,[|f(x_1)|, \ldots, |f(x_n)|]$. Then each $y$ in $F$ belongs to some interval $I_{x_k}$ in the finite subcollection, and hence $|f(y)| < 1 + |f(x_k)| \leq M$. This shows that $f$ is bounded (by $M$) on $F$.

To see that $f$ assumes its maximum on $F$, let $m = \sup\limits_{x \in F} f(x)$. Since $f$ is bounded, $m$ is finite, and our goal is to show that there is an $x_1 \in F$ such that $f(x_1) = m$. Suppose not. Then $f(x) < m$ for each $x \in F$, and by continuity there is an open interval $I_x$ containing $x$ such that $f(y) < \frac{1}{2}(f(x) + m)$ for all $y \in I_x \cap F$. Again using the Heine–Borel Theorem, we can find a finite number of these intervals $\{I_{x_1}, \ldots, I_{x_n}\}$ that cover $F$. Set $a = \max\,[f(x_1), \ldots, f(x_n)]$. Then each $y \in F$ belongs to one such interval $I_{x_k}$ and $f(y) < \frac{1}{2}[f(x_k) + m] \leq \frac{1}{2}(a + m)$. Thus $\frac{1}{2}(a + m)$ is a bound for $f$ on $F$. But this is impossible, since $\frac{1}{2}(a + m) < m$. Consequently, there must be an $x_1$ such that $f(x_1) = m$. Similarly, there is an $x_2$ at which $f$ assumes its minimum. ▌

**18. Proposition:** *Let $f$ be a real-valued function defined on $(-\infty, \infty)$. Then $f$ is continuous if and only if for each open set $O$ of real numbers $f^{-1}[O]$ is an open set.*

***Proof:*** Suppose $f^{-1}[O]$ is open for each open set $O$, and let $x$ be an arbitrary real number. Then, given $\epsilon > 0$, the interval

$I = (f(x) - \epsilon,\ f(x) + \epsilon)$ is an open set, and so its inverse image $f^{-1}[I]$ must be open. Since $x \in f^{-1}[I]$, there must be some $\delta > 0$ such that $(x - \delta,\ x + \delta) \subset f^{-1}[I]$. But this implies that if $|y - x| < \delta$, then $f(y) \in (f(x) - \epsilon, f(x) + \epsilon)$; that is, $|f(x) - f(y)| < \epsilon$. Hence $f$ is continuous at $x$. Since $x$ was arbitrary, $f$ is continuous.

Suppose now that $f$ is continuous and that $O$ is an open set. Let $x$ be a point of $f^{-1}[O]$. Then $f(x) \in O$, and there is an $\epsilon > 0$ such that $(f(x) - \epsilon, f(x) + \epsilon) \subset O$. Since $f$ is continuous at $x$, there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for $|x - y| < \delta$. Thus, for every $y \in (x - \delta,\ x + \delta)$, we have $f(y) \in (f(x) - \epsilon,\ f(x) + \epsilon) \subset O$. Hence $(x - \delta,\ x + \delta) \subset f^{-1}[O]$, and so $f^{-1}[O]$ is open. ▮

**19. Proposition (Intermediate Value Theorem):** *Let $f$ be a continuous real-valued function on $[a, b]$ and suppose that $f(a) \le \gamma \le f(b)$ [or $f(b) \le \gamma \le f(a)$]; then there is a point $c \in [a, b]$ such that $f(c) = \gamma$.*

**Definition:** *A real-valued function $f$ defined on a set $E$ is said to be* ***uniformly continuous*** *(on $E$) if given $\epsilon > 0$, there is a $\delta > 0$ such that for all $x$ and $y$ in $E$ with $|x - y| < \delta$ we have $|f(x) - f(y)| < \epsilon$.*

**20. Proposition:** *If a real-valued function $f$ is defined and continuous on a closed and bounded set $F$ of real numbers, it is uniformly continuous on $F$.*

***Proof:*** Given $\epsilon > 0$ and $x$ in $F$, there is a $\delta_x > 0$ such that $|x - y| < \delta_x$ implies $|f(x) - f(y)| < \frac{1}{2}\epsilon$. Let $O_x$ be the interval $(x - \frac{1}{2}\delta_x,\ x + \frac{1}{2}\delta_x)$. Then $\{O_x\colon x \in F\}$ is an open covering of $F$. By the Heine–Borel Theorem there is a finite subcollection $\{O_{x_1}, \ldots, O_{x_n}\}$ that covers $F$. Let $\delta = \frac{1}{2}\min(\delta_{x_1}, \ldots, \delta_{x_n})$. Then $\delta$ is positive. Given two points $y$ and $z$ in $F$ such that $|y - z| < \delta$, the point $y$ must belong to some $O_{x_i}$, and hence there is an $i$ such that $|y - x_i| < \frac{1}{2}\delta_{x_i}$. Consequently,

$$|z - x_i| \le |z - y| + |y - x_i| < \tfrac{1}{2}\delta_{x_i} + \delta \le \delta_{x_i}.$$

Hence

$$|f(y) - f(x_i)| < \frac{\epsilon}{2}$$

and

$$|f(z) - f(x_i)| < \frac{\epsilon}{2},$$

whence

$$|f(z) - f(y)| < \epsilon,$$

showing that $f$ is uniformly continuous on $F$. ▮

**Definition:** *A sequence $\langle f_n \rangle$ of functions defined on a set E is said to converge* ***pointwise*** *on E to a function f if for every x in E we have* $f(x) = \lim f_n(x)$*; that is, if, given $x \in E$ and $\epsilon > 0$, there is an N such that for all $n \geq N$, we have $|f(x) - f_n(x)| < \epsilon$.*

**Definition:** *A sequence $\langle f_n \rangle$ of functions defined on a set E is said to converge* ***uniformly*** *on E if given $\epsilon > 0$, there is an N such that for all $x \in E$ and all $n \geq N$, we have $|f(x) - f_n(x)| < \epsilon$.*

## Problems

**40.** Let $F$ be a closed set of real numbers and $f$ a real-valued function which is defined and continuous on $F$. Show that there is a function $g$ defined and continuous on $(-\infty, \infty)$ such that $f(x) = g(x)$ for each $x \in F$. [Hint: Take $g$ to be linear in each of the intervals of which $\tilde{F}$ is composed.]

**41.** Let $f$ be a real-valued function with domain $E$. Prove that $f$ is continuous if and only if for each open set $O$ there is an open set $U$ such that $f^{-1}[O] = E \cap U$.

**42.** Let $\langle f_n \rangle$ be a sequence of continuous functions defined on a set $E$. Prove that if $\langle f_n \rangle$ converges uniformly to $f$ on $E$, then $f$ is continuous on $E$.

**43.** Let $f$ be that function defined by setting

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

At what points is $f$ continuous?

**44. a.** Show that if $f$ and $g$ are continuous functions, then the functions $f + g$ and $fg$ are continuous.

**b.** Show that if $f$ and $g$ are continuous, then so is $f \circ g$.

**c.** Let $f \vee g$ be the function defined by $(f \vee g)(x) = f(x) \vee g(x)$, and define $f \wedge g$ similarly. Show that if $f$ and $g$ are continuous, so are $f \vee g$ and $f \wedge g$.

**d.** If $f$ is continuous, then so is $|f|$.

**45.** Prove Proposition 19.

**46.** A real-valued function $f$ is said to be *monotone increasing* if $f(x) \leq f(y)$ whenever $x \leq y$, and *strictly* monotone increasing if $f(x) < f(y)$

whenever $x < y$. It is called monotone (or strictly monotone) if $f$ or $-f$ is monotone (or strictly monotone) increasing. Let $f$ be a continuous function on the interval $[a, b]$. Then there is a continuous function $g$ such that $g(f(x)) = x$ for all $x \in [a, b]$ if and only if $f$ is strictly monotone. In this case we also have $f(g(y)) = y$ for each $y$ between $f(a)$ and $f(b)$. A function $f$ which has a continuous inverse is called a *homeomorphism* (between its domain and its range).

**47.** A continuous function $\varphi$ on $[a, b]$ is called *polygonal* (or *piecewise linear*) if there is a subdivision $a = x_0 < x_1 < \cdots < x_n = b$ such that $\varphi$ is linear on each interval $[x_i, x_{i+1}]$. Let $f$ be an arbitrary continuous function on $[a, b]$ and $\epsilon$ a positive number. Show that there is a polygonal function $\varphi$ on $[a, b]$ with $|f(x) - \varphi(x)| < \epsilon$ for all $x \in [a, b]$.

**48.** Let $x$ be a real number in $[0, 1]$ with the ternary expansion $\langle a_n \rangle$ (cf. Problem 22). Let $N = \infty$ if none of the $a_n$ are 1, and otherwise let $N$ be the smallest value of $n$ such that $a_n = 1$. Let $b_n = \frac{1}{2}a_n$ for $n < N$ and $b_N = 1$. Show that

$$\sum_{n=1}^{N} \frac{b_n}{2^n}$$

is independent of the ternary expansion of $x$ (if $x$ has two expansions) and that the function $f$ defined by setting

$$f(x) = \sum_{n=1}^{N} \frac{b_n}{2^n}$$

is a continuous, monotone function on the interval $[0, 1]$. Show that $f$ is constant on each interval contained in the complement of the Cantor ternary set (Problem 37), and that $f$ maps the Cantor ternary set *onto* the interval $[0, 1]$. (This function is called the *Cantor ternary function.*)

**49.** *Limit superior of a function of a real variable.* Let $f$ be a real (or extended real) valued function defined for all $x$ in an interval containing $y$. We define

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

$$\overline{\lim_{x \to y+}}\, f(x) = \inf_{\delta > 0} \sup_{0 < x-y < \delta} f(x)$$

with similar definitions for $\underline{\lim}$.

**a.** $\overline{\lim}_{x \to y} f(x) \le A$ if and only if, given $\epsilon > 0$, there is a $\delta > 0$ such that for all $x$ with $0 < |x - y| < \delta$ we have $f(x) \le A + \epsilon$.

**b.** $\overline{\lim}_{x \to y} f(x) \ge A$ if and only if, given $\epsilon > 0$ and $\delta > 0$, there is an $x$ such that $0 < |x - y| < \delta$ and $f(x) \ge A - \epsilon$.

**c.** $\underline{\lim}_{x \to y} f(x) \le \overline{\lim}_{x \to y} f(x)$ with equality (for $\overline{\lim} f \neq \pm\infty$) if and only if $\lim_{x \to y} f(x)$ exists.

**d.** If $\overline{\lim}_{x \to y} f(x) = A$ and $\langle x_n \rangle$ is a sequence with $x_n \neq y$ such that $y = \lim x_n$, then $\overline{\lim} f(x_n) \le A$.

**e.** If $\overline{\lim}_{x \to y} f(x) = A$, then there is a sequence $\langle x_n \rangle$ with $x_n \neq y$ such that $y = \lim x_n$ and $A = \lim f(x_n)$.

**f.** For a real number $l$ we have $l = \lim_{x \to y} f(x)$ if and only if $l = \lim f(x_n)$ for every sequence $\langle x_n \rangle$ with $x_n \neq y$ and $y = \lim x_n$.

**50.** *Semicontinuous functions.* An extended real-valued function $f$ is called *lower semicontinuous* at the point $y$ if $f(y) \neq -\infty$ and $f(y) \le \underline{\lim}_{x \to y} f(x)$. Similarly, $f$ is called *upper semicontinuous* at $y$ if $f(y) \neq +\infty$ and $f(y) \ge \overline{\lim}_{x \to y} f(x)$. We say that $f$ is lower (upper) semicontinuous on an interval if it is lower (upper) semicontinuous at each point of the interval. The function $f$ is upper semicontinuous if and only if the function $-f$ is lower semicontinuous.

**a.** Let $f(y)$ be finite. Prove that $f$ is lower semicontinuous at $y$ if and only if given $\epsilon > 0$, $\exists \delta > 0$ such that $f(y) \le f(x) + \epsilon$ for all $x$ with $|x - y| < \delta$.

**b.** A function $f$ is continuous (at a point or in an interval) if and only if it is both upper and lower semicontinuous (at the point or in the interval).

**c.** Show that a real-valued function $f$ is lower semicontinuous on $(a, b)$ if and only if the set $\{x : f(x) > \lambda\}$ is open for each real number $\lambda$.

**d.** Show that if $f$ and $g$ are lower semicontinuous functions, so are $f \vee g$ and $f + g$.

**e.** Let $\langle f_n \rangle$ be a sequence of lower semicontinuous functions. Show that the function $f$ defined by $f(x) = \sup_n f_n(x)$ is also lower semicontinuous.

**f.** A real-valued function $\varphi$ defined on an interval $[a, b]$ is called a **step function** if there is a partition $a = x_0 < x_1 < \cdots < x_n = b$ such that for each $i$ the function $\varphi$ assumes only one value in the interval $(x_i, x_{i+1})$. Show that a step function $\varphi$ is lower semicontinuous iff $\varphi(x_i)$ is less than or equal to the smaller of the two values assumed in $(x_{i-1}, x_i)$ and $(x_i, x_{i+1})$.

**g.** A function $f$ defined on an interval $[a, b]$ is lower semicontinuous if and only if there is a monotone increasing sequence $\langle \varphi_n \rangle$ of lower semicontinuous step functions on $[a, b]$ such that for each $x \in [a, b]$ we have $f(x) = \lim \varphi_n(x)$.

**h.** A function $f$ defined on $[a, b]$ is lower semicontinuous if and only if there is a monotone increasing sequence $\langle \psi_n \rangle$ of continuous functions such that $f(x) = \lim \psi_n(x)$ for each $x$ in $[a, b]$. [Hint: Modify the functions $\varphi_n$ in part (*g*) to make them continuous.]

**i.** Prove that a function $f$ that is defined and lower semicontinuous on a closed interval $[a, b]$ is bounded from below and assumes its minimum on $[a, b]$, that is, that there is a $y \in [a, b]$ such that $f(y) \le f(x)$ for all $x \in [a, b]$.

**51.** *Upper and lower envelopes of a function.* Let $f$ be a real-valued function defined on $[a, b]$. We define the *lower envelope* $g$ of $f$ to be the function $g$ defined by

$$g(y) = \sup_{\delta > 0} \inf_{|x-y|<\delta} f(x),$$

and the *upper envelope* $h$ by

$$h(y) = \inf_{\delta > 0} \sup_{|x-y|<\delta} f(x).$$

**a.** For each $x \in [a, b]$, $g(x) \le f(x) \le h(x)$, and $g(x) = f(x)$ if and only if $f$ is lower semicontinuous at $x$, while $g(x) = h(x)$ if and only if $f$ is continuous at $x$.

**b.** If $f$ is bounded, the function $g$ is lower semicontinuous, while $h$ is upper semicontinuous.

**c.** If $\varphi$ is any lower semicontinuous function such that $\varphi(x) \le f(x)$ for all $x \in [a, b]$, then $\varphi(x) \le g(x)$ for all $x \in [a, b]$.

## 7 Borel Sets

Although the intersection of any collection of closed sets is closed and the union of any *finite* collection of closed sets is closed, the union of a *countable* collection of closed sets need not be closed. For example, the set of rational numbers is the union of a countable collection of closed sets each of which contains exactly one number. Thus if we are interested in $\sigma$-algebras of sets that contain all of the closed sets, we must consider more general types of sets than the open and closed sets. This leads us to the following definition:

**Definition:** *The collection $\mathcal{B}$ of Borel sets is the smallest $\sigma$-algebra which contains all of the open sets.*

Such a smallest $\sigma$-algebra exists by Proposition 1.3. It is also the smallest $\sigma$-algebra that contains all closed sets and the smallest $\sigma$-algebra that contains the open intervals.

A set which is a countable union of closed sets is called an $F_\sigma$ ($F$ for closed, $\sigma$ for sum). Thus every countable set is an $F_\sigma$, as is, of course, every closed set. A countable union of sets in $F_\sigma$ is again in $F_\sigma$. Since

$$(a, b) = \bigcup_{n=1}^{\infty} \left[ a + \frac{1}{n}, b - \frac{1}{n} \right],$$

each open interval is an $F_\sigma$, and hence each open set is an $F_\sigma$.

We say that a set is a $G_\delta$ if it is the intersection of a countable collection of open sets ($G$ for open, $\delta$ for *durchschnitt*). Thus the complement of an $F_\sigma$ is a $G_\delta$, and conversely.

The $F_\sigma$ and $G_\delta$ are relatively simple types of Borel sets. We could also consider sets of type $F_{\sigma\delta}$, which are the intersections of countable collections of sets each of which is an $F_\sigma$. Similarly, we can construct the classes $G_{\delta\sigma}$, $F_{\sigma\delta\sigma}$, etc. Thus the classes in the two sequences

$$F_\sigma, F_{\sigma\delta}, F_{\sigma\delta\sigma}, \ldots, G_\delta, G_{\delta\sigma}, G_{\delta\sigma\delta}, \ldots$$

are all classes of Borel sets. However, not every Borel set belongs to one of these classes. Further theory of Borel sets can be found in Kuratowski [11], but we shall need only the properties that follow directly from the fact that they form the smallest $\sigma$-algebra containing the open and closed sets.

## Problems

**52.** Let $f$ be a lower semicontinuous function defined for all real numbers. What can you say about the sets $\{x: f(x) > \alpha\}$, $\{x: f(x) \geq \alpha\}$, $\{x: f(x) < \alpha\}$, $\{x: f(x) \leq \alpha\}$, and $\{x: f(x) = \alpha\}$?

**53.** Let $f$ be a real-valued function defined for all real numbers. Prove that the set of points at which $f$ is continuous is a $G_\delta$.

**54.** Let $\langle f_n \rangle$ be a sequence of continuous functions defined on $\mathbf{R}$. Show that the set $C$ of points where this sequence converges is an $F_{\sigma\delta}$.

# 3 Lebesgue Measure

## 1 Introduction

The length $l(I)$ of an interval $I$ is defined, as usual, to be the difference of the endpoints of the interval. Length is an example of a *set function*, that is, a function that associates an extended real number to each set in some collection of sets. In the case of length the domain is the collection of all intervals. We should like to extend the notion of length to more complicated sets than intervals. For instance, we could define the "length" of an open set to be the sum of the lengths of the open intervals of which it is composed. Since the class of open sets is still too restricted for our purposes, we would like to construct a set function $m$ that assigns to each set $E$ in some collection $\mathfrak{M}$ of sets of real numbers a nonnegative extended real number $mE$ called the measure of $E$. Ideally, we should like $m$ to have the following properties:

i. $mE$ is defined for each set $E$ of real numbers; that is, $\mathfrak{M} = \mathcal{P}(R)$.
ii. For an interval $I$, $mI = l(I)$.
iii. If $\langle E_n \rangle$ is a sequence of disjoint sets (for which $m$ is defined), $m(\bigcup E_n) = \sum mE_n$.
iv. $m$ is translation invariant; that is, if $E$ is a set for which $m$ is defined and if $E + y$ is the set $\{x + y: x \in E\}$, obtained by replacing each point $x$ in $E$ by the point $x + y$, then

$$m(E + y) = mE.$$

Unfortunately, as we shall see in Section 4, it is impossible to construct a set function having all four of these properties, and it is not known whether there is a set function satisfying the first three properties.[1] Consequently, one of these properties must be weakened, and it is most useful to retain the last three properties and to weaken the first condition so that $mE$ need not be defined for all sets $E$ of real numbers.[2] We shall want $mE$ to be defined for as many sets as possible and will find it convenient to require the family $\mathfrak{M}$ of sets for which $m$ is defined to be a $\sigma$-algebra. Thus we shall say that $m$ is a **countably additive measure** if it is a nonnegative extended real-valued function whose domain of definition is a $\sigma$-algebra $\mathfrak{M}$ of sets (of real numbers) and we have $m(\bigcup E_n) = \sum mE_n$ for each sequence $\langle E_n \rangle$ of disjoint sets in $\mathfrak{M}$. Our goal in the next two sections will be the construction of a countably additive measure which is translation invariant and has the property that $mI = l(I)$ for each interval $I$.

## Problems

Let $m$ be a countably additive measure defined for all sets in a $\sigma$-algebra $\mathfrak{M}$.

**1.** If $A$ and $B$ are two sets in $\mathfrak{M}$ with $A \subset B$, then $mA \leq mB$. This property is called *monotonicity*.

**2.** Let $\langle E_n \rangle$ be any sequence of sets in $\mathfrak{M}$. Then $m(\bigcup E_n) \leq \sum mE_n$. [Hint: Use Proposition 1.2.] This property of a measure is called *countable subadditivity*.

**3.** If there is a set $A$ in $\mathfrak{M}$ such that $mA < \infty$, then $m\emptyset = 0$.

**4.** Let $nE$ be $\infty$ for an infinite set $E$ and be equal to the number of elements in $E$ for a finite set. Show that $n$ is a countably additive set function that is translation invariant and defined for all sets of real numbers. This measure is called the **counting measure.**

[1] If we assume the continuum hypothesis (that every noncountable set of real numbers can be put in one-to-one correspondence with the set of all real numbers), then such a measure is impossible.

[2] Weakening property (i) is not the only approach; it is also possible to replace property (iii) of countable additivity by the weaker property of finite additivity: For each finite sequence $\langle E_n \rangle$ of disjoint sets we have $m(\bigcup E_n) = \sum mE_n$ (see Problem 10.21). Another possible alternative to property (iii) is countable subadditivity, which is satisfied by the outer measure we construct in the next section (see Problem 2).

## 2 Outer Measure

For each set $A$ of real numbers consider the countable collections $\{I_n\}$ of open intervals that cover $A$, that is, collections for which $A \subset \bigcup I_n$, and for each such collection consider the sum of the length of the intervals in the collection. Since the lengths are positive numbers, this sum is uniquely defined independently of the order of the terms. We define the **outer measure**[3] $m^*A$ of $A$ to be the infimum of all such sums. In an abbreviated notation

$$m^*A = \inf_{A \subset \bigcup I_n} \sum l(I_n).$$

It follows immediately from the definition of $m^*$ that $m^*\emptyset = 0$ and that if $A \subset B$, then $m^*A \leq m^*B$. Also, each set consisting of a single point has outer measure zero. We establish two propositions concerning outer measure:

**1. Proposition:** *The outer measure of an interval is its length.*

***Proof:*** We begin with the case in which we have a closed finite interval, say $[a, b]$. Since the open interval $(a - \epsilon, b + \epsilon)$ contains $[a, b]$ for each positive $\epsilon$, we have $m^*[a, b] \leq l(a - \epsilon, b + \epsilon) = b - a + 2\epsilon$. Since $m^*[a, b] \leq b - a + 2\epsilon$ for each positive $\epsilon$, we must have $m^*[a, b] \leq b - a$. Thus we have only to show that $m^*[a, b] \geq b - a$. But this is equivalent to showing that if $\{I_n\}$ is any countable collection of open intervals covering $[a, b]$, then

$$\sum l(I_n) \geq b - a. \tag{1}$$

By the Heine–Borel Theorem, any collection of open intervals covering $[a, b]$ contains a finite subcollection that also covers $[a, b]$, and since the sum of the lengths of the finite subcollection is no greater than the sum of the lengths of the original collection, it suffices to prove the inequality (1) for finite collections $\{I_n\}$ that cover $[a, b]$. Since $a$ is contained in $\bigcup I_n$, there must be one of the $I_n$'s that contains $a$. Let this be the interval $(a_1, b_1)$. We have $a_1 < a < b_1$. If $b_1 \leq b$, then $b_1 \in [a, b]$, and since $b_1 \notin (a_1, b_1)$, there must be an interval $(a_2, b_2)$ in the collection $\{I_n\}$ such that $b_1 \in (a_2, b_2)$; that is,

[3] In order to distinguish this outer measure from the more general outer measures to be considered in Chapter 12, we call this outer measure *Lebesgue* outer measure, after Henri Lebesgue. Since we consider no other outer measure in this chapter, we refer to $m^*$ simply as outer measure.

$a_2 < b_1 < b_2$. Continuing in this fashion, we obtain a sequence $(a_1, b_1), \ldots, (a_k, b_k)$ from the collection $\{I_n\}$ such that $a_i < b_{i-1} < b_i$.

Since $\{I_n\}$ is a finite collection, our process must terminate with some interval $(a_k, b_k)$. But it terminates only if $b \in (a_k, b_k)$, that is, if $a_k < b < b_k$. Thus

$$\begin{aligned}\sum l(I_n) &\geq \sum l(a_i, b_i)\\ &= (b_k - a_k) + (b_{k-1} - a_{k-1}) + \cdots + (b_1 - a_1)\\ &= b_k - (a_k - b_{k-1}) - (a_{k-1} - b_{k-2})\\ &\quad - \cdots - (a_2 - b_1) - a_1 > b_k - a_1,\end{aligned}$$

since $a_i < b_{i-1}$. But $b_k > b$ and $a_1 < a$, and so we have $b_k - a_1 > b - a$, whence $\sum l(I_n) > (b - a)$. This shows that $m^*[a, b] = b - a$.

If $I$ is any finite interval, then given $\epsilon > 0$, there is a closed interval $J \subset I$ such that $l(J) > l(I) - \epsilon$. Hence

$$l(I) - \epsilon < l(J) = m^*J \leq m^*I \leq m^*\bar{I} = l(\bar{I}) = l(I).$$

Thus for each $\epsilon > 0$,

$$l(I) - \epsilon < m^*I \leq l(I),$$

and so $m^*I = l(I)$.

If $I$ is an infinite interval, then given any real number $\Delta$, there is a closed interval $J \subset I$ with $l(J) = \Delta$. Hence $m^*I \geq m^*J = l(J) = \Delta$. Since $m^*I \geq \Delta$ for each $\Delta$, $m^*I = \infty = l(I)$. ∎

**2. Proposition:** *Let $\{A_n\}$ be a countable collection of sets of real numbers. Then*

$$m^*(\bigcup A_n) \leq \sum m^*A_n.$$

***Proof:*** If one of the sets $A_n$ has infinite outer measure, the inequality holds trivially. If $m^*A_n$ is finite, then, given $\epsilon > 0$, there is a countable collection $\{I_{n,i}\}_i$ of open intervals such that $A_n \subset \bigcup_i I_{n,i}$ and $\sum_i l(I_{n,i}) < m^*A_n + 2^{-n}\epsilon$. Now the collection $\{I_{n,i}\}_{n,i} = \bigcup_n \{I_{n,i}\}_i$ is countable, being the union of a countable number of countable collections, and covers $\bigcup A_n$. Thus

$$\begin{aligned}m^*(\bigcup A_n) &\leq \sum_{n,i} l(I_{n,i}) = \sum_n \sum_i l(I_{n,i}) < \sum_n (m^*A_n + \epsilon 2^{-n})\\ &= \sum m^*A_n + \epsilon.\end{aligned}$$

Since $\epsilon$ was an arbitrary positive number,

$$m^*(\bigcup A_n) \leq \sum m^* A_n. \quad \blacksquare$$

**3. Corollary:** *If $A$ is countable, $m^*A = 0$.*

**4. Corollary:** *The set $[0, 1]$ is not countable.*

**5. Proposition:** *Given any set $A$ and any $\epsilon > 0$, there is an open set $O$ such that $A \subset O$ and $m^*O \leq m^*A + \epsilon$. There is a $G \in G_\delta$ such that $A \subset G$* and *$m^*A = m^*G$.*

## Problems

5. Let $A$ be the set of rational numbers between 0 and 1, and let $\{I_n\}$ be a *finite* collection of open intervals covering $A$. Then $\sum l(I_n) \geq 1$.
6. Prove Proposition 5.
7. Prove that $m^*$ is translation invariant.
8. Prove that if $m^*A = 0$, then $m^*(A \cup B) = m^*B$.

## 3 Measurable Sets and Lebesgue Measure

While outer measure has the advantage that it is defined for all sets, it is not countably additive. It becomes countably additive, however, if we suitably reduce the family of sets on which it is defined. Perhaps the best way of doing this is to use the following definition due to Carathéodory:

**Definition:** *A set $E$ is said to be **measurable**[4] if for each set $A$ we have $m^*A = m^*(A \cap E) + m^*(A \cap \tilde{E})$.*

Since we always have $m^*A \leq m^*(A \cap E) + m^*(A \cap \tilde{E})$, we see that $E$ is measurable if (and only if) for each $A$ we have $m^*A \geq m^*(A \cap E) + m^*(A \cap \tilde{E})$. Since the definition of measurability is symmetric in $E$ and $\tilde{E}$, we have $\tilde{E}$ measurable whenever $E$ is. Clearly $\emptyset$ and the set $\mathbf{R}$ of all real numbers are measurable.

**6. Lemma:** *If $m^*E = 0$, then $E$ is measurable.*

[4] In the present case $m^*$ is Lebesgue outer measure, and we say that $E$ is *Lebesgue* measurable. More general notions of measurable set are considered in Chapters 11 and 12.

***Proof:*** Let $A$ be any set. Then $A \cap E \subset E$, and so $m^*(A \cap E) \leq m^*E = 0$. Also $A \supset A \cap \tilde{E}$, and so

$$m^*A \geq m^*(A \cap \tilde{E}) = m^*(A \cap \tilde{E}) + m^*(A \cap E),$$

and therefore $E$ is measurable. ▮

7. **Lemma:** *If $E_1$ and $E_2$ are measurable, so is $E_1 \cup E_2$.*

***Proof:*** Let $A$ be any set. Since $E_2$ is measurable, we have

$$m^*(A \cap \tilde{E}_1) = m^*(A \cap \tilde{E}_1 \cap E_2) + m^*(A \cap \tilde{E}_1 \cap \tilde{E}_2),$$

and since $A \cap (E_1 \cup E_2) = [A \cap E_1] \cup [A \cap E_2 \cap \tilde{E}_1]$, we have

$$m^*(A \cap [E_1 \cup E_2]) \leq m^*(A \cap E_1) + m^*(A \cap E_2 \cap \tilde{E}_1).$$

Thus

$$\begin{aligned} m^*(A \cap [E_1 \cup E_2]) + m^*(A \cap \tilde{E}_1 \cap \tilde{E}_2) &\leq m^*(A \cap E_1) \\ &\quad + m^*(A \cap E_2 \cap \tilde{E}_1) + m^*(A \cap \tilde{E}_1 \cap \tilde{E}_2) \\ &= m^*(A \cap E_1) + m^*(A \cap \tilde{E}_1) = m^*A, \end{aligned}$$

by the measurability of $E_1$. Since $\sim(E_1 \cup E_2) = \tilde{E}_1 \cap \tilde{E}_2$, this shows that $E_1 \cup E_2$ is measurable. ▮

8. **Corollary:** *The family $\mathfrak{M}$ of measurable sets is an algebra of sets.*

9. **Lemma:** *Let $A$ be any set, and $E_1, \ldots, E_n$ a finite sequence of disjoint measurable sets. Then*

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

***Proof:*** We prove the lemma by induction on $n$. It is clearly true for $n = 1$, and we assume it is true if we have $n - 1$ sets $E_i$. Since the $E_i$ are disjoint sets, we have

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

and

$$A \cap \left[\bigcup_{i=1}^{n} E_i\right] \cap \tilde{E}_n = A \cap \left[\bigcup_{i=1}^{n-1} E_i\right].$$

Hence the measurability of $E_n$ implies that

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

by our assumption of the lemma for $n - 1$ sets. ▮

**10. Theorem:** *The collection $\mathfrak{M}$ of measurable sets is a $\sigma$-algebra; that is, the complement of a measurable set is measurable and the union (and intersection) of a countable collection of measurable sets is measurable. Moreover, every set with outer measure zero is measurable.*

***Proof:*** We have already observed that $\mathfrak{M}$ is an algebra of sets, and so we have only to prove that if a set $E$ is the union of a countable collection of measurable sets it is measurable. By Proposition 1.2 such an $E$ must be the union of a sequence $\langle E_n \rangle$ of pairwise disjoint measurable sets. Let $A$ be any set, and let $F_n = \bigcup_{i=1}^{n} E_i$. Then $F_n$ is measurable, and $\tilde{F}_n \supset \tilde{E}$. Hence

$$m^*A = m^*(A \cap F_n) + m^*(A \cap \tilde{F}_n) \geq m^*(A \cap F_n) + m^*(A \cap \tilde{E}).$$

By Lemma 9

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

Thus

$$m^*A \geq \sum_{i=1}^{n} m^*(A \cap E_i) + m^*(A \cap \tilde{E}).$$

Since the left side of this inequality is independent of $n$, we have

$$m^*A \geq \sum_{i=1}^{\infty} m^*(A \cap E_i) + m^*(A \cap \tilde{E})$$
$$\geq m^*(A \cap E) + m^*(A \cap \tilde{E})$$

by the countable subadditivity of $m^*$. ▮

**11. Lemma:** *The interval $(a, \infty)$ is measurable.*

***Proof:*** Let $A$ be any set, $A_1 = A \cap (a, \infty)$, $A_2 = A \cap (-\infty, a]$. Then we must show $m^*A_1 + m^*A_2 \leq m^*A$. If $m^*A = \infty$, then there is nothing to prove. If $m^*A < \infty$, then, given $\epsilon > 0$, there is a countable collection $\{I_n\}$ of open intervals which cover $A$ and for which

$$\sum l(I_n) \leq m^*A + \epsilon.$$

Let $I_n' = I_n \cap (a, \infty)$ and $I_n'' = I_n \cap (-\infty, a]$. Then $I_n'$ and $I_n''$ are intervals (or empty) and

$$l(I_n) = l(I_n') + l(I_n'') = m^*I_n' + m^*I_n''.$$

Since $A_1 \subset \bigcup I_n'$, we have

$$m^*A_1 \leq m^*(\bigcup I_n') \leq \sum m^*I_n',$$

and, since $A_2 \subset \bigcup I_n''$, we have

$$m^*A_2 \leq m^*(\bigcup I_n'') \leq \sum m^*I_n''.$$

Thus

$$\begin{aligned} m^*A_1 + m^*A_2 &\leq \sum (m^*I_n' + m^*I_n'') \\ &\leq \sum l(I_n) \leq m^*A + \epsilon. \end{aligned}$$

But $\epsilon$ was an arbitrary positive number, and so we must have $m^*A_1 + m^*A_2 \leq m^*A$. ▮

**12. Theorem:** *Every Borel set is measurable. In particular each open set and each closed set is measurable.*

***Proof:*** Since the collection $\mathfrak{M}$ of measurable sets is a $\sigma$-algebra, we have $(-\infty, a]$ measurable for each $a$, since $(-\infty, a] = \sim(a, \infty)$. Since

$$(-\infty, b) = \bigcup_{n=1}^{\infty} \left(-\infty, b - \frac{1}{n}\right],$$

we have $(-\infty, b)$ measurable. Hence each open interval $(a, b) = (-\infty, b) \cap (a, \infty)$ is measurable. But each open set is the union of a countable number of open intervals and so must be measurable. Thus $\mathfrak{M}$ is a $\sigma$-algebra containing the open sets and must therefore contain the family $\mathfrak{B}$ of Borel sets, since $\mathfrak{B}$ is the smallest $\sigma$-algebra containing the open sets. [*Note:* The theorem also follows immediately from the fact that $\mathfrak{M}$ is a $\sigma$-algebra containing each interval of the form $(a, \infty)$ and the fact that $\mathfrak{B}$ is the smallest $\sigma$-algebra containing all such intervals.] ▮

If $E$ is a measurable set, we define the *Lebesgue measure* $mE$ to be the outer measure of $E$. Thus $m$ is the set function obtained by restricting the set function $m^*$ to the family $\mathfrak{M}$ of measurable sets. Two important properties of Lebesgue measure are summarized by the following propositions:

**13. Proposition:** *Let $\langle E_i \rangle$ be a sequence of measurable sets. Then*

$$m(\bigcup E_i) \le \sum mE_i.$$

*If the sets $E_n$ are pairwise disjoint, then*

$$m(\bigcup E_i) = \sum mE_i.$$

***Proof:*** The inequality is simply a restatement of the subadditivity of $m^*$ given by Proposition 2. If $\langle E_i \rangle$ is a finite sequence of disjoint measurable sets, then Lemma 9 with $A = \mathbf{R}$ implies that

$$m(\bigcup E_i) = \sum mE_i,$$

and so $m$ is finitely additive. Let $\langle E_i \rangle$ be an infinite sequence of pairwise disjoint measurable sets. Then

$$\bigcup_{i=1}^{\infty} E_i \supset \bigcup_{i=1}^{n} E_i,$$

and so

$$m\left(\bigcup_{i=1}^{\infty} E_i\right) \ge m\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} mE_i.$$

Since the left side of this inequality is independent of $n$, we have

$$m\left(\bigcup_{i=1}^{\infty} E_i\right) \ge \sum_{i=1}^{\infty} mE_i.$$

The reverse inequality follows from countable subadditivity, and we have

$$m\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} mE_i. \quad \blacksquare$$

**14. Proposition:** *Let $\langle E_n \rangle$ be an infinite decreasing sequence of measurable sets, that is, a sequence with $E_{n+1} \subset E_n$ for each $n$. Let $mE_1$ be finite. Then*

$$m\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{n \to \infty} mE_n.$$

***Proof:*** Let $E = \bigcap_{i=1}^{\infty} E_i$, and let $F_i = E_i \sim E_{i+1}$. Then

$$E_1 \sim E = \bigcup_{i=1}^{\infty} F_i,$$

and the sets $F_i$ are pairwise disjoint. Hence

$$m(E_1 \sim E) = \sum_{i=1}^{\infty} mF_i = \sum_{i=1}^{\infty} m(E_i \sim E_{i+1}).$$

But $mE_1 = mE + m(E_1 \sim E)$, and $mE_i = mE_{i+1} + m(E_i \sim E_{i+1})$, since $E \subset E_1$ and $E_{i+1} \subset E_i$. Since $mE_i \le mE_1 < \infty$, we have $m(E_1 \sim E) = mE_1 - mE$ and $m(E_i \sim E_{i+1}) = mE_i - mE_{i+1}$. Thus

$$\begin{aligned} mE_1 - mE &= \sum_{i=1}^{\infty} (mE_i - mE_{i+1}) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} (mE_i - mE_{i+1}) \\ &= \lim_{n\to\infty} (mE_1 - mE_n) \\ &= mE_1 - \lim_{n\to\infty} mE_n. \end{aligned}$$

Since $mE_1 < \infty$, we have

$$mE = \lim_{n\to\infty} mE_n. \quad \blacksquare$$

The following proposition expresses a number of ways in which a measurable set is very nearly a nice set. The proof is left to the reader (Problem 13).

**15. Proposition:** *Let $E$ be a given set. Then the following five statements are equivalent:*

i. *$E$ is measurable.*
ii. *Given $\epsilon > 0$, there is an open set $O \supset E$ with $m^*(O \sim E) < \epsilon$.*
iii. *Given $\epsilon > 0$, there is a closed set $F \subset E$ with $m^*(E \sim F) < \epsilon$.*
iv. *There is a $G$ in $G_\delta$ with $E \subset G$, $m^*(G \sim E) = 0$.*
v. *There is an $F$ in $F_\sigma$ with $F \subset E$, $m^*(E \sim F) = 0$.*

*If $m^*E$ is finite, the above statements are equivalent to:*

vi. *Given $\epsilon > 0$, there is a finite union $U$ of open intervals such that $m^*(U \,\Delta\, E) < \epsilon$.*

## Problems

**9.** Show that if $E$ is a measurable set, then each translate $E + y$ of $E$ is also measurable.

**10.** Show that if $E_1$ and $E_2$ are measurable, then

$$m(E_1 \cup E_2) + m(E_1 \cap E_2) = mE_1 + mE_2.$$

**11.** Show that the condition $mE_1 < \infty$ is necessary in Proposition 14 by giving a decreasing sequence $\langle E_n \rangle$ of measurable sets with $\emptyset = \bigcap E_n$ and $mE_n = \infty$ for each $n$.

**12.** Let $\langle E_i \rangle$ be a sequence of disjoint measurable sets and $A$ any set. Then $m^*\left(A \cap \bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} m^*(A \cap E_i)$.

**13.** Prove Proposition 15. [Hints:

**a.** Show that for $m^*E < \infty$, (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (vi) (cf. Proposition 5).

**b.** Use (a) to show that for arbitrary sets $E$, (i) $\Rightarrow$ (ii) $\Rightarrow$ (iv) $\Rightarrow$ (i).

**c.** Use (b) to show that (i) $\Rightarrow$ (iii) $\Rightarrow$ (v) $\Rightarrow$ (i).]

**14. a.** Show that the Cantor ternary set (Problem 2.37) has measure zero.

**b.** Let $F$ be a subset of $[0, 1]$ constructed in the same manner as the Cantor ternary set except that each of the intervals removed at the $n$th step has length $\alpha 3^{-n}$ with $0 < \alpha < 1$. Then $F$ is a closed set, $\tilde{F}$ dense in $[0, 1]$ and $mF = 1 - \alpha$. Such a set $F$ is called a *generalized Cantor set.*

## ★4 A Nonmeasurable Set

We are going to show the existence of a nonmeasurable set. If $x$ and $y$ are real numbers in $[0, 1)$, we define the *sum modulo* 1 of $x$ and $y$ to be $x + y$, if $x + y < 1$, and to be $x + y - 1$ if $x + y \geq 1$. Let us denote the sum modulo 1 of $x$ and $y$ by $x \dot{+} y$. Then $\dot{+}$ is a commutative and associative operation taking pairs of numbers in $[0, 1)$ into numbers in $[0, 1)$. If we assign to each $x \in [0, 1)$ the angle $2\pi x$, then addition modulo 1 corresponds to the addition of angles. If $E$ is a subset of $[0, 1)$, we define the translate modulo 1 of $E$ to be the set $E \dot{+} y = \{z : z = x \dot{+} y \text{ for some } x \in E\}$. If we consider addition modulo 1 as addition of angles, translation modulo 1 by $y$ corresponds to rotation through an angle of $2\pi y$. The following lemma shows that Lebesgue measure is invariant under translation modulo 1.

**16. Lemma:** *Let* $E \subset [0, 1)$ *be a measurable set. Then for each* $y \in [0, 1)$ *the set* $E \dot{+} y$ *is measurable and* $m(E \dot{+} y) = mE$.

***Proof:*** Let $E_1 = E \cap [0, 1 - y)$ and $E_2 = E \cap [1 - y, 1)$. Then $E_1$ and $E_2$ are disjoint measurable sets whose union is $E$, and so

$$mE = mE_1 + mE_2.$$

Now $E_1 \dot{+} y = E_1 + y$, and so $E_1 \dot{+} y$ is measurable and we have $m(E_1 \dot{+} y) = mE_1$, since $m$ is translation invariant. Also $E_2 \dot{+} y = E_2 + (y - 1)$, and so $E_2 \dot{+} y$ is measurable and $m(E_2 \dot{+} y) = mE_2$. But $E \dot{+} y = (E_1 \dot{+} y) \cup (E_2 \dot{+} y)$ and the sets $(E_1 \dot{+} y)$ and $(E_2 \dot{+} y)$ are disjoint measurable sets. Hence $E \dot{+} y$ is measurable and

$$\begin{aligned} m(E \dot{+} y) &= m(E_1 \dot{+} y) + m(E_2 \dot{+} y) \\ &= mE_1 + mE_2 \\ &= mE. \quad \blacksquare \end{aligned}$$

We are now in a position to define a nonmeasurable set. If $x - y$ is a rational number, we say that $x$ and $y$ are equivalent and write $x \sim y$. This is an equivalence relation and hence partitions $[0, 1)$ into equivalence classes, that is, classes such that any two elements of one class differ by a rational number, while any two elements of different classes differ by an irrational number. By the axiom of choice there is a set $P$ which contains exactly one element from each equivalence class. Let $\langle r_i \rangle_{i=0}^{\infty}$ be an enumeration of the rational numbers in $[0, 1)$ with $r_0 = 0$, and define $P_i = P \dot{+} r_i$. Then $P_0 = P$. Let $x \in P_i \cap P_j$. Then $x = p_i \dot{+} r_i = p_j \dot{+} r_j$ with $p_i$ and $p_j$ belonging to $P$. But $p_i - p_j = r_j - r_i$ is a rational number, whence $p_i \sim p_j$. Since $P$ has only one element from each equivalence class, we must have $i = j$. This implies that if $i \neq j$, $P_i \cap P_j = \emptyset$, that is, that $\langle P_i \rangle$ is a pairwise disjoint sequence of sets. On the other hand, each real number $x$ in $[0, 1)$ is in some equivalence class and so is equivalent to an element in $P$. But if $x$ differs from an element in $P$ by the rational number $r_i$, then $x \in P_i$. Thus $\bigcup P_i = [0, 1)$. Since each $P_i$ is a translation modulo 1 of $P$, each $P_i$ will be measurable if $P$ is and will have the same measure. But if this were the case,

$$m[0, 1) = \sum_{i=1}^{\infty} mP_i = \sum_{i=1}^{\infty} mP,$$

and the right side is either zero or infinite, depending on whether $mP$ is zero or positive. But this is impossible since $m[0, 1) = 1$, and consequently $P$ cannot be measurable.

While the above proof that $P$ is not measurable is a proof by contradiction, it should be noted that (until the last sentence) we have made no use of properties of Lebesgue measure other than translation invariance and countable additivity. Hence the foregoing argument gives a direct proof of the following theorem:

**17. Theorem:** *If $m$ is a countably additive, translation invariant measure defined on a $\sigma$-algebra containing the set $P$, then $m[0, 1)$ is either zero or infinite.*

The nonmeasurability of $P$ with respect to any translation invariant countability additive measure $m$ for which $m[0, 1)$ is 1 follows by contraposition.

## Problems

**15.** Show that if $E$ is measurable and $E \subset P$, then $mE = 0$. [Hint: Let $E_i = E \dot{+} r_i$. Then $\langle E_i \rangle$ is a disjoint sequence of measurable sets and $mE_i = mE$. Thus $\sum mE_i = m \bigcup E_i \leq m[0, 1]$.]

**16.** Show that, if $A$ is any set with $m^*A > 0$, then there is a nonmeasurable set $E \subset A$. [Hint: If $A \subset (0, 1)$, let $E_i = A \cap P_i$. The measurability of $E_i$ implies $mE_i = 0$, while $\sum m^*E_i \geq m^*A > 0$.]

**17. a.** Give an example where $\langle E_i \rangle$ is a disjoint sequence of sets and $m^*(\bigcup E_i) < \sum m^*E_i$.

**b.** Give an example of a sequence of sets $\langle E_i \rangle$ with $E_i \supset E_{i+1}$, $m^*E_i < \infty$, and $m^*(\bigcap E_i) < \lim m^*E_i$.

## 5 Measurable Functions

Since not all sets are measurable, it is of great importance to know that sets which arise naturally in certain constructions are measurable. If we start with a function $f$, the most important sets that arise from it are those listed in the following proposition:

**18. Proposition:** *Let $f$ be an extended real-valued function whose domain is measurable. Then the following statements are equivalent:*

i. *For each real number $\alpha$ the set $\{x : f(x) > \alpha\}$ is measurable.*
ii. *For each real number $\alpha$ the set $\{x : f(x) \geq \alpha\}$ is measurable.*
iii. *For each real number $\alpha$ the set $\{x : f(x) < \alpha\}$ is measurable.*
iv. *For each real number $\alpha$ the set $\{x : f(x) \leq \alpha\}$ is measurable.*

*These statements imply*

v. *For each extended real number $\alpha$ the set $\{x : f(x) = \alpha\}$ is measurable.*

***Proof:*** Let the domain of $f$ be $D$. We have (i) $\Rightarrow$ (iv), since $\{x : f(x) \leq \alpha\} = D \sim \{x : f(x) > \alpha\}$ and the difference of two measurable sets is measurable. Similarly, (iv) $\Rightarrow$ (i) and (ii) $\Leftrightarrow$ (iii). Now (i) $\Rightarrow$ (ii), since $\{x : f(x) \geq \alpha\} = \bigcap_{n=1}^{\infty} \{x : f(x) > \alpha - 1/n\}$, and the intersection of a sequence of measurable sets is measurable. Similarly, (ii) $\Rightarrow$ (i), since $\{x : f(x) > \alpha\} = \bigcup_{n=1}^{\infty} \{x : f(x) \geq \alpha + 1/n\}$, and the union of a sequence of measurable sets is measurable. This shows that the first four statements are equivalent. If $\alpha$ is a real number, $\{x : f(x) = \alpha\} = \{x : f(x) \geq \alpha\} \cap \{x : f(x) \leq \alpha\}$, and so (ii) and (iv) $\Rightarrow$ (v) for $\alpha$ real. Since

$$\{x : f(x) = \infty\} = \bigcap_{n=1}^{\infty} \{x : f(x) \geq n\},$$

(ii) $\Rightarrow$ (v) for $\alpha = \infty$. Similarly, (iv) $\Rightarrow$ (v) for $\alpha = -\infty$, and we have (ii) & (iv) $\Rightarrow$ (v). ▮

**Definition:** *An extended real-valued function $f$ is said to be (Lebesgue) measurable if its domain is measurable and if it satisfies one of the first four statements of Proposition* 18.

Thus if we restrict ourselves to measurable functions, the most important sets connected with them are measurable. It should be noted that a continuous function (with a measurable domain) is measurable, and of course each step function is measurable. If $f$ is a measurable function and $E$ is a measurable subset of the domain of $f$, then the function obtained by restricting $f$ to $E$ is also measurable. The following proposition tells us that certain operations performed on measurable functions lead again to measurable functions:

**19. Proposition:** *Let $c$ be a constant and $f$ and $g$ two measurable real-valued functions defined on the same domain. Then the functions $f + c$, $cf$, $f + g$, $g - f$, and $fg$ are also measurable.*

***Proof:*** We shall use condition (iii) of Proposition 18. Then

$$\{x : f(x) + c < \alpha\} = \{x : f(x) < \alpha - c\},$$

and so $f + c$ is measurable when $f$ is. A similar argument shows $cf$ to be measurable.

If $f(x) + g(x) < \alpha$, then $f(x) < \alpha - g(x)$ and by the corollary to the axiom of Archimedes there is a rational number $r$ such that

$$f(x) < r < \alpha - g(x).$$

Hence

$$\{x : f(x) + g(x) < \alpha\} = \bigcup_r (\{x : f(x) < r\} \cap \{x : g(x) < \alpha - r\}).$$

Since the rationals are countable, this set is measurable and so $f + g$ is measurable. Since $-g = (-1)g$ is measurable when $g$ is, we have $f - g$ measurable.

The function $f^2$ is measurable, since

$$\{x : f^2(x) > \alpha\} = \{x : f(x) > \sqrt{\alpha}\} \cup \{x : f(x) < -\sqrt{\alpha}\}$$

for $\alpha \geq 0$ and

$$\{x : f^2(x) > \alpha\} = D$$

if $\alpha < 0$, where $D$ is the domain of $f$. Thus

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

is measurable. ▌

We often want to use Proposition 19 for extended real-valued functions $f$ and $g$. Unfortunately, $f + g$ is not defined at points where it is of the form $\infty - \infty$. However, $fg$ is always measurable and $f + g$ is measurable if we always take the same value for $f + g$ at points where it is undefined. Also, $f + g$ is measurable no matter what values we take at the points where it is not defined, provided that these points are a set of measure zero (see Problem 22).

**20. Theorem:** *Let $\langle f_n \rangle$ be a sequence of measurable functions (with the same domain of definition). Then the functions* $\sup\{f_1, \ldots, f_n\}$, $\inf\{f_1, \ldots, f_n\}$, $\sup_n f_n$, $\inf_n f_n$, $\overline{\lim} f_n$, *and* $\underline{\lim} f_n$ *are all measurable.*

***Proof:*** If $h$ is defined by $h(x) = \sup\{f_1(x), \ldots, f_n(x)\}$, then $\{x : h(x) > \alpha\} = \bigcup_{i=1}^{n} \{x : f_i(x) > \alpha\}$. Hence the measurability of the $f_i$ implies that of $h$. Similarly, if $g$ is defined by $g(x) = \sup f_n(x)$, then

$\{x : g(x) > \alpha\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) > \alpha\}$, and so $g$ is measurable. A similar argument establishes the corresponding statements for inf. Since $\overline{\lim} f_n = \inf_n \sup_{k \geq n} f_k$, we have $\overline{\lim}\, f_n$ measurable, and similarly for $\underline{\lim}\, f_n$. ∎

A property is said to hold **almost everywhere**[5] (abbreviated a.e.) if the set of points where it fails to hold is a set of measure zero. Thus in particular we say that $f = g$ a.e. if $f$ and $g$ have the same domain and $m\{x : f(x) \neq g(x)\} = 0$. Similarly, we say that $f_n$ converges to $g$ almost everywhere if there is a set $E$ of measure zero such that $f_n(x)$ converges to $g(x)$ for each $x$ not in $E$. One consequence of equality a.e. is the following:

**21. Proposition:** *If $f$ is a measurable function and $f = g$ a.e., then $g$ is measurable.*

***Proof:*** Let $E$ be the set $\{x : f(x) \neq g(x)\}$. By hypothesis $mE = 0$. Now

$$\{x : g(x) > \alpha\} = [\{x : f(x) > \alpha\} \cup \{x \in E : g(x) > \alpha\}] \sim \{x \in E : g(x) \leq \alpha\}.$$

The first set on the right is measurable, since $f$ is a measurable function. The last two sets on the right are measurable, since they are subsets of $E$ and $mE = 0$. Thus $\{x : g(x) > \alpha\}$ is measurable for each $\alpha$, and so $g$ is measurable. ∎

The following proposition tells us that a measurable function is "almost" a continuous function. The proof is left to the reader (cf. Problem 23).

**22. Proposition:** *Let $f$ be a measurable function defined on an interval $[a, b]$, and assume that $f$ takes the values $\pm\infty$ only on a set of measure zero. Then given $\epsilon > 0$, we can find a step function $g$ and a continuous function $h$ such that*

$$|f - g| < \epsilon \quad \textit{and} \quad |f - h| < \epsilon$$

[5] French: *presque partout* (p.p).

*except on a set of measure less than* $\epsilon$; *i.e.*, $m\{x\colon |f(x) - g(x)| \geq \epsilon\} < \epsilon$ *and* $m\{x\colon |f(x) - h(x)| \geq \epsilon\} < \epsilon$. *If in addition* $m \leq f \leq M$, *then we may choose the functions* $g$ *and* $h$ *so that* $m \leq g \leq M$ *and* $m \leq h \leq M$.

If $A$ is any set, we define the **characteristic function** $\chi_A$ of the set $A$ to be the function given by

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

The function $\chi_A$ is measurable if and only if $A$ is measurable. Thus the existence of a nonmeasurable set implies the existence of a nonmeasurable function.

A real-valued function $\varphi$ is called **simple** if it is measurable and assumes only a finite number of values. If $\varphi$ is simple and has the values $\alpha_1, \ldots, \alpha_n$ then $\varphi = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$, where $A_i = \{x\colon \varphi(x) = \alpha_i\}$. The sum, product, and difference of two simple functions are simple.

## Problems

**18.** Show that (v) does not imply (iv) in Proposition 18 by constructing a function $f$ such that $\{x\colon f(x) > 0\} = E$, a given nonmeasurable set, and such that $f$ assumes each value at most once.

**19.** Let $D$ be a dense set of real numbers, that is, a set of real numbers such that every interval contains an element of $D$. Let $f$ be an extended real-valued function on $\mathbf{R}$ such that $\{x\colon f(x) > \alpha\}$ is measurable for each $\alpha \in D$. Then $f$ is measurable.

**20.** Show that the sum and product of two simple functions are simple. Show that

$$\chi_{A \cap B} = \chi_A \cdot \chi_B$$

$$\chi_{A \cup B} = \chi_A + \chi_B - \chi_A \cdot \chi_B$$

$$\chi_{\tilde{A}} = 1 - \chi_A.$$

**21. a.** Let $D$ and $E$ be measurable sets and $f$ a function with domain $D \cup E$. Show that $f$ is measurable if and only if its restrictions to $D$ and $E$ are measurable.

**b.** Let $f$ be a function with measurable domain $D$. Show that $f$ is measurable iff the function $g$ defined by $g(x) = f(x)$ for $x \in D$ and $g(x) = 0$ for $x \notin D$ is measurable.

**22. a.** Let $f$ be an extended real-valued function with measurable domain $D$, and let $D_1 = \{x: f(x) = \infty\}$, $D_2 = \{x: f(x) = -\infty\}$. Then $f$ is measurable if and only if $D_1$ and $D_2$ are measurable and the restriction of $f$ to $D \sim (D_1 \cup D_2)$ is measurable.

**b.** Prove that the product of two measurable extended real-valued functions is measurable.

**c.** If $f$ and $g$ are measurable extended real-valued functions and $\alpha$ a fixed number, then $f + g$ is measurable if we define $f + g$ to be $\alpha$ whenever it is of the form $\infty - \infty$ or $-\infty + \infty$.

**d.** Let $f$ and $g$ be measurable extended real-valued functions that are finite almost everywhere. Then $f + g$ is measurable no matter how it is defined at points where it has the form $\infty - \infty$.

**23.** Prove Proposition 22 by establishing the following lemmas:

**a.** Given a measurable function $f$ on $[a, b]$ that takes the values $\pm\infty$ only on a set of measure zero, and given $\epsilon > 0$, there is an $M$ such that $|f| \le M$ except on a set of measure less than $\epsilon/3$.

**b.** Let $f$ be a measurable function on $[a, b]$. Given $\epsilon > 0$ and $M$, there is a simple function $\varphi$ such that $|f(x) - \varphi(x)| < \epsilon$ except where $|f(x)| \ge M$. If $m \le f \le M$, then we may take $\varphi$ so that $m \le \varphi \le M$.

**c.** Given a simple function $\varphi$ on $[a, b]$, there is a step function $g$ on $[a, b]$ such that $g(x) = \varphi(x)$ except on a set of measure less than $\epsilon/3$. [Hint: Use Proposition 15.] If $m \le \varphi \le M$, then we can take $g$ so that $m \le g \le M$.

**d.** Given a step function $g$ on $[a, b]$, there is a continuous function $h$ such that $g(x) = h(x)$ except on a set of measure less than $\epsilon/3$. If $m \le g \le M$, then we may take $h$ so that $m \le h \le M$.

**24.** Let $f$ be measurable and $B$ a Borel set. Then $f^{-1}[B]$ is a measurable set. [Hint: The class of sets for which $f^{-1}[E]$ is measurable is a $\sigma$-algebra.]

**25.** Show that if $f$ is a measurable real-valued function and $g$ a continuous function defined on $(-\infty, \infty)$, then $g \circ f$ is measurable.

**26.** *Borel measurability.* A function $f$ is said to be **Borel measurable** if for each $\alpha$ the set $\{x: f(x) > \alpha\}$ is a Borel set. Verify that Propositions 18 and 19 and Theorem 20 remain valid if we replace "measurable set" by "Borel set" and "(Lebesgue) measurable" by "Borel measurable." Every Borel measurable function is Lebesgue measurable. If $f$ is Borel measurable and $B$ is a Borel set, then $f^{-1}[B]$ is a Borel set. If $f$ and $g$ are Borel measurable, so is $f \circ g$. If $f$ is Borel measurable and $g$ is Lebesgue measurable, then $f \circ g$ is Lebesgue measurable.

**27.** How much of the preceding problem can be carried out if we replace the class $\mathfrak{B}$ of Borel sets by an arbitrary $\sigma$-algebra $\mathfrak{A}$ of sets?

**28.** Let $f_1$ be the Cantor ternary function (cf. Problem 2.48), and define $f$ by $f(x) = f_1(x) + x$.

**a.** Show that $f$ is a homeomorphism of $[0, 1]$ onto $[0, 2]$.

**b.** Show that $f$ maps the Cantor set onto a set $F$ of measure 1.

**c.** Let $g = f^{-1}$. Show that there is a measurable set $A$ such that $g^{-1}[A]$ is not measurable.

**d.** Give an example of a continuous function $g$ and a measurable function $h$ such that $h \circ g$ is not measurable. Compare with Problems 25 and 26.

**e.** Show that there is a measurable set which is not a Borel set.

## 6 Littlewood's Three Principles

Speaking of the theory of functions of a real variable, J. E. Littlewood says,[6] "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 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."

We have already met two of Littlewood's principles: Various forms of the first principle are given by Proposition 15. One version of the second principle is given by Proposition 22, another version by Problem 31, and a third is given by Problem 4.15 and Proposition 6.8. The following proposition gives one version of the third principle. A slightly stronger form is given by Egoroff's theorem (Problem 30), but you will generally find the weak form adequate.

**23. Proposition:** *Let $E$ be a measurable set of finite measure, and $\langle f_n \rangle$ a sequence of measurable functions defined on $E$. Let $f$ be a real-valued function such that for each $x$ in $E$ we have $f_n(x) \to f(x)$. Then given $\epsilon > 0$ and $\delta > 0$, there is a measurable set $A \subset E$ with $mA < \delta$*

[6] Littlewood [20], p. 26.

*and an integer $N$ such that for all $x \notin A$ and all $n \geq N$,*

$$|f_n(x) - f(x)| < \epsilon.$$

***Proof:*** Let

$$G_n = \{x \in E \colon |f_n(x) - f(x)| \geq \epsilon\},$$

and set

$$E_N = \bigcup_{n=N}^{\infty} G_n = \{x \in E \colon |f_n(x) - f(x)| \geq \epsilon \text{ for some } n \geq N\}.$$

We have $E_{N+1} \subset E_N$, and for each $x \in E$ there must be some $E_N$ to which $x$ does not belong, since $f_n(x) \to f(x)$. Thus $\bigcap E_N = \emptyset$, and so, by Proposition 14, $\lim mE_N = 0$. Hence given $\delta > 0$, $\exists N$ so that $mE_N < \delta$; that is,

$$m\{x \in E \colon |f_n(x) - f(x)| \geq \epsilon \text{ for some } n \geq N\} < \delta.$$

If we write $A$ for this $E_N$, then $mA < \delta$ and

$$\tilde{A} = \{x \in E \colon |f_n(x) - f(x)| < \epsilon \text{ for all } n \geq N\}. \quad \blacksquare$$

If, as in the hypothesis of the proposition, we have $f_n(x) \to f(x)$ for each $x$, we say that the sequence $\langle f_n \rangle$ converges **pointwise** to $f$ on $E$. If there is a subset $B$ of $E$ with $mB = 0$ such that $f_n \to f$ pointwise on $E \sim B$, we say that $f_n \to f$ a.e. on $E$. We have the following trivial modification of the last proposition:

**24. Proposition:** *Let $E$ be a measurable set of finite measure, and $\langle f_n \rangle$ a sequence of measurable functions that converge to a real-valued function $f$* a.e. *on $E$. Then, given $\epsilon > 0$ and $\delta > 0$, there is a set $A \subset E$ with $mA < \delta$, and an $N$ such that for all $x \notin A$ and all $n \geq N$,*

$$|f_n(x) - f(x)| < \epsilon.$$

## Problems

**29.** Give an example to show that we must require $mE < \infty$ in Proposition 23.

**30.** Prove *Egoroff's Theorem:* If $\langle f_n \rangle$ is a sequence of measurable functions that converge to a real-valued function $f$ a.e. on a measurable set $E$ of finite measure, then given $\eta > 0$, there is a subset $A \subset E$ with $mA < \eta$ such that $f_n$ converges to $f$ *uniformly* on $E \sim A$. [Hint: Apply Proposition 24 repeatedly with $\epsilon_n = 1/n$ and $\delta_n = 2^{-n}\eta$.]

**31.** Prove *Lusin's Theorem:* Let $f$ be a measurable real-valued function on an interval $[a, b]$. Then given $\delta > 0$, there is a continuous function $\varphi$ on $[a, b]$ such that $m\{x: f(x) \neq \varphi(x)\} < \delta$. Can you do the same on the interval $(-\infty, \infty)$? [Hint: Use Egoroff's Theorem, Propositions 15 and 22, and Problem 2.40.]

**32.** Show that Proposition 23 need not be true if the integer variable $n$ is replaced by a real variable $t$; that is, construct a family $\langle f_t \rangle$ of measurable real-valued functions on $[0, 1]$ such that for each $x$ we have $\lim_{t \to 0} f_t(x) = 0$, but for some $\delta > 0$ we have $m^*\{x: f_t(x) > \frac{1}{2}\} > \delta$. [Hint: Let $P_i$ be the sets in Section 4. For $2^{-i-1} \leq t < 2^{-i}$ define $f_t$ by

$$f_t(x) = \begin{cases} 1 & \text{if } x \in P_i \text{ and } x = 2^{i+1}t - 1 \\ 0 & \text{otherwise.} \end{cases}$$

# 4 The Lebesgue Integral

## 1 The Riemann Integral

We recall a few definitions pertaining to the Riemann integral. Let $f$ be a bounded real-valued function defined on the interval $[a, b]$ and let

$$a = \xi_0 < \xi_1 < \cdots < \xi_n = b$$

be a subdivision of $[a, b]$. Then for each subdivision we can define the sums

$$S = \sum_{i=1}^{n} (\xi_i - \xi_{i-1}) M_i$$

and

$$s = \sum_{i=1}^{n} (\xi_i - \xi_{i-1}) m_i,$$

where

$$M_i = \sup_{\xi_{i-1} < x \le \xi_i} f(x), \qquad m_i = \inf_{\xi_{i-1} < x \le \xi_i} f(x).$$

We then define the upper Riemann integral of $f$ by

$$R \overline{\int_a^b} f(x)\, dx = \inf S$$

with the infimum taken over all possible subdivisions of [$a$, $b$]. Similarly, we define the lower integral

$$R \underline{\int_a^b} f(x)\, dx = \sup s.$$

The upper integral is always at least as large as the lower integral, and if the two are equal we say that $f$ is *Riemann integrable* and call this common value the Riemann integral of $f$. We shall denote it by

$$R \int_a^b f(x)\, dx$$

to distinguish it from the Lebesgue integral, which we shall consider later.

By a **step function** we mean a function $\psi$ which has the form

$$\psi(x) = c_i, \qquad \xi_{i-1} < x < \xi_i$$

for some subdivision of [$a$, $b$] and some set of constant $c_i$. Under practically anybody's definition of an integral we have

$$\int_a^b \psi(x)\, dx = \sum_{i=1}^n c_i(\xi_i - \xi_{i-1}).$$

With this in mind we see that

$$R \overline{\int_a^b} f(x)\, dx = \inf \int_a^b \psi(x)\, dx$$

for all step functions $\psi(x) \geq f(x)$. Similarly,

$$R \underline{\int_a^b} f(x)\, dx = \sup \int_a^b \varphi(x)\, dx$$

for all step functions $\varphi(x) \leq f(x)$.

## Problem

**1. a.** Show that if

$$f(x) = \begin{cases} 0 & x \text{ irrational} \\ 1 & x \text{ rational,} \end{cases}$$

then

$$R \overline{\int_a^b} f(x)\, dx = b - a \quad \text{and} \quad R \underline{\int_a^b} f(x)\, dx = 0.$$

**b.** Construct a sequence $\langle f_n \rangle$ of nonnegative, Riemann integrable functions such that $f_n$ increases monotonically to $f$. What does this imply about changing the order of integration and the limiting process?

## 2 The Lebesgue Integral of a Bounded Function over a Set of Finite Measure

Problem 1 in the preceding section shows some of the shortcomings of the Riemann integral. In particular, we would like a function that is 1 on a measurable set and zero elsewhere to be integrable and have as its integral the measure of the set.

The function $\chi_E$ defined by

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

is called the *characteristic function* of $E$. A linear combination

$$\varphi(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x)$$

is called a **simple function** if the sets $E_i$ are measurable. This representation for $\varphi$ is not unique. However, we note that a function $\varphi$ is simple if and only if it is measurable and assumes only a finite number of values. If $\varphi$ is a simple function and $\{a_1, \ldots, a_n\}$ the set of nonzero values of $\varphi$, then

$$\varphi = \sum a_i \chi_{A_i},$$

where $A_i = \{x\colon \varphi(x) = a_i\}$. This representation for $\varphi$ is called the canonical representation, and it is characterized by the fact that the $A_i$ are disjoint and the $a_i$ distinct and nonzero.

If $\varphi$ vanishes outside a set of finite measure, we define the integral of $\varphi$ by

$$\int \varphi(x)\, dx = \sum_{i=1}^{n} a_i mA_i$$

when $\varphi$ has the canonical representation $\varphi = \sum_{i=1}^{n} a_i \chi_{A_i}$. We sometimes abbreviate the expression for this integral to $\int \varphi$. If $E$ is any measurable set, we define

$$\int_E \varphi = \int \varphi \cdot \chi_E.$$

It is often convenient to use representations which are not canonical, and the following lemma is useful:

**1. Lemma:** *Let* $\varphi = \sum_{i=1}^{n} a_i \chi_{E_i}$, *with* $E_i \cap E_j = \emptyset$ *for* $i \neq j$. *Suppose each set* $E_i$ *is a measurable set of finite measure. Then*

$$\int \varphi = \sum_{i=1}^{n} a_i mE_i.$$

***Proof:*** The set $A_a = \{x\colon \varphi(x) = a\} = \bigcup_{a_i = a} E_i$. Hence $amA_a = \sum_{a_i = a} a_i mE_i$ by the additivity of $m$, and so

$$\begin{aligned} \int \varphi(x)\, dx &= \sum amA_a \\ &= \sum a_i mE_i. \quad \blacksquare \end{aligned}$$

**2. Proposition:** *Let* $\varphi$ *and* $\psi$ *be simple functions which vanish outside a set of finite measure. Then*

$$\int (a\varphi + b\psi) = a\int \varphi + b\int \psi,$$

*and if* $\varphi \geq \psi$ a.e., *then*

$$\int \varphi \geq \int \psi.$$

***Proof:*** Let $\{A_i\}$ and $\{B_i\}$ be the sets occurring in canonical representations of $\varphi$ and $\psi$. Let $A_0$ and $B_0$ be the sets where $\varphi$ and $\psi$ are zero. Then the sets $E_k$ obtained by taking the intersections $A_i \cap B_j$ form a finite disjoint collection of measurable sets, and we may write

$$\varphi = \sum_{k=1}^{N} a_k \chi_{E_k}$$

$$\psi = \sum_{k=1}^{N} b_k \chi_{E_k},$$

and so

$$a\varphi + b\psi = \sum (aa_k + bb_k)\chi_{E_k},$$

whence $\int (a\varphi + b\psi) = a\int \varphi + b\int \psi$ follows from Lemma 1. To prove the second statement, we note that

$$\int \varphi - \int \psi = \int (\varphi - \psi) \geq 0,$$

since the integral of a simple function which is greater than or equal to zero a.e. is nonnegative by the definition of the integral. ▮

It follows from this proposition that, if $\varphi = \sum_{i=1}^{n} a_i \chi_{E_i}$, then $\int \varphi = \sum a_i mE_i$, and so the restriction of Lemma 1 that the sets $E_i$ be disjoint is unnecessary.

Let $f$ be a bounded real-valued function and $E$ a measurable set of finite measure. By analogy with the Riemann integral we consider for simple functions $\varphi$ and $\psi$ the numbers

$$\inf_{\psi \geq f} \int_E \psi$$

and

$$\sup_{\varphi \leq f} \int_E \varphi,$$

and ask when these two numbers are equal. The answer is given by the following proposition:

**3. Proposition:** *Let $f$ be defined and bounded on a measurable set $E$ with $mE$ finite. In order that*

$$\inf_{f \leq \psi} \int_E \psi(x)\, dx = \sup_{f \geq \varphi} \int_E \varphi(x)\, dx$$

*for all simple functions $\varphi$ and $\psi$, it is necessary and sufficient that $f$ be measurable.*

***Proof:*** Let $f$ be bounded by $M$ and suppose that $f$ is measurable. Then the sets

$$E_k = \left\{x : \frac{kM}{n} \geq f(x) > \frac{(k-1)M}{n}\right\}, \quad -n \leq k \leq n,$$

are measurable, disjoint, and have union $E$. Thus

$$\sum_{k=-n}^{n} mE_k = mE.$$

The simple functions defined by

$$\psi_n(x) = \frac{M}{n} \sum_{k=-n}^{n} k\chi_{E_k}(x)$$

and

$$\varphi_n(x) = \frac{M}{n} \sum_{k=-n}^{n} (k-1)\chi_{E_k}(x)$$

satisfy

$$\varphi_n(x) \le f(x) \le \psi_n(x).$$

Thus

$$\inf \int_E \psi(x)\,dx \le \int_E \psi_n(x)\,dx = \frac{M}{n} \sum_{k=-n}^{n} kmE_k$$

and

$$\sup \int_E \varphi(x)\,dx \ge \int_E \varphi_n(x)\,dx = \frac{M}{n} \sum_{k=-n}^{n} (k-1)mE_k,$$

whence

$$0 \le \inf \int_E \psi(x)\,dx - \sup \int_E \varphi(x)\,dx \le \frac{M}{n} \sum_{k=-n}^{n} mE_k = \frac{M}{n} mE.$$

Since $n$ is arbitrary, we have

$$\inf \int_E \psi(x)\,dx - \sup \int_E \varphi(x)\,dx = 0,$$

and the condition is sufficient.

Suppose now that

$$\inf_{\psi \ge f} \int_E \psi(x)\,dx = \sup_{\varphi \le f} \int_E \varphi(x)\,dx.$$

Then, given $n$, there are simple functions $\varphi_n$ and $\psi_n$ such that

$$\varphi_n(x) \le f(x) \le \psi_n(x)$$

and

$$\int \psi_n(x)\,dx - \int \varphi_n(x)\,dx < \frac{1}{n}.$$

Then the functions

$$\psi^* = \inf \psi_n$$

and

$$\varphi^* = \sup \varphi_n$$

are measurable by Theorem 3.20, and

$$\varphi^*(x) \le f(x) \le \psi^*(x).$$

Now the set

$$\Delta = \{x: \varphi^*(x) < \psi^*(x)\}$$

is the union of the sets

$$\Delta_\nu = \left\{x: \varphi^*(x) < \psi^*(x) - \frac{1}{\nu}\right\}.$$

But each $\Delta_\nu$ is contained in the set $\{x: \varphi_n(x) < \psi_n(x) - 1/\nu\}$, and this latter set has measure less than $\nu/n$. Since $n$ is arbitrary, $m\Delta_\nu = 0$, and so $m\Delta = 0$. Thus $\varphi^* = \psi^*$ except on a set of measure zero, and $\varphi^* = f$ except on a set of measure zero. Thus $f$ is measurable by Proposition 3.21, and the condition is also necessary. ▮

**Definition:** *If $f$ is a bounded measurable function defined on a measurable set $E$ with $mE$ finite, we define the (Lebesgue) integral of $f$ over $E$ by*

$$\int_E f(x)\,dx = \inf \int_E \psi(x)\,dx$$

*for all simple functions $\psi \ge f$.*

We sometimes write the integral as $\int_E f$. If $E = [a, b]$, we write $\int_a^b f$ instead of $\int_{[a,b]} f$. If $f$ is a bounded measurable function that vanishes outside a set $E$ of finite measure, we write $\int f$ for $\int_E f$. Note that $\int_E f = \int f \cdot \chi_E$. The following corollary to Proposition 3 shows that the Lebesgue integral is in fact a generalization of the Riemann integral.

**4. Proposition:** *Let $f$ be a bounded function defined on $[a, b]$. If $f$ is Riemann integrable on $[a, b]$, then it is measurable and*

$$R\int_a^b f(x)\,dx = \int_a^b f(x)\,dx.$$

***Proof:*** Since every step function is also a simple function, we have

$$R\underline{\int_a^b} f(x)\,dx \le \sup_{\varphi \le f} \int_a^b \varphi(x)\,dx \le \inf_{\psi \ge f} \int_s^b \psi(x)\,dx \le R\overline{\int_a^b} f(x)\,dx.$$

Since $f$ is Riemann integrable, the inequalities are all equalities, and $f$ is measurable by Proposition 3. ▮

**5. Proposition:** *If $f$ and $g$ are bounded measurable functions defined on a set $E$ of finite measure, then:*

i. $$\int_E (af + bg) = a\int_E f + b\int_E g.$$

ii. *If* $f = g$ a.e., *then*

$$\int_E f = \int_E g.$$

iii. *If* $f \le g$ a.e., *then*

$$\int_E f \le \int_E g.$$

*Hence*

$$\left|\int f\right| \le \int |f|.$$

iv. *If $A \le f(x) \le B$, then*

$$AmE \le \int_E f \le BmE.$$

v. *If $A$ and $B$ are disjoint measurable sets of finite measure, then*

$$\int_{A \cup B} f = \int_A f + \int_B f.$$

***Proof:*** If $\psi$ is a simple function so is $a\psi$, and conversely (if $a \ne 0$). Hence for $a > 0$,

$$\int_E af = \inf_{\psi \ge f} \int_E a\psi = a \inf_{\psi \ge f} \int_E \psi = a \int_E f.$$

If $a < 0$,

$$\int_E af = \inf_{\varphi \le f} \int_E a\varphi = a \sup_{\varphi \le f} \int_E \varphi = a \inf_{\psi \ge f} \int_E \psi = a \int_E f,$$

using Proposition 3.

If $\psi_1$ is a simple function $f \le \psi_1$, and $\psi_2$ a simple function, $g \le \psi_2$, then $\psi_1 + \psi_2$ is a simple function $f + g \le \psi_1 + \psi_2$. Hence

$$\int_E f + g \le \int_E (\psi_1 + \psi_2) = \int_E \psi_1 + \int_E \psi_2 .$$

Since the infimum of the right side is $\int f + \int g$, we have

$$\int_E f + g \le \int_E f + \int_E g.$$

On the other hand, $\varphi_1 \le f$ and $\varphi_2 \le g$ imply $\varphi_1 + \varphi_2$ is a simple function not greater than $f + g$. Hence

$$\int_E f + g \ge \int_E (\varphi_1 + \varphi_2) = \int_E \varphi_1 + \int_E \varphi_2 .$$

Since now the supremum of the right side is $\int f + \int g$, we have

$$\int_E f + g \ge \int_E f + \int_E g,$$

and statement (i) of the proposition follows.

To prove (ii), it now suffices to show that

$$\int_E f - g = 0.$$

Since $f - g = 0$ a.e., it follows that if $\psi \ge f - g$, $\psi \ge 0$ a.e. From this it follows that

$$\int_E \psi \ge 0,$$

whence

$$\int_E f - g \ge 0.$$

Similarly,

$$\int_E f - g \le 0,$$

whence (ii). This proof also serves to establish (iii). Statement (iv) follows from (iii) and the fact that

$$\int_E 1 = mE.$$

Statement (v) follows from (i) and the fact that $\chi_{A \cup B} = \chi_A + \chi_B$. ▮

We next prove a proposition, which will be used to prove Theorem 16. The proposition is a special case of Theorem 16.

**6. Proposition (Bounded Convergence Theorem):** *Let $\langle f_n \rangle$ be a sequence of measurable functions defined on a set $E$ of finite measure, and suppose that there is a real number $M$ such that $|f_n(x)| \leq M$ for all $n$ and all $x$. If $f(x) = \lim f_n(x)$ for each $x$ in $E$, then*

$$\int_E f = \lim \int_E f_n .$$

***Proof:*** The proof of this proposition furnishes a nice illustration of the use of Littlewood's "three principles." The conclusion of the proposition would be trivial if $\langle f_n \rangle$ converged to $f$ uniformly. Littlewood's third principle states that if $\langle f_n \rangle$ converges to $f$ pointwise, then $\langle f_n \rangle$ is "nearly" uniformly convergent to $f$. A precise version of this principle is given by Proposition 3.23, which states that, given $\epsilon > 0$, there is an $N$ and a measurable set $A \subset E$ with $mA < \epsilon/4M$ such that for $n \geq N$ and $x \in E \sim A$ we have $|f_n(x) - f(x)| < \epsilon/2mE$. Thus

$$\begin{aligned} \left| \int_E f_n - \int_E f \right| &= \left| \int_E f_n - f \right| \\ &\leq \int_E |f_n - f| \\ &= \int_{E \sim A} |f_n - f| + \int_A |f_n - f| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Hence

$$\int_E f_n \rightarrow \int_E f. \quad \blacksquare$$

The following beautiful result gives a necessary and sufficient condition for the Riemann integrability of a bounded function. It is due to Lebesgue.

**7. Proposition:** *A bounded function $f$ on $[a, b]$ is Riemann integrable if and only if the set of points at which $f$ is discontinuous has measure zero.*

## Problem

**2. a.** Let $f$ be a bounded function on $[a, b]$, and let $h$ be the upper envelope of $f$ (cf. Problem 2.51). Then $R \bar{\int}_a^b f = \int_a^b h$. (If $\varphi \geq f$ is a step function, then $\varphi \geq h$ except at a finite number of points, and so $\int_a^b h \leq R \bar{\int}_a^b f$. But there is a sequence $\langle \varphi_n \rangle$ of step functions such that $\varphi_n \downarrow h$. By Proposition 6 we have $\int_a^b h = \lim \int_a^b \varphi_n \geq R \bar{\int}_a^b f$.)

**b.** Use part (a) to prove Proposition 7.

## 3 The Integral of a Nonnegative Function

If $f$ is a nonnegative measurable function defined on a measurable set $E$, we define

$$\int_E f = \sup_{h \leq f} \int_E h,$$

where $h$ is a bounded measurable function such that $m\{x: h(x) \neq 0\}$ is finite.

**8. Proposition:** *If $f$ and $g$ are nonegative measurable functions, then:*

i. $$\int_E cf = c \int_E f, \qquad c > 0.$$

ii. $$\int_E f + g = \int_E f + \int_E g.$$

iii. *If $f \leq g$ a.e., then*

$$\int_E f \leq \int_E g.$$

***Proof:*** Parts (i) and (iii) follow directly from Proposition 5, and we prove only (ii) in detail. If $h(x) \leq f(x)$ and $k(x) \leq g(x)$, we have

$h(x) + k(x) \le f(x) + g(x)$, and so

$$\int_E h + \int_E k \le \int_E f + g.$$

Taking suprema, we have

$$\int_E f + \int_E g \le \int_E f + g.$$

On the other hand, let $l$ be a bounded measurable function which vanishes outside a set of finite measure and which is not greater than $f + g$. Then we define the functions $h$ and $k$ by setting

$$h(x) = \min\,(f(x), l(x))$$

and

$$k(x) = l(x) - h(x).$$

We have

$$h(x) \le f(x)$$

and

$$k(x) \le g(x),$$

while $h$ and $k$ are bounded by the bound for $l$ and vanish where $l$ vanishes. Hence $\int_E l = \int_E h + \int_E k \le \int_E f + \int_E g$, and so

$$\int_E f + \int_E g \ge \int_E f + g. \quad \blacksquare$$

**9. Theorem (Fatou's Lemma):** *If $\langle f_n \rangle$ is a sequence of nonnegative measurable functions and $f_n(x) \to f(x)$ almost everywhere on a set $E$, then*

$$\int_E f \le \underline{\lim} \int_E f_n.$$

***Proof:*** Without loss of generality we may assume that the convergence is everywhere, since integrals over sets of measure zero are zero. Let $h$ be a bounded measurable function which is not greater than $f$ and which vanishes outside a set $E'$ of finite measure. Define a function $h_n$ by setting

$$h_n(x) = \min\,\{h(x), f_n(x)\}.$$

Then $h_n$ is bounded by the bound for $h$ and vanishes outside $E'$. Now $h_n(x) \to h(x)$ for each $x$ in $E'$. Thus, by Proposition 6, we have

$$\int_E h = \int_{E'} h = \lim \int_{E'} h_n \leq \underline{\lim} \int_E f_n .$$

Taking the supremum over $h$, we get

$$\int_E f \leq \underline{\lim} \int_E f_n . \quad \blacksquare$$

**10. Monotone Convergence Theorem:** *Let $\langle f_n \rangle$ be an increasing sequence of nonnegative measurable functions, and let $f = \lim f_n$ a.e. Then*

$$\int f = \lim \int f_n .$$

***Proof:*** By Theorem 9 we have

$$\int f \leq \underline{\lim} \int f_n .$$

But for each $n$ we have $f_n \leq f$, and so $\int f_n \leq \int f$. But this implies

$$\overline{\lim} \int f_n \leq \int f.$$

Hence

$$\int f = \lim \int f_n . \quad \blacksquare$$

**11. Corollary:** *Let $u_n$ be a sequence of nonnegative measurable functions, and let $f = \sum_{n=1}^{\infty} u_n$. Then*

$$\int f = \sum_{n=1}^{\infty} \int u_n .$$

**12. Proposition:** *Let $f$ be a nonnegative function and $\langle E_i \rangle$ a disjoint sequence of measurable sets. Let $E = \bigcup E_i$. Then*

$$\int_E f = \sum \int_{E_i} f.$$

***Proof:*** Let $u_i = f \cdot \chi_{E_i}$. Then $f \cdot \chi_E = \sum u_i$, and so the proposition follows from the preceding corollary. ▮

**Definition:** *A nonnegative measurable function f is called **integrable** over the measurable set E if*

$$\int_E f < \infty.$$

**13. Proposition:** *Let f and g be two nonnegative measurable functions. If f is integrable over E and $g(x) < f(x)$ on E, then g is also integrable on E, and*

$$\int_E f - g = \int_E f - \int_E g.$$

***Proof:*** By Proposition 8,

$$\int_E f = \int_E (f-g) + \int_E g.$$

Since the left side is finite, the terms on the right must also be finite, and so $g$ is integrable. ▮

**14. Proposition:** *Let f be a nonnegative function which is integrable over a set E. Then given $\epsilon > 0$ there is a $\delta > 0$ such that for every set $A \subset E$ with $mA < \delta$ we have*

$$\int_A f < \epsilon.$$

***Proof:*** The proposition would be trivial if $f$ were bounded. Set $f_n(x) = f(x)$ if $f(x) \leq n$ and $f_n(x) = n$ otherwise. Then each $f_n$ is bounded and $f_n$ converges to $f$ at each point. By the Monotone Convergence Theorem there is an $N$ such that $\int_E f_N > \int_E f - \epsilon/2$, and $\int_E f - f_N < \epsilon/2$. Choose $\delta < \epsilon/2N$. If $mA < \delta$, we have

$$\begin{aligned}\int_A f &= \int_A (f - f_N) + \int_A f_N \\ &< \int_E (f - f_N) + NmA \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$ ▮

## Problems

3. Let $f$ be a nonnegative measurable function. Show that $\int f = 0$ implies $f = 0$ a.e.

4. Let $f$ be a nonnegative measurable function.

  a. Show that there is an increasing sequence $\langle \varphi_n \rangle$ of nonnegative simple functions each of which vanishes outside a set of finite measure such that $f = \lim \varphi_n$.

  b. Show that $\int f = \sup \int \varphi$ over all simple functions $\varphi \leq f$.

5. Let $f$ be a nonnegative integrable function. Show that the function $F$ defined by

$$F(x) = \int_{-\infty}^{x} f$$

is continuous by using Theorem 10.

6. Let $\langle f_n \rangle$ be a sequence of nonnegative measurable functions that converge to $f$, and suppose $f_n \leq f$ for each $n$. Then

$$\int f = \lim \int f_n .$$

7. a. Show that we may have strict inequality in Fatou's Lemma. [Consider the sequence $\langle f_n \rangle$ defined by $f_n(x) = 1$ if $n \leq x < n + 1$, with $f_n(x) = 0$ otherwise.]

  b. Show that the Monotone Convergence Theorem need not hold for decreasing sequences of functions. [Let $f_n(x) = 0$ if $x < n$, $f_n(x) = 1$ for $x \geq n$.]

8. Prove the following generalization of Fatou's Lemma: If $\langle f_n \rangle$ is a sequence of nonnegative functions, then

$$\int \underline{\lim}\, f_n \leq \underline{\lim} \int f_n .$$

9. Let $\langle f_n \rangle$ be a sequence of nonnegative measurable functions on $(-\infty, \infty)$ such that $f_n \to f$ a.e., and suppose that $\int f_n \to \int f < \infty$. Then for each measurable set $E$ we have $\int_E f_n \to \int_E f$.

## 4 The General Lebesgue Integral

By the positive part $f^+$ of a function $f$ we mean the function $f^+ = f \vee 0$; that is,

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

Similarly, we define the negative part $f^-$ by $f^- = (-f) \vee 0$. If $f$ is measurable, so are $f^+$ and $f^-$. We have

$$f = f^+ - f^-$$

and

$$|f| = f^+ + f^-.$$

With these notions in mind we make the following definition:

**Definition:** *A measurable function f is said to be integrable over E if $f^+$ and $f^-$ are both integrable over E. In this case we define*

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

**15. Proposition:** *Let f and g be integrable over E. Then:*

i. *The function cf is integrable over E, and* $\int_E cf = c \int_E f$.

ii. *The function f + g is integrable over E, and*

$$\int_E f + g = \int_E f + \int_E g.$$

iii. *If $f \le g$ a.e., then* $\int_E f \le \int_E g$.

iv. *If A and B are disjoint measurable sets contained in E, then*

$$\int_{A \cup B} f = \int_A f + \int_B f.$$

***Proof:*** Part (i) follows directly from the definition of the integral and Proposition 8. To prove part (ii), we first note that if $f_1$ and $f_2$ are nonnegative integrable functions with $f = f_1 - f_2$, then $f^+ + f_2 = f^- + f_1$. For Proposition 8 tells us that

$$\int f^+ + \int f_2 = \int f^- + \int f_1,$$

and so

$$\int f = \int f^+ - \int f^- = \int f_1 - \int f_2.$$

But, if $f$ and $g$ are integrable, so are $f^+ + g^+$ and $f^- + g^-$, and $(f+g) = (f^+ + g^+) - (f^- + g^-)$. Hence

$$\begin{aligned}\int (f+g) &= \int (f^+ + g^+) - \int (f^- + g^-)\\ &= \int f^+ + \int g^+ - \int f^- - \int g^-\\ &= \int f + \int g.\end{aligned}$$

Part (iii) follows from part (ii) and the fact that the integral of a nonnegative integrable function is nonnegative. For (iv) we have

$$\begin{aligned}\int_{A \cup B} f &= \int f\chi_{A \cup B}\\ &= \int f\chi_A + \int f\chi_B\\ &= \int_A f + \int_B f. \quad \blacksquare\end{aligned}$$

It should be noted that $f + g$ is not defined at points where $f = \infty$ and $g = -\infty$ and where $f = -\infty$ and $g = \infty$. However, the set of such points must have measure zero, since $f$ and $g$ are integrable. Hence the integrability and the value of $\int (f + g)$ are independent of the choice of values in these ambiguous cases.

**16. Lebesgue Convergence Theorem:** *Let $g$ be integrable over $E$ and let $\langle f_n \rangle$ be a sequence of measurable functions such that $|f_n| \le g$ on $E$ and for almost all $x$ in $E$ we have $f(x) = \lim f_n(x)$. Then*

$$\int_E f = \lim \int_E f_n.$$

***Proof:*** The function $g - f_n$ is nonnegative, and so by Theorem 9

$$\int_E (g - f) \le \underline{\lim} \int_E (g - f_n).$$

Since $|f| \le g$, $f$ is integrable, and we have

$$\int_E g - \int_E f \le \int_E g - \overline{\lim} \int_E f_n,$$

whence

$$\int_E f \geq \overline{\lim} \int_E f_n .$$

Similarly, considering $g + f_n$, we get

$$\int_E f \leq \underline{\lim} \int_E f_n ,$$

and the theorem follows. ▮

The above theorem requires that the sequence $\langle f_n \rangle$ be dominated by a fixed integrable function $g$. However, the proof does not require quite so much. If we replace the appropriate $g$'s in the above proof by $g_n$'s, we get the following generalization of the Lebesgue Convergence Theorem:

**17. Theorem:** *Let $\langle g_n \rangle$ be a sequence of integrable functions which converges* a.e. *to an integrable function $g$. Let $\langle f_n \rangle$ be a sequence of measurable functions such that $|f_n| \leq g_n$ and $\langle f_n \rangle$ converges to $f$* a.e. *If*

$$\int g = \lim \int g_n ,$$

*then*

$$\int f = \lim \int f_n .$$

If $\langle f_n \rangle$ is a sequence of measurable functions that converges a.e. to $f$, then Fatou's Lemma, the Monotone Convergence Theorem, and the Lebesgue Convergence Theorem all state that under suitable hypotheses we can assert something about $\int f$ in terms of $\int f_n$. Fatou's Lemma has the weakest hypothesis: We need only have $f_n$ bounded below by zero (or more generally by an integrable function). Consequently, the conclusion of Fatou's Lemma is weaker than that of the others: We can only assert $\int f \leq \underline{\lim} \int f_n$. The Lebesgue Convergence Theorem requires that $f_n$ be bounded from above and below by fixed integrable functions and then asserts the equality of $\int f$ and $\lim \int f_n$. The Monotone Convergence Theorem (as generalized in Problem 6) is something of a hybrid: It requires that the $f_n$ be bounded from below by zero (or an integrable function) and above by the limit function $f$ itself. Of course, if $f$ is integrable, this is a special case of the Lebesgue Convergence Theorem, but the advantage of Fatou's Lemma and the Monotone Convergence Theorem is that they are applicable even if $f$ is not integrable and are often a good way of showing that $f$ is integrable. Fatou's Lemma and the Monotone Convergence Theorem are very close in the sense that each can be derived from the other using only the fact that integration is positive and linear.

## Problems

**10. a.** Show that if $f$ is integrable over $E$, then so is $|f|$ and

$$\left|\int_E f\right| \le \int_E |f|.$$

Does the integrability of $|f|$ imply that of $f$?

**b.** The improper Riemann integral of a function may exist without the function being integrable (in the sense of Lebesgue), e.g., if $f(x) = (\sin x)/x$ on $[0, \infty]$. If $f$ is integrable, show that the improper Riemann integral is equal to the Lebesgue integral whenever the former exists.

**11.** If $\varphi$ is a simple function, we have two definitions for $\int \varphi$, that on page 77 and that on page 90. Show that they are the same.

**12.** Let $g$ be an integrable function on a set $E$ and suppose that $\langle f_n \rangle$ is a sequence of measurable functions such that $|f_n(x)| \le g(x)$ a.e. on $E$. Then

$$\int_E \underline{\lim}\, f_n \le \underline{\lim} \int_E f_n \le \overline{\lim} \int_E f_n \le \int_E \overline{\lim}\, f_n.$$

**13.** Let $h$ be an integrable function and $\langle f_n \rangle$ a sequence of measurable functions with $f_n \ge -h$ and $\lim f_n = f$. Show that $\int f_n$ and $\int f$ have a meaning and that $\int f \le \underline{\lim} \int f_n$.

**14. a.** Show that under the hypotheses of Theorem 17 we have $\int |f_n - f| \to 0$.

**b.** Let $\langle f_n \rangle$ be a sequence of integrable functions such that $f_n \to f$ a.e. with $f$ integrable. Then $\int |f - f_n| \to 0$ if and only if $\int |f_n| \to \int |f|$.

**15. a.** Let $f$ be integrable over $E$. Then, given $\epsilon > 0$, there is a simple function $\varphi$ such that

$$\int_E |f - \varphi| < \epsilon.$$

[Apply Problem 4 to the positive and negative parts of $f$.]

**b.** Under the same hypothesis there is a step function $\psi$ such that

$$\int_E |f - \psi| < \varepsilon.$$

[Combine part (a) with Proposition 3.22.]

**c.** Under the same hypothesis there is a continuous function $g$ vanishing outside a finite interval such that

$$\int_E |f - g| < \epsilon.$$

**16.** Establish the *Riemann–Lebesgue Theorem:* If $f$ is an integrable function on $(-\infty, \infty)$, then $\lim\limits_{n\to\infty} \int_{-\infty}^{\infty} f(x) \cos nx \, dx = 0$. [Hint: The theorem is easy if $f$ is a step function. Use Problem 15.]

**17. a.** Let $f$ be integrable over $(-\infty, \infty)$. Then

$$\int f(x)\, dx = \int f(x + t)\, dx.$$

**b.** Let $g$ be a bounded measurable function. Then

$$\lim_{t\to 0} \int_{-\infty}^{\infty} |g(x)[f(x) - f(x + t)]| = 0.$$

[Hint: If $f$ is continuous and vanishes outside a finite interval, this result follows from the uniform continuity of $f$. Apply Problem 15.]

**18.** Let $f$ be a function of two variables $\langle x, t\rangle$ which is defined in the square $Q = \{\langle x, t\rangle : 0 \le x \le 1, 0 \le t \le 1\}$ and which is a measurable function of $x$ for each fixed value of $t$. Suppose that $\lim\limits_{t\to 0} f(x, t) = f(x)$ and that for all $t$ we have $|f(x, t)| \le g(x)$, where $g$ is an integrable function on $[0, 1]$. Then

$$\lim_{t\to 0} \int f(x, t)\, dx = \int f(x)\, dx.$$

(Problem 2.49f is useful here.) Show also that if the function $f(x, t)$ is continuous in $t$ for each $x$, then

$$h(t) = \int f(x, t)\, dx$$

is a continuous function of $t$.

**19.** Let $f$ be a function defined and bounded in the square $Q = \{\langle x, t\rangle : 0 \le x \le 1, 0 \le t \le 1\}$, and suppose that for each fixed $t$ the function $f$ is a measurable function of $x$. For each $\langle x, t\rangle \in Q$ let the partial derivative $\partial f/\partial t$ exist. Suppose that $\partial f/\partial t$ is bounded in $Q$. Then

$$\frac{d}{dt}\int_0^1 f(x, t)\,dx = \int_0^1 \frac{\partial f}{\partial t}\,dx.$$

## *5 Convergence in Measure

Suppose that $\langle f_n \rangle$ is a sequence of measurable functions such that $\int |f_n| \to 0$. What can we say about the sequence $\langle f_n \rangle$? Perhaps the most important property of such a sequence is that for each positive $\eta$ the measure of the sets $\{x\colon |f_n(x)| > \eta\}$ must tend to zero. This leads us to the following definition:

**Definition:** *A sequence $\langle f_n \rangle$ of measurable functions is said to converge to $f$ in measure if, given $\epsilon > 0$, there is an $N$ such that for all $n \geq N$ we have*

$$m\{x\colon |f(x) - f_n(x)| \geq \epsilon\} < \epsilon.$$

It follows directly from this definition and Proposition 3.23 that if $\langle f_n \rangle$ is a sequence of measurable functions defined on a measurable set $E$ of finite measure and $f_n \to f$ a.e., then $\langle f_n \rangle$ converges to $f$ in measure.

An example of a sequence $\langle f_n \rangle$ that converges to zero in measure on $[0, 1]$ but such that $\langle f_n(x) \rangle$ does not converge for any $x$ in $[0, 1]$ can be constructed as follows: Let $n = k + 2^\nu$, $0 \leq k < 2^\nu$, and set $f_n(x) = 1$ if $x \in [k2^{-\nu}, (k+1)2^{-\nu}]$ and $f_n(x) = 0$ otherwise. Then

$$m\{x\colon |f_n(x)| > \epsilon\} \leq \frac{2}{n},$$

and so $f_n \to 0$ in measure, although for any $x \in [0, 1]$, the sequence $\langle f_n(x) \rangle$ has the value 1 for arbitrarily large values of $n$ and so does not converge. We do, however, have the following proposition:

**18. Proposition:** *Let $\langle f_n \rangle$ be a sequence of measurable functions that converges in measure to $f$. Then there is a subsequence $\langle f_{n_k} \rangle$ that converges to $f$ almost everywhere.*

***Proof:*** Given $\nu$, there is an integer $n_\nu$ such that for all $n \geq n_\nu$ we have

$$m\{x\colon |f_n(x) - f(x)| \geq 2^{-\nu}\} < 2^{-\nu}.$$

Let $E_\nu = \{x: |f_{n_\nu}(x) - f(x)| \geq 2^{-\nu}\}$. Then, if $x \notin \bigcup_{\nu=k}^{\infty} E_\nu$, we must have $|f_{n_\nu}(x) - f(x)| < 2^{-\nu}$ for $\nu \geq k$, and so $f_{n_\nu}(x) \to f(x)$. Hence $f_{n_\nu}(x) \to f(x)$ for any $x \notin A = \bigcap_{k=1}^{\infty} \bigcup_{\nu=k}^{\infty} E_\nu$. But $mA \leq m\left[\bigcup_{\nu=k}^{\infty} E_\nu\right] \leq \sum_{\nu=k}^{\infty} mE_\nu = 2^{-k+1}$. Hence $mA = 0$. ▮

**19. Corollary:** *Let $\langle f_n \rangle$ be a sequence of measurable functions defined on a measurable set E of finite measure. Then $\langle f_n \rangle$ converges to f in measure if and only if every subsequence of $\langle f_n \rangle$ has in turn a subsequence that converges almost everywhere to f.*

**20. Proposition:** *Fatou's Lemma and the Monotone and Lebesgue Convergence Theorems remain valid if 'convergence* a.e.*' is replaced by 'convergence in measure'.*

## Problems

**20.** Show that if $\langle f_n \rangle$ is a sequence that converges to $f$ in measure, then each subsequence $\langle f_{n_k} \rangle$ converges to $f$ in measure.

**21.** Deduce Proposition 20 from Proposition 18 using Problems 20 and 2.12.

**22.** Prove that a sequence $\langle f_n \rangle$ of measurable functions on a set $E$ of finite measure converges to $f$ in measure if and only if every subsequence of $\langle f_n \rangle$ has in turn a subsequence that converges to $f$ in measure.

**23.** Prove Corollary 19.

**24.** Use Proposition 14 to prove directly that if $f_n \to f$ in measure and if there is an integrable function $g$ such that for all $n$ we have $|f_n| \leq g$, then $\int |f_n - f| \to 0$.

**25.** A sequence $\langle f_n \rangle$ of measurable functions is said to be a Cauchy sequence in measure if given $\epsilon > 0$ there is an $N$ such that for all $m, n \geq N$ we have

$$m\{x: |f_n(x) - f_m(x)| \geq \epsilon\} < \epsilon.$$

Show that if $\langle f_n \rangle$ is a Cauchy sequence in measure, then there is a function $f$ to which the sequence $\langle f_n \rangle$ converges in measure. [Choose $n_{\nu+1} > n_\nu$ so that $m\{x: |f_{n_\nu} - f_{n_{\nu+1}}| > 2^{-\nu}\} < 2^{-\nu}$. Then the series $\sum (f_{n_{\nu+1}} - f_{n_\nu})$ converges almost everywhere to a function $g$. Let $f = g + f_{n_1}$. Then $f_{n_\nu} \to f$ in measure, and one can show that consequently $f_n \to f$ in measure.]

# 5 Differentiation and Integration

In this chapter we consider the sense in which differentiation is the inverse of integration. In particular we shall be concerned with the following questions:

When does

$$\int_a^b f'(x)\,dx = f(b) - f(a)?$$

When does

$$\frac{d}{dx}\int_a^x f(y)\,dy = f(x)?$$

From the theory of Riemann integration it is known that the second relation is true if $x$ is a point of continuity of $f$. We shall show that more generally this relation holds almost everywhere. Thus differentiation is the inverse of Lebesgue integration. The first question, however, is more dificult, and even using the Lebesgue integral it is true only for a certain class of functions, which we shall characterize.

## 1 Differentiation of Monotone Functions

Let $\mathcal{I}$ be a collection of intervals. Then we say that $\mathcal{I}$ covers a set $E$ *in the sense of Vitali*, if for each $\epsilon > 0$ and any $x$ in $E$, there is an interval $I \in \mathcal{I}$ such that $x \in I$ and $l(I) < \epsilon$. The intervals may be open, closed or half-open, but we do *not* allow degenerate intervals consisting of only one point.

**1. Lemma (Vitali):** *Let $E$ be a set of finite outer measure and $\mathscr{I}$ a collection of intervals that cover $E$ in the sense of Vitali. Then, given $\epsilon > 0$, there is a finite disjoint collection $\{I_1, \ldots, I_N\}$ of intervals in $\mathscr{I}$ such that*

$$m^*\left[E \sim \bigcup_{n=1}^{N} I_n\right] < \epsilon.$$

***Proof:*** It suffices to prove the lemma in the case that each interval in $\mathscr{I}$ is closed, for otherwise we replace each interval by its closure and observe that the set of endpoints of $I_1, \ldots, I_N$ has measure zero.

Let $O$ be an open set of finite measure containing $E$. Since $\mathscr{I}$ is a Vitali covering of $E$, we may assume without loss of generality that each $I$ of $\mathscr{I}$ is contained in $O$. We choose a sequence $\langle I_n \rangle$ of disjoint intervals of $\mathscr{I}$ by induction as follows: Let $I_1$ be any interval in $\mathscr{I}$, and suppose $I_1, \ldots, I_n$ have already been chosen. Let $k_n$ be the supremum of the lengths of the intervals of $\mathscr{I}$ that do not meet any of the intervals $I_1, \ldots, I_n$. Since each $I$ is contained in $O$, we have $k_n \leq mO < \infty$. Unless $E \subset \bigcup_{i=1}^{n} I_i$, we can find $I_{n+1}$ in $\mathscr{I}$ with $l(I_{n+1}) > \frac{1}{2}k_n$ and $I_{n+1}$ disjoint from $I_1, \ldots, I_n$.

Thus we have a sequence $\langle I_n \rangle$ of disjoint intervals of $\mathscr{I}$, and since $\bigcup I_n \subset O$, we have $\sum l(I_n) \leq mO < \infty$. Hence we can find an integer $N$ such that

$$\sum_{N+1}^{\infty} l(I_n) < \frac{\epsilon}{5}.$$

Let

$$R = E \sim \bigcup_{n=1}^{N} I_n.$$

The lemma will be established if we can show that $m^*R < \epsilon$. Let $x$ be an arbitrary point of $R$. Since $\bigcup_{n=1}^{N} I_n$ is a closed set not containing $x$, we can find an interval $I$ in $\mathscr{I}$ which contains $x$ and whose length is so small that $I$ does not meet any of the intervals $I_1, \ldots, I_N$. If now $I \cap I_i = \emptyset$ for $i \leq n$, we must have $l(I) \leq k_n < 2l(I_{n+1})$. Since

$\lim l(I_n) = 0$, the interval $I$ must meet at least one of the intervals $I_n$. Let $n$ be the smallest integer such that $I$ meets $I_n$. We have $n > N$, and $l(I) \le k_{n-1} \le 2l(I_n)$. Since $x$ is in $I$, and $I$ has a point in common with $I_n$, it follows that the distance from $x$ to the midpoint of $I_n$ is at most $l(I) + \frac{1}{2}l(I_n) \le \frac{5}{2}l(I_n)$. Thus $x$ belongs to the interval $J_n$ having the same midpoint as $I_n$ and five times the length. Thus we have shown that

$$R \subset \bigcup_{N+1}^{\infty} J_n.$$

Hence

$$m^*R \le \sum_{N+1}^{\infty} l(J_n) = 5 \sum_{N+1}^{\infty} l(I_n) < \epsilon. \quad \blacksquare$$

In order to talk about the derivatives of a function $f$, we first define a set of four quantities called the **derivates** of $f$ at $x$ as follows:

$$D^+f(x) = \varlimsup_{h \to 0+} \frac{f(x+h) - f(x)}{h},$$

$$D^-f(x) = \varlimsup_{h \to 0+} \frac{f(x) - f(x-h)}{h},$$

$$D_+f(x) = \varliminf_{h \to 0+} \frac{f(x+h) - f(x)}{h},$$

$$D_-f(x) = \varliminf_{h \to 0+} \frac{f(x) - f(x-h)}{h}.$$

Clearly, we have $D^+f(x) \ge D_+f(x)$ and $D^-f(x) \ge D_-f(x)$. If $D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x) \ne \pm\infty$, we say that $f$ is differentiable at $x$ and define $f'(x)$ to be the common value of the derivates at $x$. If $D^+f(x) = D_+f(x)$, we say that $f$ has a right-hand derivative at $x$, and define $f'(x+)$ to be their common value. Similarly for $f'(x-)$.

**2. Proposition:** *If $f$ is continuous on $[a, b]$ and one of its derivates (say $D^+$) is everywhere nonnegative on $(a, b)$, then $f$ is nondecreasing on $[a, b]$*; i.e., $f(x) \le f(y)$ for $x \le y$.

**3. Theorem:** *Let $f$ be an increasing real-valued function on the interval $[a, b]$. Then $f$ is differentiable almost everywhere. The derivative $f'$ is measurable, and*

$$\int_a^b f'(x)\,dx \leq f(b) - f(a).$$

***Proof:*** Let us show that the sets where any two derivates are unequal have measure zero. We consider only the set $E$ where $D^+f(x) > D_-f(x)$, the sets arising from other combinations of derivates being similarly handled. Now the set $E$ is the union of the sets

$$E_{u,v} = \{x\colon D^+f(x) > u > v > D_-f(x)\}$$

for all rational $u$ and $v$. Hence it suffices to prove that $m^*E_{u,v} = 0$. Let $s = m^*E_{u,v}$ and, choosing $\epsilon > 0$, enclose $E_{u,v}$ in an open set $O$ with $mO < s + \epsilon$. For each point $x$ in $E_{u,v}$, there is an arbitrarily small interval $[x - h, x]$ contained in $O$ such that

$$f(x) - f(x - h) < vh.$$

By Lemma 1 we can choose a finite collection $\{I_1, \ldots, I_N\}$ of them whose interiors cover a subset $A$ of $E_{u,v}$ of outer measure greater than $s - \epsilon$. Then, summing over these intervals, we have

$$\begin{aligned}\sum_{n=1}^{N} [f(x_n) - f(x_n - h_n)] &< v \sum_{n=1}^{N} h_n \\ &< vmO \\ &< v(s + \epsilon).\end{aligned}$$

Now each point $y \varepsilon A$ is the left endpoint of an arbitrarily small interval $(y, y + k)$ that is contained in some $I_n$ and for which $f(y + k) - f(y) > uk$. Using Lemma 1 again, we can pick out a finite collection $\{J_1, \ldots, J_M\}$ of such intervals such that their union contains a subset of $A$ of outer measure greater than $s - 2\epsilon$. Then summing over these intervals yields

$$\begin{aligned}\sum_{i=1}^{M} f(y_i + k_i) - f(y_i) &> u \sum k_i \\ &> u(s - 2\epsilon).\end{aligned}$$

Each interval $J_i$ is contained in some interval $I_n$, and if we sum over those $i$ for which $J_i \subset I_n$, we have

$$\sum f(y_i + k_i) - f(y_i) \leq f(x_n) - f(x_n - h_n),$$

since $f$ is increasing. Thus

$$\sum_{n=1}^{N} f(x_n) - f(x_n - h_n) \geq \sum_{i=1}^{M} f(y_i + k_i) - f(y_i),$$

and so

$$v(s + \epsilon) > u(s - 2\epsilon).$$

Since this is true for each positive $\epsilon$, we have $vs \geq us$. But $u > v$, and so $s$ must be zero.

This shows that

$$g(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

is defined almost everywhere and that $f$ is differentiable wherever $g$ is finite. Let

$$g_n(x) = n[f(x + 1/n) - f(x)],$$

where we set $f(x) = f(b)$ for $x \geq b$. Then $g_n(x) \to g(x)$ for almost all $x$, and so $g$ is measurable. Since $f$ is increasing, we have $g_n \geq 0$. Hence by Fatou's Lemma

$$\begin{aligned}\int_a^b g \leq \underline{\lim} \int_a^b g_n &= \underline{\lim}\, n \int_a^b [f(x + 1/n) - f(x)]\, dx \\ &= \underline{\lim} \left[ n \int_b^{b+1/n} f - n \int_a^{a+1/n} f \right] \\ &= \underline{\lim} \left[ f(b) - n \int_a^{a+1/n} f \right] \\ &\leq f(b) - f(a).\end{aligned}$$

This shows that $g$ is integrable and hence finite almost everywhere. Thus $f$ is differentiable a.e. and $g = f'$ a.e. ∎

## Problems

**1.** Let $f$ be the function defined by $f(0) = 0$ and $f(x) = x \sin(1/x)$ for $x \neq 0$. Find $D^+f(0)$, $D_+f(0)$, $D^-f(0)$, and $D_-f(0)$.

**2. a.** Show that $D^+[-f(x)] = -D_+f(x)$.

**b.** If $g(x) = f(-x)$, then $D^+g(x) = -D_-f(-x)$.

**3. a.** If $f$ is continuous on $[a, b]$ and assumes a local maximum at $c \in (a, b)$, then

$$D_-f(c) \leq D^-f(c) \leq 0 \leq D_+f(c) \leq D^+f(c).$$

**b.** What if $f$ has a local maximum at $a$ or $b$?

**4.** Prove Proposition 2. [Hint: First show this for a function $g$ for which $D^+g \geq \varepsilon > 0$. Apply this to the function $g(x) = f(x) + \epsilon x$.]

**5. a.** Show that $D^+(f + g) \leq D^+f + D^+g$.

**b.** State and prove similar inequalities for the other derivates.

**c.** Let $f$ and $g$ be nonnegative and continuous at $c$. Then

$$D^+(f \cdot g)(c) \leq f(c)D^+g(c) + g(c)D^+f(c).$$

**6.** Let $f$ be defined on $[a, b]$ and $g$ a continuous function on $[\alpha, \beta]$ that is differentiable at $\gamma$ with $g(\gamma) = c \in (a, b)$.

**a.** If $g'(\gamma) > 0$, then $D^+(f \circ g)(\gamma) = D^+f(c) \cdot g'(\gamma)$.

**b.** If $g'(\gamma) < 0$, then $D^+(f \circ g)(\gamma) = D_-f(c) \cdot g'(\gamma)$.

**c.** If $g'(\gamma) = 0$ and all derivates of $f$ are finite at $c$, then

$$D^+(f \circ g)(\gamma) = 0.$$

## 2 Functions of Bounded Variation

Let $f$ be a real-valued function defined on the interval $[a, b]$, and let $a = x_0 < x_1 < \cdots < x_k = b$ be any subdivision of $[a, b]$. Define

$$p = \sum_{i=1}^{k} [f(x_i) - f(x_{i-1})]^+$$

$$n = \sum_{i=1}^{k} [f(x_i) - f(x_{i-1})]^-$$

$$t = n + p = \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})|,$$

where we use $r^+$ to denote $r$, if $r \geq 0$ and $0$, if $r \leq 0$, and set $r^- = |r| - r^+$. We have $f(b) - f(a) = p - n$. Set

$$P = \sup p,$$

$$N = \sup n,$$

$$T = \sup t,$$

where we take the suprema over all possible subdivisions of $[a, b]$.

We clearly have $P \leq T \leq P + N$. We call $P$, $N$, $T$ the positive, negative, and total variations of $f$ over $[a, b]$. We sometimes write $T_a^b$, $T_a^b(f)$, etc., to denote the dependence on the interval $[a, b]$ or on the function $f$. If $T < \infty$, we say that $f$ is of **bounded variation** over $[a, b]$. This notion is sometimes abbreviated by writing $f \in BV$.

**4. Lemma:** *If $f$ is of bounded variation on $[a, b]$, then*

$$T_a^b = P_a^b + N_a^b$$

*and*

$$f(b) - f(a) = P_a^b - N_a^b.$$

***Proof:*** For any subdivision of $[a, b]$

$$p = n + f(b) - f(a),$$

and taking suprema over all possible subdivisions, we obtain

$$P = N + f(b) - f(a).$$

Also

$$t = p + n = p + p - \{f(b) - f(a)\}$$

Taking suprema, we obtain

$$T = 2P - \{f(b) - f(a)\} = P + N \quad \blacksquare$$

**5. Theorem:** *A function $f$ is of bounded variation on $[a, b]$ if and only if $f$ is the difference of two monotone real-valued functions on $[a, b]$.*

***Proof:*** Let $f$ be of bounded variation, and set $g(x) = P_a^x$ and $h(x) = N_a^x$. Then $g$ and $h$ are monotone increasing functions which are real valued, since $0 \leq P_a^x \leq T_a^x \leq T_a^b < \infty$ and $0 \leq N_a^x \leq T_a^x \leq T_a^b < \infty$. But $f(x) = g(x) - h(x) + f(a)$ by Lemma 4. Since $h - f(a)$ is a monotone function, we have $f$ expressed as the difference of two monotone functions.

On the other hand, if $f = g - h$ on $[a, b]$ with $g$ and $h$ increasing, then for any subdivision we have

$$\begin{aligned}\sum |f(x_i) - f(x_{i-1})| &\leq \sum [g(x_i) - g(x_{i-1})] + \sum [h(x_i) - h(x_{i-1})] \\ &= g(b) - g(a) + h(b) - h(a).\end{aligned}$$

Hence

$$T_a^b(f) \le g(b) + h(b) - g(a) - h(a). \quad \blacksquare$$

**6. Corollary:** *If $f$ is of bounded variation on $[a, b]$, then $f'(x)$ exists for almost all $x$ in $[a, b]$.*

## Problems

**7. a.** Let $f$ be of bounded variation on $[a, b]$. Show that for each $c \in (a, b)$ the limit of $f(x)$ exists as $x \to c-$ and also as $x \to c+$. Prove that a monotone function (and hence a function of bounded variation) can have only a countable number of discontinuities. [Hint: If $f$ is monotone, the number of points where $|f(c+) - f(c-)| > 1/n$ is finite.]

**b.** Construct a monotone function on $[0, 1]$ which is discontinuous at each rational point.

**8. a** Show that if $a \le c \le b$, then $T_a^b = T_a^c + T_c^b$ and that hence $T_a^c \le T_a^b$.

**b.** Show that $T_a^b(f + g) \le T_a^b(f) + T_a^b(g)$, and $T_a^b(cf) = |c| T_a^b(f)$.

**9.** Let $\langle f_n \rangle$ be a sequence of functions on $[a, b]$ that converges at each point of $[a, b]$ to a function $f$. Then $T_a^b(f) \le \underline{\lim}\, T_a^b(f_n)$.

**10. a.** Let $f$ be defined by $f(0) = 0$ and $f(x) = x^2 \sin(1/x^2)$ for $x \ne 0$. Is $f$ of bounded variation on $[-1, 1]$?

**b.** Let $g$ be defined by $g(0) = 0$ and $g(x) = x^2 \sin(1/x)$ for $x \ne 0$. Is $g$ of bounded variation on $[-1, 1]$?

**11.** Let $f$ be of bounded variation on $[a, b]$. Show that

$$\int_a^b |f'| \le T_a^b(f).$$

## 3 Differentiation of an Integral

If $f$ is an integrable function on $[a, b]$, we define its indefinite integral to be the function $F$ defined on $[a, b]$ by

$$F(x) = \int_a^x f(t)\, dt.$$

In this section we show that the derivative of the indefinite integral of an integrable function is equal to the integrand almost everywhere. We begin by establishing some lemmas.

**7. Lemma:** *If $f$ is integrable on $[a, b]$, then the function $F$ defined by*

$$F(x) = \int_a^x f(t)\,dt$$

*is a continuous function of bounded variation on $[a, b]$.*

***Proof:*** The continuity follows from Proposition 4.14. To show that $F$ is of bounded variation, let $a = x_0 < x_1 < \cdots < x_k = b$ be any subdivision of $[a, b]$. Then

$$\sum_{i=1}^{k} |F(x_i) - F(x_{i-1})| = \sum_{i=1}^{k} \left| \int_{x_{i-1}}^{x_i} f(t)\,dt \right| \le \sum_{i=1}^{k} \int_{x_{i-1}}^{x_i} |f(t)|\,dt$$
$$= \int_a^b |f(t)|\,dt.$$

Thus

$$T_a^b(F) \le \int_a^b |f(t)|\,dt < \infty. \quad \blacksquare$$

**8. Lemma:** *If $f$ is integrable on $[a, b]$ and*

$$\int_a^x f(t)\,dt = 0$$

*for all $x \in [a, b]$, then $f(t) = 0$ a.e. in $[a, b]$.*

***Proof:*** Suppose $f(x) > 0$ on a set $E$ of positive measure. Then by Proposition 3.15 there is a closed set $F \subset E$ with $mF > 0$. Let $O = (a, b) \sim F$. Then either $\int_a^b f \neq 0$, or else

$$0 = \int_a^b f = \int_F f + \int_O f,$$

and

$$\int_O f = -\int_F f \neq 0.$$

But $O$ is the disjoint union of a countable collection $\{(a_n, b_n)\}$ of open intervals, and so by Proposition 4.12

$$\int_O f = \sum \int_{a_n}^{b_n} f.$$

Thus for some $n$ we have

$$\int_{a_n}^{b_n} f \neq 0,$$

and so either

$$\int_a^{a_n} f \neq 0$$

or

$$\int_a^{b_n} f \neq 0.$$

In any case we see that if $f$ is positive on a set of positive measure, then for some $x \varepsilon [a, b]$ we have

$$\int_a^x f \neq 0.$$

Similarly for $f$ negative on a set of positive measure, and the lemma follows by contraposition. ∎

**9. Lemma:** *If $f$ is bounded and measurable on $[a, b]$ and*

$$F(x) = \int_a^x f(t)\, dt + F(a),$$

*then $F'(x) = f(x)$ for almost all $x$ in $[a, b]$.*

***Proof:*** By Lemma 7, $F$ is of bounded variation over $[a, b]$, and so $F'(x)$ exists for almost all $x$ in $[a, b]$. Let $|f| \leq K$. Then setting

$$f_n(x) = \frac{F(x+h) - F(x)}{h},$$

with $h = 1/n$, we have

$$f_n(x) = \frac{1}{h} \int_x^{x+h} f(t)\, dt,$$

and so

$$|f_n| \leq K.$$

Since

$$f_n(x) \to F'(x) \quad \text{a.e.},$$

the bounded convergence theorem implies that

$$\int_a^c F'(x)\,dx = \lim \int_a^c f_n(x)\,dx = \lim_{h\to 0} \frac{1}{h}\int_a^c (F(x+h) - F(x))\,dx$$

$$= \lim \left[\frac{1}{h}\int_c^{c+h} F(x)\,dx - \frac{1}{h}\int_a^{a+h} F(x)\,dx\right]$$

$$= F(c) - F(a) = \int_a^c f(x)\,dx,$$

since $F$ is continuous. Hence

$$\int_a^c \{F'(x) - f(x)\}\,dx = 0$$

for all $c \in [a, b]$, and so

$$F'(x) = f(x) \quad \text{a.e.}$$

by Lemma 8. ▮

**10. Theorem:** *Let $f$ be an integrable function on $[a, b]$, and suppose that*

$$F(x) = F(a) + \int_a^x f(t)\,dt.$$

*Then $F'(x) = f(x)$ for almost all $x$ in $[a, b]$.*

***Proof:*** Without loss of generality we may assume $f \geq 0$. Let $f_n$ be defined by $f_n(x) = f(x)$, if $f(x) \leq n$, and $f_n(x) = n$ if $f(x) > n$. Then $f - f_n \geq 0$, and so

$$G_n(x) = \int_a^x f - f_n$$

is an increasing function of $x$, which must have a derivative almost everywhere, and this derivative will be nonnegative. Now by Lemma 9

$$\frac{d}{dx}\int_a^x f_n = f_n(x) \quad \text{a.e.},$$

and so

$$F'(x) = \frac{d}{dx} G_n + \frac{d}{dx}\int_a^x f_n$$
$$\geq f_n(x) \quad \text{a.e.}$$

Since $n$ is arbitrary,

$$F'(x) \geq f(x) \quad \text{a.e.}$$

Consequently,

$$\int_a^b F'(x)\,dx \geq \int_a^b f(x)\,dx = F(b) - F(a).$$

Thus by Theorem 3 we have

$$\int_a^b F'(x)\,dx = F(b) - F(a) = \int_a^b f(x)\,dx$$

and

$$\int_a^b (F'(x) - f(x))\,dx = 0.$$

Since $F'(x) - f(x) \geq 0$, this implies that $F'(x) - f(x) = 0$ a.e., and so $F'(x) = f(x)$ a.e. ▮

## 4 Absolute Continuity

A real-valued function $f$ defined on $[a, b]$ is said to be **absolutely continuous** on $[a, b]$ if, given $\epsilon > 0$, there is a $\delta > 0$ such that

$$\sum_{i=1}^{n} |f(x_i') - f(x_i)| < \epsilon$$

for every finite collection $\{(x_i, x_i')\}$ of nonoverlapping intervals with

$$\sum_{i=1}^{n} |x_i' - x_i| < \delta.$$

An absolutely continuous function is continuous, and it follows from Proposition 4.14 that every indefinite integral is absolutely continuous. Note that the sum and difference of two absolutely continuous functions is absolutely continuous.

**11. Lemma:** *If $f$ is absolutely continuous on $[a, b]$, then it is of bounded variation on $[a, b]$.*

***Proof:*** Let $\delta$ be the $\delta$ in the definition of absolute continuity that corresponds to $\epsilon = 1$. Then any subdivision of $[a, b]$ can be split (by inserting fresh division points, if necessary) into $K$ sets of intervals, each of total length less than $\delta$, where $K$ is the largest integer less than $1 + (b - a)/\delta$. Thus for any subdivision we have $t \le K$, and so $T \le K$. ▮

**12. Corollary:** *If $f$ is absolutely continuous, then $f$ has a derivative almost everywhere.*

**13. Lemma:** *If $f$ is absolutely continuous on $[a, b]$ and $f'(x) = 0$* a.e., *then $f$ is constant.*

***Proof:*** We wish to show that $f(a) = f(c)$ for any $c \in [a, b]$. Let $E \subset (a, c)$ be the set of measure $c - a$ in which $f'(x) = 0$, and let $\epsilon$ and $\eta$ be arbitrary positive numbers. To each $x$ in $E$ there is an arbitrarily small interval $[x, x + h]$ contained in $[a, c]$ such that $|f(x + h) - f(x)| < \eta h$. By Lemma 1 we can find a finite collection $\{[x_k, y_k]\}$ of nonoverlapping intervals of this sort which cover all of $E$ except for a set of measure less than $\delta$, where $\delta$ is the positive number corresponding to $\epsilon$ in the definition of the absolute continuity of $f$. If we label the $x_k$ so that $x_k \le x_{k+1}$, we have

$$y_0 = a \le x_1 < y_1 \le x_2 < \cdots \le y_n \le c = x_{n+1}$$

and

$$\sum_{k=0}^{n} |x_{k+1} - y_k| < \delta.$$

Now

$$\sum_{k=1}^{n} |f(y_k) - f(x_k)| \le \eta \sum (y_k - x_k)$$
$$< \eta(c - a)$$

by the way the intervals $\{[x_k, y_k]\}$ were constructed, and

$$\sum_{k=0}^{n} |f(x_{k+1}) - f(y_k)| < \epsilon$$

by the absolute continuity of $f$. Thus

$$|f(c) - f(a)| = \left| \sum_{k=0}^{n} [f(x_{k+1}) - f(y_k)] + \sum_{k=1}^{n} [f(y_k) - f(x_k)] \right|$$
$$\le \epsilon + \eta(c - a).$$

Since $\epsilon$ and $\eta$ are arbitrary positive numbers, $f(c) - f(a) = 0$. ▮

**14. Theorem:** *A function $F$ is an indefinite integral if and only if it is absolutely continuous.*

***Proof:*** If $F$ is an indefinite integral, then $F$ is absolutely continuous by Proposition 4.14. Suppose on the other hand that $F$ is absolutely continuous on $[a, b]$. Then $F$ is of bounded variation, and we may write

$$F(x) = F_1(x) - F_2(x),$$

where the functions $F_i$ are monotone increasing. Hence $F'(x)$ exists almost everywhere and

$$|F'(x)| \leq F_1'(x) + F_2'(x).$$

Thus

$$\int |F'(x)|\, dx \leq F_1(b) + F_2(b) - F_1(a) - F_2(a)$$

by Theorem 3, and $F'(x)$ is integrable. Let

$$G(x) = \int_a^x F'(t)\, dt.$$

Then $G$ is absolutely continuous and so is the function $f = F - G$. It follows from Theorem 10 that $f'(x) = F'(x) - G'(x) = 0$ a.e., and so $f$ is constant by Lemma 13. Thus

$$F(x) = \int_a^x F'(t)\, dt + F(a). \quad \blacksquare$$

**15. Corollary:** *Every absolutely continuous function is the indefinite integral of its derivative.*

## Problems

**12.** Let $f$ be absolutely continuous in the interval $[\epsilon, 1]$ for each $\epsilon > 0$. Does the continuity of $f$ at 0 imply that $f$ is absolutely continuous on $[0, 1]$? What if $f$ is also of bounded variation on $[0, 1]$?

**13.** Let $f$ be absolutely continuous on $[a, b]$. Show that

$$T_a^b(f) = \int_a^b |f'| \quad \text{and} \quad P_a^b(f) = \int_a^b [f']^+.$$

**14. a.** Show that the sum and difference of two absolutely continuous functions are also absolutely continuous.

**b.** Show that the product of two absolutely continuous functions is absolutely continuous. [Hint: Make use of the fact that they are bounded.]

**c.** If $f$ is absolutely continuous on $[a, b]$ and if $f$ is never zero there, then the function $g = 1/f$ is also absolutely continuous on $[a, b]$.

**15.** The Cantor ternary function (Problem 2.48) is continuous and monotone but not absolutely continuous.

**16.** A monotone function $f$ on $[a, b]$ is called **singular** if $f' = 0$ a.e.

**a.** Show that any monotone increasing function is the sum of an absolutely continuous function and a singular function.

**b.** Let $f$ be a nondecreasing singular function on $[a, b]$. Then $f$ has the following property: ($S$) Given $\epsilon > 0$, $\delta > 0$, there is a finite collection $\{[y_k, x_k]\}$ of nonoverlapping intervals such that

$$\sum |x_k - y_k| < \delta$$

and

$$\sum f(x_k) - f(y_k) > f(b) - f(a) - \epsilon.$$

[Hint: See proof of Lemma 13.]

**c.** Let $f$ be a nondecreasing function on $[a, b]$ with property ($S$) of part (b). Then $f$ is singular. [Hint: Use part (a).]

**d.** Let $\langle f_n \rangle$ be a sequence of nondecreasing singular functions on $[a, b]$ such that the function

$$f(x) = \sum f_n(x)$$

is everywhere finite. Then $f$ is also singular.

**e.** Show that there is a strictly increasing singular function on $[0, 1]$.

**17. a.** Let $F$ be absolutely continuous on $[c, d]$ and $g$ be absolutely continuous with $c \leq g \leq d$ on $[a, b]$. Then $F \circ g$ is absolutely continuous on $[a, b]$.

**b.** Let $E = \{x: g'(x) = 0\}$. Then $m(g[E]) = 0$.

**18.** Let $g$ be an absolutely continuous monotone function on $[0, 1]$ and $E$ a set of measure zero. Then $g[E]$ has measure zero.

**19. a.** Construct an absolutely continuous strictly monotone function $g$ on $[0, 1]$ such that $g' = 0$ on a set of positive measure. [Hint: Let $G$ be the complement of a generalized Cantor set of positive measure (Problem 3.14), and let $g$ be the indefinite integral of $\chi_G$.]

**b.** Show that there is a set $E$ of measure zero such that $g^{-1}[E]$ is not measurable. How does this example compare with that of Problem 3.28?

**20.** A function $f$ is said to satisfy a Lipschitz condition on an interval if there is a constant $M$ such that $|f(x) - f(y)| \leq M|x - y|$ for all $x$ and $y$ in the interval.

**a.** Show that a function satisfying a Lipschitz condition is absolutely continuous.

**b.** Show that an absolutely continuous function $f$ satisfies a Lipschitz condition if and only if $|f'|$ is bounded.

**c.** Show that $f$ satisfies a Lipschitz condition if one of its derivates (say $D^+$) is bounded.

**21.** *Change of variable.* Let $g$ be a monotone increasing absolutely continuous function on $[a, b]$ with $g(a) = c$, $g(b) = d$.

**a.** Show that for any open set $O \subset [c, d]$

$$mO = \int_{g^{-1}[O]} g'(x)\, dx.$$

**b.** Let $H = \{x\colon g'(x) \neq 0\}$. If $E$ is a subset of $[c, d]$ with $mE = 0$, then $g^{-1}(E) \cap H$ has measure zero.

**c.** If $E$ is a measurable subset of $[c, d]$, then $F = g^{-1}[E] \cap H$ is measurable and

$$mE = \int_F g' = \int_a^b \chi_E(g(x))g'(x)\, dx.$$

**d.** If $f$ is a nonnegative measurable function on $[c, d]$, then $(f \circ g)\, g'$ is measurable on $[a, b]$ and

$$\int_c^d f(y)\, dy = \int_a^b f(g(x))g'(x)\, dx.$$

**22.** *Change of variable II.* Let $g$ be a monotone increasing absolutely continuous function on $[a, b]$ with $g(a) = c$, $g(b) = d$, and let $f$ be an integrable function on $[c, d]$. Let

$$F(y) = \int_c^y f(t)\, dt,$$

and set $H(x) = F(g(x))$.

**a.** Show that $H$ is absolutely continuous (Problem 17a) and that $F'(g(x))$ exists wherever $H'$ and $g'$ exist and $g'(x) \neq 0$. Thus

$$H'(x) = F'(g(x))g'(x)$$

almost everywhere except on the set $E$ where $g'(x) = 0$.

**b.** Let

$$f_0(y) = \begin{cases} f(y) & y \notin g[E] \\ 0 & y \in g[E]. \end{cases}$$

Then $f_0 = f$ a.e. (See Problem 17.b). Hence

$$H'(x) = f_0(g(x))g'(x) \text{ a.e.}$$

**c.** Show that

$$\int_c^d F(y)\,dy = \int_a^b F(g(x))g'(x)\,dx.$$

**d.** Is the expression in part (*c*) still true if $g$ is not assumed to be monotone? Assume that $g$ is bounded and that $f$ is integrable on an interval containing the range of $E$.

## 5 Convex Functions

A function $\varphi$ defined on an open interval $(a, b)$ is said to be *convex* if for each $x, y \in (a, b)$ and each $\lambda$, $0 \le \lambda \le 1$ we have

$$\varphi(\lambda x + (1 - \lambda)y) \le \lambda\varphi(x) + (1 - \lambda)\varphi(y).$$

If we look at the graph of $\varphi$ in $\mathbf{R}^2$, this condition can be formulated geometrically by saying that each point on the chord between $\langle x, \varphi(x)\rangle$ and $\langle y, \varphi(y)\rangle$ is above the graph of $\varphi$. An important property of the chords of a convex function is given by the following lemma, whose proof is left to the reader.

**16. Lemma:** *If $\varphi$ is convex on $(a, b)$ and if $x, y, x', y'$ are points of $(a, b)$ with $x \le x' < y'$ and $x < y \le y'$, then the chord over $(x', y')$ has larger slope than the chord over $(x, y)$; that is,*

$$\frac{\varphi(y) - \varphi(x)}{y - x} \le \frac{\varphi(y') - \varphi(x')}{y' - x'}.$$

If the upper and lower left-hand derivates $D^- f$ and $D_- f$ of a function $f$ are equal and finite at a point $x$, we say that $f$ is differentiable on the left at $x$ and call this common value the left-hand derivative at $x$. Similarly, we say that $f$ is differentiable on the right at $x$ if $D^+ f$ and $D_+ f$ are equal there. Some of the continuity and differentiability properties of convex functions are given by the following proposition.

**17. Proposition:** *If $\varphi$ is convex on $(a, b)$, then $\varphi$ is absolutely continuous on each closed subinterval of $(a, b)$. The right- and left-hand derivatives of $\varphi$ exist at each point of $(a, b)$ and are equal to each other* *except on a countable set. The left- and right-hand derivatives are monotone increasing functions, and at each point the left-hand derivative is less than or equal to the right-hand derivative.*

***Proof:*** Let $[c, d] \subset (a, b)$. Then, by Lemma 16, we have

$$\frac{\varphi(c) - \varphi(a)}{c - a} \leq \frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(b) - \varphi(d)}{b - d}$$

for $x, y$ in $[c, d]$. Thus $|\varphi(y) - \varphi(x)| \leq M|x - y|$ in $[c, d]$, and so $\varphi$ is absolutely continuous there (cf. Problem 20).

If $x_0 \in (a, b)$, then $[\varphi(x) - \varphi(x_0)]/(x - x_0)$ is an increasing function of $x$ by Lemma 16, and so the limits as $x$ approaches $x_0$ from the right and from the left exist and are finite. Thus $\varphi$ is differentiable on the right and on the left at each point, and the left-hand derivative is less than or equal to the right-hand derivative. If $x_0 < y_0$, $x < y_0$, and $x_0 < y$, then

$$\frac{\varphi(x) - \varphi(x_0)}{x - x_0} \leq \frac{\varphi(y) - \varphi(y_0)}{y - y_0},$$

and either derivative at $x_0$ is less than or equal to either derivative at $y_0$. Consequently, each derivative is monotone, and they are equal at a point if one of them is continuous there. Since a monotone function can have only a countable number of discontinuities, they are equal except on a countable set. ▮

The following proposition is a partial converse of the previous proposition.

**18. Proposition:** *If $\varphi$ is a continuous function on $(a, b)$ and if one derivate (say $D^+$) of $\varphi$ is nondecreasing, then $\varphi$ is convex.*

***Proof:*** Given $x$, $y$ with $a < x < y < b$, define a function $\psi$ on $[0, 1]$ by

$$\psi(t) = \varphi[ty + (1 - t)x] - t\varphi(y) - (1 - t)\varphi(x).$$

Our goal is to show that $\psi$ is nonpositive on $[0, 1]$. Now $\psi$ is continuous, and $\psi(0) = \psi(1) = 0$. Moreover,

$$D^+\psi = (y - x)D^+\varphi - \varphi(y) + \varphi(x),$$

and so $D^+\psi$ is nondecreasing on $[0, 1)$.

Let $\gamma$ be a point where $\psi$ assumes its maximum on $[0, 1]$. If $\gamma = 1$, then $\psi(t) \leq \psi(1) = 0$ on $[0, 1]$. Hence suppose that $\gamma \in [0, 1)$.

Since $\psi$ has a local maximum at $\gamma$, we have $D^+\psi(\gamma) \le 0$. But $D^+\psi$ was nondecreasing, and so $D^+\psi \le 0$ on $[0, \gamma]$. Consequently, $\psi$ is nonincreasing on $[0, \gamma]$, and hence $\psi(\gamma) \le \psi(0) = 0$. Thus the maximum of $\psi$ on $[0, 1]$ is nonpositive, and so $\psi \le 0$ on $[0, 1]$. ▮

**19. Corollary:** *Let $\varphi$ have a second derivative at each point of $(a, b)$. Then $\varphi$ is convex on $(a, b)$ if and only if $\varphi''(x) \ge 0$ for each $x \in (a, b)$.*

Let $\varphi$ be a convex function on $(a, b)$ and $x_0 \in (a, b)$. The line $y = m(x - x_0) + \varphi(x_0)$ through $\langle x_0, \varphi(x_0)\rangle$ is called a *supporting line* at $x_0$ if it always lies below the graph of $\varphi$, that is, if

$$\varphi(x) \ge m(x - x_0) + \varphi(x_0).$$

It follows from Lemma 16 that such a line is a supporting line if and only if its slope $m$ lies between the left- and right-hand derivatives at $x_0$. Thus, in particular, there is always at least one supporting line at each point. This notion enables us to give a short proof for the following proposition:

**20. Proposition (Jensen Inequality):** *Let $\varphi$ be a convex function on $(-\infty, \infty)$ and $f$ an integrable function on* $[0, 1]$. *Then*

$$\int \varphi(f(t))\,dt \ge \varphi\left[\int f(t)\,dt\right].$$

***Proof:*** Let $\alpha = \int f(t)\,dt$, and let $y = m(x - \alpha) + \varphi(\alpha)$ be the equation of a supporting line at $\alpha$. Then

$$\varphi(f(t)) \ge m(f(t) - \alpha) + \varphi(\alpha).$$

Integrating both sides with respect to $t$ gives the proposition. ▮

This inequality has a geometric interpretation worth mentioning. Since the point $\lambda x_1 + (1 - \lambda)x_2$ is the centroid of masses $\lambda$ and $(1 - \lambda)$ at $x_1$ and $x_2$, we can say that a function is convex if its value at the centroid of a two-point mass is less than the weighted average of its values at the two points. The Jensen inequality is a generalization of this fact: If we define a mass distribution $\mu$ on the line by setting $\mu(a, b] = m\{t\colon a < f(t) \le b\}$, then $\int f(t)\,dt$ is the centroid of this mass and $\int \varphi(f(t))\,dt = \int \varphi(x)\,d\mu$ is the weighted average of $\varphi$.

An important application of the Jensen inequality is obtained by taking for our convex function $\varphi$ the function exp defined by

$\exp x = e^x$. The inequality then becomes a generalization of the inequality between the arithmetic and geometric mean:

**21. Corollary:** *Let $f$ be an integrable function on* $[0, 1]$. *Then*

$$\int \exp(f(t))\,dt \ge \exp\left[\int f(t)\,dt\right].$$

We conclude this section with two further definitions. We say that a function $\varphi$ is strictly convex if

$$\varphi(\lambda x + (1-\lambda)y) < \lambda\varphi(x) + (1-\lambda)\varphi(y)$$

for all $x, y \in (a, b)$ and all $\lambda \in (0, 1)$.

We sometimes say that a function $\varphi$ is *concave* to mean that $-\varphi$ is convex. The only functions that are both convex and concave are the linear functions.

If $I$ is any interval, open, closed, or half-open, we say that $\varphi$ is convex on $I$ if $\varphi$ is continuous on $I$ and convex in the interior. See Problem 23c for an important property of such functions.

## Problems

**23. a.** Let $\varphi$ be a convex function on a finite interval $[a, b)$. Then $\varphi$ is bounded from below.

**b.** Show that if $\varphi$ is convex on $(a, b)$, then $\varphi(x)$ has a limit (possibly infinite) as $x$ approaches $a$ (or $b$) from within $(a, b)$. If $a$ (or $b$) is finite, then the limit at $a$ (or $b$) may be $+\infty$ but not $-\infty$.

**c.** Let $\varphi$ be continuous on an interval $I$ (open, closed, half-open) and convex on the interior of $I$. Then we have

$$\varphi(tx + (1-t)y) \le t\varphi(x) + (1-t)\varphi(y)$$

for all $x, y \in I$ and all $t \in [0, 1]$.

**24.** Prove Corollary 19.

**25. a.** Suppose $a \ge 0$ and $b > 0$. Then the function $\varphi(t) = (a + bt)^p$ is convex on $[0, \infty)$ for $1 \le p < \infty$ and concave for $0 < p \le 1$.

**b.** Show that $\varphi$ is strictly convex for $p > 1$ and strictly concave for $0 < p < 1$.

**26.** When do we have equality in Corollary 21?

**27.** Let $\langle \alpha_n \rangle$ be a sequence of nonnegative numbers whose sum is 1 and $\langle \xi_n \rangle$ a sequence of positive numbers. Then

$$\prod_{n=1}^{\infty} \xi_n^{\alpha_n} \leq \sum_{n=1}^{\infty} \alpha_n \xi_n .$$

**28.** Let $g$ be a nonnegative measurable function on $[0, 1]$. Then $\log \int g(t)\,dt \geq \int \log (g(t))\,dt$ whenever the right side is defined.

# 6 The Classical Banach Spaces

## 1 The $L^p$ Spaces

In this chapter we study some spaces of functions of a real variable. Let $p$ be a positive real number. A measurable function defined on $[0, 1]$ is said to belong to the space $L^p = L^p[0, 1]$ if $\int_0^1 |f|^p < \infty$. Thus $L^1$ consists precisely of the Lebesgue integrable functions on $[0, 1]$. Since $|f+g|^p \leq 2^p(|f|^p + |g|^p)$, we see that the sum of two functions in $L^p$ is again in $L^p$. Since $\alpha f$ is in $L^p$ whenever $f$ is, we have $\alpha f + \beta g$ in $L^p$ whenever $f$ and $g$ are. A space $X$ of real-valued functions is called a *linear space* (or vector space) if it has the property that $\alpha f + \beta g$ belongs to $X$ for each pair $f$ and $g$ belonging to $X$ and for each pair of constants $\alpha$ and $\beta$. Thus the $L^p$ spaces are linear spaces.

For a function $f \in L^p$, we define

$$\|f\| = \|f\|_p = \left\{\int_0^1 |f|^p\right\}^{1/p}.$$

We see that $\|f\| = 0$ if and only if $f = 0$ a.e. If $\alpha$ is a constant, then $\|\alpha f\| = |\alpha|\,\|f\|$. In the next section we derive two inequalities, the first of which states that $\|f+g\| \leq \|f\| + \|g\|$, if $p \geq 1$. This property is called subadditivity. Henceforth, we shall assume that $p \geq 1$, unless otherwise specified. A linear space is said to be a normed linear space if we have assigned a nonnegative real number $\|f\|$ to each $f$ such that $\|\alpha f\| = |\alpha|\,\|f\|$, $\|f+g\| \leq \|f\| + \|g\|$, and with $\|f\| = 0 \Leftrightarrow f \equiv 0$. Unfortunately, norms for the $L^p$ spaces do not satisfy the last requirement, for from $\|f\| = 0$ we can only conclude that $f = 0$ a.e. We shall, however, consider two measurable functions to be equivalent if they are equal almost everywhere; and, if we do not distinguish between equivalent functions, then the $L^p$ spaces are normed linear spaces.[1]

It is convenient to denote by $L^\infty$ the space of all bounded measurable functions on [0, 1] (or rather all measurable functions which are bounded except possibly on a subset of measure zero). Again we identify functions which are equivalent. Then $L^\infty$ is a linear space, and it becomes a normed linear space if we define

$$\|f\| = \|f\|_\infty = \text{ess sup } |f(t)|,$$

where ess sup $f(t)$ is the infimum of sup $g(t)$ as $g$ ranges over all functions which are equal to $f$ almost everywhere. Thus

$$\text{ess sup } f(t) = \inf \{M: m\{t: f(t) > M\} = 0\}.$$

## Problems

1. Show that $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.
2. Let $f$ be a bounded measurable function on [0, 1]. Then $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$.
3. Prove that $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$.
4. If $f \varepsilon L^1$ and $g \varepsilon L^\infty$, then

$$\int |fg| \leq \|f\|_1 \cdot \|g\|_\infty.$$

## 2 The Minkowski and Hölder Inequalities

The purpose of this section is to establish several inequalities involving the norm $\|\cdot\|_p$ in the $L^p$ spaces. The first is the subadditivity of the norm when $1 \leq p \leq \infty$.

[1] To be pedantic, we should say that the elements of $L^p$ are not functions but, rather, equivalence classes of functions (cf. Problem 10.10).

**1. Minkowski Inequality:** *If $f$ and $g$ are in $L^p$ with $1 \le p \le \infty$, then so is $f + g$ and*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p.$$

*If $1 < p < \infty$, then equality can hold only if there are nonnegative constants $\alpha$ and $\beta$ such that $\beta f = \alpha g$.*

***Proof:*** The case when $p = \infty$ is elementary (Problem 1), as are the cases when $\|f\| = 0$ or $\|g\| = 0$. Thus we assume that $1 \le p < \infty$ and $\|f\| = \alpha \ne 0$, $\|g\| = \beta \ne 0$. Then there are functions $f_0$ and $g_0$ such that $|f| = \alpha f_0$, $|g| = \beta g_0$, and $\|f_0\| = \|g_0\| = 1$. Set $\lambda = \alpha/(\alpha + \beta)$. Then $(1 - \lambda) = \beta/(\alpha + \beta)$, and we have

$$\begin{aligned}|f(x) + g(x)|^p \le (|f(x)| + |g(x)|)^p &= [\alpha f_0(x) + \beta g_0(x)]^p \\ &= (\alpha + \beta)^p[\lambda f_0(x) + (1 - \lambda)g_0(x)]^p \\ &\le (\alpha + \beta)^p[\lambda f_0(x)^p + (1 - \lambda)g_0(x)^p]\end{aligned}$$

by the convexity of the function $\varphi(t) = t^p$ on $[0, \infty)$ for $1 \le p < \infty$ (cf. Problem 5.25). If $1 < p < \infty$, this inequality is strict unless $f_0(x) = g_0(x)$ and $\operatorname{sgn} f(x) = \operatorname{sgn} g(x)$. Integrating both sides of this inequality gives

$$\begin{aligned}\|f + g\|^p &\le (\alpha + \beta)^p[\lambda\|f_0\|^p + (1 - \lambda)\|g_0\|^p] \\ &\le (\alpha + \beta)^p = (\|f\| + \|g\|)^p.\end{aligned}$$

Taking $p$-th roots gives

$$\|f + g\| \le \|f\| + \|g\|.$$

If $1 < p < \infty$, the inequality is strict unless $f_0 = g_0$ a.e. and $\operatorname{sgn} f = \operatorname{sgn} g$ a.e. But this means that the inequality is strict unless $\beta f = \alpha g$. ▮

Note that the function $\varphi(t) = t^p$ is concave on $[0, \infty)$ for $0 < p < 1$. Hence the proof above gives, *mutatis mutandis*, the following inequality:

**2. Minkowski Inequality for $0 < p < 1$:** *Let $f$ and $g$ be two nonnegative functions which belong to the space $L^p$ with $0 < p < 1$. Then*

$$\|f + g\| \ge \|f\| + \|g\|.$$

We next establish a differential form of the Minkowski inequality. We begin with the following lemma.

**3. Lemma:** *Let $1 \le p < \infty$. Then for a, b, t nonnegative we have*

$$(a + tb)^p \ge a^p + ptba^{p-1}.$$

***Proof:*** Set

$$\varphi(t) = (a + tb)^p - a^p - ptba^{p-1}.$$

Then $\varphi(0) = 0$, and

$$\varphi'(t) = pb[(a + tb)^{p-1} - a^{p-1}] \ge 0$$

for $p \ge 1$ and $a, b, t \ge 0$. Thus $\varphi(t)$ is increasing and hence nonnegative for $t > 0$. ▮

**4. Hölder Inequality:** *If p and q are nonnegative extended real numbers such that*

$$\frac{1}{p} + \frac{1}{q} = 1,$$

*and if $f \in L^p$ and $g \in L^q$, then $f \cdot g \in L^1$ and*

$$\int |fg| \le \|f\|_p \cdot \|g\|_q.$$

*Equality holds if and only if for some constants $\alpha$ and $\beta$, not both zero, we have* $\alpha|f|^p = \beta|g|^q$ a.e.

***Proof:*** The case $p = 1$, $q = \infty$ is straightforward and is left to the reader. Hence we assume that $1 < p < \infty$ and consequently $1 < q < \infty$.

Replacing $f$ and $g$ with $|f|$ and $|g|$, if necessary, we need only to consider the case $f \ge 0$ and $g \ge 0$. Set

$$h(x) = g(x)^{q-1} = g(x)^{q/p},$$

since $q - 1 = q/p$. Also,

$$g(x) = h(x)^{p-1} = h(x)^{p/q}.$$

Thus

$$ptf(x)g(x) = ptf(x)h(x)^{p-1} \le (h(x) + tf(x))^p - h(x)^p.$$

Hence

$$pt \int fg \le \int |h + tf|^p - \int h^p = \|h + tf\|^p - \|h\|^p$$

and

$$pt \int fg \leq (\|h\| + t\|f\|)^p - \|h\|^p.$$

Differentiating both sides with respect to $t$ at $t = 0$, we get

$$p \int fg \leq p\|f\| \; \|h\|_p^{p-1} = p\|f\| \; \|g\|.$$

The condition for equality follows from the condition for equality in the Minkowski inequality. ∎

## Problems

5. a. Prove the Minkowski inequality for $0 < p < 1$.

b. Show that if $f \in L^p$, $g \in L^p$, then $f + g \in L^p$ even for $0 < p < 1$. [Hint: $\|f + g\|^p \leq 2^p(\|f\|^p + \|g\|^p)$.]

6. State and prove a version of the Hölder inequality for $0 < p < 1$.

7. a. For $1 \leq p < \infty$, we denote by $l^p$ the space of all sequences $\langle \xi_\nu \rangle_{\nu=1}^\infty$ such that $\sum_{\nu=1}^\infty |\xi_\nu|^p < \infty$. Prove the Minkowski inequality for sequences:

$$\|\langle \xi_\nu + \eta_\nu \rangle\|_p \leq \|\langle \xi_\nu \rangle\|_p + \|\langle \eta_\nu \rangle\|_p.$$

Here we have $1 \leq p \leq \infty$,

$$(\|\langle \xi_\nu \rangle\|_p)^p = \sum_{\nu=1}^\infty |\xi_\nu|^p$$

and

$$\|\langle \eta_\nu \rangle\|_\infty = \sup |\eta_\nu|.$$

b. Show that if $\langle \xi_\nu \rangle \in l^p$ and $\langle \eta_\nu \rangle \in l^q$ with

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then

$$\sum_{\nu=1}^\infty |\xi_\nu \eta_\nu| \leq \|\langle \xi_\nu \rangle\|_p \cdot \|\langle \eta_\nu \rangle\|_q.$$

This is the Hölder inequality for sequences.

**8. a.** Let $a$, $b$ be nonnegative, $1 < p < \infty$, $1/p + 1/q = 1$. Establish Young's inequality

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}.$$

[There are several conceptually different ways of doing this.]

**b.** Use Young's inequality to give a proof of the Hölder inequality different from that given in the text.

**c.** Show that the inequality sign in Young's inequality is reversed if $0 < p < 1$.

**d.** Establish a version of the Hölder inequality for $0 < p < 1$.

## 3 Convergence and Completeness

The notion of convergence for a sequence of real numbers generalizes to give us a notion of convergence for sequences in a normed linear space.

**Definition:** *A sequence $\langle f_n \rangle$ in a normed linear space is said to converge to an element $f$ in the space if given $\epsilon > 0$, there is an $N$ such that for all $n > N$ we have $\|f - f_n\| < \epsilon$. If $f_n$ converges to $f$, we write $f = \lim f_n$ or $f_n \to f$.*

Another way of formulating the convergence of $f_n$ to $f$ is by noting that $f_n \to f$ if $\|f_n - f\| \to 0$. Convergence in the space $L^p$, $1 \le p < \infty$, is often referred to as **convergence in the mean of order $p$.** Thus a sequence of functions $\langle f_n \rangle$ is said to converge to $f$ in the mean of order $p$ if each $f_n$ belongs to $L^p$ and $\|f - f_n\|_p \to 0$. Convergence in $L^\infty$ is nearly uniform convergence (Problem 10).

Since we are dealing with a linear space $X$ of functions, we must be careful to distinguish the above notion of convergence in $X$ from the notion of a sequence of functions which converges at each point. We shall call this latter type of convergence **pointwise convergence,** and say that $\langle f_n \rangle$ converges pointwise to $f$ if for each $x$ we have $f(x) = \lim f_n(x)$. If there is a set $E$ of measure zero such that for each $x$ in $\tilde{E}$ we have $f(x) = \lim f_n(x)$, then we say that $f_n$ converges to $f$ almost everywhere.

Just as for the case of sequences of real numbers, we say that a sequence $\langle f_n \rangle$ in a normed linear space is a **Cauchy sequence** if given $\epsilon > 0$, there is an $N$ such that for all $n \ge N$ and all $m \ge N$ we have

$\|f_n - f_m\| < \epsilon$. It is easily verified that each convergent sequence is a Cauchy sequence.

**Definition:** *A normed linear space is called* **complete** *if every Cauchy sequence in the space converges, that is, if for each Cauchy sequence $\langle f_n \rangle$ in the space there is an element $f$ in the space such that $f_n \to f$. A complete normed linear space is called a* **Banach space.**

A series $\langle f_n \rangle$ in a normed linear space is said to be **summable** to a sum $s$ if $s$ is in the space and the sequence of partial sums of the series converges to $s$; that is,

$$\left\| s - \sum_{i=1}^{n} f_i \right\| \to 0.$$

If this is the case, we write $s = \sum_{i=1}^{\infty} f_i$. The series $\langle f_n \rangle$ is said to be absolutely summable if $\sum_{n=1}^{\infty} \|f_n\| < \infty$.

For a series of real numbers we know that absolute summability implies that the series is summable. While this is not true in general for series of elements in a normed linear space, the following proposition shows that this implication holds if the space is complete.

**5. Proposition:** *A normed linear space $X$ is complete if and only if every absolutely summable series is summable.*

***Proof:*** $\Rightarrow$: Let $X$ be complete and $\langle f_n \rangle$ an absolutely summable series of elements of $X$. Since $\sum \|f_n\| = M < \infty$, there is, for each $\epsilon > 0$, an $N$ such that $\sum_{n=N}^{\infty} \|f_n\| < \epsilon$. Let $s_n = \sum_{i=1}^{n} f_i$ be the partial sum of the series $\langle f_n \rangle$. Then for $n \geq m \geq N$ we have

$$\|s_n - s_m\| = \left\| \sum_{i=m}^{n} f_i \right\| \leq \sum_{i=m}^{\infty} \|f_i\| \leq \sum_{i=N}^{\infty} \|f_i\| < \epsilon.$$

Hence the sequence $\langle s_n \rangle$ of partial sums is a Cauchy sequence and must converge to an element $s$ in $X$, since $X$ is complete.

$\Leftarrow$: Let $\langle f_n \rangle$ be a Cauchy sequence in $X$. For each integer $k$ there is an integer $n_k$ such that $\|f_n - f_m\| < 2^{-k}$ for all $n$ and $m$ greater than $n_k$, and we may choose the $n_k$'s so that $n_{k+1} > n_k$. Then $\langle f_{n_k} \rangle_{k=1}^{\infty}$ is a subsequence of $\langle f_n \rangle$, and if we then set $g_1 = f_{n_1}$, and

$g_k = f_{n_k} - f_{n_{k-1}}$ for $k > 1$ we obtain a series $\langle g_k \rangle$ whose $k$-th partial sum is $f_{n_k}$. But we have $\|g_k\| \leq 2^{-k+1}$ if $k > 1$. Thus

$$\sum \|g_k\| \leq \|g_1\| + \sum 2^{-k+1} = \|g_1\| + 1.$$

Hence the series $\langle g_k \rangle$ is absolutely summable, and so by our hypothesis there is an element $f$ in $X$ to which the partial sums of the series converge. Thus the subsequence $\langle f_{n_k} \rangle$ converges to $f$.

We shall now show that $f = \lim f_n$. Since $\langle f_n \rangle$ is a Cauchy sequence, given $\epsilon > 0$, there is an $N$ such that $\|f_n - f_m\| < \epsilon/2$ for all $n$ and $m$ larger than $N$. Since $f_{n_k} \to f$, there is a $K$ such that for all $k \geq K$ we have $\|f_{n_k} - f\| < \epsilon/2$. Let us take $k$ so large that $k > K$ and $n_k \geq N$. Then

$$\|f_n - f\| \leq \|f_n - f_{n_k}\| + \|f_{n_k} - f\| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus for all $n > N$ we have $\|f_n - f\| < \epsilon$, and so $f_n \to f$. ▮

**6. Theorem (Riesz–Fischer):** *The $L^p$ spaces are complete.*

***Proof:*** Since the case $p = \infty$ is elementary, it is left to the reader, and we assume $1 \leq p < \infty$. By virtue of the preceding proposition we need only show that each absolutely summable series in $L^p$ is summable in $L^p$ to some element of $L^p$.

Let $\langle f_n \rangle$ be a sequence in $L^p$ with $\sum_{n=1}^{\infty} \|f_n\| = M < \infty$, and define functions $g_n$ by setting $g_n(x) = \sum_{k=1}^{n} |f_k(x)|$. From the Minkowski inequality we have

$$\|g_n\| \leq \sum_{k=1}^{n} \|f_k\| \leq M.$$

Hence

$$\int (g_n)^p \leq M^p.$$

For each $x$, $\langle g_n(x) \rangle$ is an increasing sequence of (extended) real numbers and so must converge to an extended real number $g(x)$. The function $g$ so defined is measurable, and, since $g_n \geq 0$, we have

$$\int g^p \leq M^p$$

by Fatou's Lemma. Hence $g^p$ is integrable, and $g(x)$ is finite for almost all $x$.

For each $x$ such that $g(x)$ is finite the series $\sum_{k=1}^{\infty} f_k(x)$ is an absolutely summable series of real numbers and so must be summable to a real number $s(x)$. If we set $s(x) = 0$ for those $x$ where $g(x) = \infty$, we have defined a function $s$ which is the limit almost everywhere of the partial sums $s_n = \sum_{k=1}^{n} f_k$. Hence $s$ is measurable. Since $|s_n(x)| \le g(x)$, we have $|s(x)| \le g(x)$. Consequently, $s$ is in $L^p$ and we have

$$|s_n(x) - s(x)|^p \le 2^p[g(x)]^p.$$

Since $2^p g^p$ is integrable and $|s_n(x) - s(x)|^p$ converges to 0 for almost all $x$, we have

$$\int |s_n - s|^p \to 0$$

by the Lebesgue Convergence Theorem. Thus $\|s_n - s\|^p \to 0$, whence $\|s_n - s\| \to 0$. Consequently, the series $\langle f_n \rangle$ has in $L^p$ the sum $s$. ∎

## Problems

**9.** Prove that every convergent sequence is a Cauchy sequence.

**10.** Let $\langle f_n \rangle$ be a sequence of functions in $L^\infty$. Prove that $\langle f_n \rangle$ converges to $f$ in $L^\infty$ if and only if there is a set $E$ of measure zero such that $f_n$ converges to $f$ uniformly on $\tilde{E}$.

**11.** Prove that $L^\infty$ is complete.

**12.** Prove that $l^p$ is complete $(1 \le p < \infty)$ (see Problem 7).

**13.** Let $C = C[0, 1]$ be the space of all continuous functions on $[0, 1]$ and define $\|f\| = \max |f(x)|$. Show that $C$ is a Banach space.

**14.** We denote by $l^\infty$ the space of all bounded sequences of real numbers and define $\|\langle \xi_\nu \rangle\|_\infty = \sup |\xi_\nu|$. Show that $l^\infty$ is a Banach space.

**15.** Show that the space $c$ of all convergent sequences of real numbers and the space $c_0$ of all sequences which converge to zero are Banach spaces (with the $l^\infty$ norm).

**16.** Let $\langle f_n \rangle$ be a sequence of functions in $L^p$, $1 \le p < \infty$, which converge almost everywhere to a function $f$ in $L^p$. Show that $\langle f_n \rangle$ converges to $f$ in $L^p$ if and only if $\|f_n\| \to \|f\|$. (For $p = 1$ this is just Problem 4.14b.)

**17.** Let $\langle f_n \rangle$ be a sequence of functions in $L^p$, $1 < p < \infty$, which converge almost everywhere to a function $f$ in $L^p$, and suppose that there is a constant $M$ such that $\|f_n\| \le M$ for all $n$. Then for each function $g$ in $L^q$ we have

$$\int fg = \lim \int f_n g.$$

Is the result true for $p = 1$?

**18.** Let $f_n \to f$ in $L^p$, $1 \le p < \infty$, and let $\langle g_n \rangle$ be a sequence of measurable functions such that $|g_n| \le M$, all $n$, and $g_n \to g$ a.e. Then $g_n f_n \to gf$ in $L^p$.

## 4 Approximation in $L^p$

In this section we establish versions of Littlewood's second principle which says that every function $f$ in $L^p$, $1 \le p < \infty$ is 'nearly' a step function and 'nearly' a continuous function: That is, given $f$ and $\epsilon > 0$, there is a step function $\varphi$ and a continuous function $\psi$ with $\|f - \varphi\|_p < \epsilon$ and $\|f - \psi\|_p < \epsilon$.

If $\Delta = \{\xi_0, \ldots, \xi_n\}$ is a subdivision, $0 = \xi_0 < \xi_1 < \cdots < \xi_n = 1$, of $[0, 1]$, we define the step function $\varphi_\Delta$ by taking $\varphi_\Delta$ to be constant on each interval $[\xi_k, \xi_{k-1})$ of the subdivision and equal there to the average of $f$ over that interval. We will show that $\|f - \varphi_\Delta\| \to 0$ as the length $\delta$ of the largest subinterval of $\Delta$ goes to zero. That we can approximate $f$ by the step functions generated in this particular manner is useful in geometric applications.

We begin with some preliminary results.

**7. Lemma:** *Given $f \in L^p$, $1 \le p < \infty$, and $\epsilon > 0$, there is a bounded measurable function $f_M$ with $|f_M| \le M$ and $\|f - f_M\| < \epsilon$.*

***Proof:*** Let

$$f_N(x) = \begin{cases} N & N \le f(x) \\ f(x) & -N \le f(x) \le N \\ -N & f(x) \le -N. \end{cases}$$

Then $|f_N| \le N$, and $\langle f_N \rangle$ converges to $f$ almost everywhere, and so $|f - f_N|^p \to 0$ almost everywhere. Since

$$|f - f_N|^p \le |f|^p,$$

and $|f|^p$ is integrable, we have

$$\|f - f_N\|^p = \int |f - f_N|^p \to 0,$$

as $N \to \infty$. Thus $\|f - f_N\| \to 0$, and there is an $M$ with $\|f - f_M\| < \epsilon$. ▮

**8. Proposition:** *Given $f \in L^p$, $1 < p \leq \infty$ and $\epsilon > 0$, there is a step function $\varphi$ and a continuous function $\psi$ such that $\|f - \varphi\|_p < \epsilon$ and $\|f - \psi\|_p < \epsilon$.*

***Proof:*** By Lemma 7 we may choose a bounded function $f_M$ with $\|f - f_M\| < \epsilon/2$. By Proposition 3.22 we can find a step function $\varphi$ such that

$$|f_M - \varphi| < \frac{\epsilon}{4}$$

except on a set $E$ of measure less than $\delta$, where

$$\delta = \left(\frac{\epsilon}{4M}\right)^p.$$

Then

$$\begin{aligned}\|f_M - \varphi\|^p &= \int_0^1 |f_M - \varphi|^p \\ &= \int_{[0,1]\sim E} |f_M - \varphi|^p + \int_E |f_M - \varphi|^p \\ &< \frac{\epsilon^p}{4^p} + \frac{M^p\epsilon^p}{4^pM^p} \leq \frac{\epsilon^p}{2^p}.\end{aligned}$$

Consequently, $\|f_M - \varphi\| < \epsilon/2$, and so

$$\|f - \varphi\| \leq \|f - f_M\| + \|f_M - \varphi\| < \epsilon$$

by the Minkowski inequality.

The existence of $\psi$ follows from the fact that any step function $\varphi$ can be approximated in $L^p$ by a continuous function. ▮

**Definition:** *Let $\Delta = \{\xi_0, \xi_1, \ldots, \xi_m\}$ be a subdivision of the finite interval $[a, b]$* and *$f$ an integrable function on $[a, b]$. The function $\varphi_\Delta$ on $[a, b]$ defined by*

$$\varphi_\Delta(x) = \frac{1}{\xi_{k+1} - \xi_k} \int_{\xi_k}^{\xi_{k+1}} f(t)\, dt$$

*is called the $\Delta$-approximant to $f$ in mean.*

**9. Proposition:** *Let $f \varepsilon L^p$. Then the $\Delta$-approximant $\varphi_\Delta$ to $f$ converges to $f$ in $L^p$, i.e., $\|f - \varphi_\Delta\| \to 0$, as the length $\delta$ of the longest subinterval in $\Delta$ approaches zero.*

***Proof:*** For any $\Delta$ and any $f \varepsilon L^p$ we denote the $\Delta$-approximant to $f$ by $T_\Delta(f)$. Then $T_\Delta$ is a mapping from $L^p$ to the space of step functions whose set of discontinuities are the division points of $\Delta$. It is readily verified that

$$T_\Delta(f + g) = T_\Delta(f) + T_\Delta(g)$$
$$T_\Delta(\alpha f) = \alpha T_\Delta(f)$$

and

$$\| T_\Delta(f) \|_p \leq \| f \|_p .$$

By Proposition 8, there is, given $\epsilon > 0$, a step function $\varphi$ such that $\| f - \varphi \|_p < \epsilon/3$. Now $T_\Delta(\varphi)$ differs from $\varphi$ only on those subintervals of $\Delta$ that contain points of discontinuity of $\varphi$. Let $\varphi$ have $l$ points of discontinuity and let $M = \max |\varphi(x)|$. Then

$$\| \varphi - T(\varphi) \|_p^p = \int | \varphi - T(\varphi)|^p \leq \delta l (2M)^p,$$

and so

$$\| \varphi - T_\Delta \varphi \|_p \leq 2M(\delta l)^{1/p},$$

where $\delta$ is the length of the longest subinterval of $\Delta$. By the Minkowski inequality

$$\begin{aligned} \| f - T_\Delta f \| &\leq \| f - \varphi \| + \| \varphi - T_\Delta \varphi \| + \| T_\Delta(\varphi - f) \| \\ &\leq 2 \| f - \varphi \| + \| \varphi - T_\Delta \varphi \| \\ &< \frac{2}{3} \epsilon + 2M(\delta l)^{1/p} \\ &< \epsilon \end{aligned}$$

whenever

$$\delta < \frac{1}{(2M)^p l} . \quad \blacksquare$$

**10. Corollary:** *The $\Delta$-approximant $\varphi_\Delta$ converges to $f$ in measure as $\delta \to 0$.*

## Problems

**19.** Show that $\|T_\Delta f\|_p \le \|f\|_p$.

**20.** Prove Corollary 10.

## 5 Bounded Linear Functionals on the $L^p$ Spaces

We define a **linear functional** on a normed linear space $X$ to be a mapping $F$ of the space $X$ into the set of real numbers *such that* $F(\alpha f + \beta g) = \alpha F(f) + \beta F(g)$. We say that the linear functional is **bounded** if there is a constant $M$ such that $|F(f)| \le M \cdot \|f\|$ for all $f$ in $X$. The smallest constant $M$ for which this inequality is true is called the norm of $F$. Thus

$$\|F\| = \sup \frac{|F(f)|}{\|f\|},$$

as $f$ ranges over all nonzero elements of $X$.

If $g$ is a function in $L^q$, we can define a bounded linear functional $F$ on $L^p$ by setting

$$F(f) = \int fg.$$

The functional $F$ is clearly linear, and the Hölder inequality states that $\|F\| \le \|g\|_q$. In fact, we actually have $\|F\| = \|g\|_q$. To see this for the case $1 < p < \infty$, we set[2]

$$f = |g|^{q/p} \operatorname{sgn} g.$$

Then $|f|^p = |g|^q = fg$. Hence $f$ is in $L^p$ and $\|f\|_p = (\|g\|_q)^{q/p}$. Now

$$F(f) = \int fg = \int |g|^q$$
$$= (\|g\|_q)^q = \|g\|_q \|f\|_p,$$

and so $\|F\|$ must be at least as great as $\|g\|_q$. We state this result as a proposition; the cases $p = 1$ and $p = \infty$ are left to the reader (see Problem 21).

[2] For any real number $x$ we define $\operatorname{sgn} x = 1$ if $x > 0$, $\operatorname{sgn} 0 = 0$, and $\operatorname{sgn} x = -1$ if $x < 0$.

**11. Proposition:** *Each function g in $L^q$ defines a bounded linear functional F on $L^p$ by*

$$F(f) = \int fg.$$

*We have* $\|F\| = \|g\|_q$.

The goal of the present section is to show that for $1 \le p < \infty$ the converse of this proposition holds, that is, that we obtain every bounded linear functional on $L^p$ in this manner. We shall find it useful to first establish the following lemma.

**12. Lemma:** *Let g be an integrable function on* $[0, 1]$, *and suppose that there is a constant M such that*

$$\left|\int fg\right| \le M\|f\|_p$$

*for all bounded measurable functions f. Then g is in $L^q$, and* $\|g\|_q \le M$.

***Proof:*** Let us first assume $1 < p < \infty$. We define a sequence of bounded measurable functions by setting

$$g_n(x) = \begin{cases} g(x) & \text{if } |g(x)| \le n \\ 0 & \text{if } |g(x)| > n, \end{cases}$$

and letting

$$f_n = |g_n|^{q/p} \operatorname{sgn} g_n.$$

Now $\|f_n\|_p = (\|g_n\|_q)^{q/p}$, and $|g_n|^q = f_n \cdot g_n = f_n \cdot g$. Hence

$$(\|g_n\|_q)^q = \int f_n g \le M\|f_n\|_p = M(\|g_n\|_q)^{q/p}.$$

Since $q - q/p = 1$,

$$\|g_n\|_q \le M$$

and

$$\int |g_n|^q \le M^q.$$

Since $|g_n|^q$ converges to $|g|^q$ almost everywhere, we have

$$\int |g|^q \le \lim \int |g_n|^q \le M^q$$

by Fatou's Lemma. Thus $g \in L^q$, and $\|g\|_q \le M$.

For the case $p = 1$, let $E = \{x\colon |g(x)| \geq M + \epsilon\}$, and set $f = (\operatorname{sgn} g)\chi_E$. Then $\|f\|_1 = mE$, and

$$MmE = M\|f\|_1 \geq \left|\int fg\right| \geq (M + \epsilon)mE.$$

Thus $mE = 0$, and $\|g\|_\infty \leq M$. ∎

We are now in a position to give the following characterization of the bounded linear functionals on $L^p$ for $1 \leq p < \infty$:

**13. Riesz Representation Theorem:** *Let $F$ be a bounded linear functional on $L^p$, $1 \leq p < \infty$. Then there is a function $g$ in $L^q$ such that*

$$F(f) = \int fg.$$

*We also have $\|F\| = \|g\|_q$.*

***Proof:*** Let $\chi_s$ be the characteristic function of the interval $[0, s]$. We begin our investigation of $F$ by observing what it does to $\chi_s$. For each $s$ the value of $F(\chi_s)$ is a real number $\Phi(s)$, and this defines a function $\Phi$ on $[0, 1]$. Now I maintain that $\Phi$ is absolutely continuous. For let $\{(s_i, s_i')\}$ be any finite collection of nonoverlapping subintervals of $[0, 1]$ of total length less than $\delta$. Then

$$\sum_i |\Phi(s_i') - \Phi(s_i)| = F(f),$$

where

$$f = \sum_i (\chi_{s'_i} - \chi_{s_i}) \operatorname{sgn} (\Phi(s_i') - \Phi(s_i)).$$

Since $(\|f\|_p)^p < \delta$, we have

$$\begin{aligned} \sum_i |\Phi(s_i') - \Phi(s_i)| &= F(f) \\ &\leq \|F\|\,\|f\|_p \\ &< \|F\|\delta^{1/p}. \end{aligned}$$

This shows that the total variation of $\Phi$ is less than $\epsilon$ over any finite collection of disjoint intervals of total length $\delta$ if we take $\delta = \epsilon^p/\|F\|^p$. Thus $\Phi$ is absolutely continuous.

By Theorem 5.14 there is an integrable function $g$ on [0, 1] such that

$$\Phi(s) = \int_0^s g.$$

Thus

$$F(\chi_s) = \int_0^1 g \cdot \chi_s.$$

Since every step function on [0, 1] is (equal except at a finite number of points to) a suitable linear combination $\sum c_i \chi_{s_i}$, we must have

$$F(\psi) = \int_0^1 g\psi$$

for each step function $\psi$, by the linearity of $F$ and of the integral.

Let $f$ be any bounded measurable function on [0, 1]. Then it follows from Proposition 3.22 that there is a bounded sequence $\langle \psi_n \rangle$ of step functions which converge almost everywhere to $f$. Since the sequence $\langle |f - \psi_n|^p \rangle$ is uniformly bounded and tends to zero almost everywhere, the bounded convergence theorem implies that $\|f - \psi_n\|_p \to 0$. Since $F$ is bounded and

$$\begin{aligned} |F(f) - F(\psi_n)| &= |F(f - \psi_n)| \\ &\le \|F\| \, \|f - \psi_n\|_p, \end{aligned}$$

we must have

$$F(f) = \lim F(\psi_n).$$

Since $g\psi_n$ is always less than $|g|$ times the uniform bound for the sequence $\langle \psi_n \rangle$, we have

$$\int fg = \lim \int g\psi_n$$

by the Lebesgue Convergence Theorem. Consequently, we must have

$$\int fg = F(f)$$

for each bounded measurable function $f$. Since $|F(f)| \le \|F\| \, \|f\|_p$, we have $g$ in $L^q$ and $\|g\|_q \le \|F\|$ by Lemma 12.

Thus we have only to show that $F(f) = \int fg$ for each $f$ in $L^p$. Let $f$ be an arbitrary function in $L^p$. Then by Proposition 8 there is for each $\epsilon > 0$ a step function $\psi$ such that $\|f - \psi\|_p < \epsilon$. Since $\psi$ is bounded we have

$$F(\psi) = \int \psi g.$$

Hence

$$\begin{aligned} |F(f) - \int fg| &= |F(f) - F(\psi) + \int \psi g - \int fg| \\ &\leq |F(f - \psi)| + |\int (\psi - f)g| \\ &\leq \|F\| \; \|f - \psi\|_p + \|g\|_q \|f - \psi\|_p \\ &< (\|F\| + \|g\|_q)\epsilon. \end{aligned}$$

Since $\epsilon$ is an arbitrary number, we must have

$$F(f) = \int fg.$$

The equality $\|F\| = \|g\|_q$ follows from Proposition 11. ∎

In the problems the reader is asked to carry out a similar representation for the bounded linear functionals on $l^p$, $1 \leq p < \infty$, $c$, and $c_0$. In Theorem 13.23 we give a representation for the bounded linear functionals on $C$. Unfortunately, the bounded linear functionals on $L^\infty$ (and on $l^\infty$) do not admit of a similar representation.

## Problems

**21. a.** Let $g$ be an integrable function on [0, 1]. Show that there is a bounded measurable function $f$ such that $\|f\| \neq 0$ and

$$\int fg = \|g\|_1 \cdot \|f\|_\infty.$$

**b.** Let $g$ be a bounded measurable function. Show that for each $\epsilon > 0$ there is an integrable function $f$ such that

$$\int fg \geq (\|g\|_\infty - \epsilon)\|f\|_1.$$

[Hint: $f$ may be taken to be a suitable characteristic function.]

**22.** Find a representation for the bounded linear functionals on $l^p$, $1 \leq p < \infty$.

**23.** Find a representation for the bounded linear functionals on $c$ and on $c_0$. (*Caveat:* These representations are different.)

**24.** Show that the *element* $g$ in $L^q$ given by Theorem 13 is unique.

# Part Two

## ABSTRACT SPACES

# 7 Metric Spaces

## 1 Introduction

The system of real numbers has two types of properties. The first type consists of the algebraic, dealing with addition, multiplication, etc. The other type consists of properties having to do with the notion of distance between two numbers and with the concept of a limit. The latter properties are called topological or metric, and the object of the present chapter is to study these properties in a general space in which the notion of distance is defined. We make the following definition:

**Definition:** *A **metric space** $\langle X, \rho\rangle$ is a nonempty set $X$ of elements (which we call points) together with a real-valued function $\rho$ defined on $X \times X$ such 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 function $\rho$ is called a **metric.***

An obvious example of a metric space is the set $\mathbf{R}$ of all numbers with $\rho(x, y) = |x - y|$. A second example is the $n$-dimensional Euclidean space $\mathbf{R}^n$ whose points are the $n$-tuples $x = \langle x_1, \ldots, x_n\rangle$ of real numbers and

$$\rho(x, y) = [(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2]^{1/2}.$$

For $\mathbf{R}^n$, property (iv) of the metric is merely the statement that the length of a side of a triangle is less than the sum of the lengths of the other two sides. Consequently, (iv) is usually called the triangle inequality.

Other examples of metric spaces are the normed linear spaces of the last chapter with

$$\rho(x, y) = \|x - y\|,$$

the triangle inequality being equivalent to $\|x + y\| \le \|x\| + \|y\|$.

It should be emphasized that a metric space is not the set $X$ of its points, since it is in fact the pair $\langle X, \rho\rangle$ consisting of the set of its points together with the metric $\rho$. For example, the set of all $n$-tuples of real numbers can also be made into a metric space by use of the metric $\rho^*$ given by

$$\rho^*(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|,$$

and this is not the metric space $\mathbf{R}^n$ (if $n > 1$). Often we are interested in only one metric 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 metric space $\langle X, \rho\rangle$.

If we have two metric spaces $\langle X, \rho\rangle$ and $\langle Y, \sigma\rangle$, we can form a new metric space called the **Cartesian product** $X \times Y$ whose set of points is the set $X \times Y = \{\langle x, y\rangle : x \in X, y \in Y\}$ and whose metric $\tau$ is given by

$$\tau(\langle x_1, y_1\rangle, \langle x_2, y_2\rangle) = [\rho(x_1, x_2)^2 + \sigma(y_1, y_2)^2]^{1/2}.$$

It is readily verified that $\tau$ has all the properties required of a metric and that $\mathbf{R}^m \times \mathbf{R}^n = \mathbf{R}^{m+n}$. Other metrics for $X \times Y$ are discussed in Section 5.

If $E$ is any nonempty subset of a metric space $(X, \rho)$, we define the **diameter** of $E$ to be the extended real number $\sup \{\rho(x, y) : x, y \in E\}$.

Any nonempty subset of a metric space is itself a metric space if we restrict the metric to it. For example, the space $C$ of the last chapter is a subspace of $L^\infty$.

It is sometimes convenient to relax the condition that $\rho(x, y) = 0$ only if $x = y$. When we allow the possibility that $\rho(x, y) = 0$ for some $x \neq y$, we call $\rho$ a **pseudometric.** Thus the $L^p$ norms are pseudometrics on the spaces of measurable functions whose $p$-th powers are integrable. It is also convenient on occasions to allow $\rho$ to assume the value $+\infty$. In this case $\rho$ is called an **extended metric** (or extended pseudometric). Some properties and examples of pseudometrics and extended metrics can be found in Problems 3, 8, and 11.

## Problems

**1. a.** Show that the set of all $n$-tuples of real numbers becomes a metric space under each of the following metrics:

$$\rho^*(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|$$

$$\rho^+(x, y) = \max \{|x_1 - y_1|, \ldots, |x_n - y_n|\}.$$

**b.** For $n = 2$ and $n = 3$ describe the sets $\{x: \rho(x, y) < 1\}$, $\{x: \rho^*(x, y) < 1\}$, and $\{x: \rho^+(x, y) < 1\}$.

**2.** By the **ball** (or spheroid) centered at $x$ and having radius $\delta$ we mean the set

$$S_{x,\delta} = \{y: \rho(x, y) < \delta\}.$$

Prove that if $0 < \epsilon < \delta - \rho(x, z)$, then $S_{z,\epsilon} \subset S_{x,\delta}$.

**3. a.** *Pseudometrics.* A pair $\langle X, \rho \rangle$ is called a *pseudometric space* if $\rho$ is a pseudometric. Show that $\rho(x, y) = 0$ is an equivalence relation and that if $X^*$ is the set of equivalence classes under this relation, then $\rho(x, y)$ depends only on the equivalence classes of $x$ and $y$ and defines a metric on $X^*$.

**b.** *Extended metrics.* Let $\rho$ be an extended metric on a set $X$. Show that the relation $\rho(x, y) < \infty$ is an equivalence relation, i.e., that it is reflexive, symmetric, and transitive. The equivalence classes $X_\alpha$ of $X$ under this metric are sometimes called parts of the extended metric space $(X, \rho)$. Show that each part is both open and closed.

## 2 Open and Closed Sets

We shall find that a number of the properties of sets of real numbers apply immediately to sets in a metric space. Throughout the present section all sets mentioned are subsets of a given metric space $\langle X, \rho \rangle$. The following propositions and definitions correspond to those in Section 5 of Chapter 2, and the reader is asked to check that the proofs given there are valid for metric spaces.

**Definition:** *A set $O$ is called* ***open*** *if for every $x \in O$, $\exists \delta > 0$ such that all $y$ with $\rho(x, y) < \delta$ belong to $O$.*

**1. Proposition:** *The set $X$ and $\emptyset$ are open; the intersection of any two open sets is open; and the union of any collection of open sets is open.*

**Definition:** *A point $x \in X$ is called a* **point of closure** *of the set $E$ if for every $\delta > 0$ there is a point $y \in E$ such that $\rho(x, y) < \delta$. We use $\bar{E}$ to denote the set of points of closure of $E$. Clearly, $E \subset \bar{E}$.*

**2. Proposition:** *If $A \subset B$, then $\bar{A} \subset \bar{B}$. Also, $\overline{(A \cup B)} = \bar{A} \cup \bar{B}$, and $\overline{(A \cap B)} \subset \bar{A} \cap \bar{B}$.*

**Definition:** *A set $F$ is called* **closed** *if $\bar{F} = F$.*

**3. Proposition:** *The closure $\bar{E}$ of any set $E$ is closed; that is, $\bar{E} = \bar{\bar{E}}$.*

**4. Proposition:** *The sets $\emptyset$ and $X$ are closed; the union of any two closed sets is closed; and the intersection of any collection of closed sets is closed.*

**5. Proposition:** *The complement of an open set is closed; the complement of a closed set is open.*

**Definition:** *A metric space $X$ is called* **separable** *if it has a subset $D$ which has a countable number of points and which is dense in $X$, that is, for which $\bar{D} = X$.*

Since the set of rational numbers is a countable dense subset of **R**, we see that **R** is separable. The following proposition shows that the Lindelöf theorem holds for a metric space if and only if it is separable.

**6. Proposition:** *A metric space $X$ is separable if and only if there is a countable family $\{O_i\}$ of open sets such that for any* open *set $O \subset X$,*

$$O = \bigcup_{O_i \subset O} O_i.$$

***Proof:*** If $X$ is separable, let $D$ be a countable dense set. By the ball at $x$ with radius $\delta$ we mean the set

$$S_{x,\delta} = \{y: \rho(x, y) < \delta\}.$$

Let $\{O_i\}$ consist of those balls $S_{x,\delta}$ for which $x$ is in $D$ and $\delta$ is rational. Then $\{O_i\}$ is a countable collection of open sets. If $O$ is any open set and $y \in O$, then we want to show that for some $O_i$ we have $y \in O_i \subset O$. Since $O$ is open, there is a ball $S_{y,\delta}$ such that $S_{y,\delta} \subset O$. By taking $\delta$ even smaller, we may assume $\delta$ is rational. Since $y$ is a point of closure of $D$, there is a point $x \in D$ such that $\rho(x, y) < \delta/2$. Hence

$$S_{x,\delta/2} \subset S_{y,\delta} \subset O.$$

But $S_{x,\delta/2}$ is one of the $\{O_i\}$, and the "only if" part of the theorem is proved.

Suppose, on the other hand, we are given the countable collection $\{O_i\}$. Let $x_i$ be a point of $O_i$, and let $D$ be the set of all these points $x_i$. We shall now see that $D$ is dense. Let $x$ be any point of $X$ and $S$ any spheroid centered at $x$. Then we must show that $S$ contains a point of $D$. But $S$ is an open set (Problem 6), and so we must have some $O_i$ so $x \in O_i \subset S$. Hence $x_i \in S$, and we see that $x \in \bar{D}$. ∎

By a **neighborhood** of a point $x$ in a metric space we shall mean a ball (or more generally an open set) containing $x$.

## Problems

**4.** **a.** Show that $C$ is a closed subset of $L^\infty$ (see Chapter 6).

**b.** Show that the set of all integrable functions that vanish for $0 \le t < \frac{1}{2}$ is a closed subset of $L^1$.

**c.** Show that the set of all measurable functions $x(t)$ with $\int x < 1$ is an open subset of $L^1$.

**5.** Show that $\bar{E} = \bigcap_{E \subset F} F$ for closed sets $F$. The *interior* $E^\circ$ of a set $E$ is the set of those $y \in E$ for which there is a $\delta > 0$ such that $\rho(y, z) < \delta \Rightarrow z \in E$.

**a.** Show that

$$E^\circ = \bigcup_{O \subset E} O.$$

**b.** Show that

$$\sim(\bar{E}) = (\sim E^\circ).$$

**6.** **a.** Show that each ball is open.

**b.** Show that the sets $\{x : \rho(x, y) \le \delta\}$ are closed.

c. Is the set in (b) always the closure of the ball

$$\{x: \rho(x, y) < \delta\}?$$

7. Which of the spaces $\mathbf{R}^n$, $C$, $L^\infty$, $L^1$ are separable?

## 3 Continuous Functions and Homeomorphisms

A function $f$ on a metric space $\langle X, \rho\rangle$ into a metric space $\langle Y, \sigma\rangle$ is a rule that associates to each $x \varepsilon X$ a unique $y \varepsilon Y$. We also call $f$ a mapping of $X$ into $Y$ and shall use the terms function and mapping interchangeably. By a mapping $f$ of $X$ **onto** $Y$ we mean, as usual, that for any $y \varepsilon Y$ there is some $x \varepsilon X$ such that $f(x) = y$; that is to say, that $Y$ is the range of $f$. Then as in Section 6 of Chapter 2, we have the following definition and propositions:

**Definition:** *The function $f$ is said to be continuous at $x$ if, for every $\epsilon > 0$, there is a $\delta > 0$ so that if $\rho(x, y) < \delta$, then $\sigma[f(x), f(y)] < \epsilon$. The function $f$ is called* **continuous** *if it is continuous at each $x \varepsilon X$.*

7. **Proposition:** *A function $f$ from a metric space $X$ to a metric space $Y$ is continuous if and only if for each open set $O$ in $Y$ the set $f^{-1}[O]$ is an open set in $X$.*

8. **Proposition:** *If $f$ is a continuous mapping from $X$ to $Y$ and if $g$ is a continuous mapping from $Y$ to $Z$, then the mapping $g \circ f$ from $X$ to $Z$ is also continuous.*

A one-to-one mapping $f$ of $X$ **onto** $Y$ is called a **homeomorphism** between $X$ and $Y$ if $f$ is continuous and the mapping $f^{-1}$ inverse to $f$ is also continuous. The spaces $X$ and $Y$ are said to be homeomorphic if there is a homeomorphism between them. The study of topology is essentially the study of those properties which are unaltered by homeomorphisms, and such properties are called topological. By virtue of Proposition 7 a one-to-one correspondence between $X$ and $Y$ is a homeomorphism if and only if it makes open sets in $X$ correspond to open sets in $Y$ and conversely. Thus the property of being an open subset of a space is a topological one. Since a closed set is the complement of an open set, it follows that being a closed subset of $X$ is a topological property. In fact, every property that can be defined by means of open sets is a topological one, and hence so is the property of being a continuous function: that is, if $f$ is a continuous function on $X$ and $h: X \to Y$ is a homeomorphism between $X$ and $Y$, then $f \circ h^{-1}$ is a continuous function on $Y$.

Not all properties in a metric space are preserved under a homeomorphism, however. For example, the distance between two points is usually altered by a homeomorphism. A homeomorphism that leaves distance unchanged, that is, one for which

$$\sigma[h(x_1), h(x_2)] = \rho(x_1, x_2)$$

for all $x_1$ and $x_2$ in $X$, is called an **isometry** between $X$ and $Y$. The spaces $X$ and $Y$ are called isometric if there is an isometry between them. From an abstract point of view two isometric metric spaces are exactly the same, an isometry amounting merely to a relabeling of the points. A notion that arises naturally in this connection is that of equivalent metrics: Two metrics $p$ and $\sigma$ on the same set $X$ of points are called **equivalent** if the identity mapping of $\langle X, \rho\rangle$ onto $\langle X, \sigma\rangle$ is a homeomorphism. Thus two metrics are equivalent if and only if they define the same open sets, that is, if a set is open with respect to one whenever it is open with respect to the other.

## Problems

**8.** Show that the function $h$ on $[0, 1)$ given by $h(x) = x/(1 - x)$ is a homeomorphism between $[0, 1)$ and $[0, \infty)$.

**9.** Let $E$ be a set and $x$ a point in a metric space. Define

$$\rho(x, E) = \inf_{y \varepsilon E} \rho(x, y).$$

**a.** Show that for a fixed $E$ the function $f$ given by $f(x) = \rho(x, E)$ is continuous.

**b.** Show that $\{x: \rho(x, E) = 0\} = \bar{E}$.

**10. a.** Prove that two metrics on a set $X$ are equivalent if and only if given $x \varepsilon X$ and $\epsilon > 0$ there is a $\delta > 0$ such that for all $y \varepsilon X$

$$\rho(x, y) < \delta \Rightarrow \sigma(x, y) < \epsilon$$

and

$$\sigma(x, y) < \delta \Rightarrow \rho(x, y) < \epsilon.$$

**b.** Show that the following set of metrics for the set of $n$-tuples of real numbers are equivalent:

$$\rho(x, y) = [(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2]^{1/2}$$
$$\rho^*(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|$$
$$\rho^+(x, y) = \max \{|x_1 - y_1|, \ldots, |x_n - y_n|\}.$$

**c.** Find a metric for the set of $n$-tuples that is not equivalent to these.

**11. a.** Show that if $\rho$ is any metric on a set $X$, then $\sigma = \rho/(1 + \rho)$ is an equivalent metric for $X$. Prove that $\langle X, \sigma\rangle$ is a bounded metric space; that is, $\sigma(x, y) \leq 1$ for all $x$ and $y$ in $X$.

**b.** Show that this holds for extended metrics and pseudometrics.

## 4 Convergence and Completeness

Just as for the case of real numbers, we say that a sequence $\langle x_n\rangle$ from a metric space $X$ **converges** to the point $x$ in $X$ (or has $x$ as a limit), if given $\epsilon > 0$, there is an $N$ such that $\rho(x, x_n) < \epsilon$ for all $n \geq N$. This definition can be rephrased in geometric terms by saying that $\langle x_n\rangle$ converges to $x$ if every ball about $x$ contains all but a finite number of terms of the sequence.

We often write $x = \lim x_n$ or $x_n \to x$ to mean that $x$ is the limit of the sequence $\langle x_n\rangle$. If we have merely the weaker condition that each ball about $x$ contains infinitely many terms of the sequence, then we say that $x$ is a cluster point of the sequence $\langle x_n\rangle$. Thus $x$ is a cluster point of $\langle x_n\rangle$ if given $\epsilon > 0$ and given $N$, there is an $n \geq N$ such that $\rho(x, x_n) < \epsilon$. Thus if $x$ is the limit of $\langle x_n\rangle$, then $x$ is a cluster point of $\langle x_n\rangle$, but the converse is not necessarily true. A number of properties concerning limits and cluster points are given in the problems.

A sequence $\langle x_n\rangle$ from a metric space is called a **Cauchy sequence**, if given $\epsilon > 0$, there is an $N$ such that for all $n$ and $m$ larger than $N$ we have $\rho(x_n, x_m) < \epsilon$. If $\langle x_n\rangle$ converges to some point $x$, then given $\epsilon > 0$ we may choose $N$ so large that $\rho(x_n, x) < \epsilon/2$ for $n \geq N$. Hence for all $n, m \geq N$ we have

$$\rho(x_n, x_m) \leq \rho(x_n, x) + \rho(x_m, x) < \epsilon,$$

and $\langle x_n\rangle$ is a Cauchy sequence. The converse statement that every Cauchy sequence converges is not true in an arbitrary metric space (for example, in the space of rational numbers with the usual metric). If a metric space has the property that every Cauchy sequence converges (to some point of the space), we say that the space is **complete.** The Cauchy criterion for real numbers merely states that the space **R** of real numbers is complete. Other examples of complete spaces are given by the Banach spaces discussed in Chapter 6.

If $X$ is an incomplete metric space, it can always be enlarged to become complete. A precise statement of this fact is given by the following theorem. One proof of this theorem is outlined in Problem 17 and another in Problem 10.16.

**9. Theorem:** *If $\langle X, \rho \rangle$ is an incomplete metric space, it is possible to find a complete metric space $X^*$ in which $X$ is isometrically embedded as a dense subset. If $X$ is contained in an arbitrary complete space $Y$, then $X^*$ is isometric with the closure of $X$ in $Y$.*

## Problems

**12.** Show that a sequence $\langle x_n \rangle$ in a metric space has $x$ as a cluster point if and only if there is a subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ that converges to $x$.

**13.** Show that a sequence $\langle x_n \rangle$ in a metric space converges to $x$ if and only if every subsequence of $\langle x_n \rangle$ has $x$ as a cluster point. Hence $\langle x_n \rangle$ converges to $x$ if every subsequence has in turn a subsequence that converges to $x$.

**14.** Let $E$ be a set in a metric space $X$. If $x$ is a cluster point of a sequence from $E$, then $x \in \bar{E}$, while, if $x \in \bar{E}$, there is a sequence from $E$ that converges to $x$.

**15.** If a Cauchy sequence $\langle x_n \rangle$ in a metric space has a cluster point $x$, then $\langle x_n \rangle$ converges to $x$.

**16.** If $X$ and $Y$ are metric spaces and $f$ a mapping from $X$ to $Y$, then $f$ is continuous at $x$ if and only if for each sequence $\langle x_n \rangle$ in $X$ converging to $x$ we have $\langle f(x_n) \rangle$ converging to $f(x)$ in $Y$.

**17. a.** If $\langle x_n \rangle$ and $\langle y_n \rangle$ are Cauchy sequences from a metric space $X$, then $\rho(x_n, y_n)$ converges.

**b.** The set of all Cauchy sequences from a metric space $X$ becomes a pseudometric space (cf. Problem 3) if $\rho^*(\langle x_n \rangle, \langle y_n \rangle) = \lim \rho(x_n, y_n)$.

**c.** This pseudometric space becomes (as in Problem 3) a metric space $X^*$ when we identify elements for which $\rho^* = 0$, and $X$ is isometrically embedded in $X^*$.

**d.** The metric space $\langle X^*, \rho^* \rangle$ is complete. [Hint: If $\langle x_n \rangle$ is a Cauchy sequence from $X$, we may assume (by taking subsequences) that $\rho(x_n, x_{n+1}) < 2^{-n}$. If $\langle\langle x_{n,m} \rangle_{n=1}^{\infty} \rangle_{m=1}^{\infty}$ is a sequence of such Cauchy sequences which represents a Cauchy sequence in $X^*$, then the sequence $\langle x_{n,n}\rangle_{n=1}^{\infty}$ is a Cauchy sequence from $X$ which represents the limit of the Cauchy sequences from $X^*$.]

**e.** Use Propositions 10, 11, and 12 of the next two sections to complete the proof of Theorem 9.

**18.** Prove that the Cartesian product of two complete metric spaces is complete.

## 5 Uniform Continuity and Uniformity

Let $f$ be a mapping from the metric space $\langle X, \rho\rangle$ to the metric space $\langle Y, \sigma\rangle$. We say that $f$ is **uniformly continuous** if given $\epsilon > 0$ there is a $\delta > 0$ such that for all $x$ and $x'$ in $X$ with $\rho(x, x') < \delta$ we have $\sigma(f(x), f(x')) < \epsilon$. The function $h$ defined on $[0, 1)$ by $h(x) = x/(1 - x)$ is continuous but not uniformly continuous. Moreover, this function $h$ takes the Cauchy sequence given by $x_n = 1 - 1/n$ into the sequence $y_n = n - 1$, which is not a Cauchy sequence. Thus the image of a Cauchy sequence under a continuous function need not be a Cauchy sequence. However, we have the following proposition:

**10. Proposition:** *Let $f$ be a uniformly continuous mapping of the metric space $X$ into the metric space $Y$. If $\langle x_n\rangle$ is a Cauchy sequence in $X$, then $\langle f(x_n)\rangle$ is a Cauchy sequence in $Y$.*

A homeomorphism $f$ between metric spaces $X$ and $Y$ is called a **uniform homeomorphism** if both $f$ and $f^{-1}$ are uniformly continuous. It follows from Proposition 10 that the property of being a Cauchy sequence is preserved under uniform homeomorphism, as is the property of being complete. Properties that are preserved under uniform homeomorphisms are called uniform properties. In addition to the properties of being a Cauchy sequence and of completeness, we also have uniform continuity as a uniform property. These three properties are not topological properties, since the function $h$ defined by $h(x) = x/(1 - x)$ is a homeomorphism between the incomplete space $[0, 1)$ and the complete space $[0, \infty)$, which takes a Cauchy sequence into a sequence which is not a Cauchy sequence and its inverse carries the uniformly continuous function sin back into a function which is not uniformly continuous on $[0, 1)$.

Two metrics $\rho$ and $\sigma$ for a set $X$ of points are said to be **uniformly equivalent** if the identity map from $\langle X, \rho\rangle$ to $\langle X, \sigma\rangle$ is a uniform homeomorphism. Thus $\sigma$ and $\rho$ are uniformly equivalent if given $\epsilon > 0$, there is a $\delta > 0$ so that for all $x$ and $y$ we have $\rho(x, y) < \delta \Rightarrow \sigma(x, y) < \epsilon$ and $\sigma(x, y) < \delta \Rightarrow \rho(x, y) < \epsilon$.

Let $(x, \rho)$ and $(Y, \sigma)$ be two metric spaces. We define the metrics $\rho_1$ and $\rho_\infty$ on the Cartesian product $X \times Y$ by

$$\rho_1(\langle x, y\rangle, \langle x', y'\rangle) = \rho(x, x') + \sigma(y, y')$$
$$\rho_\infty(\langle x, y\rangle, \langle x', y'\rangle) = \max\,[\rho(x, x') + \sigma(y, y')].$$

The product $X \times Y$ with one of these metrics is said to be the product with the $l_1$ (or $l_\infty$) product metric. Each of these metrics is readily seen to be **uniformly** equivalent to the usual product metric on $X \times Y$. These metrics are often easier to use than the usual product metric, which was defined so as to make $\mathbf{R}^n \times \mathbf{R}^m$ isometric to $\mathbf{R}^{m+n}$.

We conclude this section by stating the following useful extension theorem for uniformly continuous mappings. Its proof is left to the reader.

**11. Proposition:** *Let $\langle X, \rho\rangle$ and $\langle Y, \sigma\rangle$ be metric spaces with $Y$ complete. Let $f$ be a uniformly continuous mapping from a subset $E$ of $X$ into $Y$. Then there is a unique continuous extension $g$ of $f$ from $E$ to $\bar{E}$; that is, there is a unique continuous mapping $g$ from $\bar{E}$ into $Y$ such that $g(x) = f(x)$ for $x \in E$. Moreover, $g$ is uniformly continuous.*

## Problems

**19.** Show that $\rho_1$ and $\rho_\infty$ are metrics and that each is uniformly equivalent to the usual product metric on $X \times Y$.

**20.** Prove Proposition 10.

**21.** Prove Proposition 11 by the following steps:

**a.** If $\langle x_n\rangle$ is a sequence from $E$ that converges to a point $x \in \bar{E}$, then $\langle f(x_n)\rangle$ converges to a point $y \in Y$ (cf. Proposition 10).

**b.** The point $y$ in (a) depends only on $x$ and not on the sequence $\langle x_n\rangle$. Thus, if we define $y = g(x)$, we have defined a function $g$ on $\bar{E}$ which is an extension of $f$.

**c.** The function $g$ is uniformly continuous on $\bar{E}$.

**d.** If $h$ is any continuous function from $\bar{E}$ to $Y$ that agrees with $f$ on $E$, then $h \equiv g$.

**22. a.** Show that the metrics in Problem 10b are uniformly equivalent.

**b.** Find a metric for the set of $n$-tuples of real numbers that is equivalent but not uniformly equivalent to the usual metric.

**c.** If $\langle X, \rho \rangle$ is any metric space, the metric $\sigma = \rho/(1 + \rho)$ is uniformly equivalent to $\rho$.

**23. a.** Boundedness is a metric but not a uniform property (see Problem 22c).

**b.** A metric space $X$ is said to be totally bounded if given $\epsilon > 0$ there are a finite number of balls of radius $\epsilon$ that cover (that is, whose union is) $X$. Show that total boundedness is a uniform property.

**c.** Show that total boundedness is not a topological property. [Consider $[0, 1)$ and $[0, \infty)$.]

**d.** Show that every totally bounded metric space is separable.

**24.** Let $(X_k, \rho_k)$ be a sequence of metric spaces. We define their *direct product*

$$Z = \bigtimes_{k=1}^{\infty} X_k$$

by letting $Z$ be the space of all sequences $\langle x_k \rangle$ with $x_k \in X_k$ and defining a metric $\tau$ on $Z$ by setting

$$\tau(x, y) = \sum_{k=1}^{\infty} 2^{-k} \rho_k^*(x_k, y_k),$$

where $x = \langle x_k \rangle$, $y = \langle y_k \rangle$, and $\rho_k^* = \rho_k/1 + \rho_k$. In the special case when $(X_k, \rho_k)$ are all the same space $(X, \rho)$, so that we are dealing with sequences from $X$), we often write

$$X^{\omega} \quad \text{or} \quad X^{\mathbf{N}}$$

instead of $\bigtimes_{k=1}^{\infty} X_k$.

**a.** Show that $\tau$ is a metric and that a sequence $\langle x^{(n)} \rangle$ in $Z$ converges to $x \in Z$ if and only if $x_k^{(n)}$ converges to $x_k$ for each $k$.

**b.** Show that $(Z, \tau)$ is complete if each $(X_k, \rho_k)$ is.

**c.** Assume that for each $k$ the spaces $(X_k, \rho_k)$ and $(Y_k, \sigma_k)$ are homeomorphic; then so are the spaces

$$\bigtimes X_k \quad \text{and} \quad \bigtimes Y_k.$$

**d.** Extend (c) to the case of uniform homeomorphisms.

## 6 Subspaces

If $\langle X, \rho\rangle$ is a metric space and $S$ is a subset of $X$, then $S$ becomes a metric space if we restrict $\rho$ to $S$, that is to say, if we take as the distance between two points of $S$ their distance as points of $X$. When we consider $S$ as a metric space with this metric, we call $S$ a subspace of $X$. For example, the rationals are a subspace of $\mathbf{R}$, and the set $\{\langle x, 0\rangle\}$ in $\mathbf{R}^2$ is a subspace isometric with $\mathbf{R}$. The space $C$ is a subspace of $L^\infty$.

If $E$ is a subset of $S$, then we may consider the closure of $E$ in $S$ or in $X$; that is, we may wish to consider the set of all points of $X$ which are points of closure of $E$ or else the set of all points of $S$ which are points of closure of $E$. These sets are in general different. For example, let $X$ be the space $\mathbf{R}$ and $S$ the interval $(0, 1)$. Then if $E$ is the interval $(0, \frac{1}{2}]$, the closure of $E$ in $\mathbf{R}$ is the interval $[0, \frac{1}{2}]$, while its closure in $S$ is just $(0, \frac{1}{2}]$; that is, $E$ is closed relative to $S$. Thus we see that the closure of a set as well as the properties of a set being closed or open are all relative to the space containing the set. However, we do have the following relations between these notions.

**12. Proposition:** *Let $X$ be a metric space and $S$ a subspace of it. Then the closure of $E$ relative to $S$ is $\bar{E} \cap S$, where $\bar{E}$ denotes the closure of $E$ in $X$. A set $A \subset S$ is closed relative to $S$ if and only if $A = S \cap F$ with $F$ closed in $X$. A set $A \subset S$ is open relative to $S$ if and only if $A = S \cap O$ with $O$ open in $X$.*

***Proof:*** If $x$ is a point of closure of $E$ in $X$, then it is a point of closure of $E$ in $S$ if it belongs to $S$. Hence the closure in $S$ of $E$ is $\bar{E} \cap S$. If $A$ is closed in $S$, we must have $A = S \cap \bar{A}$, while on the other hand if $F$ is closed in $X$, then the closure in $S$ of $S \cap F$ is

$$S \cap (\overline{S \cap F}) \subset S \cap (\bar{S} \cap \bar{F}) \subset S \cap F,$$

whence $S \cap F$ is closed relative to $S$.

If $A$ is open relative to $S$, then $S \sim A$ is closed relative to $S$ and we have $S \sim A = S \cap F$, or $S \cap A = S \cap (\sim F)$ and $\sim F$ is open in $X$. Similarly, if $O$ is open, then $S \cap O$ is the complement in $S$ of $S \cap (\sim O)$, which is closed in $S$. ▮

**13. Proposition:** *Every subspace of a separable metric space is separable.*

***Proof:*** Let $X$ be a separable metric space and $S$ a subspace. Then by Proposition 6 there is a countable collection $\{O_i\}$ of open sets in $X$ such that each open set in $X$ is a union of some subcollection of $\{O_i\}$. By Proposition 12 the collection $\{O_i \cap S\}$ is a countable collection of open subsets of $S$ such that every open subset of $S$ is a union of a subcollection of them. Hence $S$ is separable by Proposition 6. ∎

In contrast to the relative properties discussed above, there are some properties that are intrinsic. For example, the property of $x$ being a point of closure of $E$ holds in any subspace of $X$ containing $x$ and $E$ as soon as it holds in one of them. Another such property is that of being complete, since the definition of completeness of a space is given in terms of points in the space. However, the following proposition gives some relations between complete sets and closed sets.

**14. Proposition:** *If a subset $A$ of a metric space $X$ is complete, then it is closed. On the other hand, a closed subset of a complete metric space is itself complete.*

## Problems

**25.** Prove Proposition 14. [Hint: Problem 14 is helpful.]

**26.** Let $O$ be an open subset of a complete metric space $(X, \rho)$. Show that there is a bounded metric $\sigma$ for $O$ which is equivalent to $\rho$ on $O$ and for which $(O, \sigma)$ is a *complete* metric space. [Hint: Let $\varphi(x) = [\rho(x, \tilde{O})]^{-1}$ for each $x \in O$. Then $\{\langle x, y\rangle : x \in O, y = \varphi(x)\}$ is a closed, and therefore complete, subset of $X \times \mathbf{R}$. Use Problems 19 and 22c.]

## 7 Compact Metric Spaces

Many of the important properties of the interval $[0, 1]$ follow from the Heine–Borel Theorem. We introduce a class of metric spaces in which the conclusion of the Heine–Borel Theorem is valid and show that many properties of $[0, 1]$ are also true for these spaces. These spaces are called compact spaces. To give an exact definition, we say that a collection $\mathcal{U}$ of open sets in a metric space is an open covering for a set $K$ if $K$ is contained in the union of the sets in $\mathcal{U}$. A metric space $X$ is said to be **compact** if every open covering $\mathcal{U}$ of $X$ has a finite subcovering, that is, if there is a finite collection $\{O_1, O_2, \ldots, O_N\} \subset \mathcal{U}$ such that $X = \bigcup_{i=1}^{N} O_i$. A subset $K$ of a metric space is called compact if it is compact as a subspace of $X$. In view of the last statement of Proposition 12, this is equivalent to saying that a subset $K$ of $X$ is compact if every covering $\mathcal{U}$ of $K$ by open sets of $X$ has a finite subcovering. The Heine–Borel Theorem states that every closed and bounded subset of real numbers is compact.

If $\mathcal{U}$ is an open covering of a space $X$, then the collection $\mathcal{F}$ of complements of sets in $\mathcal{U}$ is a collection of closed sets whose intersection is empty, and conversely. Thus a space $X$ is compact if and only if every collection of closed sets with an empty intersection has a finite subcollection whose intersection is empty. A collection $\mathcal{F}$ of sets in $X$ is said to have the *finite intersection property* if any finite subcollection of $\mathcal{F}$ has a nonempty intersection. Hence we have the following proposition:

**15. Proposition:** *A metric space $X$ is compact if and only if every collection $\mathcal{F}$ of closed sets with the finite intersection property has a nonempty intersection.*

Another important property of the interval $[0, 1]$ which generalizes is the following: A space $X$ is said to have the **Bolzano–Weierstrass property** if every infinite sequence $\langle x_n \rangle$ in $X$ has at least one cluster point, that is, if there is an $x \in X$ such that each open set containing $x$ and for each $N$ there is an $n \geq N$ with $x_n \in O$.

**16. Lemma:** *A compact space has the Bolzano–Weierstrass property.*

***Proof:*** Let $\langle x_i \rangle$ be a sequence from $X$, and set $B_n = \{x_n, x_{n+1}, \ldots\}$. Then $\{\bar{B}_n\}$ is a collection of closed sets with the finite intersection property, and so there is an $x \in \bigcap \bar{B}_n$. This $x$ is a cluster point for the sequence $\langle x_i \rangle$. ∎

Another related property is sequential compactness: A space $X$ is said to be *sequentially compact* if every sequence $\langle x_n \rangle$ from $X$ contains a convergent subsequence $\langle x_{n_k} \rangle$.

**17. Lemma:** *A metric space $X$ has the Bolzano–Weierstrass property if and only if $X$ is sequentially compact.*

***Proof:*** Since every limit of a subsequence of $\langle x_n \rangle$ is a cluster point of $\langle x_n \rangle$, sequential compactness implies the Bolzano–Weierstrass property. Conversely, if $\langle x_n \rangle$ has $x$ for a cluster point, then for each $k$ we can find an $n_k > n_{k-1}$ such that the ball of radius $1/k$ about $x$ contains $x_{n_k}$. Then $x_{n_k} \to x$. Thus a metric space with the Bolzano–Weierstrass property is sequentially compact. ∎

We shall later show (Theorem 21) that a sequentially compact metric space is compact so that the three notions of compactness, sequential compactness, and the Bolzano–Weierstrass property all coincide for metric spaces. They will turn out to be distinct notions, however, when we consider general topological spaces. We next establish some important properties of compact metric spaces. Until we get to Theorem 21, it is important to note that these depend only on the apparently weaker hypothesis of sequential compactness.

**18. Theorem:** *Let $f$ be a continuous real-valued function on a (sequentially) compact space. Then $f$ is bounded and assumes its maximum and minimum.*

***Proof:*** Let $M = \sup \{f(x): x \in X\}$. Then (even if $M = \infty$) there is a sequence $\langle x_n \rangle$ of points of $X$ such that $M = \lim f(x_n)$. Since $X$ is sequentially compact, there is a subsequence $\langle x_{n_k} \rangle$ that converges to a point $z \in X$. Then $f(z) = \lim f(x_{n_k}) = M$. Thus $M$ is a real number (i.e., not $\infty$) and $f(z) = M$. Hence $M$ is the maximum of $f$ and is assumed at $z$. The other half of the theorem follows by replacing $f$ by $-f$. ∎

A metric space $X$ is said to be *totally bounded* if, for each $\epsilon > 0$, there is a finite collection of points $\{x_1, \ldots, x_n\}$ such that each $x \in X$ is within a distance of $\epsilon$ of one of the $x_k$. This is equivalent to saying that for each $\epsilon > 0$ the space $X$ is covered by a finite number of balls of radius $\epsilon$.

**19. Lemma:** *A sequentially compact metric space is totally bounded.*

***Proof:*** Suppose that $X$ is not totally bounded. Then for some $\epsilon > 0$ we cannot cover $X$ by a finite number of balls of radius $\epsilon$.

Hence we can pick $x_1, x_2, \ldots, x_n$ in turn so that $\rho(x_n, x_k) \geq \epsilon$ for $n > k$. Then the sequence $\langle x_n \rangle$ can have no convergent subsequence, since any two different terms in it are at distance $\epsilon$ or more. Thus $X$ is not sequentially compact, and the lemma follows by contraposition. ▮

If $\mathcal{U}$ is an open covering of a metric space $X$, then each $x \in X$ is contained in some open set $O \in \mathcal{U}$, and hence for some $\delta > 0$ there is a ball $B_{x,\delta}$ about $x$ that is contained in $O$. The following proposition tells us that for a (sequentially) compact metric space this property holds uniformly in the sense that we can find $\delta$ independently of $x$. The number $\epsilon$ with the property stated in the proposition is called the *Lebesgue number* of the covering $\mathcal{U}$. The proof of the proposition is left to the reader (Problem 29).

**20. Proposition:** *Let $\mathcal{U}$ be an open cover of a (sequentially) compact metric space $X$. Then there is a number $\epsilon > 0$ such that for each $x \in X$ and each $\delta < \epsilon$ the ball $B_{x,\delta}$ is contained in some open set $O \in \mathcal{U}$.*

**21. Theorem (Borel–Lebesgue):** *Let $X$ be a metric space. Then the following are equivalent:*

i. *$X$ is compact.*
ii. *$X$ has the Bolzano–Weierstrass property.*
iii. *$X$ is sequentially compact.*

***Proof:*** We have already shown that (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii) by Lemmas 16 and 17. To show that (iii) $\Rightarrow$ (i), assume $X$ is sequentially compact and $\mathcal{U}$ is an open covering of $X$. Let $\epsilon$ be the Lebesgue number of $\mathcal{U}$ given by Proposition 20, and take $\delta$ with $0 < \delta < \epsilon$. Since $X$ is totally bounded by Lemma 19, we can find a finite number of balls $B_{x_1,\delta}, \ldots, B_{x_n,\delta}$ of radius $\delta$ which cover $X$. Since each ball $B_{x_k,\delta}$ is contained in a set $O_k \in \mathcal{U}$, the collection $\{O_1, \ldots, O_n\}$ is a finite subcollection of $\mathcal{U}$ which covers $X$. ▮

We conclude this section with several propositions expressing useful properties of compact metric spaces. We first observe that the notion of compactness is intimately connected with that of being closed, as the following proposition shows. Thus compactness may be viewed as an absolute type of closedness.

**22. Proposition:** *A closed subset of a compact space is compact. A compact subset of a metric space is closed and bounded.*

***Proof:*** Let $X$ be compact, $F$ a closed subset of $X$, and $\mathcal{U}$ an open covering for $F$. Then $\mathcal{U} \cup \{\tilde{F}\}$ is an open covering for $X$ and so must have a finite subcovering $\{\tilde{F}, O_1, \ldots, O_N\}$. Then the sets $O_1$, $O_2, \ldots, O_N$ cover $F$, and so $\mathcal{U}$ has a finite subcovering.

Suppose that $K$ is a compact subset of a metric space $X$, and let $y$ be a point of closure of $K$. Then $f(x) = \rho(x, y)$ is a continuous function on $K$ whose infimum is 0. Since $f$ assumes its minimum on $K$, there is a $z \in K$ such that $f(z) = \min f(x) = 0$. Hence $\rho(z, y) = 0$, and so $y = z \in K$. Thus $K$ is closed. To see that $K$ is bounded, we observe that $\rho(x_0, x)$ is a continuous function of $x$ on $K$ and so bounded by some number $M$. Then $K \subset B_{x_0, M}$. ▮

**23. Corollary:** *Every compact set of real numbers is closed and bounded.*

**24. Proposition:** *The continuous image of a compact set is compact.*

***Proof:*** Let $f$ be a continuous function mapping the compact set $K$ *onto* a space $Y$. If $\mathcal{U}$ is an open covering for $Y$, then the collection of sets $f^{-1}[O]$ for all $O \in \mathcal{U}$ is an open covering of $K$. By the compactness of $K$, there are a finite number $O_1, \ldots, O_n$ of sets of $\mathcal{U}$ such that the sets $f^{-1}[O_i]$ cover $K$. Since $f$ is onto, the sets $O_1, \ldots, O_n$ cover $Y$. ▮

**25. Proposition:** *A metric space $X$ is compact if and only if it is both complete and totally bounded.*

***Proof:*** If $X$ is compact, it is trivially totally bounded. If $\langle x_n \rangle$ is a Cauchy sequence in $X$, then $\langle x_n \rangle$ must have a cluster point. But a Cauchy sequence which has a cluster point converges to the cluster point. Thus $X$ is complete.

Suppose that $X$ is complete and totally bounded. To show that $X$ is compact, it suffices to show that each infinite sequence $\langle x_n \rangle$ has a convergent subsequence. Since $X$ is totally bounded, we may cover $X$ by a finite number of spheres of radius 1. Among these spheres there must be a sphere $S_1$ that contains infinitely many terms of the sequence $\langle x_n \rangle$. Covering $X$ by a finite number of spheres of radius $\frac{1}{2}$, we can find among them a sphere $S_2$ such that $S_1 \cap S_2$ contains infinitely many terms of the sequence $\langle x_n \rangle$. Continuing, we obtain a sequence $\langle S_k \rangle$ of spheres, $S_k$ having radius $1/k$, such that $S_1 \cap \cdots \cap S_k$ contains infinitely many terms of the sequence. Since there are infinitely many terms of the sequence in $S_1 \cap \cdots \cap S_k$, we may choose $n_k$ so that $n_k > n_{k-1}$ and $x_{n_k} \in S_1 \cap \cdots \cap S_k$. Then $\langle x_{n_k} \rangle$ is a subsequence of $\langle x_n \rangle$, and it must be a Cauchy sequence, since $\rho(x_{n_k}, x_{n_l}) \le 2/N$ for $k, l \ge N$. Since $X$ is complete, this subsequence converges. ▮

**26. Proposition:** *Let $f$ be a continuous mapping of a compact metric space $X$ into a metric space $Y$. Then $f$ is uniformly continuous.*

*Proof:* Given $\epsilon > 0$ and $x \in X$, there is a $\delta_x > 0$ such that $\rho(x, y) < \delta_x$ implies that $\sigma(f(x), f(y)) < \epsilon/2$. Let $O_x$ be the spheroid $\{y: \rho(x, y) < \frac{1}{2}\delta_x\}$. Then $\{O_x: x \in X\}$ is an open covering of $X$ and so has a finite subcovering $\{O_{x_1}, \ldots, O_{x_n}\}$. Let $\delta = \frac{1}{2} \min \{\delta_{x_1}, \ldots, \delta_{x_n}\}$. Then $\delta > 0$. Given two points $y$ and $z$ in $X$ such that $\rho(y, z) < \delta$, the point $y$ must belong to some $O_{x_i}$, and hence $\rho(y, x_i) < \frac{1}{2}\delta_{x_i}$. Consequently, $\rho(z, x_i) \le \rho(z, y) + \rho(y, x_i) < \frac{1}{2}\delta_{x_i} + \delta \le \delta_{x_i}$. Thus we have $\sigma(f(y), f(x_i)) < \epsilon/2$ and $\sigma(f(z), f(x_i)) < \epsilon/2$. This implies that $\sigma(f(z), f(y)) < \epsilon$, showing that $f$ is uniformly continuous on $X$. ▮

## Problems

**27.** Let $X$ be a metric space, $K$ a compact subset, and $F$ a closed subset. Then $F \cap K = \emptyset$ if and only if $\rho(F, K) > 0$, i.e., if there is a $\delta > 0$ such that $\rho(x, y) > \delta$ for all $x \in F$ and all $y \in K$. [Consider the function $\rho(x, F) = \inf_{y \in F} \rho(x, y)$.]

**28. a.** Let $X$ be a totally bounded metric space, and $f: X \to Y$ a uniformly continuous map onto $Y$. Then $Y$ is totally bounded.

**b.** Is this result still true if $f$ is only required to be continuous?

**29.** Prove Proposition 20. [Hints:

**a.** We need only consider the case when $X \notin \mathcal{U}$. Set $\varphi(x) = \sup \{r: \exists O \in \mathcal{U} \text{ with } B_{x,r} \subset O\}$. Show that $0 < \varphi(x) < \infty$.

**b.** Show that for each $x$ and $y$

$$\varphi(y) \ge \varphi(x) - \rho(x, y).$$

**c.** Show that $\varphi$ is a continuous function on $X$.

**d.** If $X$ is sequentially compact, then $\epsilon = \inf \varphi$ is positive.

**e.** This $\epsilon$ satisfies the conclusion of the proposition.]

**30. a.** Let $Z = \bigtimes_{k=1}^{\infty} X_k$ (see Problem 24). Show that $Z$ is totally bounded if each $X_k$ is.

**b.** Show that if each $X_k$ is a compact metric space, so is their direct product.

## 8 Baire Category

In this section we go more deeply into certain aspects of complete metric spaces, namely, the Baire theory of category. The applications of this theory give some of the deepest results of a topological nature which are useful in analysis. We begin with the classical theorem of Baire.

**27. Theorem (Baire):** *Let $X$ be a complete metric space and $\{O_k\}$ a countable collection of dense open subsets of $X$. Then $\bigcap O_n$ is dense.*

***Proof:*** Given an open set $U$, let $x_1$ be a point of $O_1 \cap U$ and $S_1$ a ball (of radius $r_1$) centered at $x_1$ and contained in $O_1 \cap U$. Since $O_2$ is dense, there must be a point $x_2$ in $O_2 \cap S_1$. Since $O_2$ is open, there is a small ball $S_2$ centered at $x_2$ and contained in $O_2$, and we may take the radius $r_2$ of $S_2$ to be smaller than $\frac{1}{2}r_1$ and smaller than $r_1 - \rho(x_1, x_2)$. Then $\bar{S}_2 \subset S_1$. Proceeding inductively, we obtain a sequence $\langle S_n \rangle$ of balls such that $\bar{S}_n \subset S_{n-1}$ and $S_n \subset O_n$ and whose radii $\langle r_n \rangle$ tend to zero. Let $\langle x_n \rangle$ be the sequence of centers of these balls. Then for $n, m \geq N$ we have $x_n \in S_N$ and $x_m \in S_N$. Therefore, $\rho(x_n, x_m) \leq 2r_N$, and $\langle x_n \rangle$ is a Cauchy sequence, since $r_n \to 0$. By the completeness of $X$ there is a point $x$ such that $x_n \to x$. Since $x_n \in S_{N+1}$ for $n > N$, we have $x \in \bar{S}_{N+1} \subset S_N \subset O_N$. Consequently, $x \in \bigcap O_n$, and $x \in U$. Since $U$ was an arbitrary open, $\bigcap O_n$ is dense in $X$. ▮

We introduce some terminology. We begin by saying that set $E$ is *nowhere dense* if $\sim(\bar{E})$ is dense. This is equivalent to saying that $\bar{E}$ contains no nonempty open set. A set $E$ is said to be of the **first category** (or *meager*) if $E$ is the union of a countable collection of nowhere dense sets. A set that is not of first category is said to be of **second category** (or *nonmeager*), and the complement of a set of first category is called **residual** (or *co-meager*).

With this terminology the theorem of Baire can be reformulated as follows:

**28. Corollary (Baire Category Theorem):** *Let $X$ be a complete metric space. Then no nonempty open subset of $X$ is of first category, i.e., the union of a countable collection of nowhere dense subsets.*

***Proof:*** Let $E_n$ be a countable collection of nowhere dense subsets and $U$ a nonempty open set. Then $O_n = \sim\bar{E}_n$ is a dense open subset of $X$, and by the Theorem of Baire there is a point $x \in U$ which is in $\bigcap O_n$. But this means that $x \notin \bigcup E_n$, and so $U$ is not contained in $\bigcup E_n$. ∎

The Theorem of Baire and its corollary are unusual in that the hypothesis involves a uniform property (the completeness of $X$) and the conclusion is of a purely topological nature, for, if $X$ and $Y$ are homeomorphic and if no nonempty open subset of $X$ is of first category, then neither is any nonempty open subset of $Y$. Metric (or topological) spaces with this property are said to be of *second category everywhere* (with respect to themselves). For these spaces we have the conclusions of all of the propositions in this section as well as many of the properties explored in the problems. In addition to the complete metric spaces, we shall see that also locally compact Hausdorff spaces are of second category everywhere. For these spaces, however, many of the results of category theory follow directly from local compactness.

There is some similarity between the properties of the class of sets of first category in a topological space and the class of sets of measure zero in a complete measure space: Both classes are closed under the operation of taking countable unions, and any subset of a set in one of these classes is again in the class. Any collection of sets with these two properties is called a *σ-ideal* of sets. Sets of first category can be regarded as "small" sets. If $X$ is a complete metric space, then a set of first category can have no interior points.

On the other hand, a set of first category in $[0, 1]$ can have Lebesgue measure 1, and hence have a complement which is a residual set of measure zero.

The following Propositions express some characterizations and properties that are useful to have in mind when proving statements involving category theory.

**29. Proposition:** *If $O$ is open and $F$ closed, then the sets $\bar{O} \sim O$ and $F \sim F^\circ$ are nowhere dense. If $F$ is closed and of first category in a complete metric space, then $F$ is nowhere dense.*

**30. Proposition:** *A subset of a complete metric space is residual if and only if it contains a dense $G_\delta$. Hence a subset of a complete metric space is of first category if and only if it is contained in an $F_\sigma$ whose complement is dense.*

The utility of this last proposition lies in the fact that if we know *a priori* that a set $E$ is a $G_\delta$, then it suffices to show that $E$ is dense in order to show that $E$ is residual.

**31. Proposition:** *Let $\{F_n\}$ be a countable collection of closed sets with $X = \bigcup F_n$. Then $O = \bigcup F_n^\circ$ is a residual open set. If $X$ is a complete metric space, $O$ is dense.*

***Proof:*** The sets $E_n = F_n \sim F_n^\circ$ are nowhere dense, so $E = \bigcup E_n$ is a set of first category. But $X \sim O \subset E$, and so $O$ is residual. The second statement now follows from Proposition 30. ▮

Some of the applications of the theory of category to analysis seem almost too good to be believed. One useful application is the following theorem, which is known as the *uniform boundedness principle.* Others are given in the following problems, in Section 10.4 and in Problem 10.44e.

**32. Theorem:** *Let $\mathcal{F}$ be a family of real-valued continuous functions on a complete metric space $X$, and suppose that for each $x \in X$ there is a number $M_x$ such that $|f(x)| \le M_x$ for all $f \in \mathcal{F}$. Then there is a nonempty open set $O \subset X$ and a constant $M$ such that $|f(x)| \le M$ for all $f \in \mathcal{F}$ and all $x \in O$.*

***Proof:*** For each integer $m$, let $E_{m,f} = \{x\colon |f(x)| \le m\}$, and set $E_m = \bigcap_{\mathcal{F}} E_{m,f}$. Since each $f$ is continuous, $E_{m,f}$ is closed, and consequently $E_m$ is closed. For each $x \in X$, there is an $m$ such that $|f(x)| \le m$ for all $f \in \mathcal{F}$; that is, there is an $m$ such that $x \in E_m$. Hence

$$X = \bigcup_{m=1}^{\infty} E_m.$$

Since $X$ is a complete metric space, there is a set $E_m$ which is not nowhere dense. Since this $E_m$ is a closed set, it must contain some ball $O$. But for every $x \in O$ we have $|f(x)| \le m$ for all $f \in \mathcal{F}$. ▮

## Problems

**31. a.** Prove that a closed set $F$ is nowhere dense if and only if it contains no open set.

**b.** Prove that $E$ is nowhere dense if and only if for any nonempty open set $O$ there is a ball contained in $O \sim E$.

**32. a.** Prove that, if $E$ is of first category and $A \subset E$, then $A$ is also of first category.

**b.** Prove that if $\langle E_n \rangle$ is a sequence of sets of first category, then $\bigcup E_n$ is also of first category.

**33. a.** Show that on $[0, 1]$ there is a nowhere dense closed set having Lebesgue measure $1 - 1/n$.

**b.** Construct a set of first category on $[0, 1]$ that has measure 1.

**34.** Prove Proposition 29.

**35.** Prove Proposition 30.

**36. a.** A point $x$ in a metric space is called isolated if the set $\{x\}$ is open. Prove that a complete metric space without isolated points has an uncountable number of points.

**b.** Prove that $[0, 1]$ is uncountable. Compare this proof with that of Corollary 3.4.

**37.** Let $E$ be a subset of a complete metric space.

**a.** If $\tilde{E}$ is dense and $F$ a closed set contained in $E$, then $F$ is nowhere dense.

**b.** If $E$ and $\tilde{E}$ are both dense, then at most one of them is an $F_\sigma$.

**c.** The set of rational numbers in $[0, 1]$ is not a $G_\delta$.

**d.** Is there a real-valued function on $[0, 1]$ which is continuous on the rationals and discontinuous on the irrationals?

**38.** Let $C$ be the space of continuous functions on $[0, 1]$ and set $F_n = \{f\colon (\exists x_0)$ with $0 \le x_0 \le 1 - 1/n$ and $|f(x) - f(x_0)| \le n(x - x_0)$ for all $x$, $x_0 \le x < 1\}$.

**a.** Show that $F_n$ is a closed subset of $C$.

**b.** Show that $F_n$ is nowhere dense. [Hint: Any $g \in C$ can be approximated to within $\epsilon/2$ by a polygonal function $\varphi$, and $\varphi$ can be approximated to within $\epsilon/2$ by a polygonal function $\psi$ whose right-hand derivative is everywhere greater than $n$ in absolute value.]

**c.** Show that the set $D$ of continuous functions which have a finite derivate on the right for at least one point of $[0, 1]$ is a set of first category in $C$.

**d.** There is a nowhere differentiable continuous function on $[0, 1]$.

**39.** Show that under the hypotheses of Theorem 32 there is a dense open set $O \subset X$ such that each $x \varepsilon O$ has a neighborhood $U$ on which $\mathcal{F}$ is uniformly bounded.

**40.** Let $X$ and $Y$ be metric spaces. Carathéodory introduced a notion called continuous convergence: A sequence $\langle f_n \rangle$ of maps from $X$ to $Y$ is said to **converge continuously** *at a point* $x$ to a map $f$ of $X$ into $Y$ if, for each sequence $\langle x_n \rangle$ with $x = \lim x_n$, we have $f(x) = \lim f_n(x_n)$. We say that $f_n$ converges continuously to $f$ on $X$ if $f_n$ converges continuously to $f$ at each $x \varepsilon X$.

**a.** Show that $\langle f_n \rangle$ converges continuously to $f$ at $x$ if and only if, given $\epsilon > 0$, there is an integer $N$ and a neighborhood $U$ of $x$ such that $\sigma(f_n(x'), f(x)) < \epsilon$ for all $n \geq N$ and all $x' \varepsilon U$. [The 'if' part is straightforward; the 'only if' can be shown by contraposition.]

**b.** Let $Z = \{1/n\} \cup \{0\}$ have the metric it inherits as a subspace of $[0, 1]$. Define a map $g: X \times Z \to Y$ by setting

$$g\left(x, \frac{1}{n}\right) = f_n(x)$$

$$g(x, 0) = f(x).$$

Show that $g$ is continuous at a point $\langle x_0, 0 \rangle$ in the product metric if and only if $\langle f_n \rangle$ converges continuously at $x_0$.

**c.** Let $f_n$ converge continuously to $f$ at $x$. Then $f$ is continuous at $x$.

**d.** A sequence $\langle f_n \rangle$ of *continuous* maps of $X$ into $Y$ converges continuously to $f$ at $x$ if and only if given $\epsilon > 0$, there is an integer $N$ and a neighborhood $U$ of $x$ such that $\sigma(f_n(x'), f(x')) < \epsilon$ for all $n \geq N$ and all $x' \varepsilon U$.

**e.** A sequence $\langle f_n \rangle$ of continuous maps of $X$ into $Y$ converges continuously to $f$ on $X$ if and only if $\langle f_n \rangle$ converges to $f$ uniformly on each compact subset of $X$.

**f.** Let $X$ be a complete metric space and $\langle f_n \rangle$ a sequence of continuous maps of $X$ into a metric space $Y$ such that $f(x) = \lim f_n(x)$ for each $x \varepsilon X$. For each pair $m, n \varepsilon \mathbf{N}$ define

$$F_{m,n} = \left\{x \varepsilon X : \sigma(f_k(x), f_l(x)) \leq \frac{1}{m} \quad \text{for all } k, l \geq n\right\}.$$

Then $F_{m,n}$ are closed subsets of $X$ and $X = \bigcup_{n=1}^{\infty} F_{m,n}$. Also, $O_m = \bigcup_{n=1}^{\infty} F^{\circ}_{m,n}$ is a dense open subset of $X$.

**g.** Let $O_m$ be the sets defined in (*f*). Then for each $x \varepsilon O_m$, there is a neighborhood $U$ of $x$ and an $n$ such that

$$\sigma(f_k(x'), f_l(x')) \leq \frac{1}{m}$$

for all $k, l \geq n$ and all $x' \varepsilon U$. Hence also

$$\sigma(f_k(x'), f(x')) \leq \frac{1}{m}$$

for all $k \geq n$ and all $x' \varepsilon U$.

**h.** Let $E = \bigcap O_m$. Then $E$ is a dense $G_\delta$. If $x \varepsilon E$, then given $\epsilon > 0$ there is a neighborhood $U$ of $x$ and an $n$ such that $\sigma(f_k(x'), f(x')) < \epsilon$ for all $k \geq n$ and all $x' \varepsilon U$.

**i.** Let $X$ be a complete metric space and $\langle f_n \rangle$ a sequence of continuous maps of $X$ into a metric space $Y$. Suppose that $\langle f_n \rangle$ converges pointwise to a map $f$, i.e. $f(x) = \lim f_n(x)$ for each $x \varepsilon X$. Then there is a dense $G_\delta$ in $X$ on which $\langle f_n \rangle$ converges to $f$ continuously.

**41.** Let $(X, \rho)$ and $(Y, \sigma)$ be complete metric spaces and $f: X \times Y \to Z$ be a mapping into a metric space $(Z, \tau)$ that is continuous in each variable separately, i.e., such that

$$f(\cdot, y): X \to Z$$

is continuous for each $y \varepsilon Y$, and

$$f(x, \cdot): Y \to Z$$

is continuous for each $x \varepsilon Y$.

**a.** Fix $y_0 \varepsilon Y$. Set

$$F_{m,n} = \left\{x \varepsilon X : \tau[f(x, y), f(x, y_0)] \leq \frac{1}{m} \quad \text{for all } y \text{ with } \sigma(y, y_0) \leq \frac{1}{m}\right\}.$$

Then each $F_{m,n}$ is closed, and

$$X = \bigcup_{n=1}^{\infty} F_{m,n}.$$

**b.** Let $O_m = \bigcup_{n=1}^{\infty} F^{\circ}_{m,n}$. Then $O_m$ is a dense subset of $X$, and given $x \varepsilon O_m$ there is a neighborhood $U$ of $\langle x, y_0 \rangle$ in $X \times Y$ such that

$$\tau[f(p), f(x, y_0)] \leq \frac{1}{m}$$

for all $p \varepsilon U$.

**c.** Let $G = \bigcap O_m$. Then $G$ is a dense $G_\delta$ in $X$ and $f$ is continuous at $\langle x, y_0 \rangle$ for each $x \varepsilon G$.

**d.** There is a set $E \varepsilon X \times Y$ which is a dense $G_\delta$ such that $f$ is continuous at each $p \varepsilon E$.

**e.** Generalize to a product $X_1 \times X_2 \times \cdots \times X_p$ of a finite number of spaces.

**42. a.** Let $X$ and $Y$ be complete metric spaces, and let $G \subset X$ and $H \subset Y$ be dense $G_\delta$'s. Then $G \times H$ is a dense $G_\delta$ in $X \times Y$.

**b.** Let $O$ be a dense open set in $X \times Y$. Then there is a $G \subset X$ which is a dense $G_\delta$ such that

$$E_x = \{y \varepsilon Y: \langle x, y \rangle \varepsilon O\}$$

is a dense open subset of $Y$ for each $x \varepsilon G$.

**c.** Let $E$ be a dense $G_\delta$ in $X \times Y$. Then there is a $G \subset X$ which is a dense $G_\delta$ such that

$$E_x = \{y \varepsilon Y: \langle x, y \rangle \varepsilon O\}$$

is a dense $G_\delta$ for each $x \varepsilon G$.

**d.** Show that we cannot always take the set $G$ to be open.

**43.** Let $(X, \rho)$ be a complete metric space and $f$: $X \to \mathbf{R}$ be an upper semicontinuous function. Then the set $E$ of points at which $f$ is continuous is a dense $G_\delta$ in $X$.

## ★9 Absolute $G_\delta$'s

The last section has shown some of the importance of sets which are $G_\delta$'s in a complete metric space. Among other things they can readily be shown to be of second category everywhere with respect to themselves (or with respect to their closures). It is the purpose of this section to give a characterization of those spaces that are homeomorphic to a $G_\delta$ in a complete metric space: They are the spaces for which there is a metric giving the topology of the space and for which the space is complete. They have the property that any homeomorphic image of such a space in any metric space is a $G_\delta$. For later use the conclusions of several of the propositions given here are in terms of topological spaces. That the proofs given here for metric spaces extend to the more general case is straightforward (cf. Problem 8.31).

We begin with a proposition about $G_\delta$'s in complete metric spaces.

**33. Proposition:** *Let $(X, \rho)$ be a complete metric space and $E \subset X$ a $G_\delta$. Then there is a metric $\sigma$ for $E$, equivalent on $E$ to $\rho$, and such that $(E, \sigma)$ is a complete metric space.*

***Proof:*** Let $E = \bigcap_{n=1}^{\infty} O_n$ with $O_n$ open. Let $\sigma_n$ be a metric for $O_n$ which is bounded by 1, equivalent on $O_n$ to $\rho$, and for which $(O_n, \rho_n)$ is complete (cf. Problem 26). Set

$$\sigma = \sum 2^{-n}\sigma_n.$$

Then $\sigma$ is a metric on $E$ which is equivalent to $\rho$.

Let $\langle x_k \rangle$ be a Cauchy sequence in $\sigma$. Then, since $2^n\sigma \geq \sigma_n$, $\langle x_k \rangle$ must be a Cauchy sequence for $\sigma_n$. Hence $\langle x_k \rangle$ converges in $(O_n, \sigma_n)$ to a point $y_n \in O_n$. Since $\sigma_n$ and $\rho$ are equivalent on $O_n$, $x_k \to y_n$ in $(X, \rho)$. Since limits of sequences in metric spaces are unique, we have $y_n = y_m$ for all $m$, $n$. Call this common value $y$. Then $x_n \to y$ in $X$. Since $y = y_n \in O_n$, $y \in E = \bigcap O_n$. The equivalence of $\rho$ and $\sigma$ implies that $\langle x_k \rangle$ converges to $y$ in $(E, \sigma)$, and so $(E, \sigma)$ is complete. ▮

**34. Proposition:** *Let $E$ be a subset of a topological space $X$ and $f$: $E \to Y$ a continuous map into a complete metric space $(Y, \sigma)$. Then $f$ may be extended to a continuous function $f^*$: $E^* \to Y$ defined on a set $E^*$ that is a $G_\delta$ containing $E$.*

***Proof:*** Let $O_n = \{x \in X : \exists$ a neighborhood $U$ of $x$ such that $U \cap E \neq \emptyset$ and the diameter of $f[U \cap E]$ is less than $1/n\}$. Then $O_n$ is open and $E \subset O_n$. Let $E^* = \bigcap O_n$. Then $E^*$ is a $G_\delta$ and $E \subset E^*$.

Let $x \in E^*$. Then for each $n$ there is a neighborhood $U$ of $x$ with $U \cap E \neq \emptyset$ and diam $f[E \cap U]$ less than $1/n$. Choose $x_n \in U$. Then the sequence $\langle x_k \rangle$ has the property that $\sigma(f(x_k), f(x_l)) \leq 1/n$ for $k$, $l \geq n$. Hence $\langle f(x_k) \rangle$ is a Cauchy sequence and so converges to a point $y \in Y$. Define $f^*(x)$ to be $y$. Then the value of $f^*(x)$ is readily seen to be independent of the choice of the sequence $\langle x_k \rangle$. Thus $f^*(x) = f(x)$ for $x \in E$, and $f^*$ is continuous. ▮

**35. Proposition:** *Let $E$ be a dense subset of a Hausdorff topological space, and suppose that $E$ is homeomorphic to a complete metric space $(Y, \sigma)$. Then $E$ is a $G_\delta$.*

***Proof:*** Let $f$: $E \to Y$ be the homeomorphism and let $g = f^{-1}$. Let $f^*$ be the extension of $f$ to $E^*$ given by Proposition 34.

Then $g \circ f$: $E \to X$ is the identity map and $\mathrm{id}_E$ on $E$. Since $E$ is dense in $X$, $E^* \subset \bar{E}$, and so $g \circ f^*$ and $\mathrm{id}_{E^*}$ are continuous maps of $E^*$ into $X$ that agree on the dense subset $E$ of $E^*$. Thus $g \circ f^* = \mathrm{id}_{E^*}$ (cf. Problem 44 for $X$ metric and Problem 8.30 for $X$ Hausdorff).

Thus

$$f \circ (\mathrm{id}_{E^*}) = f \circ (g \circ f^*) = (f \circ g) \circ f^*$$
$$= \mathrm{id}_Y \circ f^* = f^*.$$

Hence $E^* = \operatorname{range} f^* \subset \operatorname{range} f = E$, and so $E^* = E$. ∎

The following Corollary is a converse of Proposition 33.

**36. Corollary:** *Let $E$ be a subset of a metric space $X$, and suppose that $E$ is homeomorphic to a complete metric space $Y$. Then $E$ is a $G_\delta$.*

***Proof:*** By Proposition 35 the set $E$ is a $G_\delta$ relative to the metric space $\bar{E}$. But $\bar{E}$ is a closed subset of a metric space and hence a $G_\delta$ in $X$. Let $E = \bigcap U_n$ open in $X$ and $E = \bigcap (O_n \cap \bar{E})$ with $O_n \cap \bar{E}$ open in $\bar{E}$. Then $E = \bigcap (O_n \cap U_n)$, and so $E$ is a $G_\delta$. ∎

## Problems

**44.** Show that the function $\sigma$ defined in the proof of Proposition 33 is a metric and is equivalent to $\rho$. Is it uniformly equivalent?

**45.** Let $A \subset B \subset \bar{A}$ be subsets of a metric space and let $g$ and $h$ be continuous maps of $B$ into a metric space $X$. If $g(u) = h(u)$ for all $u \in A$, then $g \equiv h$.

**46.** The purpose of this problem is to show that every subset $E \subset (0, 1)$ that is a dense $G_\delta$ and whose complement is dense is homeomorphic to the irrationals on $[0, 1]$ and to the product space $\mathbf{N}^\omega$ of a countable number of copies of the natural numbers.

**a.** Let $I$ be an open interval and $E_I$ a $G_\delta$ which is dense in $I$ and whose complement is dense in $I$. Let $\sigma$ be an equivalent metric for $E_I$ which makes $E_I$ complete. Then given $\epsilon > 0$, there are a countably infinite number of subintervals $I_1, \ldots, I_k, \ldots$ of $I$ that are pairwise disjoint, whose union contains $E_I$, and such that the diameter of $E_I \cap I_k$ in the metric $\sigma$ is less than $\epsilon$.

**b.** Starting with $E$, find disjoint open intervals $I_1, \ldots, I_j, \ldots$ satisfying conditions of part (a) with $\epsilon = \frac{1}{2}$. For $E_{I_j} = E \cap I_j$ repeat the process with $\epsilon = \frac{1}{4}$, getting intervals $I_{j,1}, \ldots, I_{j,k}, \ldots$. Continuing, at the $n$-th stage getting intervals $I_{j,k,\ldots,l}$ with $\epsilon = 2^{-n}$. Show that for each $x \in E$ there is a unique sequence of integers $k_1, \ldots, k_j, \ldots$ such that for each $n$ we have

$$x \in I_{k_1, \ldots, k_n}.$$

c. Show that for each sequence of integers $k_1, k_2, \ldots, k_j, \ldots$ there is a unique $x \in E$ such that for any $n$ we have

$$x \in I_{k_1, \ldots, k_n}.$$

d. Let $\mathbf{N}^\omega$ be the space of infinite sequences of integers, and make $\mathbf{N}$ into a metric space by setting $\rho(i, j) = \delta_{ij}$. Let $\tau$ be the product metric on $\mathbf{N}^\omega$ (Problem 24). Show that the correspondence $h$ between $\mathbf{N}^\omega$ and $E$ given by parts (b) and (c) is a uniform homeomorphism between $(\mathbf{N}^\omega, \tau)$ and $(E, \sigma)$.

e. Show that any dense $G_\delta$ in $(0, 1)$ whose complement is dense is homeomorphic to the set $\mathscr{I}$ of irrationals in $(0, 1)$.

## 10 The Ascoli–Arzelà Theorem

It is often useful in analysis to know conditions under which a sequence of functions has a subsequence that is convergent in some sense. The following notion plays a central role in such questions: A family $\mathfrak{F}$ of functions from a metric space $X$ to a metric space $\langle Y, \sigma\rangle$ is called **equicontinuous** at the point $x \in X$ if given $\epsilon > 0$ there is an open set $O$ containing $x$ such that $\sigma[f(x), f(y)] < \epsilon$ for all $y$ in $O$ and all $f \in \mathfrak{F}$. The family is said to be equicontinuous on $X$ if it is equicontinuous at each point $x$ in $X$. We are going to show that if $\langle f_n\rangle$ is a sequence from an equicontinuous family of functions, and if at each point $x$ of $X$ there is a convergent subsequence of $\langle f_n(x)\rangle$, then $\langle f_n\rangle$ has a subsequence which converges uniformly on each compact subset of $X$. We begin with several lemmas.

**37. Lemma:** *Let $\langle f_n\rangle$ be a sequence of mappings of a countable set $D$ into a metric space $Y$ such that for each $x \in D$ the closure of the set $\{f_n(x)\colon 0 \le n < \infty\}$ is compact. Then there is a subsequence $\langle f_{n_k}\rangle$ that converges for each $x$ in $D$.*

***Proof:*** Let $D = \{x_k\}$. By the sequential compactness of the closure of $\{f_n(x_1)\colon 0 \le n < \infty\}$ we can pick a subsequence $\langle f_{1n}\rangle$ of $\langle f_n\rangle$ such that $\langle f_{1n}(x_1)\rangle$ converges. We can now pick a subsequence $\langle f_{2n}\rangle$ of $\langle f_{1n}\rangle$ such that $\langle f_{2n}(x_2)\rangle$ converges. Continuing in this fashion, we obtain a subsequence $\langle f_{jn}\rangle$ convergent on $x_j$. Consider the "diagonal" sequence $\langle f_{nn}\rangle$. We have $\langle f_{nn}\rangle_{n=j}^{\infty}$ a subsequence of $\langle f_{jn}\rangle$, and so $\langle f_{nn}(x_j)\rangle$ converges. ▮

**38. Lemma:** *Let $\langle f_n\rangle$ be an equicontinuous sequence of mappings from a metric space $X$ to a complete metric space $Y$. If the sequences*

*$\langle f_n(x)\rangle$ converge for each point $x$ of a dense subset $D$ of $X$, then $\langle f_n\rangle$ converges at each point of $X$, and the limit function is continuous.*

***Proof:*** Given $x$ in $X$ and $\epsilon > 0$, we can find an open set $O$ containing $x$ such that $\sigma[f_n(x), f_n(y)] < \epsilon/3$ for all $y$ in $O$. Since $D$ is dense, there must be a point $y \in D \cap O$, and since $\langle f_n(y)\rangle$ converges, it must be a Cauchy sequence, and we may choose $N$ so large that

$$\sigma[f_n(y), f_m(y)] < \epsilon/3$$

for all $m, n \geq N$. Then

$$\begin{aligned}\sigma[f_n(x), f_m(x)] &\leq \sigma[f_n(x), f_n(y)] + \sigma[f_n(y), f_m(y)] + \sigma[f_m(x), f_m(y)] \\ &< \epsilon\end{aligned}$$

for all $m, n \geq N$. Thus $\langle f_n(x)\rangle$ is a Cauchy sequence and converges by the completeness of $Y$.

Let $f(x) = \lim f_n(x)$. To see that $f$ is continuous at $x$, let $\epsilon > 0$ be given. By equicontinuity there is an open set $O$ containing $x$ such that $\sigma[f_n(x), f_n(y)] < \epsilon$ for all $n$ and all $y$ in $O$. Hence for all $y$ in $O$ we have

$$\sigma[f(x), f(y)] = \lim \sigma[f_n(x), f_n(y)] \leq \epsilon,$$

and $f$ is continuous at $x$. ▮

**39. Lemma:** *Let $K$ be a compact metric space and $\langle f_n\rangle$ an equicontinuous sequence of functions to a metric space $Y$ that converges at each point of $K$ to a function $f$. Then $\langle f_n\rangle$ converges to $f$ uniformly on $K$.*

***Proof:*** Choose $\epsilon > 0$. By equicontinuity each $x$ in $K$ is contained in an open set $O_x$ such that $\sigma[f_n(x), f_n(y)] < \epsilon/3$ for all $y$ in $O_x$ and all $n$. From this it follows that also $\sigma[f(x), f(y)] \leq \epsilon/3$ for all $y$ in $O_x$.

By the compactness of $K$ there is a finite collection $\{O_{x_1}, \ldots, O_{x_k}\}$ of these sets which covers $K$. Choose $N$ so large that for all $n \geq N$ we have $\sigma[f_n(x_i), f(x_i)] < \epsilon/3$ for each $x_i$ corresponding to this finite collection. Then for any $y$ in $K$ there is an $i \leq k$ such that $y \in O_{x_i}$. Hence

$$\begin{aligned}\sigma[f_n(y), f(y)] &\leq \sigma[f_n(y), f_n(x_i)] + \sigma[f_n(x_i), f(x_i)] + \sigma[f(y), f(x_i)] \\ &< \epsilon\end{aligned}$$

for $n \geq N$. Thus $\langle f_n\rangle$ converges to $f$ uniformly on $K$. ▮

These three lemmas, taken together with the fact that a compact subset of a metric space is complete, imply the following theorem, which is variously attributed to Arzelá or Ascoli.

**40. Theorem (Ascoli–Arzelá):** *Let $\mathcal{F}$ be an equicontinuous family of functions from a separable space $X$ to a metric space $Y$. Let $\langle f_n \rangle$ be a sequence in $\mathcal{F}$ such that for each $x$ in $X$ the closure of the set $\{f_n(x): 0 \leq n < \infty\}$ is compact. Then there is a subsequence $\langle f_{n_k} \rangle$ that converges pointwise to a continuous function $f$, and the convergence is uniform on each compact subset of $X$.*

**41. Corollary:** *Let $\mathcal{F}$ be an equicontinuous family of real-valued functions on a separable space $X$. Then each sequence $\langle f_n \rangle$ in $\mathcal{F}$ which is bounded at each point (of a dense subset) has a subsequence $\langle f_{n_k} \rangle$ that converges pointwise to a continuous function, the convergence being uniform on each compact subset of $X$.*

## Problems

**47.** Let $X$ be a metric space. Let $\langle f_n \rangle$ be a sequence of continuous functions from $X$ to a metric space $Y$ which converge to a function $f$ uniformly on each compact subset $K$ of $X$. Then $f$ is continuous.

**48.** Let $X$ be a separable, locally compact metric space, and $\langle Y, \sigma \rangle$ any metric space. Show that:

**a.** There is a countable collection $\{O_n\}$ of open subsets of $X$ such that $\bar{O}_n$ is compact and $X = \bigcup O_n$.

**b.** The set of functions from $X$ into $Y$ becomes a metric space if we define

$$\sigma^*(f, g) = \sum 2^{-n} \sigma_n^*(f, g),$$

where

$$\sigma_n^*(f, g) = \sup_{\bar{O}_n} \frac{\sigma[f(x), g(x)]}{1 + \sigma[f(x), g(x)]}.$$

**49.** Let $\mathcal{F}$ be an equicontinuous family of functions from $X$ to $Y$, and let $\mathcal{F}^+$ be the family of all pointwise limits of functions in $\mathcal{F}$, that is of $f$ for which there is a sequence $\langle f_n \rangle$ from $\mathcal{F}$ such that $f(x) = \lim f_n(x)$ for each $x \in X$. Show that $\mathcal{F}^+$ is also an equicontinuous family of functions.

**50.** A real-valued function $f$ on $[0, 1]$ is said to be Hölder continuous of order $\alpha$ if there is a constant $C$ such that $|f(x) - f(y)| \le C|x - y|^\alpha$. Define $\|f\|_\alpha = \max|f(x)| + \sup|f(x) - f(y)|/|x - y|^\alpha$.

Show that for $0 < \alpha \le 1$, the set of functions with $\|f\|_\alpha \le 1$ is a compact subset of $C[0, 1]$.

**51.** Let $\mathcal{F}$ be the family of functions that are holomorphic on the unit disk $\Delta = \{z: |z| < 1\}$ with $|f(z)| \le 1$.

**a.** Show that $\mathcal{F}$ is equicontinuous. [Use Cauchy's formula to estimate $|f(z) - f(z')|$.]

**b.** Show that any sequence $\langle f_n \rangle$ of functions in $\mathcal{F}$ has a subsequence $\langle f_{n_j} \rangle$ that converges uniformly on each compact subset of $\Delta$ to a function $f$ holomorphic on $\Delta$.

**c.** Prove *Osgood's Theorem:* Let $f_n$ be a sequence of holomorphic functions on $\Delta$ such that $f_n(z) \to f(z)$ for each $z \in \Delta$. Then there is a dense open set $O \subset \Delta$ on which $f$ is holomorphic. [Use the uniform boundedness principle.]

# 8 Topological Spaces

## 1 Fundamental Notions

In Chapter 7 we discussed properties of metric spaces and found that a number of theorems depended only on the properties of open and closed sets. In the present chapter we study spaces in which the notion of an open set is fundamental and other notions are defined in terms of it. Such spaces are called topological spaces and are more general than metric spaces. Perhaps you ask: Why not stick to metric spaces? It is true that metric spaces are simpler, but there are many examples of spaces of functions where certain topological notions have a natural meaning not consistent with the topological concepts derived from any metric that might be put on the space. Important examples are given by the weak topologies in Banach spaces. We proceed with a formal definition.

**Definition:** *A topological space* $\langle X, \mathfrak{T} \rangle$ *is a nonempty set* $X$ *of points together with a family* $\mathfrak{T}$ *of subsets (which we shall call open) possessing the following properties:*

i. $X \varepsilon \mathfrak{T}, \emptyset \varepsilon \mathfrak{T}$.
ii. $O_1 \varepsilon \mathfrak{T}$ *and* $O_2 \varepsilon \mathfrak{T}$ *imply* $O_1 \cap O_2 \varepsilon \mathfrak{T}$.
iii. $O_\alpha \varepsilon \mathfrak{T}$ *implies* $\bigcup_\alpha O_\alpha \varepsilon \mathfrak{T}$.

*The family* $\mathfrak{T}$ *is called a* ***topology*** *for the set* $X$.

The properties in this definition are all satisfied by open sets in a metric space $\langle X, \rho \rangle$, and hence to each metric space $\langle X, \rho \rangle$ we can associate a topological space $\langle X, \mathfrak{J} \rangle$, where $\mathfrak{J}$ is the family of open sets in $\langle X, \rho \rangle$. A topological space which is associated in this manner with some metric space is called *metrizable*, and the metric $\rho$ is said to be a metric for the topological space. From a logical standpoint the distinction between a metric space and its associated topological space is essential, since different metric spaces can give rise to the same topological space. Two such metric spaces are of course equivalent, and we often disregard the distinction between a metric space and its associated topological space in cases where no confusion is likely to arise. In some cases we shall not trouble even to distinguish between a topological space $\langle X, \mathfrak{J} \rangle$ and the set $X$ of its points, using $X$ to denote both. It is to be remembered, however, that metric and topological spaces are couples, and in many cases it is necessary to express this fact explicitly.

Given any set $X$ of points, there are always two topologies that can be defined on $X$. One is the trivial topology in which the only open sets are $\emptyset$ and $X$. A second possible topology is the discrete topology: Every subset of $X$ is an open set.

In terms of the notion of open set we may define other topological properties, for example: A subset $F$ of $X$ is called **closed** if $\tilde{F}$ is open.

**1. Proposition:** *The sets $\emptyset$ and $X$ are closed. The union of any two closed sets is closed. The intersection of any collection of closed sets is closed.*

For an arbitrary set $E$ the intersection of all the closed sets containing $E$ is again a closed set called the *closure* of $E$ and denoted by $\bar{E}$. It is the smallest closed set containing $E$. A point $x \in X$ is said to be a *point of closure* of $E$ if every open set $O$ containing $x$ meets $E$, i.e., has a nonempty intersection with $E$.

The union of all the open sets contained in $E$ is again an open set called the *interior* of $E$ and denoted by $E^\circ$. A point $x$ is said to be an *interior point* of $E$ if there is an open set with $x \in O \subset E$. The following proposition, whose proof is left to the reader, lists some of the properties of closure and interior.

**2. Proposition:** *We have $E \subset \bar{E}$, $\bar{\bar{E}} = \bar{E}$, $\overline{A \cup B} = \bar{A} \cup \bar{B}$, and a set $F$ is closed if and only if $F = \bar{F}$. Also, $\bar{E}$ is the set of points of closure of*

*$E$. Moreover, $E^\circ \subset E$, $E^{\circ\circ} = E^\circ$, $(A \cap B)^\circ = A^\circ \cap B^\circ$ and $E^\circ$ is the set of interior points of $E$. Finally, $(\tilde{E})^\circ = \sim(\bar{E})$.*

If $A$ is a subset of a topological space $\langle X, \mathfrak{T}\rangle$, we can define a topology $\mathcal{S}$ for $A$ by taking for $\mathcal{S}$ those subsets $B$ of $A$ for which there is a set $O \varepsilon \mathfrak{T}$ such that $B = A \cap O$. We call $\mathcal{S}$ the topology inherited from $\mathfrak{T}$ and call the topological space $\langle A, \mathcal{S}\rangle$ a subspace of $\langle X, \mathfrak{T}\rangle$. This is consistent with our usage concerning metric spaces.

A sequence $\langle x_n\rangle$ in a topological space is said to converge to the point $x$, or to have the limit $x$, if given any open set $O$ containing $x$ there is an integer $N$ such that $x_n \varepsilon O$ for all $n \geq N$. Similarly, a sequence $\langle x_n\rangle$ is said to have $x$ as a cluster point if given any open set $O$ containing $x$ and any integer $N$ there is an integer $n \geq N$ such that $x_n \varepsilon O$. Thus if $\langle x_n\rangle$ has a subsequence converging to $x$, then $x$ is a cluster point of $\langle x_n\rangle$. The converse of this statement is not always true in an arbitrary topological space.

In view of Proposition 7.7 we may give a definition of a continuous function on a topological space that agrees with the usual concept of continuity in the case in which the topological space is also a metric space.

**Definition:** *A mapping $f$ of a topological space $\langle X, \mathfrak{T}\rangle$ into a topological space $\langle Y, \mathcal{S}\rangle$ is said to be **continuous** if the inverse image of every open set is open, that is, if $O \varepsilon \mathcal{S} \Rightarrow f^{-1}[O] \varepsilon \mathfrak{T}$.*

It should be noted that if $f$ is a continuous function on a space $X$, then the restriction $f_1$ of $f$ to a subspace $A$ of $X$ is a continuous function on $A$, for $f_1^{-1}[O] = A \cap f^{-1}[O]$ and must be open for open $O$ by the continuity of $f$ and the definition of open sets in $A$.

We say that a map $f$ from $X$ to $Y$ is *continuous at a point* $x_0 \varepsilon X$ if, given any open set $O$ in $Y$ containing $f(x_0)$, there is an open set $U$ in $X$ containing $x_0$ such that $f[U] \subset O$. A map from $X$ to $Y$ is continuous if and only if it is continuous at each point of $X$.

We sometimes construct continuous functions by piecing together two continuous functions. The following proposition gives conditions for this.

**3. Proposition:** *Let the subset $A$ of a topological space $X$ be the union of two sets $A_1$ and $A_2$ both of which are closed (or both of which are open). If $f$ is a map of $A$ into a topological space $Y$ such that the restrictions $f|A_1$ and $f|A_2$ are each continuous, then $f$ is continuous.*

**Definition:** *A homeomorphism between two topological spaces is a one-to-one continuous mapping of $X$* **onto** *$Y$ for which $f^{-1}$ is continuous. The spaces $X$ and $Y$ are said to be homeomorphic if there is a homeomorphism between them.*

From an abstract point of view two homeomorphic topological spaces are indistinguishable, the homeomorphism amounting to a mere relabeling of the points of one set by the points of a second set. Thus the concept of homeomorphism plays the same role for topological spaces that isometry plays for metric spaces and isomorphism plays for algebraic systems.

Suppose that $\mathfrak{I}$ and $\mathcal{S}$ are two topologies for the same set $X$. Then $\mathcal{S}$ is said to be *stronger* than $\mathfrak{I}$ if $\mathcal{S} \supset \mathfrak{I}$. In this case we also say that $\mathfrak{I}$ is *weaker* than $\mathcal{S}$. Thus $\mathcal{S}$ is stronger than $\mathfrak{I}$ if and only if the identity mapping of $(X, \mathcal{S})$ into $(X, \mathfrak{I})$ is continuous. The trivial topology for a set $X$ is the weakest possible topology on $X$, while the discrete topology is the strongest possible topology. Sometimes the terms *finer* and *coarser* are used for *stronger* and *weaker*, respectively.

If $\mathcal{S}$ and $\mathfrak{I}$ are two topologies for a set $X$, then $\mathcal{S} \cap \mathfrak{I}$ is also a topology. In fact, if $\{\mathfrak{I}_\alpha\}$ is any collection of topologies, then $\bigcap_\alpha \mathfrak{I}_\alpha$ is a topology. Thus if $\mathcal{C}$ is any collection of subsets of $X$, then the intersection of all topologies containing $\mathcal{C}$ is a topology containing $\mathcal{C}$. This topology is the weakest topology such that all of the sets of $\mathcal{C}$ are open.

**4. Proposition:** *Let $X$ be a nonempty set of points and $\mathcal{C}$ any collection of subsets of $X$. Then there is a weakest topology $\mathfrak{I}$ that contains $\mathcal{C}$.*

## Problems

**1. a.** Given a set $X$, can you define a metric on $X$ so that the associated topological space is discrete? Trivial?

**b.** Let $X$ be a space with a trivial topology. Find all continuous mappings of $X$ into $\mathbf{R}$.

**c.** Let $X$ be a space with a discrete topology. Find all continuous mappings of $X$ into $\mathbf{R}$.

**2.** Prove Proposition 2.

**3.** Prove that a set $A \subset X$ is open if and only if, given $x \in A$, there is an open set $O$ such that $x \in O \subset A$.

**4.** Prove that a mapping of $X$ into $Y$ is continuous if and only if the inverse image of every closed set is closed.

**5.** Show that if $f$ is a continuous mapping of $X$ into $Y$ and $g$ a continuous mapping of $Y$ into $Z$, then $g \circ f$ is a continuous mapping of $X$ into $Z$.

**6.** Prove that the sum and product of two real-valued continuous functions are themselves continuous.

**7. a.** Let $F$ be a closed subset of a topological space and $\langle x_n \rangle$ a sequence of points from $F$. Show that, if $x$ is a cluster point of $\langle x_n \rangle$, then $x \in F$.

**b.** Show that, if $f$ is continuous and $x = \lim x_n$, then $\langle f(x_n) \rangle$ has the limit $f(x)$.

**c.** Show that, if $f$ is continuous and $x$ is a cluster point of $\langle x_n \rangle$, then $f(x)$ is a cluster point of $\langle f(x_n) \rangle$.

**8.** Kuratowski fourteen-subset problem:

**a.** Let $E$ be an arbitrary set in a topological space $X$. Show that at most 14 different sets can be obtained from $E$ by repeated use of complementation and closure. (This includes $E = \sim(\sim E)$.)

**b.** Give an example in $\mathbf{R}^2$ where there are 14 different sets coming from a suitable $E$. (One can give an example in $\mathbf{R}$ but it is easier to draw in $\mathbf{R}^2$.)

**9.** Show that a function $f$ from a topological space $X$ to a topological space $Y$ is continuous if and only if it is continuous at each point of $X$.

**10. a.** Prove Proposition 3.

**b.** Show that the proposition is not true if $A_1$ and $A_2$ are not required to be both open or both closed.

**c.** Show, however, that it is still true if $\overline{(A_1 \sim A_2)} \cap (A_2 \sim A_1) = \emptyset$ and $\overline{(A_2 \sim A_1)} \cap (A_1 \sim A_2) = \emptyset$.

## 2 Bases and Countability

A collection $\mathcal{B}$ of open subsets of a topological space $X$ is called a **base** for the topology $\mathfrak{I}$ of $X$ if for each open set $O$ in $X$ and each $x \in O$ there is a set $B \in \mathcal{B}$ such that $x \in B \subset O$. A collection $\mathcal{B}_x$ of open sets containing a point $x$ is called a base at $x$ if for each open set $O$ containing $x$ there is a $B \in \mathcal{B}_x$ such that $x \in B \subset O$. Thus a collection $\mathcal{B}$ of open sets is a base if and only if it contains a base at each point $x \in X$. If $X$ is a metric space, the balls form a base, and the balls centered at $x$ form a base at $x$.

If $\mathcal{B}$ is a base for the topology $\mathcal{T}$, then $O \ \varepsilon \ \mathcal{T}$ if and only if for each $x \ \varepsilon \ O$ there is a $B \ \varepsilon \ \mathcal{B}$ with $x \ \varepsilon \ B \subset O$. The 'only if' part of the statement follows from the definition of a base, while if for each $x$ in $O$ we have $x \ \varepsilon \ B \subset O$, then $O$ must be the union of those $B$ in $\mathcal{B}$ with $B \subset O$, and $O$ is open, since it is a union of open sets.

We often find it convenient to specify a topology for a set $X$ by specifying a base $\mathcal{B}$ of open sets and using the preceding criterion to define the open sets. Conditions on a collection $\mathcal{B}$ in order that it be a base for some topology are given by the following proposition:

**5. Proposition:** *A collection $\mathcal{B}$ of subsets of a set $X$ is a base for some topology on $X$ if and only if each $x$ in $X$ is contained in some $B$, and if $x \ \varepsilon \ B_1 \cap B_2$ then there is a $B_3 \ \varepsilon \ \mathcal{B}$ such that $x \ \varepsilon \ B_3 \subset B_1 \cap B_2$.*

***Proof:*** That these conditions are necessary follows from the definition of a base and the fact that $X$ and $B_1 \cap B_2$ must be open. Suppose now that $\mathcal{B}$ satisfies these conditions. If we then set $\mathcal{T} = \{O\colon (x \ \varepsilon \ O)(\exists B \ \varepsilon \ \mathcal{B})(x \ \varepsilon \ B \subset O)\}$, then $\emptyset \ \varepsilon$ , and the union of sets in $\mathcal{T}$ will be in $\mathcal{T}$, and the first condition on $\mathcal{B}$ implies that $X \ \varepsilon \ \mathcal{T}$. To show that the intersection of two sets $O_1$ and $O_2$ in $\mathcal{T}$ is itself in $\mathcal{T}$, let $x \ \varepsilon \ O_1 \cap O_2$. Then there are sets $B_1$ and $B_2$ in $\mathcal{B}$ such that $x \ \varepsilon \ B_1 \subset O_1$ and $x \ \varepsilon \ B_2 \subset O_2$. Let $B_3$ be a set in $\mathcal{B}$ with $x \ \varepsilon \ B_3 \subset B_1 \cap B_2$. Then $x \ \varepsilon \ B_3 \subset O_1 \cap O_2$, and it follows that $O_1 \cap O_2$ is open. ▮

The elements of a base $\mathcal{B}_x$ at $x$ are often referred to as neighborhoods of $x$. If no base is given, either explicitly or implicitly, then by a neighborhood at $x$ we simply mean an open set containing $x$. The advantage of using bases is that many topological properties can be determined by using only the elements of a base rather than all open sets. Examples are given in Problems 11 and 12.

We have taken the elements in a base to be open sets. Some authors, including Bourbaki and Kelley, only require a neighborhood of $x$ to be a set $N$ with $x$ in the interior of $x$ and a base $\mathcal{B}_x$ at $x$ to be a collection of neighborhoods of $x$ such that, given any open set $O$ containing $x$, there is an $N \ \varepsilon \ \mathcal{B}_x$ with $x \ \varepsilon \ N^\circ \subset O$.

It is sometimes convenient to define a topology in terms of a *subbase.* A collection $\mathcal{C}$ of open sets is called a subbase for a topology $\mathcal{T}$ if given $x \ \varepsilon \ X$ and $O \ \varepsilon \ \mathcal{T}$, there are sets $C_1, \ldots, C_n \ \varepsilon \ \mathcal{C}$ such that $x \ \varepsilon \ C_1 \cap \cdots \cap C_n \subset O$. If $\mathcal{C}$ is a subbase for $\mathcal{T}$, then the collection of all finite intersections of sets from $\mathcal{C}$ is a base for $\mathcal{T}$. If $\mathcal{C}$ is any collection of subsets of $X$, then $\mathcal{C}$ is a subbase for the weakest topology on $X$ for which the sets in $\mathcal{C}$ are all open.

A topological space is said to satisfy the *first axiom of countability* if there is a countable base at each point. Every metric space satisfies the first axiom of countability, since the balls centered at $x$ and having rational radii are countable in number and form a base at $x$. A space is said to satisfy the *second axiom of countability* if there is a countable base for the topology. Thus Proposition 7.6 states that a metric space satisfies the second axiom of countability if and only if it is separable.

## Problems

**11. a.** Let $\mathcal{B}$ be a base for the topological space $\langle X, \mathfrak{T}\rangle$. Then $x \varepsilon \bar{E}$ if and only if for every $B \varepsilon \mathcal{B}$ with $x \varepsilon B$, there is a $y \varepsilon B \cap E$.

**b.** Let $X$ satisfy the first axiom of countability. Then $x \varepsilon \bar{E}$ if and only if there is a sequence from $E$ that converges to $x$.

**c.** Let $X$ satisfy the first axiom of countability. Then $x$ is a cluster point of a sequence $\langle x_n \rangle$ from $X$ if and only if $\langle x_n \rangle$ has a subsequence that converges to $x$.

**12. a.** Show that $f$ is continuous if and only if it is continuous at each point of $X$.

**b.** Let $\mathcal{B}_x$ be a base at $x$ and $\mathcal{C}_y$ be a base at $y = f(x)$. Then $f$ is continuous at $x$ if and only if for each $C \varepsilon \mathcal{C}_y$ there is a $B \varepsilon \mathcal{B}_x$ such that $B \subset f^{-1}[C]$.

**13.** Let $\mathcal{C}$ be any collection of subsets of $X$. Let $\mathcal{B}$ consist of $X$ and all finite intersections of sets in $\mathcal{C}$. Show that $\mathcal{B}$ is a base for the weakest topology that contains $\mathcal{C}$.

**14.** Let $X$ be an uncountable set of points, and let $\mathfrak{T}$ consist of the empty set and all subsets of $X$ whose complement is finite. Show that $\mathfrak{T}$ is a topology for $X$ and that the space $\langle X, \mathfrak{T}\rangle$ does not satisfy the first axiom of countability.

**15.** Let $X$ be the set of real numbers, and let $\mathcal{B}$ be the set of all intervals of the form $[a, b)$. Show that $\mathcal{B}$ is the base of a topology $\mathfrak{T}$ for $X$. (This topology is called the *half-open interval topology*.) Show that $\langle X, \mathfrak{T}\rangle$ satisfies the first but not the second axiom of countability and that the rationals are dense in $X$. Is $\langle X, \mathfrak{T}\rangle$ metrizable?

**16.** A topological space is said to be a *Lindelöf* space or to have the *Lindelöf property* if each open cover of $X$ has a countable subcover. Show that if $X$ is second countable, it is *Lindelöf*.

**17.** Let $X_1 = \mathbf{N} \times \mathbf{N}$ and take $X = X_1 \cup \{\omega\}$, the set obtained by adding one more point. Thus $X$ consists of $\{\omega\}$ and all pairs $\langle j, k\rangle$ of natural numbers. For each sequence $s = \langle m_k \rangle$ of natural numbers define

$$B_{s,n} = \{\omega\} \cup \{\langle j, k\rangle : j \geq m_k \text{ all } k \geq m\}.$$

**a.** Show that the sets $B_{s,n}$ together with the sets $\{\langle j, k\rangle\}$ form a base for a topology on $X$.

**b.** Show that $\omega$ is a point of closure of $X_1$ even though no sequence $\langle x_n \rangle$ from $X_1$ has $\omega$ as a cluster point.

**c.** The space $X$ is separable but satisfies neither countability axiom.

**d.** Is $X$ a Lindelöf space?

## 3 The Separation Axioms and Continuous Real-Valued Functions

The properties of topological spaces are in general quite different from those of metric spaces, and it is often convenient to suppose that our topological space satisfies some additional conditions which are true in metric spaces. Consider the following set of conditions on a topological space:

$T_1$: *Given two distinct points $x$ and $y$, there is an open set that contains $y$ but not $x$.*

$T_2$: *Given two distinct points $x$ and $y$, there are disjoint open sets $O_1$ and $O_2$ such that $x \in O_1$ and $y \in O_2$.*

$T_3$: *In addition to $T_1$, given a closed set $F$ and a point $x$ not in $F$, there are disjoint open sets $O_1$ and $O_2$ such that $x \in O_1$ and $F \subset O_2$.*

$T_4$: *In addition to $T_1$, given two disjoint closed sets $F_1$ and $F_2$, there are disjoint open sets $O_1$ and $O_2$ such that $F_1 \subset O_1$ and $F_2 \subset O_2$.*

These are called separation axioms, and all are satisfied in a metric space. A topological space satisfying $T_1$ is called a **Tychonoff** space. A topological space which satisfies $T_2$ is called a **Hausdorff** space, one which satisfies $T_3$ is called a **regular** space, and one which satisfies $T_4$ is called a **normal** space. The following proposition tells us that the condition $T_1$ is equivalent to the statement that each set consisting of a single point is closed. With this in mind we see that $T_4 \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1$.

**6. Proposition:** *A topological space $X$ satisfies* $T_1$ *if and only if every set consisting of a single point is closed.*

***Proof:*** If each set $\{x\}$ is closed, given two distinct points $x$ and $y$, we may take $O = \sim\{x\}$. Then $O$ is an open set containing $y$ but not $x$. Suppose that $T_1$ holds. Each $y \in \sim\{x\}$ is contained in an open set $O \subset \sim\{x\}$. Thus the set $\sim\{x\}$ is the union of the open sets contained in it and so must be open. Hence $\{x\}$ is closed. ▮

Important consequences of normality are the following propositions, whose proofs are left to the reader (Problems 23 and 24).

**7. Urysohn's Lemma:** *Let $A$ and $B$ be disjoint closed subsets of a normal space $X$. Then there is a continuous real-valued function $f$ defined on $X$ such that $0 \leq f \leq 1$ on $X$ while $f \equiv 0$ on $A$ and $f \equiv 1$ on $B$.*

**8. Tietze's Extension Theorem:** *Let $X$ be a normal topological space, $A$ a closed subset of $X$, and $f$ a continuous real-valued function on $A$. Then there is a continuous real-valued function $g$ on $X$ such that $g(x) = f(x)$ for $x \in A$.*

The following theorem characterizes separable metric spaces. It can be proved using the concepts of the next section (see Problem 30).

**9. Urysohn Metrization Theorem:** *Every normal topological space satisfying the second axiom of countability is metrizable.*

If $X$ is any set of points and $\mathcal{F}$ is any collection of real-valued functions on $X$, there is always a weakest topology on $X$ such that every function in $\mathcal{F}$ is continuous, for let $\mathcal{C} = \{E: E = f^{-1}[O], f \in \mathcal{F}$ and $O$ an open subset of $\mathbf{R}\}$ and apply Proposition 5. This topology is called the **weak topology** generated (or induced) by $\mathcal{F}$. If the functions in $\mathcal{F}$ are all continuous in some topology $\mathfrak{J}$, then the weak topology generated by $\mathcal{F}$ is in general weaker than $\mathfrak{J}$. For these topologies to coincide, there must be enough continuous real-valued functions. One condition that guarantees this is the following: Given a closed set $F$ and a point $x \notin F$, there is a function $f \in \mathcal{F}$ with $f(x) = 1$ and $f \equiv 0$ on $F$. If this condition is satisfied when $\mathcal{F}$ is the space $C(X)$ of all continuous real-valued functions on $X$, then we say that $X$ is **completely regular,** provided that $X$ satisfies $T_1$. It follows from Urysohn's lemma that each normal space is completely regular, and it is easy to see that complete regularity implies regularity. For this reason the condition defining complete regularity is sometimes called $T_{3\frac{1}{2}}$.

Properties of very general topological spaces can be quite bizarre, but such spaces do not seem to be much needed in analysis. The only topological spaces, beyond the metrizable ones, that I have found useful for analysis are the locally compact Hausdorff spaces and the topological vector spaces. All of these spaces are completely regular, as we shall see in later chapters. Algebraic geometers, however, make use of the Zariski topology on affine or projective space (see Problem 29), a topology which gives us a compact $T_1$ space that is not Hausdorff.

## Problems

**18. a.** Show that every metric space is Hausdorff.

**b.** Show that every metric space is normal. [Hint: If $F_1$ and $F_2$ are disjoint closed sets, let $O_1 = \{x: \rho(x, F_1) < \rho(x, F_2)\}$ and $O_2 = \{x: \rho(x, F_2) < \rho(x, F_1)\}$.]

**19.** Let $X$ consist of the numbers $[0, 1]$ and an element $0'$, and take as a basis for a topology the sets $(\alpha, \beta)$, $[0, \beta)$, $(\alpha, 1]$, and $\{0'\} \cup (0, \beta)$. Show that these are a base for a topology on $X$ that is $T_1$ but not Hausdorff.

**20.** Let $f$ be a real-valued function on a topological space $X$. Show that $f$ is continuous iff for each real number $a$ the sets $\{x: f(x) < a\}$ and $\{x: f(x) > a\}$ are open. Show that $f$ is continuous iff for each real number $a$ the set $\{x: f(x) > a\}$ is open and the set $\{x: f(x) \geq a\}$ is closed.

**21.** If $f$ and $g$ are continuous real-valued functions on a topological space $X$, then the functions $f + g$, $fg$, $f \vee g$, and $f \wedge g$ are continuous. [Here, as usual, $(f \vee g)(x) = \max f(x), g(x)$.]

**22.** Let $\langle f_n \rangle$ be a sequence of continuous functions from a topological space $X$ to a metric space $Y$. If $\langle f_n \rangle$ converges uniformly to a function $f$, then $f$ is continuous.

**23. a.** Show that a Hausdorff space is normal if and only if given a closed set $F$ and an open set $O$ containing $F$ there is an open set $U$ such that $F \subset U$ and $\bar{U} \subset O$.

**b.** Let $F$ be a closed subset of a normal space contained in an open set $O$. By repeating the result in part (a) indefinitely, show that it is possible to construct a family $\{U_r\}$ of open sets, one corresponding to each rational in $(0, 1)$ of the form $r = p \cdot 2^{-n}$, such that $F \subset U_r \subset O$ and $\bar{U}_r \subset U_s$ for $r < s$.

**c.** Let $\{U_r\}$ be the family constructed in (b) with $U_1 = X$. Let $f$ be the real-valued function on $X$ defined by $f(x) = \inf \{r: x \in U_r\}$. Then $f$ is a continuous function, $0 \le f \le 1$, with $f \equiv 0$ on $F$ and $f \equiv 1$ on $\tilde{O}$.

**d.** Let $X$ be a Hausdorff space. Prove that $X$ is normal if and only if for every pair of disjoint closed sets $A$ and $B$ on $X$ there is a continuous real-valued function $f$ on $X$ such that $0 \le f \le 1, f \equiv 0$ on $A$ and $f \equiv 1$ on $B$.

**24.** Prove Tietze's Extension Theorem by using the following steps:

**a.** Let $h = f/(1 + |f|)$. Then $|h| < 1$.

**b.** Let $B = \{x: h(x) \le -\frac{1}{3}\}$, $C = \{x: h(x) \ge \frac{1}{3}\}$. Then by the Urysohn Lemma there is a continuous real-valued function $h_1$ on $X$ which is $-\frac{1}{3}$ on $B$ and $\frac{1}{3}$ on $C$, while $|h_1(x)| \le \frac{1}{3}$ for all $x \in X$. Clearly, $|h(x) - h_1(x)| < \frac{2}{3}$ for all $x \in A$.

**c.** By induction there is a continuous function $h_n$ on $X$ such that

$|h_n(x)| < 2^{n-1}/3^n$ for all $x \in X$ and $|h(x) - \sum_{i=1}^{n} h_i(x)| < 2^n/3^n$ for all $x \in A$.

**d.** The sequence $\langle h_n \rangle$ is uniformly summable to a continuous function $k$ on $X$, $|k| \le 1$ and $k = h$ on $A$.

**e.** There is a continuous function $\varphi$ on $X$ which is 1 on $A$ and 0 on $\{x: k(x) = 1\}$.

**f.** Set $g = \varphi k/(1 - \varphi k)$.

**25.** Let $\mathcal{F}$ be a family of real-valued functions on a set $X$. Show that a base for the weak topology on $X$ generated by $\mathcal{F}$ is given by the sets of the form $\{x: |f_i(x) - f_i(y)| < \epsilon$, for some $\epsilon > 0$, some $y \in X$, and some finite set $f_1, \ldots, f_n$ of functions in $\mathcal{F}\}$. Show that this topology is Hausdorff if and only if given any pair $\{x, y\}$ of distinct points in $X$ there is an $f \in \mathcal{F}$ with $f(x) \neq f(y)$.

**26.** Let $\mathcal{F}$ be a family of real-valued continuous functions on a topological space $\langle X, \mathfrak{J} \rangle$. Show that the weak topology generated by $\mathcal{F}$ is $\mathfrak{J}$ if for each closed set $F$ and each $x \notin F$ there is an $f \in \mathcal{F}$ with $f(x) = 1$ and $f \equiv 0$ on $F$.

**27.** Show that every completely regular space is regular.

**28.** Prove that every subset of a Hausdorff space is Hausdorff.

**29.** *Zariski Topology.* In $\mathbf{R}^n$ let $\mathcal{B}$ be the family of sets $\{x: p(x) \neq 0\}$, where $p$ is a polynomial in $n$ variables. Let $\mathfrak{J}$ be the family of all finite intersections $O = B_1 \cap \cdots \cap B_k$ from $\mathcal{B}$. Show that $\mathfrak{J}$ gives a topology for $\mathbf{R}^n$ which is $T_1$ and compact but not $T_2$.

**30.** Let $A \subset B \subset \bar{A}$ be subsets of a Hausdorff space, and let $f$ and $g$ be two continuous maps of $B$ into a topological space $X$ with $f(u) = g(u)$ for all $u \in A$. Then $f \equiv g$.

**31. a.** Verify that the proof given for Proposition 7.34 is valid for an arbitrary topological space $X$.

**b.** Show that the proof of Proposition 7.35 is valid when $X$ is an arbitrary Hausdorff space. Why is Hausdorff necessary?

## 4 Connectedness

A topological space $X$ is said to be *connected* if there do not exist two nonempty disjoint open sets $O_1$ and $O_2$ such that $X = O_1 \cup O_2$. Such a pair of open sets is called a *separation* of $X$. Since each set is the complement of the other, they are closed sets as well as open sets. Any pair of disjoint nonempty closed sets whose union is $X$ is a separation for $X$, since each of these sets must also be open. A space $X$ is connected if and only if the only subsets of $X$ that are both open and closed are the sets $\emptyset$ and $X$. A subset $E$ of $X$ is said to be connected when it is a connected space in the topology it inherits from $X$: Thus $E$ is connected if there do not exist open sets $O_1$ and $O_2$ in $X$, with $E \subset O_1 \cup O_2$ and $E \cap O_1 \cap O_2 = \emptyset$.

**10. Proposition:** *Let $f$ be a continuous mapping of a connected space $X$ onto a topological space $Y$. Then $Y$ is connected.*

***Proof:*** Let $O_1$ and $O_2$ be a separation of $Y$. Then $f^{-1}[O_1]$ and $f^{-1}[O_2]$ are disjoint open sets in $X$ whose union is $X$. Since $f$ is onto, neither $f^{-1}[O_1]$ nor $f^{-1}[O_2]$ is empty, and so this pair is a separation of $X$. Thus if $Y$ is not connected, $X$ is not connected, and the proposition follows by contraposition. ▮

The following theorem generalizes the Intermediate Value Theorem:

**11. Proposition:** *Let $f$ be a real-valued continuous function on a connected space $X$. Let $x$ and $y$ be two points in $X$ and $c$ a real number such that $f(x) < c < f(y)$. Then there is a $z \in X$ such that $f(z) = c$.*

***Proof:*** If $f$ does not assume the value $c$, then $f^{-1}[(-\infty, c)]$ and $f^{-1}[(c, \infty)]$ are disjoint open sets whose union is $X$. They are nonempty, since $x$ belongs to the first and $y$ to the second. Thus $X$ is not connected. ▮

**12. Proposition:** *A subset $E$ of* **R** *is connected if and only if it is either an interval or a single point.*

Let $x_0$ be a point of a topological space $X$. By the component $C$ of $X$ containing $x_0$ we mean the union of all connected sets containing $x_0$. It is connected (by Problem 32) and closed (by Problem 33). If two components of $X$ have a point in common, they coincide. Thus $X$ is the disjoint union of its components.

A space $X$ is said to be **locally connected** if we can find a base for $X$ consisting of connected sets. The components of a locally connected space are open. A space may be connected but not locally connected (Problem 35).

## Problems

**32.** Let $\{C_\alpha\}$ be a collection of connected sets and suppose that any two of them have a point in common. Then their union $G = \bigcup C_\alpha$ is connected.

**33.** Let $A$ be a connected subset of a topological space, and suppose that $A \subset B \subset \bar{A}$. Then $B$ is connected.

**34. a.** Let $E$ be a connected subset of **R** having more than one point. Prove that $E$ is an interval. [If $x$ and $y$ are in $E$, $x < y$, then $[x, y] \subset E$. Let $a = \inf E$, $b = \sup E$. Then $(a, b) \subset E \subset [a, b]$.]

**b.** Prove that an interval in **R** is connected. [Let $I = (a, b)$ and let $O$ be a subset of $I$ that is both open and closed in $I$. Show that sup $\{y\colon (x, y) \subset O\} = b$, and use Problem 33 to take care of nonopen intervals.]

**35.** A space $X$ is said to be arcwise connected if given two points $x$ and $y$ in $X$ there is a continuous map $f$ of $[0, 1]$ into $X$ such that $f(0) = x$ and $f(1) = y$.

**a.** Show that an arcwise connected space is connected.

**b.** In the plane $\mathbf{R}^2$ consider the subspace

$$X = \{\langle x, y\rangle\colon x = 0,\ -1 \le y \le 1\} \cup \{\langle x, y\rangle\colon y = \sin 1/x,\ 0 < x \le 1\}.$$

Show that $X$ is connected but not arcwise connected.

**c.** Show that each connected open set $G$ in $\mathbf{R}^n$ is arcwise connected. [Let $x \in G$ and let $H$ be the set of points $G$ that can be connected to $x$ by a polygonal arc. Then $H$ is open and closed in $G$.]

**36.** Show that each component of a locally connected space is open.

**37.** Show that the set $X$ in Problem 35b is connected but not locally connected.

## 5 Products and Direct Unions of Topological Spaces

If $\langle X, \mathfrak{I}\rangle$ and $\langle Y, \mathcal{S}\rangle$ are two topological spaces, we define a topology on the product $X \times Y$ by taking as a base the collection of all sets of the form $O_1 \times O_2$, where $O_1 \in \mathfrak{I}$ and $O_2 \in \mathcal{S}$. This is called the **product topology** for $X \times Y$. If $X$ and $Y$ are metric spaces, then the product topology agrees with the topology given by the product metric. If $\langle X_\alpha, \mathfrak{I}_\alpha\rangle$ is any indexed family of topological spaces, we define the product topology on $\bigtimes_\alpha X_\alpha$ by taking as a base all sets of the form $\bigtimes_\alpha O_\alpha$, where $O_\alpha \in \mathfrak{I}_\alpha$ and $O_\alpha = X_\alpha$ except for a finite number of $\alpha$. If the $X_\alpha$ are all the same space $X$ and indexed by an index set $A$, we write $X^A$ for $\bigtimes_\alpha X_\alpha$.

If $\langle X_\alpha, \mathfrak{I}_\alpha\rangle$ is a collection of topological spaces and $Y$ their product, we define for each $\alpha$ a mapping $\pi_\alpha$ (called a *projection*) of $Y$ into $X_\alpha$ by letting $\pi_\alpha(x)$ be the $\alpha$-th coordinate of $x$. Each $\pi_\alpha$ is continuous, and the product topology in $Y$ is the weakest topology such that each $\pi_\alpha$ is continuous.

If $A$ is countable and $X$ is metrizable, then $X^A$ is metrizable. Since only the number of elements in $A$ is important in determining $X^A$, we usually write $X^\omega$ (or $X^{\mathbf{N}}$) for a countable product. If we denote the discrete space with two elements by **2**, then $\mathbf{2}^\omega$ is homeomorphic to the Cantor set. If we use $\mathbf{N}$ to denote not only a countable set but also a countable set with discrete topology, then $\mathbf{N}^\omega$ or $\mathbf{N}^{\mathbf{N}}$ is a topological space which is homeomorphic to the set of irrational numbers.

If $I = [0, 1]$, then $I^A$ is called a *cube*. The cube $I^\omega$ is metrizable, and is called the *Hilbert cube*. Let $X$ be any set and $\mathcal{F}$ a family of functions $f$ on $X$ with $0 \le f \le 1$ and such that for any two distinct points $x$ and $y$ in $X$ there is an $f \in \mathcal{F}$ with $f(x) \ne f(y)$. Then $\mathcal{F}$ can be identified with a subset of $I^X$ if we let each $f$ correspond to the element whose $x$-th coordinate is $f(x)$. The mapping of $\mathcal{F}$ into $I$ which takes $f$ into $f(x)$ is simply the projection $\pi_x$ restricted to the image of $\mathcal{F}$. The topology that $\mathcal{F}$ inherits as a subspace of $I^X$ is called the *topology of pointwise convergence*.

On the other hand, $X$ can be identified with a subset of $I^{\mathcal{F}}$ by letting each $x$ correspond to the element whose $f$-th coordinate is $f(x)$. The topology of $X$ as a subspace of $I^{\mathcal{F}}$ is the weak topology generated by $\mathcal{F}$. If $X$ is a topological space and each $f$ in $\mathcal{F}$ is continuous, then the mapping of $X$ onto its image in $I^{\mathcal{F}}$ is continuous, and, if $\mathcal{F}$ has the property that for each closed subset $F \subset X$ and each $x \notin F$ there is an $f \varepsilon \mathcal{F}$ with $f(x) = 1$ and $f \equiv 0$ on $F$, then $X$ is homeomorphic with its image in $I^{\mathcal{F}}$.

If $\langle X, \mathfrak{I}\rangle$ and $\langle Y, \mathfrak{S}\rangle$ are two topological spaces with $X$ and $Y$ disjoint, we define a topology on the union $Z = X \cup Y$ by taking as our open sets all sets $O \subset Z$ for which $O \cap X \varepsilon \mathfrak{I}$ and $O \cap Y = \mathfrak{S}$. We call the space $\mathfrak{I}$ with this topology the *direct union* of $X$ and $Y$ and denote it by $X \mathring{\cup} Y$. If $\langle X_\alpha, \ _\alpha\rangle$ is any indexed family of topological spaces with $X_\alpha \cap X_\beta = \emptyset$ for $\alpha \neq \beta$, we define their direct union $Z = \mathring{\bigcup} X_\alpha$ to be the union of the $X_\alpha$ with $O \subset Z$ defined to be open if $O \cap X_\alpha \varepsilon \mathfrak{I}_\alpha$ for each $\alpha$. Each space $X_\alpha$ is referred to as a *direct summand.*[1] If the sets $X_\alpha$ are not all disjoint, we may take $X_\alpha^* = X_\alpha \times \{\alpha\}$ and consider the direct sum $\mathring{\bigcup} X_\alpha^*$.

As we shall see in the problems, the topological properties of a direct union depend only on what is happening in each of the direct summands: There is no coupling between what is happening in different summands. Thus, if we can show that a topological space is a direct union, then we can study its properties by studying those of the individual summands.

Suppose that a topological space $X$ is the union of a disjoint collection of open sets $\{X_\alpha\}$. Then $X$ is the direct union of the $X_\alpha$. To see that $X$ and $\mathring{\bigcup} X_\alpha$ have the same topology, we observe that, if $O$ is an open set of $X$, then $O \cap X_\alpha$ is open for each $\alpha$, and so $O$ is an open subset of $\mathring{\bigcup} X_\alpha$. Conversely, if $O$ is an open subset of $\mathring{\bigcup} X_\alpha$, then $O = \bigcup (O \cap X_\alpha)$, and each $O \cap X_\alpha$ is an open subset of $X_\alpha$.

## Problems

**38.** Let $Z = \mathring{\bigcup} X_\alpha$.

**a.** Show that map $f\colon Z \to Y$ is continuous iff each restriction $f|X_\alpha$ is continuous.

**b.** Show that a set $F \subset Z$ is closed if and only if $F \cap X_\alpha$ is closed for each $\alpha$.

**c.** Show that $Z$ is Hausdorff iff each of the spaces $X_\alpha$ is.

**d.** Show that $Z$ is normal iff each of the spaces $X_\alpha$ is.

**39. a.** Show that if a subset $X_1$ of a topological space $X$ is a direct summand (in some direct union), then $X_1$ is both open and closed.

**b.** If $X_1$ is a subset of $X$ that is both open and closed, then $X = X_1 \mathring{\cup} X_2$, where $X_2 = X \sim X_1$.

[1] This terminology is left over from the time when unions were called sums.

**40.** A property $P$ of topological spaces is said to be a *local property*, provided that a space $X$ has property $P$ whenever $X$ has a base each element of which has property $P$.

**a.** Show that the properties of regularity, complete regularity and being Tychonoff are local properties.

**b.** If $P$ is a local property and a space $X$ has $P$, then each open subset of $X$ has $P$.

**c.** Let $X = \mathring{\bigcup} X_\alpha$ be a direct union. Then $X$ has a local property $P$ if and only if each $X_\alpha$ has $P$.

**41.** Let $(X, \rho)$ be a metric space with an extended real-valued metric and $X_\alpha$ its parts, i.e. equivalence classes under the equivalence relation $\rho(x, y) < \infty$ (cf. Problem 7.3b). Show that $X = \bigcup X_\alpha$ is a direct union.

**42.** Show that the direct product of Hausdorff spaces is a Hausdorff space.

**43.** Prove that the collections taken for bases in defining product topologies satisfy the conditions of Proposition 3. Show that if $\langle X, \rho \rangle$ and $\langle Y, \sigma \rangle$ are two metric spaces, then the product topology on $X \times Y$ is the same as the topology induced by the product metric.

**44.** Show that $X^A$ is the set of all functions mapping $A$ into $X$ with a base for its topology given by open sets of the form $\{f: f(\alpha_1) \in O_1, f(\alpha_2) \in O_2, \dots, f(\alpha_n) \in O_n\}$, where $\{\alpha_1, \dots, \alpha_n\}$ is some finite subset of $A$ and $\{O_1, \dots, O_n\}$ a finite collection of open subsets of $X$. Prove that a sequence $\langle f_n \rangle$ converges to $f$ in $X^A$ if and only if $f_n(\alpha)$ converges to $f(\alpha)$ for each $\alpha$ in $A$.

**45.** Show that if $X$ is metrizable and $A$ is countable, then $X^A$ is metrizable. [Hint: $X$ can always be metrized by a bounded metric $\rho$. Define a metric $\sigma$ on $X^A$ by $\sigma(x, y) = \sum_{\alpha \in A} 2^{-n} \rho(x_\alpha, y_\alpha)$.]

**46.** Show that each projection $\pi_\alpha$ is continuous and that the product topology on $X^A$ is the weakest topology such that each $\pi_\alpha$ is continuous.

**47.** Show that $2^\omega$ is homeomorphic to the Cantor ternary set.

**48. a.** Show that the correspondence described in the text between a topological space $X$ and its image in $I^{\mathcal{F}}$ is a homeomorphism if $\mathcal{F}$ has the property that given a closed $F$ and $x \notin F$, there is an $f \in \mathcal{F}$ with $f[F] = 0$ and $f(x) = 1$.

**b.** Show that if $X$ is a normal space satisfying the second axiom of countability, then we can find a countable family $\mathcal{F}$ of continuous functions with the property in part (a).

**c.** Prove the Urysohn Metrization Theorem.

**49.** Prove that the product of connected spaces is connected.

## *6 Topological and Uniform Properties

Metric spaces have three types of properties: topological properties, which are preserved under homeomorphisms; uniform properties, which are preserved under uniform homeomorphisms; and metric properties, which are preserved under isometries. Metric properties are by their nature restricted to the category of metric spaces. Any property that can be defined solely in terms of open sets is a topological property (and conversely), and these properties usually extend to the category of topological spaces. Topological properties and notions include continuity, convergence, closure, etc.

The uniform properties, such as uniform continuity, uniform convergence, equicontinuity, total boundedness, and completeness, fall between the topological and the metric properties. Their definitions depend on the ability to compare the sizes of neighborhoods at different points. In metric spaces this is done by taking all balls of a given radius to be the same size. This notion can be generalized to give us a category of spaces called **uniform spaces.** In these spaces we define a **uniform structure** on a set $X$ by specifying a family of entourages, each entourage being a family of neighborhoods, one for each point of $X$. We regard two neighborhoods from the same entourage as being of the same size. Thus for each $\epsilon > 0$ the balls of radius $\epsilon$ form an entourage in a metric space. The general theory of uniform structures lies outside the scope of this book, and we refer the interested reader to Kelley[2] [9] and Bourbaki [15].

The cases of uniform structure of interest here are those where $X$ has an algebraic structure that allows us to have a natural comparison of neighborhoods at different points. For example, if $X$ is both a vector space and a topological space for which the translations are continuous self-maps, then we get a natural uniform structure by saying that neighborhoods $N_x$ of $x$ and $N_y$ of $y$ are the same size if $N_y$ is the translate of $N_x$ by the vector $y - x$. Thus an entourage is given by specifying a neighborhood $N$ of the origin and taking $N_x$ to be $N + x$. Similar considerations apply if $X$ is a topological group. When $X$ is a noncommutative topological group, we may get different uniform structures depending on whether we use left translations or right translations.

Some notions, such as uniform convergence and equicontinuity, are mixed: They concern mappings from a space $X$ to a space $Y$ and

[2] See Chapter 6. Note Problem 6H, in particular.

depend on the uniform structure of $Y$ but require only the topological structure of $X$. Thus a sequence of maps $f_n$ from a set $X$ to a metric space $(Y, \sigma)$ is said to **converge uniformly** to a map $f$ if, given $\epsilon > 0$, there is an $N$ such that $\sigma(f_n(x), f(x)) < \epsilon$ for all $n \geq N$ and all $x \in X$. For this notion we do not need *any* structure on $X$. Similarly, a family $\mathfrak{F}$ of maps of a topological space $X$ into a metric space $(Y, \sigma)$ is said to be **equicontinuous** at a point $x_0 \in X$ if, given $\epsilon > 0$, there is a neighborhood $O$ of $x_0$ such that $\sigma(f(x), f(x_0)) < \epsilon$ for each $x \in O$ and each $f \in \mathfrak{F}$. We say that $\mathfrak{F}$ is equicontinuous on $X$ if it is equicontinuous at each point.

There also exist some purely topological notions corresponding to uniform convergence on compact sets (the compact-open topology, as in Problem 9.8) and equicontinuity (topological equicontinuity, as in Section 2 of Chapter 14, or even continuity, as Kelley [9], Chapter 7).

## Problems

**50. a.** Let $\langle f_n \rangle$ be a sequence of continuous maps from a topological space $X$ to a metric space $(Y, \sigma)$ that converges uniformly to a map $f$. Show that $f$ is continuous.

**b.** A sequence $\langle f_n \rangle$ of continuous maps from a topological space $X$ to a metric space $\langle Y, \sigma \rangle$ is said to be a **uniform Cauchy sequence** if, given $\epsilon > 0$, there is an $N$ such that $\sigma(f_n(x), f_m(x)) < \epsilon$ for all $n, m \geq N$ and all $x \in X$. Show that if $\langle f_n \rangle$ is a uniform Cauchy sequence and if $Y$ is complete, then there is continuous map $f$ to which $\langle f_n \rangle$ converges uniformly.

**51.** Show that the Ascoli–Arzelá Theorem (7.40) and its corollary are still true if we only assume that $X$ is a separable topological space. [For the uniform convergence statement, see Problem 9.8b.]

## ★7 Nets

By a directed system we mean a set $A$ together with a relation $\prec$ satisfying the following conditions:

i. If $\alpha \prec \beta$ and $\beta \prec \gamma$, then $\alpha \prec \gamma$.
ii. If $\alpha, \beta \in A$, there is a $\gamma \in A$ with $\alpha \prec \gamma$ and $\beta \prec \gamma$.

One example of a directed system is the set $N$ of positive integers with $\prec$ replaced by $\leq$. Another commonly used directed set is the set of all open sets containing a point $x$, with $O_1 \prec O_2$ defined to mean $O_1 \supset O_2$.

A *net* is a mapping of a directed system into a topological space $X$. If the directed system is the integers, we have a sequence, and nets may be thought of as generalizations of sequences. We usually write $x_\alpha$ for the value of the net at $\alpha$ and $\langle x_\alpha \rangle$ for the net itself. A point $x \varepsilon X$ is said to be the *limit* of a net $\langle x_\alpha \rangle$ if for each open set $O$ containing $x$ there is an $\alpha_0 \varepsilon A$ such that $x_\alpha \varepsilon O$ for all $\alpha \succ \alpha_0$. A point $x$ is called a *cluster point* of $\langle x_\alpha \rangle$ if given $O$ containing $x$ and given $\alpha \varepsilon A$ there is a $\beta \succ \alpha$ such that $x_\beta \varepsilon O$. For sequences these notions coincide with our earlier notions of limit and cluster point.

**13. Proposition:** *A point $x$ is a point of closure of a set $E$ if and only if it is the limit of a net $\langle x_\alpha \rangle$ from $E$.*

***Proof:*** The 'if' part follows directly from the definitions of limit and point of closure. Hence we assume $x$ is a point of closure of $E$. We take as our directed system $A$ the collection of open sets which contain $x$ and set $O_1 \prec O_2$ if $O_1 \supset O_2$. Since $x$ is a point of closure of $E$, for each $O \varepsilon A$ there is a point $x_O$ in $O \cap E$. Then $\langle x_O \rangle$ is a net from $E$, and it converges to $x$, since, given $O$ containing $x$, we have $x_{O'} \varepsilon O$ for all $O' \succ O$. ▌

## Problems

**52.** Prove that $X$ is Hausdorff if and only if every net in $X$ has at most one limit. [To prove the 'if' part, let $x$ and $y$ be two points that cannot be separated and let the directed system be the collection of all pairs $\langle A, B \rangle$ of open sets with $x \varepsilon A$, $y \varepsilon B$. Choose $x_{\langle A,B \rangle}$ to be in $A \cap B$ and show that both $x$ and $y$ are limits of this net.]

**53.** Prove that a function $f$ from $X$ to $Y$ is continuous at $x$, if and only if, for each net $\langle x_\alpha \rangle$ that converges to $x$, the net $\langle f(x_\alpha) \rangle$ converges to $f(x)$.

**54.** Let $X$ be any set and $f$ a real-valued function on $X$. Let $A$ be the system consisting of all finite subsets of $X$, with $F \prec G$ meaning $F \subset G$. For each $F \varepsilon A$, let $y_F = \sum_{x \varepsilon F} f(x)$. Prove that the net $\langle y_F \rangle$ has a limit if and only if $f(x) = 0$ except for $x$ in a countable subset $\{x_n\}$ and $\sum |f(x_n)| < \infty$. In this case, $\lim y_F = \sum_{n=1}^{\infty} f(x_n)$.

**55.** Let $X = \bigtimes_\alpha X_\alpha$. Then a net $\langle x_\beta \rangle$ in $X$ converges to $x$ if and only if each coordinate of $x_\beta$ converges to the corresponding coordinate of $x$.

# 9 Compact and Locally Compact Spaces

## 1 Compact Spaces

The concept of compactness which we have studied for metric spaces generalizes to the class of topological spaces. Thus we say that a collection $\mathcal{U}$ of open sets in a topological space is an open covering for a set $K$ if $K$ is contained in the union of the sets in $\mathcal{U}$. A topological space $X$ is said to be **compact** if every open covering $\mathcal{U}$ of $X$ has a finite subcovering, that is, if we can find a finite collection $\{O_1, O_2, \dots, O_N\} \subset \mathcal{U}$ such that

$$X = \bigcup_{i=1}^{N} O_i.$$

A subset $K$ of a topological space is called compact if it is compact as a subspace of $X$. In view of the definition of the topology of a subspace, this is equivalent to saying that a subset $K$ of $X$ is compact if every covering $\mathcal{U}$ of $K$ by open sets of $X$ has a finite subcovering. The Heine–Borel Theorem states that every closed and bounded subset of real numbers is compact.

If $\mathcal{U}$ is an open covering of a space $X$, then the collection $\mathcal{F}$ of complements of sets in $\mathcal{U}$ is a collection of closed sets whose intersection is empty, and conversely. Thus a space $X$ is compact if and only if every collection of closed sets with an empty intersection has a finite subcollection whose intersection is empty. A collection $\mathcal{F}$ of sets in $X$ is said to have the *finite intersection property* if any finite subcollection of $\mathcal{F}$ has a nonempty intersection. Hence we have the following proposition:

**1. Proposition:** *A topological space $X$ is compact if and only if every collection $\mathcal{F}$ of closed sets with the finite intersection property has a nonempty intersection.*

The notion of compactness is intimately connected with that of being closed, as the following proposition shows. Thus compactness may be viewed as an absolute type of closedness.

**2. Proposition:** *A closed subset of a compact space is compact. A compact subset of a Hausdorff space is closed.*

***Proof:*** Let $X$ be compact, $F$ a closed subset of $X$, and $\mathcal{U}$ an open covering for $F$. Thus $\mathcal{U} \cup \{\tilde{F}\}$ is an open covering for $X$ and so must have a finite subcovering $\{\tilde{F}, O_1, \ldots, O_N\}$. Then the sets $O_1, O_2, \ldots, O_N$ cover $F$, and so $\mathcal{U}$ has a finite subcovering.

Suppose now that $X$ is a Hausdorff space and $K$ a compact subset of $X$. We shall show that $\tilde{K}$ is open. Let $y \in \tilde{K}$. Since $X$ is Hausdorff, for each $x \in K$ there are disjoint open sets $O_x$ and $N_x$ such that $x \in O_x$ and $y \in N_x$. The sets $\{O_x: x \in K\}$ form an open covering of $K$, and so there is a finite subcollection $\{O_{x_1}, O_{x_2}, \ldots, O_{x_n}\}$ which covers $K$. Let

$$N = \bigcap_{i=1}^{n} N_{x_i}.$$

Then $N$ is an open set containing $y$ and not meeting any of the sets $O_{x_i}$. Since $K \subset \bigcup O_{x_i}$, $N$ does not meet $K$ and so is contained in $\tilde{K}$. Thus $\tilde{K}$ is open and $K$ closed. ▌

**3. Corollary:** *Every compact set of real numbers is closed and bounded.*

***Proof:*** Since $\mathbf{R}$ is Hausdorff, a compact subset $K$ of $\mathbf{R}$ must be closed. Moreover, the intervals $I_n = (-n, n)$ form an open covering of $K$, and so a finite number of them must cover $K$. Hence $K$ must be bounded. ▌

**4. Proposition:** *The continuous image of a compact set is compact.*

***Proof:*** Let $f$ be a continuous function that maps the compact set $K$ *onto* a topological space $Y$. If $\mathcal{U}$ is an open covering for $Y$, then the collection of sets $f^{-1}[O]$ for all $O \in \mathcal{U}$ is an open covering of $K$.

By the compactness of $K$, there are a finite number $O_1, \ldots, O_n$ of sets of $\mathcal{U}$ such that the sets $f^{-1}[O_i]$ cover $K$. Since $f$ is onto, the sets $O_1, \ldots, O_n$ cover $Y$. ▮

**5. Proposition:** *A one-to-one continuous mapping of a compact space onto a Hausdorff space is a homeomorphism.*

***Proof:*** Let $X$ be compact, $Y$ Hausdorff, and $f$ a one-to-one continuous mapping onto $Y$. In order to show that $f$ is a homeomorphism, it is only necessary to show that it carries open sets into open sets or equivalently closed sets into closed sets. But if $F$ is a closed subset of $X$, it is compact by Proposition 2, and so $f[F]$ is compact by Proposition 4 and so must be closed by Proposition 2. ▮

The property of compactness is sometimes expressed in terms of refinements instead of subcoverings. An open covering $\mathcal{V}$ of $X$ is said to be a **refinement** of the open covering $\mathcal{U}$ (or to *refine* $\mathcal{U}$) if every element of $\mathcal{V}$ is a subset of an element of $\mathcal{U}$. Then it is not difficult to see that $X$ is compact if and only if every open covering of $X$ has a finite refinement. We also note for future use that any two open covers $\mathcal{U}$ and $\mathcal{V}$ for $X$ have a common refinement, namely, the collection of all sets of the form $U \cap V$ with $U \in \mathcal{U}$ and $V \in \mathcal{V}$.

## Problems

1. Show that $X$ is compact iff every open cover has a finite refinement.

2. Let $\langle K_n \rangle$ be a decreasing sequence compact sets, that is, $K_{n+1} \subset K_n$. Let $O$ be an open set with $\bigcap_{n=1}^{\infty} K_n \subset O$. Then $K_n \subset O$ for some $n$.

3. Prove that a compact Hausdorff space is regular.

4. Prove that a compact Hausdorff space is normal.

5. Let $f$ be a continuous mapping of the compact space $X$ onto the Hausdorff space $Y$. Then any mapping $g$ of $Y$ into $Z$ for which $g \circ f$ is continuous must itself be continuous.

**a.** Prove that if $(X, \mathfrak{T})$ is a compact space, then $(X, \mathfrak{T}_1)$ is compact for all $\mathfrak{T}_1$ weaker than $\mathfrak{T}$.

**b.** Show that if $(X, \mathfrak{T})$ is a Hausdorff space, then $(X, \mathfrak{T}_2)$ is a Hausdorff space for all $\mathfrak{T}_2$ stronger than $\mathfrak{T}$.

**c.** Show that if $(X, \mathfrak{T})$ is a compact Hausdorff space, then any weaker topology is not Hausdorff and any stronger topology is not compact.

**6.** Let $X$ be a compact space and $\mathcal{F}$ an equicontinuous family of maps from $X$ to a metric space $\langle Y, \sigma\rangle$. Let $\langle f_n\rangle$ be a sequence from $\mathcal{F}$ such that $f_n(x) \to f(x)$ for each $x \in X$. Then $\langle f_n\rangle$ converges to $f$ uniformly on $X$.

**7.** Let $X$ be a topological space and $\langle C_n\rangle$ a decreasing sequence of compact and connected sets. Then $\bigcap C_n$ is compact and connected. A compact connected set is called a *continuum* if it has more than one point.

**8.** *The Compact-Open Topology.* Let $X$ and $Y$ be topological spaces and $X^Y$ the space of maps from $X$ into $Y$. On $X^Y$ we define a topology, called the compact-open topology, by taking as a subbase sets of the form $N_{K,O} = \{f \in X^Y : f[K] \subset O\}$, where $K$ is a compact subset of $X$ and $O$ is an open subset of $Y$. Thus the compact-open topology is the weakest topology on $X^Y$ such that the sets $N_{K,O}$ are open.

**a.** Let $\langle f_n\rangle$ be a sequence of maps from $X$ to $Y$ that converge in the compact-open topology to $f: X \to Y$. Then for each $x \in X$ we have $f(x) = \lim f_n(x)$.

**b.** Let $\langle f_n\rangle$ be a sequence of continuous maps from a topological space $X$ to a metric space $(Y, \sigma)$. Show that $\langle f_n\rangle$ converges in the compact-open topology to a map $f: X \to Y$ if and only if $\langle f_n\rangle$ converges to $f$ uniformly on each compact subset $C$ of $X$.

## 2 Countable Compactness and the Bolzano–Weierstrass Property

A weaker notion than compactness is countable compactness: A space $X$ is said to be **countably compact** if every countable open covering has a finite subcovering. We defined a topological space to be Lindelöf if every open covering of it had a countable subcovering. Thus a topological space is compact iff it is both Lindelöf and countably compact. Since every second countable space is Lindelöf, it follows that countable compactness is equivalent to compactness in the presence of the second axiom of countability. The proof of Proposition 4 applies in the countably compact case to give the following proposition:

**6. Proposition:** *The continuous image of a countably compact space is countably compact.*

A topological space $X$ is said to have the **Bolzano–Weierstrass property** if every sequence $\langle x_n\rangle$ in $X$ has at least one cluster point, that is, if there is an $x \in X$ such that for each open set $O$ containing $x$ and for each $N$ there is an $n \geq N$ with $x_n \in O$.

**7. Proposition:** *A topological space has the Bolzano–Weierstrass property if and only if it is countably compact.*

***Proof:*** We first observe that $X$ is countably compact if and only if every countable family $\mathcal{F}$ of closed sets with the finite intersection property has a nonempty intersection. Suppose now that $X$ has the Bolzano–Weierstrass property and that $\mathcal{F} = \{F_i\}$ is a countable family of closed sets with the finite intersection property. Since the intersection $H_n = \bigcap_{k=1}^{n} F_k$ is empty for no $n$, we may choose for each $n$ an element $x_n \in H_n$. By the Bolzano–Weierstrass property, the sequence $\langle x_n \rangle$ has a cluster point $x$. But $x_n \in F_i$ for all $n \geq i$, and so $x$ must belong to $F_i$, since $F_i$ is closed. Thus $x$ belongs to every $F_i$ and so to their intersection.

Suppose, on the other hand, that $X$ is countably compact and that $\langle x_i \rangle$ is a sequence from $X$. Let $B_n$ be the set $\{x_n, x_{n+1}, \dots, \}$. Then $\{\bar{B}_n\}$ is a countable collection of closed sets with the finite intersection property, and so there is a point $x$ which belongs to $\bigcap \bar{B}_n$. The point $x$ is a cluster point of the sequence, since given $N$ and any open set $O$ containing $x$, we have $x \in \bar{B}_N$, and so there must be an $x_n \in O$ with $n \geq N$. ▌

A concept that resembles the Bolzano–Weierstrass property is sequential compactness. A space $X$ is said to be **sequentially compact** if every infinite sequence from $X$ has a convergent subsequence. For metric spaces the concepts of compactness, countable compactness, and sequential compactness all coincide as a consequence of the Borel–Lebesgue Theorem. In general, we must distinguish between these notions: Problem 11 gives an example of a space that is sequentially compact but not compact, and Problem 41 an example of a compact space that is not sequentially compact. Problem 12 shows that a space may be compact without being either separable or first countable.

**8. Proposition:** *A sequentially compact space is countably compact. A countably compact space satisfying the first countability axiom is sequentially compact.*

***Proof:*** Sequential compactness implies the Bolzano–Weierstrass property, which is equivalent to countable compactness. The second part is an immediate consequence of Problem 8.11. ▌

**9. Proposition:** *Let $f$ be a continuous real-valued function on a countably compact space $X$. Then $f$ is bounded and assumes its maximum and minimum.*

This proposition can be proved by using Proposition 6 and the fact that every countably compact subset of $\mathbf{R}$ is closed and bounded. However, we can give a direct proof which proves more. A real-valued function $f$ on a topological space is called **upper semicontinuous** if for each real number $\alpha$ the set $\{x: f(x) < \alpha\}$ is open. If $f$ is continuous, both $f$ and $-f$ are upper semicontinuous. This implies that Proposition 9 is a corollary of the following proposition:

**10. Proposition:** *Let $f$ be an upper semicontinuous real-valued function on a countably compact space $X$. Then $f$ is bounded (from above) and assumes its maximum.*

***Proof:*** The sets $O_n = \{x: f(x) < n\}$ form a countable open covering for $X$, and so there must be a finite subcovering $\{O_1, \ldots, O_N\}$. But this implies $X \subset O_N$. Hence $f(x) < N$ for all $x$, and $f$ is bounded from above. Let $\beta = \sup \{f(x): x \in X\}$. Then the sets

$$F_n = \left\{x: f(x) \geq \beta - \frac{1}{n}\right\}$$

form a countable collection of closed sets with the finite intersection property. Hence there is a $y$ belonging to every $F_n$. Then $f(y) = \beta$, and $f$ assumes its maximum at $y$. ∎

**11. Proposition (Dini):** *Let $\langle f_n \rangle$ be a sequence of upper semicontinuous real-valued functions on a countably compact space $X$, and suppose that for each $x \in X$ the sequence $\langle f_n(x) \rangle$ decreases monotonically to zero. Then $\langle f_n \rangle$ converges to zero uniformly.*

***Proof:*** Choose $\epsilon > 0$, and let $O_n = \{x: f_n(x) < \epsilon\}$. Since $f_n$ is upper semicontinuous, $O_n$ is open. Since $f_n(x) \to 0$ for each $x$, we have $X \subset \bigcup O_n$. By the countable compactness of $X$, there are a finite number of open sets $\{O_1, \ldots, O_N\}$ whose union contains $X$. But this implies that $O_N = X$, and hence $f_N(x) < \epsilon$ for all $x$. If $n \geq N$, we have $0 \leq f_n(x) \leq f_N(x) < \epsilon$, and the sequence $\langle f_n \rangle$ converges to 0 uniformly. ∎

## Problems

**9. a.** A real-valued function $f$ is called lower semicontinuous if $-f$ is upper semicontinuous. Show that a real-valued function on a space $X$ is continuous if and only if it is both upper and lower semicontinuous.

**b.** Show that if $f$ and $g$ are upper semicontinuous, so is $f + g$.

**c.** Let $\langle f_n \rangle$ be a decreasing sequence of upper semicontinuous functions which converge pointwise to a real-valued function $f$. Then $f$ is upper semicontinuous.

**d.** Let $\langle f_n \rangle$ be a decreasing sequence of upper semicontinuous functions on a countably compact space, and suppose that $\lim f_n(x) = f(x)$, where $f$ is a lower semicontinuous real-valued function. Then $f$ is continuous and $\langle f_n \rangle$ converges to $f$ uniformly.

**e.** Show that if a sequence $\langle f_n \rangle$ of upper semicontinuous functions converges uniformly to a function $f$, then $f$ is also upper semicontinuous.

**10.** Let $X$ be a normal topological space. Then the following are equivalent:

i. $X$ is countably compact.
ii. Every continuous real-valued function on $X$ is bounded.
iii. Every bounded continuous real-valued function on $X$ assumes its maximum.

**11.** Let $X$ be the set of ordinals less than the first uncountable ordinal, and let $\mathcal{B}$ be the collection of sets of the form $\{x: x < a\}$, $\{x: a < x < b\}$, and $\{x: a < x\}$.

**a.** Show that $\mathcal{B}$ is a base for a topology for $X$.

**b.** Show that $X$ is sequentially compact but not compact. [Hint: Use the well ordering of the ordinals.]

**c.** Show that if $f$ is a continuous real-valued function on $X$, then there is an $x_0$ such that $f(x) = f(x_0)$ for all $x \geq x_0$. [Hint: Show that the set of $x$ for which $f(x) < \overline{\lim} f$ is countable.]

**12.** Let $Y$ be the set of ordinals less than or equal to the first uncountable ordinal $\Omega$, and let $\mathcal{B}$ be the collection of sets of the form $\{x: x < a\}$, $\{x: a < x < b\}$, and $\{x: a < x\}$.

**a.** Show that $\mathcal{B}$ is a base for a topology for $X$.

**b.** Show that $X$ is compact but neither separable nor first countable.

## 3 Products of Compact Spaces

In this section we prove the Theorem of Tychonoff—that a product of compact spaces is compact. It is probably the most important theorem in general topology. Most applications in analysis need only the special case of a product of (closed) intervals, but this special case does not seem to be easier to prove than the general case. We begin with two lemmas concerning the finite intersection property.

**12. Lemma:** *Let $\mathcal{A}$ be a collection of subsets of a set $X$, and suppose that $\mathcal{A}$ has the finite intersection property. Then there is a collection $\mathcal{B} \supset \mathcal{A}$ such that $\mathcal{B}$ has the finite intersection property and is maximal with respect to this property; that is, no collection properly containing $\mathcal{B}$ has the finite intersection property.*

***Proof:*** Consider the family of all collections containing $\mathcal{A}$ and having the finite intersection property. This family is partially ordered by inclusion. By the Hausdorff maximal principle there is a maximal linearly ordered subfamily $\mathcal{F}$. Let $\mathcal{B}$ be the union of the collections in $\mathcal{F}$. If $B_1, \ldots, B_n$ are in $\mathcal{B}$, then each $B_i$ belongs to some $\mathcal{C}_i \in \mathcal{F}$. Since $\mathcal{F}$ is linearly ordered by inclusion, one of the collections $\mathcal{C}_k$ contains the others, and so all $B_i$ belong to $\mathcal{C}_k$. Since $\mathcal{C}_k$ has the finite intersection property, $\bigcap B_i \neq \emptyset$. Thus $\mathcal{B}$ has the finite intersection property. If $\mathcal{B}' \supset \mathcal{B}$ and $\mathcal{B}'$ has the finite intersection property, then $\mathcal{B}'$ contains every $\mathcal{C}$ in $\mathcal{F}$ and so must belong to $\mathcal{F}$ by the maximality of $\mathcal{F}$. Thus $\mathcal{B}$ is a union of collections one of which is $\mathcal{B}'$, and so $\mathcal{B}' \subset \mathcal{B}$. This shows that $\mathcal{B}$ is maximal with respect to the finite intersection property. ▮

**13. Lemma:** *Let $\mathcal{B}$ be a collection of subsets of $X$ 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 set that meets 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 of $\mathcal{B}$.

Suppose that a set $C$ meets each element 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 maximality $\mathcal{B} \cup \{C\} = \mathcal{B}$, and so $C \in \mathcal{B}$. ▮

**14. Theorem (Tychonoff):** *Let $\langle X_\alpha \rangle$ be an indexed family of compact topological spaces. Then the product space $\bigtimes_\alpha X_\alpha$ is compact in the product topology.*

***Proof* (*Bourbaki*):** Let $\pi_\alpha$ be the mapping of $X$ to $X_\alpha$ which assigns to each $x \in X$ its $\alpha$-th coordinate. Then the sets obtained by taking finite intersections of sets of the form $\pi_\alpha^{-1}[O_\alpha]$ with $O_\alpha$ open in $X_\alpha$ form a base $\mathcal{N}$ for the topology of $X$.

Let $\mathcal{A}$ be any collection of closed subsets of $X$ with the finite intersection property, and let $\mathcal{B}$ be a collection of (not necessarily closed) sets that contains $\mathcal{A}$ and is maximal with respect to the finite intersection property. Let $\mathcal{B}_\alpha$ be the collection of subsets of $X_\alpha$ of the form $\pi_\alpha(B)$ with $B \in \mathcal{B}$. Then $\mathcal{B}_\alpha$ has the finite intersection property, and by the compactness of $X_\alpha$ it is possible to choose a point $x_\alpha$ belonging to $\bigcap_{\mathcal{B}} [\overline{\pi_\alpha(B)}]$, that is, a point $x_\alpha$ which is a point of closure of each set $\pi_\alpha[B]$. Let $x$ be that point in $X$ whose $\alpha$-th coordinate is $x_\alpha$.

Consider a set $S$ that is of the form $\pi_\alpha^{-1}[O_\alpha]$ for some $\alpha$ and some open set $O_\alpha$ in $X_\alpha$ with $x_\alpha \in O_\alpha$. Since $x_\alpha$ is a point of closure of $\pi_\alpha[B]$ for each $B$ in $\mathcal{B}$, the set $S$ must intersect each $B$ in $\mathcal{B}$. By Lemma 13 we must have $S \in \mathcal{B}$. Each set containing $x$ in the base $\mathcal{N}$ for the topology of $X$ is a finite intersection of sets of this form and so must be in $\mathcal{B}$ by Lemma 13. Let $F$ be a closed set in $\mathcal{B}$. Then $F$ meets each $N \in \mathcal{N}$ with $x \in N$. Consequently, $x$ is a point of closure of $F$ and so is in $F$. Hence $x$ belongs to each set in $\mathcal{A}$, and $\mathcal{A}$ has nonempty intersection. ∎

## Problems

**13.** Each closed and bounded set in $\mathbf{R}^n$ is compact.

**14.** Prove without using the axiom of choice that, if $X$ is compact and $I$ is a closed interval, then $X \times I$ is compact. [Hint: 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 $\mathcal{U}$.]

**15.** Prove that the product of a countable number of sequentially compact spaces is sequentially compact. [If $\langle x_n \rangle$ is a sequence in the product, choose a subsequence $\langle x_n^1 \rangle$ whose first coordinate converges, choose a subsequence $\langle x_n^2 \rangle$ of this whose second coordinate converges, etc. Then the "diagonal" sequence $\langle x_n^n \rangle$ converges in the product space.]

**16.** 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 $Q = \bigtimes_{f \in \mathcal{F}} I_f$. Then the mapping $g$ of $X$ onto $Q$ that takes $x$ into the point whose $f$-th coordinate is $f(x)$ is one-to-one into $Q$ and continuous.]

**17.** 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 $Q$ such that $|f - g| < \epsilon$ and $g$ is a function of only a finite number of coordinates. [Hint: Cover the range of $f$ by a finite number of intervals of length $\epsilon$ and look at the inverse images of these intervals.]

## 4 Locally Compact Spaces

A topological space $X$ is called **locally compact** if for each $x \in X$ there is an open set $O$ containing $x$ such that $\bar{O}$ is compact. Thus $X$ is locally compact iff the collection of open sets with compact closures forms a base for the topology of $X$. Every compact space is locally compact, while the Euclidean spaces $\mathbf{R}^n$ are examples of spaces which are locally compact but not compact.

Locally compact Hausdorff spaces constitute one of the most important classes of topological spaces. This section is devoted to establishing some of their basic properties. The behavior of locally compact spaces subject to additional assumptions is the topic of several later sections of this chapter, and Chapter 13 deals with further topics on these spaces. Throughout these sections we assume that $X$ is locally compact *and* Hausdorff.[1]

The following propositions provide useful properties of locally compact Hausdorff spaces. The first assures us that there are an adequate number of real-valued functions on $X$. Its proof is left to the reader (Problems 18 and 19).

**15. Proposition:** *Let $K$ be a compact subset of a locally compact Hausdorff space $X$. Then there is an open set $O$ containing $K$ with $\bar{O}$ compact. Given such a set $O$, there is a continuous nonnegative function $f$ on $X$ which vanishes outside $O$ and is identically 1 on $K$. If $K$ is also a $G_\delta$, we may take $f < 1$ in $\tilde{K}$.*

[1] So useful is the combination of the Hausdorff separation axiom in connection with compactness that French usage (following Bourbaki) reserves the term '*compact space*' for those that are compact *and* Hausdorff, using the term '*pseudocompact*' for those that are not Hausdorff.

We recall that, if $f$ is a real-valued function on a topological space, the **support** of $f$ is the closure of the set $\{x: f(x) \neq 0\}$. Thus

$$\text{support } f = \overline{\{x: f(x) \neq 0\}}.$$

We say that a collection $\{\varphi_\alpha\}$ of real-valued functions on $X$ is **subordinate** to a covering $\{O_\lambda\}$ of $X$ if the support of each $\varphi_\alpha$ is contained in some $O_\lambda$.

**16. Proposition:** *Let $\{O_\lambda\}$ be an open covering of a compact subset $K$ of a locally compact Hausdorff space $X$. Then there is a finite collection $\{\varphi_1, \ldots, \varphi_n\}$ of continuous nonnegative real-valued functions subordinate to the collection $\{O_\lambda\}$ and such that*

$$\varphi_1 + \varphi_2 + \cdots + \varphi_n \equiv 1$$

*on $K$.*

***Proof:*** Let $O$ be an open set with $K \subset O$ and $\bar{O}$ compact. For each $x_0 \in K$ there is a continuous real-valued function $f_{x_0}$ with $f_{x_0}(x_0) = 1$, $0 \leq f_{x_0} \leq 1$, and support $f_{x_0} \subset O \cap O_\lambda$ for some $\lambda$. For each $x_0 \in \bar{O} \sim K$ let $g_{x_0}$ be a continuous real-valued function with $g_{x_0}(x_0) = 1$, $0 \leq g_{x_0} \leq 1$, and support $g_{x_0} \subset \sim K$. By the compactness of $\bar{O}$ we may choose a finite number $f_1, \ldots, f_n, g_1, \ldots, g_m$ of these functions such that the sets where they are positive cover $\bar{O}$. Set

$$f = \sum_{i=1}^{n} f_i$$

$$g = \sum_{j=1}^{m} g_j.$$

Then $f > 0$ on $K$, support $f \subset O$, $f + g > 0$ on $\bar{O}$, and $g \equiv 0$ on $K$. Thus $f/(f+g)$ is continuous and $\equiv 1$ on $K$. Take $\varphi_i = f_i/(f+g)$. ∎

The proof of this proposition can be modified so that the functions $\varphi_i$ come from a restricted class of functions, say differentiable functions if $X$ is a differentiable manifold (see Problem 25).

The next proposition, whose proof we leave to the reader (see Problem 26), is the analogue of the Theorem of Baire for complete metric spaces.

**17. Proposition:** *Let $X$ be a locally compact space and $\{O_n\}$ a countable collection of dense open sets. Then $\bigcap O_n$ is dense in $X$.*

Thus each locally compact space is locally of second Baire category with respect to itself, and the various theorems of Baire category theory apply to locally compact Hausdorff spaces as well as to complete metric spaces. We state one such consequence which will be useful later.

**18. Proposition:** *Let $X$ be a locally compact space. If $O$ is an open subset that is contained in a countable union $\bigcup F_n$ of closed sets, then the union of their interiors $\bigcup F_n^\circ$ is an open set dense in $O$.*

The following proposition and corollary, whose proof is left to the reader, give a characterization of those subsets of a locally compact space that are themselves locally compact.

**19. Proposition:** *Let $Y$ be a dense subset of a Hausdorff space $X$, and suppose that $Y$ with its subspace topology is locally compact. Then $Y$ is an open subset of $X$.*

**20. Corollary:** *A subset $Y$ of a locally compact Hausdorff space $X$ is locally compact in its subspace topology if and only if $Y$ is a relatively open subset of $\overline{Y}$.*

If $X$ is a locally compact Hausdorff space, we can form a new space $X^*$ by adding to $X$ a single point $\omega$ not in $X$ and taking a set in $X^*$ to be open if 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 $\omega$ is often referred to as the point at infinity in $X^*$.

A continuous mapping $f$ from a topological space $X$ into a topological space $Y$ is said to be **proper** if $f^{-1}[K]$ is a compact set for each compact set $K \subset Y$. Proper maps from a locally compact Hausdorff space $X$ into a locally compact Hausdorff space $Y$ are precisely those continuous maps of $X$ into $Y$ that can be extended to continuous maps $f^*$ of $X^*$ into $Y^*$ by taking the point at infinity in $X^*$ to the point at infinity in $Y^*$.

## Problems

**18.** Let $X$ be a locally compact space and $K$ a compact subset of it. Show that there is an open set $O \supset K$ such that $\overline{O}$ is compact. [Hint: For each point $x \varepsilon K$ there is an $O_x$ containing $x$ with $\overline{O}_x$ compact. Let $O$ be the union of a finite number of these $O_x$ which cover $K$.]

**19. a.** Let $X$ be a locally compact Hausdorff space and $K$ a compact set. Then there is a continuous real-valued function on $X$ that is identically 1 on $K$ and for which the set $O = \{x: f(x) \neq 0\}$ has compact closure. [Hint: Use Problem 16, the Urysohn Lemma, and Proposition 8.3.]

**b.** Prove Proposition 15.

**20. a.** Let $X^*$ be the Alexandroff one-point compactification of $X$. Prove that the subsets of $X^*$ which are either open subsets of $X$ or the complements of compact subsets of $X$ form a topology for $X^*$, that is, that the intersection of two such sets is such a set and the union of any collection of such sets is such a set.

**b.** Show that the identity mapping from $X$ to the subspace $X^* \sim \{\omega\}$ is a homeomorphism.

**c.** Show that $X^*$ is compact and Hausdorff.

**21. a.** Show that the Alexandroff one-point compactification of $\mathbf{R}^n$ is homeomorphic to the boundary of a sphere in $\mathbf{R}^{n+1}$.

**b.** Show that the one-point compactification of the space $X$ in Problem 11 is the space $Y$ in Problem 12.

**22. a.** Let $O$ be an open subset of a compact Hausdorff space. Then $O$ is locally compact.

**b.** Let $O$ be an open set in a compact Hausdorff space $X$. Then the mapping of $X$ to the one-point compactification of $O$ which is the identity on $O$ and takes each point in $X \sim O$ into $\omega$ is continuous.

**23.** Let $X$ and $Y$ be locally compact Hausdorff spaces, and $f$ a continuous mapping of $X$ into $Y$. Let $X^*$ and $Y^*$ be the one-point compactifications of $X$ and $Y$, and $f^*$ the mapping of $X^*$ into $Y^*$ whose restriction to $X$ is $f$ and which takes the point at infinity in $X^*$ into the point at infinity in $Y^*$. Then $f$ is proper if and only if $f^*$ is continuous.

**24. a.** Let $X$ be a locally compact space. A subset $F$ of $X$ is closed if and only if $F \cap K$ is closed for each closed compact set $K$.

**b.** The above conclusion is true if $X$ is a Hausdorff space satisfying the first axiom of countability instead of being locally compact.

**25.** Let $\mathcal{F}$ be a family of real-valued continuous functions on a locally compact Hausdorff space $X$, and suppose that $\mathcal{F}$ has the following properties:

i. If $f \varepsilon \mathcal{F}$ and $g \varepsilon \mathcal{F}$, then $f + g \varepsilon \mathcal{F}$.
ii. If $f \varepsilon \mathcal{F}$ and $g \varepsilon \mathcal{F}$, then $f/g \varepsilon \mathcal{F}$, provided that support $f \subset \{x \varepsilon X: g(x) \neq 0\}$.

iii. Given an open set $O \subset X$ and $x_0 \in O$, there is an $f \in \mathcal{F}$ with $f(x_0) = 1$, $0 \le f \le 1$, and support $f \subset O$.

Show that Proposition 16 is still true if we require the functions $\varphi_i$ be in $\mathcal{F}$.

**26.** Prove Proposition 17. [Hint: Use the proof of the Theorem of Baire, with the finite intersection property substituting for the completeness of $X$.]

**27.** Prove Proposition 18.

**28.** Prove Proposition 19.

**29. a.** Show that a closed subset of a locally compact space is locally compact.

**b.** Show that an open subset of a locally compact Hausdorff space is locally compact.

**c.** Prove Corollary 20. [Hint: $Y$ is dense in $\overline{Y}$.]

## 5 σ-Compact Spaces

A topological space $X$ is said to be **σ-compact** if it is the union of a countable collection of compact sets. In the presence of local compactness there are a number of other properties that are equivalent to $\sigma$-compactness.

**21. Theorem:** *Let $X$ be a locally compact Hausdorff space. Then the following statements are equivalent:*

i. *$X$ is Lindelöf.*
ii. *$X$ is $\sigma$-compact.*
iii. *There is a sequence $\langle O_n \rangle$ of open sets with $\overline{O}_n$ compact, $\overline{O}_n \subset O_{n+1}$ and $X = \bigcup O_n$.*
iv. *There is a proper continuous map $\varphi \colon X \to (0, \infty)$.*

***Proof:*** To see that (i) implies (ii), we observe that $X$ is covered by open sets whose closures are compact. If $X$ is Lindelöf, we have a countable subcovering $\{U_n\}$, whence $X = \bigcup \overline{U}_n$.

(ii) $\Rightarrow$ (iii): Let $X = \bigcup K_n$ with $K_n$ compact. Let $O_1$ be an open set with $\overline{O}_1$ compact and $K_1 \subset O_1$. Define the sequence $\langle O_n \rangle$ inductively by letting $O_n$ be an open set with $\overline{O}_n$ compact and $O_n$ containing the compact set $K_n \cup \overline{O}_{n-1}$. Then $\langle O_n \rangle$ is the desired sequence.

(iii) $\Rightarrow$ (iv): Let $\varphi_n$ be a continuous real-valued function with $\varphi_n \equiv 1$ on $\bar{O}_{n-1}$ and support $\varphi_n \subset O_n$. Let

$$\varphi = \sum_{n=1}^{\infty} (1 - \varphi_n).$$

Then $\varphi$ is a proper continuous map of $X$ into $[0, \infty)$.

(iv) $\Rightarrow$ (i): Let $\varphi$ be a proper map of $X$ into $[0, \infty)$. Then $X = \bigcup K_n$, where $K_n = \varphi^{-1}[[0, n]]$. Since $\varphi$ is proper, $K_n$ is compact. Each open covering $\mathfrak{U}$ of $X$ has a finite subcollection $\mathfrak{U}_n$ which covers $K_n$. Then $\mathfrak{U} = \bigcup \mathfrak{U}_n$ is a countable subcovering of $X$. ∎

A sequence $\langle O_n \rangle$ with the properties in (iii) is called an **exhaustion** of $X$.

## Problems

**30.** Show that the function $\varphi$ constructed in the proof of Theorem 21 is continuous and proper.

**31.** A locally compact metric space $(X, \rho)$ is said to be proper or to have a **proper metric** $\rho$ if the closed balls $\{x: \rho(x, x_0) \leq a\}$ are compact for some $x_0$ and all $a \in (0, \infty)$.

**a.** Let $(X, \rho)$ be a proper locally compact metric space. Then a subset $K$ is compact if and only if it is both closed and bounded.

**b.** Every proper locally compact metric space is $\sigma$-compact.

**c.** Every $\sigma$-compact and locally compact metric space $(X, \rho)$ has an equivalent metric $\rho^*$ that is proper. [Hint: Try $\rho^*(x, y) = \rho(x, y) + |\varphi(x) - \varphi(y)|$, where $\varphi$ is a proper real-valued function on $X$.]

## ⋆6 Paracompact Spaces

We say that a collection $\mathcal{C}$ of subsets in a topological space $X$ is **locally finite** if each $x \in X$ has a neighborhood $U$ that meets only a finite number of sets of $\mathcal{C}$. The following lemmas, whose proofs are left to the reader, are useful in dealing with locally finite collections. They show that such collections have some properties of finite collections.

**22. Lemma:** *Let $\{E_\lambda\}$ be a locally finite collection of subsets of a topological space $X$, and set $E = \bigcup E_\lambda$. Then $\bar{E} = \bigcup \bar{E}_\lambda$.*

**23. Lemma:** *Let $\{E_\lambda\}$ be a locally finite collection of subsets of $X$ and $K$ a compact subset of $X$. Then $K$ meets only a finite number of sets in $\{E_\lambda\}$.*

A topological space $X$ is said to be **paracompact** if every open cover of $X$ has a locally finite open refinement.

A famous theorem of Arthur Stone asserts that every metric space is paracompact. The interested reader is referred to Mary Ellen Rudin [26] for a reasonably short proof. In this section we establish the equivalence, in the class of locally compact Hausdorff spaces, of a number of other properties with paracompactness.

A notion that is sometimes confused with that of local finiteness in the case of open coverings is star finiteness: A collection $\{E_\lambda\}$ of subsets of $X$ is said to be **star-finite** if each $E_\lambda$ meets only a finite number of other elements of the collection. A star-finite collection of open sets is locally finite, but the converse need not hold as is shown by the example $\{O_n\}$ with $O_n = (n, \infty)$, $n \geq 0$. This is a locally finite cover of $[1, \infty)$, but not star-finite.

**24. Proposition:** *A $\sigma$-compact locally compact space is paracompact.*

***Proof:*** Let $\mathcal{U}$ be an open covering of $X$ and $\langle O_n \rangle$ an exhaustion of $X$ as given in (iii) of Theorem 21. Let $\mathcal{U}_n$ be the collection of sets of the form $U \cap (O_{n+1} \sim \bar{O}_{n-2})$. Then each $\mathcal{U}_n$ is a refinement of $\mathcal{U}$, and $\mathcal{U}_n$ covers the compact set $K_n = \bar{O}_n \sim O_{n-1}$. Thus there is a finite subcover $\mathcal{U}_n^*$ that covers $K_n$. Since $X = \bigcup K_n$, the collection $\mathcal{V} = \bigcup \mathcal{U}_n^*$ covers $X$ and is a refinement of $\mathcal{U}$. Now each $x \in X$ belongs to the open set $O_n \sim \bar{O}_{n-2}$ for some $n$. But this open set can meet members of only four of the collections $\mathcal{U}_k^*$. Since each $\mathcal{U}_k^*$ is finite, $O_n \sim \bar{O}_{n-2}$ meets only a finite number of members of $\mathcal{V}$. This shows that $\mathcal{V}$ is locally finite. ▌

**25. Theorem:** *Let $X$ be a locally compact Hausdorff space. Then the following statements are equivalent:*

i. *$X$ is paracompact.*
ii. *Every open cover of $X$ has a star-finite open refinement.*
iii. *$X$ is the direct union of $\sigma$-compact spaces.*

***Proof:*** Clearly, (ii) $\Rightarrow$ (i). To see the converse, let $\mathcal{U}$ be an open covering of $X$. Since $X$ is locally compact $\mathcal{U}$ has a refinement consisting of open sets with compact closures. By paracompactness this covering has a locally finite refinement $\mathcal{V}$. Since each member of $\mathcal{V}$ has a compact closure, it can meet only a finite number of other members of $\mathcal{V}$. Thus $\mathcal{V}$ is star-finite.

Since each $\sigma$-compact locally compact space is paracompact and the direct union of paracompact spaces is paracompact, we have (iii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

Thus it remains only to show that (ii) $\Rightarrow$ (iii). To do this, let $\mathcal{U}$ be a star-finite covering of $X$ by open sets with compact closures. Define an equivalence relation in $X$ by setting $x \equiv y$ if there is a finite chain $U_0, \ldots, U_n$ of sets from $\mathcal{U}$ such that $x \in U_0$, $U_k \cap U_{k+1} \neq \emptyset$, and $y \in U_n$. Let $\{X_\alpha\}$ be the decomposition of $X$ into equivalence classes with respect to $\equiv$. Then $X = \bigcup X_\alpha$, and this is a disjoint union. Given $x \in X$, let $V_n$ be the set of $y \in X$ that can be connected to $x$ by a chain of $n + 1$ or fewer elements in $\mathcal{U}$. Then each $V_n$ is open, and because $\mathcal{U}$ is star-finite, each $V_n$ is the union of a finite number of elements of $\mathcal{U}$. Also $\bar{V}_n \subset V_{n+1}$. Thus the equivalence class containing $x$ is $\bigcup V_n$ and hence a $\sigma$-compact open set. Thus $X$ is the direct union of $\sigma$-compact sets. ∎

**26. Corollary:** *A connected paracompact and locally compact space is $\sigma$-compact.*

***Proof:*** A direct union can only be connected if there is just one direct summand. ∎

## Problems

**32.** Prove Lemma 22.

**33.** Prove Lemma 23.

**34. a.** Show that a paracompact Hausdorff space is regular. [Hint: Vary the proof of Problem 3 by using Lemma 22.]

**b.** Show that a paracompact Hausdorff space is normal.

**35.** An extended metric $\rho$ for a locally compact metrizable space is said to be proper if for each $x_0 \in X$ the sets $\{x\colon \rho(x, x_0) \leq a\}$ are compact for each $a \in (0, \infty)$. Show that a locally compact metrizable space is paracompact if and only if it can be metrized by a proper extended metric.

## 7 Manifolds

By an $n$-dimensional **manifold** we mean a connected Hausdorff space $M$ such that each point has a neighborhood which is homeomorphic to a ball in $\mathbf{R}^n$. We sometimes express this by saying that a manifold is a connected Hausdorff space which is locally Euclidean. It follows from the definition that a manifold has all of the local properties of Euclidean space. In particular, it is locally compact and locally connected.

Each neighborhood which is homeomorphic to a ball is called a *coordinate* neighborhood or a coordinate ball. A pair $\langle U, \varphi\rangle$ consisting of a coordinate ball $U$ and a homeomorphism $\varphi$ of $U$ onto a ball in $\mathbf{R}^n$ is called a *coordinate chart*, and $\varphi$ is called a coordinate map. The coordinates (in $\mathbf{R}^n$) of a point $x \in U$ under $\varphi$ are said to be the coordinates of $x$ in this chart.

The following theorem gives the equivalence of a number of properties for manifolds.

**27. Theorem:** *Let $M$ be a manifold. Then the following statements are equivalent:*

i. *$M$ is paracompact.*
ii. *$M$ is $\sigma$-compact.*
iii. *$M$ is Lindelöf.*
iv. *Every open cover of $M$ has a star-finite open refinement.*
v. *There is a sequence $\langle O_n\rangle$ of open subsets of $M$ with $\bar{O}_n$ compact, $\bar{O}_n \subset O_{n+1}$, and $M = \bigcup O_n$.*
vi. *There is a proper continuous map $\varphi: M \to [0, \infty)$.*
vii. *$M$ is second countable.*

***Proof:*** Since $M$ is connected and Hausdorff, Theorems 21, 25 and Corollary 26 imply that the first six conditions are equivalent. For any topological space (vii) implies (iii). Suppose, on the other hand, that $M$ is Lindelöf. Then $M$ can be covered by a countable number of coordinate balls each of which has a countable base. The union of these bases is a countable base for $M$. Thus (iii) implies (vii) for manifolds. ▌

Manifolds thus fall into three general classes: The compact manifolds; those noncompact manifolds that satisfy one, hence all, of the conditions of Theorem 27, and those that satisfy none of these conditions. The first two clases are customarily referred to as paracompact manifolds.

By an *atlas* for a manifold $M$ we mean a collection $\{\langle U_\alpha, \varphi_\alpha\rangle\}$ of coordinate charts such that $\{U_\alpha\}$ covers $M$. An atlas for $M$ is said to be a *differentiable atlas* if each map $\varphi_\alpha \circ \varphi_\beta^{-1}$ is a differentiable map of $\varphi_\beta[U_\alpha \cap U_\beta]$ into $\mathbf{R}^n$, whenever $U_\alpha \cap U_\beta \neq \emptyset$.

Let $M$ be an $m$-dimensional manifold with a differentiable atlas $\mathcal{A}$ and $N$ an $n$-dimensional manifold with a differentiable atlas $\mathcal{B}$. A map $f$ from $M$ to $N$ is said to be a differentiable map from $\langle M, \mathcal{A}\rangle$ to $\langle N, \mathcal{B}\rangle$ if $\psi_\beta \circ f \circ \varphi_\alpha^{-1}$ is a differentiable map on the set $\varphi_\alpha[U_\alpha \cap f^{-1}[V_\beta]]$ for each $\langle U_\alpha, \varphi_\alpha\rangle \in \mathcal{A}$ and $\langle V_\beta, \psi_\beta\rangle \in \mathcal{B}$. Two differentiable atlases $\mathcal{A}$ and $\mathcal{B}$ on the manifold $M$ are said to be equivalent if the identity map on $M$ is a differentiable map from $\langle M, \mathcal{A}\rangle$ to $\langle M, \mathcal{B}\rangle$ and from $\langle M, \mathcal{B}\rangle$ to $\langle M, \mathcal{A}\rangle$. By a differential structure on $M$ we mean an equivalence class of differentiable atlases, and we call a manifold with a differentiable structure on it a differentiable manifold. Thus a differentiable structure for a manifold $M$ is given when we are given a differentiable atlas for $M$. The concept of a differentiable map from one differentiable manifold to another depends only on the differential structures on them, not the particular atlas giving the structure.

In the case of a differentiable manifold the function $\varphi$ in (vi) of Theorem 27 may be taken to be differentiable.

## Problems

**36.** Let $X = (-1, 1) \cup [2, 3)$, and make $X$ into a topological space by taking as a base all open intervals $(a, b) \subset X$ and all sets of the form $(-\epsilon, 0) \cup [2, 2 + \epsilon)$ for $0 < \epsilon < 1$. Show that $X$ is a locally Euclidean space which is not Hausdorff.

**37.** Some authors do not require manifolds to be connected. Let us call these generalized manifold *not necessarily connected* manifolds.

**a.** Show that every not necessarily connected manifold is the direct union of (connected) manifolds.

**b.** Modify the statement of Theorem 27 to describe the case of not necessarily connected manifolds.

**38.** *The Moore Manifold.* Let $X$ be the set whose elements are the points of the open right half-plane (i.e. $\{\langle x, y\rangle: x > 0\}$) and the lines in the plane with nonnegative slope. Write $m(l)$ and $b(l)$ for the slope and $y$-intercept of the line $l$. We define a base for a topology for $X$ by taking the open disks in the half-plane and the sets

$$V_\epsilon = \{l: |m(l) - m_0| < \epsilon, \quad b(l) = b_0\} \cup \{\langle x, y\rangle: |(y - b_0)/x \pm m_0| < \epsilon, \quad x < \epsilon\}.$$

**a.** Show that $X$ is a connected Hausdorff space and that each point of $X$ is contained in an open set which is homeomorphic to an open subset of $\mathbf{R}^2$. (Such a space is called a two-dimensional manifold or surface.)

**b.** Show that $X$ (or any manifold) is locally compact, completely regular, and satisfies the first axiom of countability.

**c.** Show that $X$ has a countable dense subset but does not satisfy the second axiom of countability.

**d.** Show that $X$ is not normal.

## ★8 The Stone–Čech Compactification

Let $X$ be a completely regular topological space and $\mathcal{F}$ the family of continuous real-valued functions $f$ on $X$ with $|f| \leq 1$. If we let $I = [-1, 1]$, then it follows from Problem 8.48 that $X$ is homeomorphic with a set $E \subset I^{\mathcal{F}}$. Let $F = \bar{E}$. Since $I^{\mathcal{F}}$ is a compact Hausdorff space, the set $F$ is a compact Hausdorff space, and, identifying $X$ with $E$, we have $X$ a dense open subset of $F$. The space $F$ is called the Stone–Čech compactification of $X$ and is denoted by $\beta(X)$. We summarize some of its properties in the following proposition.

**28. Proposition:** *Let $X$ be a completely regular topological space. Then there is a unique compact Hausdorff space $\beta(X)$ with the following properties:*

i. *The space $X$ is a dense subset of $\beta(X)$.*
ii. *Each bounded continuous real-valued function on $X$ extends to a continuous function on $\beta(X)$.*
iii. *If $X$ is a dense open subset of a compact Hausdorff space $Y$, then there is a unique continuous mapping $\varphi$ of $\beta(X)$ onto $Y$ such that $\varphi(x) = x$ for all $x \in X$.*

*If $X$ is locally compact, then $X$ is an open subset of $\beta(X)$.*

### Problems

**39.** Prove Proposition 28:

**a.** If $f$ is a bounded continuous real-valued function on $X$ with $|f| \leq 1$, then $f$ is the restriction to $X$ of $\pi_f$, and $\pi_f$ is continuous on $\beta(X)$.

**b.** Use the fact that $Y$ is a subset of $I^{\mathcal{G}}$, where $\mathcal{G}$ is the space of continuous $g$ on $Y$ with $|g| \leq 1$, to show (iii).

**c.** Show that $\beta(X)$ is unique in the sense that if $Z$ is another space with the same properties, then there is a homeomorphism $\varphi$ of $Z$ with $\beta(X)$ such that $\varphi(x) = x$ for all $x \in X$.

**40.** Let $X$ and $Y$ be the spaces in Problems 11 and 12. Show that $\beta(X) = Y$.

**41.** Let $\mathbf{N}$ be the set of natural numbers. Discuss $\beta(\mathbf{N})$. Show that a sequence from $\mathbf{N}$ converges in $\beta(\mathbf{N})$ if and only if it converges in $\mathbf{N}$. Hence $\beta(\mathbf{N})$ is compact but not sequentially compact.

## 9 The Stone–Weierstrass Theorem

Let $X$ be a compact Hausdorff space. We denote by $C(X)$ the set of all continuous real-valued functions on $X$. Since $X$ is normal, it follows from Urysohn's Lemma that there are enough functions in $C(X)$ to separate points; that is, given two distinct points $x$ and $y$ in $X$, we can find an $f$ in $C(X)$ such that $f(x) \neq f(y)$. The set $C(X)$ is a linear space, since any constant multiple of a continuous real-valued function is continuous and the sum of two continuous functions is continuous. The space $C(X)$ becomes a normed linear space if we define $\|f\| = \max |f(x)|$, and a metric space if we set $\rho(f, g) = \|f - g\|$. As a metric space $C(X)$ is complete.

The space $C(X)$ has also a ring structure: The product $fg$ of two functions $f$ and $g$ in $C(X)$ is again in $C(X)$. A linear space $A$ of functions in $C(X)$ is called an **algebra** if the product of any two elements in $A$ is again in $A$. Thus $A$ is an algebra if for any two functions $f$ and $g$ in $A$ and any real numbers $a$ and $b$ we have $af + bg$ in $A$ and $fg$ in $A$. A family $A$ of functions on $X$ is said to separate points if given distinct points $x$ and $y$ of $X$ there is an $f$ in $A$ such that $f(x) \neq f(y)$. In the present section we study the closed subalgebras of $C(X)$ and prove that if $A$ is a subalgebra of $C(X)$ that separates points, contains the constant functions, and is closed, then $A = C(X)$.

The space $C(X)$ also has a lattice structure: If $f$ and $g$ are in $C(X)$, so is the function $f \wedge g$ defined by $(f \wedge g)(x) = \min [f(x), g(x)]$ and the function $f \vee g$ defined by $(f \vee g)(x) = \max [f(x), g(x)]$. A subset $L$ of $C(X)$ is called a **lattice** if for every pair of functions $f$ and $g$ in $L$ we also have $f \vee g$ and $f \wedge g$ in $L$. It is convenient to investigate subalgebras of $C(X)$ by first investigating lattices of functions. The following proposition can be thought of as a generalization of the Dini Theorem:

**29. Proposition:** *Let $L$ be a lattice of continuous real-valued functions on a compact space $X$, and suppose that the function $h$ defined by*

$$h(x) = \inf_{f \varepsilon L} f(x)$$

*is continuous. Then, given $\epsilon > 0$, there is a $g$ in $L$ such that $0 \leq g(x) - h(x) < \epsilon$ for all $x$ in $X$.*

***Proof:*** For each $x$ in $X$ there is a function $f_x$ in $L$ such that $f_x(x) < h(x) + \epsilon/3$. Since $f_x$ and $h$ are continuous, there is an open set $O_x$ containing $x$ such that

$$|f_x(y) - f_x(x)| < \frac{\epsilon}{3} \quad \text{and} \quad |h(y) - h(x)| < \frac{\epsilon}{3}$$

for all $y \varepsilon O_x$. Hence $f_x(y) - h(y) < \epsilon$ for all $y$ in $O_x$. Now the sets $O_x$ cover $X$, and by compactness there are a finite number of them, say $\{O_{x_1}, \ldots, O_{x_n}\}$, which cover $X$. Let $g = f_{x_1} \wedge f_{x_2} \wedge \cdots \wedge f_{x_n}$. Then $g \varepsilon L$, and given $y$ in $X$ we may choose $i$ so that $y \varepsilon O_{x_i}$, whence

$$g(y) - h(y) \leq f_{x_i}(y) - h(y) < \epsilon. \quad \blacksquare$$

**30. Proposition:** *Let $X$ be a compact space and $L$ a lattice of continuous real-valued functions on $X$ with the following properties:*

i. *$L$ separates points; that is, if $x \neq y$, there is an $f \varepsilon L$ with $f(x) \neq f(y)$.*
ii. *If $f \varepsilon L$, and $c$ is any real number, then $cf$ and $c + f$ also belong to $L$.*

*Then given any continuous real-valued function $h$ on $X$ and any $\epsilon > 0$, there is a function $g \varepsilon L$ such that for all $x \varepsilon X$*

$$0 \leq g(x) - h(x) < \epsilon.$$

Before proving the proposition, we first establish two lemmas.

**31. Lemma:** *Let $L$ be a family of real-valued functions on a compact space $X$ that satisfies properties (i) and (ii) of Proposition 30. Then given any two real numbers $a$ and $b$ and any pair $x$ and $y$ of distinct points of $X$, there is an $f \varepsilon L$ such that $f(x) = a$ and $f(y) = b$.*

***Proof:*** Let $g$ be a function in $L$ such that $g(x) \neq g(y)$. Let

$$f = \frac{a - b}{g(x) - g(y)}\, g + \frac{bg(x) - ag(y)}{g(x) - g(y)}.$$

Then $f \in L$, by property (ii), and $f(x) = a, f(y) = b$. ▮

**32. Lemma:** *Let $L$ be as in Proposition* 30, *$a$ and $b$ real numbers with $a \leq b$, $F$ a closed subset of $X$, and $p$ a point not in $F$. Then there is a function $f$ in $L$ such that $f \geq a, f(p) = a$, and $f(x) > b$ for all $x \in F$.*

***Proof:*** By Lemma 31 we can choose, for each $x \in F$, a function $f_x$ such that $f_x(p) = a$ and $f_x(x) = b + 1$. Let $O_x = \{y: f_x(y) > b\}$. Then the sets $\{O_x\}$ cover $F$, and since $F$ is compact, there are a finite number $\{O_{x_1}, \ldots, O_{x_n}\}$ that cover $F$. Let $f = f_{x_1} \vee \cdots \vee f_{x_n}$. Then $f \in L, f(p) = a$, and $f > b$ on $F$. If we replace $f$ by $f \vee a$, then we also have $f \geq a$ on $X$. ▮

***Proof of Proposition 30:*** Since $L$ is nonempty, it follows from (ii) that the constant functions belong to $L$. Given $g \in C(X)$, let $L' = \{f: f \in L \text{ and } f \geq g\}$. Proposition 30 will follow from Proposition 29 if we can show that for each $p \in X$ we have $g(p) = \inf f(p), f \in L'$. Choose a positive real number $\eta$. Since $g$ is continuous, the set

$$F = \{x: g(x) \geq g(p) + \eta\}$$

is closed. Since $X$ is compact, $g$ is bounded on $X$, say by $M$. By Lemma 32 we can find a function $f \in L$ such that $f \geq g(p) + \eta$, $f(p) = g(p) + \eta$, and $f(x) > M$ on $F$. Since $g < g(p) + \eta$ on $\tilde{F}$, we have $g < f$ on $X$. Thus $f \in L'$, and $f(p) \leq g(p) + \eta$. Since $\eta$ was an arbitrary positive number, we have $g(p) = \inf f(p), f \in L'$. ▮

**33. Lemma:** *Given $\epsilon > 0$, there is a polynomial $P$ in one variable such that for all $s \in [-1, 1]$ we have $|P(s) - |s|| < \epsilon$.*

***Proof:*** Let $\sum_{n=0}^{\infty} c_n t^n$ be the binomial series for $(1 - t)^{1/2}$. This series converges uniformly for $t$ in the interval $[0, 1]$. Hence, given $\epsilon > 0$, we can choose $N$ so that for all $t \in [0, 1]$ we have

$$|(1 - t)^{1/2} - Q_N(t)| < \epsilon,$$

where $Q_N = \sum_{n=0}^{N} c_n t^n$. Let $P(s) = Q_N(1 - s^2)$. Then $P$ is a polynomial in $s$, and $||s| - P(s)| < \epsilon$ for $s \in [-1, 1]$. ▮

**34. Theorem (Stone–Weierstrass):** *Let $X$ be a compact space and $A$ an algebra of continuous real-valued functions on $X$ that separates* *the points of $X$ and contains the constant functions. Then given any continuous real-valued function $f$ on $X$ and any $\epsilon > 0$ there is a function $g$ in $A$ such that for all $x$ in $X$ we have $|g(x) - f(x)| < \epsilon$. In other words, $A$ is a dense subset of $C(X)$.*

***Proof:*** Let $\bar{A}$ denote the closure of $A$ considered as a subset of $C(X)$. Thus $\bar{A}$ consists of those functions on $X$ that are uniform limits of sequences of functions from $A$. It is easy to verify that $\bar{A}$ is itself an algebra of continuous real-valued functions on $X$. The theorem is equivalent to the statement that $\bar{A} = C(X)$. This will follow from Proposition 30 if we can show that $\bar{A}$ is a lattice. Let $f \in \bar{A}$ and $\|f\| \leq 1$. Then given $\epsilon > 0$, $\||f| - P(f)\| < \epsilon$, where $P$ is the polynomial given in Lemma 33. Since $\bar{A}$ is an algebra containing the constants, $P(f) \in \bar{A}$, and since $\bar{A}$ is a closed subset of $C(X)$, we have $|f| \in A$. If now $f$ is any function in $A$, then $f/\|f\|$ has norm 1, and so $|f|/\|f\|$ and hence also $|f|$ belong to $\bar{A}$. Thus $\bar{A}$ contains the absolute value of each function which is in $\bar{A}$. But

$$f \vee g = \tfrac{1}{2}(f + g) + \tfrac{1}{2}|f - g|$$

and

$$f \wedge g = \tfrac{1}{2}(f + g) - \tfrac{1}{2}|f - g|.$$

Thus $\bar{A}$ is a lattice and must be $C(X)$ by Proposition 30. ▮

**35. Corollary:** *Every continuous function on a closed bounded set $X$ in $\mathbf{R}^n$ can be uniformly approximated on $X$ by a polynomial (in the coordinates).*

***Proof:*** The set of all polynomials in the coordinate functions forms an algebra containing the constants. It separates points, since given two distinct points in $\mathbf{R}^n$, one of the coordinate functions takes different values on these points. Hence Theorem 34 applies. ▮

## Problems

**42.** Let $f$ be a continuous periodic real-valued function on $\mathbf{R}$ with period $2\pi$; that is, $f(x + 2\pi) = f(x)$. Show that, given $\epsilon > 0$, there is a finite Fourier series $\varphi$, given by $\varphi(x) = a_0 + \sum_{n=1}^{N} (a_n \cos nx + b_n \sin nx)$, such that $|\varphi(x) - f(x)| < \epsilon$ for all $x$. [Hint: Note that periodic functions are really functions on the circumference of the unit circle, and that $\cos mx \cos nx = \frac{1}{2}\{\cos(m+n)x - \cos(m-n)x\}$, etc.]

**43.** Let $A$ be an algebra of continuous real-valued functions on a compact space $X$, and assume that $A$ separates the points of $X$. Then either $\bar{A} = C(X)$ or there is a point $p \in X$ and $\bar{A} = \{f: f \in C(X), f(p) = 0\}$. [Hint: If $1 \in \bar{A}$, we are done. If $\exists f \in A$, which is nowhere zero, then $1 \in \bar{A}$. If for each $x \in X$ there is an $f \in A$ with $f(x) \neq 0$, then $\exists g \in A$ with $g > 0$ everywhere.]

**44.** Let $\mathcal{F}$ be a family of continuous real-valued functions on a compact Hausdorff space $X$, and suppose that $\mathcal{F}$ separates the points of $X$. Then every continuous real-valued function on $X$ can be uniformly approximated by a polynomial in a finite number of functions of $\mathcal{F}$.

**45. a.** Let $X$ be a topological space and $A$ a set of real-valued continuous functions on $X$. Define $x \equiv y$ if $f(x) = f(y)$ for all $f \in A$. Show that $\equiv$ is an equivalence relation.

**b.** Let $\tilde{X}$ be the set of equivalence classes of $\equiv$ and $\varphi$ the natural map of $X$ into $\tilde{X}$. Show that for each $f \in A$ there is a unique real-valued function $\tilde{f}$ on $\tilde{X}$ such that $f = \tilde{f} \circ \varphi$.

**c.** Let $\tilde{X}$ have the weak topology generated by these $\tilde{f}$. Then $\varphi$ is continuous.

**d.** If $X$ is compact, then so is $\tilde{X}$, and the functions $\tilde{f}$ in part (b) are continuous.

**e.** Let $X$ be a compact space and $A$ a closed subalgebra of $C(X)$ containing the constant functions. Then there is a compact Hausdorff space $\tilde{X}$ and a mapping $\varphi$ of $X$ onto $\tilde{X}$ such that $A$ is the set of all functions $f$ of the form $\tilde{f} \circ \varphi$ with $\tilde{f} \in C(\tilde{X})$.

**46.** Let $X$ and $Y$ be compact spaces. Then for each continuous real-valued function $f$ on $X \times Y$ and each $\epsilon > 0$, we can find continuous real-valued functions $g_1, \ldots, g_n$ on $X$ and $h_1, \ldots, h_n$ on $Y$ such that for each $\langle x, y \rangle \in X \times Y$ we have

$$\left| f(x, y) - \sum_{i=1}^{n} g_i(x) h_i(y) \right| < \epsilon.$$

**47.** The Weierstrass theorem, which states that a continuous function on a cube in $\mathbf{R}^n$ can be uniformly appropriated by a polynomial, can be proved directly by giving an integral formula for the approximating polynomials.[2] Show that this special case implies the general Stone–Weierstrass Theorem by showing that the functions of norm 1 in the algebra $\mathfrak{A}$ give a mapping of $X$ into the infinite-dimensional cube $\bigtimes \{I_f : f \in \mathfrak{A}, \|f\| = 1\}$. Use the Tietze Extension Theorem and Problem 17 to show that each continuous function

[2] See, for example, R. Courant and D. Hilbert, *Mathematische Physik*, Bd. I (Berlin: Springer-Verlag, 1931), pp. 55–57.

on the image of X can be approximated by a polynomial in (a finite number of) the coordinate functions.

The Stone–Weierstrass Theorem gives precise information about approximation by functions in an algebra of continuous functions. A natural question is the nature of functions that can be approximated by a ring of real-valued continuous functions, that is to say, when we no longer postulate the possibility of multiplying by arbitrary real numbers. The next three problems give some results in this direction.

**48.** Let $I$ be the interval $[-1, 1]$ in $\mathbf{R}$ and $f$ a continuous real-valued function on $I$ such that $f(-1)$, $f(0)$, and $f(1)$ are integers and $f(1) \equiv f(-1) \bmod 2$. Then given $\epsilon > 0$, there is a polynomial $P$ with integral coefficients such that $|f(x) - P(x)| < \epsilon$ for all $x \in I$. Hints:

**a.** Let $\varphi$ be the polynomial defined by $\varphi(x) = x + x(1 - 2x)(1 - x)$. Then $\varphi$ is a monotone increasing function whose fixed points are $0, \frac{1}{2}$, and 1.

**b.** Choose $\epsilon > 0$. Then some iterate $\varphi_n$ of $\varphi$ is a polynomial with integral coefficients which is monotone increasing on $[0, 1]$ and such that $|\varphi_n(x) - \frac{1}{2}| < \epsilon$ for $x \in [\epsilon, 1 - \epsilon]$.

**c.** Given a number $\alpha$, $0 < \alpha < 1$, and any $\epsilon > 0$, then there is a polynomial $\psi$ with integral coefficients (and no constant term) such that $0 \leq \psi(x) \leq 1$ in $[0, 1]$ and $|\psi(x) - \alpha| < \epsilon$ for all $x$ in $[\epsilon, 1 - \epsilon]$.

**d.** Let $P$ be a polynomial with integral coefficients, and suppose that $P(-1) = P(0) = P(1) = 0$. Let $\beta$ be any real number. Then for each $\epsilon > 0$, $\beta P$ can be uniformly approximated to within $\epsilon$ on $[-1, 1]$ by a polynomial having integral coefficients and no constant term.

**e.** Reduce the statement of the problem to (d) and the Stone–Weierstrass Theorem.

**49. a.** Let $X$ be a set and $R$ a ring of real-valued functions on $X$. Let $\bar{R}$ be the ring of all real-valued functions that can be uniformly approximated by functions in $R$. If $f \in \bar{R}$, and $\sup_X |f(x)| < 1$, then $cf \in \bar{R}$ for each real number $c$.

**b.** Let $X$ be a compact Hausdorff space and $R$ a ring of continuous real-valued functions on $X$ such that $1 \in R$, and for each pair of distinct points $x$ and $y$ there is a function $f$ in $R$ such that $f(x) \neq f(y)$, and $|f(z)| < 1$ for all $z \in X$. Then every continuous real-valued function on $X$ can be uniformly approximated by functions in $R$.

**50 a.** The statement of Problem 48 can be improved slightly. Show, for example, that we may take the interval $I$ to be any closed interval contained in $(-\sqrt{2}, \sqrt{2})$. [The polynomial $x^2 - 1$ has absolute value at most 1 on $I$. Apply Problem 49b.]

**b.** Show that we cannot take $I$ in Problem 48 to be $[-2, 2]$. [Hint: If $P$ is a polynomial with integral coefficients,

$$\frac{1}{\pi}\int_{-2}^{2} P(x)(4 - x^2)^{-1/2}\, dx$$

is an integer.]

# 10 Banach Spaces

## 1 Introduction

We are going to study a class of spaces that are endowed with both a topological and an algebraic structure. A set $X$ of elements is called a **vector space** (or linear space, or linear vector space) over the reals if we have a function $+$ on $X \times X$ to $X$ and a function $\cdot$ on $\mathbf{R} \times X$ to $X$ that satisfy the following conditions:

i. $x + y = y + x$.
ii. $(x + y) + z = x + (y + z)$.
iii. There is a vector $\theta$ in $X$ such that $x + \theta = x$ for all $x$ in $X$.
iv. $\lambda(x + y) = \lambda x + \lambda y$; $\lambda \in \mathbf{R}$, $x, y \in X$.
v. $(\lambda + \mu)x = \lambda x + \mu x$; $\lambda, \mu \in \mathbf{R}$, $x \in X$.
vi. $\lambda(\mu x) = (\lambda\mu)x$; $\lambda, \mu \in \mathbf{R}$, $x \in X$.
vii. $0 \cdot x = \theta$, $1 \cdot x = x$.

We call $+$ addition and $\cdot$ multiplication by scalars. It should be noted that the element $\theta$ defined in (iii) is unique, for if $\theta'$ also has this property, then $\theta = \theta + \theta' = \theta' + \theta = \theta'$. The element $(-1)x$ is called the negative of $x$ and written $-x$. We have $x + (-x) = 1 \cdot x + (-1)x = (1 - 1)x = 0 \cdot x = \theta$.

A nonnegative real-valued function $\| \ \|$ defined on a vector space is called a **norm** if

i. $\|x\| = 0 \Leftrightarrow x = \theta$.
ii. $\|x + y\| \le \|x\| + \|y\|$.
iii. $\|\alpha x\| = |\alpha| \|x\|$.

A normed vector space becomes a metric space if we define a metric $\rho$ by $\rho(x, y) = \|x - y\|$. When we speak about metric properties in a normed space, we are referring to this metric.

If a normed vector space is complete in this metric, it is called a Banach space. Examples of Banach spaces were given in Chapter 6. Another example is $C(X)$, the space of all continuous real-valued functions on a compact space $X$. We restate here Proposition 6.5, and note that the proof given in Chapter 6 is valid in any normed vector space.

**1. Proposition:** *A normed vector space is complete if and only if every absolutely summable sequence is summable.*

A nonempty subset $S$ of a vector space $X$ is a **subspace** or **linear manifold** if $\lambda_1 x_1 + \lambda_2 x_2$ belongs to $S$ whenever $x_1$ and $x_2$ do. If $S$ is also closed as a subset of $X$, then it is called a closed linear manifold. The intersection of any family of linear manifolds is a linear manifold. Hence, given a set $A$ in $X$, there is always a smallest linear manifold containing $A$. We often denote this manifold by $\{A\}$.

If $A$ is any set in $X$, we use $A + x$ to denote the set of all elements $z$ of the form $z = x + y$, $y \in A$. The set $A + x$ is called the translate of $A$ by $x$. The set $\lambda A$ is the set of all elements of the form $\lambda x$ with $x \in A$, and $A + B$ is the set of all elements of the form $x + y$ with $x$ in $A$ and $y$ in $B$.

## Problems

**1.** Show that if $x_n \to x$, then $\|x_n\| \to \|x\|$.

**2.** Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are called equivalent if there is a positive constant $K$ such that $K^{-1}\|x\|_1 \le \|x\|_2 \le K\|x\|_1$. If the norms are equivalent, then the metrics derived from them are uniformly equivalent. Show that the metrics introduced in Problem 7.10b for $\mathbf{R}^n$ are derived from norms for $\mathbf{R}^n$, and these norms are all equivalent.

**3.** Show that $+$ is a continuous function from $X \times X$ into $X$ and that $\cdot$ is a continuous function from $\mathbf{R} \times X$ into $X$.

**4.** Show that a nonempty set $M$ is a linear manifold if and only if $M + M = M$ and $\lambda M = M$ for each $\lambda$.

**5. a.** Prove that the intersection of a family of linear manifolds is a manifold.

**b.** Prove that there is a smallest linear manifold $\{A\}$ containing a given set $A$.

**c.** Show that $\{A\}$ consists of all finite linear combinations of the form $\lambda_1 x_1 + \cdots + \lambda_n x_n$ with $x_i \in A$.

**6. a.** If $M$ and $N$ are linear manifolds, so is $M + N$, and $M + N = \{M \cup N\}$.

**b.** If $M$ is a linear manifold, so is $\overline{M}$.

**7.** Show that the set $P$ of all polynomials on $[0, 1]$ is a linear manifold in $C[0, 1]$. Is it closed? Give an example of a closed linear manifold in $C[0, 1]$.

**8.** A linear manifold $M$ is said to be finite-dimensional if there are a finite number of elements $x_1, \ldots, x_n$ such that $M = \{x_1, \ldots, x_n\}$. Prove that every finite-dimensional linear manifold in a normed vector space $X$ must be closed.

**9.** Let $S$ be the spheroid of radius 1 centered at $\theta$; that is, $S = \{x: \|x\| < 1\}$. Prove that $S$ is open and that

$$\bar{S} = \{x: \|x\| \leq 1\}.$$

We call $S$ the open unit *sphere* (or *ball*) and $\bar{S}$ the closed unit sphere (or ball).

**10.** A nonnegative real-valued function $\| \ \|$ defined on a vector space $X$ is called a **pseudonorm** if $\|x + y\| \leq \|x\| + \|y\|$ and $\|\alpha x\| = |\alpha| \|x\|$. Show that the relation $x \equiv y$ defined by $\|x - y\| = 0$ is an equivalence relation compatible with addition and multiplication by scalars, and that, if $x \equiv y$, then $\|x\| = \|y\|$. Let $X'$ be the set of equivalence classes of $X$ under $\equiv$. Then $X'$ becomes a normed vector space if we define $\alpha x' + \beta y'$ as the (unique) equivalence class which contains $\alpha x + \beta y$ for $x \in x'$ and $y \in y'$ and define $\|x'\| = \|x\|$ for $x \in x'$. The mapping $\varphi$ of $X$ onto $X'$ that takes each element of $X$ into the equivalence class to which it belongs is a homomorphism (called the natural homomorphism) of $X$ onto $X$. What is the kernel of $\varphi$? Illustrate this procedure with the $L^p$ spaces on $[0, 1]$.

**11.** Let X be a normed linear space (with norm $\| \cdot \|$) and $M$ a linear manifold in $X$. Show that $\|x\|_1 = \inf_{m \in M} \|x - m\|$ defines a pseudonorm on $X$. Let $X'$ be the normed linear space derived from $X$ and the pseudonorm $\| \cdot \|_1$ using the process described in Problem 10. The natural map $\varphi$ of $X$ onto $X'$ has kernel $\bar{M}$. Prove that $\varphi$ takes open sets into open sets. The space $X'$ is usually denoted by $X/\bar{M}$ and called the *quotient space* of $X$ modulo $\overline{M}$.

**12.** Show that, if $X$ is complete and $M$ is a closed linear manifold of $X$, then $X/M$ is also complete. [Hint: Use Proposition 1.]

## 2 Linear Operators

A mapping $A$ of a vector space $X$ into a vector space $Y$ is called a linear mapping, a *linear operator*, or a *linear transformation* if

$$A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A x_1 + \alpha_2 A x_2$$

for all $x_1$, $x_2$ in $X$ and all real $\alpha_1$, $\alpha_2$. If $X$ and $Y$ are normed vector spaces, we call a linear operator $A$ *bounded* if there is a constant $M$ such that for all $x$ we have $\|Ax\| \leq M\|x\|$. We call the least such $M$ the *norm* of $A$ and denote it by $\|A\|$. Thus

$$\|A\| = \sup_{\substack{x \in X \\ x \neq \theta}} \frac{\|Ax\|}{\|x\|}.$$

Since $A(\alpha x) = \alpha A x$, we have also

$$\|A\| = \sup_{\|x\| = 1} \|Ax\| = \sup_{\|x\| \leq 1} \|Ax\|.$$

A bounded linear transformation $A$ from $X$ to $Y$ is called an *isomorphism* between $X$ and $Y$ if there is a bounded linear transformation $B$ from $Y$ to $X$ such that $AB$ is the identity on $Y$ and $BA$ the identity on $X$. The following proposition relates the notions of boundeness and continuity for linear operators:

**2. Proposition:** *A bounded linear operator is uniformly continuous. If a linear operator is continuous at one point, it is bounded.*

***Proof:*** Suppose $A$ is bounded. Then

$$\|Ax_1 - Ax_2\| \leq \|A\| \cdot \|x_1 - x_2\| < \epsilon$$

for all $x_1$ and $x_2$ in $X$ with $\|x_1 - x_2\| < \epsilon/\|A\|$. Thus $A$ is uniformly continuous.

Suppose now that $A$ is a linear operator that is continuous at $x_0$. Then there is a $\delta > 0$ such that $\|Ax - Ax_0\| < 1$ for all $x$ such that $\|x - x_0\| < \delta$. For any $z$ in $X$ with $z \neq \theta$, set $w = \eta z/\|z\|$, where $0 < \eta < \delta$. Then

$$\frac{\eta}{\|z\|} Az = Aw = A(w + x_0) - A(x_0)$$

and

$$\frac{\eta}{\|z\|} \|Az\| = \|A(w + x_0) - A(x_0)\| < 1,$$

since $\|w + x_0 - x_0\| = \|w\| = \eta < \delta$. Consequently, $\|Az\| \leq \eta^{-1}\|z\|$, and $A$ is bounded. ∎

**3. Proposition:** *The space $\mathfrak{B}$ of all bounded linear operators from a normed vector space $X$ to a Banach space $Y$ is itself a Banach space.*

***Proof:*** If $A$ and $B$ are in $\mathfrak{B}$, we define $\alpha A + \beta B$ by $(\alpha A + \beta B)x = \alpha Ax + \beta Bx$; it is easily seen to be a linear operator. Now

$$\|\lambda A\| = \sup_{\|x\|=1} \|\lambda Ax\| = |\lambda| \sup_{\|x\|=1} \|Ax\| = |\lambda| \|A\|$$

and

$$\|A + B\| = \sup_{\|x\|=1} \|Ax + Bx\| \leq \sup_{\|x\|=1} (\|Ax\| + \|Bx\|) \leq \|A\| + \|B\|.$$

Thus any linear combination of two bounded linear operators is again a bounded linear operator. If $\|A\| = 0$, then

$$\|Ax\| \leq \|A\| \cdot \|x\| = 0,$$

and so $Ax = \theta$. Thus $\|A\| = 0$ only for the operator 0, which maps every $x$ into $\theta$. Thus $\| \ \|$ satisfies all the requirements of a norm and we have only to show that if $Y$ is complete so is $\mathfrak{B}$.

Let $\langle A_n \rangle$ be a Cauchy sequence from $\mathfrak{B}$. For each $x \in X$ we have $\|A_n x - A_m x\| \leq \|A_n - A_m\| \cdot \|x\|$, and so $\langle A_n x \rangle$ is a Cauchy sequence in $Y$ and must converge to an element $y$ in $Y$. Call this element $Ax$. It follows from the definition of $Ax$ that $A(\lambda x) = \lambda Ax$ and $A(x_1 + x_2) = Ax_1 + Ax_2$.

To show that the linear operator $A$ is bounded, we observe that, given $\epsilon > 0$, there is an $N$ such that for all $m$, $n \geq N$ we have $\|A_n - A_m\| < \epsilon$. Hence $\|A_n\| < \|A_N\| + \epsilon$ for all $n \geq N$, and so

$$\|Ax\| = \lim \|A_n x\| \leq (\|A_N\| + \epsilon)\|x\|.$$

Thus $A$ is bounded. For each $x$ in $X$ we have

$$\begin{aligned} \|A_n x - Ax\| &= \lim_{m\to\infty} \|A_n x - A_m x\| \\ &\leq \overline{\lim_{m\to\infty}} \|A_n - A_m\| \|x\| \\ &\leq \epsilon \|x\| \end{aligned}$$

for all $n \geq N$. Thus for $n \geq N$,

$$\|A_n - A\| = \sup_{\|x\|=1} \|(A_n - A)x\| \leq \epsilon.$$

Thus $A_n \to A$ and $\mathcal{B}$ is complete. ∎

## Problems

**13.** Show that if $A_n \to A$ and $x_n \to x$, then $A_n x_n \to Ax$.

**14.** The *kernel* of an operator $A$ is the set $\{x\colon Ax = \theta\}$. Prove that the kernel of a linear operator is a linear manifold and that the kernel of a continuous operator is closed.

**15. a.** Let $X$ be a normed linear space and $M$ a closed linear manifold. Then the natural homomorphism $\varphi$ of $X$ onto $X/M$ has norm 1.

**b.** Let $X$ and $Y$ be normed linear spaces and $A$ a bounded linear operator from $X$ into $Y$ whose kernel is $M$. Then there is a unique bounded linear operator $B$ from $X/M$ into $Y$ such that $A = B \circ \varphi$. Moreover, $\|A\| = \|B\|$.

**16.** Let $X$ be a metric space, and let $Y$ be the space consisting of those real-valued functions $f$ on $X$ that vanish at a fixed point $x_0 \in X$ and satisfy $|f(x) - f(y)| \leq M\rho(x, y)$ for some $M$ (depending on $f$). Define

$$\|f\| = \sup \frac{|f(x) - f(y)|}{\rho(x, y)}.$$

Then $Y$ is a normed linear space. For each $x \in X$, the functional $F_x$ defined by $F_x(f) = f(x)$ is a bounded linear functional on $Y$, and $\|F_x - F_y\| = \rho(x, y)$. Thus $X$ is isometric to a subset of the space $Y^*$ of bounded linear operators for $Y$ to $\mathbf{R}$. Since $Y^*$ is complete by Proposition 3, the closure of this subset gives a completion of $Y$, and we have another proof of Theorem 7.9.

## 3 Linear Functionals and the Hahn–Banach Theorem

A linear functional on a vector space $X$ is a linear operator from $X$ to the space $\mathbf{R}$ of real numbers. Thus a linear functional is a real-valued function $f$ on $X$ such that $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$. The first question with which we shall be concerned is that of extending a linear functional from a subspace to the whole space $X$ in such a manner that various properties of the functional are preserved. The principal result in this direction is the following:

**4. Theorem (Hahn–Banach):** *Let $p$ be a real-valued function defined on the vector space $X$ satisfying $p(x + y) \le p(x) + p(y)$ and $p(\alpha x) = \alpha p(x)$ for each $\alpha \ge 0$. Suppose that $f$ is a linear functional defined on a subspace $S$ and that $f(s) \le p(s)$ for all $s$ in $S$. Then there is a linear functional $F$ defined on $X$ such that $F(x) \le p(x)$ for all $x$, and $F(s) = f(s)$ for all $s$ in $S$.*

***Proof:*** Consider all linear functionals $g$ defined on a subspace of $X$ and satisfying $g(x) \le p(x)$ whenever $g(x)$ is defined. This set is partially ordered by setting $g_1 \prec g_2$ if $g_2$ is an extension of $g_1$, that is, if the domain of $g_1$ is contained in the domain of $g_2$ and $g_1 = g_2$ on the domain of $g_1$.

By the Hausdorff Maximal Principle there is a maximal linearly ordered subfamily $\{g_\alpha\}$ that contains the given functional $f$. We define a functional $F$ on the union of the domains of the $g_\alpha$ by setting $F(x) = g_\alpha(x)$ if $x$ is in the domain of $g_\alpha$. Since $\{g_\alpha\}$ is linearly ordered, this definition is independent of the $\alpha$ chosen. The domain of $F$ is a subspace and $F$ is a linear functional, for if $x$ and $y$ are in the domain of $F$, then $x \in$ domain $g_\alpha$ and $y \in$ domain $g_\beta$ for some $\alpha$, $\beta$. By the linear ordering of $\{g_\alpha\}$, we have either $g_\alpha \prec g_\beta$ or $g_\beta \prec g_\alpha$, say the former. Then $x$ and $y$ are in the domain of $g_\beta$, and so $\lambda x + \mu y$ is in the domain of $g_\beta$ and so in the domain of $F$, and $F(\lambda x + \mu y) = g_\beta(\lambda x + \mu y) = \lambda g_\beta(x) + \mu g_\beta(y) = \lambda F(x) + \mu F(y)$. Thus $F$ is an extension of $f$. Moreover, $F$ is a maximal extension. For if $G$ is any extension of $F$, $g_\alpha \prec F \prec G$ implies that $G$ must belong to $\{g_\alpha\}$ by the maximality of $\{g_\alpha\}$. Hence $G \prec F$, and so $G = F$.

It remains only to show that $F$ is defined for all $x \in X$. Since $F$ is maximal, this will follow if we can show that each $g$ that is defined on a proper subspace $T$ of $X$ and satisfies $g(t) \le p(t)$ has a proper extension $h$.

Let $y$ be an element in $X \sim T$. We shall show that $g$ may be extended to the subspace $U$ spanned by $T$ and $y$, that is, to the subspace consisting of elements of the form $\lambda y + t$ with $t \in T$. If $h$ is an extension of $g$, we must have

$$h(\lambda y + t) = \lambda h(y) + h(t) = \lambda h(y) + g(t),$$

and so $h$ is defined as soon as we specify $h(y)$.

For $t_1, t_2, \in T$ we have

$$g(t_1) + g(t_2) = g(t_1 + t_2) \le p(t_1 + t_2) \le p(t_1 - y) + p(t_2 + y).$$

Hence

$$-p(t_1 - y) + g(t_1) \le p(t_2 + y) - g(t_2),$$

and so

$$\sup_{t \in T} [-p(t - y) + g(t)] \le \inf_{t \in T} [p(t + y) - g(t)].$$

Define $h(y) = \alpha$, where $\alpha$ is a real number such that

$$\sup [-p(t - y) + g(t)] \le \alpha \le \inf [p(t + y) - g(t)].$$

We must now show that

$$h(\lambda y + t) = \lambda\alpha + g(t) \le p(\lambda y + t).$$

If $\lambda > 0$, then

$$\begin{aligned}\lambda\alpha + g(t) &= \lambda[\alpha + g(t/\lambda)] \\ &\le \lambda[\{p(t/\lambda + y) - g(t/\lambda)\} + g(t/\lambda)] \\ &= \lambda p(t/\lambda + y) = p(t + \lambda y).\end{aligned}$$

If $\lambda = -\mu < 0$, then

$$\begin{aligned}-\alpha\mu + g(t) &= \mu(-\alpha + g(t/\mu)) \\ &\le \mu(\{p(t/\mu - y) - g(t/\mu)\} + g(t/\mu)) \\ &= \mu p(t/\mu - y) = p(t - \mu y).\end{aligned}$$

Thus $h(\lambda y + t) \le p(\lambda y + t)$ for all $\lambda$, and $h$ is a proper extension of $g$. ▮

The Hahn–Banach Theorem has a wide range of applications, many of them involving a clever choice of the subadditive function $p$. Propositions 6 and 7 and Theorem 20 are applications of this sort. The following proposition is a generalization of the Hahn–Banach Theorem which is useful in certain applications (cf. Problems 20 and 21). By an Abelian semigroup of linear operators on a vector space $X$ we mean a collection $G$ of linear operators from $X$ to $X$ such that if $A$ and $B$ are in $G$, then $AB = BA$ and $AB$ is in $G$. We also assume that the identity operator belongs to $G$.

**5. Proposition:** *Let $X$, $S$, $p$, and $f$ be as in Theorem 4, and let $G$ be an Abelian semigroup of linear operators on $X$ such that for every $A$ in $G$ we have $p(Ax) \le p(x)$ for all $x$ in $X$, while for each $s$ in $S$ we have $As$ in $S$ and $f(As) = f(s)$. Then there is an extension $F$ of $f$ to a linear functional on $X$ such that $F(x) \le p(x)$ and $F(Ax) = F(x)$ for all $x$ in $X$.*

***Proof:*** Define a function $q$ on $X$ by setting

$$q(x) = \inf \frac{1}{n} p(A_1 x + \cdots + A_n x),$$

where the inf is taken over all finite sequences $\langle A_1, \ldots, A_n \rangle$ from $G$. We clearly have $q(x) \le p(x)$ and $q(\alpha x) = \alpha q(x)$ for $\alpha \ge 0$. For any $x$ and $y$ in $X$ and any pairs $\langle A_1, \ldots, A_n \rangle$ and $\langle B_1, \ldots, B_m \rangle$ of finite sequences from $G$ we have

$$\begin{aligned} q(x+y) &\le \frac{1}{nm} p\left(\sum_{i=1}^{n} \sum_{j=1}^{m} A_i B_j(x+y)\right) \\ &\le \frac{1}{nm} p\left(\sum_{j=1}^{m} B_j\left(\sum_{i=1}^{n} A_i x\right)\right) + \frac{1}{nm} p\left(\sum_{i=1}^{n} A_i\left(\sum_{j=1}^{m} B_j y\right)\right) \\ &\le \frac{1}{n} p\left(\sum_{i=1}^{n} A_i x\right) + \frac{1}{m} p\left(\sum_{j=1}^{m} B_j y\right). \end{aligned}$$

Taking infima over every pair $\langle A_i \rangle$, $\langle B_j \rangle$, we obtain

$$q(x+y) \le q(x) + q(y).$$

Since $q(\theta) = p(\theta) = 0$, we have

$$\begin{aligned} 0 = q(x - x) &\le q(x) + q(-x) \\ &\le q(x) + p(-x). \end{aligned}$$

Thus $q(x)$ cannot be $-\infty$, and $q$ is real valued.

For $s$ in $S$,

$$f(s) = \frac{1}{n} f(A_1 s + \cdots + A_n s) \le \frac{1}{n} p(A_1 s + \cdots + A_n s).$$

Hence $f(s) \le q(s)$, and we may apply Theorem 4 with $p$ replaced by $q$ to obtain an extension $F$ of $f$ to all $X$ such that $F(x) \le q(x) \le p(x)$. It remains only to show that $F(Ax) = F(x)$. Now

$$\begin{aligned} q(x - Ax) &\le \frac{1}{n} p((x - Ax) + A(x - Ax) + \cdots + A^n(x - Ax)) \\ &= \frac{1}{n} p(x - A^{n+1}x) \le \frac{1}{n} [p(x) + p(-x)]. \end{aligned}$$

Since this is true for each $n$, we have $q(x - Ax) \le 0$. Since

$$F(x) - F(Ax) = F(x - Ax) \le q(x - Ax) \le 0,$$

we have $F(x) \le F(Ax)$, and applying this to $-x$, we get $F(x) = F(Ax)$. ∎

**6. Proposition:** *Let $x$ be an element in a normed vector space $X$. Then there is a bounded linear functional $f$ on $X$ such that $f(x) = \|f\|\|x\|$.*

***Proof:*** Let $S$ be the subspace consisting of all multiples of $x$, and define $f$ on $S$ by $f(\lambda x) = \lambda\|x\|$ and set $p(y) = \|y\|$. Then by the Hahn–Banach Theorem there is an extension of $f$ to be a linear functional on $X$ such that $f(y) \leq \|y\|$. Since $f(-y) \leq \|y\|$, we have $\|f\| \leq 1$. Also $f(x) = \|x\| \leq \|f\| \cdot \|x\|$. Thus $\|f\| = 1$ and $f(x) = \|f\| \cdot \|x\|$. ▮

**7. Proposition:** *Let $T$ be a linear subspace of a normed linear space $X$ and $y$ an element of $X$ whose distance to $T$ is at least $\delta$, that is, an element such that $\|y - t\| \geq \delta$ for all $t \in T$. Then there is a bounded linear functional $f$ on $X$ with $\|f\| \leq 1, f(y) = \delta$, and such that $f(t) = 0$ for all $t$ in $T$.*

***Proof:*** Let $S$ be the subspace spanned by $T$ and $y$, that is, the subspace consisting of all elements of the form $\alpha y + t$ with $t \in T$. Define $f(\alpha y + t) = \alpha\delta$. Then $f$ is a linear functional on $S$, and since $\|\alpha y + t\| = |\alpha| \cdot \|y + t/\alpha\| \geq \alpha\delta$, we have $f(s) \leq \|s\|$ on $S$. By the Hahn–Banach Theorem we may extend $f$ to all of $X$ so that $f(x) \leq \|x\|$. But this implies that $\|f\| \leq 1$. By the definition of $f$ on $S$, we have $f(t) = 0$ for $t \in T$ and $f(y) = \delta$. ▮

The space of bounded linear functionals on a normed space $X$ is called the **dual** (or conjugate) of $X$ and is denoted by $X^*$. Since $\mathbf{R}$ is complete, the dual $X^*$ of any normed space $X$ is a Banach space by Proposition 3. Two normed vector spaces are said to be isometrically isomorphic if there is a one-to-one linear mapping of one of them onto the other which preserves norms. From an abstract point of view, isometrically isomorphic spaces are identical, the isomorphism merely amounting to a renaming of the elements. In Chapter 6 we saw that the dual of $L^p$ was (isometrically isomorphic to) $L^q$ for $1 \leq p < \infty$ and that there was a natural representation of the bounded linear functionals on $L^p$ by elements of $L^q$.

We are now in a position to show that a similar representation does not hold for bounded linear functionals on $L^\infty[0, 1]$. We note that $C[0, 1]$ is a closed subspace of $L^\infty[0, 1]$. Let $f$ be that linear functional on $C[0, 1]$ which assigns to each $x$ in $C[0, 1]$ the value $x(0)$ of $x$ at 0. It has norm 1 on $C$, and so can be extended to a bounded linear functional $F$ on $L^\infty[0, 1]$. Now there is no $y$ in

$L^1[0, 1]$ such that $F(x) = \int_0^1 xy\,dt$ for all $x$ in $C$, for let $\langle x_n \rangle$ be a sequence of continuous functions on $[0, 1]$ that are bounded by 1, have $x_n(0) = 1$, and are such that $x_n(t) \to 0$ for all $t \neq 0$. Then, for each $y \in L^1$, $\int x_n y \to 0$, while $F(x_n) = 1$.

If we consider the dual $X^{**}$ of $X^*$, then to each $x$ in $X$ there corresponds an element $\varphi x$ in $X^{**}$ defined by $(\varphi x)(f) = f(x)$. We have $\|\varphi x\| = \sup_{\|f\|=1} f(x)$. Since $f(x) \leq \|f\|\,\|x\|$, we have $\|\varphi x\| \leq \|x\|$, while by Proposition 6 we have an $f$ of norm 1 with $f(x) = \|x\|$. Hence $\|\varphi x\| = \|x\|$. Since $\varphi$ is clearly a linear mapping, $\varphi$ is an isometric isomorphism of $X$ onto some linear subspace $\varphi[X]$ of $X^{**}$. The mapping $\varphi$ is called the natural isomorphism of $X$ into $X^{**}$, and if $\varphi[X] = X^{**}$ we say that $X$ is *reflexive.*

Thus $L^p$ is reflexive if $1 < p < \infty$. Since there are functionals on $L^\infty$ that are not given by integration with respect to a function in $L^1$, it follows that $L^1$ is not reflexive. This and Problem 22 show that $L^\infty$ is not reflexive. It should be observed that $X$ may be isometric with $X^{**}$ without being reflexive.

By Proposition 3, the space $X^{**}$ is complete, and so the closure $\overline{\varphi[X]}$ in $X^{**}$ must be complete by Proposition 7.14. Thus each normed vector space is isometrically isomorphic to a dense subset of a Banach space.

Before closing this section, we add a word about the Hahn–Banach Theorem for complex vector spaces. A complex vector space is a vector space in which we allow multiplication by complex scalars. The following extension of the Hahn–Banach Theorem for complex spaces is due to Bohnenblust and Sobczyk:

**8. Theorem:** *Let $X$ be a complex vector space, $S$ a linear subspace, $p$ a real-valued function on $X$ such that $p(x + y) \leq p(x) + p(y)$, and $p(\alpha x) = |\alpha| p(x)$. Let $f$ be a (complex) linear functional on $S$ such that $|f(s)| \leq p(s)$ for all $s$ in $S$. Then there is a linear functional $F$ defined on $X$ such that $F(s) = f(s)$ for $s$ in $S$ and $|F(x)| \leq p(x)$ for all $x$ in $X$.*

***Proof:*** We first note that $X$ can be considered as a real vector space if we simply ignore the possibility of multiplying by complex constants. A mapping $F$ from $X$ to the complex numbers that is linear in the real sense is linear in the complex sense if and only if $F(ix) = iF(x)$ for each $x$. On $S$ define $g$ and $h$ by taking $g(s)$ to be the real part of $f(s)$ and $h(s)$ the imaginary part. Then $g$ and $h$ are linear in the real sense and $f = g + ih$. Since $f$ is linear in the complex sense, $g(is) + ih(is) = f(is) = if(s) = ig(s) - h(s)$, and so $h(s) = -g(is)$.

Since $g(s) \leq |f(s)| \leq p(s)$, we can extend $g$ to a functional $G$ on $X$ that is linear in the real sense and satisfies $G(x) \leq p(x)$. Let $F(x) = G(x) - iG(ix)$. Then $F(s) = f(s)$ for $s$ in $S$. Since $F(ix) = G(ix) - iG(ix) = i[G(x) - iG(ix)]$, we have $F$ linear in the complex sense. For any $x$, choose $\omega$ to be a complex number such that $|\omega| = 1$ and $\omega F(x) = |F(x)|$. Then

$$|F(x)| = \omega F(x) = F(\omega x) = G(\omega x) \leq p(\omega x) = p(x). \quad \blacksquare$$

## Problems

**17.** Show that a linear functional $f$ on a normed linear space is bounded iff its kernel is closed. [The kernel of $f$ is $\{x : f(x) = 0\}$.]

**18.** Let $T$ be a linear subspace of a normed linear space $X$ and $y$ a given element of $X$. Show that $\inf_{t \in T} \|y - t\| = \sup \{f(y): \|f\| = 1, f(t) = 0$ all $t \in T\}$.

**19.** Prove Proposition 7 by taking $S$ to be the subspace consisting of multiples of $y$, $f(\lambda y) = \lambda\delta$, and $p(x) = \inf_{t \in T} \|x - t\|$.

**20.** Let $l^\infty$ be the space of all bounded sequences. Use Proposition 5 to show that there is a linear functional $F$ on $l^\infty$ with the following properties:

i. $\underline{\lim}\ \xi_n \leq F[\langle \xi_n \rangle] \leq \overline{\lim}\ \xi_n$.
ii. $F[\langle \xi_n + \eta_n \rangle] = F[\langle \xi_n \rangle] + F[\langle \eta_n \rangle]$.
iii. $F[\langle \alpha\xi_n \rangle] = \alpha F[\langle \xi_n \rangle]$.
iv. If $\eta_n = \xi_{n+1}$, then $F[\langle \eta_n \rangle] = F[\langle \xi_n \rangle]$.

The functional $F$ is called a *Banach limit* and is often denoted by Lim.

**21.** Use Proposition 5 to show that there is a set function $\mu$ defined for all bounded sets of **R** with the following properties:

i. If $A \cap B = \emptyset$, $\mu(A \cup B) = \mu A + \mu B$.
ii. $\mu(A + t) = \mu A$.
iii. If $A \subset B$, $\mu A \leq \mu B$.
iv. If $A$ is Lebesgue measurable, then $\mu A$ is the Lebesgue measure of $A$.

[Hint: It is easier to work with integrals than sets.]

**22.** Show that a Banach space $X$ is reflexive iff $X^*$ is reflexive. [Hint: If $\varphi[X]$ is not all of $X^{**}$, then there is a nonzero function $y \in X^{***}$ such that $y(x) = 0$ for all $x \in \varphi[X]$.]

**23.** If $S$ is a linear subspace of a Banach space $X$, we define the *annihilator* $S^0$ of $S$ to be the subset $S^0 = \{y \in X^*: y(s) = 0$ for all $s \in S\}$. If $T$ is a subspace of $X^*$, we define $T^0 = \{x \in X: t(x) = 0$ for all $t \in T\}$.

**a.** Show that $S^0$ is a closed linear subspace of $X^*$.

**b.** Show that $S^{00} = \bar{S}$.

**c.** If $S$ is a closed subspace of $X$, then $S^*$ is isomorphic to $X^*/S^0$.

**d.** If $S$ is a closed subspace of a reflexive Banach space $X$, then $S$ is reflexive.

**24.** Let $X$ be a vector space and $P$ a subset of $X$ such that $x, y \in P$ implies $x + y \in P$ and $\alpha x \in P$ for $\alpha > 0$. Define a partial order in $X$ by defining $x \leq y$ to mean $y - x \in P$. A linear functional $f$ on $X$ is said to be positive (with respect to $P$) if $f(x) \geq 0$ for all $x \in P$. Let $S$ be any subspace of $X$ with the property that for each $x \in X$ there is an $s \in S$ with $x \leq s$. Then each positive linear functional on $S$ can be extended to a positive linear functional on $X$. [Hint: The transfinite part of the proof is the same as that for the Hahn–Banach Theorem. The possibility of extending a functional to a space containing one more element is even simpler than for the Hahn–Banach Theorem.]

**25.** Let $f$ be a mapping of the unit ball $S = \{x: \|x\| \leq 1\}$ into $\mathbf{R}$ such that $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ whenever $x$, $y$ and $\alpha x + \beta y$ are in $S$. Show that $f$ may be extended to all of $X$ so that it is a linear functional.

## 4 The Closed Graph Theorem

A mapping from one topological space to another is called an *open mapping* if the image of each open set is open. Thus a one-to-one continuous open mapping is a homeomorphism. We shall show that a continuous linear transformation of one Banach space *onto* another is always an open mapping and use this to give criteria for the continuity of a linear transformation. We begin with a lemma.

**9. Lemma:** *Let $A$ be a continuous linear transformation of the Banach space $X$ onto the Banach space $Y$. Then the image by $A$ of the unit sphere in $X$ contains a sphere about the origin in $Y$.*

***Proof:*** Let $S_n = \{x: \|x\| < 1/2^n\}$. Since $A$ is onto and

$$X = \bigcup_{k=1}^{\infty} kS_1,$$

we have

$$Y = \bigcup_{k=1}^{\infty} kA(S_1).$$

But $Y$ is a complete metric space, and so is not of first category in itself. Consequently, $A(S_1)$ cannot be nowhere dense, and $\overline{A(S_1)}$ contains some sphere, say

$$\{y: \|y - p\| < \eta\}.$$

Then $\overline{A(S_1)} - p$ contains the sphere

$$\{y: \|y\| < \eta\}.$$

But

$$\overline{A(S_1)} - p \subset \overline{A(S_1)} - \overline{A(S_1)} \subset 2\overline{A(S_1)} = \overline{A(S_0)}.$$

Thus $\overline{A(S_0)}$ contains a sphere about the origin of radius $\eta$ and so by the linearity of $A$, $\overline{A(S_n)}$ contains a sphere about the origin of radius $\eta/2^n$.

We now proceed to show that $A(S_0)$ contains a sphere of radius $\eta/2$ about the origin. Let $y$ be an arbitrary point of $Y$ with $\|y\| < \eta/2$. Since $y \in \overline{A(S_1)}$, we can choose $x_1 \in S_1$ such that

$$\|y - A(x_1)\| < \frac{\eta}{4}.$$

Similarly, we may choose $x_2 \in S_2$ such that

$$\|y - A(x_1) - A(x_2)\| < \frac{\eta}{8},$$

and we continue so that we choose $x_n \in S_n$ with

$$\left\| y - \sum_{k=1}^{n} A(x_k) \right\| < \frac{\eta}{2^{n+1}}.$$

Since $\|x_k\| < 1/2^k$, $\sum_{k=1}^{\infty} x_k$ is absolutely convergent and $x = \sum_{k=1}^{\infty} x_k$ belongs to $S_0$. Moreover,

$$A(x) = A\left(\sum_{k=1}^{\infty} x_k\right) = \sum_{k=1}^{\infty} A(x_k) = y.$$

Thus $y \in A(S_0)$ and so

$$\left\{y: \|y\| < \frac{\eta}{2}\right\} \subset A(S_0). \quad \blacksquare$$

**10. Proposition:** *A continuous linear transformation $A$ of a Banach space $X$ onto a Banach space $Y$ is an open mapping. Thus in particular, if $A$ is one-to-one, it is an isomorphism.*

***Proof:*** Let $O$ be an arbitrary open set of $X$ and $y$ any point of $A[O]$. Then there is some $x \in O$ such that $y = A(x)$. Since $O$ is open there is a sphere $S$ containing $x$ and contained in $O$. But Lemma 9 states that $A[S - x]$ must contain a sphere about the origin or that $A[S]$ must contain a sphere about $y$. Thus $y$ is contained in a sphere that is contained in $A[O]$, and so $A[O]$ is open. ▮

**11. Proposition:** *Let $X$ be a linear vector space that is complete in each of the norms $\| \ \|$ and $||| \ |||$, and suppose there is a constant $C$ such that*

$$\|x\| \leq C|||x|||$$

*for all $x \in X$. Then the norms are equivalent. That is, there is a second constant $C'$ such that*

$$|||x||| \leq C'\|x\|$$

*for all $x \in X$.*

***Proof:*** The identity map of $(X, ||| \ |||)$ onto $(X, \| \ \|)$ is a one-to-one continuous linear transformation and so must be an isomorphism by Proposition 10. Therefore, the inverse mapping must be bounded. ▮

**12. Closed Graph Theorem:** *Let $A$ be a linear transformation on a Banach space $X$ to a Banach space $Y$. Suppose that $A$ has the property that, whenever $\langle x_n \rangle$ is a sequence in $X$ that converges to some point $x$ and $\langle Ax_n \rangle$ converges in $Y$ to a point $y$, then $y = Ax$. Then $A$ is continuous.*

***Proof:*** Define a new norm in $X$ by

$$|||x||| = \|x\| + \|Ax\|.$$

Then $X$ is complete in the norm $||| \ |||$. For if $|||x_p - x_q||| \to 0$, then $\|x_p - x_q\| \to 0$ and $\|Ax_p - Ax_q\| \to 0$. Hence by the completeness of $X$ and $Y$ we have $x \in X$ and $y \in Y$ such that $\|x_p - x\| \to 0$ and $\|Ax_p - y\| \to 0$. By the hypothesis of our theorem $y = Ax$. Hence $|||x_p - x||| \to 0$, and $X$ is complete with respect to $||| \ |||$. But now Proposition 11 applies, and there is a $C'$ such that

$$\|x\| + \|Ax\| \leq C'\|x\|.$$

Thus

$$\|Ax\| \leq C'\|x\|,$$

and $A$ is bounded. ▮

The graph of a mapping of $X$ into $Y$ is just the set of all $\langle x, Ax \rangle$ in $X \times Y$. The hypothesis of Theorem 12 merely states that the graph of $A$ is closed.

Another consequence of the theory of category is the following proposition, which is known as the principle of *uniform boundedness.*

**13. Proposition:** *Let $X$ be a Banach space and $\mathfrak{F}$ a family of bounded linear operators from $X$ to a normed space $Y$. Suppose that for each $x$ in $X$ there is a constant $M_x$ such that $\|Tx\| \le M_x$ for all $T$ in $\mathfrak{F}$. Then the operators in $\mathfrak{F}$ are uniformly bounded; that is, there is a constant $M$ such that $\|T\| \le M$ for all $T$ in $\mathfrak{F}$.*

***Proof:*** For each $T$ the function $f$ defined by $f(x) = \|Tx\|$ is a real-valued continuous function on $X$. Since the family of these functions is bounded at each $x$ in $X$ and $X$ is complete, there is by Theorem 7.32 an open set $O$ in $X$ on which these functions are uniformly bounded. Thus there is a constant $M'$ such that $\|Tx\| \le M'$ for all $x \in O$. Let $y$ be a point in $O$. Since $O$ is open, there is a sphere $S = \{x\colon \|x - y\| < \delta\}$ of some radius $\delta$ centered at $y$ and contained in $O$. If $\|z\| \le \delta$, then $Tz = T(y + z) - Ty$ with $y + z$ in $S \subset O$. Hence $\|Tz\| \le \|T(y + z)\| + \|Ty\| \le M' + M_y$. Consequently, $\|T\| \le (M' + M_y)/\delta$ for all $T$ in $\mathfrak{F}$. ▮

## Problems

**26.** Let $\langle T_n \rangle$ be a sequence of continuous linear operators on a Banach space $X$ to a normed vector space $Y$, and suppose that for each $x$ in $X$ the sequence $\langle T_n x \rangle$ converges to a value $Tx$. Then $T$ is a bounded linear operator.

**27.** Let $A$ be a bounded linear transformation from a Banach space $X$ to a Banach space $Y$, and let $M$ be the kernel and $S$ the range of $A$. Then $S$ is isomorphic to $X/M$ if and only if $S$ is closed.

**28.** Let $S$ be a linear subspace of $C[0, 1]$ that is closed as a subspace of $L^2[0, 1]$.

**a.** Show that $S$ is a closed subspace of $C[0, 1]$.

**b.** Show that there is a constant $M$ such that for all $f \in S$ we have $\|f\|_2 \le \|f\|_\infty$ and $\|f\|_\infty \le M\|f\|_2$.

**c.** Show that for each $y \in [0, 1]$ there is a function $k_y$ in $L^2$ such that for each $f \in S$ we have $f(y) = \int k_y(x) f(x)\, dx$.

The space $S$ can be shown to be finite dimensional (see Problem 41 or 55).

**29.** **a.** Give an example of a discontinuous operator $A$ from a normed linear space $X$ to a Banach space $Y$ such that $A$ has a closed graph. [Hint: Let $f$ be a discontinuous linear functional on a Banach space $X$. Let $X = \{x \in Y \oplus \mathbf{R}: x = \langle y, f(y)\rangle\}$.]

**b.** Give an example of a discontinuous operator $A$ from a Banach space $X$ to a normed linear space $Y$ such that $A$ has a closed graph.

## 5 Topological Vector Spaces

Just as the notion of a metric space generalizes to that of a topological space, so the notion of a normed linear space generalizes to that of a topological vector space: A linear vector space $X$ with a topology $\mathfrak{J}$ on it is called a *topological vector space* if addition is a continuous function from $X \times X$ into $X$ and multiplication by scalars is a continuous function from $\mathbf{R} \times X$ into $X$. It follows from the continuity of addition that translation by an element $x$ is a homeomorphism and that the translate $x + O$ of an open set $O$ is open. Any topology with this property on a vector space is said to be translation invariant. If $\mathfrak{J}$ is a translation invariant topology on $X$ and $\mathcal{B}$ a base for $\mathfrak{J}$ at $\theta$, then the sets of the form $x + U$, $U \in \mathcal{B}$, form a base for $\mathfrak{J}$ at $x$. Thus it suffices to give a base at $\theta$ in order to determine a translation invariant topology. A base at $\theta$ is often called a *local base*. The following proposition gives conditions on a set $\mathcal{B}$ which guarantee that it will be a base for a topology for a topological vector space and states that we can always find a base satisfying these conditions.

**14. Proposition:** *Let $X$ be a topological vector space. Then we can find a base $\mathcal{B}$ at $\theta$ that satisfies the following:*

i. *If $U, V \in \mathcal{B}$, then there is a $W \in \mathcal{B}$ with $W \subset U \cap V$.*
ii. *If $U \in \mathcal{B}$ and $x \in U$, there is a $V \in \mathcal{B}$ such that $x + V \subset U$.*
iii. *If $U \in \mathcal{B}$, there is a $V \in \mathcal{B}$ such that $V + V \subset U$.*
iv. *If $U \in \mathcal{B}$ and $x \in X$, there is an $\alpha \in \mathbf{R}$ such that $x \in \alpha U$.*
v. *If $U \in \mathcal{B}$ and $0 < |\alpha| \leq 1$, then $\alpha U \subset U$ and $\alpha U \in \mathcal{B}$.*

*Conversely, given a collection $\mathcal{B}$ of subsets containing $\theta$ and satisfying the above conditions, there is a topology for $X$ making $X$ a topological vector space and having $\mathcal{B}$ as a base at $\theta$. The topology will be Hausdorff if and only if*

vi. $\bigcap \{U \in \mathcal{B}\} = \{\theta\}$.

The proof of the proposition is left to the reader. We note that, if $X$ is a normed linear space, we may take $\mathcal{B}$ to be the set of spheres about $\theta$, and the proposition gives us a base for the general case which has many of the properties possessed by the collection of spheres.

In a topological vector space we can compare the neighborhoods at one point with those at another by translation. Thus it is possible to speak of uniform properties: A mapping $f$ from a topological vector space $X$ into a topological vector space $Y$ is said to be uniformly continuous if for any open set $O$ containing the origin in $Y$ there is an open set $U$ containing the origin in $X$ such that for each $x \in X$ we have $f[x + U] \subset f(x) + O$. It is readily seen that a linear transformation from $X$ to $Y$ is uniformly continuous if it is continuous at one point. A one-to-one linear mapping $\varphi$ of $X$ onto $Y$ is called a (topological) *isomorphism* if $\varphi$ and $\varphi^{-1}$ are both continuous. From an abstract point of view isomorphic spaces are the same. The following theorem tells us that the only topology on a finite-dimensional vector space that makes it into a topological vector space is the usual one.

**15. Proposition (Tychonoff):** *Let $X$ be a finite-dimensional Hausdorff topological vector space. Then $X$ is topologically isomorphic to $\mathbf{R}^n$ for some $n$.*

Some suggestions for the proof of this proposition are given in Problem 33. Useful corollaries are given in Problems 34, 35, and 36.

## Problems

**30.** Prove Proposition 14:

**a.** A collection $\mathcal{B}$ of subsets containing $\theta$ is a base at $\theta$ for a translation invariant topology if and only if (i) and (ii) hold.

**b.** Addition is continuous from $X \times X$ to $X$ if and only if (iii).

**c.** If multiplication by scalars is continuous (at $\langle 0, \theta\rangle$) from $\mathbf{R} \times X$ to $X$, then (iv).

**d.** If $X$ is a topological vector space, then the family $\mathcal{B}$ of all open sets $U$ that contain $\theta$ and such that $\alpha U \subset U$ for all $\alpha$ with $|\alpha| < 1$ is a local base for the topology and satisfies (v). [If $O$ is any open set containing $\theta$, the continuity of multiplication implies that there is an open set $V$ containing $\theta$ and an $\epsilon > 0$ such that $\lambda V \subset O$ for all $|\lambda| < \epsilon$. Then $U = \bigcup_{|\lambda| < \epsilon} \lambda V$ is open, $\theta \in U \subset O$, and $\alpha U \subset U$ for each $\alpha$ with $|\alpha| < 1$.]

**e.** If $\mathcal{B}$ satisfies the conditions of the proposition, then it generates a topology in which multiplication by scalars is continuous from $\mathbf{R} \times X$ to $X$. [Show that (iv) and (v) imply continuity at $\langle 0, x\rangle$ and $\langle \alpha, \theta\rangle$, and use (iii).]

**f.** If $X$ is $T_1$, then (vi) holds. If (vi) and (iii) hold, then $X$ is Hausdorff.

**13. a.** Show that a linear transformation from one topological vector space to another is uniformly continuous if it is continuous at one point.

**b.** Show that a linear functional $f$ on $X$ is continuous iff there is an open set $O$ such that $f[O] \neq \mathbf{R}$. [Hint: You can take $O$ to satisfy property (v) of Proposition 14.]

**32.** Let $X$ be a topological vector space and $M$ a closed linear subspace. Let $\varphi$ be the natural homomorphism of $X$ onto $X/M$, and define a topology on $X/M$ by taking $O$ to be open if and only if $\varphi^{-1}[O]$ is open in $X$. Then this topology makes $X/M$ a topological vector space and $\varphi$ a continuous open map. When we speak of $X/M$ as a topological vector space, we always mean $X/M$ with this topology.

**33.** Prove Proposition 15:

**a.** If $X$ is $n$-dimensional, there is a continuous one-to-one linear map $\varphi$ of $\mathbf{R}^n$ onto $X$.

**b.** Let $S$ and $B$ be the subsets of $\mathbf{R}^n$ defined by $S = \{y: \|y\| = 1\}$, $B = \{y: \|y\| < 1\}$. Then $\varphi[S]$ is closed and $X \sim \varphi[S]$ open.

**c.** There is an open set $U$ of $X$ containing $\theta$ such that $\alpha U \subset U$ for each $|\alpha| < 1$ and $U \subset X \sim \varphi[S]$.

**d.** The set $U$ in part (c) is contained in $\varphi[B]$, and so $\varphi^{-1}$ is continuous.

**34.** Show that a finite-dimensional subspace $M$ of a Hausdorff topological vector space $X$ is closed. [Hint: If $x \notin M$, let $N$ be the finite-dimensional subspace spanned by $x$ and $M$. Then $N$ has the usual topology, and so $x$ is not a point of closure of $M$.]

**35.** Let $A$ be a linear mapping of a finite-dimensional Hausdorff vector space $X$ into a topological vector space $Y$. Then $A$ is continuous. [Hint: The range of $A$ is finite-dimensional and hence has the usual topology.]

**36.** A linear mapping $A$ from a topological vector space $X$ to a finite-dimensional topological space $Y$ is continuous iff its kernel $M$ is closed. [Hint: $A = B \circ \varphi$, where $\varphi$ is the natural mapping of $X$ on $X/M$, and $B$ is continuous by Problem 35.]

**37.** Prove that every locally compact Hausdorff vector space $X$ is finite-dimensional. [Hint: Let $V$ be a neighborhood of $\theta$ with $\bar{V}$ compact and

$\alpha V \subset V$ for each $\alpha$, $|\alpha| < 1$. If we cover $\bar{V}$ by a finite number of translates $x_1 + \frac{1}{3}V, \ldots, x_n + \frac{1}{3}V$, then $x_1, \ldots, x_n$ form a basis for $X$.]

## 6 Weak Topologies

If $X$ is any vector space and $\mathcal{F}$ a collection of linear functionals on $X$, we define the *weak topology* generated by $\mathcal{F}$ to be the weakest topology such that each $f$ in $\mathcal{F}$ is continuous (cf. Problem 8.25). This topology is easily seen to be translation invariant, and a base at $\theta$ for this topology is given by the sets $\{x\colon |f_i(x)| < \epsilon,\ i = 1, \ldots, n\}$, where $\epsilon > 0$ and $\{f_1, \ldots, f_n\}$ is a finite subset of $\mathcal{F}$. Since the family of all such sets satisfy the conditions of Proposition 14, this topology makes $X$ a topological vector space. A sequence (or net) $\langle x_n \rangle$ converges to $x$ in this topology if and only if $f(x_n) \to f(x)$ for each $f \in \mathcal{F}$.

If $X$ is a normed vector space and the functionals in $\mathcal{F}$ are all continuous (i.e., if $\mathcal{F} \subset X^*$), then the weak topology generated by $\mathcal{F}$ is weaker (has fewer open sets) than the norm topology of $X$. We usually call the metric topology generated by the norm the strong topology of $X$ and the weak topology on $X$ generated by $X^*$ the *weak topology* of $X$. Thus we speak of strongly closed and strongly open sets when referring to the strong topology and weakly open and weakly closed sets for the weak topology. Every weakly closed set is strongly closed but not conversely. Every strongly convergent sequence (or net) is weakly convergent. While not every strongly closed set is weakly closed, we do have the following proposition, a generalization of which is given by Corollary 23.

**16. Proposition:** *A linear manifold $M$ is weakly closed if and only if it is strongly closed.*

***Proof:*** Since every weakly closed set is strongly closed, we have only to show that, if $M$ is strongly closed, it is also weakly closed. Suppose that $M$ is strongly closed and that $x$ is a point not on $M$. We must show that $x$ is not a point of closure of $M$ in the weak topology. Since $x$ is not a point of closure of $M$ in the strong (metric) topology, we have

$$\inf_{s \in M} \|x - s\| \geq \delta > 0.$$

Hence by Proposition 7 there is a continuous linear functional $f$ that vanishes on $M$ and does not vanish at $x$. Now $\{y\colon f(y) \neq 0\}$ is an open set in the weak topology that contains $x$ but does not meet $M$. Hence $x$ is not a weak point of closure of $M$. ▌

If we apply the notion of weak topology to the dual $X^*$ of a normed space $X$, we see that the weak topology of $X^*$ is the weakest topology for $X^*$ such that all of the functionals in $X^{**}$ are continuous. The weak topology for $X^*$ turns out to be less useful than the weak topology for $X^*$ generated by $X$ (or more precisely by $\varphi[X]$, where $\varphi$ is the natural embedding of $X$ into $X^{**}$). This topology is called the *weak* topology* for $X^*$ and is even weaker than the weak topology. Thus a weak* closed subset of $X^*$ is weakly closed, and weak convergence implies weak* convergence. A base at $\theta$ for the weak* topology is given by sets of the form $\{f: |f(x_i)| < \epsilon, i = 1, \ldots, n\}$, where $\{x_1, \ldots, x_n\}$ is a finite subset of $X$. If $X$ is reflexive, then the weak and weak* topologies of $X^*$ coincide. Some of the importance of the weak* topology stems from the following theorem:

**17. Theorem (Alaoglu):** *The unit ball* $S^* = \{f: \|f\| \le 1\}$ *of* $X^*$ *is compact in the weak* topology.*

***Proof:*** If $f \in S^*$, then $|f(x)| \le \|x\|$, and so $f(x) \in [-\|x\|, \|x\|]$. Let $I_x = [-\|x\|, \|x\|]$. Then each $f \in S^*$ corresponds to a point in

$$P = \bigtimes_{x \in X} I_x,$$

since the latter is by definition the set of all functions $f$ on $X$ such that $f(x) \in I_x$. Thus we may think of $S^*$ as a subset of $P$, and the definition of the topology for $P$ shows that the topology $S^*$ inherits as a subspace of $P$ is the weak* topology of $S^*$. Since $P$ is compact by the Tychonoff Theorem, $S^*$ will be compact if it is a closed subset of $P$. Let $f$ be a point of closure of $S^*$ in $P$. Then $f$ is a mapping of $X$ into $\mathbf{R}$. Since $|g(x)| \le \|x\|$ for $g \in S^*$, and evaluation at $x$ is a continuous function on $P$, we have $|f(x)| \le \|x\|$. Let $x$, $y$, and $z$ be three points of $X$ such that $z = \alpha x + \beta y$. For each $\epsilon > 0$ the set

$$N = \{g \in P: |g(x) - f(x)| < \epsilon, |g(y) - f(y)| < \epsilon, |g(z) - f(z)| < \epsilon\}$$

is an open subset of $P$ containing $f$. Since $f$ is a point of closure of $S^*$, we can find a $g$ in $S^* \cap N$. Since this $g$ is linear (being in $S^*$), we have $g(z) = \alpha g(x) + \beta g(y)$. Therefore, it follows that you have $|f(z) - \alpha f(x) - \beta f(y)| < \epsilon(1 + |\alpha| + |\beta|)$. This inequality holding for each positive $\epsilon$, we have $f(z) = \alpha f(x) + \beta f(y)$, and $f$ is linear on $X$. Thus $f$ is in $S^*$, and so $S^*$ is closed. ▌

## Problems

**38. a.** Show that if $x_n \to x$ weakly, then $\langle \|x_n\| \rangle$ is bounded.

**b.** Let $\langle x_n \rangle$ be a sequence in $l^p$, $1 < p < \infty$, and let $x_n = \langle \xi_{m,n} \rangle_{m=1}^{\infty}$. Show that $\langle x_n \rangle$ converges weakly to $x = \langle \xi_m \rangle$ if and only if $\langle \|x_n\| \rangle$ is bounded and for each $m$ we have $\xi_{m,n} \to \xi_m$.

**c.** Let $\langle x_n \rangle$ be a sequence in $L^p[0, 1]$, $1 < p < \infty$. Show that $\langle x_n \rangle$ converges weakly to $x$ if $\langle \|x_n\| \rangle$ is bounded and $\langle x_n \rangle$ converges to $x$ in measure (cf. Problem 6.17).

**d.** Show that part (c) is false if $p = 1$.

**e.** In $l^p$, $1 < p < \infty$, let $x_n$ be that sequence whose $n$-th term is one and whose remaining terms are zero. Then $\langle x_n \rangle$ does not converge in the strong topology, but $x_n \to 0$ in the weak topology.

**f.** Let $\langle x_n \rangle$ be as in (e), and define $y_{n,m} = x_n + nx_m$. Then the set $F = \{y_{n,m}: m > n\}$ is strongly closed. [Hint: The distance between any two points of $F$ is at least 1. Hence $F$ contains no nonconstant sequences that converge in the strong topology.]

**g.** The set $F$ in (f) has $\theta$ as a weak closure point. However, there is no sequence $\langle z_n \rangle$ from $F$ that converges weakly to zero.

**h.** Describe the weak and weak* topologies of $l^1$ considered as the dual of $c_0$.

**39. a.** Let $S$ be a bounded subset of a normed space $X$. Let $\mathcal{F}$ be a set of functionals in $X^*$ and let $\mathcal{F}_0$ be a dense subset of $\mathcal{F}$ (dense in the sense of the norm topology on $X^*$). Then $\mathcal{F}$ and $\mathcal{F}_0$ may generate different weak topologies for $X$, but these topologies are the same on $S$; that is, $S$ inherits the same topology from each.

**b.** Let $S^*$ be the unit sphere in the dual $X^*$ of a separable Banach space $X$. Then the weak* topology on $S^*$ is metrizable. (*Caution:* This does not mean the weak* topology on $X^*$ is metrizable!)

**40.** Show that every weakly compact set is bounded in the norm topology.

**41.** Let $S$ be the linear subspace of $C[0, 1]$ given in Problem 28.

**a.** Show that if $f_n \to f$ weakly in $L^2$, then $f_n(y) \to f(y)$ for each $y \in [0, 1]$.

**b.** If $f_n \to f$ weakly in $L^2$, then $\|f_n\|_\infty$ is bounded, and hence $f_n \to f$ strongly in $L^2$ by the Lebesgue convergence theorem.

**c.** The space $S$ is a locally compact subspace of $L^2$ and hence finite-dimensional.

## 7 Convexity

A subset $K$ of a vector space $X$ is said to be **convex** if, whenever it contains $x$ and $y$, it also contains $\lambda x + (1 - \lambda)y$ for $0 \le \lambda \le 1$. The set $\{z: z = \lambda x + (1 - \lambda)y \text{ for } 0 \le \lambda \le 1\}$ is called the line segment joining $x$ and $y$. The points $x$ and $y$ are its endpoints, and a $z$ for which $0 < \lambda < 1$ is called an interior point of the segment. Thus a set $K$ is convex iff whenever it contains $x$ and $y$ it contains the line segment joining $x$ and $y$. Every linear manifold is convex, and the unit ball in a normed space is convex. The following lemma gives some basic properties of convex sets. Further properties are given in Problems 42, 43, and 44. The proof of the lemma is straightforward and is omitted.

**18. Lemma:** *If $K_1$ and $K_2$ are convex sets, so also are the sets $K_1 \cap K_2$, $\lambda K_1$, and $K_1 + K_2$.*

A point $x_0$ is said to be an **internal point** of a set $K$ if the intersection with $K$ of each line through $x_0$ contains an open interval about $x_0$. Thus $x_0$ is internal to $K$ if, given $x \in X$, there is an $\epsilon > 0$ such that $x_0 + \lambda x \in K$ for all $\lambda$ with $|\lambda| < \epsilon$. Let $K$ be a convex set that contains $\theta$ as an internal point. Then we define the support function $p$ of $K$ (with respect to $\theta$) by $p(x) = \inf \{\lambda: \lambda^{-1}x \in K, \lambda > 0\}$. We have the following properties for this support function:

**19. Lemma:** *If $K$ is a convex set containing $\theta$ as an internal point, then the support function $p$ has the following properties:*

i. $p(\lambda x) = \lambda p(x)$ *for* $\lambda \ge 0$.
ii. $p(x + y) \le p(x) + p(y)$.
iii. $\{x: p(x) < 1\} \subset K \subset \{x: p(x) \le 1\}$.

***Proof:*** The first and third properties follow immediately from the definition of $p$. To prove the second, suppose that $\lambda^{-1}x$ and $\mu^{-1}y$ belong to $K$. Then

$$(\lambda + \mu)^{-1}(x + y) = \lambda(\lambda + \mu)^{-1}(\lambda^{-1}x) + \mu(\lambda + \mu)^{-1}(\mu^{-1}y)$$

belongs to $K$, since $K$ is convex. Thus $p(x + y) \le \lambda + \mu$, and taking infima over all admissible $\lambda$ and $\mu$ we obtain $p(x + y) \le p(x) + p(y)$. ∎

Two convex sets $K_1$ and $K_2$ are said to be **separated** by a linear functional $f$ if there is a real number $\alpha$ such that $f(x) \le \alpha$ on $K_1$ and $f(x) \ge \alpha$ on $K_2$.

**20. Theorem:** *Let $K_1$ and $K_2$ be two disjoint convex sets in a vector space $X$, and suppose that one of them has an internal point. Then there is a nonzero linear functional $f$ that separates $K_1$ and $K_2$.*

***Proof:*** Let $x_1$ be an internal point of $K_1$. Then $K_1 - K_2$ is convex, and the point $x_0 = x_1 - x_2$ is an internal point of $K_1 - K_2$ for any $x_2$ in $K_2$. Let $K = K_1 - K_2 - x_0$. Then $K$ is a convex set containing $\theta$ as an internal point. Since $K_1$ and $K_2$ are disjoint, $\theta \notin K_1 - K_2$, and so $-x_0 \notin K$.

Let $p$ be the support function of $K$ (with respect to $\theta$). Then $p(-x_0) \ge 1$. Let $S$ be the one-dimensional subspace of $X$ consisting of all multiples of $x_0$. On $S$ define $f$ by $f(\alpha x_0) = -\alpha$. Then $f(s) \le p(s)$ and $p$ satisfies the condition of the Hahn–Banach Theorem by Lemma 19. Thus we may extend $f$ to a linear functional defined on all of $X$ so that $f(x) \le p(x)$ for all $x$. Thus if $x \in K$, we have $f(x) \le 1$.

Let $x \in K_1$ and $y \in K_2$. Then $x - y - x_0 \in K$, and we have

$$f(x) - f(y) - f(x_0) = f(x - y - x_0) \le 1.$$

Since $f(x_0) = -1$, we have $f(x) \le f(y)$. This being true for each $x \in K_1$ and each $y$ in $K_2$, we have $\sup_{x \in K_1} f(x) \le \inf_{y \in K_2} f(y)$. Thus $f$ separates $K_1$ and $K_2$, and $f$ is a nonzero functional, since $f(x_0) = -1$. ∎

A topological vector space is called **locally convex** if we can find a base for the topology consisting of convex sets. The following proposition gives a convenient criterion for ensuring that a given topology $\mathfrak{J}$ for a vector space $X$ will make $X$ into a locally convex topological vector space. Suggestions for its proof are given in Problem 46.

**21. Proposition:** *Let $\mathfrak{N}$ be a family of convex sets in a vector space $X$. Then the following conditions are sufficient for the translates of sets in $\mathfrak{N}$ to form a base for a topology that makes $X$ into a locally convex topological vector space:*

i. *If $N \in \mathfrak{N}$, each point of $N$ is internal.*
ii. *For $N_1$ and $N_2$ in $\mathfrak{N}$, there is an $N_3$ in $\mathfrak{N}$ with $N_3 \subset N_1 \cap N_2$.*
iii. *If $N$ is in $\mathfrak{N}$, then for each $\alpha$ with $0 < |\alpha| < 1$ we have $\alpha N \in \mathfrak{N}$.*

*Moreover, in each locally convex topological vector space there is a base $\mathfrak{N}$ at $\theta$ satisfying these conditions.*

It follows from this proposition that the weak topology on a vector space $X$ that is generated by a family of linear functionals makes $X$ into a locally convex topological vector space. Also a normed vector space is a locally convex topological vector space.

**22. Proposition:** *Let $X$ be a locally convex topological vector space and $F$ a closed convex subset. Let $x_0$ be a point of $X$ not in $F$. Then there is a continuous linear functional $f$ on $X$ such that*

$$f(x_0) < \inf_{x \in F} f(x).$$

***Proof:*** By translating by $-x_0$, we reduce the proposition to the case that $x_0 = \theta$. Since $\theta$ is not a point of closure of $F$, there is a convex open set $N$ that contains $\theta$ but does not meet $F$. Let $O = N \cap (-N)$. Then $O$ is an open convex set containing $\theta$ and disjoint from $F$, and $-O = O$. Since $\theta$ is an internal point of $O$ (see Problem 44a), there is by Theorem 20 a nonzero linear functional $f$ such that $\sup_{x \in O} f(x) \leq \inf_{y \in F} f(y) = \alpha$. Thus $f(x) \leq \alpha$ for $x$ in $O$, and since $x \in O$ implies $-x \in O$, we have $-f(x) \leq \alpha$ on $O$, whence $|f(x)| \leq \alpha$ on $O$. Thus for each $\epsilon > 0$, we have $|f(x)| < \epsilon$ on the set $O' = (\epsilon\alpha^{-1})O$. But $O'$ is an open set containing $\theta$, and so $f$ is continuous at $\theta$. Since $f$ is linear and continuous at $\theta$, it is continuous everywhere.

It remains only to show that $\alpha > 0$. Since $f$ is a nonzero functional, there is an $x$ such that $f(x) > 0$. Since $\theta$ is an internal point of $O$, we can choose $\lambda > 0$ so that $\lambda x$ is in $O$. Then

$$0 < \lambda f(x) = f(\lambda x) \leq \alpha. \quad \blacksquare$$

**23. Corollary:** *Let $K$ be a convex set in a locally convex topological space. Then $K$ is strongly closed if and only if it is weakly closed.*

**24. Corollary:** *Let $x$ and $y$ be distinct points of a locally convex Hausdorff vector space $X$. Then there is a continuous linear functional $f$ such that $f(x) \neq f(y)$.*

Let $K$ be a convex subset of a vector space $X$. A point $x$ in $K$ is called an **extreme** point if it is not an interior point of any line segment lying in $K$. Thus $x$ is extreme iff whenever $x = \lambda y + (1 - \lambda)z$ with $0 < \lambda < 1$, we have $y \notin K$ or $z \notin K$. We are going to show that every compact convex set has extreme points, but we first consider some preliminary notions. A subset $S$ of a convex set $K$ is called a supporting set of $K$ if it is closed and convex and has the property that, if an interior point of a line segment in $K$ belongs to $S$, then the entire line segment belongs to $S$. Thus an extreme point is a supporting set consisting of exactly one point.

**25. Lemma:** *Let $f$ be a continuous linear functional on a closed convex set $K$. Then the set $S$ of points where $f$ assumes its maximum on $K$ is a supporting set of $K$.*

***Proof:*** The set $S$ is convex, for if $f(x) = m$ and $f(y) = m$, then $f(\lambda x + (1 - \lambda)y) = m$. If the line segment joining $x$ to $y$ is in $K$ and $f$ assumes its maximum $m$ at the point $z = \lambda x + (1 - \lambda)y$, then $m = \lambda f(x) + (1 - \lambda)f(y)$, and since $f(x)$ and $f(y)$ are not greater than $m$, we must have $f(x) = f(y) = m$. But this implies that $f(\mu x + (1 - \mu)y) = m$, and so the entire line segment joining $x$ to $y$ is in $S$. ∎

The intersection of all convex sets containing a set $E$ is a convex set which contains $E$ and which is contained in every convex set containing $E$. This set is called the **convex hull** of $E$. The intersection of all closed convex sets containing $E$ is a closed convex set which contains $E$ and which is contained in every closed convex set containing $E$. This set is called the **closed convex hull** of $E$.

**26. Theorem (Krein–Milman):** *Let $K$ be a compact convex set in a locally convex topological vector space $X$. Then $K$ is the closed convex hull of its extreme points.*

***Proof (Kelley):*** We assume that $K$ is not empty. It follows from the definition of supporting sets that the intersection of a collection of supporting sets of $K$ is a supporting set of $K$ and that, if $S$ is a supporting set of $K$ and $T$ a supporting set of $S$, then $T$ is a supporting set of $K$.

Given any nonempty supporting set $S$ of $K$, the family of all nonempty supporting sets of $K$ is partially ordered by inclusion, and by the Hausdorff maximal principle there is a maximal linearly ordered family $\mathcal{S}$ of nonempty supporting sets with $S$ in $\mathcal{S}$. Since $K$ is compact, the intersection $T$ of all members of $\mathcal{S}$ is nonempty and hence is itself a nonempty supporting set of $K$. It is, moreover, a minimal nonempty supporting set, for, if $T$ properly contained a supporting set, the family $\mathcal{S}$ would not be maximal. Thus any supporting set contains a minimal nonempty supporting set. Now a minimal nonempty supporting set can contain only one point. For if a supporting set $S$ contains the distinct points $x$ and $y$, there is a continuous linear functional $f$ such that $f(x) > f(y)$. Then the subset of $S$ where $f$ attains its maximum is by Lemma 25 a supporting subset of $S$ and hence of $K$. Since $K$ is compact, it is a nonempty supporting subset of $K$ that does not contain $y$.

If a supporting set consists of just one point, that point must be extreme. Hence we have shown that every nonempty supporting set contains an extreme point.

Since the subset of $K$ where a linear functional assumes its maximum is a nonempty supporting set, we conclude that the maximum of a continuous linear functional on $K$ is equal to its maximum on the set $E$ of extreme points of $K$. Let $C$ be the closed convex hull of the extreme points of $K$, and suppose $x \notin C$. By Proposition 22 there is a continuous linear functional $f$ such that $f(x) > \max_{y \in C} f(y) = \max_{y \in K} f(y)$. Thus $x \notin K$, and we have $K \subset C$. Hence $K = C$. ▮

## Problems

**42.** Let $A$ be a linear operator from the vector space $X$ to the vector space $Y$. Then the image of each convex set (or linear manifold) in $X$ is a convex set (or linear manifold) in $Y$ and the inverse image of a convex set (or linear manifold) in $Y$ is a convex set (or linear manifold) in $X$. Give an example to show that a nonconvex set may have a convex image.

**43.** Show that the closure of a convex set $K$ in a topological vector space is convex.

**44. a.** Show that each interior point of a subset of a topological vector space is internal. [Hint: Use the continuity of multiplication.]

**b.** Show that in $\mathbf{R}^n$ each internal point of a convex set is an interior point.

**c.** Give an example of a set in the plane that has an internal point that is not an interior point.

**d.** If a convex set has an interior point, then each internal point is interior.

**e.** Let $X$ be a topological vector space that is of second Baire category with respect to itself. Then, if a closed convex subset of $X$ has an internal point, it has an interior point.

**45.** Let $K$ be a convex set containing $\theta$, and suppose that $x$ is an internal point of $K$. Then for some $\lambda > 0$ the set $x + \lambda K$ is contained in $K$. [Hint: Choose $\lambda > 0$ so that $(1 - \lambda)^{-1}x$ is in $K$.]

**46.** Prove Proposition 21. [Hint: If $N$ is convex, then $\frac{1}{2}N + \frac{1}{2}N \subset N$. Use Proposition 14 and its proof.]

**47.** *Strongest locally convex topology.* Let $X$ be a vector space and $\mathcal{B}$ the collection of all convex sets $V$ containing $\theta$ such that for each $x \in X$ there is an $\alpha > 0$ with $\alpha x \in V$. Then $\mathcal{B}$ is a local base for a locally convex topology on $X$, and this topology is stronger than any other locally convex topology on $X$.

**48. a.** In $L^p[0, 1]$, $1 < p < \infty$, every $x$ with $\|x\| = 1$ is an extreme point of the unit sphere $S = \{x\colon \|x\| \le 1\}$.

**b.** In $L^\infty[0, 1]$ the extreme points of the unit sphere are those $x$ such that $|x(t)| = 1$ a.e.

**c.** The unit sphere in $L^1[0, 1]$ has no extreme points.

**d.** $L^1[0, 1]$ is not the dual of any normed linear space.

**e.** What are the extreme points of the unit sphere in $l^p$?

**f.** What are the extreme points of the unit sphere in $C(X)$, $X$ a compact Hausdorff space? Show that $C[0, 1]$ is not the dual of any normed linear space.

**49.** Let $X$ be the vector space of all measurable real-valued functions on $[0, 1]$ with addition and multiplication by scalars defined in the usual way. Define

$$\sigma(x) = \int_0^1 \frac{|x(t)|}{1 + |x(t)|}\, dt.$$

**a.** We have $\sigma(x + y) \le \sigma(x) + \sigma(y)$. Hence if we define $\rho(x, y) = \sigma(x - y)$, $\rho$ is a metric for $X$.

**b.** In this metric $x_n \to x$ if and only if $x_n \to x$ in measure.

**c.** $X$ is a complete metric space (cf. Problem 4.25).

**d.** Addition is a continuous mapping of $X \times X$ into $X$.

**e.** Multiplication is a continuous mapping of $\mathbf{R} \times X$ into $X$. [Since $X$ is a metric space, it suffices to prove that if $x_n \to x$ and $\lambda_n \to \lambda$, then $\lambda_n x_n \to \lambda x$. This follows from the Bounded Convergence Theorem for convergence in measure.]

**f.** Show that the set of step functions is dense in $X$.

**g.** There is no nonzero continuous linear functional on $X$. [Show that there is an $n$ such that $f(x) = 0$ whenever $x$ is the characteristic function of an interval of length less than $1/n$. Hence $f(x) = 0$ for all step functions $x$.]

**h.** The space $X$ is a topological vector space that is not locally convex.

**i.** Let $s$ be the space of all sequences of real numbers, and define

$$\sigma(\langle\xi_\nu\rangle) = \sum \frac{2^{-\nu}|\xi_\nu|}{1 + |\xi_\nu|}.$$

Prove the analogues of (a), (c), (d), and (e). What is the most general continuous linear functional on $s$?

## 8 Hilbert Space

By a Hilbert space we mean a Banach space $H$ in which there is defined a function $(x, y)$ on $H \times H$ to $\mathbf{R}$ with the following properties.[1]

i. $(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1(x_1, y) + \alpha_2(x_2, y)$.
ii. $(x, y) = (y, x)$.
iii. $(x, x) = \|x\|^2$.

We call $(x, y)$ the inner product of $x$ and $y$. Two examples are immediate: One is the space $\mathbf{R}^n$ with

$$(x, y) = \sum_{i=1}^{n} x_i y_i.$$

The other is the space $L^2$ with

$$(x, y) = \int x(t)y(t)\, dt.$$

Since $\|x\| \geq 0$ with equality only for $x = \theta$, we have

$$\begin{aligned} 0 \leq \|x - \lambda y\|^2 &= (x - \lambda y, x - \lambda y) \\ &= (x, x) - 2\lambda(x, y) + \lambda^2(y, y). \end{aligned}$$

If $\lambda > 0$, we have

$$2(x, y) \leq \lambda^{-1}\|x\|^2 + \lambda\|y\|^2.$$

Setting $\lambda = \|x\|/\|y\|$, we obtain

$$(x, y) \leq \|x\| \cdot \|y\|,$$

[1] We have defined here a Hilbert space. In analysis it is generally more convenient to deal with a complex Hilbert space, that is, one in which the scalars are complex numbers, the inner product is complex-valued, and (ii) is replaced by (ii′): $(x, y) = \overline{(y, x)}$.

and we see that equality can only occur when $y = \theta$ or $x = \lambda y$ for some $\lambda \geq 0$. This inequality is variously known as the Schwarz, Cauchy–Schwarz, or Cauchy–Buniakowsky–Schwarz inequality. A consequence of this inequality is that the linear functional $f$ defined by $f(x) = (x, y)$ is bounded by $\|y\|$, and from this it follows that $(x, y)$ is a continuous function from $H \times H$ to $\mathbf{R}$.

We say that two elements $x$ and $y$ of $H$ are *orthogonal* if $(x, y) = 0$. We write $x \perp y$ to mean $x$ and $y$ are orthogonal. A set $\mathcal{S}$ in $H$ is called an *orthogonal system* if any two different elements $\varphi$ and $\psi$ of $\mathcal{S}$ are orthogonal, that is, $(\varphi, \psi) = 0$. An orthogonal system $\mathcal{S}$ is called *orthonormal* if $\|\varphi\| = 1$ for each $\varphi$ in $\mathcal{S}$. Any two elements of an orthonormal system are at distance $\sqrt{2}$ from each other. Hence if $H$ is separable, every orthonormal system in $H$ must be countable.

Henceforth we shall deal only with separable Hilbert spaces. Thus each orthonormal system may be expressed as a sequence $\langle \varphi_\nu \rangle$, which may be finite or infinite. We define the Fourier coefficients (with respect to $\langle \varphi_\nu \rangle$) of an element $x$ in $H$ to be $a_\nu = (x, \varphi_\nu)$. For any $n$ we have[2]

$$0 \leq \left\| x - \sum_{\nu=1}^{n} a_\nu \varphi_\nu \right\|^2 = \|x\|^2 - 2 \sum_{\nu=1}^{n} a_\nu (x, \varphi_\nu) + \sum_{\nu=1}^{n} \sum_{\mu=1}^{n} a_\nu a_\mu (\varphi_\nu, \varphi_\mu)$$

$$= \|x\|^2 - \sum_{\nu=1}^{n} a_\nu^2.$$

Thus

$$\sum_{\nu=1}^{n} a_\nu^2 \leq \|x\|^2,$$

and, since $n$ was arbitrary, we have Bessel's inequality:

$$\sum_{\nu=1}^{\infty} a_\nu^2 \leq \|x\|^2.$$

On the other hand, let $\langle a_\nu \rangle$ be any sequence of real numbers with $\sum_{\nu=1}^{\infty} a_\nu^2 < \infty$. Then the sequence

$$z_n = \sum_{\nu=1}^{n} a_\nu \varphi_\nu$$

[2] If there are only a finite number $N$ of elements in $\langle \varphi_\nu \rangle$, we adopt the convention that $\sum_{\nu=1}^{n}$ means $\sum_{\nu=1}^{N}$ for $N \leq n \leq \infty$.

is a Cauchy sequence, since for $m \geq n$

$$z_m - z_n = \sum_{n+1}^{m} a_\nu \varphi_\nu,$$

and we have

$$\| z_m - z_n \|^2 = \sum_{n+1}^{m} a_\nu^2,$$

which must tend to zero, since $\sum a_\nu^2$ converges. By the completeness of $H$ there is an element $y$ such that $y = \lim z_n$, and we write

$$y = \sum_{\nu=1}^{\infty} a_\nu \varphi_\nu.$$

Since the inner product is continuous, we have

$$(y, \varphi_\nu) = \lim (z_n, \varphi_\nu) = a_\nu.$$

We have thus shown that for any $x$ there is a $y$ of the form $y = \sum_{\nu=1}^{\infty} a_\nu \varphi_\nu$ that has the same Fourier coefficients as $x$.

When does this $y$ equal $x$? If we look at $y - x$, we see that all its Fourier coefficients are zero. Hence $y = x$ if the orthonormal system $\langle \varphi_\nu \rangle$ has the property that, if $(z, \varphi_\nu) = 0$ for all $\nu$, then $z = \theta$. An orthonormal system with this property is called *complete* (or total). A complete orthonormal system is clearly a maximal one, while if $\langle \varphi_\nu \rangle$ is a maximal orthonormal system, it must be complete. For if $(z, \varphi_\nu) = 0$ all $\nu$ and $z \neq \theta$, then $z/\|z\|$ can be added to $\langle \varphi_\nu \rangle$. The Hausdorff Maximal Principle implies the existence of a maximal orthonormal system. We have thus established the following proposition:

**27. Proposition:** *In a separable Hilbert space every orthonormal system is countable and there is a complete orthonormal system. If $\langle \varphi_\nu \rangle$ is any complete orthonormal system and $x$ any element of $H$, we have*

$$x = \sum_{\nu=1}^{\infty} a_\nu \varphi_\nu,$$

*where $a_\nu = (x, \varphi_\nu)$. Moreover,* $\|x\|^2 = \sum_{\nu=1}^{\infty} a_\nu^2$.

In a separable Hilbert space there are two alternatives: Either every complete orthonormal system has an infinite number of elements or else there is a complete orthonormal system with a finite number $N$ of elements. In the latter case such a system is a basis (in the vector space sense) of $H$ by Proposition 27. Hence $H$ is a finite-dimensional vector space, and any system of $N + 1$ elements is linearly dependent. Consequently, every orthonormal system can have at most $N$ elements in it. From this it follows that every complete orthogonal system must have $N$ elements. We have thus proved that in a separable Hilbert space $H$ every complete orthonormal system has the same number of elements. We call this number the dimension of $H$. (Thus if $H$ has a countably infinite complete orthonormal system, we say that dim $H = \aleph_0$.)

An isomorphism $\Phi$ of a Hilbert space $H$ onto a Hilbert space $H'$ is a linear mapping of $H$ onto $H'$ such that $(\Phi x, \Phi y) = (x, y)$. Thus an isomorphism between Hilbert spaces is an isometry. Each $n$-dimensional Hilbert space is isomorphic to $\mathbf{R}^n$, since the mapping defined by $\Phi(x) = \langle a_1, \dots, a_n \rangle$, where $a_\nu = (x, \varphi_\nu)$ is an isomorphism. Similarly, every $\aleph_0$-dimensional Hilbert is isomorphic to $l^2$. Since $L^2[0, \pi]$ is separable and $\{\cos \nu t\}$ is an infinite orthogonal system, we see that the dimension of $L^2$ is $\aleph_0$, and so $L^2$ is isomorphic to $l^2$.

**28. Proposition:** *Let $f$ be a bounded linear functional on the Hilbert space $H$. Then there is a unique $y \in H$ such that $f(x) = (x, y)$ for all $x$. Moreover, $\|f\| = \|y\|$.*

***Proof:*** (We consider only the case of separable $H$. For non-separable $H$ see Problem 52.) Let $\langle \varphi_\nu \rangle$ be a complete orthonormal system for $H$, and set $b_\nu = f(\varphi_\nu)$. Then for each $n$ we have

$$\sum_{\nu=1}^{n} b_\nu^2 = f\left(\sum_{\nu=1}^{n} b_\nu \varphi_\nu\right) \leq \|f\| \cdot \left\| \sum_{\nu=1}^{n} b_\nu \varphi_\nu \right\|$$

$$\leq \|f\| \left[\sum_{\nu=1}^{n} b_\nu^2\right]^{1/2}.$$

Thus $\sum_{\nu=1}^{n} b_\nu^2 \leq \|f\|^2$, and so $\sum_{\nu=1}^{\infty} b_\nu^2 \leq \|f\|^2 < \infty$. Hence there is an element $y = \sum_{\nu=1}^{\infty} b_\nu \varphi_\nu$. We have $\|y\| \leq \|f\|$.

Let $x$ be any element of $H$. Then $\sum_{\nu=1}^{n} a_\nu \varphi_\nu \to x$, and so

$$\begin{aligned} f(x) &= \lim f\left(\sum_{\nu=1}^{n} a_\nu \varphi_\nu\right) = \lim \sum_{\nu=1}^{n} a_\nu b_\nu \\ &= \sum_{\nu=1}^{\infty} a_\nu b_\nu \\ &= (x, y). \end{aligned}$$

By the Schwarz inequality $\|f\| \le \|y\|$. ▮

## Problems

**50.** Show that the inner product is continuous; that is, if $x_n \to x$ and $y_n \to y$, then $(x_n, y_n) \to (x, y)$.

**51. a.** Show that $\{\cos \nu t, \sin \nu t\}$ is (when suitably normalized) a complete orthonormal system for $L^2[0, 2\pi]$ (cf. Proposition 6.8 and Problem 9.42).

**b.** Every function in $L^2[0, 2\pi]$ is the limit in mean (of order 2) of its Fourier series (cf. Section 6.3).

**52. a.** Show that for each $x$ in a nonseparable Hilbert space there are only a countable number of nonzero Fourier coefficients (with respect to a fixed orthonormal system).

**b.** Show that Proposition 27 is still valid in a nonseparable Hilbert space except that every complete orthonormal system is uncountable.

**c.** Show that Proposition 28 is still valid in a nonseparable Hilbert space.

**d.** Show that if $H$ is any infinite-dimensional Hilbert space, then the number $\mathbf{n}$ of elements in a complete orthonormal system in $H$ is the smallest cardinal $\mathbf{n}$ such that there is a dense subset of $H$ with $\mathbf{n}$ elements. Hence every complete orthonormal system in $H$ has the same number of elements. We call this number the dimension of $H$.

**e.** Show that two Hilbert spaces are isomorphic if and only if they have the same dimension.

**f.** Show that there is a Hilbert space of each dimension.

**53.** Let $P$ be a subset of $H$. By the orthogonal complement $P^\perp$ of $P$ we mean the set $\{y: y \perp x \text{ all } x \in P\}$.

**a.** Show that $P^\perp$ is always a closed linear manifold.

**b.** Show that $P^{\perp\perp}$ is the smallest closed linear manifold containing $P$.

**c.** Let $M$ be a closed linear manifold. Then each $x \in H$ can be written uniquely in the form $x = y + z$ with $y \in M$ and $z \in M^{\perp}$. Moreover, $\|x\|^2 = \|y\|^2 + \|z\|^2$.

**54.** Let $\langle x_n \rangle$ be a bounded sequence of elements in a separable Hilbert space. Then $\langle x_n \rangle$ contains a subsequence which converges weakly.

**55.** Let $S$ be a subspace of $L^2[0, 1]$, and suppose that there is a constant $K$ such that $|f(x)| \leq K\|f\|$ for all $x \in [0, 1]$. Then the dimension of $S$ is at most $K^2$. [Hint: If $\langle f_1, \ldots, f_n \rangle$ is any finite orthonormal sequence in $S$, then $\sum_{i=1}^{n} |f_i(x)|^2 \leq K^2$.]

# Part Three

## GENERAL MEASURE AND INTEGRATION THEORY

# 11 Measure and Integration

## 1 Measure Spaces

The purpose of the present chapter is to abstract the most important properties of Lebesgue measure and Lebesgue integration. We shall do this by giving certain axioms which Lebesgue measure satisfies and base our integration theory on these axioms. As a consequence our theory will be valid for every system satisfying the given axioms.

We begin by recalling that a $\sigma$-algebra $\mathcal{B}$ is a family of subsets of a given set $X$ which contains $\emptyset$ and is closed with respect to complements and with respect to countable unions. By a set function $\mu$ we mean a function which assigns an extended real number to certain sets. With this in mind we make the following definitions:

**Definition:** *By a measurable space we mean a couple $(X, \mathcal{B})$ consisting of a set $X$ and $\sigma$-algebra $\mathcal{B}$ of subsets of $X$. A subset $A$ of $X$ is called measurable (or measurable with respect to $\mathcal{B}$) if $A \in \mathcal{B}$.*

**Definition:** *By a measure $\mu$ on a measurable space $(X, \mathcal{B})$ we mean a nonnegative set function defined for all sets of $\mathcal{B}$ and satisfying $\mu(\emptyset) = 0$ and*

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu E_i$$

*for any sequence $E_i$ of disjoint measurable sets. By a measure space $(X, \mathcal{B}, \mu)$ we mean a measurable space $(X, \mathcal{B})$ together with a measure $\mu$ defined on $\mathcal{B}$.*

This second property of $\mu$ is often referred to by saying that $\mu$ is countably additive. We also have that $\mu$ is "finitely additive"; that is,

$$\mu\left(\bigcup_{i=1}^{N} E_i\right) = \sum_{i=1}^{N} \mu E_i,$$

for disjoint sets $E_i$ belonging to $\mathcal{B}$, since we may set $E_i = \emptyset$ for $i > N$.[1]

One example of a measure space is $(\mathbf{R}, \mathfrak{M}, m)$, where $\mathbf{R}$ is the set of real numbers, $\mathfrak{M}$ the Lebesgue measurable sets of real numbers, and $m$ Lebesgue measure. Another measure space results if we replace $\mathbf{R}$ by the interval $[0, 1]$ and $\mathfrak{M}$ by the measurable subsets of $[0, 1]$. A third example is $(\mathbf{R}, \mathcal{B}, m)$, where $\mathcal{B}$ is the class of Borel sets and $m$ is again Lebesgue measure. Another example is the counting measure (see Problem 3.4). A slightly bizarre example is the following. Let $X$ be any uncountable set, $\mathcal{B}$ the family of those subsets which are either countable or the complement of a countable set. Then $\mathcal{B}$ is a $\sigma$-algebra and we can define a measure on it by setting $\mu A = 0$ for each countable set and $\mu B = 1$ for each set whose complement is countable.

Two further properties of measures are given by the following propositions:

**1. Proposition:** *If $A \in \mathcal{B}$, $B \in \mathcal{B}$, and $A \subset B$, then*

$$\mu A \leq \mu B.$$

***Proof:*** Since

$$B = A \cup [B \sim A]$$

is a disjoint union, we have

$$\mu B = \mu A + \mu(B \sim A) \geq \mu A. \quad \blacksquare$$

[1] A set function $\mu$ defined on an algebra of sets and satisfying $\mu(\emptyset) = 0$ and $\mu(A \cup B) = \mu A + \mu B$ for disjoint sets $A$ and $B$ in the algebra is called a *finitely additive measure.* Since our definition (and the usual usage) requires a measure to be countably additive, it follows that a finitely additive measure is not in general a measure, although every measure is, of course, a finitely additive measure.

**2. Proposition:** *If* $E_i \in \mathcal{B}$, $\mu E_1 < \infty$ *and* $E_i \supset E_{i+1}$, *then*

$$\mu\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{n\to\infty} \mu E_n.$$

***Proof:*** Set $E = \bigcap_{i=1}^{\infty} E_i$. Then

$$E_1 = E \cup \bigcup_{i=1}^{\infty} (E_i \sim E_{i+1}),$$

and this is a disjoint union. Hence

$$\mu E_1 = \mu E + \sum_{i=1}^{\infty} \mu(E_i \sim E_{i+1}).$$

Since

$$E_i = E_{i+1} \cup (E_i \sim E_{i+1})$$

is a disjoint union, we have

$$\mu(E_i \sim E_{i+1}) = \mu E_i - \mu E_{i+1}.$$

Hence

$$\begin{aligned} \mu E_1 &= \mu E + \sum_{i=1}^{\infty} (\mu E_i - \mu E_{i+1}) \\ &= \mu E + \lim_{n\to\infty} \sum_{i=1}^{n-1} (\mu E_i - \mu E_{i+1}) \\ &= \mu E + \mu E_1 - \lim \mu E_n, \end{aligned}$$

whence the proposition follows. ∎

**3. Proposition:** *If* $E_i \in \mathcal{B}$, *then*

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) \le \sum_{i=1}^{\infty} \mu E_i.$$

***Proof:*** Let

$$G_n = E_n \sim \left[\bigcup_{i=1}^{n-1} E_i\right].$$

Then $G_n \subset E_n$ and the sets $G_n$ are disjoint. Hence

$$\mu G_n \le \mu E_n,$$

while

$$\mu(\bigcup E_i) = \sum_{n=1}^{\infty} \mu G_n \leq \sum_{n=1}^{\infty} \mu E_n. \quad \blacksquare$$

A measure $\mu$ is called **finite** if $\mu(X) < \infty$. It is called **σ-finite** if there is a sequence $\langle X_n \rangle$ of measurable sets in $\mathcal{B}$ such that

$$X = \bigcup_{n=1}^{\infty} X_n$$

and $\mu X_n < \infty$. By virtue of Proposition 1.2 we may always take $\langle X_n \rangle$ to be a disjoint sequence of sets. Lebesgue measure on $[0, 1]$ is an example of a finite measure, while Lebesgue measure on $(-\infty, \infty)$ is an example of a $\sigma$-finite measure. The counting measure on an uncountable set is a measure that is not $\sigma$-finite. Other examples are given in Problem 48.

A set $E$ is said to be of **finite measure** if $E \in \mathcal{B}$ and $\mu E < \infty$. A set $E$ is said to be of **σ-finite measure** if $E$ is the union of a countable collection of measurable sets of finite measure. Any measurable set contained in a set of $\sigma$-finite measure is itself of $\sigma$-finite measure, and the union of a countable collection of sets of $\sigma$-finite measure is again of $\sigma$-finite measure. If $\mu$ is $\sigma$-finite, then every measurable set is of $\sigma$-finite measure.

Roughly speaking, nearly all the familiar properties of Lebesgue measure and Lebesgue integration hold for arbitrary $\sigma$-finite measures, and many treatments of abstract measure theory limit themselves to $\sigma$-finite measures. However, many parts of the general theory do not require the assumption of $\sigma$-finiteness, and it seems undesirable to have a development that is unnecessarily restrictive. A notion weaker than that of $\sigma$-finiteness is semifiniteness. A measure $\mu$ is said to be **semifinite** if each measurable set of infinite measure contains measurable sets of arbitrary large finite measure. Thus every $\sigma$-finite measure is semifinite, while the measure that assigns 0 to countable subsets of an uncountable set $X$ and $\infty$ to the uncountable sets is not semifinite. Measures that are not semifinite tend to be somewhat bizarre, and there is a temptation to restrict oneself to the consideration of semifinite measures. Unfortunately, many of the natural operations on measures, such as the extension procedures of the next chapter, may take us out of the class of semifinite measures.

A measure space $(X, \mathcal{B}, \mu)$ is said to be **complete** if $\mathcal{B}$ contains all subsets of sets of measure zero, that is, if $B \in \mathcal{B}$, $\mu B = 0$, and $A \subset B$

imply $A \in \mathcal{B}$. Thus Lebesgue measure is complete, while Lebesgue measure restricted to the $\sigma$-algebra of Borel sets is not complete. The following proposition, whose proof is left to the reader (Problem 7), shows that each measure space can be completed by the addition of subsets of sets of measure zero. The measure space $(X, \mathcal{B}_0, \mu_0)$ given in the proposition is called the *completion* of $(X, \mathcal{B}, \mu)$.

**4. Proposition:** *If $(X, \mathcal{B}, \mu)$ is a measure space, then we can find a complete measure space $(X, \mathcal{B}_0, \mu_0)$ such that*

i. $\mathcal{B} \subset \mathcal{B}_0$.
ii. $E \in \mathcal{B} \Rightarrow \mu E = \mu_0 E$.
iii. $E \in \mathcal{B}_0 \Leftrightarrow E = A \cup B$ *where* $B \in \mathcal{B}$ *and* $A \subset C$, $C \in \mathcal{B}$, $\mu C = 0$.

If $(X, \mathcal{B}, \mu)$ is a measure space, we say that a subset $E$ of $X$ is locally measurable if $E \cap B \in \mathcal{B}$ for each $B \in \mathcal{B}$ with $\mu B < \infty$. The collection $\mathcal{C}$ of all locally measurable sets is a $\sigma$-algebra containing $\mathcal{B}$. The measure $\mu$ is called **saturated** if every locally measurable set is measurable (i.e., is in $\mathcal{B}$). Every $\sigma$-finite measure is saturated. A measure can always be extended to a saturated measure, but unlike the process of completion, the process of saturation is not uniquely determined: There is no difficulty with the $\sigma$-algebra, for we simply take the $\sigma$-algebra $\mathcal{C}$ of all locally measurable sets, but the measure $\mu$ can often be extended to $\mathcal{C}$ in different ways. Some details are given in Problem 8, and we shall return to this topic in the next chapter.

## Problems

**1.** Let $\{A_n\}$ be a countable collection of measurable sets. Then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} A_k\right).$$

**2.** Let $\{(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha)\}$ be a collection of measure spaces, and suppose that the sets $\{X_\alpha\}$ are disjoint. Then we can form a new measure space (called their union) by letting $X = \bigcup X_\alpha$, $\mathcal{B} = \{B: (\alpha)[B \cap X_\alpha \in \mathcal{B}_\alpha]\}$, and defining $\mu(B) = \sum \mu_\alpha(B \cap X_\alpha)$.

**a.** Show that $\mathcal{B}$ is a $\sigma$-algebra.

**b.** Show that $\mu$ is a measure.

**c.** Show that $\mu$ is $\sigma$-finite if and only if all but a countable number of the $\mu_\alpha$ are zero and the remainder are $\sigma$-finite.

**3. a.** Show that $\mu(E_1 \Delta E_2) = 0$ implies $\mu E_1 = \mu E_2$ provided that $E_1$ and $E_2 \in \mathcal{B}$.

**b.** Show that if $\mu$ is complete, then $E_1 \in \mathcal{B}$ and $\mu(E_1 \Delta E_2) = 0$ imply $E_2 \in \mathcal{B}$.

**4.** Let $(X, \mathcal{B}, \mu)$ be a measure space and $Y \in \mathcal{B}$. Let $\mathcal{B}_Y$ consist of those sets of $\mathcal{B}$ that are contained in $Y$. Set $\mu_Y E = \mu E$ if $E \in \mathcal{B}_Y$. Then $(Y, \mathcal{B}_Y, \mu_Y)$ is a measure space. The measure $\mu_Y$ is called the **restriction** of $\mu$ to $Y$.

**5.** Let $(X, \mathcal{B})$ be a measurable space.

**a.** If $\mu$ and $\nu$ are measures defined on $\mathcal{B}$, then the set function $\lambda$ defined on $\mathcal{B}$ by $\lambda E = \mu E + \nu E$ is also a measure. We denote $\lambda$ by $\mu + \nu$.

**b.** If $\mu$ and $\nu$ are measures on $\mathcal{B}$ and $\mu \geq \nu$, then there is a measure $\lambda$ on $\mathcal{B}$ such that $\mu = \nu + \lambda$.

**c.** If $\nu$ is $\sigma$-finite, the measure $\lambda$ in (b) is unique.

**d.** Show that in general the measure $\lambda$ need not be unique but that there is always a smallest such $\lambda$.

**6. a.** Show that each $\sigma$-finite measure is semifinite.

**b.** Show that each measure $\mu$ is the sum $\mu_1 + \mu_2$ of a semifinite measure $\mu_1$ and a measure $\mu_2$ that assumes only the values 0 and $\infty$.

**c.** The measures $\mu_1$ and $\mu_2$ are not unique, but there is a smallest $\mu_1$ and for that $\mu_1$ a smallest $\mu_2$.

**7.** Prove Proposition 4. [First show that the family $\mathcal{B}_0$ defined by (iii) is a $\sigma$-algebra. If $E \in \mathcal{B}_0$, show that $\mu A$ is the same for all sets $A \in \mathcal{B}$ such that $E = A \cup B$ with $B$ a subset of a set of measure zero. Use this fact to define $\mu_0$ and show $\mu_0$ is a measure.]

**8. a.** Show that each $\sigma$-finite measure is saturated.

**b.** Show that the collection of locally measurable sets is a $\sigma$-algebra.

**c.** Let $(X, \mathcal{B}, \mu)$ be a measure space and $\mathcal{C}$ the $\sigma$-algebra of locally measurable sets. For $E \in \mathcal{C}$ set $\bar{\mu} E = \mu E$ if $E \in \mathcal{B}$ and $\bar{\mu} E = \infty$ if $E \notin \mathcal{B}$. Show that $(X, \mathcal{C}, \bar{\mu})$ is a saturated measure space.

**d.** If $\mu$ is semifinite and $E \in \mathcal{C}$, set $\underline{\mu} E = \sup \{\mu B \colon B \in \mathcal{B},\ B \subset E\}$. Show that $(X, \mathcal{C}, \underline{\mu})$ is a saturated measure space and that $\underline{\mu}$ is an extension of $\mu$. Give an example to show that $\bar{\mu}$ and $\underline{\mu}$ may be different.

**9.** *$\sigma$-Rings and $\sigma$-algebras.* Some authors prefer to define a measure on a $\sigma$-ring $\mathcal{R}$, that is, a collection of subsets of $X$ such that if $A$ and $B$ are in $\mathcal{R}$ so is $A \sim B$, and if $\langle A_n \rangle$ is a sequence from $\mathcal{R}$ then $\bigcup A_n$ is in $\mathcal{R}$. Thus a $\sigma$-ring $\mathcal{R}$ is a $\sigma$-algebra if and only if $X \in \mathcal{R}$. Some of the relations between $\sigma$-rings and $\sigma$-algebras are given below:

**a.** Let $\mathcal{R}$ be a $\sigma$-ring which is not a $\sigma$-algebra, let $\mathcal{B}$ be the smallest $\sigma$-algebra containing $\mathcal{R}$, and set $\mathcal{R}' = \{E \colon \tilde{E} \in \mathcal{R}\}$. Then $\mathcal{B} = \mathcal{R} \cup \mathcal{R}'$ and $\mathcal{R} \cap \mathcal{R}' = \emptyset$.

**b.** If $\mu$ is a measure on $\mathfrak{R}$, define $\bar{\mu}$ on $\mathfrak{B}$ by $\bar{\mu}E = \mu E$ if $E \in \mathfrak{R}$ and $\bar{\mu}E = \infty$ if $E \in \mathfrak{R}'$. Then $\bar{\mu}$ is a measure on $\mathfrak{B}$.

**c.** Define $\underline{\mu}$ on $\mathfrak{B}$ by $\underline{\mu}E = \mu E$ if $E \in \mathfrak{R}$ and

$$\underline{\mu}E = \sup \{\mu A: A \subset E, A \in \mathfrak{R}\}$$

if $E \in \mathfrak{R}'$. Then $\underline{\mu}$ is also a measure on $\mathfrak{B}$.

**d.** More generally, for any nonnegative extended real number $\beta$ define $\mu_\beta$ on $\mathfrak{B}$ by $\mu_\beta E = \mu E$ if $E \in \mathfrak{R}$ and $\mu_\beta E = \underline{\mu}E + \beta$ if $E \in \mathfrak{R}'$. Then $\mu_\beta$ is a measure on $\mathfrak{B}$.

**e.** If $\nu$ is any measure on $\mathfrak{B}$ which agrees with $\mu$ on $\mathfrak{R}$, then $\nu = \mu_\beta$ for some $\beta \geq 0$.

## 2 Measurable Functions

The concept of a measurable function on an abstract measurable space is almost identical with that for functions of a real variable. Consequently, those propositions and theorems whose proofs are essentially the same as those in Section 5 of Chapter 3 are stated without proof. The reader should verify that the proofs in Chapter 3 can be carried out in the abstract case. Throughout this section we assume that a fixed measurable space $(X, \mathfrak{B})$ is given.

**5. Proposition:** *Let $f$ be an extended real-valued function defined on $X$. Then the following statements are equivalent:*

i. $\{x: f(x) < \alpha\} \in \mathfrak{B}$ *for each* $\alpha$.
ii. $\{x: f(x) \leq \alpha\} \in \mathfrak{B}$ *for each* $\alpha$.
iii. $\{x: f(x) > \alpha\} \in \mathfrak{B}$ *for each* $\alpha$.
iv. $\{x: f(x) \geq \alpha\} \in \mathfrak{B}$ *for each* $\alpha$.

(See Proposition 3.18.)

**Definition:** *The extended real-valued function $f$ defined on $X$ is called measurable (or measurable with respect to $\mathfrak{B}$) if any one of the statements of Proposition 5 holds.*

**6. Theorem:** *If $c$ is a constant and the functions $f$ and $g$ are measurable, then so are the functions $f + c$, $cf$, $f + g$, $f \cdot g$, and $f \vee g$. Moreover, if $\langle f_n \rangle$ is a sequence of measurable functions, then* $\sup f_n$, $\inf f_n$, $\overline{\lim}\, f_n$, *and* $\underline{\lim}\, f_n$ *are all measurable.*

(See Proposition 3.19 and Theorem 3.20.)

By a simple function we mean as before a finite linear combination

$$\varphi(x) = \sum_{i=1}^{n} c_i \chi_{E_i}(x)$$

of characteristic functions of measurable sets $E_i$.

**7. Proposition:** *Let $f$ be a nonnegative measurable function. Then there is a sequence $\langle \varphi_n \rangle$ of simple functions with $\varphi_{n+1} \geq \varphi_n$ such that $f = \lim \varphi_n$ at each point of $X$. If $f$ is defined on a $\sigma$-finite measure space, then we may choose the functions $\varphi_n$ so that each vanishes outside a set of finite measure.*

**8. Proposition:** *If $\mu$ is a complete measure and $f$ is a measurable function, then $f = g$ a.e. implies $g$ is measurable.*

(See Proposition 3.21.)

The sets $\{x : f(x) < \alpha\}$ are sometimes called ordinate sets for $f$. They increase with $\alpha$. The following lemma states that, given a collection $\{B_\alpha\}$ of measurable sets that increase with $\alpha$, we can find a measurable function $f$ which nearly has these for ordinate sets in the sense that

$$\{x : f(x) < \alpha\} \subset B_\alpha \subset \{x : f(x) \leq \alpha\}.$$

**9. Lemma:** *Suppose that to each $\alpha$ in a dense set $D$ of real numbers there is assigned a set $B_\alpha \in \mathcal{B}$ such that $B_\alpha \subset B_\beta$ for $\alpha < \beta$. Then there is a unique measurable extended real-valued function $f$ on $X$ such that $f \leq \alpha$ on $B_\alpha$ and $f \geq \alpha$ on $X \sim B_\alpha$.*

***Proof:*** For each $x \in X$ define

$$f(x) = \inf \{\alpha \in D : x \in B_\alpha\}$$

where, as usual, $\inf \emptyset = \infty$. If $x \in B_\alpha$, then $f(x) \leq \alpha$. If $x \notin B_\alpha$, then $x \notin B_\beta$ for each $\beta < \alpha$, and so $f(x) \geq \alpha$. To show that $f$ is measurable, we take $\lambda \in \mathbf{R}$ and choose a sequence $\langle \alpha_n \rangle$ from $D$ with $\alpha_n < \lambda$ and $\lambda = \lim \alpha_n$. Then

$$\{x : f(x) < \lambda\} = \bigcup_{n=1}^{\infty} B_{\alpha_n}.$$

For if $f(x) < \lambda$, then $f(x) < \alpha_n$ for some $n$, and so $x \in B_{\alpha_n}$. If $x \in B_{\alpha_n}$ for any $n$, then $f(x) \leq \alpha_n < \lambda$. Thus the sets $\{x : f(x) < \lambda\}$ are all measurable, and so $f$ is measurable.

To prove the unicity of $f$, let $g$ be any extended real-valued function with $g \leq \alpha$ on $B_\alpha$ and $g \geq \alpha$ on $\tilde{B}_\alpha$. Then $x \in B_\alpha$ implies $g(x) \leq \alpha$, and so

$$\{\alpha \in D : x \in B_\alpha\} \subset \{\alpha \in D : \alpha \geq g(x)\}.$$

Since $g(x) < \alpha$ implies that $x \in B_\alpha$ we have

$$\{\alpha \in D : \alpha > g(x)\} \subset \{\alpha \in D : x \in B_\alpha\}.$$

Because of the density of $D$ we have

$$\begin{aligned} g(x) &= \inf \{\alpha \in D : \alpha > g(x)\} = \inf \{\alpha \in D : \alpha \geq g(x)\} \\ &= \inf \{\alpha \in D : x \in B_\alpha\} = f(x). \quad \blacksquare \end{aligned}$$

The preceding lemma shows that a function is uniquely determined by its ordinate sets and that these ordinate sets may be taken to be an arbitrary increasing family $\{B_\alpha\}$, provided we are flexible about whether or not an $x$ with $f(x) = \alpha$ belongs to $B_\alpha$.

In practice we often have to deal with classes of measurable functions (under equality almost everywhere). In this case the ordinate sets are only defined modulo null sets; that is, the sets $\{B_\alpha\}$ satisfy $f(x) \geq \alpha$ a.e. on $B_\alpha$ and $f(x) \leq \alpha$ a.e. on $\tilde{B}_\alpha$. For a family $\{B_\alpha\}$ of sets to be a family of ordinate sets in this sense, we need only require that $\{B_\alpha\}$ is almost increasing; i.e. $\mu\{B_\alpha \sim B_\beta\} = 0$ for $\beta > \alpha$. This is the content of the following proposition. It will prove useful in Section 6 and again in Chapter 15.

**10. Proposition:** *Suppose that for each $\alpha$ in a dense set $D$ of real numbers there is assigned a set $B_\alpha \in \mathcal{B}$ such that $\mu(B_\alpha \sim B_\beta) = 0$ for $\alpha < \beta$. Then there is a measurable function $f$ such that $f \leq \alpha$* a.e. *on $B_\alpha$ and $f \geq \alpha$* a.e. *on $X \sim B_\alpha$. If $g$ is any other function with this property, then $g = f$* a.e.

***Proof:*** Let $C$ be a countable dense subset of $D$, and set $N = \bigcup (B_\alpha \sim B_\beta)$ for $\alpha$ and $\beta$ in $C$ with $\alpha < \beta$. Then $N$ is the countable union of sets of measure zero and so is itself a set of measure zero. Let $B'_\alpha = B_\alpha \cup N$. For $\alpha$ and $\beta$ in $C$ with $\alpha < \beta$ we have

$$B'_\alpha \sim B'_\beta = (B_\alpha \sim B_\beta) \sim N = \emptyset.$$

Thus $B'_\alpha \subset B'_\beta$. By Lemma 9 there is a measurable function $f$ such that $f \leq \gamma$ on $B'_\gamma$ and $f \geq \gamma$ on $\tilde{B}'_\gamma$.

Let $\alpha \in D$, and choose a sequence $\langle \gamma_n \rangle$ from $C$ with $\alpha < \gamma_n$ and $\alpha = \lim \gamma_n$. Then

$$B_\alpha \sim B'_{\gamma_n} \subset (B_\alpha \sim B_{\gamma_n}).$$

Thus $P = \bigcup_n (B_\alpha \sim B'_{\gamma_n})$ is a countable union of null sets and so a null set. Let $A = \bigcap B'_{\gamma_n}$. Then $f \le \inf \gamma_n = \alpha$ on $A$, and $A \sim B_\alpha \subset P$. Thus $f \le \alpha$ almost everywhere on $B_\alpha$. A similar argument shows that $f \ge \alpha$ almost everywhere on $\tilde{B}_\alpha$.

Let $g$ be an extended real-valued function with $g \le \gamma$ a.e. on $B_\gamma$ and $g \ge \gamma$ on $\tilde{B}_\gamma$ for each $\gamma \in C$. Then $g \le \gamma$ on $B'_\gamma$ and $g \ge \gamma$ on $\tilde{B}'_\gamma$ except for $x$ in a null set $Q_\gamma$. Thus $Q = \bigcup Q_\gamma$ is a null set and we must have $f = g$ on $X \sim Q$. ▮

## Problems

**10.** Prove Proposition 7. [For each pair $\langle n, k \rangle$ of integers let

$$E_{n,k} = \{x\colon k2^{-n} \le f(x) < (k+1)2^{-n}\}, \text{ and set } \varphi_n = 2^{-n} \sum_{k=0}^{2^{2n}} k\chi_{E_{n,k}}.]$$

**11.** Prove Proposition 8, and show that it is false if the word 'complete' is omitted.

**12.** Let $\langle f_n \rangle$ be a sequence of measurable functions that converge to a function $f$ except at the points of set $E$ of measure zero. Then $f$ is a measurable function if $\mu$ is complete.

**13. a.** A sequence $\langle f_n \rangle$ of measurable real-valued functions is said to *converge in measure* to a function $f$ if given $\epsilon > 0$, there is an integer $N$ and a measurable set $E$ with $\mu E < \epsilon$ such that

$$|f_n(x) - f(x)| < \epsilon$$

for all $n \ge N$ and all $x \notin E$. Show that, if $f_n$ converges to $f$ in measure, then there is a subsequence $f_{n_k}$ that converges to $f$ almost everywhere.

**b.** Suppose that $\langle f_n \rangle$ is a sequence of measurable functions each of which vanishes outside a fixed measurable set $A$ with $\mu A < \infty$. Suppose that $f_n(x) \to f(x)$ for almost all $x$. Then $\langle f_n \rangle$ converges to $f$ in measure.

**c.** A sequence $\langle f_n \rangle$ of measurable functions is said to be Cauchy in measure if given $\epsilon > 0$, there is an integer $N$ and a measurable set $E$ with $\mu E < \epsilon$ such that

$$|f_n(x) - f_m(x)| < \epsilon$$

for all $n, m \ge N$ and all $x \notin E$. Show that, if $\langle f_n \rangle$ is Cauchy in measure, then there is a function $f$ to which $\langle f_n \rangle$ converges in measure.

**14.** Let $(X, \mathcal{B}, \mu)$ be a measure space and $(X, \mathcal{B}_0, \mu_0)$ its completion. Then a function $f$ is measurable with respect to $\mathcal{B}_0$ if and only if there is a function $g$ measurable with respect to $\mathcal{B}$ such that $f = g$ almost everywhere in the sense that there is a set $E \in \mathcal{B}$ with $\mu E = 0$ and $f = g$ on $X \sim E$. Note that this does not require the set $\{x : f(x) \neq g(x)\}$ to be in $\mathcal{B}$.

**15.** Let $D$ be the rationals and $f$ the function in Lemma 9. Express the sets $\{x : f(x) < \alpha\}$, $\{x : f(x) \leq \alpha\}$, $\{x : f(x) = \alpha\}$ in terms of the sets $\{B_\beta\}$.

**16.** Prove Egoroff's Theorem (Problem 3.30) in the context of general measure spaces.

## 3 Integration

Many definitions and proofs of Chapter 4 depend on only those properties of Lebesgue measure which are also true for an arbitrary measure in an abstract measure space and carry over to this case. If $E$ is a measurable set and $\varphi$ a nonnegative simple function, we define

$$\int_E \varphi \, d\mu = \sum_{i=1}^{n} c_i \mu(E_i \cap E),$$

where

$$\varphi(x) = \sum_{i=1}^{n} c_i \chi_{E_i}(x).$$

It is easily seen that the value of this integral is independent of the representation of $\varphi$ which we use. If $a$ and $b$ are positive numbers and $\varphi$ and $\psi$ nonnegative simple functions, then

$$\int a\varphi + b\psi = a \int \varphi + b \int \psi.$$

Let $f$ be a bounded measurable function which is identically zero outside a measurable set $E$ of finite measure. Then we can show, just as in Proposition 4.3, that $\inf\limits_{f \leq \psi} \int \psi = \sup\limits_{\varphi \leq f} \int \varphi$ for all simple functions $\varphi$ and $\psi$ if and only if $f = g$ almost everywhere for some measurable function $g$. Equivalently, this equality holds if and only if $f$ is measurable with respect to the completion $\bar{\mu}$ of $\mu$. This result indicates that the natural class of functions to consider for integration theory are those which are measurable with respect to the completion of $\mu$. Hence we shall assume for the rest of this chapter that $\mu$ is a complete measure, unless otherwise specified. In future chapters we define integrals with respect to $\mu$ to be the integral with respect to the completion of $\mu$.

In Chapter 4 we first defined the integral for bounded measurable functions and then defined the integral of a nonnegative measurable function $f$ to be the supremum of $\int g\, d\mu$ as $g$ ranged over all bounded measurable functions which vanished outside a set of finite measure. Unfortunately, this definition is not quite suitable in the case of measures which are not semi-finite. For if $\mathcal{B} = \{X, \emptyset\}$ and $\mu\emptyset = 0$, $\mu X = \infty$, then we certainly want $\int 1\, d\mu = \infty$. But the only measurable function $g$ which vanishes outside a set of finite measure is $g \equiv 0$, and hence $\sup \int g\, d\mu = 0$. To avoid this difficulty, we define the integral of a nonnegative measurable function directly in terms of integrals of nonnegative simple functions.

**Definition:** *Let $f$ be a nonnegative extended real-valued measurable function on the measure space $(X, \mathcal{B}, \mu)$. Then $\int f\, d\mu$ is the supremum of the integrals $\int \varphi\, d\mu$ as $\varphi$ ranges over all simple functions with $0 \le \varphi \le f$.*

It follows immediately from this definition that $f \le g$ implies that $\int f \le \int g$ and $\int cf = c \int f$. In other respects, however, this definition is somewhat awkward to apply, since we are taking a supremum over a large collection of simple functions, and it is not apparent from the definition that $\int (f + g) = \int f + \int g$. Consequently, we begin our treatment of the integral by first establishing the convergence theorems. These then enable us to determine the value of $\int f$ by taking the limit of $\int \varphi_n$ for any increasing sequence $\langle \varphi_n \rangle$ of simple functions which converge to $f$. We begin with Fatou's Lemma:

**11. Theorem (Fatou's Lemma):** *Let $\langle f_n \rangle$ be a sequence of nonnegative measurable functions that converge almost everywhere on a set $E$ to a function $f$. Then*

$$\int_E f \le \underline{\lim} \int_E f_n .$$

***Proof:*** Without loss of generality we may assume that $f_n(x) \to f(x)$ for each $x \in E$. From the definition of $\int f$ it suffices to show that, if $\varphi$ is any nonnegative simple function with $\varphi \le f$, then $\int_E \varphi \le \underline{\lim} \int_E f_n$.

If $\int_E \varphi = \infty$, then there is a measurable set $A \subset E$ with $\mu A = \infty$ such that $\varphi > a > 0$ on $A$. Set $A_n = \{x \in E : f_k(x) > a \text{ for all } k \ge n\}$. Then $\langle A_n \rangle$ is an increasing sequence of measurable sets whose union contains $A$, since $\varphi \le \lim f_n$. Thus $\lim \mu A_n = \infty$. Since $\int_E f_n \ge a\mu A_n$, we have $\lim \int_E f_n = \infty = \int_E \varphi$.

If $\int_E \varphi < \infty$, then the set $A = \{x \,\varepsilon\, E: \varphi(x) > 0\}$ is a measurable set of finite measure. Let $M$ be the maximum of $\varphi$, let $\epsilon$ be a given positive number, and set

$$A_n = \{x \,\varepsilon\, E: f_k(x) > (1 - \epsilon)\varphi(x) \text{ for all } k \geq n\}.$$

Then $\langle A_n \rangle$ is an increasing sequence of sets whose union contains $A$, and so $\langle A \sim A_n \rangle$ is a decreasing sequence of sets whose intersection is empty. By Proposition 2 $\lim \mu(A \sim A_n) = 0$, and so we can find an $n$ such that $\mu(A \sim A_k) < \epsilon$ for all $k \geq n$. Thus for $k \geq n$

$$\begin{aligned}\int_E f_k \geq \int_{A_k} f_k \geq (1 - \epsilon) \int_{A_k} \varphi \\ \geq (1 - \epsilon) \left[\int_E \varphi - \int_{A \sim A_k} \varphi\right] \\ \geq \int_E \varphi - \epsilon\left[\int_E \varphi + M\right].\end{aligned}$$

Hence

$$\underline{\lim} \int_E f_n \geq \int_E \varphi - \epsilon\left[\int_E \varphi + M\right].$$

Since $\epsilon$ is arbitrary,

$$\underline{\lim} \int_E f_n \geq \int_E \varphi. \quad \blacksquare$$

**12. Monotone Convergence Theorem:** *Let $\langle f_n \rangle$ be a sequence of nonnegative measurable functions which converge almost everywhere to a function $f$ and suppose that $f_n \leq f$ for all $n$. Then*

$$\int f = \lim \int f_n .$$

***Proof:*** Since $f_n \leq f$, we have $\int f_n \leq \int f$. Thus by Fatou's Lemma we have

$$\int f \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n \leq \int f. \quad \blacksquare$$

We are now able to establish some of the standard properties of the integral.

**13. Proposition:** *If $f$ and $g$ are nonnegative measurable functions and $a$ and $b$ nonnegative constants, then*

$$\int af + bg = a\int f + b\int g.$$

*We have*

$$\int f \geq 0$$

*with equality only if $f = 0$* a.e.

***Proof:*** To prove the first statement, let $\langle\varphi_n\rangle$ and $\langle\psi_n\rangle$ be increasing sequences of simple functions which converge to $f$ and $g$. Then $\langle a\varphi_n + b\psi_n\rangle$ is an increasing sequence of simple functions which converge to $af + bg$. By the Monotone Convergence Theorem

$$\begin{aligned}\int af + bg &= \lim \int a\varphi_n + b\psi_n \\ &= \lim \left(a\int \varphi_n + b\int \psi_n\right) \\ &= a\int f + b\int g.\end{aligned}$$

Clearly $\int f \geq 0$. If $\int f = 0$, let $A_n = \{x : f(x) \geq 1/n\}$. Then we have $f \geq (1/n)\chi_{A_n}$, and so $\mu A_n = \int \chi_{A_n} = 0$. Since the set where $f > 0$ is the union of the sets $A_n$, it has measure zero. ▮

Taking this proposition together with the Monotone Convergence Theorem gives us the following corollary:

**14. Corollary:** *Let $\langle f_n\rangle$ be a sequence of nonnegative measurable functions. Then*

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

A nonnegative function $f$ is called **integrable** (over a measurable set $E$ with respect to $\mu$) if it is measurable and

$$\int_E f\,d\mu < \infty.$$

An arbitrary function $f$ is said to be integrable if both $f^+$ and $f^-$ are integrable. In this case we define

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

Some of the properties of the integral are contained in the following proposition, whose proof is left to the reader.

**15. Proposition:** *If $f$ and $g$ are integrable functions and $E$ is a measurable set, then*

i. $\int_E (c_1 f + c_2 g) = c_1 \int_E f + c_2 \int_E g$.
ii. *If $|h| \le |f|$ and $h$ is measurable then $h$ is integrable.*
iii. *If $f \ge g$ a.e., then $\int f \ge \int g$.*

**16. Lebesgue Convergence Theorem:** *Let $g$ be integrable over $E$, and suppose that $\langle f_n \rangle$ is a sequence of measurable functions such that on $E$*

$$|f_n(x)| \le g(x)$$

*and such that almost everywhere on $E$*

$$f_n(x) \to f(x).$$

*Then*

$$\int_E f = \lim \int_E f_n.$$

***Proof:*** Apply Fatou's Lemma to sequences $g + f_n$ and $g - f_n$. ∎

## Problems

**17.** Prove Proposition 15.

**18.** Suppose that $\mu$ is not complete, but that we define a bounded function $f$ to be integrable over a set $E$ of finite measure if

$$\sup_{\varphi \le f} \int_E \varphi \, d\mu = \inf_{\psi \ge f} \int_E \psi \, d\mu$$

for all simple functions $\varphi$ and $\psi$. Show that $f$ is integrable if and only if it is measurable with respect to the completion of $\mu$.

**19.** Let $f$ be an integrable function on the measure space $(X, \mathcal{B}, \mu)$. Show that given $\epsilon > 0$, there is a $\delta > 0$ such that for each measurable set $E$ with

$\mu E < \delta$ we have

$$\left|\int_E f\right| < \epsilon.$$

**20.** Show that almost everywhere convergence can be replaced by convergence in measure in the convergence theorems (see Problem 13).

**21. a.** Show that, if $f$ is integrable, then the set $\{x: f(x) \neq 0\}$ is of $\sigma$-finite measure.

**b.** Show that if $f$ is integrable, $f \geq 0$, then $f = \lim \varphi_n$ for some increasing sequence of simple functions each of which vanishes outside a set of finite measure.

**c.** Show that, if $f$ is integrable with respect to $\mu$, then given $\epsilon > 0$ there is a simple function $\varphi$ such that

$$\int |f - \varphi|\, d\mu < \epsilon.$$

**22. a.** Let $(X, \mathcal{B}, \mu)$ be a measure space and $g$ a nonnegative measurable function on $X$. Set $\nu E = \int_E g\, d\mu$. Show that $\nu$ is a measure on $\mathcal{B}$.

**b.** Let $f$ be a nonnegative measurable function on $X$. Then

$$\int f\, d\nu = \int fg\, d\mu.$$

[Hint: First establish this for the case when $f$ is simple and then use the Monotone Convergence Theorem.]

**23.** A function $f$ on a measure space $(X, \mathcal{B}, \mu)$ is called locally measurable if the restriction of $f$ to each $E$ in $\mathcal{B}$ with $\mu E < \infty$ is measurable, that is, if $f\chi_E$ is measurable.

**a.** Show that $f$ is locally measurable if and only if it is measurable with respect to the $\sigma$-algebra of locally measurable sets.

**b.** Let $\mu$ be a $\sigma$-finite measure. Define integration for nonnegative locally measurable functions $f$ by taking $\int f$ to be the supremum of $\int \varphi$ as $\varphi$ ranges over all simple functions less than $f$, then $\int f = \int f\, d\underline{\mu}$, where $\underline{\mu}$ is the extension of $\mu$ given in Problem 8d.

## 4 General Convergence Theorems

In the preceding section we discussed the behavior of the integrals of a convergent sequence of functions. These were all integrals with respect to a fixed measure $\mu$. We can generalize by allowing the measure to vary also. Let $(X, \mathcal{B})$ be a measurable space and $\langle \mu_n \rangle$ a sequence of set functions defined on $\mathcal{B}$. We say that $\mu_n$ *converges setwise* to the set function $\mu$ if for each $E \in \mathcal{B}$ we have $\mu E = \lim \mu_n E$. With this notion we have the following two propositions, which generalize Fatou's Lemma and the Lebesgue Convergence Theorem.

**17. Proposition:** *Let $(X, \mathcal{B})$ be a measurable space, $\langle \mu_n \rangle$ a sequence of measures that converge setwise to a measure $\mu$, and $\langle f_n \rangle$ a sequence of nonnegative measurable functions that converge pointwise to the function $f$. Then*

$$\int f\,d\mu \le \varliminf \int f_n\,d\mu_n.$$

***Proof:*** Setwise convergence of $\mu_n$ to $\mu$ implies that

$$\int \varphi\,d\mu = \lim \int \varphi\,d\mu_n$$

for any simple function $\varphi$. From the definition of $\int f\,d\mu$ it suffices to prove that $\int \varphi\,d\mu \le \varliminf \int f_n\,d\mu_n$ for any simple function $\varphi \le f$.

Suppose $\int \varphi\,d\mu < \infty$. Then $\varphi$ vanishes outside a set $E$ of finite measure. Let $\epsilon$ be a positive number, and set

$$E_n = \{x : f_k(x) \ge (1 - \epsilon)\varphi(x) \text{ for all } k \ge n\}.$$

Then $\langle E_n \rangle$ is an increasing sequence of sets whose union contains $E$, and so $\langle E \sim E_n \rangle$ is a decreasing sequence of measurable sets whose intersection is empty. Thus by Proposition 2 there is an $m$ such that $\mu(E \sim E_m) < \epsilon$. Since $\mu(E \sim E_m) = \lim \mu_k(E \sim E_m)$, we may choose $n \ge m$ so that $\mu_k(E \sim E_m) < \epsilon$ for $k \ge n$. Since $E \sim E_k \subset E \sim E_m$, we have $\mu_k(E \sim E_k) < \epsilon$ for $k \ge n$. Thus

$$\begin{aligned}\int f_k\,d\mu_k \ge \int_{E_k} f_k\,d\mu_k &\ge (1 - \epsilon) \int_{E_k} \varphi\,d\mu_k \\ &\ge (1 - \epsilon) \int_E \varphi\,d\mu_k - \int_{E \sim E_k} \varphi\,d\mu_k \\ &\ge (1 - \epsilon) \int_E \varphi\,d\mu_k - M\epsilon,\end{aligned}$$

where $M$ is the maximum of $\varphi$. Thus

$$\varliminf \int f_k\,d\mu_k \ge \int_E \varphi\,d\mu - \epsilon\left[M + \int \varphi\,d\mu\right].$$

Since $\epsilon$ was arbitrary,

$$\int_E \varphi \, d\mu \leq \varliminf \int f_k \, d\mu_k .$$

The case when $\int \varphi \, d\mu = \infty$ is handled similarly. ▮

**18. Proposition:** *Let $(X, \mathcal{B})$ be a measurable space and $\langle \mu_n \rangle$ a sequence of measures on $\mathcal{B}$ that converge setwise to a measure $\mu$. Let $\langle f_n \rangle$ and $\langle g_n \rangle$ be two sequences of measurable functions that converge pointwise to $f$ and $g$. Suppose that $|f_n| \leq g_n$ and that*

$$\lim \int g_n \, d\mu_n = \int g \, d\mu < \infty.$$

*Then*

$$\lim \int f_n \, d\mu_n = \int f \, d\mu.$$

***Proof:*** Apply Proposition 17 to sequences $g_n + f_n$ and $g_n - f_n$. ▮

## Problems

**24.** Let $(X, \mathcal{B})$ be a measurable space and $\langle \mu_n \rangle$ a sequence of measures on $\mathcal{B}$ such that for each $E \in \mathcal{B}$, $\mu_{n+1} E \geq \mu_n E$. Let $\mu E = \lim \mu_n E$. Then $\mu$ is a measure on $\mathcal{B}$.

**25.** Give an example of a decreasing sequence $\langle \mu_n \rangle$ of measures on a measurable space such that the set function $\mu$ defined by $\mu E = \lim \mu_n E$ is not a measure.

**26.** Let $(X, \mathcal{B})$ be a measurable space and $\langle \mu_n \rangle$ a sequence of measures on $\mathcal{B}$ that converge setwise to a set function $\mu$. If $\mu X < \infty$, then $\mu$ is a measure.

## 5 Signed Measures

In this section we consider some of the possibilities that may arise if a measure is allowed to take on both positive and negative values. We first note that if $\mu_1$ and $\mu_2$ are two measures defined on the same measurable space $(X, \mathcal{B})$, then we may define a new measure $\mu_3$ on

$(X, \mathcal{B})$ by setting

$$\mu_3(E) = c_1\mu_1(E) + c_2\mu_2(E) \qquad c_1, c_2 \geq 0.$$

What happens if we try to define a measure by

$$\nu E = \mu_1 E - \mu_2 E?$$

The first thing that may occur is that $\nu$ is not always nonnegative, and this leads to the consideration of signed measures, which we shall define later. A more serious difficulty comes from the fact that $\nu$ is not defined when $\mu_1 E = \mu_2 E = \infty$. For this reason we should have either $\mu_1$ or $\mu_2$ finite. With these considerations in mind we make the following definition:

**Definition:** *By a signed measure on the measurable space $(X, \mathcal{B})$ we mean an extended real-valued set function $\nu$ defined for the sets of $\mathcal{B}$ and satisfying the following conditions:*

i. *$\nu$ assumes at most one of the values $+\infty$, $-\infty$.*
ii. $\nu(\emptyset) = 0$.
iii. $\nu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \nu E_i$ *for any sequence $E_i$ of disjoint measurable sets, the equality taken to mean that the series on the right converges absolutely if $\nu(\bigcup E_i)$ is finite and that it properly diverges otherwise.*

Thus a measure is a special case of a signed measure, but a signed measure is not in general a measure. We say that a set $A$ is a **positive set** with respect to a signed measure $\nu$ if $A$ is measurable and for every measurable subset $E$ of $A$ we have $\nu E \geq 0$. Every measurable subset of a positive set is again positive, and if we take the restriction of $\nu$ to a positive set we obtain a measure. Similarly, a set $B$ is called a negative set if it is measurable and every measurable subset of it has nonpositive $\nu$ measure. A set that is both positive and negative with respect to $\nu$ is called a **null set.** A measurable set is a null set if and only if every measurable subset of it has $\nu$ measure zero. The reader should carefully note the distinction between a null set and a set of measure zero: While every null set must have measure zero, a set of measure zero may well be a union of two sets whose measures are not zero but are negatives of each other. Similarly, a positive set is not to be confused with a set that merely has positive measure. We have the following lemmas concerning positive sets. Similar statements hold, of course, for negative sets.

**19. Lemma:** *Every measurable subset of a positive set is itself positive. The union of a countable collection of positive sets is positive.*

***Proof:*** The first statement is trivially true by the definition of a positive set. To prove the second statement, let $A$ be the union of a sequence $\langle A_n \rangle$ of positive sets. If $E$ is any measurable subset of $A$, set

$$E_n = E \cap A_n \cap \tilde{A}_{n-1} \cap \cdots \cap \tilde{A}_1.$$

Then $E_n$ is a measurable subset of $A_n$ and so $\nu E_n \geq 0$. Since the $E_n$ are disjoint and $E = \bigcup E_n$, we have

$$\nu E = \sum_{n=1}^{\infty} \nu E_n \geq 0.$$

Thus $A$ is a positive set. ▮

**20. Lemma:** *Let $E$ be a measurable set such that $0 < \nu E < \infty$. Then there is a positive set $A$ contained in $E$ with $\nu A > 0$.*

***Proof:*** Either $E$ itself is a positive set or it contains sets of negative measure. In the latter case let $n_1$ be the smallest positive integer such that there is a measurable set $E_1 \subset E$ with

$$\nu E_1 < -\frac{1}{n_1}.$$

Proceeding inductively, if $E \sim \bigcup_{j=1}^{k-1} E_k$ is not already a positive set, let $n_k$ be the smallest positive integer for which there is a measurable set $E_k$ such that

$$E_k \subset E \sim \left[ \bigcup_{j=1}^{k-1} E_j \right]$$

and

$$\nu E_k < -\frac{1}{n_k}.$$

If we set

$$A = E \sim \bigcup_{k=1}^{\infty} E_k,$$

then

$$E = A \cup \left[ \bigcup_{k=1}^{\infty} E_k \right].$$

Since this is a disjoint union, we have

$$\nu E = \nu A + \sum_{k=1}^{\infty} \nu E_k$$

with the series on the right absolutely convergent, since $\nu E$ is finite. Thus $\sum 1/n_k$ converges, and we have $n_k \to \infty$. Since $\nu E_k \leq 0$ and $\nu E > 0$, we must have $\nu A > 0$.

To show that $A$ is a positive set, let $\epsilon > 0$ be given. Since $n_k \to \infty$, we may choose $k$ so large that $(n_k - 1)^{-1} < \epsilon$. Since

$$A \subset E \sim \left[ \bigcup_{j=1}^{k} E_j \right],$$

$A$ can contain no measurable sets with measure less than $-(n_k - 1)^{-1}$, which is greater than $-\epsilon$. Thus $A$ contains no measurable sets of measure less than $-\epsilon$. Since $\epsilon$ is an arbitrary positive number, it follows that $A$ can contain no sets of negative measure and so must be a positive set. ∎

**21. Proposition (Hahn Decomposition Theorem):** *Let $\nu$ be a signed measure on the measurable space $(X, \mathfrak{B})$. Then there is a positive set $A$ and a negative set $B$ such that $X = A \cup B$ and $A \cap B = \emptyset$.*

***Proof:*** Without loss of generality we may assume that $+\infty$ is the infinite value omitted by $\nu$. Let $\lambda$ be the supremum of $\nu A$ over all sets $A$ that are positive with respect to $\nu$. Since the empty set is positive, $\lambda \geq 0$. Let $\langle A_i \rangle$ be a sequence of positive sets such that

$$\lambda = \lim_{i \to \infty} \nu A_i,$$

and set

$$A = \bigcup_{i=1}^{\infty} A_i.$$

By Lemma 19 the set $A$ is itself a positive set, and so $\lambda \geq \nu A$. But $A \sim A_i \subset A$ and so $\nu(A \sim A_i) \geq 0$. Thus

$$\nu A = \nu A_i + \nu(A \sim A_i) \geq \nu A_i.$$

Hence $\nu A \geq \lambda$, and so $\nu A = \lambda$, and $\lambda < \infty$.

Let $B = \sim A$, and suppose that $E$ is a positive subset of $B$. Then $E$ and $A$ are disjoint and $E \cup A$ is a positive set. Hence

$$\lambda \geq \nu(E \cup A) = \nu E + \nu A = \nu E + \lambda,$$

whence $\nu E = 0$, since $0 \leq \lambda < \infty$. Thus $B$ contains no positive subsets of positive measure and hence no subsets of positive measure by Lemma 20. Consequently, $B$ is a negative set. ∎

A decomposition of $X$ into two disjoint sets $A$ and $B$ such that $A$ is positive for $\nu$ and $B$ negative is called a **Hahn decomposition** for $\nu$. Proposition 21 states the existence of a Hahn decomposition for each signed measure. Unfortunately, a Hahn decomposition is not unique.

If $\{A, B\}$ is a Hahn decomposition for $\nu$, then we may define two measures $\nu^+$ and $\nu^-$ with $\nu = \nu^+ - \nu^-$ by setting

$$\nu^+(E) = \nu(E \cap A)$$

and

$$\nu^-(E) = -\nu(E \cap B).$$

Two measures $\nu_1$ and $\nu_2$ on $(X, \mathcal{B})$ are said to be mutually singular (in symbols $\nu_1 \perp \nu_2$) if there are disjoint measurable sets $A$ and $B$ with $X = A \cup B$ such that $\nu_1(A) = \nu_2(B) = 0$. Thus the measures $\nu^+$ and $\nu^-$ defined above are mutually singular. We have thus established the existence part of the following proposition. The uniqueness part is left to the reader.

**22. Proposition:** *Let $\nu$ be a signed measure on the measurable space $(X, \mathcal{B})$. Then there are two mutually singular measures $\nu^+$ and $\nu^-$ on $(X, \mathcal{B})$ such that $\nu = \nu^+ - \nu^-$. Moreover, there is only one such pair of mutually singular measures.*

The decomposition of $\nu$ given by the proposition is called the **Jordan decomposition** of $\nu$. The measures $\nu^+$ and $\nu^-$ are called the positive and negative parts (or variations) of $\nu$. Since $\nu$ assumes at most one of the values $+\infty$ and $-\infty$, either $\nu^+$ or $\nu^-$ must be finite. If they are both finite, we call $\nu$ a finite signed measure.

The measure $|\nu|$ defined by

$$|\nu|(E) = \nu^+E + \nu^-E$$

is called the **absolute value** or **total variation** of $\nu$. A set $E$ is positive for $\nu$ if $\nu^- E = 0$. It is a null set if $|\nu|(E) = 0$.

## Problems

**27. a.** Give an example to show that the Hahn decomposition need not be unique.

**b.** Show that the Hahn decomposition is unique except for null sets.

**28.** Show that there is only one pair of mutually singular measures $\nu^+$ and $\nu^-$ such that $\nu = \nu^+ - \nu^-$. [Hint: Show that any such pair determines a Hahn decomposition and apply the results of Problem 27b.]

**29.** Show that if $E$ is any measurable set, then

$$-\nu^- E \leq \nu E \leq \nu^+ E$$

and

$$|\nu E| \leq |\nu|(E).$$

**30.** Show that if $\nu_1$ and $\nu_2$ are any two finite signed measures, then so is $\alpha\nu_1 + \beta\nu_2$, where $\alpha$ and $\beta$ are real numbers. Show that

$$|\alpha\nu| = |\alpha|\,|\nu|$$

and

$$|\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|,$$

where $\nu \leq \mu$ means $\nu E \leq \mu E$ for all measurable sets $E$.

**31.** We define integration with respect to a signed measure $\nu$ by defining

$$\int f\,d\nu = \int f\,d\nu^+ - \int f\,d\nu^-.$$

If $|f| \leq M$,

$$\left|\int_E f\,d\nu\right| \leq M|\nu|(E).$$

Moreover, there is a measurable function $f$ with $|f| \leq 1$ such that

$$\int_E f\,d\nu = |\nu|(E).$$

**32. a.** Let $\mu$ and $\nu$ be finite signed measures. Show that there is a signed measure $\mu \wedge \nu$ which is smaller than $\mu$ and $\nu$ but larger than any other signed measure which is smaller than $\mu$ and $\nu$. [Hint: Use the fact that $\mu \wedge \nu = \frac{1}{2}(\mu + \nu - |\mu - \nu|)$.]

**b.** Show that there is a measure $\mu \vee \nu$ which is larger than $\mu$ and $\nu$ but smaller than any other measure which is larger than $\mu$ and $\nu$. Also, $\mu \vee \nu + \mu \wedge \nu = \mu + \nu$.

**c.** If $\mu$ and $\nu$ are positive measures, then they are mutually singular if and only if $\mu \wedge \nu = 0$.

## 6 The Radon–Nikodym Theorem

Let $(X, \mathcal{B})$ be a fixed measurable space. If $\mu$ and $\nu$ are two measures defined on $(X, \mathcal{B})$, we say that $\mu$ and $\nu$ are **mutually singular** (and write $\mu \perp \nu$) if there are disjoint sets $A$ and $B$ in $\mathcal{B}$ such that $X = A \cup B$ and $\nu A = \mu B = 0$. Despite the fact that the notion of singularity is symmetric in $\nu$ and $\mu$, we sometimes say that $\nu$ is singular with respect to $\mu$. The notion antithetical to singularity is absolute continuity. A measure $\nu$ is said to be **absolutely continuous** with respect to the measure $\mu$ if $\nu A = 0$ for each set $A$ for which $\mu A = 0$. We use the symbolism $\nu \ll \mu$ for $\nu$ absolutely continuous with respect to $\mu$.

In the case of signed measures $\mu$ and $\nu$, we say $\nu \ll \mu$ if $|\nu| \ll |\mu|$ and $\nu \perp \mu$ if $|\nu| \perp |\mu|$.

Whenever we are dealing with more than one measure on a measurable space $(X, \mathcal{B})$, the term 'almost everywhere' becomes ambiguous, and we must specify almost everywhere with respect to $\mu$ or almost everywhere with respect to $\nu$, etc. These are usually abbreviated a.e. $[\mu]$ and a.e. $[\nu]$. If $\nu \ll \mu$ and a property holds a.e. $[\mu]$, then it holds a.e. $[\nu]$.

Let $\mu$ be a measure and $f$ a nonnegative measurable function on $X$. For $E$ in $\mathcal{B}$, set

$$\nu E = \int_E f \, d\mu.$$

Then $\nu$ is a set function defined on $\mathcal{B}$, and it follows from Corollary 14 that $\nu$ is countably additive and hence a measure. The measure $\nu$ will be finite if and only if $f$ is integrable. Since the integral over a set of $\mu$-measure zero is zero, we have $\nu$ absolutely continuous with respect to $\mu$. The next theorem shows that, subject to $\sigma$-finiteness restrictions, every absolutely continuous measure $\nu$ is obtained in this fashion.

**23. Theorem (Radon–Nikodym):** *Let $(X, \mathcal{B}, \mu)$ be a $\sigma$-finite measure space, and let $\nu$ be a measure defined on $\mathcal{B}$ which is absolutely* *continuous with respect to $\mu$. Then there is a nonnegative measurable function $f$ such that for each set $E$ in $\mathcal{B}$ we have*

$$\nu E = \int_E f\,d\mu.$$

*The function $f$ is unique in the sense that if $g$ is any measurable function with this property then $g = f$ a.e. $[\mu]$.*

***Proof:*** The extension from the finite to the $\sigma$-finite case is not difficult and is left to the reader. Thus we shall assume that $\mu$ is finite. Then $\nu - \alpha\mu$ is a signed measure for each rational number $\alpha$. Let $(A_\alpha, B_\alpha)$ be a Hahn decomposition for $\nu - \alpha\mu$, and take $A_0 = X$, $B_0 = \emptyset$.

Now $B_\alpha \sim B_\beta = B_\alpha \cap A_\beta$. Thus $(\nu - \alpha\mu)(B_\alpha \sim B_\beta) \leq 0$, and hence $(\nu - \beta\mu)(B_\alpha \sim B_\beta) \geq 0$. If $\beta > \alpha$, these imply $\mu(B_\alpha \sim B_\beta) = 0$, and so by Proposition 10 there is a measurable function $f$ such that for each rational $\alpha$ we have $f \geq \alpha$ a.e. on $A_\alpha$ and $f \leq \alpha$ a.e. on $B_\alpha$. Since $B_0 = \emptyset$, we may take $f$ to be nonnegative.

Let $E$ be an arbitrary set in $\mathcal{B}$, and set

$$E_k = E \cap (B_{(k+1)/N} \sim B_{k/N}), \qquad E_\infty = E \sim \bigcup B_{k/N}.$$

Then $E = E_\infty \cup \bigcup_{k=0}^{\infty} E_k$, and this union is disjoint modulo null sets. Thus

$$\nu E = \nu E_\infty + \sum_{k=0}^{\infty} \nu E_k.$$

Since $E_k \subset B_{(k+1)/N} \cap A_{k/N}$, we have $\frac{k}{N} \leq f \leq \frac{k+1}{N}$ on $E_k$, and so

$$\frac{k}{N}\,\mu E_k \leq \int_{E_k} f\,d\mu \leq \frac{k+1}{N}\,\mu E_k.$$

Since $\frac{k}{N}\,\mu E_k \leq \nu E_k \leq \frac{k+1}{N}\,\mu E_k$, we have

$$\nu E_k - \frac{1}{N}\,\mu E_k \leq \int_{E_k} f\,d\mu \leq \nu E_k + \frac{1}{N}\,\mu E_k.$$

On $E_\infty$ we have $f = \infty$ a.e. If $\mu E_\infty > 0$, we must have $\nu E_\infty = \infty$, since $(\nu - \alpha\mu)\,E_\infty$ is positive for each $\alpha$. If $\mu E_\infty = 0$, we have $\nu E_\infty = 0$, since $\nu \ll \mu$. In either case

$$\nu E_\infty = \int_{E_\infty} f\, d\mu.$$

Adding this equality and our previous inequalities gives

$$\nu E - \frac{1}{N}\mu E \le \int_E f\, d\mu \le \nu E + \frac{1}{N}\mu E.$$

Since $\mu E$ is finite and $N$ arbitrary, we must have

$$\nu E = \int_E f\, d\mu. \quad \blacksquare$$

The function $f$ given by Theorem 23 is called the **Radon–Nikodym derivative** of $\nu$ with respect to $\mu$. It is sometimes denoted by $\left[\dfrac{d\nu}{d\mu}\right]$.

**24. Proposition (Lebesgue Decomposition):** *Let $(X, \mathcal{B}, \mu)$ be a $\sigma$-finite measure space and $\nu$ a $\sigma$-finite measure defined on $\mathcal{B}$. Then we can find a measure $\nu_0$, singular with respect to $\mu$, and a measure $\nu_1$, absolutely continuous with respect to $\mu$, such that $\nu = \nu_0 + \nu_1$. The measures $\nu_0$ and $\nu_1$ are unique.*

***Proof:*** Since $\mu$ and $\nu$ are $\sigma$-finite measures, so is the measure $\lambda = \mu + \nu$. Since both $\mu$ and $\nu$ are absolutely continuous with respect to $\lambda$, the Radon–Nikodym Theorem asserts the existence of nonnegative measurable functions $f$ and $g$ such that for each $E \in \mathcal{B}$

$$\mu E = \int_E f\, d\lambda, \qquad \nu E = \int_E g\, d\lambda.$$

Let $A = \{x : f(x) > 0\}$ and $B = \{x : f(x) = 0\}$. Then $X$ is the disjoint union of $A$ and $B$, $\mu B = 0$. If we define $\nu_0$ by

$$\nu_0 E = \nu(E \cap B),$$

we have $\nu_0(A) = 0$ and so $\nu_0 \perp \mu$. Let

$$\nu_1(E) = \nu(E \cap A) = \int_{E \cap A} g\, d\lambda.$$

Then $\nu = \nu_0 + \nu_1$, and we have only to show that $\nu_1 \ll \mu$. Let $E$ be a set of $\mu$-measure zero. Then

$$0 = \mu E = \int_E f\, d\lambda,$$

and $f = 0$ a.e. $[\lambda]$ on $E$. Since $f > 0$ on $A \cap E$, we must have $\lambda(A \cap E) = 0$. Hence $\nu(A \cap E) = 0$, and so $\nu_1(E) = \nu(A \cap E) = 0$. This establishes the proposition except for the uniqueness, which is left to the reader. ▮

## Problems

**33. a.** Show that the Radon–Nikodym Theorem for a finite measure $\mu$ implies the theorem for a $\sigma$-finite measure $\mu$. [Hint: Decompose $X$ into a countable union of sets $X_i$ of finite $\mu$-measure and apply the Radon–Nikodym Theorem to each $X_i$ to obtain $f$. Show $f$ to have the required properties.]

**b.** Show the uniqueness of the function $f$ in the Radon–Nikodym Theorem.

**34.** *Radon–Nikodym derivatives.* Let $\mu$, $\nu$, and $\lambda$ be $\sigma$-finite. Show that the Radon–Nikodym derivative $[d\nu/d\mu]$ has the following properties:

**a.** If $\nu \ll \mu$ and $f$ is a nonnegative measurable function, then

$$\int f\,d\nu = \int f\left[\frac{d\nu}{d\mu}\right] d\mu.$$

**b.**
$$\left[\frac{d(\nu_1 + \nu_2)}{d\mu}\right] = \left[\frac{d\nu_1}{d\mu}\right] + \left[\frac{d\nu_2}{d\mu}\right].$$

**c.** If $\nu \ll \mu \ll \lambda$, then

$$\left[\frac{d\nu}{d\lambda}\right] = \left[\frac{d\nu}{d\mu}\right]\left[\frac{d\mu}{d\lambda}\right].$$

**d.** If $\nu \ll \mu$ and $\mu \ll \nu$, then

$$\left[\frac{d\nu}{d\mu}\right] = \left[\frac{d\mu}{d\nu}\right]^{-1}.$$

**35. a.** Show that if $\nu$ is a signed measure such that $\nu \perp \mu$ and $\nu \ll \mu$, then $\nu = 0$.

**b.** Show that if $\nu_1$ and $\nu_2$ are singular with respect to $\mu$, then so is $c_1\nu_1 + c_2\nu_2$.

**c.** Show that if $\nu_1$ and $\nu_2$ are absolutely continuous with respect to $\mu$, so is $c_1\nu_1 + c_2\nu_2$.

**d.** Prove the uniqueness assertion in the Lebesgue decomposition.

**36.** Extend the Radon–Nikodym Theorem to the case of signed measures.

**37.** *Complex measures.* A set function $\nu$ that assigns a complex number $\nu E$ to each $E$ in a $\sigma$-algebra $\mathcal{B}$ is called a complex measure if $\nu\emptyset = 0$ and for each countable disjoint union $\bigcup E_i$ of sets in $\mathcal{B}$ we have

$$\nu(\bigcup E_i) = \sum_{i=1}^{\infty} \nu E_i$$

with absolute convergence on the right.

**a.** Show that each complex measure $\nu$ may be expressed as $\nu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$, where $\mu_1$, $\mu_2$, $\mu_3$, and $\mu_4$ are finite measures.

**b.** Show that for each complex measure $\nu$ there is a measure $\mu$ and a complex-valued measurable function $\varphi$ with $|\varphi| = 1$ such that for each set $E$ in $\mathcal{B}$,

$$\nu E = \int_E \varphi \, d\mu.$$

[Hint: Apply the Radon–Nikodym Theorem to the measures $\mu_i$ with respect to the measure $\mu_1 + \mu_2 + \mu_3 + \mu_4$.]

**c.** Show that the measure $\mu$ in (b) is unique and that $\varphi$ is uniquely determined to within sets of $\mu$ measure zero.

**d.** The measure $\mu$ in (b) is called the *total variation* or *absolute value* of $\nu$ and is sometimes denoted by $|\nu|$. Show that the results of Problem 31 hold for complex measures.

**e.** If $\nu$ is a complex measure with $|\nu|(X) = 1$ and $\nu(X) = 1$, then $\nu$ is a positive real measure.

**38.** *Alternate proof of the Radon–Nikodym Theorem.* We can give a proof of the Radon–Nikodym Theorem that is independent of the Hahn Decomposition Theorem by using Proposition 10.28, which states that for each bounded linear functional $F$ on a Hilbert space $H$ there is a $g$ in $H$ such that $F(f) = (f, g)$ for all $f$ in $H$. This proof is due to von Neumann. The details are outlined below:

**a.** Let $\mu$ and $\nu$ be finite measures on a measurable space $(X, \mathcal{B})$ and set $\lambda = \mu + \nu$. Define $F(f) = \int f \, d\mu$. Then $F$ is a bounded linear functional on $L^2(\lambda)$.

**b.** The function $g \in L^2(\lambda)$ such that $F(f) = (f, g)$ has the property that $0 \le g \le 1$, and

$$\mu(E) = \int_E g \, d\lambda$$

$$\nu(E) = \int_E (1 - g) \, d\lambda.$$

**c.** If $\nu \ll \mu$, then $\lambda \ll \mu$, and $g = 0$ only on a set of $\mu$-measure zero. In this case

$$\lambda(E) = \int_E g^{-1}\, d\mu.$$

[Hint: Consider Problem 22.]

**d.** If $\nu \ll \mu$, then $(1 - g)g^{-1}$ is integrable with respect to $\mu$ and

$$\nu(E) = \int_E (1 - g)g^{-1}\, d\mu.$$

**39.** Use the following example to show that the hypothesis in the Radon–Nikodym Theorem that $\mu$ is $\sigma$-finite cannot be omitted. Let $X = [0, 1]$, $\mathcal{B}$ the class of Lebesgue measurable subsets of $[0, 1]$, and take $\nu$ to be Lebesgue measure and $\mu$ to be the counting measure on $\mathcal{B}$. Then $\nu$ is finite and absolutely continuous with respect to $\mu$, but there is no function $f$ such that $\nu E = \int_E f\, d\mu$ for all $E \in \mathcal{B}$. At what point does the proof of Theorem 23 break down for this example?

**40.** *Decomposable measures.* Let $(X, \mathcal{B}, \mu)$ be a measure space. A collection $\{X_\alpha\}$ of disjoint measurable subsets of $X$ is called a **decomposition** for $\mu$ if $\mu X_\alpha < \infty$ for each $\alpha$ and if $\mu E = 0$ for every measurable set $E$ such that $\mu(E \cap X_\alpha) = 0$ for all $\alpha$. A measure $\mu$ is called **decomposable** if it has a decomposition.

**a.** If $\{X_\alpha\}$ is a decomposition for $\mu$, and if $E$ is any measurable set, then $\mu E = \sum \mu(X_\alpha \cap E)$, where $\sum$ is meant in the sense of Problem 2.21.

**b.** If $\{X_\alpha\}$ is a decomposition for a complete measure $\mu$, then $f$ is locally measurable if and only if the restriction of $f$ to $X_\alpha$ is measurable for each $\alpha$. If $f$ is a nonnegative locally measurable function of $X$, then

$$\int_X f\, d\mu = \sum_\alpha \int_{X_\alpha} f\, d\mu.$$

**c.** Let $\nu$ be absolutely continuous with respect to $\mu$, and suppose that there is a collection $\{X_\alpha\}$ which is a decomposition for both $\mu$ and $\nu$. Then there is a nonnegative, locally measurable real-valued function $f$ such that for every measurable set $E$ we have

$$\nu E = \int_E f\, d\mu.$$

**d.** The conclusion of (c) is still valid if instead of assuming $\{X_\alpha\}$ to be a decomposition for $\nu$, we merely assume that if $E \in \mathcal{B}$ and $\nu(E \cap X_\alpha) = 0$ for all $\alpha$, then $\nu E = 0$.

## 7 The $L^p$ Spaces

If $(X, \mathcal{B}, \mu)$ is a complete measure space, we denote by $L^p(\mu)$ the space of all measurable functions on $X$ for which $\int |f|^p \, d\mu < \infty$, considering two functions in $L^p$ to be equivalent if they are equal almost everywhere. As in Chapter 6, we define $L^\infty(\mu)$ to be the space of bounded measurable functions. For $1 \le p < \infty$ we set

$$\|f\|_p = \left\{\int |f|^p \, d\mu\right\}^{1/p},$$

and for $p = \infty$ we set

$$\|f\|_\infty = \text{ess sup } |f|.$$

Note that the space $L^\infty(\mu)$ depends on the choice of $\mu$ to determine the norm and the classes of equivalent functions but that this only requires knowing what the sets of measure zero are.

The Hölder and Minkowski inequalities and the Riesz–Fischer Theorem follow just as in Chapter 6, and we summarize them in the following theorem:

**25. Theorem:** *For $1 \le p \le \infty$ the spaces $L^p(\mu)$ are Banach spaces, and if $f \in L^p(\mu)$, $g \in L^q(\mu)$, with $1/p + 1/q = 1$, then $fg \in L^1(\mu)$ and*

$$\int |fg| \, d\mu \le \|f\|_p \|g\|_q.$$

The following proposition, whose proof is left to the reader, is a version of Littlewood's second principle:

**26. Proposition:** *Let $f \in L^p(\mu)$, $1 \le p < \infty$. Then, given $\epsilon > 0$, there is a simple function $\varphi$ vanishing outside a set of finite measure such that $\|f - \varphi\|_p < \epsilon$.*

It follows from the Hölder inequality that each $g \in L^q$ defines a linear functional $F$ on $L^p$ by setting

$$F(f) = \int fg \, d\mu,$$

and it is not difficult to show that $\|F\| = \|g\|$. Our goal in the remainder of this section is to establish the converse (Riesz Representation Theorem), which states that each linear functional on $L^p(\mu)$ for $1 < p < \infty$ is of this form, and if $\mu$ is $\sigma$-finite, each linear functional on $L^1(\mu)$ is of this form. We begin by establishing a lemma.

**27. Lemma:** *Let $(X, \mathcal{B}, \mu)$ be a finite measure space and $g$ an integrable function such that for some constant $M$,*

$$\left| \int g\varphi \, d\mu \right| \leq M \|\varphi\|_p$$

*for all simple functions $\varphi$. Then $g \in L^q$.*

***Proof:*** Assume $p > 1$, and let $\langle \psi_n \rangle$ be a sequence of nonnegative simple functions which increase to $|g|^q$. Set

$$\varphi_n = (\psi_n)^{1/p} \operatorname{sgn} g.$$

Then $\varphi_n$ is a simple function, and

$$\|\varphi_n\|_p = \left\{ \int \psi_n \, d\mu \right\}^{1/p}.$$

Since $\varphi_n g \geq |\varphi_n| \, |\psi_n|^{1/q} = |\psi_n|^{1/p + 1/q} = \psi_n$, we have

$$\begin{aligned} \int \psi_n \, d\mu &\leq \int \varphi_n g \, d\mu \\ &\leq M \|\varphi_n\|_p \\ &\leq M \left\{ \int \psi_n \, d\mu \right\}^{1/p}. \end{aligned}$$

Since $1 - \dfrac{1}{p} = \dfrac{1}{q}$,

$$\left\{ \int \psi_n \, d\mu \right\}^{1/q} \leq M$$

or

$$\int \psi_n \, d\mu \leq M^q,$$

and by the Monotone Convergence Theorem

$$\int |g|^q \, d\mu \leq M^q.$$

The case $p = 1$ is left to the reader. ▮

We shall also use the following lemma, whose proof is left to the reader.

**28. Lemma:** *Let $\langle E_n \rangle$ be a sequence of disjoint measurable sets, and for each $n$ let $f_n$ be a function in $L^p(1 \le p < \infty)$ that vanishes outside $E_n$. Set $f = \sum_{n=1}^{\infty} f_n$. Then $f \in L^p$ if and only if $\sum \|f_n\|^p < \infty$. In this case $f = \sum f_n$ in $L^p$; that is,*

$$\|f - \sum_{i=1}^{n} f_i\|_p \to 0$$

*and*

$$\|f\|^p = \sum_{n=1}^{\infty} \|f_n\|^p.$$

**29. Riesz Representation Theorem:** *Let $F$ be a bounded linear functional on $L^p(\mu)$ with $1 \le p < \infty$ and $\mu$ a $\sigma$-finite measure. Then there is a unique element $g$ in $L^q$, where $1/q + 1/p = 1$, such that*

$$F(f) = \int fg \, d\mu.$$

*We have also $\|F\| = \|g\|_q$.*

***Proof:*** Let us first consider the case of a finite measure $\mu$. Then every bounded measurable function is in $L^p(\mu)$. Define a set function $\nu$ on the measurable sets by setting

$$\nu E = F(\chi_E).$$

If $E$ is the union of a sequence $\langle E_n \rangle$ of disjoint measurable sets, let $\alpha_n = \operatorname{sgn} F(\chi_{E_n})$, and set $f = \sum \alpha_n \chi_{E_n}$. Then by Lemma 28 and the boundedness of $F$ we have

$$\sum_{n=1}^{\infty} |\nu E_n| = F(f) < \infty$$

and

$$\sum_{n=1}^{\infty} \nu E_n = F(\chi_E) = \nu E.$$

Thus $\nu$ is a signed measure and it follows from its definition that it is absolutely continuous with respect to $\mu$. By the Radon–Nikodym Theorem there is a measurable function $g$ such that for each measurable set $E$ we have $\nu E = \int_E g \, d\mu$. Since $\nu$ is always finite, $g$ is integrable.

If $\varphi$ is a simple function, the linearity of $F$ and of the integral imply that

$$F(\varphi) = \int \varphi g \, d\mu.$$

Since the left side is bounded by $\|F\| \|\varphi\|_p$, we have $g \in L^q$ by Lemma 27. Let $G$ be the bounded linear functional defined on $L^p$ by

$$G(f) = \int fg \, d\mu.$$

Then $G - F$ is a bounded linear functional which vanishes on the subspace of simple functions. By Proposition 26 the simple functions are dense in $L^p$, and so $G - F = 0$. Thus for all $f \in L^p$,

$$F(f) = \int fg \, d\mu,$$

and it is readily verified that $\|F\| = \|G\| = \|g\|_q$.

The function $g$ must determine a unique element of $L^q$, for if $g_1$ and $g_2$ determine the same functional $F$, then $g_1 - g_2$ must give the zero functional, and so $\|g_1 - g_2\|_q = 0$. Thus $g_1 = g_2$ a.e.

To extend the theorem to the $\sigma$-finite case, let $\langle X_n \rangle$ be an increasing sequence of measurable sets of finite measure whose union is $X$. The theorem for finite measure spaces implies that for each $n$ there is a function $g_n$ in $L^q$ such that $g_n$ vanishes outside $X_n$ and

$$F(f) = \int fg_n \, d\mu$$

for all $f \in L^p$ that vanish outside $X_n$. Moreover, $\|g_n\|_q \leq \|F\|$. Since any function $g_n$ with this property is uniquely determined on $X_n$ except for changes on sets of measure zero and since $g_{n+1}$ also has this property, we may assume $g_{n+1} = g_n$ on $X_n$. For $x \in X_n$ set $g(x) = g_n(x)$. Then $g$ is a well-defined measurable function and $|g_n|$ increases pointwise to $|g|$. Thus by the Monotone Convergence

Theorem

$$\int |g|^q \, d\mu = \lim \int |g_n|^q \, d\mu$$
$$\leq \|F\|^q,$$

and $g \in L^q$.

For $f \in L^p$, let $f_n = f$ on $X_n$, and $f_n = 0$ on $\tilde{X}_n$. Then $f_n \to f$ pointwise and in $L^p$. Since $|fg|$ is integrable and $|f_n g| \leq |fg|$, the Lebesgue Convergence Theorem gives

$$\int fg \, d\mu = \lim \int f_n g \, d\mu$$
$$= \lim \int f_n g_n \, d\mu$$
$$= \lim F(f_n)$$
$$= F(f). \quad \blacksquare$$

If $p = 1$, the requirement that $\mu$ be $\sigma$-finite is necessary. Some extension and counterexamples are given in Problems 47 and 48. If $p > 1$, the $\sigma$-finiteness is not required:

**30. Theorem:** *Let $F$ be a bounded linear functional on $L^p(\mu)$ with $1 < p < \infty$. Then there is a unique element $g \in L^q$ such that*

$$F(f) = \int fg \, d\mu.$$

*We have $\|F\| = \|g\|_q$.*

***Proof:*** It follows from the preceding theorem that, if $E$ is any measurable set of $\sigma$-finite measure, then there is a unique $g_E \in L^q$, vanishing outside $E$, such that

$$F(f) = \int g_E f \, d\mu$$

for each $f \in L^p$ that vanishes outside $E$. The uniqueness of $g_E$ implies that, if $A \subset E$, then $g_A = g_E$ a.e. on $A$. For each $E$ of $\sigma$-finite measure set $\lambda(E) = \int |g_E|^q \, d\mu$. Then for $A \subset E$, we have $\lambda(A) \leq \lambda(E) \leq \|F\|^q$. Let $\langle E_n \rangle$ be a sequence of sets of $\sigma$-finite measure such that $\lambda(E_n)$ tends to the maximum value $m$ of $\lambda$. Then $H = \bigcup E_n$ is a set of $\sigma$-finite measure, and by the monotonicity of $\lambda$ we have $\lambda(H) = m$.

Let $g$ be defined to be $g_H$ on $H$ and 0 elsewhere. Then $g \in L^q$. If $E$

is any set of $\sigma$-finite measure containing $H$, then $g_E = g_H$ a.e. on $H$. But $\int |g_E|^q = \lambda(E) \le \lambda(H) = \int |g_H|^q$, and so $g_E = 0$ a.e. on $E \sim H$. Thus $g_E = g$ almost everywhere on $E$.

If $f \in L^p$, then the set $N = \{x : f(x) \neq 0\}$ is a set of $\sigma$-finite measure, and so is the set $E = N \cup H$. Thus

$$F(f) = \int fg_E \, d\mu = \int fg \, d\mu.$$

The equality of $\|F\|$ and $\|g\|_q$ follows, as in the preceding theorem. ▮

## Problems

**41.** Prove Proposition 26.

**42.** Prove Lemma 27 for the case $p = 1$.

**43.** Show that Lemma 27 remains true if we only assume $\mu$ to be $\sigma$-finite.

**44.** Prove Lemma 28.

**45.** For $g \in L^q$, let $F$ be the linear functional on $L^p$ defined by

$$F(f) = \int fg \, d\mu.$$

Show that $\|F\| = \|g\|_q$.

**46. a.** Let $\mu$ be the counting measure on a countable set $X$. Show that $L^p(\mu) = l^p$.

**b.** Discuss the spaces $l^p(X) = L^p(\mu)$, where $\mu$ is the counting measure on a not necessarily countable set $X$.

**47.** Let $(X, \mathcal{B}, \mu)$ be a decomposable measure space (cf. Problem 40).

**a.** Show that for each bounded linear functional $F$ on $L^1(\mu)$ there is a bounded locally measurable function $g$ such that $F(f) = \int fg \, d\mu$.

**b.** Show that the dual space to $L^1(\mu)$ is the space $L^\infty(\underline{\mu})$, where $\underline{\mu}$ is the saturation of $\mu$ given in Problem 8d.

**48.** Let $A$ and $B$ be uncountable sets with different numbers of elements, and let $X = A \times B$. A set of the form $\{\langle x, y\rangle : x = a\}$ is called a vertical line and a set of the form $\{\langle x, y\rangle : y = b\}$ a horizontal line. Let $\mathcal{B}$ be the collection of subsets $E$ of $X$ such that for every horizontal or vertical line $L$ either $E \cap L$ or $\tilde{E} \cap L$ is countable. Then $\mathcal{B}$ is a $\sigma$-algebra. Let $\mu E$ be the number of horizontal and vertical lines $L$ for which $\tilde{E} \cap L$ is countable and $\nu E$ the number of horizontal lines with $\tilde{E} \cap L$ countable. Then $\mu$ and $\nu$ are measures on $\mathcal{B}$, and we can define a bounded linear functional $F$ on $L^1(\mu)$ by setting $F(f) = \int f \, d\nu$. There is no locally measurable function $g$ such that $F(f) = \int fg \, d\mu$.

# 12 Measure and Outer Measure

In this chapter we first consider some of the ways in which a measure can be defined on a $\sigma$-algebra. In the case of Lebesgue measure we defined measure for open sets and used this to define outer measure, from which we obtain the notion of measurable set and Lebesgue measure. Such a procedure is feasible in general. In the first section we discuss the process of deriving a measure from an outer measure, and in the second section we derive an outer measure from a measure that is defined only on an algebra of sets. The remainder of the chapter is devoted to some applications of this process.

## 1 Outer Measure and Measurability

By an outer measure $\mu^*$ we mean a nonnegative extended real-valued set function defined on all subsets of a space $X$ and having the following properties:

i. $\mu^*\emptyset = 0$.
ii. $A \subset B \Rightarrow \mu^*A \leq \mu^*B$.
iii. $E \subset \bigcup_{i=1}^{\infty} E_i \Rightarrow \mu^*E \leq \sum_{i=1}^{\infty} \mu^*E_i$.

The second property is called monotonicity and the third countable subadditivity. In view of (i) finite subadditivity follows from (iii).

Because of (ii), property (iii) can be replaced by

$$\text{iii. } E = \bigcup_{i=1}^{\infty} E_i,\ E_i \text{ disjoint} \Rightarrow \mu^* E \le \sum_{i=1}^{\infty} \mu^* E_i.$$

The outer measure $\mu^*$ is called finite if $\mu^* X < \infty$.

By analogy with the case of Lebesgue measure we define a set $E$ to be **measurable** with respect to $\mu^*$ if for every set $A$ we have

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap \tilde{E}).$$

Since $\mu^*$ is subadditive, it is only necessary to show that

$$\mu^*(A) \ge \mu^*(A \cap E) + \mu^*(A \cap \tilde{E})$$

for every $A$ in order to prove that $E$ is measurable. This inequality is trivially true when $\mu^* A = \infty$, and so we need only establish it for sets $A$ with $\mu^* A$ finite.

**1. Theorem:** *The class* $\mathcal{B}$ of *$\mu^*$-measurable sets is a $\sigma$-algebra. If $\bar{\mu}$ is $\mu^*$ restricted to* $\mathcal{B}$, *then $\bar{\mu}$ is a complete measure on* $\mathcal{B}$.

***Proof:*** Trivially, the empty set is measurable. The symmetry of the definition of measurability in $E$ and $\tilde{E}$ shows that $\tilde{E}$ is measurable whenever $E$ is.

Let $E_1$ and $E_2$ be measurable sets. From the measurability of $E_2$,

$$\mu^*(A) = \mu^*(A \cap E_2) + \mu^*(A \cap \tilde{E}_2)$$

and

$$\mu^*(A) = \mu^*(A \cap E_2) + \mu^*(A \cap \tilde{E}_2 \cap E_1) + \mu^*(A \cap \tilde{E}_1 \cap \tilde{E}_2)$$

by the measurability of $E_1$. Since

$$A \cap [E_1 \cup E_2] = [A \cap E_2] \cup [A \cap E_1 \cap \tilde{E}_2],$$

we have

$$\mu^*(A \cap [E_1 \cup E_2]) \le \mu^*(A \cap E_2) + \mu^*(A \cap \tilde{E}_2 \cap E_1)$$

by subadditivity, and so

$$\mu^* A \ge \mu^*(A \cap [E_1 \cup E_2]) + \mu^*(A \cap \tilde{E}_1 \cap \tilde{E}_2).$$

This means that $E_1 \cup E_2$ is measurable, since

$$\sim(E_1 \cup E_2) = \tilde{E}_1 \cap \tilde{E}_2.$$

Thus the union of two measurable sets is measurable, and by induction the union of any finite number of measurable sets is measurable, showing that $\mathcal{B}$ is an algebra of sets.

Assume that $E = \bigcup E_i$, where $\langle E_i \rangle$ is a disjoint sequence of measurable sets, and set

$$G_n = \bigcup_{i=1}^{n} E_i.$$

Then $G_n$ is measurable, and

$$\mu^*(A) = \mu^*(A \cap G_n) + \mu^*(A \cap \tilde{G}_n) \geq \mu^*(A \cap G_n) + \mu^*(A \cap \tilde{E}),$$

since $\tilde{E} \subset \tilde{G}_n$. Now $G_n \cap E_n = E_n$ and $G_n \cap \tilde{E}_n = G_{n-1}$, and by the measurability of $E_n$ we have

$$\mu^*(A \cap G_n) = \mu^*(A \cap E_n) + \mu^*(A \cap G_{n-1}).$$

By induction

$$\mu^*(A \cap G_n) = \sum_{i=1}^{n} \mu^*(A \cap E_i),$$

and so

$$\begin{aligned} \mu^*(A) &\geq \mu^*(A \cap \tilde{E}) + \sum_{i=1}^{\infty} \mu^*(A \cap E_i) \\ &\geq \mu^*(A \cap \tilde{E}) + \mu^*(A \cap E), \end{aligned}$$

since

$$A \cap E \subset \bigcup_{i=1}^{\infty} (A \cap E_i).$$

Thus $E$ is measurable. Since the union of any sequence of sets in an algebra can be replaced by a disjoint union of sets in the algebra, it follows that $\mathcal{B}$ is a $\sigma$-algebra.

We next demonstrate the finite additivity of $\bar{\mu}$. Let $E_1$ and $E_2$ be disjoint measurable sets. Then the measurability of $E_2$ implies that

$$\begin{aligned} \bar{\mu}(E_1 \cup E_2) &= \mu^*(E_1 \cup E_2) \\ &= \mu^*([E_1 \cup E_2] \cap E_2) + \mu^*([E_1 \cup E_2] \cap \tilde{E}_2) \\ &= \mu^* E_2 + \mu^* E_1. \end{aligned}$$

Finite additivity follows by induction.

If $E$ is the disjoint union of the measurable sets $\{E_i\}$, then

$$\bar{\mu}(E) \geq \bar{\mu}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \bar{\mu}(E_i),$$

and so

$$\bar{\mu}(E) \geq \sum_{i=1}^{\infty} \bar{\mu}(E_i).$$

But

$$\bar{\mu}(E) \leq \sum_{i=1}^{\infty} \bar{\mu}(E_i)$$

by the subadditivity of $\mu^*$. Hence $\bar{\mu}$ is countably additive and thus a measure since it is nonnegative and $\bar{\mu}\emptyset = \mu^*\emptyset = 0$. ∎

## Problems

**1.** Prove the completeness of $\bar{\mu}$.

**2.** Assume that $\langle E_i \rangle$ is a sequence of disjoint measurable sets and $E = \bigcup E_i$. Then for any set $A$ we have

$$\mu^*(A \cap E) = \sum \mu^*(A \cap E_i).$$

## 2 The Extension Theorem

By a **measure on an algebra** we mean a nonnegative extended real-valued set function $\mu$ defined on an algebra $\mathfrak{A}$ of sets such that:

i. $\mu(\emptyset) = 0$.
ii. If $\langle A_i \rangle$ is a disjoint sequence of sets in $\mathfrak{A}$ whose union is also in $\mathfrak{A}$, then

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

Thus a measure on an algebra $\mathfrak{A}$ is a measure if and only if $\mathfrak{A}$ is a $\sigma$-algebra. The purpose of this section is to show that, if we start with a measure on an algebra $\mathfrak{A}$ of sets, we may extend it to a measure defined on a $\sigma$-algebra $\mathfrak{B}$ containing $\mathfrak{A}$. We shall do this by using the measure on the algebra to construct an outer measure $\mu^*$ and show that the measure $\bar{\mu}$ induced by $\mu^*$ is the desired extension of $\mu$. The process by which we construct $\mu^*$ from $\mu$ is analogous to that by which we constructed Lebesgue outer measure from the lengths of intervals: We define

$$\mu^*E = \inf \sum_{i=1}^{\infty} \mu A_i, \tag{1}$$

where $\langle A_i \rangle$ ranges over all sequences from $\mathfrak{A}$ such that $E \subset \bigcup_{i=1}^{\infty} A_i$. We first establish some lemmas concerning $\mu^*$.

**2. Lemma:** *If $A \in \mathfrak{A}$ and if $\langle A_i \rangle$ is any sequence of sets in $\mathfrak{A}$ such that $A \subset \bigcup_{i=1}^{\infty} A_i$, then $\mu A \leq \sum_{i=1}^{\infty} \mu A_i$.*

***Proof:*** Set

$$B_n = A \cap A_n \cap \tilde{A}_{n-1} \cap \cdots \cap \tilde{A}_1.$$

Then $B_n \in \mathfrak{A}$ and $B_n \subset A_n$. But $A$ is the disjoint union of the sequence $\langle B_n \rangle$, and so by countable additivity

$$\mu A = \sum_{n=1}^{\infty} \mu B_n \leq \sum_{n=1}^{\infty} \mu A_n. \quad \blacksquare$$

**3. Corollary:** *If $A \in \mathfrak{A}$, $\mu^* A = \mu A$.*

**4. Lemma:** *The set function $\mu^*$ is an outer measure.*

***Proof:*** Since $\mu^*$ is clearly a monotone nonnegative set function defined for all sets and $\mu^*\emptyset = 0$, we have only to show that it is countably subadditive. Let $E \subset \bigcup_{i=1}^{\infty} E_i$. If $\mu^*E_i = \infty$ for any $i$, we have $\mu^*E \leq \sum \mu^*E_i = \infty$. If not, given $\epsilon > 0$, there is for each $i$ a sequence $\langle A_{ij} \rangle_{j=1}^{\infty}$ of sets in $\mathfrak{A}$ such that $E_i \subset \bigcup_{j=1}^{\infty} A_{ij}$ and

$$\sum_{j=1}^{\infty} \mu A_{ij} < \mu^*E_i + \frac{\epsilon}{2^i}.$$

Then

$$\mu^*E \leq \sum_{ij} \mu A_{ij} < \sum_{i=1}^{\infty} \mu^*E_i + \epsilon.$$

Since $\epsilon$ was an arbitrary positive number,

$$\mu^*E \leq \sum_{i=1}^{\infty} \mu^*E_i,$$

and $\mu^*$ is subadditive. ▮

**5. Lemma:** *If $A \in \mathfrak{A}$, then $A$ is measurable with respect to $\mu^*$.*

***Proof:*** Let $E$ be an arbitrary set of finite outer measure and $\epsilon$ a positive number. Then there is a sequence $\langle A_i \rangle$ from $\mathfrak{A}$ such that $E \subset \bigcup A_i$ and

$$\sum \mu A_i < \mu^*E + \epsilon.$$

By the additivity of $\mu$ on $\mathfrak{A}$ we have

$$\mu(A_i) = \mu(A_i \cap A) + \mu(A_i \cap \tilde{A}).$$

Hence

$$\begin{aligned} \mu^*E + \epsilon &> \sum_{i=1}^{\infty} \mu(A_i \cap A) + \sum_{i=1}^{\infty} \mu(A_i \cap \tilde{A}) \\ &\geq \mu^*(E \cap A) + \mu^*(E \cap \tilde{A}), \end{aligned}$$

since

$$E \cap A \subset \bigcup (A_i \cap A)$$

and

$$E \cap \tilde{A} \subset \bigcup (A_i \cap \tilde{A}).$$

Since $\epsilon$ was an arbitrary positive number,

$$\mu^*E \geq \mu^*(E \cap A) + \mu^*(E \cap \tilde{A}),$$

and $A$ is measurable. ▮

The outer measure $\mu^*$ that we have defined is called the outer measure induced by $\mu$. For a given algebra $\mathfrak{A}$ of sets we use $\mathfrak{A}_\sigma$ to denote those sets that are countable unions of sets of $\mathfrak{A}$ and use $\mathfrak{A}_{\sigma\delta}$ to denote those sets that are countable intersections of sets in $\mathfrak{A}_\sigma$.

**6. Proposition:** *Let $\mu$ be a measure on an algebra $\mathfrak{A}$, $\mu^*$ the outer measure induced by $\mu$, and $E$ any set. Then for $\epsilon > 0$, there is a set $A \in \mathfrak{A}_\sigma$ with $E \subset A$ and*

$$\mu^*A \leq \mu^*E + \epsilon.$$

*There is also a set $B \in \mathfrak{A}_{\sigma\delta}$ with $E \subset B$ and $\mu^*E = \mu^*B$.*

***Proof:*** By the definition of $\mu^*$ there is a sequence $\langle A_i \rangle$ from $\mathfrak{A}$ such that $E \subset \bigcup A_i$ and

$$\sum_{i=1}^{\infty} \mu A_i \leq \mu^* E + \epsilon.$$

Set $A = \bigcup A_i$. Then $\mu^* A \leq \sum \mu^* A_i = \sum \mu A_i$.

To prove the second statement, we note that for each positive integer $n$ there is a set $A_n$ in $\mathfrak{A}_\sigma$ with $E \subset A_n$ and $\mu^* A_n < \mu^* E + 1/n$. Let $B = \bigcap A_n$. Then $B \in \mathfrak{A}_{\sigma\delta}$ and $E \subset B$. Since $B \subset A_n$, $\mu^* B \leq \mu^* A_n < \mu^* E + 1/n$. Since $n$ is arbitrary, $\mu^* B \leq \mu^* E$. But $E \subset B$, and so $\mu^* B \geq \mu^* E$ by monotonicity. Hence $\mu^* B = \mu^* E$. ▮

An outer measure $\mu^*$ is said to be *regular* if given any subset $E$ of $X$ and any $\epsilon > 0$, there is a $\mu^*$-measurable set $A$ with $E \subset A$ and

$$\mu^* A \leq \mu^* E + \epsilon.$$

It follows from Lemma 5 and Proposition 6 that every outer measure induced by a measure on an algebra is a regular outer measure.

If we apply this proposition in the case that $E$ is a measurable set of finite measure, we see that $E$ must be the difference of a set $B$ in $\mathfrak{A}_{\sigma\delta}$ and a set of measure zero. This gives us the structure of the measurable sets of finite measure, and the next proposition extends this to the $\sigma$-finite case. It can be considered a generalization of the first principle of Littlewood. It is a key element in the proof of a number of our theorems. Other forms of this principle are given by Problems 7 and 10. Versions of Littlewood's other principles are given by Propositions 11.7 and 11.26 and by Problems 11, 11.16, and 11.21c.

**7. Proposition:** *Let $\mu$ be a $\sigma$-finite measure on an algebra $\mathfrak{A}$, and let $\mu^*$ be the outer measure generated by $\mu$. A set $E$ is $\mu^*$ measurable if and only if $E$ is the proper difference $A \sim B$ of a set $A$ in $\mathfrak{A}_{\sigma\delta}$ and a set $B$ with $\mu^* B = 0$. Each set $B$ with $\mu^* B = 0$ is contained in a set $C$ in $\mathfrak{A}_{\sigma\delta}$ with $\mu^* C = 0$.*

***Proof:*** The "if" part of the proposition follows from the fact that each set in $\mathfrak{A}_{\sigma\delta}$ must be measurable, since the measurable sets form a $\sigma$-algebra, while each set of $\mu^*$-measure zero must be measurable, since $\bar{\mu}$ is complete.

To prove the "only if" part of the proposition, let $\{X_i\}$ be a countable disjoint collection of sets in $\mathfrak{A}$ with $\mu X_i$ finite and $X = \bigcup X_i$. If $E$ is measurable, then $E$ is the disjoint union of the measurable sets $E_i = X_i \cap E$. By Proposition 6 we can find for each positive integer $n$, a set $A_{ni}$ in $\mathfrak{A}_\sigma$ such that $E_i \subset A_{ni}$ and

$$\bar{\mu}A_{ni} \leq \bar{\mu}E_i + \frac{1}{n2^i}.$$

Set

$$A_n = \bigcup_{i=1}^{\infty} A_{ni}.$$

Then $E \subset A_n$, and $A_n \sim E \subset \bigcup_{i=1}^{\infty} [A_{ni} \sim E_i]$. Hence

$$\begin{aligned} \bar{\mu}(A_n \sim E) &\leq \sum_{i=1}^{\infty} \bar{\mu}(A_{ni} \sim E_i) \\ &\leq \sum_{i=1}^{\infty} \frac{1}{n2^i} = \frac{1}{n}. \end{aligned}$$

Since $A_n \in \mathfrak{A}_\sigma$, the set $A = \bigcap_{n=1}^{\infty} A_n$ is in $\mathfrak{A}_{\sigma\delta}$, and for each $n$

$$A \sim E \subset A_n \sim E.$$

Hence

$$\bar{\mu}(A \sim E) \leq \bar{\mu}(A_n \sim E) \leq \frac{1}{n}.$$

Since this holds for each positive integer $n$, we must have

$$\bar{\mu}(A \sim E) = 0. \quad \blacksquare$$

We summarize the results of this section in the following theorem.

**8. Theorem (Carathéodory):** *Let $\mu$ be a measure on an algebra $\mathfrak{A}$, and $\mu^*$ the outer measure induced by $\mu$. Then the restriction $\bar{\mu}$ of $\mu^*$ to the $\mu^*$-measurable sets is an extension of $\mu$ to a $\sigma$-algebra containing $\mathfrak{A}$. If $\mu$ is finite (or $\sigma$-finite) so is $\bar{\mu}$. If $\mu$ is $\sigma$-finite, then $\bar{\mu}$ is the only measure on the smallest $\sigma$-algebra containing $\mathfrak{A}$ which is an extension of $\mu$.*

***Proof:*** The fact that $\bar{\mu}$ is an extension of $\mu$ from $\mathfrak{A}$ to be a measure on a $\sigma$-algebra containing $\mathfrak{A}$ follows directly from Corollary 3, Lemma 5, and Theorem 1, and it is readily verified that $\bar{\mu}$ is finite or $\sigma$-finite whenever $\mu$ is.

To show the unicity of $\bar{\mu}$ when $\mu$ is $\sigma$-finite, we let $\mathcal{B}$ be the smallest $\sigma$-algebra containing $\mathcal{A}$ and $\tilde{\mu}$ some measure on $\mathcal{B}$ that agrees with $\mu$ on $\mathcal{A}$.

Since each set in $\mathcal{A}_\sigma$ can be expressed as a disjoint countable union of sets in $\mathcal{A}$, the measure $\tilde{\mu}$ must agree with $\bar{\mu}$ on $\mathcal{A}_\sigma$. Let $B$ be any set in $\mathcal{B}$ with finite outer measure. Then by Proposition 6 there is an $A$ in $\mathcal{A}_\sigma$ such that $B \subset A$ and

$$\mu^*A \leq \mu^*B + \epsilon.$$

Since $B \subset A$,

$$\tilde{\mu}B \leq \tilde{\mu}A = \mu^*A \leq \mu^*B + \epsilon.$$

Since $\epsilon$ is an arbitrary positive number, we have

$$\tilde{\mu}B \leq \mu^*B$$

for each $B \in \mathcal{B}$.

Since the class of sets measurable with respect to $\mu^*$ is a $\sigma$-algebra containing $\mathcal{A}$, each $B$ in $\mathcal{B}$ must be measurable. If $B$ is measurable and $A$ is in $\mathcal{A}_\sigma$ with $B \subset A$ and $\mu^*A \leq \mu^*B + \epsilon$, then

$$\mu^*A = \mu^*B + \mu^*(A \sim B),$$

and so

$$\tilde{\mu}(A \sim B) \leq \mu^*(A \sim B) \leq \epsilon,$$

if $\mu^*B < \infty$. Hence

$$\begin{aligned}\mu^*B \leq \mu^*A &= \tilde{\mu}A \\ &= \tilde{\mu}B + \tilde{\mu}(A \sim B) \\ &\leq \tilde{\mu}B + \epsilon.\end{aligned}$$

Since $\epsilon$ is arbitrary, we have

$$\mu^*B \leq \tilde{\mu}B$$

and so

$$\mu^*B = \tilde{\mu}B.$$

If $\mu$ is a $\sigma$-finite measure, let $\{X_i\}$ be a countable disjoint collection of sets in $\mathcal{A}$ with $X = \bigcup X_i$ and $\mu X_i$ finite. If $B$ is any set in $\mathcal{B}$, then

$$B = \bigcup (X_i \cap B)$$

and this is a countable disjoint union of sets in $\mathcal{B}$, and so we have

$$\tilde{\mu}B = \sum \tilde{\mu}(X_i \cap B)$$

and

$$\bar{\mu}B = \sum \bar{\mu}(X_i \cap B).$$

Since $\mu^*(X_i \cap B) < \infty$, we have

$$\bar{\mu}(X_i \cap B) = \tilde{\mu}(X_i \cap B). \quad \blacksquare$$

This extension procedure not only extends $\mu$ to a measure on the smallest $\sigma$-algebra $\mathcal{B}$ containing $\mathcal{A}$, but also completes and saturates the measure. If $\mu$ is $\sigma$-finite, the extension to $\mathcal{B}$ is already saturated, and the extension to the $\mu^*$-measurable sets is merely the completion of the extension of $\bar{\mu}$ on $\mathcal{B}$. If $\mu$ is not $\sigma$-finite, then the extension to $\mu^*$-measurable sets also saturates $\bar{\mu}$. It should be observed that in this case the extension of $\mu$ to $\mathcal{B}$ need not be unique (Problem 3), although any extension $\tilde{\mu}$ must agree with $\bar{\mu}$ for each set $B$ of $\mathcal{B}$ for which $\bar{\mu}B < \infty$, and we always have $\tilde{\mu}B \leq \mu^*B$. We shall return to the question of extension and unicity in Sections 6 and 7.

It is often convenient to start with a set function on a collection $\mathcal{C}$ of sets having less structure than an algebra of sets. We say that a collection $\mathcal{C}$ of subsets of $X$ is a **semialgebra** of sets if the intersection of any two sets in $\mathcal{C}$ is again in $\mathcal{C}$ and the complement of any set in $\mathcal{C}$ is a finite disjoint union of sets in $\mathcal{C}$. If $\mathcal{C}$ is any semialgebra of sets, then the collection $\mathcal{A}$ consisting of the empty set and all finite disjoint unions of sets in $\mathcal{C}$ is an algebra of sets which is called the algebra generated by $\mathcal{C}$. If $\mu$ is a set function defined on $\mathcal{C}$, it is natural to attempt to define a finitely additive set function on $\mathcal{A}$ by setting

$$\mu A = \sum_{i=1}^{n} \mu E_i$$

whenever $A$ is the disjoint union of the set $E_i$ in $\mathcal{C}$. Since a set $A$ in $\mathcal{A}$ may possibly be represented in several ways as a disjoint union of sets in $\mathcal{C}$, we must be certain that such a procedure leads to a unique value for $\mu A$. The following proposition gives conditions under which this procedure can be carried out and will give a measure on the algebra $\mathcal{A}$.

**9. Proposition:** *Let $\mathcal{C}$ be a semialgebra of sets and $\mu$ a nonnegative set function defined on $\mathcal{C}$ with $\mu\emptyset = 0$ (if $\emptyset \in \mathcal{C}$). Then $\mu$ has a unique* *extension to a measure on the algebra $\mathfrak{A}$ generated by $\mathcal{C}$ if the following conditions are satisfied:*

i. *If a set $C$ in $\mathcal{C}$ is the union of a finite disjoint collection $\{C_i\}$ of sets in $\mathcal{C}$, then $\mu C = \sum \mu C_i$.*
ii. *If a set $C$ in $\mathcal{C}$ is the union of a countable disjoint collection $\{C_i\}$ of sets in $\mathcal{C}$, then $\mu C \leq \sum \mu C_i$.*

## Problems

**3.** Let $X$ be the set of rational numbers and $\mathfrak{A}$ the algebra of finite unions of intervals of the form $(a, b]$ with $\mu(a, b] = \infty$ and $\mu\emptyset = 0$. The extension of $\mu$ to the smallest $\sigma$-algebra containing $\mathfrak{A}$ is not unique.

**4.** Prove Proposition 9 by showing:

**a.** Condition (i) implies that if $A$ is the union of each of two finite disjoint collections $\{C_i\}$ and $\{D_j\}$ of sets in $\mathcal{C}$, then $\sum \mu C_i = \sum \mu D_j$. [Hint: $\mu C_i = \sum\limits_j \mu(C_i \cap D_j)$.]

**b.** Condition (ii) implies that $\mu$ is countably additive on $\mathfrak{A}$ (for finite additivity and monotonicity already imply the reverse inequality).

**5.** Let $\mathcal{C}$ be a semialgebra of sets and $\mathfrak{A}$ the smallest algebra of sets containing $\mathcal{C}$.

**a.** Show that $\mathfrak{A}$ is comprised of sets of the form $A = \bigcup\limits_{i=1}^{n} C_i$ with $C_i \in \mathcal{C}$.

**b.** Show $\mathfrak{A}_\sigma = \mathcal{C}_\sigma$, so that $\mathfrak{A}_\sigma$ and $\mathfrak{A}_{\sigma\delta}$ may be replaced in theorems by $\mathcal{C}_\sigma$ and $\mathcal{C}_{\sigma\delta}$, respectively.

**6.** Let $\mathfrak{A}$ be a collection of sets which is closed under finite unions and finite intersections; an algebra of sets, for example.

**a.** Show that $\mathfrak{A}_\sigma$ is closed under countable unions and finite intersections.

**b.** Show that each set in $\mathfrak{A}_{\sigma\delta}$ is the intersection of a *decreasing* sequence of sets in $\mathfrak{A}_\sigma$.

**7.** Let $\mu$ be a finite measure on an algebra $\mathfrak{A}$, and $\mu^*$ the induced outer measure. Show that a set $E$ is measurable if and only if for each $\epsilon > 0$ there is a set $A \in \mathfrak{A}_\delta$, $A \subset E$, such that $\mu^*(E \sim A) < \epsilon$.

**8.** If we start with an outer measure $\mu^*$ on $X$ and form the induced measure $\bar{\mu}$ on the $\mu^*$-measurable sets, we can use $\bar{\mu}$ to induce an outer measure $\mu^+$.

**a.** Show that for each set $E$ we have $\mu^+ E \geq \mu^* E$.

**b.** For a given set $E$ we have $\mu^+E = \mu^*E$ if and only if there is a $\mu^*$-measurable set $A \supset E$ with $\mu^*A = \mu^*E$.

**c.** Show that $\mu^+E = \mu^*E$ for every $E$ if and only if $\mu^*$ is regular.

**d.** Show that an outer measure $\mu^*$ is regular if and only if it is induced by a measure on an algebra.

**e.** Let $X$ be a set consisting of two points. Construct an outer measure on $X$ which is not regular.

**9.** Let $\mu^*$ be a regular outer measure.

**a.** Show that the measure $\bar{\mu}$ induced by $\mu^*$ is complete and saturated.

**b.** Let $(X, \mathfrak{A}, \mu)$ be a complete measure space. Let $\bar{\mu}$ be the extension of $\mu$ obtained by the Carathéodory process. Then $\bar{\mu}$ is the same as the extension given in Problem 11.8c.

**10.** Let $\mu$ be a measure on an algebra $\mathfrak{A}$ and $\bar{\mu}$ the extension of it given by the Caratheodory process. Let $E$ be measurable with respect to $\bar{\mu}$ and $\bar{\mu}E < \infty$. Then given $\epsilon > 0$, there is an $A \in \mathfrak{A}$ with

$$\bar{\mu}(A \,\Delta\, E) < \epsilon.$$

**11.** We say that a function $\varphi$ is $\mathfrak{A}$-simple if $\varphi = \sum a_i \chi_{A_i}$, where $A_i \in \mathfrak{A}$. Let $\mu$ be a measure on $\mathfrak{A}$ and $\bar{\mu}$ its extension.

**a.** Given $\epsilon > 0$ and a $\bar{\mu}$ integrable function $f$, there is an $\mathfrak{A}$-simple function $\varphi$ such that

$$\int |f - \varphi| \, d\bar{\mu} < \epsilon.$$

**b.** Show that the function $\varphi$ in Problem 11.21c can be taken to be $\mathfrak{A}$-simple.

## ⋆3 The Lebesgue–Stieltjes Integral

Let $X$ be the set of real numbers and $\mathfrak{B}$ the class of all Borel sets. A measure $\mu$ defined on $\mathfrak{B}$ and finite for bounded sets is called a Baire *measure* (on the real line). To each finite Baire measure we associate a function $F$ by setting

$$F(x) = \mu(-\infty, x].$$

The function $F$ is called the *cumulative distribution function* of $\mu$ and is real-valued and monotone increasing. We have

$$\mu(a, b] = F(b) - F(a).$$

Since $(a, b]$ is the intersection of the sets $(a, b + 1/n]$, Proposition 11.2 implies that

$$\mu(a, b] = \lim_{n\to\infty} \mu\left(a, b + \frac{1}{n}\right],$$

and so

$$F(b) = \lim_{n\to\infty} F\left(b + \frac{1}{n}\right) = F(b+).$$

Thus a cumulative distribution function is continuous on the right. Similarly,

$$\begin{aligned} \mu\{b\} &= \lim_{n\to\infty} \mu\left(b - \frac{1}{n}, b\right] \\ &= \lim_{n\to\infty} F(b) - F\left(b - \frac{1}{n}\right) \\ &= F(b) - F(b-). \end{aligned}$$

Hence $F$ is continuous at $b$ if and only if the set $\{b\}$ consisting of $b$ alone has measure zero. Since $\emptyset = \bigcap\, (-\infty, -n]$, we have

$$\lim_{n\to-\infty} F(n) = 0,$$

and hence

$$\lim_{x\to-\infty} F(x) = 0,$$

because of the monotonicity of $F$. We summarize these properties in the following lemma:

**10. Lemma:** *If $\mu$ is a finite Baire measure on the real line, then its cumulative distribution function $F$ is a monotone increasing bounded function which is continuous on the right. Moreover,* $\lim_{x\to-\infty} F(x) = 0$.

Suppose that we begin with a monotone increasing function $F$ which is continuous on the right. Then we shall show that there is a unique Baire measure $\mu$ such that

$$\mu(a, b] = F(b) - F(a) \tag{2}$$

for all intervals of the form $(a, b]$, where we define $F(\infty) = \lim_{x \to \infty} F(x)$ and $F(-\infty) = \lim_{x \to -\infty} F(x)$. We begin with the following lemma, whose proof is left to the reader (Problem 12):

**11. Lemma:** *Let $F$ be a monotone increasing function continuous on the right. If $(a, b] \subset \bigcup_{i=1}^{\infty} (a_i, b_i]$, then*

$$F(b) - F(a) \leq \sum_{i=1}^{\infty} F(b_i) - F(a_i).$$

If we let $\mathcal{C}$ be the semialgebra consisting of all intervals of the form $(a, b]$ or $(a, \infty)$ and set $\mu(a, b] = F(b) - F(a)$, then $\mu$ is easily seen to satisfy condition (i) of Proposition 9, and since Lemma 11 is precisely the second condition, we see that $\mu$ admits a unique extension to a measure on the algebra generated by $\mathcal{B}$. By Theorem 8 this $\mu$ can be extended to a $\sigma$-algebra containing $\mathcal{C}$. Since the class $\mathcal{B}$ of Borel sets is the smallest $\sigma$-algebra containing $\mathcal{C}$, we have an extension of $\mu$ to a Baire measure. The measure $\mu$ is $\sigma$-finite, since $X$ is the union of the intervals $(n, n + 1]$ and each has finite measure. Thus the extension of $\mu$ to $\mathcal{B}$ is unique, and we have the following proposition:

**12. Proposition:** *Let $F$ be a monotone increasing function which is continuous on the right. Then there is a unique Baire measure $\mu$ such that for all $a$ and $b$ we have*

$$\mu(a, b] = F(b) - F(a).$$

**13. Corollary:** *Each bounded monotone function which is continuous on the right is the cumulative distribution function of a unique finite Baire measure provided $F(-\infty) = 0$.*

If $\varphi$ is a nonnegative Borel measurable function and $F$ is a monotone increasing function which is continuous on the right, we define the Lebesgue–Stieltjes integral of $\varphi$ with respect to $F$ to be

$$\int \varphi \, dF = \int \varphi \, d\mu,$$

where $\mu$ is the Baire measure having $F$ as its cumulative distribution function. If $\varphi$ is both positive and negative, we say that it is integrable with respect to $F$ if it is integrable with respect to $\mu$.

If $F$ is any monotone increasing function, then there is a unique function $F^*$ which is monotone increasing, continuous on the right, and agrees with $F$ wherever $F$ is continuous on the right (Problem 13), and we define the Lebesgue–Stieltjes integral of $\varphi$ with respect to $F$ by

$$\int \varphi \, dF = \int \varphi \, dF^*.$$

If $F$ is a monotone function, continuous on the right, then $\int_a^b \varphi \, dF$ agrees with the Riemann–Stieltjes integral whenever the latter is defined. The Lebesgue–Stieltjes integral is only defined when $F$ is monotone (or more generally of bounded variation, as in Problem 14c), while the Riemann–Stieltjes integral can exist when $F$ is not of bounded variation, say when $F$ is continuous and $\varphi$ is of bounded variation.

## Problems

**12.** Prove Lemma 11. [Choose $\epsilon > 0$. By the continuity on the right of $F$, choose $\eta_i > 0$ so that $F(b_i + \eta_i) < F(b_i) + \epsilon 2^{-i}$, and choose $\delta > 0$ so that $F(a + \delta) < F(a) + \epsilon$. Then the open intervals $(a_i, b_i + \eta_i)$ cover the closed interval $[a + \delta, b]$, and the proof proceeds like that of Proposition 3.1. A little extra care must be taken when $(a, b]$ is infinite.]

**13.** Let $F$ be a monotone increasing function, and define

$$F^*(x) = \lim_{y \to x+} F(y).$$

Then $F^*$ is a monotone increasing function which is continuous on the right and agrees with $F$ wherever $F$ is continuous on the right. We have $(F^*)^* = F^*$, and if $F$ and $G$ are monotone increasing functions which agree wherever they are both continuous, then $F^* = G^*$.

**14. a.** Show that each bounded function $F$ of bounded variation gives rise to a finite signed Baire measure $\nu$ such that

$$\nu(a, b] = F(b+) - F(a+).$$

**b.** Show that $v^+$ and $v^-$ in the Jordan decomposition correspond to the positive and negative variations of $F$.

**c.** Extend the definition of the Lebesgue–Stieltjes integral $\int \varphi \, dF$ to functions $F$ of bounded variation.

**d.** Show that if $|\varphi| \le M$ and if the total variation of $F$ is $T$, then $|\int \varphi \, dF| \le MT$.

**15. a.** Let $F$ be the cumulative distribution function of the Baire measure $\nu$, and assume that $F$ is continuous. Then for any Borel set $E$ contained in the range of $F$, we have $mE = \nu[F^{-1}(E)]$, with $m$ Lebesgue measure. [Hint: This is true for intervals, and the uniqueness part of Theorem 8 can be used to derive its truth in general.]

**b.** Generalize to the case of discontinuous cumulative distribution functions.

**16.** Let $F$ be a continuous increasing function on $[a, b]$ with $F(a) = c$, $F(b) = d$, and let $\varphi$ be a nonnegative Borel measurable function on $[c, d]$. Then $\int_a^b \varphi(F(x))\, dF(x) = \int_c^d \varphi(y)\, dy$. [Hint: Use Problem 15a to take care of the case when $\varphi$ is a characteristic function and generalize first to simple $\varphi$ and then to general $\varphi$.]

**17. a.** Show that a measure $\mu$ is absolutely continuous with respect to Lebesgue measure if and only if its cumulative distribution function is absolutely continuous.

**b.** If $\mu$ is absolutely continuous with respect to Lebesgue measure, then its Radon–Nikodym derivative is the derivative of its cumulative distribution function.

**c.** If $F$ is absolutely continuous, then

$$\int f\, dF = \int fF'\, dx.$$

**18.** *Riemann's Convergence Criterion.* Let $f$ be a nonnegative monotone decreasing function on $(0, \infty)$, $g$ a nonnegative monotone increasing function on $(0, \infty)$, and $\langle a_n \rangle$ a nonnegative sequence. Suppose that for each $x \in (0, \infty)$ the number of $n$ such that $a_n \geq f(x)$ is at most $g(x)$. Then we have $\sum a_n < \infty$ if $\int_b^\infty f\, dg < \infty$.

## 4 Product Measures

Let $(X, \mathfrak{A}, \mu)$ and $(Y, \mathfrak{B}, \nu)$ be two complete measure spaces, and consider the direct product $X \times Y$ of $X$ and $Y$. If $A \subset X$ and $B \subset Y$, we call $A \times B$ a rectangle. If $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$, we call $A \times B$ a measurable rectangle. The collection $\mathfrak{R}$ of measurable rectangles is a semialgebra, since

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$$

and

$$\sim(A \times B) = (\tilde{A} \times B) \cup (A \times \tilde{B}) \cup (\tilde{A} \times \tilde{B}).$$

If $A \times B$ is a measurable rectangle, we set

$$\lambda(A \times B) = \mu A \cdot \nu B.$$

**14. Lemma:** *Let $\{(A_i \times B_i)\}$ be a countable disjoint collection of measurable rectangles whose union is a measurable rectangle $A \times B$. Then*

$$\lambda(A \times B) = \sum \lambda(A_i \times B_i).$$

***Proof:*** Fix a point $x \in A$. Then for each $y \in B$, the point $\langle x, y\rangle$ belongs to exactly one rectangle $A_i \times B_i$. Thus $B$ is the disjoint union of those $B_i$ such that $x$ is in the corresponding $A_i$. Hence

$$\sum \nu B_i \cdot \chi_{A_i}(x) = \nu B \cdot \chi_A(x),$$

since $\nu$ is countably additive. Thus by the corollary of the Monotone Convergence Theorem (11.14), we have

$$\sum \int \nu B_i \cdot \chi_{A_i}\, d\mu = \int \nu(B) \cdot \chi_A\, d\mu$$

or

$$\sum \nu B_i \cdot \mu A_i = \nu B \cdot \mu A. \quad \blacksquare$$

The lemma implies that $\lambda$ satisfies the conditions of Proposition 9 and hence has a unique extension to a measure on the algebra $\mathcal{R}'$ consisting of all finite disjoint unions of sets in $\mathcal{R}$. Theorem 8 allows us to extend $\lambda$ to be a complete measure on a $\sigma$-algebra $\mathcal{S}$ containing $\mathcal{R}$. This extended measure is called the product measure of $\mu$ and $\nu$ and is denoted by $\mu \times \nu$. If $\mu$ and $\nu$ are finite (or $\sigma$-finite), so is $\mu \times \nu$. If $X$ and $Y$ are the real line and $\mu$ and $\nu$ are both Lebesgue measure, then $\mu \times \nu$ is called two-dimensional Lebesgue measure for the plane.

The purpose of the next few lemmas is to describe the structure of the sets which are measurable with respect to the product measure $\mu \times \nu$. If $E$ is any subset of $X \times Y$ and $x$ a point of $X$, we define the *x cross section* $E_x$ by

$$E_x = \{y: \langle x, y\rangle \in E\},$$

and similarly for the $y$ cross section for $y$ in $Y$. The characteristic function of $E_x$ is related to that of $E$ by

$$\chi_{E_x}(y) = \chi_E(x, y).$$

We also have $(\tilde{E})_x = \sim(E_x)$ and $(\bigcup E_\alpha)_x = \bigcup (E_\alpha)_x$ for any collection $\{E_\alpha\}$.

**15. Lemma:** *Let $x$ be a point of $X$ and $E$ a set in $\mathcal{R}_{\sigma\delta}$. Then $E_x$ is a measurable subset of $Y$.*

***Proof:*** The lemma is trivially true if $E$ is in the class $\mathcal{R}$ of measurable rectangles. We next show it to be true for $E$ in $\mathcal{R}_\sigma$. Let $E = \bigcup_{i=1}^{\infty} E_i$, where each $E_i$ is a measurable rectangle. Then

$$\begin{aligned}\chi_{E_x}(y) &= \chi_E(x, y)\\ &= \sup_i \chi_{E_i}(x, y)\\ &= \sup_i \chi_{(E_i)_x}(y).\end{aligned}$$

Since each $E_i$ is a measurable rectangle, $\chi_{(E_i)_x}(y)$ is a measurable function of $y$, and so $\chi_{E_x}$ must also be measurable, whence $E_x$ is measurable.

Suppose now that $E = \bigcap_{i=1}^{\infty} E_i$ with $E_i \in \mathcal{R}_\sigma$. Then

$$\begin{aligned}\chi_{E_x} &= \chi_E(x, y)\\ &= \inf_i \chi_{E_i}(x, y)\\ &= \inf_i \chi_{(E_i)_x}(y),\end{aligned}$$

and we see that $\chi_{E_x}$ is measurable. Thus $E_x$ is measurable for any $E \in \mathcal{R}_{\sigma\delta}$. ▮

**16. Lemma:** *Let $E$ be a set in $\mathcal{R}_{\sigma\delta}$ with $\mu \times \nu(E) < \infty$. Then the function $g$ defined by*

$$g(x) = \nu E_x$$

*is a measurable function of $x$ and*

$$\int g \, d\mu = \mu \times \nu(E).$$

***Proof:*** The lemma is trivially true if $E$ is a measurable rectangle. We first note that any set in $\mathcal{R}_\sigma$ is a disjoint union of measurable rectangles. Let $\langle E_i \rangle$ be a disjoint sequence of measurable rectangles, and let $E = \bigcup E_i$. Set

$$g_i(x) = \nu[(E_i)_x].$$

Then each $g_i$ is a nonnegative measurable function, and

$$g = \sum g_i .$$

Thus $g$ is measurable, and by the corollary of the Monotone Convergence Theorem (11.14), we have

$$\begin{aligned} \int g \, d\mu &= \sum \int g_i \, d\mu \\ &= \sum \mu \times \nu(E_i) \\ &= \mu \times \nu(E). \end{aligned}$$

Consequently, the lemma holds for $E \in \mathcal{R}_\sigma$.

Let $E$ be a set of finite measure in $\mathcal{R}_{\sigma\delta}$. Then there is a sequence $\langle E_i \rangle$ of sets in $\mathcal{R}_\sigma$ such that $E_{i+1} \subset E_i$ and $E = \bigcap E_i$. It follows from Proposition 6 that we may take $\mu \times \nu(E_1) < \infty$. Let $g_i(x) = \nu[(E_i)_x]$. Since

$$\int g_1 \, d\mu = \mu \times \nu(E_1) < \infty,$$

we have $g_1(x) < \infty$ for almost all $x$. For an $x$ with $g_1(x) < \infty$, we have $\langle (E_i)_x \rangle$ a decreasing sequence of measurable sets of finite measure whose intersection is $E_x$.

Thus by Proposition 11.2 we have

$$\begin{aligned} g(x) = \nu(E_x) &= \lim \nu[(E_i)_x] \\ &= \lim g_i(x). \end{aligned}$$

Hence

$$g_i \to g \text{ a.e.,}$$

and so $g$ is measurable. Since $0 \le g_i \le g_1$, the Lebesgue Convergence Theorem implies that

$$\begin{aligned} \int g \, d\mu &= \lim \int g_i \, d\mu \\ &= \lim \mu \times \nu(E_i) \\ &= \mu \times \nu(E). \end{aligned}$$

the last equality following from Proposition 11.2. ▮

**17. Lemma:** *Let $E$ be a set for which $\mu \times \nu(E) = 0$. Then for almost all $x$ we have $\nu(E_x) = 0$.*

***Proof:*** By Proposition 6 there is a set $F$ in $\mathcal{R}_{\sigma\delta}$ such that $E \subset F$ and $\mu \times \nu(F) = 0$. It follows from Lemma 16 that for almost all $x$ we have $\nu(F_x) = 0$. But $E_x \subset F_x$, and so $\nu E_x = 0$ for almost all $x$ since $\nu$ is complete. ▮

**18. Proposition:** *Let E be a measurable subset of $X \times Y$ such that $\mu \times \nu(E)$ is finite. Then for almost all x the set $E_x$ is a measurable subset of Y. The function g defined by*

$$g(x) = v(E_x)$$

*is a measurable function defined for almost all x and*

$$\int g \, d\mu = \mu \times \nu(E).$$

***Proof:*** By Proposition 6 there is a set $F$ in $\mathcal{R}_{\sigma\delta}$ such that $E \subset F$ and $\mu \times \nu(F) = \mu \times \nu(E)$. Let $G = F \sim E$. Since $E$ and $F$ are measurable, so is $G$, and

$$\mu \times \nu(F) = \mu \times \nu(E) + \mu \times \nu(G).$$

Since $\mu \times \nu(E)$ is finite and equal to $\mu \times \nu(F)$, we have $\mu \times \nu(G) = 0$. Thus by Lemma 17 we have $\nu G_x = 0$ for almost all $x$. Hence

$$g(x) = \nu E_x = \nu F_x \text{ a.e.};$$

so $g$ is a measurable function by Lemma 16. Again by Lemma 16

$$\int g \, d\mu = \mu \times \nu(F)$$
$$= \mu \times \nu(E). \quad ▮$$

The following two theorems enable us to interchange the order of integration and to calculate integrals with respect to product measures by iteration.

**19. Theorem (Fubini):** *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two complete measure spaces and f an integrable function on $X \times Y$. Then*

i. *For almost all x the function $f_x$ defined by $f_x(y) = f(x, y)$ is an integrable function on Y.*

i′. *For almost all y the function $f^y$ defined by $f^y(x) = f(x, y)$ is an integrable function on X.*

ii. $\int_Y f(x, y)\, d\nu(y)$ *is an integrable function on* $X$.

ii′. $\int_X f(x, y)\, d\mu(x)$ *is an integrable function on* $Y$.

iii. $\displaystyle \int_X \left[\int_Y f\, d\nu\right] d\mu = \int_{X\times Y} f\, d(\mu \times \nu) = \int_Y \left[\int_X f\, d\mu\right] d\nu.$

***Proof:*** Because of the symmetry between $x$ and $y$ it suffices to prove (i), (ii), and the first half of (iii). If the conclusion of the theorem holds for each of two functions, it also holds for their difference, and hence it is sufficient to consider the case when $f$ is nonnegative. Proposition 18 asserts that the theorem is true if $f$ is the characteristic function of a measurable set of finite measure, and hence the theorem must be true if $f$ is a simple function which vanishes outside a set of finite measure. Proposition 11.7 asserts that each nonnegative integrable function $f$ is the limit of an increasing sequence $\langle\varphi_n\rangle$ of nonnegative simple functions, and, since each $\varphi_n$ is integrable and simple, it must vanish outside a set of finite measure. Thus $f_x$ is the limit of the increasing sequence $\langle(\varphi_n)_x\rangle$ and is measurable. By the Monotone Convergence Theorem

$$\int_Y f(x, y)\, d\nu(y) = \lim \int_Y \varphi_n(x, y)\, d\nu(y),$$

and so this integral is a measurable function of $x$. Again by the Monotone Convergence Theorem

$$\begin{aligned}\int_X \left[\int_Y f\, d\nu\right] d\mu &= \lim \int_X \left[\int_Y \varphi_n\, d\nu\right] d\mu \\ &= \lim \int_{X\times Y} \varphi_n\, d(\mu\times\nu) \\ &= \int_{X\times Y} f\, d(\mu \times \nu). \quad \blacksquare\end{aligned}$$

In order to apply the Fubini Theorem, one must first verify that $f$ is integrable with respect to $\mu \times \nu$; that is, one must show that $f$ is a measurable function on $X \times Y$ and that $\int |f|\, d(\mu \times \nu) < \infty$. The measurability of $f$ on $X \times Y$ is sometimes difficult to establish, but in many cases we can establish it by topological considerations (cf. Problem 21). In the case when $\mu$ and $\nu$ are $\sigma$-finite, the integrability of $f$ can be determined by iterated integration using the following theorem:

**20. Theorem (Tonelli):** *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two $\sigma$-finite measure spaces, and let $f$ be a nonnegative measurable function on $X \times Y$. Then*

i. *For almost all $x$ the function $f_x$ defined by $f_x(y) = f(x, y)$ is a measurable function on $Y$.*

i′. *For almost all $y$ the function $f^y$ defined by $f^y(x) = f(x, y)$ is a measurable function on $X$.*

ii. $\int_Y f(x, y)\, d\nu(y)$ *is a measurable function on $X$.*

ii′. $\int_X f(x, y)\, d\mu(x)$ *is a measurable function on $Y$.*

iii. $$\int_X \left[\int_Y f\, d\nu\right] d\mu = \int_{X \times Y} f\, d(\mu \times \nu) = \int_Y \left[\int_X f\, d\mu\right] d\nu.$$

***Proof:*** For a nonnegative measurable function $f$ the only point in the proof of Theorem 19 where the integrability of $f$ was used was to infer the existence of an increasing sequence $\langle \varphi_n \rangle$ of simple functions each vanishing outside a set of finite measure such that $f = \lim \varphi_n$. But if $\mu$ and $\nu$ are $\sigma$-finite, then so is $\mu \times \nu$, and any nonnegative measurable function on $X \times Y$ can be so approximated by Proposition 11.7. ▮

If $\mathcal{A}$ and $\mathcal{B}$ are $\sigma$-algebras on $X$ and $Y$, then the smallest $\sigma$-algebra containing the measurable rectangles is denoted by $\mathcal{A} \times \mathcal{B}$. Thus the product measure is defined on a $\sigma$-algebra containing $\mathcal{A} \times \mathcal{B}$, and since $\mu \times \nu$ is obtained by the Carathéodory extension process, it is both complete and saturated. If $\mu$ and $\nu$ are both $\sigma$-finite, then the product measure on $\mathcal{A} \times \mathcal{B}$ is already saturated and the measurable sets for $\mu \times \nu$ are those which differ from sets in $\mathcal{A} \times \mathcal{B}$ by sets of measure zero.

Many authors prefer to define product measure to be the restriction of $\mu \times \nu$ to $\mathcal{A} \times \mathcal{B}$. The advantage of taking $\mu \times \nu$ to be complete, as we have done here, is that this does what we want it to for Lebesgue measure: The product of $n$-dimensional Lebesgue measure with $m$-dimensional Lebesgue measure is $(n + m)$-dimensional Lebesgue measure. Since our hypotheses for the Fubini and Tonelli theorems require only measurability with respect to the complete product measure, they are weaker than requiring measurability with respect to $\mathcal{A} \times \mathcal{B}$. The price for using these weaker hypotheses is the necessity of including the "almost all" phrases in the conclusion of the theorems. This has to be expected, since changing $f$ arbitrarily for $x$ in a set of measure zero does not change the measurability or integrability of $f$, but $f_x$ can be arbitrary for those $x$. If however, $f$ is measurable with respect to $\mathcal{A} \times \mathcal{B}$, then $f_x$ is measurable for each $x$.

We have also used the completeness of $\mu$ to show that $\int f(x, y)\, d\nu(y)$ is measurable, for if $\mu$ were not complete we could only conclude that this was a function which differed on a subset of a set of measure zero from a measurable function. If, however, $f$ is measurable with respect to $\mathcal{A} \times \mathcal{B}$, then it turns out that $\int f(x, y)\, d\nu(y)$ is measurable with respect to $\mathcal{A}$ even if $\mu$ is not complete (provided $f$ is integrable), but the proof of this is surprisingly intricate. For a proof see Halmos [5], p. 143.

The examples in the problems show that we cannot omit the hypothesis of the integrability of $f$ from the Fubini Theorem or the hypotheses of $\sigma$-finiteness and nonnegativity from the Tonelli Theorem. Problem 26 shows the essential role played by the measurability of $f$ in these theorems: If we omit this assumption, even for bounded functions and finite measures, we may have the iterated integrals $\int [\int f\, d\nu]\, d\mu$ and $\int [\int f\, d\mu]\, d\nu$ well defined but unequal.

## Problems

**19.** Let $X = Y$ be the set of positive integers, $\mathcal{A} = \mathcal{B} = \mathcal{P}(X)$, and let $\nu = \mu$ be the measure defined by setting $\mu(E)$ equal to the number of points in $E$ if $E$ is finite and $\infty$ if $E$ is an infinite set. (This measure is called the counting measure.) State the Fubini and Tonelli Theorems explicitly for this case.

**20.** Let $(X, \mathcal{A}, \mu)$ be any $\sigma$-finite measure space and $Y$ the set of positive integers with $\nu$ the counting measure (Problem 19). Then Theorem 20 and Corollary 11.14 state the same conclusion. However, Corollary 11.14 is valid even if $\mu$ is not $\sigma$-finite, and hence the Tonelli Theorem is true without $\sigma$-finiteness if $(Y, \mathcal{B}, \nu)$ is this special measure space.

**21.** Let $X = Y = [0, 1]$, and let $\mu = \nu$ be Lebesgue measure. Show that each open set in $X \times Y$ is measurable, and hence each Borel set in $X \times Y$ is measurable.

**22.** Let $h$ and $g$ be integrable functions on $X$ and $Y$, and define $f(x, y) = h(x)g(y)$. Then $f$ is integrable on $X \times Y$ and

$$\int_{X \times Y} f \, d(\mu \times \nu) = \int_X h \, d\mu \int_Y g \, d\nu.$$

(*Note:* We do not need to assume that $\mu$ and $\nu$ are $\sigma$-finite.)

**23.** Show that Tonelli's Theorem is still true if, instead of assuming $\mu$ and $\nu$ to be $\sigma$-finite, we merely assume that $\{\langle x, y\rangle : f(x, y) \neq 0\}$ is a set of $\sigma$-finite measure.

**24.** The following example shows that we cannot remove the hypothesis that $f$ be nonnegative from the Tonelli Theorem or that $f$ be integrable from the Fubini Theorem. Let $X = Y$ be the positive integers and $\mu = \nu$ be the counting measure. Let

$$f(x, y) = \begin{cases} 2 - 2^{-x} & \text{if } x = y \\ -2 + 2^{-x} & \text{if } x = y + 1 \\ 0 & \text{otherwise.} \end{cases}$$

**25.** The following example shows that we cannot remove the hypothesis that $f$ be integrable from the Fubini Theorem or that $\mu$ and $\nu$ are $\sigma$-finite from the Tonelli Theorem: Let $X = Y$ be the interval $[0, 1]$, with $\mathfrak{A} = \mathfrak{B}$ the class of Borel sets. Let $\mu$ be Lebesgue measure and $\nu$ the counting measure. Then the diagonal $\Delta = \{\langle x, y\rangle \in X \times Y : x = y\}$ is measurable (is an $\mathcal{R}_{\sigma\delta}$, in fact), but its characteristic function fails to satisfy any of the equalities in condition (iii) of the Fubini and Tonelli Theorems.

**26.** The following example shows that the hypothesis that $f$ be measurable with respect to the product measure cannot be omitted from the Fubini and Tonelli Theorems even if we assume the measurability of $f^y$ and $f_x$ and the integrability of $\int f(x, y)\, d\nu(y)$ and $\int f(x, y)\, d\mu(x)$. Let $X = Y =$ the set of ordinals less than or equal to the first uncountable ordinal $\Omega$. Let $\mathfrak{A} = \mathfrak{B}$ be the $\sigma$-algebra consisting of all countable sets and their complements. Define $\mu = \nu$ by letting $\mu E = 0$ if $E$ countable, $\mu E = 1$ otherwise. Define a subset $S$ of $X \times Y$ by $S = \{\langle x, y\rangle : x < y\}$. Then $S_x$ and $S_y$ are measurable for each $x$ and $y$, but if $f$ is the characteristic function of $S$ we have

$$\int \left[ \int f(x, y)\, d\mu(x) \right] d\nu(y) \neq \int \left[ \int f(x, y)\, d\nu(y) \right] d\mu(x).$$

If we assume the continuum hypothesis, that is, that $X$ can be put in one-to-one correspondence with $[0, 1]$, then we can take $f$ to be a function on the unit square such that $f_x$ and $f^y$ are bounded and measurable for each $x$ and $y$ but such that the conclusion of the Fubini and Tonelli Theorems do not hold.

**27.** Show that if $(X, \mathfrak{A}, \mu)$ and $(Y, \mathfrak{B}, \nu)$ are two $\sigma$-finite measure spaces, then $\mu \times \nu$ is the only measure on $\mathfrak{A} \times \mathfrak{B}$ which assigns the value $\mu A\ \nu B$ to each measurable rectangle $A \times B$. Show that a measure on $\mathfrak{A} \times \mathfrak{B}$ with this property need not be unique, if we do not have $\sigma$-finiteness.

**28. a.** Show that if $E \in \mathfrak{A} \times \mathfrak{B}$, then $E_x \in \mathfrak{B}$ for each $x$.

**b.** If $f$ is measurable with respect to $\mathfrak{A} \times \mathfrak{B}$, then $f_x$ is measurable with respect to $\mathfrak{B}$ for each $x$.

**29.** Let $X = Y = \mathbf{R}$ and let $\mu = \nu =$ Lebesgue measure. Then $\mu \times \nu$ is two-dimensional Lebesgue measure on $X \times Y = \mathbf{R}^2$. We often write $dx\, dy$ for $d(\mu \times \nu)$.

**a.** For each measurable subset $E$ of $\mathbf{R}$, let

$$\sigma(E) = \{\langle x, y\rangle : x - y \in E\}.$$

Show that $\sigma(E)$ is a measurable subset of $\mathbf{R}^2$. [Hint: Consider first the cases when $E$ open, $E$ a $G_\delta$, $E$ of measure zero, and $E$ measurable.]

**b.** If $f$ is a measurable function on $\mathbf{R}$, the function $F$ defined by $F(x, y) = f(x - y)$ is a measurable function on $\mathbf{R}^2$.

**c.** If $f$ and $g$ are integrable functions on $\mathbf{R}$, then for almost all $x$ the function $\varphi$ given by $\varphi(y) = f(x - y)g(y)$ is integrable. If we denote its integral by $h(x)$, then $h$ is integrable and

$$\int |h| \le \int |f| \int |g|.$$

**30.** Let $f$ and $g$ be functions in $L^1(-\infty, \infty)$, and define $f * g$ to be the function $h$ defined by $h(y) = \int f(y - x)g(x)\, dx$.

**a.** Show that $f * g = g * f$.

**b.** Show that $(f * g) * h = f * (g * h)$.

**c.** For $f \in L^1$, define $\hat{f}$ by $\hat{f}(s) = \int e^{ist} f(t)\, dt$. Then $\hat{f}$ is a bounded complex function and

$$\widehat{f * g} = \hat{f}\hat{g}.$$

**31.** Let $f$ be a nonnegative integrable function on $(-\infty, \infty)$, and let $m_2$ be two-dimensional Lebesgue measure on $\mathbf{R}^2$. Then

$$m_2\{\langle x, y\rangle : 0 \le y \le f(x)\} = m_2\{\langle x, y\rangle : 0 < y < f(x)\} = \int f(x)\, dx.$$

Let $\varphi(t) = m\{x : f(x) \ge t\}$. Then $\varphi$ is a decreasing function and

$$\int_0^\infty \varphi(t)\, dt = \int f(x)\, dx.$$

**32.** If $\langle (X_i, \mathfrak{A}_i, \mu_i)\rangle_{i=1}^n$ is a finite collection of measure spaces, we can form the product measure $\mu_1 \times \cdots \times \mu_n$ on the space $X_1 \times \cdots \times X_n$ by starting with the semialgebra of rectangles of the form $R = A_1 \times \cdots \times A_n$ and $\mu(R) = \prod \mu_i A_i$, and using the Carathéodory extension procedure. Show that if we identify $(X_1 \times \cdots \times X_p) \times (X_{p+1} \times \cdots \times X_n)$ with $(X_1 \times \cdots \times X_n)$, then $(\mu_1 \times \cdots \times \mu_p) \times (\mu_{p+1} \times \cdots \times \mu_n) = \mu_1 \times \cdots \times \mu_n$.

**33.** A measure $\mu$ with $\mu X = 1$ is often called a *probability measure*. Let $\{(X_\lambda, \mathfrak{A}_\lambda, \mu_\lambda)\}$ be a collection of probability measure spaces. Show that we can define a probability measure

$$\mu = \prod_\lambda \mu_\lambda$$

on a suitable $\sigma$-algebra on the space $\bigtimes_\lambda X_\lambda$ so that

$$\mu A = \prod \mu_\lambda A_\lambda$$

when $A = \bigtimes_\lambda A_\lambda$. (Note that $\mu A$ can only be nonzero if all but a countable number of the $A_\lambda$ have $\mu A_\lambda = 1$.)

## 5 Integral Operators

In this section we study a class of integral operators which define linear transformations from $L^q(\nu)$ to $L^p(\mu)$. We let the letters $p$, $q$, and $r$ stand for extended real numbers $1 \le p \le \infty$, and so on, and use $p^*$ for the conjugate exponent $p/(p-1)$ so that $1/p + 1/p^* = 1$. We shall often denote $1/p$, $1/q$, and $1/r$ by $\alpha$, $\beta$, and $\gamma$, respectively. Thus $\alpha^* = 1 - \alpha$.

Let $(X, \mathfrak{A}, \mu)$ and $(Y, \mathfrak{B}, \nu)$ be two $\sigma$-finite measure spaces and $k = k(x, y)$ a nonnegative measurable function on $X \times Y$. We define

$$M^*_{\gamma,\beta} = \sup \iint h(x)k(x, y)g(y)\, d(\mu \times \nu)$$

as $h$ and $g$ range over all functions of norm at most one in $L^r(\mu)$ and $L^q(\nu)$, where $\gamma = 1/r$ and $\beta = 1/q$. Since $k \ge 0$, it suffices to consider only nonnegative $h$ and $g$. If $M^*_{\gamma,\beta} < \infty$, we say that $k$ is an **integral kernel** of **covariant type** $(r, q)$, and call $M^*_{\gamma,\beta}$ its covariant norm. We also say that $k$ is an integral kernel of **operator type** $(p, q)$, where $p = r^*$. We write

$$M_{\alpha,\beta} = M^*_{(1-\alpha),\beta} = \|k\|_{p,q}.$$

In the notion of covariant type one thinks of $k$ as defining a bilinear form

$$[h, g] = \iint h(x)k(x, y)g(y)\, d(\mu \times \nu)$$

between the elements of $L^r(\mu)$ and $L^q(\nu)$, with $M^*_{\gamma,\beta}$ the norm of the bilinear form. The relation of this to the operator version is given by the following proposition.

**21. Proposition:** *Let $k$ be a nonnegative measurable function on $X \times Y$ of covariant type $(p^*, q)$ and $g \in L^q(\nu)$. Then for almost all $x \in X$ the integral*

$$f(x) = \int_Y k(x, y)g(y)\, d\nu$$

*exists, and the function $f$ belongs to $L^p(\mu)$* with

$$\|f\|_p \le M_{1/p,\,1/q}\, \|g\|_q.$$

***Proof:*** Since $\mu$ is $\sigma$-finite, there is a function $h \in L^{p^*}(\mu)$ with $h(x) > 0$ everywhere. Since

$$\iint_{X \times Y} h(x)k(x, y)|g(y)|\, d(\mu \times \nu) \le M_{\alpha,\beta}\, \|h\|_{p^*}\, \|g\|_q < \infty,$$

we see that

$$\int_Y h(x)k(x, y)|g(y)|\, d\nu = h(x) \int_Y k(x, y)|g(y)|\, d\nu$$

for almost all $x \in X$ by Tonelli's Theorem.

Thus $f(x)$ exists for almost all $x$. Let $h$ be an arbitrary function in $L^{p^*}(\mu)$. Then

$$\int_X |h(x)f(x)|\, d\mu = \iint_{X \times Y} |h(x)k(x, y)g(y)|\, d(\mu \times \nu)$$

by the Fubini Theorem, since $|hkg|$ is integrable. Consequently, $hf$ is integrable and

$$\left| \int hf\, d\mu \right| \le M^*_{1-\alpha,\beta} \|h\|_{p^*} \|g\|_q.$$

By Lemma 7.27 we have $f \in L^p$ and

$$\|f\|_p \leq M_{\alpha,\beta}\|g\|_q. \quad \blacksquare$$

This proposition shows that we have defined a linear operator $T: L^q(\nu) \to L^p(\mu)$ by taking $Tg = f$ where

$$f(x) = \int_Y k(x, y)g(y)\, d\nu.$$

Moreover, the operator norm $\|T\|$ of $T$ can be shown to be $M_{\alpha,\beta} = \|k\|_{p,q}$.

More generally, we call a measurable function $k(x, y)$ on $X \times Y$ an integral operator of *absolute operator type* $(p, q)$ if $|k|$ is of operator type $(p, q)$. The proposition can be rephrased for such kernels:

**22. Corollary:** *Let $k(x, y)$ be a measurable function on $X \times Y$ of absolute operator type $(p, q)$ and $g \in L^q(\nu)$. Then for almost all $x \in X$ the integral*

$$f(x) = \int_Y k(x, y)g(y)\, d\nu$$

*exists, and the function $f$ belongs to $L^p(\mu)$ with*

$$\|f\|_p \leq \|\,|k|\,\|_{p,q}\,\|g\|_q.$$

The following useful theorem is due to M. Riesz.

**23. Theorem:** *Let $k$ be a nonnegative measurable function on $X \times Y$, and set*

$$M^*_{\gamma,\beta} = \sup \iint_{X \times Y} h(x)k(x, y)g(y)\, d(\mu \times \nu),$$

*where $f$ and $g$ range over the unit balls in $L^r(\mu)$ and $L^q(\nu)$, respectively. Then the function $\log M^*_{\gamma,\beta}$ is a convex function of $\gamma$ and $\beta$ in the square $0 \leq \gamma \leq 1$, $0 \leq \beta \leq 1$.*

***Proof:*** We have to verify that if $0 \leq \lambda \leq 1$, $\gamma = \lambda\gamma_1 + (1 - \lambda)\gamma_2$, and $\beta = \lambda\beta_1 + (1 - \lambda)\beta_2$, then

$$M^*_{\gamma,\beta} \leq (M^*_{\gamma_1,\beta_1})^{\lambda}(M^*_{\gamma_2,\beta_2})^{1-\lambda}.$$

Let $h$ and $g$ be arbitrary nonnegative functions in the unit balls of $L^r(\mu)$ and $L^q(\nu)$. Set

$$h_1 = h^{\gamma_1/\gamma}, \quad h_2 = h^{\gamma_2/\gamma}, \quad g_1 = g^{\beta_1/\beta}, \quad g_2 = g^{\beta_2/\beta}.$$

Then $h_1$, $h_2$, $g_1$, and $g_2$ are in the unit balls of $L^{r_1}(\mu)$, $L^{r_2}(\mu)$, $L^{q_1}(\nu)$, and $L^{q_2}(\nu)$, respectively. Also,

$$\iint hkg = \iint (h_1 kg)^\lambda (h_2 kg_2)^{1-\lambda}$$
$$\leq \left[\iint h_1 kg_1\right]^\lambda \cdot \left[\iint h_2 kg\right]^{1-\lambda}$$

by the Hölder inequality for $u = 1/\lambda$ and $u^* = 1/(1-\lambda)$. Hence

$$\iint hkg \leq (M^*_{\gamma_1,\beta_1})^\lambda (M^*_{\gamma_2,\beta_2})^{1-\lambda}$$

and the result follows by taking the supremum over $h$ and $g$. ∎

**24. Corollary:** *The function* $\log M_{\alpha,\beta}$ *is a convex function of* $\alpha$ *and* $\beta$ *in the square* $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$.

***Proof:*** We have $M_{\alpha,\beta} = M^*_{1-\alpha,\beta}$. ∎

The preceding Theorem of Riesz requires that $k$ be nonnegative. A deeper theorem, also due to Riesz, asserts that for a kernel of mixed sign (or even complex valued) the operator norm $M_{\alpha,\beta}$ of the corresponding integral operator is logarithmically convex on the square $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$. The interested reader will find a proof in Dunford and Schwartz [4], p. 525 or Hardy, Littlewood, and Pólya [19], p. 214.

When $(X, \mathcal{A}, \mu) = (Y, \mathcal{B}, \nu) = (\mathbf{R}^n, \mathfrak{M}, m)$, where $m$ is Lebesgue measure, we obtain a special class of integral operators by taking $k(x, y) = k(x - y)$ for some $k \in L^r(m)$. Such operators are called *convolution operators*. It is readily verified in this case that $k$ is of covariant types $(1, r^*)$ and $(r^*, 1)$. It follows from Proposition 21 that $k$ is also of convariant type $(p, q)$ when

$$\frac{1}{p} = 1 - \frac{1-\lambda}{r}, \qquad \frac{1}{q} = 1 - \frac{\lambda}{r}$$

for $0 \leq \lambda \leq 1$. This gives us the following propositions.

**25. Proposition:** *Let* $g$, $h$, *and* $k$ *be functions on* $\mathbf{R}^n$ *of class* $L^q$, $L^p$, *and* $L^r$, *respectively, with* $1/p + 1/q + 1/r = 2$. *Then*

$$\iint_{\mathbf{R}^{2n}} |h(x)k(x-y)g(y)|\,dx\,dy \leq \|h\|_p \|k\|_r \|g\|_q.$$

**26. Proposition:** *Let $g \in L^q$ and $k \in L^r$, with $1/q + 1/r > 1$. Then the function*

$$f(x) = \int_{\mathbf{R}^n} k(x - y)g(y)\,dy$$

*is defined for almost all x and*

$$\|f\|_p \leq \|k\|_r \|g\|_q,$$

*where*
$$\frac{1}{p} = \frac{1}{q} + \frac{1}{r} - 1.$$

## Problems

**34.** Show that for the operator $T$ defined by Proposition 21 we have $\|T\| = M_{\alpha,\beta}$.

**35.** Prove Corollary 22.

**36.** Prove Proposition 25.

**37.** Prove Proposition 26.

**38.** Let $g$, $h$, and $k$ be functions on $\mathbf{R}^n$ of class $L^q$, $L^p$, and $L^r$, with $1/p + 1/q + 1/r \leq 2$. Then $h(x)k(x - y)g(y)$ belongs to $L^u$ on $\mathbf{R}^{2n}$, where

$$\frac{2}{u} = \frac{1}{p} + \frac{1}{q} + \frac{1}{r}.$$

## ★6 Inner Measure

Let $\mu$ be a measure on an algebra $\mathfrak{A}$ and $\mu^*$ the induced outer measure. Then $\mu^*E$ may be thought of as the largest possible measure for $E$ compatible with $\mu$. We can also define an inner measure $\mu_*$ which assigns to a given set $E$ the smallest measure compatible with $\mu$:

**Definition:** *Let $\mu$ be a measure on an algebra $\mathfrak{A}$ and $\mu^*$ the induced outer measure. We define the inner measure $\mu_*$ induced by $\mu$ by setting*

$$\mu_* E = \sup\,[\mu A - \mu^*(A \sim E)],$$

*where the supremum is taken over all sets $A \in \mathfrak{A}$ for which $\mu^*(A \sim E) < \infty$.*

Inner measure was important historically because the measurability of a set was originally characterized using both inner and outer measure. In the historical context inner measure was first defined for bounded subsets of **R**. For such sets the definition above is equivalent to the historical one:

$$\mu_* E = l(I) - \mu(I \sim E),$$

where $I$ is a finite interval containing $E$ (see Lemma 29). A bounded set $E$ was then defined to be measurable if $\mu_* E = \mu^* E$, and the measurability of unbounded sets was defined in terms of their intersections with finite intervals. Even in the case of a bounded set this procedure is more cumbersome than the elegant approach of Carathéodory, which we have followed in this chapter. Apart from this historical importance, inner measure is useful for the extension of $\mu$ from $\mathfrak{A}$ to an algebra containing $\mathfrak{A}$ and a given set $E$ (which need not be measurable) and for determining the freedom we have in extending $\mu$ to a $\sigma$-algebra containing $\mathfrak{A}$. The purpose of this section is to do this and to develop the basic properties of inner measure.

**27. Lemma:** *We have $\mu_* E \leq \mu^* E$. If $E \in \mathfrak{A}$, $\mu_* E = \mu E$.*

***Proof:*** Since

$$\mu A \leq \mu^*(A \cap E) + \mu^*(A \cap \tilde{E}),$$

we have

$$\mu A - \mu^*(A \cap \tilde{E}) \leq \mu^*(A \cap E) \leq \mu^* E.$$

Consequently, $\mu_* E \leq \mu^* E$.

If $E \in \mathfrak{A}$, take $A = E$ in the definition of $\mu_*$. Then $\mu_* E \geq \mu E = \mu^* E$. ▌

**28. Lemma:** *If $E \subset F$, then $\mu_* E \leq \mu_* F$.*

***Proof:*** If $\mu^*(A \sim E) < \infty$, then $\mu^*(A \sim F) < \infty$, and so

$$\mu_* F \geq \mu A - \mu^*(A \sim F) \geq \mu A - \mu^*(A \sim E).$$

Taking suprema over $A \in \mathfrak{A}$ with $\mu^*(A \sim E) < \infty$, we have $\mu_* F \geq \mu_* E$. ▌

One of the difficulties of using the definition of inner measure is that we must take the supremum of $\mu A - \mu^*(A \sim E)$ over all $A$ with

$\mu^*(A \sim E) < \infty$. The next lemma shows that this expression is monotone in $A$ and enables us to calculate $\mu_* E$ more easily.

**29. Lemma:** *Let $A$ and $B$ be two sets in $\mathfrak{A}$ with $\mu^*(A \sim E) < \infty$ and $\mu^*(B \sim E) < \infty$. If $A \subset B$, we have $\mu A - \mu^*(A \sim E) \leq \mu B - \mu^*(B \sim E)$. If also $E \subset A$, we have equality, and hence $\mu_* E = \mu A - \mu^*(A \sim E)$.*

***Proof:*** Since $B \sim E \subset (B \sim A) \cup (A \sim E)$, we have $\mu^*(B \sim E) \leq \mu(B \sim A) + \mu^*(A \sim E)$. If $E \subset A$, this union is a disjoint one and the measurability of $B \sim A$ gives us equality. Since $\mu B = \mu A + \mu(B \sim A)$, we have

$$\mu B - \mu^*(B \sim E) \geq \mu A - \mu^*(A \sim E)$$

with equality if $E \subset A$. ▮

This lemma and its corollary show that if $\mu$ is a finite measure, then $\mu_* E = \mu X - \mu^* \tilde{E}$. In this case the development of the theory and properties of inner measure are relatively straightforward. The complexity of the treatment of inner measure in this section is caused by having the concept apply to measures that are not $\sigma$-finite.

**30. Corollary:** *If $A \in \mathfrak{A}$, then $\mu A = \mu_*(A \cap E) + \mu^*(A \cap \tilde{E})$.*

***Proof:*** If $\mu^*(A \cap \tilde{E}) = \infty$, then $\mu A = \infty$ and there is nothing to prove. Otherwise, set $F = A \cap E$. Then $A \sim F = A \cap \tilde{E}$, and $\mu_* F = \mu A - \mu^*(A \cap \tilde{E})$, since $F \subset A$ and $\mu^*(A \sim F) < \infty$. ▮

**31. Lemma:** *Let $B$ be a $\mu^*$-measurable set with $\mu^* B < \infty$. Then $\mu_* B = \mu^* B$.*

***Proof:*** Since $\mu^* B < \infty$, given $\epsilon > 0$, there is an $A \in \mathfrak{A}$ with $\mu^*(B \sim A) < \epsilon$ (Problem 7). Since $A$ is measurable,

$$\mu^* B = \mu^*(A \cap B) + \mu^*(B \cap \tilde{A})$$

and so

$$\mu^*(A \cap B) > \mu^* B - \epsilon.$$

Now

$$\mu_* B \geq \mu A - \mu^*(A \cap \tilde{B}) = \mu^*(A \cap B),$$

since $B$ is measurable. Thus

$$\mu^*B \geq \mu_* B > \mu^*B - \epsilon,$$

and the lemma follows by letting $\epsilon \to 0$. ∎

Proposition 6 states that every set with finite outer measure is contained in an $\mathfrak{A}_{\sigma\delta}$ with the same outer measure. The following proposition is the analogue for inner measure.

**32. Proposition:** *Let $E$ be a set with $\mu_* E < \infty$. Then there is a set $H \in \mathfrak{A}_{\delta\sigma}$ such that $H \subset E$ and $\bar{\mu}H = \mu_* E$.*

***Proof:*** Let $A_n$ be a set in $\mathfrak{A}$ with $\mu^*(A_n \sim E) < \infty$ such that $\mu A_n - \mu^*(A_n \sim E) > \mu_* E - 1/n$. By Proposition 6 there is a $G_n \in \mathfrak{A}_{\sigma\delta}$ such that $G_n \supset A_n \sim E$ and $\bar{\mu}G_n = \mu^*(A_n \sim E)$. Let $H_n = A_n \sim G_n$. Then $H_n \in \mathfrak{A}_{\delta\sigma}$, and $H_n \subset E$. Moreover, $\bar{\mu}H_n = \mu A_n - \bar{\mu}G_n > \mu_* E - 1/n$. Thus the proposition is proved if we take $H = \bigcup H_n$. ∎

**33. Corollary:** *If $\mu_* E < \infty$, then*

$$\mu_* E = \sup \{\bar{\mu}B \colon B \subset E, B \text{ measurable, and } \bar{\mu}B < \infty\}.$$

**34. Proposition:** *Suppose $\mu^*E < \infty$. Then $E$ is measurable if and only if $\mu_* E = \mu^*E$.*

***Proof:*** Suppose $\mu_* E = \mu^*E < \infty$. Then Propositions 6 and 32 give us measurable sets $G$ and $H$ with $G \subset E \subset H$ and $\bar{\mu}H = \bar{\mu}G$. Thus $E$ differs from a measurable set by a set of measure zero and hence is measurable. The converse statement is just Lemma 31. ∎

**35. Theorem:** *Let $E$ and $F$ be disjoint sets. Then*

$$\mu_* E + \mu_* F \leq \mu_*(E \cup F) \leq \mu_* E + \mu^*F \leq \mu^*(E \cup F) \leq \mu^*E + \mu^*F.$$

***Proof:*** If $\mu_* E$ or $\mu_* F$ is infinite, the first inequality follows from the monotonicity of $\mu_*$. If $\mu_* E$ and $\mu_* F$ are both finite, let $G$ and $H$ be measurable sets with $G \subset E$ and $H \subset F$ such that $\bar{\mu}G = \mu_* E$ and $\bar{\mu}H = \mu_* F$. Then $G \cup H$ is a measurable set of finite outer measure contained in $E \cup F$. Thus

$$\begin{aligned}\mu_*(E \cup F) \geq \mu_*(G \cup H) &= \bar{\mu}(G \cup H)\\ &= \bar{\mu}G + \bar{\mu}H\\ &= \mu_* E + \mu_* F,\end{aligned}$$

proving the first inequality.

If $\mu^*F = \infty$, the second inequality is trivial. If $\mu^*F < \infty$, let $A$ be in $\mathfrak{A}$ with $\mu^*(A \sim (E \cup F)) < \infty$. Since $A \sim E \subset [A \sim (E \cup F)] \cup F$, we have

$$\mu^*(A \sim E) \leq \mu^*(A \sim (E \cup F)) + \mu^*F.$$

Thus $\mu^*(A \sim E) < \infty$, and

$$\begin{aligned}\mu A - \mu^*(A \sim (E \cup F)) &\leq \mu A - \mu^*(A \sim E) + \mu^*F\\ &\leq \mu_* E + \mu^*F.\end{aligned}$$

Taking the supremum over $A$, we get

$$\mu_*(E \cup F) \leq \mu_* E + \mu^*F.$$

To prove the third inequality, we choose a measurable set $G \subset E$ with $\bar{\mu}G = \mu_* E$. Then the measurability of $G$ implies that

$$\begin{aligned}\mu_* E + \mu^*F &= \bar{\mu}G + \mu^*F\\ &= \mu^*(G \cup F)\\ &\leq \mu^*(E \cup F).\end{aligned}$$

The last inequality is just the subadditivity of outer measure. ▮

**36. Corollary:** *If $\langle E_i \rangle$ is any disjoint sequence of sets, then*

$$\sum_{i=1}^{\infty} \mu_* E_i \leq \mu_*\left(\bigcup_{i=1}^{\infty} E_i\right).$$

***Proof:*** Set $E = \bigcup E_i$. Repeated application of the first inequality in Theorem 35 gives us

$$\begin{aligned}\sum_{i=1}^{n} \mu_* E_i &\leq \mu_*\left(\bigcup_{i=1}^{n} E_i\right)\\ &\leq \mu_* E.\end{aligned}$$

The corollary follows by letting $n$ tend to $\infty$. ▮

The following lemma expresses the additivity of $\mu_*$ on decompositions given by disjoint sequences of sets in $\mathfrak{A}$. The corresponding result for outer measure is true even if the sets $A_i$ are only assumed to be measurable (see Problems 2 and 46e).

**37. Lemma:** *Let $\langle A_i \rangle$ be a disjoint sequence of sets in $\mathfrak{A}$. Then*

$$\mu_*\left(E \cap \bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu_*(E \cap A_i).$$

***Proof:*** Since we may replace $E$ by $E \cap \bigcup_{i=1}^{\infty} A_i$, we may suppose that $E \subset \bigcup_{i=1}^{\infty} A_i = C$. Let $B \in \mathfrak{A}$ with $\mu^*(B \sim E) < \infty$. Since $C$ is $\mu^*$-measurable,

$$\mu B = \mu^*(B \cap C) + \mu^*(B \cap \tilde{C})$$

and

$$\begin{aligned}\mu^*(B \cap \tilde{E}) &= \mu^*(B \cap C \cap \tilde{E}) + \mu^*(B \cap \tilde{E} \cap \tilde{C}) \\ &= \mu^*(B \cap C \cap \tilde{E}) + \mu^*(B \cap \tilde{C}),\end{aligned}$$

since $\tilde{C} \subset \tilde{E}$. Thus $\mu^*(B \cap \tilde{C}) \leq \mu^*(B \cap \tilde{E}) < \infty$, and so

$$\begin{aligned}\mu B - \mu^*(B \sim E) &= \mu^*(B \cap C) - \mu^*(B \cap \tilde{E} \cap C) \\ &= \sum \mu(A_i \cap B) - \mu^*(B \cap \tilde{E} \cap C)\end{aligned}$$

by Problem 2. Hence

$$\mu B - \mu^*(B \sim E) \leq \sum \mu_*(A_i \cap E).$$

Taking the supremum over $B$ gives

$$\mu_* E \leq \sum \mu_*(A_i \cap E).$$

The opposite inequality follows from Corollary 36. ∎

**38. Theorem:** *Let $\mu$ be a measure on an algebra $\mathfrak{A}$ of subsets of $X$ and $E$ any subset of $X$. If $\mathfrak{B}$ is the algebra generated by $\mathfrak{A}$ and $E$ and if $\tilde{\mu}$ is any extension of $\mu$ to $\mathfrak{B}$, then*

$$\mu_* E \leq \tilde{\mu} E \leq \mu^* E.$$

*Moreover, there are extensions $\bar{\mu}$ and $\underline{\mu}$ of $\mu$ to $\mathfrak{B}$ (and hence also to the $\sigma$-algebra generated by $\mathfrak{B}$) such that $\bar{\mu} E = \mu^* E$ and $\underline{\mu} E = \mu_* E$.*

***Proof:*** If $\langle A_i \rangle$ is any disjoint sequence of sets from $\mathfrak{A}$ with $E \subset \bigcup A_i$, then $E = \bigcup (A_i \cap E)$, and so

$$\tilde{\mu}E = \sum_{i=1}^{\infty} \tilde{\mu}(A_i \cap E) \leq \sum_{i=1}^{\infty} \mu A_i .$$

Thus $\tilde{\mu}E \leq \mu^*E$.

If $A$ is any set in $\mathfrak{A}$ with $\mu^*(A \sim E) < \infty$, then $\tilde{\mu}(A \sim E) \leq \mu^*(A \sim E)$, and

$$\begin{aligned} \mu A - \mu^*(A \sim E) \leq \mu A - \tilde{\mu}(A \sim E) &= \tilde{\mu}(E \cap A) \\ &\leq \tilde{\mu}E. \end{aligned}$$

Hence $\mu_* E \leq \tilde{\mu}E$.

The sets $B$ in $\mathfrak{B}$ are the sets of the form $B = (A \cap E) \cup (A' \cap \tilde{E})$ with $A$ and $A'$ in $\mathfrak{A}$, since the collection of all sets of this form is an algebra contained in $\mathfrak{B}$ and containing $\mathfrak{A}$ and $E$. For each $B \in \mathfrak{B}$ define $\bar{\mu}$ and $\underline{\mu}$ by

$$\bar{\mu}B = \mu^*(B \cap E) + \mu_*(B \cap \tilde{E})$$

and

$$\underline{\mu}B = \mu_*(B \cap E) + \mu^*(B \cap \tilde{E}).$$

Then $\bar{\mu}$ and $\underline{\mu}$ are monotone, nonnegative functions defined on $\mathfrak{B}$, and it follows from Corollary 30 that $\bar{\mu}A = \underline{\mu}A = \mu A$ for $A \in \mathfrak{A}$. Thus the theorem will be proved if we show that $\bar{\mu}$ and $\underline{\mu}$ are countably additive on $\mathfrak{B}$. But this follows easily from Problem 2 and Lemma 37. ▮

## Problems

**39. a.** If $\mu X < \infty$, then $\mu_* E = \mu X - \mu^*(\tilde{E})$.

**b.** If $\mathfrak{A}$ is a $\sigma$-algebra, then

$$\mu^*E = \inf \{\mu A \colon E \subset A,\ A \in \mathfrak{A}\}$$

and

$$\mu_* E = \sup \{\mu A \colon A \subset E,\ A \in \mathfrak{A}\}.$$

**c.** If $\mu$ is Lebesgue measure on $\mathbf{R}$, then $\mu_* E = \sup \{\mu F \colon F \subset E,\ F \text{ closed}\}$.

**40.** Show that the measures $\bar{\mu}$ and $\underline{\mu}$ in Theorem 38 are countable additive on $\mathfrak{B}$.

**41.** Let $\mu$ be a measure on an algebra $\mathfrak{A}$, and let $E$ be a $\mu^*$ measurable set. Show that the measure $\bar{\mu}$ defined in Theorem 38 has the property that $\bar{\mu}B = \mu^*B$ for all $B \in \mathfrak{B}$. Thus $\bar{\mu}$ agrees on $\mathfrak{B}$ with the measure given by the Carathéodory extension process.

**42.** Let $E$ be any set. A measurable set $G \subset E$ is called a *measurable kernel* for $E$ if $\mu_*(E \sim G) = 0$. A measurable set $H \supset E$ is called a *measurable cover* for $E$ if $\mu_*(H \sim E) = 0$.

**a.** Show that any two measurable kernels (or covers) for $E$ differ by a null set.

**b.** Show that if $E$ is a set of $\sigma$-finite outer measure, then $E$ has a measurable kernel and a measurable cover.

**43. a.** Let $P$ be the nonmeasurable set constructed in Section 4 of Chapter 3. Show that $\mu_* P = 0$ (cf. Problem 3.15).

**b.** Let $E = [0, 1] \sim P$. Then $\mu^*E = 1$, and if $A \cap [0, 1]$ is a measurable set, then $m^*(A \cap E) = mE$.

**c.** If $A \subset [0, 1]$ is a measurable set, then the measurable hull of $A \cap E$ is $A$.

**44.** Let $\mu$ be a measure on an algebra $\mathfrak{A}$ and $E$ a set with $\mu^*E < \infty$. If $\beta$ is any real number with $\mu_* E \leq \beta \leq \mu^*E$, then there is an extension $\tilde{\mu}$ of $\mu$ to a $\sigma$-algebra containing $\mathfrak{A}$ and $E$ such that $\tilde{\mu}E = \beta$.

**45. a.** If $\mu$ is a semifinite measure on an algebra $\mathfrak{A}$ and if $\mathfrak{B}$ is the smallest $\sigma$-algebra containing $\mathfrak{A}$, then $\mu_*$ restricted to $\mathfrak{B}$ is a semifinite measure and is the smallest extension of $\mu$ to $\mathfrak{B}$.

**b.** There may be more than one semifinite extension to $\mathfrak{B}$ (for example, let $X = \mathbf{N} \cup \{\omega\}$ and $\mathfrak{A}$ the algebra generated by the finite subsets of $\mathbf{N}$ with $\mu$ the continuing measure).

**46.** Let $\mathfrak{A}$ be the algebra of finite unions of half-open intervals of $\mathbf{R}$ and let $\mu\emptyset = 0$ and $\mu A = \infty$ for $A \neq 0$. The class $\mathfrak{B}$ of Borel sets is the smallest $\sigma$-algebra containing $\mathfrak{A}$.

**a.** Show that $\mu^*E = \infty$ if $E \neq \emptyset$.

**b.** Show that $\mu_* E = 0$ if $E$ contains no interval and that $\mu_* E = \infty$ if $E$ contains an interval.

**c.** The restriction of $\mu_*$ to $\mathfrak{B}$ is not a measure. Hence there is no smallest extension of $\mu$ to $\mathfrak{B}$.

**d.** The counting measure on $\mathfrak{B}$ is an extension of $\mu$ to $\mathfrak{B}$.

**e.** Show that Lemma 37 is no longer true if we replace "sets in $\mathfrak{A}$" by "measurable sets".

**47.** If $\mu^*$ is not a regular outer measure (that is, if it does not come from a measure on an algebra), then we do not get a reasonable theory of inner measure by setting $\mu_* E = \mu^*X - \mu^*(\tilde{E})$. Let $X = \{a, b, c\}$, and set

$\mu^*X = 2$, $\mu^*\emptyset = 0$, and $\mu^*E = 1$ if $E$ is not $X$ or $\emptyset$.

**a.** Calculate $\mu_*$.

**b.** What are the measurable subsets of $X$?

**c.** Show that there is a nonmeasurable set $E$ with $\mu_*E = \mu^*E$.

**d.** Show that the first and third inequalities of Theorem 35 fail.

**48.** Let $X = \mathbf{R}^2$ and $\mathfrak{A}$ the algebra consisting of all disjoint unions of vertical intervals of the form $I = \{\langle x, y\rangle : a < y \leq b\}$. Let $\mu A$ be the sum of the lengths of the intervals of which $A$ is composed. Then $\mu$ is a measure on $\mathfrak{A}$. Let $E = \{\langle x, y\rangle : y = 0\}$.

**a.** Show that $\mu^*E = \infty$, $\mu_* E = 0$.

**b.** Show that every subset of $E$ is an $\mathfrak{A}_\delta$.

**c.** Assuming that there is no finite nonatomic measure defined on $\mathcal{P}(\mathbf{R})$, show that each extension of $\mu$ to a measure on a $\sigma$-algebra must assign to $E$ the value 0 or $\infty$.

## ★7 Extension by Sets of Measure Zero

The results of Section 2 allow us to extend a measure $\mu$ on an algebra $\mathfrak{A}$ to a $\sigma$-algebra containing $\mathfrak{A}$ and those of Section 6 provide for the extension from $\mathfrak{A}$ to a $\sigma$-algebra containing $\mathfrak{A}$ and one additional set. It is sometimes useful to be able to extend to a $\sigma$-algebra containing $\mathfrak{A}$ and some collection $\mathfrak{M}$ of subsets of $X$ and to extend in such a way that each of the sets in $\mathfrak{M}$ has measure zero. A necessary condition for this to be possible is that whenever we have a set $A \in \mathfrak{A}$ such that $A \subset M \in \mathfrak{M}$, then $\mu A = 0$. This condition is not in general sufficient, since a countable union of sets in $\mathfrak{M}$ may contain an $A$ with positive measure, but, if we assume that $\mathfrak{M}$ is closed under countable unions, then the condition is sufficient.

**39. Proposition:** *Let $\mu$ be a measure on a $\sigma$-algebra $\mathfrak{A}$ of subsets of $X$, and let $\mathfrak{M}$ be a collection of subsets of $X$ which is closed under countable unions and which has the property that for each $A \in \mathfrak{A}$ with $A \subset M \in \mathfrak{M}$ we have $\mu A = 0$. Then there is an extension $\bar{\mu}$ of $\mu$ to the smallest $\sigma$-algebra $\mathfrak{B}$ containing $\mathfrak{A}$ and $\mathfrak{M}$ such that $\bar{\mu}M = 0$ for each $M \in \mathfrak{M}$.*

***Proof:*** Since the collection of sets which are subsets of a set in $\mathfrak{M}$ satisfies the same hypothesis as $\mathfrak{M}$, we may assume that each subset of a set in $\mathfrak{M}$ is itself in $\mathfrak{M}$. With this assumption the collection

$\mathcal{B} = \{B: B = A \,\Delta\, M,\ A \in \mathcal{A},\ M \in \mathfrak{M}\}$ is a $\sigma$-algebra containing $\mathcal{A}$ and $\mathfrak{M}$, and since each $\sigma$-algebra containing $\mathcal{A}$ and $\mathfrak{M}$ contains $\mathcal{B}$, $\mathcal{B}$ is the smallest $\sigma$-algebra containing $\mathcal{A}$ and $\mathfrak{M}$.

If $B = A_1 \,\Delta\, M_1 = A_2 \,\Delta\, M_2$, then $A_1 \,\Delta\, A_2 = M_1 \,\Delta\, M_2$, and so $\mu(A_1 \,\Delta\, A_2) = 0$. Thus $\mu A_1 = \mu A_2$, and, if we define $\bar{\mu} B$ to be $\mu A_1$, then $\bar{\mu}$ is well defined on $\mathcal{B}$ and is an extension of $\mu$. It remains only to show that $\bar{\mu}$ is countably additive.

Let $B = \bigcup B_i$, $B_i \cap B_j = \emptyset$. If $B_i = A_i \,\Delta\, M_i$, then $A_i \,\Delta\, A_j \in \mathfrak{M}$. Setting $A'_n = A_n \cap \tilde{A}_1 \cap \cdots \cap \tilde{A}_{n-1}$, we have $A'_i \cap A'_j = \emptyset$, and $A_n \,\Delta\, A'_n \in \mathfrak{M}$. Thus $B_i = A'_i \,\Delta\, M'_i$, and $B = A \,\Delta\, M$, where $A = \bigcup A'_i$ and $M \subset \bigcup M'_i$. Thus $\bar{\mu} B = \mu A = \sum \mu A'_i = \sum \bar{\mu} B_i$. ∎

We observe that the condition that $\mu A = 0$ for each $A \in \mathcal{A}$ with $A \subset M$ simply states that $\mu_* M = 0$. Thus the proposition states that we can extend the domain of $\mu$ to include any collection $\mathfrak{M}$ of sets of inner measure zero provided that $\mathfrak{M}$ is closed under countable unions. Note that on the $\sigma$-algebra generated by $\mathcal{A}$ and $\mathfrak{M}$ we have $\bar{\mu} = \mu_*$. Thus this proposition gives a generalization of the process of completion which extends the domain of a measure by adding sets of outer measure zero.

## Problems

**49.** Let $\mathcal{A}$ be a $\sigma$-algebra on $X$ and $\mathfrak{M}$ a collection of subsets of $X$ which is closed under countable unions and which has the property that each subset of a set in $\mathfrak{M}$ is in $\mathfrak{M}$. Show that the collection

$$\mathcal{B} = \{B: B = A \,\Delta\, M,\ A \in \mathcal{A} \text{ and } M \in \mathfrak{M}\}$$

is a $\sigma$-algebra.

**50.** Show that Proposition 39 need not be true if $\mathcal{A}$ is only an algebra of sets.

## 8 Carathéodory Outer Measure

Let $X$ be a set of points and $\Gamma$ a set of real-valued functions on $X$. It is often of interest to know conditions under which an outer measure $\mu^*$ will have the property that every function in $\Gamma$ will be measurable, and it is the purpose of the present section to prove the sufficiency of one criterion for this.

Two sets are said to be *separated by the function* $\varphi$ if there are numbers $a$ and $b$ with $a > b$ such that $\varphi$ is greater than $a$ on one and less than $b$ on the other. An outer measure $\mu^*$ is called a **Carathéodory outer measure** with respect to $\Gamma$ if it satisfies the following axiom:

iv. *If $A$ and $B$ are two sets which are separated by some function in $\Gamma$, then $\mu^*(A \cup B) = \mu^*A + \mu^*B$.*

**40. Proposition:** *If $\mu^*$ is a Carathéodory outer measure with respect to $\Gamma$, then every function in $\Gamma$ is $\mu^*$-measurable.*

***Proof:*** Given the real number $a$ and the function $\varphi \in \Gamma$, we must show that the set

$$E = \{x\colon \varphi(x) > a\}$$

is $\mu^*$-measurable or, equivalently, that given any set $A$,

$$\mu^*A \geq \mu^*(A \cap E) + \mu^*(A \cap \tilde{E}).$$

Since this inequality is trivial if $\mu^*A = \infty$, we suppose $\mu^*A < \infty$.

We begin by setting $B = E \cap A$, $C = \tilde{E} \cap A$, and

$$B_n = \left\{x\colon (x \in B) \;\&\; \left(\varphi(x) > a + \frac{1}{n}\right)\right\}.$$

Defining $R_n = B_n \sim B_{n-1}$, 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 $\varphi$ separates $R_n$ and $B_{n-2}$ and hence separates $R_{2k}$ and $\bigcup_{j=1}^{k-1} R_{2j}$, since the latter set is contained in $B_{2k-2}$. Consequently,

$$\mu^*\left[\bigcup_{j=1}^{k} R_{2j}\right] = \mu^*R_{2k} + \mu^*\left[\bigcup_{j=1}^{k-1} R_{2j}\right]$$

$$= \sum_{j=1}^{k} \mu^*R_{2j},$$

by induction. Since

$$\sum_{j=1}^{k} R_{2j} \subset B \subset A,$$

we have

$$\sum_{j=1}^{k} \mu^* R_{2j} \le \mu^* A,$$

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 therefore also the series

$$\sum_{k=1}^{\infty} \mu^* R_k .$$

From this it follows that, given $\epsilon > 0$, we can choose $n$ so large that

$$\sum_{k=n}^{\infty} \mu^* R_k < \epsilon.$$

Then by the subadditivity of $\mu^*$,

$$\begin{aligned} \mu^* B &\le \mu^* B_n + \sum_{k=n+1}^{\infty} \mu^* R_k \\ &< \mu^* B_n + \epsilon \end{aligned}$$

or

$$\mu^* B_n > \mu^* B - \epsilon.$$

Now

$$\begin{aligned} \mu^* A &\ge \mu^*(B_n \cup C) \\ &= \mu^* B_n + \mu^* C \end{aligned}$$

since $\varphi$ separates $B_n$ and $C$. Consequently,

$$\mu^* A \ge \mu^* B + \mu^* C - \epsilon.$$

Since $\epsilon$ is an arbitrary positive quantity,

$$\mu^* A \le \mu^* B + \mu^* C. \quad \blacksquare$$

**41. Proposition:** *Let $(X, \rho)$ be a metric space, and let $\mu^*$ be an outer measure on $X$ with the property that $\mu^*(A \cup B) = \mu^* A + \mu^* B$ whenever $\rho(A, B) > 0$. Then every closed set (and hence every Borel set) is measurable with respect to $\mu^*$.*

An outer measure $\mu^*$ on subsets of a metric space $X$ with the property that $\mu^*(A \cup B) = \mu^*A + \mu^*B$ whenever $\rho(A, B) > 0$ is called a *Carathéodory outer measure* for $X$ or a metric outer measure.

## Problems

**51.** Prove Proposition 41. [Let $\Gamma$ be the set of functions $\varphi$ of the form $\varphi(x) = \rho(x, E)$. Show that $\mu^*$ satisfies (iv) with respect to $\Gamma$, and note that for a closed set $F$ we have $F = \{x: \rho(x, F) \leq 0\}$.]

## 9 Hausdorff Measures

By a *Borel measure* on a metric space $X$ we mean a measure that is defined on some $\sigma$-algebra containing the $\sigma$-algebra of Borel sets in $X$. For each positive real number $\alpha$ we will define a particular Borel measure $m_\alpha$ called the Hausdorff measure on $X$ of dimension $\alpha$. These measures are particularly important for the Euclidean spaces $\mathbf{R}^n$, but much of their theory goes through just as easily for an arbitrary metric space $X$.

To define $m_\alpha$, we take $\epsilon > 0$ and set

$$\lambda_\alpha^{(\epsilon)}(E) = \inf \sum_{i=1}^{\infty} r_i^\alpha,$$

where the $\langle r_i \rangle$ are the radii of a sequence of balls $\langle B_i \rangle$ that cover $E$ and for which $r_i < \epsilon$. Observe that $\lambda_\alpha^{(\epsilon)}$ increases as $\epsilon$ decreases. Set

$$m_\alpha^* E = \sup \lambda_\alpha^{(\epsilon)}(E)$$

as $\epsilon \to 0$. Then we have

$$m_\alpha^* E = \lim_{\epsilon \to 0} \lambda_\alpha^{(\epsilon)}(E).$$

It is readily verified that $m_\alpha^*$ is countably subadditive and thus an outer measure. If $E$ and $F$ are two subsets of $X$ with $\rho(E, F) > \delta$, then

$$\lambda_\alpha^{(\epsilon)}(E \cup F) \geq \lambda_\alpha^{(\epsilon)}(E) + \lambda_\alpha^{(\epsilon)}(F)$$

as soon as $\epsilon < \delta$: For if $\langle B_i \rangle$ is a sequence of balls of radii less than $\epsilon$ covering $E \cup F$, no ball can meet both $E$ and $F$. Taking limits as

$\epsilon \to 0$, we have

$$m_\alpha^*(E \cup F) \geq m_\alpha^* E + m_\alpha^* F.$$

Thus $m_\alpha^*$ induces a Borel measure $m_\alpha$ on $X$ by Proposition 41. The measure $m_\alpha$ is called Hausdorff $\alpha$-dimensional measure.

It is customary to normalize $m_\alpha$ by dividing by the quantity

$$\pi_\alpha = \frac{2\pi^{\alpha/2}}{\alpha\Gamma\left(\dfrac{\alpha}{2}\right)}.$$

Thus $\pi_1 = 2$, $\pi_2 = \pi$, $\pi_3 = 4\pi/3$, and $\pi_n$ is the volume of the unit ball in $\mathbf{R}^n$. We refer to this measure as normalized Hausdorff measure. In $\mathbf{R}^n$ the normalized Hausdorff measure $m_n$ is equal to Lebesgue measure.

## Problems

**52.** Show that $\mu_\alpha^*$ is countably subadditive.

**53.** **a.** Show that the Hausdorff measure $m_\alpha$ in $\mathbf{R}^n$ are translation invariant; that is, if $F = E + y$, $E$ a Borel set, then $m_\alpha F = m_\alpha E$.

**b.** Show that in $\mathbf{R}^n$ the measures $m_\alpha$ are invariant under Euclidean motions (rotations and translations).

**54.** Let $E$ be the one-to-one continuous image of $[0, 1]$ in $\mathrm{R}^n$ (or a metric space). Then the normalized one-dimensional Hausdorff measure of $E$ is its length.

**55.** Let $E$ be a Borel subset of some metric space $X$.

**a.** Show that if $m_\alpha E$ is finite for some $\alpha$, then $m_\beta E = 0$ for all $\beta$ with $0 < \beta < \alpha$.

**b.** Show that if $m_\alpha E > 0$ for some $\alpha$, then $m_\beta E = \infty$ for all $\beta > \alpha$.

**c.** The Hausdorff dimension of $E$ is defined by

$$\dim_{\text{Haus}} E = \inf \{\alpha \colon m_\alpha E > \infty\}$$

or

$$\dim_{\text{Haus}} E = \sup \{\beta \colon m_\beta E = 0\}.$$

Show that the two definitions are the same.

**d.** Show that the Hausdorff dimension of the Cantor ternary set is $\log 2/\log 3$.

# 13 Measure and Topology

We are often concerned with measures on a set $X$ which is also a topological space, and it is natural to consider conditions on the measure so that it is connected with the topological structure. There seem to be two classes of topological spaces for which it is possible to carry out a reasonable theory. One is the class of locally compact Hausdorff spaces, and the present chapter develops the theory for this class. The other is the class of complete metric spaces, and some of the significance of this class for measure theory is discussed in Chapter 15.

## 1 Baire Sets and Borel Sets

Let $X$ be a locally compact Hausdorff space. From the point of view of integration theory the most useful family of functions on $X$ is the family $C_c(X)$ consisting of all continuous real-valued functions that vanish outside a compact subset of $X$. If $f$ is a real-valued function, the *support* of $f$ is the closure of the set $\{x : f(x) \neq 0\}$. Thus $C_c(X)$ is the class of all continuous real-valued functions on $X$ with compact support. The class of **Baire sets** is defined to be the smallest $\sigma$-algebra $\mathcal{B}$ of subsets of $X$ such that each function in $C_c(X)$ is measurable with respect to $\mathcal{B}$. Thus $\mathcal{B}$ is the $\sigma$-algebra generated by the sets $\{x : f(x) \geq \alpha\}$ with $f \in C_c(X)$. If $\alpha > 0$, these sets are compact $G_\delta$'s. It follows from Proposition 9.15 that each compact $G_\delta$ is a Baire set. Consequently, $\mathcal{B}$ is the $\sigma$-algebra generated by the compact $G_\delta$'s.

If $X$ is any topological space, the smallest $\sigma$-algebra containing the closed sets is called the class of **Borel sets.** Thus, if $X$ is locally compact, every Baire set is a Borel set. The converse is true when $X$ is a locally compact separable metric space, but there are compact spaces where the class of Borel sets is larger than the class of Baire sets (Problem 6).

We shall often denote the classes of Baire sets and of Borel sets on $X$ by $\mathcal{B}a(X)$ and $\mathcal{B}o(X)$.

Throughout the remainder of this chapter we always assume that $X$ is a locally compact Hausdorff space.

By a **Baire measure** on $X$ we mean a measure defined for all Baire sets and finite for each compact Baire set. Since we prefer to integrate with respect to complete measures, we shall often deal with the completion of such a measure, which we again call a Baire measure. By a **Borel measure** we mean a measure defined on the $\sigma$-algebra of Borel sets or the completion of such a measure. All Borel measures considered in this chapter will also be assumed to have finite values on compact sets.

A set $E$ in a locally compact Hausdorff space is said to be (topologically) **bounded** if $E$ is contained in some compact set, i.e., if $\bar{E}$ is compact. This notion should be distinguished from the concept of (metric) boundedness for subsets of a metric space. The two concepts agree for subsets of $\mathbf{R}^n$ and more generally for any metric space for which the metric is proper (see Problem 9.31). A set $E$ is said to be **$\sigma$-bounded** if it is the union of a countable collection of bounded sets. Recall that a set is $\sigma$-compact if it is the countable union of compact sets. It is easy to see that a set $E$ is $\sigma$-bounded if and only if it is contained in a $\sigma$-compact set.

We state a number of lemmas that are useful in dealing with Baire and Borel sets. The proofs of many of them are left to the reader.

**1. Lemma:** *Let $K$ be a compact set, $O$ an open set with $K \subset O$. Then*

$$K \subset U \subset H \subset O.$$

*where $U$ is a $\sigma$-compact open set and $H$ is a compact $G_\delta$.*

**2. Lemma:** *Every $\sigma$-compact open set is the union of a countable collection of compact $G_\delta$'s and hence a Baire set.*

**3. Lemma:** *Every bounded set is contained in a compact $G_\delta$. Every $\sigma$-bounded set $E$ is contained in a $\sigma$-compact open set $O$. If $E$ is bounded, we may take $\bar{O}$ to be compact.*

**4. Lemma:** *Let $\mathcal{R}$ be a ring of sets, and let $\mathcal{R}' = \{E: \tilde{E} \in \mathcal{R}\}$. Then either $\mathcal{R} = \mathcal{R}'$ or else $\mathcal{R} \cap \mathcal{R}' = 0$. In the latter case $\mathcal{R} \cup \mathcal{R}'$ is the smallest algebra containing $\mathcal{R}$. If $\mathcal{R}$ is a $\sigma$-ring, then $\mathcal{R} \cup \mathcal{R}'$ is a $\sigma$-algebra.*

This is essentially Problem 11.9a. The reader should study the material developed in Problem 11.9 for use in this chapter.

**5. Lemma:** *If $E$ is a Baire set, then $E$ or $\tilde{E}$ is $\sigma$-bounded. Both are $\sigma$-bounded if and only if $X$ is $\sigma$-compact.*

**6. Lemma:** *The class of $\sigma$-bounded Baire sets is the smallest $\sigma$-ring containing the compact $G_\delta$'s.*

**7. Lemma:** *Each $\sigma$-bounded Baire set is the union of a countable disjoint union of bounded Baire sets.*

The following Proposition gives us a useful means of applying theorems about Baire and Borel sets in compact spaces to bounded Baire and Borel sets in locally compact spaces.

**8. Proposition:** *Let $F$ be a closed subset of $X$. Then $F$ is a locally compact Hausdorff space, and the Baire sets of $F$ are those sets of the form $B \cap F$, where $B$ is a Baire set in $X$. Thus if $F$ is a closed Baire set, the Baire subsets of $F$ are just those Baire subsets of $X$ which are contained in $F$. The Borel sets of $F$ are those Borel sets of $X$ which are contained in $F$.*

***Proof:*** Let $\mathcal{A} = \{E: E = B \cap F,\ B \in \mathcal{B}a(X)\}$. Then $\mathcal{A}$ is a $\sigma$-algebra which includes all compact $G_\delta$'s contained in $F$. Thus $\mathcal{B}a(F) \subset \mathcal{A}$, and each Baire set of $F$ is of the form $B \cap F$. Let $\mathcal{B} = \{E \subset X: E \cap F \in \mathcal{B}a(F)\}$. Then $\mathcal{B}$ is a $\sigma$-algebra. Let $K$ be a compact $G_\delta$ in $X$. Then $K \cap F$ is a closed subset of $K$ and hence compact. Since $K$ is a $G_\delta$ in $X$, $K \cap F$ is a $G_\delta$ in $F$. Thus $K \cap F$ is a compact $G_\delta$ of $F$ and so is in $\mathcal{B}a(F)$. Consequently, $\mathcal{B}a(X) \subset \mathcal{B}$, and so each Baire set of $X$ intersects $F$ in a Baire set of $F$.

If $F$ is a closed Baire subset of $X$, then $B \cap F$ is a Baire subset of $X$ whenever $B$ is. Thus each Baire subset of $F$ is of this form. On the other hand, for each Baire subset $B$ of $X$ with $B \subset F$ we have $B = B \cap F$, and so $B$ is a Baire subset of $F$.

The statement about Borel sets is established similarly, using $\mathcal{B}\mathrm{o}$ instead of $\mathcal{B}\mathrm{a}$, and using the fact that $F$ is a Borel set. ∎

The following lemma is useful in showing that a set $E$ is a Borel set.

**9. Lemma:** *Let $E$ be a subset of $X$ such that $E \cap K$ is a Borel set for each compact set $K$. Then $E$ is a Borel set.*

The reader should be warned that the terminology regarding Baire and Borel sets is not entirely standardized. Some authors take the class of Baire sets to be the smallest $\sigma$-algebra $\mathcal{B}\mathrm{c}(X)$ such that all continuous real-valued functions on $X$ are measurable. Others restrict the class of Borel sets to be the smallest $\sigma$-algebra $\mathcal{B}\mathrm{k}(X)$ which contains the compact sets. Authors (such as Halmos [5]) who do measure theory on $\sigma$-rings often take the Baire sets to be the smallest $\sigma$-ring $\mathcal{R}$ containing the compact $G_\delta$'s and the Borel sets to be the smallest $\sigma$-ring $\mathcal{S}$ containing the compact sets. In reading works dealing with Baire and Borel sets or measures, it is imperative to check carefully what the author means by a Baire or a Borel set. A given statement may be true for one usage and false for another.

Some of the inclusion relations between the various classes are given by the following formula:

$$\begin{tikzcd}[column sep=small, row sep=small]
 & & \mathcal{B}\mathrm{c} \arrow[dr, phantom, sloped, "\subset"] & \\
\mathcal{R} \subset & \mathcal{B}\mathrm{a} \arrow[ur, phantom, sloped, "\subset"] \arrow[dr, phantom, sloped, "\subset"] & & \mathcal{B}\mathrm{o} \\
 & & \mathcal{B}\mathrm{k} \arrow[ur, phantom, sloped, "\subset"] &
\end{tikzcd}$$

We have $\mathcal{B}\mathrm{a} = \mathcal{B}\mathrm{c} \cap \mathcal{B}\mathrm{k}$, and $\mathcal{B}\mathrm{o}$ is the smallest $\sigma$-algebra containing $\mathcal{B}\mathrm{c}$ and $\mathcal{B}\mathrm{k}$. The $\sigma$-ring $\mathcal{R}$ is precisely the $\sigma$-ring of $\sigma$-bounded Baire sets, and $\mathcal{S}$ the class of $\sigma$-bounded Borel sets. If $X$ is compact (or $\sigma$-compact and locally compact), then

$$\mathcal{R} = \mathcal{B}\mathrm{a} = \mathcal{B}\mathrm{c} \quad \text{and} \quad \mathcal{S} = \mathcal{B}\mathrm{k} = \mathcal{B}\mathrm{o}.$$

If $X$ is metrizable (or even locally metrizable like the manifolds), we have

$$\mathcal{R} = \mathcal{S}, \qquad \mathcal{B}\mathrm{a} = \mathcal{B}\mathrm{k}, \quad \text{and} \quad \mathcal{B}\mathrm{c} = \mathcal{B}\mathrm{o}.$$

If $X$ is a locally compact separable metric space, then all classes coincide.

The definition adopted here for the class of Borel sets seems to me to be the most useful and convenient one. It is widely used and is standard for metric spaces. Our definition of the Baire sets, on the other hand, is probably not quite as convenient as the definition that takes them to be the smallest $\sigma$-ring containing the compact $G_\delta$'s. We have not done so here because we chose to do measure theory on $\sigma$-algebras rather than $\sigma$-rings. One price we pay is the innumerable references to the class of $\sigma$-bounded Baire sets. Moreover, the Baire measures are extremely well behaved on the $\sigma$-bounded Baire sets. Our complications with them in the next section arise because we insist on defining them on Baire sets that are not $\sigma$-bounded, sets on which they have no business being defined.

## Problems

**1.** Let $X$ be a separable locally compact metric space. Show that the class of Baire sets is the same as the class of Borel sets.

**2. a.** Prove Lemma 1. (See Proposition 9.15.)

**b.** Prove Lemma 2.

**c.** Prove Lemma 3.

**3. a.** Prove Lemma 5.

**b.** Prove Lemma 6.

**c.** Prove Lemma 7. (Lemma 2 is useful.)

**4. a.** Let $X$ be a locally compact Hausdorff space, and $F$ a set whose intersection with each compact subset of $X$ is closed. Then $F$ is closed. [Hint: Use contraposition.]

**b.** Prove Lemma 9.

**5.** Let $X$ be a locally compact Hausdorff space, and let $C_0(X)$ be the space of all uniform limits of functions in $C_c(X)$.

**a.** 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 $\{x: |f(x)| \geq \alpha\}$ is compact.

**b.** Let $X^*$ be the one-point compactification of $X$. Then $C_0(X)$ consists precisely of the restrictions to $X$ of those functions in $C(X^*)$ which vanish at $\infty$.

**c.** If $B$ is a Baire set in $X^*$, then $B \cap X$ is a Baire set in $X$.

**6.** Let $X$ be an uncountable set with the discrete topology.

**a.** What are $C_c(X)$ and $C_0(X)$?

**b.** What are the Baire sets in $X$?

**c.** Let $X^*$ be the one-point compactification of $X$. What is $C(X^*)$?

**d.** What are the Baire subsets of $X^*$?

**e.** Let $\omega$ be the "point at infinity" in $X^*$. Then $\{\omega\}$ is a compact Borel set which contains no nonempty Baire set.

**f.** Show that there is a Baire measure $\mu$ on $X$ such that $\mu(X) = 1$ and $f\, d\mu = 0$ for each $f$ in $C_0(X)$.

**7.** Let $X = \mathbf{R}^2$ with the topology generated by a base of sets of the form

$$\{\langle x, y\rangle \varepsilon \mathbf{R}^2: a < x < b, y = c\}.$$

Then $X$ is the direct union of the horizontal lines, and $X = \mathbf{R} \times Y$, where $Y$ is just $\mathbf{R}$ with the discrete topology. What are the Baire sets of $X$? What are the Borel sets of $X$?

**8.** Let $Y$ be an uncountable set with the discrete topology, $Z$ the one-point compactification of $Y$. What are the Baire sets of $X = Y \times Z$? What are the Borel sets?

**9. a.** Let $O$ be an open subset of a locally compact Hausdorff space. Then $O$ is a locally compact Hausdorff space and every Baire set $E$ of $O$ is of the form $O \cap$ B, where $B$ is a Baire set of $X$.

**b.** Not every set $O \cap B$ with $B$ a Baire set of $X$ is a Baire set of $O$. (Let $Y$ and $Z$ be uncountable sets with the discrete topology, $Y^*$ and $Z^*$ their one-point compactifications, $X$ the direct union of $Y^*$ and $Z^*$, and $O$ the direct union of $Y$ and $Z$. Let $B = Z^*$. Then $B$ is a compact open set of $X$, but $O \cap B$ is not a Baire subset of $O$.)

**c.** Let $O$ be a $\sigma$-compact open subset of $X$. Then the Baire subsets of $O$ are just the Baire subsets of $X$ that are contained in $O$.

**10. a.** Show that the relations between the classes $\mathfrak{R}$, $\mathfrak{B}$a, $\mathfrak{B}$c, $\mathfrak{B}$k, $\mathfrak{B}$o, and $\mathcal{S}$ described at the end of the section are correct.

**b.** Give examples where the classes are all distinct.

**11.** Let $X$ and $Y$ be two locally compact Hausdorff spaces.

**a.** Show that each $f \varepsilon C_c(X \times Y)$ is the limit of sums of the form

$$\sum_{i=1}^{n} \varphi_i(x)\psi_i(y)$$

where $\varphi_i \varepsilon C_c(X)$ and $\psi_i = C_c(Y)$. [The Stone–Weierstrass Theorem is useful.]

**b.** Show that $\mathfrak{B}a(X \times Y) \subset \mathfrak{B}a(X) \times \mathfrak{B}a(Y)$.

**c.** We have $\mathfrak{B}a(X \times Y) = \mathfrak{B}a(X) \times \mathfrak{B}a(Y)$ if and only if $X$ or $Y$ is $\sigma$-compact.

**d.** Let $X$ be a set with more than $\mathfrak{c}$ elements, where $\mathfrak{c}$ is the cardinal of $\mathbf{R}$, and let $X$ have the discrete topology. Set $Z = X \times X$. Then $Z$ has the discrete topology and every subset is a Borel set. Show that $D = \{\langle x, y\rangle \in Z: x = y\}$ is not in the $\sigma$-algebra $\mathcal{B}o(X) \times \mathcal{B}o(X)$. [Problem 1.20 is useful.]

**12. a.** For a map $f: X \to Y$ and collection $\mathcal{C}$ of subsets of $Y$ we define $f^*\mathcal{C}$ to be the collection of subsets of $X$ given by

$$f^*\mathcal{C} = \{E: E = f^{-1}[C] \text{ for some } C \in \mathcal{C}\}.$$

Show that if $\mathcal{A}$ is the $\sigma$-algebra generated by $\mathcal{C}$, then $f^*\mathcal{A}$ is the $\sigma$-algebra generated by $f^*\mathcal{C}$.

**b.** Given $f: X \to Y$ and a collection $\mathcal{C}$ of subsets of $X$, let $\mathcal{A}$ be the $\sigma$-algebra generated by $\mathcal{C}$. If $f^{-1}[f[C]] = C$ for each $C \in \mathcal{C}$, then $f^{-1}[f[A]] = A$ for each $A \in \mathcal{A}$.

**c.** Let $X$ be a locally compact Hausdorff space and $K$ a Baire subset of $X$ which is compact. Then there is a sequence $\langle K_n\rangle$ of compact $G_\delta$'s such that $K$ belongs to the $\sigma$-algebra $\mathcal{A}$ generated by the $\langle K_n\rangle$. [See Problem 1.9]

**d.** Let $X$, $K$, and $\langle K_n\rangle$ be as in (c). Show that there is a continuous map $f$ of $X$ into the metric space $I^\omega$ such that $f^{-1}[f[K_n]] = K_n$ for each $n$. [See Proposition 9.15.]

**e.** Prove that a Baire subset $K$ of $X$ which is compact is a $G_\delta$. [Note that $f[K]$ is a closed subset of a metric space.]

**f.** Prove that a Baire subset of $X$ which is open is also an $\mathcal{F}_\sigma$.

## 2 The Regularity of Baire and Borel Measures

Let $\mu$ be a measure defined on a $\sigma$-algebra $\mathcal{M}$ of subsets of $X$ and suppose that $\mathcal{M}$ contains the Baire sets. A set $E \in \mathcal{M}$ is said to be **outer regular** for $\mu$ (or $\mu$ is outer regular for $E$) if

$$\mu E = \inf \{\mu O: E \subset O,\ O \text{ open},\ O \in \mathcal{M}\}.$$

It is said to be **inner regular** if

$$\mu E = \sup \{\mu K: K \subset E,\ K \text{ compact},\ K \in \mathcal{M}\}.$$

The set $E$ is said to be **regular** for $\mu$ if it is both inner and outer regular for $\mu$.

We say that the measure $\mu$ is inner regular (outer regular, regular) if it is inner regular (outer regular, regular) for each set $E \in \mathcal{M}$. Proposition 3.15 asserts that Lebesgue measure is a regular measure.

For compact spaces $X$ there is complete symmetry between inner regularity and outer regularity: A measurable set $E$ is outer regular if and only if its complement is inner regular. A finite measure on $X$ is inner regular if and only if it is outer regular (and hence regular). It is not difficult to show that every Baire measure is regular when $X$ is compact.

When $X$ is no longer compact, we lose this symmetry because the complement of an open set need not be compact. We can still establish a number of desirable regularity properties for the case when $X$ is $\sigma$-compact, and some of these extended to the class of $\sigma$-bounded sets in an arbitrary locally compact $X$. We begin with a useful observation about finite measures.

**10. Proposition:** *Let $\mu$ be a finite measure defined on a $\sigma$-algebra $\mathfrak{M}$ which contains all the Baire sets of a locally compact space $X$. If $\mu$ is inner regular, it is regular.*

***Proof:*** Let $E \in \mathfrak{M}$. Then

$$\mu\tilde{E} = \sup \{\mu K \colon K \subset \tilde{E},\ K \in \mathfrak{M}, \text{ and } K \text{ compact}\}.$$

But for such a $K$ we have $\tilde{K}$ open and $E \subset \tilde{K}$. Hence

$$\begin{aligned} \mu E = \mu X - \mu\tilde{E} &= \inf \{\mu X - \mu K\} \\ &= \inf \mu\tilde{K} \\ &\geq \inf \{\mu O \colon E \subset O\}. \end{aligned}$$

Thus $\mu E = \inf \{\mu O \colon E \subset O,\ O \text{ open and } O \in \mathfrak{M}\}$. ▮

**11. Theorem:** *Let $\mu$ be a Baire measure on a locally compact space $X$ and $E$ a $\sigma$-bounded Baire set in $X$. Then for $\epsilon > 0$:*

i. *There is a $\sigma$-compact open set $O$ with $E \subset O$ and $\mu(O \sim E) < \epsilon$.*
ii. $\mu E = \sup \{\mu K \colon K \subset E,\ K$ *a compact* $G_\delta\}$.

***Proof:*** Let $\mathfrak{R}$ be the class of sets $E$ that satisfy (i) and (ii) for each $\epsilon > 0$. Suppose $E = \bigcup E_n$, where $E_n \in \mathfrak{R}$. Then for each $n$ there is a $\sigma$-compact open set $O_n$ with $E_n \subset O_n$ and $\mu(O_n \sim E_n) < 2^{-n}\epsilon$. Then $O = \bigcup O_n$ is again a $\sigma$-compact open set with

$$\mu(O \sim E) \subset \bigcup (O_n \sim E_n),$$

and so

$$\mu(O \sim E) \leq \sum \mu(O_n \sim E_n) < \epsilon.$$

Thus $E$ satisfies (i).

If for some $n$ we have $\mu E_n = \infty$, then there are compact $G_\delta$'s of arbitrary large finite measure contained in $E_n \subset E$. Hence (ii) holds for $E$. If $\mu E_n < \infty$ for each $n$, there is a $K_n \subset E_n$, $K_n$ a compact $G_\delta$ and

$$\mu(E_n \sim K_n) < 2^{-n}\epsilon.$$

Then

$$\begin{aligned} \mu E &= \sup_N \mu\left(\bigcup_{n=1}^N E_n\right) \\ &\leq \sup_N \mu\left(\bigcup_{n=1}^N K_n\right) + \epsilon. \end{aligned}$$

Thus $E$ satisfies (ii).

If $E$ is a compact $G_\delta$, then there is a continuous real-valued function $\varphi$ with compact support such that $0 \leq \varphi \leq 1$ and $E = \{x: \varphi(x) = 1\}$. Let $O_n = \{x: \varphi(x) > 1 - 1/n\}$. Then $O_n$ is a $\sigma$-compact open set with $\bar{O}_n$ compact. Since $\mu O_1 < \infty$, we have $\mu E = \inf \mu O_n$ by Proposition 11.2. Thus each compact $G_\delta$ satisfies (i), and it trivially satisfies (ii).

Let $X$ be compact. Then $E$ satisfies (i) if and only if $\tilde{E}$ satisfies (ii), and so the collection $\mathfrak{R}$ of sets satisfying (i) and (ii) is a $\sigma$-algebra containing the compact $G_\delta$'s. Thus $\mathfrak{R}$ contains all Baire sets, and the Proposition holds when $X$ is compact.

For an arbitrary locally compact space $X$ and bounded Baire set $E$, we can take $H$ to be a compact $G_\delta$ and $U$ to be a $\sigma$-compact open subset of $X$ such that

$$\bar{E} \subset U \subset H$$

Then $E$ is a Baire subset of $H$ by Proposition 8, and so

$$\mu(W \sim E) < \epsilon.$$

Since $W$ and $U$ are $\sigma$-compact, so is $O = W \cap U$. Thus $O$ is a $\sigma$-compact open set with

$$E \subset O \subset W.$$

Hence

$$O \sim E \subset W \sim E,$$

and

$$\mu(O \sim E) < \epsilon.$$

Thus $E$ satisfies (i). Therefore, all bounded Baire sets are in $\mathcal{R}$.

Since $\mathcal{R}$ is closed under countable unions and each $\sigma$-bounded Baire set is a countable union of bounded Baire sets, we see that every $\sigma$-bounded Baire set belongs to $\mathcal{R}$. ∎

If we had defined the class of Baire sets to be the smallest $\sigma$-ring containing the compact $G_\delta$'s and taken a Baire measure to be defined on this $\sigma$-ring, then the Theorem takes the elegant form: "Every Baire measure is regular". If $X$ is $\sigma$-compact, the $\sigma$-ring and the $\sigma$-algebra generated by the compact $G_\delta$'s are the same. Hence we have the following Corollary:

**12. Corollary:** *If $X$ is $\sigma$-compact, every Baire measure on $X$ is regular.*

At first sight it may appear that even in the case of a $\sigma$-compact space Corollary 12 is weaker than Theorem 11 since inner regularity for Baire measures means only that the measure of a Baire set $E$ is the supremum of the measures of the compact Baire sets contained in $E$, while the theorem says that it is the supremum of the measures of the compact $G_\delta$'s contained in $E$. This weakening is only apparent, however, since it can be shown that each compact Baire set is in fact a $G_\delta$. This result is somewhat intricate, and we shall not use it here; the reader is referred to Problem 12 for suggestions for its proof.

If $X$ is not compact, the appropriate notion dual to that of inner regularity for measures is that of quasi regularity: A measure $\mu$ defined on a $\sigma$-algebra $\mathcal{M}$ which contains the Baire sets is said to be **quasi regular** if it is outer regular and each open set $O \in \mathcal{M}$ is inner regular for $\mu$.

A Baire measure on a space which is not $\sigma$-compact need not be regular (Problems 13 and 21), but we can require it to be inner regular or quasi regular without changing its values on the

$\sigma$-bounded Baire sets. We summarize this material in the following Proposition, whose proof is left to the reader (Problem 15).

**13. Proposition:** *Let $\mu$ be a Baire measure on $X$. Then there is a unique quasi regular Baire measure $\bar{\mu}$ on $X$ and a unique inner regular Baire measure $\underline{\mu}$ on $X$ such that $\mu E = \bar{\mu} E = \underline{\mu} E$ for every $\sigma$-bounded Baire set $E$.*

We see by Proposition 10 that $\underline{\mu}$ is regular if $\underline{\mu}$ is finite. In this case $\bar{\mu} = \underline{\mu}$. If there is a regular Baire measure that agrees with $\mu$ on the $\sigma$-bounded Baire sets, it must be equal to $\bar{\mu}$ and $\underline{\mu}$ by the uniqueness of the latter measures. Unfortunately, we can have a $\sigma$-finite Baire measure for which $\bar{\mu}$ and $\underline{\mu}$ are different (Problem 21).

The next three propositions investigate regularity and unicity for Borel measures. The proof of the first is left to the reader.

**14. Proposition:** *Let $\mu$ be a measure defined on a $\sigma$-algebra $\mathfrak{M}$ containing the Borel sets. If $\mu$ is outer regular for each compact set or if $\mu$ is inner regular for each bounded open set, then $\mu$ is regular for each $\sigma$-bounded set in $\mathfrak{M}$.*

**15. Proposition:** *Let $\mu$ be a measure defined on a $\sigma$-algebra $\mathfrak{M}$ containing the Baire sets. Assume either that $\mu$ is quasi regular or that $\mu$ is inner regular. Then for each $E \in \mathfrak{M}$ with $\mu E < \infty$ there is a Baire set $B$ with*

$$\mu(E \,\Delta\, B) = 0.$$

***Proof:*** We treat only the quasi regular case and leave the inner regular case to the reader (Problem 19).

Let $E$ be a measurable set of finite measure. Since $\mu$ is outer regular, we can find a sequence $\langle O_n \rangle$ of open sets with

$$O_n \supset O_{n+1} \supset E$$

and

$$\mu O_n < \mu E + 2^{-n}.$$

Since $\mu$ is quasi regular, there is a compact set $K_m \subset O_m$ with

$$\mu K_m > \mu O_m - 2^{-m},$$

and we may take $K_m$ to be a $G_\delta$ by virtue of Lemma 1. Now

$$\begin{aligned}\mu K_m &> \mu O_m - 2^{-m} \geq \mu E - 2^{-m} \\ &> \mu O_n - 2^{-n} - 2^{-m}.\end{aligned}$$

Set

$$H_m = \bigcup_{j=m}^{\infty} K_j.$$

Then $H_m$ is a Baire set, $H_m \subset O_m \subset O_n$ for $m \geq n$. Also, $H_m \supset H_{m+1}$, and

$$\mu H_m \geq \mu K_m > \mu O_n - 2^{-n} - 2^{-m}.$$

Let $B = \bigcap H_m$. Then $B$ is a Baire set, $B \subset O_n$, and $\mu B = \lim \mu H_m$. Thus

$$\mu B \geq \mu O_n - 2^{-n}.$$

Since $B \subset O_n$ and $E \subset O_n$, we have

$$B \,\Delta\, E \subset (O_n \sim B) \cup (O_n \sim E)$$

and so

$$\begin{aligned}\mu(B \,\Delta\, E) &\leq \mu(O_n \sim B) + \mu(O_n \sim E) \\ &< 2^{-n} + 2^{-n} = 2^{-n+1}.\end{aligned}$$

This is true for any $n$ and so

$$\mu(B \,\Delta\, E) = 0. \quad \blacksquare$$

**16. Proposition:** *Let $\mu_1$ and $\mu_2$ be complete saturated measures defined on $\sigma$-algebras $\mathfrak{M}_1$ and $\mathfrak{M}_2$ containing the Borel sets. Suppose either that $\mu_1$ and $\mu_2$ are quasi regular or that both are inner regular. If $\mu_1 K = \mu_2 K$ whenever $K$ is a compact $G_\delta$, then $\mathfrak{M}_1 = \mathfrak{M}_2$ and $\mu_1 = \mu_2$ for every set in $\mathfrak{M}_i$.*

***Proof:*** We treat only the quasi regular case, leaving the inner regular case to the reader (Problem 20).

We begin by showing that $\mu_1 = \mu_2$ on the $\sigma$-algebra $\mathfrak{M} = \mathfrak{M}_1 \cap \mathfrak{M}_2$. Since each $\mu_i$ is inner on open sets and they agree on the compact $G_\delta$'s we must have $\mu_1 O = \mu_2 O$ for each open $O \in \mathfrak{M}$. But outer regularity implies $\mu_1 E = \mu_2 E$ for each $E \in \mathfrak{M}$.

Each set of $\mu_i$ measure zero is contained in a Borel set of measure zero, and $\mu_1 = \mu_2$ on the Borel sets. Thus $\mu_1$ and $\mu_2$ have the same sets of measure zero because they are complete.

It follows from Proposition 14 that the class of sets in $\mathfrak{M}_1$ with $\mu_1$ finite is the same as the class of sets in $\mathfrak{M}_2$ with $\mu_2$ finite. Consequently, the class of sets locally measurable with respect to $\langle \mu_1, \mathfrak{M}_1 \rangle$ is the same as those with respect to $\langle \mu_2, \mathfrak{M}_2 \rangle$. Since $\mu_1$ and $\mu_2$ are saturated, those classes coincide with $\mathfrak{M}_1$ and $\mathfrak{M}_2$, respectively. ▌

The content of Proposition 15 is not as surprising as it first seems. In the next section we shall see that it is always possible to extend a Baire measure to an inner regular Borel measure and to a quasi regular Borel measure which agree with the original measure on the $\sigma$-bounded Baire sets. The extension to an inner regular measure has a tendency to assign measure zero to as many sets as possible. Thus there are many sets of measure zero, and that makes it easy for $E \,\Delta\, B$ to have measure zero. The extension by quasi regularity again assigns measure zero to those sets that it must by outer regularity but has a propensity to assign infinite measure to most other ambiguous cases. These sets of infinite measure are excluded from the conclusion of Proposition 15.

The example of Problem 7, where we take $X$ to be $\mathbf{R}^2$ with a topology that gives the usual topology on each horizontal line and makes $X$ the direct union of the horizontal lines, is characteristic. A set $E$ is a $\sigma$-bounded Baire set iff it only intersects a countable number of horizontal lines and intersects each of these in a Borel set (of $\mathbf{R}$). The complements of the $\sigma$-bounded sets intersect all but a countable number of horizontal lines in the entire horizontal lines. Set

$$\mu E = \sum_y m(E \cap (R \times \{y\})),$$

where $m$ is Lebesgue measure. Then $\mu$ is a regular Baire measure.

A set $E$ in $X$ is a Borel set iff it intersects each horizontal line in a Borel set of $\mathbf{R}$. The inner regular extension of $\mu$ is the measure

$$\underline{\mu} E = \sum_y m(E \cap (R \times \{y\})).$$

The quasi regular extension of $\mu$ is the measure $\bar{\mu}$ that equals $\mu$ on the $\sigma$-bounded Baire sets (same as the $\sigma$-bounded Borel sets in this example) and is infinite on each Borel set that is not $\sigma$-bounded.

Thus the critical class of sets on which $\bar{\mu}$ and $\underline{\mu}$ differ are those that intersect each horizontal line in a set of Lebesgue measure zero but which have a nonempty intersection with uncountably many horizontal lines. For such sets we have $\underline{\mu} E = 0$ and $\bar{\mu} E = \infty$.

Before concluding this section, we remark that there are Borel measures on compact spaces which are not regular. The construction of such an example is somewhat intricate, and we will not do it here. The interested reader will find suggestions for the construction of such an example in Halmos [5], Problem 10, page 231.

## Problems

**13.** Let $X$ be an uncountable set with the discrete topology and let $X_0 = \{x_k\}$ a countable subset. Set

$$\mu_0 E = \sum_{x_n \varepsilon E} 2^{-n}.$$

**a.** Show that $\mu_0$ is a finite regular measure defined on all subsets of $X$.

**b.** Define $\mu E = \mu_0 E$ if $E$ is $\sigma$-bounded, and $\mu E = 3$ if $\sim E$ is $\sigma$-bounded. Then $\mu$ is a finite outer regular Baire measure that is not inner regular.

**14.** **a.** Show that the intersection of two $\sigma$-compact sets is $\sigma$-compact.

**b.** Give an example of an open subset $O$ of a compact space $X$ where $O$ is not $\sigma$-compact.

**15.** Prove Proposition 13. (See Problem 11.9.)

**16.** Let $X$ be the space of Problem 7 and $\mu$, $\bar{\mu}$, and $\underline{\mu}$ the measures on it defined in the text.

**a.** Show that $\mu$ is a regular Baire measure.

**b.** Show that $\underline{\mu}$ is an inner regular Borel measure.

**c.** Show that $\bar{\mu}$ is a quasi regular Borel measure.

**d.** There is no regular Borel measure on $X$ that agrees with $\mu$ on the $\sigma$-bounded Baire sets.

**17.** Let $\mu$ be an inner regular measure defined on a $\sigma$-algebra $\mathfrak{M}$ containing the Baire sets.

**a.** Each compact $K \varepsilon \mathfrak{M}$ is outer regular (and hence regular) for $\mu$. (Let $O$ be a bounded open set containing $K$. Use the inner regularity of $O \sim K$.)

**b.** For each open set $O \varepsilon \mathfrak{M}$ we have

$$\mu O = \sup \{\mu K \colon K \subset O,\ K \text{ a compact } G_\delta\}.$$

**18.** Prove Proposition 14. (See Problem 17a.)

**19. a.** Prove Proposition 15 for the case when $\mu$ is inner regular.

**b.** Show that the conclusion of Proposition 15 holds for measurable sets $E$ that are of $\sigma$-finite measure.

**20.** Prove Proposition 16 for the case when $\mu$ is inner regular.

**21.** Let $X_0 = \mathbf{R} \times \{0\}$, and $X = X_0 \cup D$, where

$$D = \{\langle x, y\rangle \in \mathbf{R}^2 : x = q2^{-n},\ y = 2^{-n},\ n \in \mathbf{N},\ q \in \mathbf{Z}\}.$$

We define a topology for $X$ by taking as sets in our neighborhood base any subset of $D$ and any set of the form

$$N_{x_0 n_0} = \{\langle x, y\rangle \in X : |x - x_0| < y \text{ and } y \leq 2^{-n_0}\}.$$

**a.** Show that the sets $N_{x_0 n_0}$ are compact in the topology for $X$ and that $X$ is a locally compact Hausdorff space.

**b.** Each compact set contained in $X_0$ is finite.

**c.** Define $\mu\{\langle q2^{-n}, 2^{-n}\rangle\} = 2^{-n}$, and define

$$\mu E = \sum_{z \in E \cap D} \mu\{z\}$$

for any subset $E \subset X$. Show that $\mu$ is a countably additive measure defined for all subsets of $X$.

**d.** Show that $\mu N_{x_0 n_0} < \infty$.

**e.** Show that $\mu$ is a $\sigma$-finite inner regular Borel measure.

**f.** Say that a point $\langle x, y\rangle \in D$ *covers* a point $\langle x_0, 0\rangle \in X_0$ if $|x - x_0| < y$. The Lebesgue measure of a set $G \subset X_0$ covered by a point $\langle x, y\rangle \in D$ is at most $\mu\{\langle x, y\rangle\}$.

**g.** Say that $E \subset D$ covers $G \subset X_0$ if each point of $G$ is covered by a point of $E$. Then the Lebesgue measure of $G$ satisfies

$$mG \leq \mu E.$$

**h.** The set $X_0$ is not outer regular for $\mu$. [We have $\mu X_0 = 0$. If $O$ is an open set with $O \supset X_0$, then $O \cap D$ covers $X_0$.]

**i.** There is no regular Baire measure on $X$ that agrees with $\mu$ on all compact sets.

**j.** What is $\bar{\mu}$ for this example?

## 3 The Construction of Borel Measures

The only nontrivial Borel measures we have met so far are Lebesgue measure and the measures of Chapter 12. In this section we generalize the procedures used in obtaining them to give us general methods of constructing Borel measures. The first method is to start with a suitable outer measure and take the measurable sets with respect to this outer measure. In order that all Borel sets be measurable, the outer measure must have certain properties. One convenient set of properties to require is given in the following definition:

**Definition:** *An outer measure* $\mu^*$ *on a locally compact Hausdorff space is said to be (topologically) regular if*

i. *For each* $E \subset X$, $\mu^*E = \inf\{\mu^*O\colon O \text{ open}, E \subset O\}$.
ii. $\mu^*(O_1 \cup O_2) = \mu^*O_1 + \mu^*O_2$ *if* $O_1$ *and* $O_2$ *are disjoint open sets.*
iii. $\mu^*(O) = \sup\{\mu^*K\colon K \subset O, K \text{ compact}\}$, *for* $O$ *open.*

This topological notion of regularity should be distinguished from the purely measure-theoretic concept of regularity introduced in Chapter 12. We shall see (Theorem 19) that if $\mu^*$ is topologically regular, then all Borel sets are $\mu^*$-measurable. Thus open sets are measurable, and hence each set $E$ with $\mu^*E < \infty$ has a $G_\delta$ for a measurable cover. Consequently, $\mu^*$ is also regular in the measure-theoretic sense.

We begin with two lemmas, whose proofs are left to the reader.

**17. Lemma:** *If* $\mu^*$ *is an outer measure on* $X$, *then each of the following is equivalent to (iii) in the definition of regularity for* $\mu^*$:

iii′. $\mu^*O = \sup\{\mu^*K\colon K \subset O, K \text{ a compact } G_\delta\}$ *for* $O$ *open.*
iii″. $\mu^*O = \sup\{\mu^*U\colon \bar{U} \subset O, \bar{U} \text{ compact}, U \text{ open}\}$ *for* $O$ *open.*

**18. Lemma:** *An arbitrary set* $E \subset X$ *is* $\mu^*$*-measurable if and only if*

$$\mu^*O \geq \mu^*(O \cap E) + \mu^*(O \cap \tilde{E})$$

*for each open* $O$ *with* $\mu^*O < \infty$.

**19. Theorem:** *Let* $\mu^*$ *be a topologically regular outer measure on* $X$. *Then each Borel set is* $\mu^*$*-measurable.*

***Proof:*** Since the $\mu^*$-measurable sets from a $\sigma$-algebra, it suffices to show that each closed set $F$ is $\mu^*$-measurable.

Let $O$ be any open set with $\mu^*O < \infty$ and $\epsilon$ an arbitrary positive number. Then $O \cap \tilde{F}$ is an open set of finite outer measure. By property (iii″) of Lemma 17 there is an open set $U$ with $\bar{U} \subset O \cap \tilde{F}$ and $\mu^*U > \mu^*(O \cap \tilde{F}) - \epsilon$.

Set $V = O \sim \bar{U}$. Then $V \cap U = \emptyset$ and $O \cap F \subset V$. Hence

$$\begin{aligned}\mu^*(O \cap F) + \mu^*(O \cap \tilde{F}) &< \mu^*V + \mu^*U + \epsilon\\ &< \mu^*(U \cup V) + \epsilon\\ &< \mu^*O + \epsilon,\end{aligned}$$

by (ii). Since $\epsilon$ was arbitrary,

$$\mu^*(O \cap F) + \mu^*(O \cap \tilde{F}) \leq \mu^*O,$$

and so $F$ is $\mu^*$ measurable. ▮

We often want to define a regular outer measure by starting with a suitable extended real-valued function on the open sets. The following proposition gives conditions under which this will lead to a regular outer measure.

**20. Proposition:** *Let $\bar{\mu}$ be a nonnegative extended real-valued function defined on the class of open subsets of $X$ and satisfying*

i. $\bar{\mu}O < \infty$ *if* $\bar{O}$ *compact,*
ii. $\bar{\mu}O_1 \leq \bar{\mu}O_2$, *if* $O_1 \subset O_2$,
iii. $\bar{\mu}(O_1 \cup O_2) = \bar{\mu}O_1 + \bar{\mu}O_2$ *if* $O_1 \cap O_2 = \emptyset$,
iv. $\bar{\mu}(\bigcup O_i) \leq \sum \mu O_i$,
v. $\bar{\mu}(O) = \sup \{\mu U\colon \bar{U} \subset O, \bar{U}$ *compact*$\}$.

*Then the set function $\mu^*$ defined by*

$$\mu^*E = \inf \{\bar{\mu}O\colon E \subset O\}$$

*is a topologically regular outer measure.*

*Proof:* The monotonicity and countable subadditivity of $\mu^*$ follow directly from (ii) and (iv) and the definition of $\mu^*$. Also $\mu^*O = \bar{\mu}O$ for $O$ open, and so condition (iii″) of Lemma 17 follows from (v). Condition (ii) of the definition of regularity follows from hypothesis (iii) of the proposition and condition (i) from the definition of $\mu^*$. Since $\bar{\mu}O < \infty$ for $\bar{O}$ compact, we have $\mu^*E < \infty$ for each bounded set $E$. ▮

A set function $\bar{\mu}$ defined on open sets and satisfying the conditions of the preceding Proposition is sometimes called an inner content.

The Proposition asserts that such a set function can always be extended to a Borel measure. A dual procedure is to start with a suitable set function on compact sets or compact $G_\delta$'s. This leads us to the following definition.

**Definition:** *Let $\mathcal{K}$ be a family of compact sets containing the compact $G_\delta$'s and having the property that $K_1 \cup K_2 \in \mathcal{K}$ and $K_1 \cap K_2 \in \mathcal{K}$ whenever $K_1 \in \mathcal{K}$ and $K_2 \in \mathcal{K}$. A nonnegative real-valued function $\lambda$ defined on $\mathcal{K}$ is called a content if*

i. $\lambda K_1 \leq \lambda K_2$ *when* $K_1 \subset K_2$.
ii. $\lambda(K_1 \cup K_2) = \lambda K_1 + \lambda K_2$ *when* $K_1 \cap K_2 = \emptyset$.

*It is called a regular content if*

iii. $\lambda K = \inf \{\lambda H: K \subset H^\circ, H \in \mathcal{K}\}$.

21. **Proposition:** *Let $\lambda$ be a content on a class $\mathcal{K}$ of compact sets. Then there is a unique quasi regular Borel measure $\bar{\mu}$ such that for each open set $O$ we have*

$$\bar{\mu} O = \sup \{\lambda K: K \subset O, K \in \mathcal{K}\}$$

*Moreover,*

$$\bar{\mu} K^\circ \leq \lambda K \leq \bar{\mu} K$$

*for $K \in \mathcal{K}$. If $\lambda$ is a regular content, then $\bar{\mu} K = \lambda K$ for all $K \in \mathcal{K}$.*

***Proof:*** For each open set $O$ let

$$\bar{\mu} O = \sup \{\lambda K: K \subset O, K \in \mathcal{K}\}.$$

Then properties (i), (ii), (iii), and (v) of Proposition 20 follow directly from the definition of $\bar{\mu}$ and the definition of $\lambda$ as a content. To see that (iv) also holds, let

$$O = \bigcup_{i=1}^{\infty} O_i,$$

and take a compact set $K \in \mathcal{K}$ with $K \subset O$. By Proposition 9.16 there are nonnegative continuous real-valued functions $\varphi_1, \ldots, \varphi_n$ with compact support such that

$$\sum_{i=1}^{n} \varphi_i \equiv 1 \text{ on } K,$$

and supp $\varphi_1 \subset O_i$. Let $G_i = \{x: \varphi_i(x) \geq 1/n\}$. Then each $G_i$ is a compact $G_\delta$ and thus in $\mathcal{K}$. Consequently, each $G_i \cap K$ is in $\mathcal{K}$. We also have

$$K = \bigcup_{i=1}^{n} G_i \cap K$$

and $G_i \cap K \subset O_i$. Thus

$$\begin{aligned} \lambda K &\leq \sum_{i=1}^{n} \lambda(G_i \cap K) \\ &\leq \sum_{i=1}^{n} \bar{\mu} O_i \leq \sum_{i=1}^{\infty} \bar{\mu} O_i. \end{aligned}$$

Taking the supremum over all such $K$ gives (iv).

By Proposition 20 and Theorem 19 the set function $\bar{\mu}$ extends to a quasi regular Borel measure $\bar{\mu}$. Since $\bar{\mu}$ is outer regular, we have

$$\begin{aligned} \bar{\mu} K &= \inf \{\bar{\mu} O,\ O \text{ open},\ K \subset O\} \\ &\geq \lambda K. \end{aligned}$$

Also,

$$\begin{aligned} \bar{\mu} K^\circ &= \sup \{\lambda H : H \ \varepsilon\ \mathcal{K},\ H \subset K^\circ\} \\ &\leq \lambda K. \end{aligned}$$

Thus

$$\bar{\mu} K^\circ \leq \lambda K \leq \bar{\mu} K.$$

If $\lambda$ is regular,

$$\begin{aligned} \bar{\mu} K &= \inf \{\bar{\mu} O,\ K \subset O,\ O \text{ open}\} \\ &\leq \inf \{\bar{\mu} H^\circ : K \subset H^\circ,\ H \ \varepsilon\ \mathcal{K}\} \\ &\leq \inf \{\lambda H : K \subset H^\circ,\ H \ \varepsilon\ \mathcal{K}\} \\ &= \lambda K. \end{aligned}$$

Thus $\bar{\mu} K = \lambda K$ for all $K \ \varepsilon\ \mathcal{K}$. ∎

**22. Theorem:** *Let $\mu$ be a Baire measure on $X$. Then there are complete saturated measures $\bar{\mu}$ and $\underline{\mu}$ defined on a $\sigma$-algebra containing the Borel sets with $\bar{\mu}$ quasi regular, $\underline{\mu}$ inner regular, and*

$$\bar{\mu} E = \underline{\mu} E = \mu E$$

*for each $\sigma$-bounded Baire set. The measures $\bar{\mu}$ and $\underline{\mu}$ are unique.*

***Proof:*** Let $\mathcal{K}$ be the class of compact $G_\delta$'s. Then the restriction of $\mu$ to $\mathcal{K}$ is a regular content which extends to a quasi regular Borel measure $\bar{\mu}$ by Proposition 21. Since $\bar{\mu}$ is inner regular for open sets and $\mu$ is inner regular for $\sigma$-bounded Baire sets, they agree for $\sigma$-bounded open Baire sets. Since both are outer regular on $\sigma$-bounded Baire sets, we have $\bar{\mu}E = \mu E$ for each $\sigma$-bounded Baire set. Since $\bar{\mu}$ comes from a regular outer measure, it is complete and saturated.

Let $\mathfrak{M}$ be the class of $\mu^*$ measurable sets for $\mu^*$. Define $\underline{\mu}$ on $\mathfrak{M}$ by

$$\underline{\mu}E = \sup \{\bar{\mu}B: B \subset E, B \in \mathfrak{M}, B \ \sigma\text{-bounded}\}.$$

Since $\bar{\mu}$ is regular on the $\sigma$-bounded sets by Proposition 14, we have $\underline{\mu}E = \bar{\mu}E$ for all $\sigma$-bounded sets in $\mathfrak{M}$ and $\underline{\mu}E = \bar{\mu}E = \mu E$ for all $\sigma$-bounded Baire sets.

To see that $\underline{\mu}$ is countably additive, let $\langle E_j \rangle$ be a disjoint sequence of sets in $\mathfrak{M}$ with $E = \bigcup E_j$. Then for any $\sigma$-bounded set $B$ of $\mathfrak{M}$ with $B \subset E$, we have $B_j = B \cap E_j$ a $\sigma$-bounded set in $\mathfrak{M}$ and $B_j \subset E_j$. Thus

$$\bar{\mu}B = \sum \bar{\mu}B_j \leq \sum \underline{\mu}E_j.$$

Hence

$$\underline{\mu}E \leq \sum \underline{\mu}E_j,$$

and $\underline{\mu}$ is countably subadditive. If $\underline{\mu}E_j = \infty$ for some $j$, $\underline{\mu}E = \infty$, and so $\underline{\mu}E = \sum \underline{\mu}E_j$. If each $\underline{\mu}E_j$ is finite, then given $\epsilon > 0$, there is a $\sigma$-bounded set $B_j \in \mathfrak{M}$, $B_j \subset E_j$, such that

$$\underline{\mu}E_j < \bar{\mu}B_j + 2^{-j}\epsilon.$$

Then $B = \bigcup B_j \subset E$, and $B$ is $\sigma$-bounded. Also,

$$\underline{\mu}E \geq \bar{\mu}B = \sum \bar{\mu}B_j \geq \sum \underline{\mu}E_j - \epsilon.$$

Since $\epsilon$ was arbitrary, $\underline{\mu}E \geq \sum \underline{\mu}E_j$, and $\underline{\mu}$ is countable additive.

Since $\bar{\mu}$ is inner regular on the $\sigma$-bounded sets of $\mathfrak{M}$, $\underline{\mu}$ is inner regular on $\mathfrak{M}$.

The unicity of $\bar{\mu}$ and $\underline{\mu}$ follows from Proposition 16. ∎

## Problems

**22.** Let $\mu$ and $\nu$ be Borel measures on $X$ and $Y$ with each of them either inner or quasi regular.

**a.** Each $\sigma$-bounded Borel set $E$ of $X \times Y$ is in $\mathcal{B}o(X) \times \mathcal{B}o(Y)$, and, if $E$ is $\sigma$-bounded,

$$\overline{(\mu \times \nu)}E = (\bar{\mu} \times \bar{\nu})E = (\underline{\mu} \times \underline{\nu})E = (\mu \times \nu)E = (\underline{\mu \times \nu})E.$$

**b.** Show that $\bar{\mu} \times \bar{\nu} \geq \overline{\mu \times \nu}$.

**c.** Show that $\underline{\mu} \times \underline{\nu} \geq \underline{\mu \times \nu}$.

**23.** Let $\mu$ be Lebesgue measure on $X = [0, 1]$ and $\nu$ the counting measure on $Y = [0, 1]$ with the discrete topology. Then $\mu = \bar{\mu} = \underline{\mu}$ and $\nu = \bar{\nu} = \underline{\nu}$.

**a.** Show that

$$(\mu \times \nu)E = \overline{(\mu \times \nu)}E = \begin{cases} \infty & \text{if } E \text{ not } \sigma\text{-bounded} \\ \sum_\alpha m(E \cap [X \times \{\alpha\}]) & \text{if } E \text{ is } \sigma\text{-bounded.} \end{cases}$$

**b.** Show that

$$(\underline{\mu \times \nu})E = \sum_\alpha m(E \cap [X \times \{\alpha\}]).$$

**c.** Let $D = \{\langle x, y\rangle \in X \times Y : x = y\}$. Then $D$ is closed and hence a Borel subset of $X \times Y$.

**d.** The set $D$ in (c) is in $\mathcal{B}o(X) \times \mathcal{B}o(Y)$.

**e.** Let $f(x, y) = \chi_D$. Then $f$ is Borel measurable, and

$$\int \left( \int f \, d\mu \right) d\nu = 0$$

$$\int \left( \int f \, d\nu \right) d\mu = 1.$$

**f.** What are $\int f \, d(\mu \times \nu)$ and $\int f \, d(\underline{\mu \times \nu})$?

**24.** Let $\mu$ be a Baire measure on a locally compact space $X$. Let $U$ be the union of all open Baire sets $O$ with $\mu O = 0$. The complement $F = \tilde{U}$ of $U$ is a closed set called the **support** (or carrier) of $\mu$.

**a.** If $O$ is an open Baire set with $O \cap F \neq \emptyset$, then $\mu O > 0$.

**b.** If $K$ is a compact Baire set with $K \cap F = \emptyset$, then $\mu K = 0$. (Each point of $K$ is contained in an open set of measure zero. Thus by compactness $K$ is contained in an open set of measure zero.)

**c.** If $E$ is a $\sigma$-bounded Baire set with $E \cap F = \emptyset$, then $\mu E = 0$.

**d.** If $f \in C_c(X)$ and $f \geq 0$, then $\int f\,d\mu = 0$ if and only if $f \equiv 0$ on $F$. [Hint: The set $\{x: f(x) > 0\}$ is a $\sigma$-bounded Baire set.]

**e.** Give an example to show that $F$ need not be a Baire set.

**f.** It follows from (c) that if $X$ is compact (or $\sigma$-compact), then $\mu E = 0$ for each Baire set with $E \cap F = \emptyset$. Construct an example to show that this need not be true if $X$ is not $\sigma$-compact (see Problem 5e).

## 4 Positive Linear Functionals and Borel Measures

Let $X$ be a locally compact Hausdorff space. By $C_c(X)$ we denote, as usual, the space of continuous real-valued functions with compact support. A real-valued linear functional $I$ on $C_c(X)$ is said to be positive if $I(f) \geq 0$ whenever $f \geq 0$. The purpose of the present section is to prove that every positive linear functional on $C_c(X)$ is represented by integration with respect to a suitable Borel (or Baire) measure. In particular we have the following theorem:

**23. Theorem (Riesz–Markov):** *Let $X$ be a locally compact Hausdorff space and $I$ a positive linear functional on $C_c(X)$. Then there is a Borel measure $\mu$ on $X$ such that*

$$I(f) = \int f\,d\mu$$

*for each $f \in C_c(X)$. The measure $\mu$ may be taken to be quasi regular or to be inner regular. In each of these cases it is then unique.*

***Proof:*** For each open set $O$ define $\bar{\mu}O$ by

$$\bar{\mu}O = \sup\,\{I(f): f \in C_c(X),\ 0 \leq f \leq 1,\ \operatorname{supp} f \subset O\}.$$

Then $\bar{\mu}$ is an extended real-valued function defined on all open sets and is readily seen to be monotone, finite on bounded sets, and to satisfy the regularity (v) of Proposition 20. To see that $\bar{\mu}$ is countably subadditive on open sets, let $O = \bigcup O_i$ and let $f$ be any function in $C_c(X)$ with $0 \leq f \leq 1$ and $\operatorname{supp} f \subset O$. By Proposition 9.16 there are nonnegative functions $\varphi_1, \ldots, \varphi_n$ in $C_c(X)$ with $\operatorname{supp} \varphi_i \subset O_i$, and

$$\sum_{i=1}^{n} \varphi_i \equiv 1$$

on supp $f$. Then $f = \sum \varphi_i f$, $0 \le \varphi_i f \le 1$, and supp $(\varphi_i f) \subset O_i$. Thus

$$If = \sum_{i=1}^{n} I(\varphi_i f) \le \sum_{i=1}^{n} \bar{\mu} O_i$$
$$\le \sum_{i=1}^{\infty} \bar{\mu} O_i.$$

Taking the supremum over all such $f$ gives

$$\bar{\mu} O \le \sum_{i=1}^{\infty} \bar{\mu} O_i,$$

and $\bar{\mu}$ is countably subadditive.

If $O = O_1 \cup O_2$ with $O_1 \cap O_2 = \emptyset$ and $f_i \in C_c(X)$, $0 \le f_i \le 1$, and supp $f_i \subset O_i$, then the function $f = f_1 + f_2$ has supp $f \subset O$ and $0 \le f \le 1$. Thus

$$If_1 + If_2 \le \bar{\mu} O.$$

Since $f_1$ and $f_2$ can be chosen arbitrarily, subject to $0 \le f_i \le 1$ and supp $f_i \subset O_i$, we have

$$\bar{\mu} O_1 + \bar{\mu} O_2 \le \bar{\mu} O,$$

whence

$$\bar{\mu} O_1 + \bar{\mu} O_2 = \bar{\mu} O.$$

Thus $\bar{\mu}$ satisfies the hypotheses of Proposition 20, and so $\bar{\mu}$ extends to a quasi regular Borel measure.

We next proceed to show that $If = \int f\,d\mu$ for each $f \in C_c(X)$. Since $f$ is the difference of two nonnegative functions in $C_c(X)$, it suffices to consider $f \ge 0$. By linearity we may also take $f \le 1$.

Choose a bounded open set $O$ with supp $f \subset O$. Set

$$O_k = \{x \colon nf(x) > k - 1\}$$

and $O_0 = O$. Then $O_{n+1} = \emptyset$, and $\bar{O}_{k+1} \subset O_k$. Define

$$\varphi_k = \begin{cases} 1 & \text{in } O_{k+1} \\ nf(x) - k + 1 & \text{in } O_k \sim O_{k+1} \\ 0 & \text{in } \tilde{O}_k. \end{cases}$$

Then

$$f = \frac{1}{n} \sum_{k=1}^{n} \varphi_k.$$

We also have supp $\varphi_k \subset \bar{O}_k \subset O_{k-1}$ and $\varphi_k \equiv 1$ on $O_{k+1}$. Thus

$$\bar{\mu}O_{k+1} \leq I\varphi_k \leq \bar{\mu}O_{k-1}$$

for $k \geq 1$. Also,

$$\bar{\mu}O_{k+1} \leq \int \varphi_k \, d\bar{\mu} \leq \bar{\mu}O_k$$

for $k \geq 1$. Hence

$$-\mu O_1 \leq \sum_{k=1}^{n} \left( I\varphi_k - \int \varphi_k \right) \leq \bar{\mu}O_0 + \bar{\mu}O_1.$$

Consequently,

$$\left| If - \int f \, d\mu \right| \leq \frac{2}{n} \bar{\mu}O.$$

Since $n$ is arbitrary,

$$If = \int f \, d\bar{\mu}.$$

It follows from Theorem 22 that there is an inner regular Borel measure $\underset{\sim}{\mu}$ which agrees with $\bar{\mu}$ on the $\sigma$-bounded Borel sets. Since only the values of $\mu$ on $\sigma$-bounded Baire sets enter into $\int f \, d\mu$, we have

$$If = \int f \, d\underset{\sim}{\mu}.$$

The unicity of $\bar{\mu}$ and $\underset{\sim}{\mu}$ is left to the reader. ▮

Although this theorem is often called the Riesz–Markov Theorem, it was only established by Markov [23] for the case when $I$ is a *bounded* positive linear functional. It was established in 1909 by F. Riesz [24] for the case when $X = [a, b]$, by Radon in 1913 when $X$ is a compact subset contained in $\mathbf{R}^n$, and by Banach in 1937 when $X$ is a compact metric space. The earliest statement I have seen in print of the theorem in the generality given here is Halmos [5].

## Problems

**25.** Prove the unicity of the measure $\bar{\mu}$ and $\underset{\sim}{\mu}$ in Theorem 23.

**26.** Let $k(x, y)$ be a bounded Borel measurable function on $X \times Y$, and let $\mu$ and $\nu$ be Borel measures on $X$ and $Y$ that are either inner or quasi regular.

**a.** Show that

$$\iint \varphi(x)k(x, y)\psi(y)\, d(\mu \times \nu) = \int\left[\int \varphi(x)k(x, y)\, d\mu\right]\psi(y)\, d\nu$$

$$= \int \varphi(x)\left[\int k(x, y)\psi(y)\, d\nu\right] d\mu$$

for all $\varphi \in C_c(X)$ and $\psi \in C_c(Y)$.

**b.** Show that (a) holds for $\varphi \in C_0(X)$ and $\psi \in C_0(Y)$.

**c.** If the integral in (a) is zero for all $\varphi$ and $\psi$ in $C_c(X)$ and $C_c(Y)$, then $k = 0$ a.e. $[\mu \times \nu]$.

## 5 Bounded Linear Functionals on $C(X)$

Let $X$ be a compact Hausdorff space and $C(X)$ the space of real-valued continuous functions on $X$. In Section 4 we described the positive linear functionals on $C(X)$, and in this section we consider the bounded linear functionals on $C(X)$. We first note that if $F$ is a positive linear functional on $C(X)$ and if $|f| \le 1$, then

$$|F(f)| \le F(|f|) \le F(1).$$

Hence it follows that

$$\|F\| = F(1).$$

The next proposition shows that every bounded linear functional on $C(X)$ is the difference of two positive ones. Since the proof makes no use of particular properties of $C(X)$ other than that it is a vector lattice of bounded functions containing 1, we state the proposition in this generality. Recall that a vector space of real-valued functions on $X$ is a vector lattice if $f \vee g$ and $f \wedge g$ belong to $L$ whenever $f$ and $g$ do.

The lattice $L$ becomes a normed linear space when we define $\|f\| = \sup |f(x)|$. A linear functional is bounded if there is an $M$ such that

$$|F(f)| \le M\|f\|,$$

and we define as usual

$$\|F\| = \sup_{\|f\| \le 1} F(f).$$

**24. Proposition:** *Let $L$ be a vector lattice of bounded functions on a set $X$, and suppose $1 \in L$. Then for each bounded linear functional $F$*

*on $L$, there are two positive linear functionals $F_+$ and $F_-$ such that $F = F_+ - F_-$ and $\|F\| = F_+(1) + F_-(1)$.*

***Proof:*** For each nonnegative $f$ in $L$ define

$$F_+(f) = \sup_{0 \le \varphi \le f} F(\varphi).$$

Then $F_+(f) \ge 0$, and $F_+(f) \ge F(f)$. Moreover, $F_+(cf) = cF_+(f)$ for $c \ge 0$. Let $f$ and $g$ be two nonnegative functions in $L$. If $0 \le \varphi \le f$ and $0 \le \psi \le g$, then $0 \le \varphi + \psi \le f + g$, and so

$$F_+(f + g) \ge F(\varphi) + F(\psi).$$

Taking suprema over all such $\varphi$ and $\psi$, we obtain

$$F_+(f + g) \ge F_+(f) + F_+(g).$$

On the other hand, if $0 \le \psi \le f + g$, then $0 \le \psi \wedge f \le f$ and thus $0 \le \psi - (\psi \wedge f) \le g$, whence

$$\begin{aligned} F(\psi) &= F(\psi \wedge f) + F(\psi - [\psi \wedge f]) \\ &\le F_+(f) + F_+(g). \end{aligned}$$

Taking the supremum over all such $\psi$, we get

$$F_+(f + g) \le F_+(f) + F_+(g),$$

and so

$$F_+(f + g) = F_+(f) + F_+(g).$$

Let $f$ be an arbitrary function in $L$, and let $M$ and $N$ be two nonnegative constants such that $f + M$ and $f + N$ are nonnegative. Then

$$\begin{aligned} F_+(f + M + N) &= F_+(f + M) + F_+(N) \\ &= F_+(f + N) + F_+(M). \end{aligned}$$

Hence

$$F_+(f + M) - F_+(M) = F_+(f + N) - F_+(N).$$

Thus the value of $F_+(f + M) - F_+(M)$ is independent of the choice of $M$, and we define $F_+(f)$ to be this value. The functional $F_+$ is defined on all of $L$, and we have $F_+(f + g) = F_+(f) + F_+(g)$ and $F_+(cf) = cF_+(f)$ for $c \ge 0$. Since $F_+(-f) + F_+(f) = F_+(0) = 0$, we have $F_+(-f) = -F_+(f)$, and $F_+$ is a linear functional on $L$.

Since $0 \le F_+(f)$ and $F(f) \le F_+(f)$ for $f \ge 0$, both $F_+$ and the linear functional $F_- = F_+ - F$ are positive linear functionals, and $F = F_+ - F_-$.

We always have $\|F\| \leq \|F_+\| + \|F_-\| = F_+(1) + F_-(1)$. To establish the inequality in the opposite direction, let $\varphi$ be any function in $L$ such that $0 \leq \varphi \leq 1$. Then $|2\varphi - 1| \leq 1$, and

$$\|F\| \geq F(2\varphi - 1) = 2F(\varphi) - F(1).$$

Taking the supremum over all such $\varphi$, we have

$$\begin{aligned}\|F\| &\geq 2F_+(1) - F(1)\\ &= F_+(1) + F_-(1).\end{aligned}$$

Hence $\|F\| = F_+(1) + F_-(1)$. ∎

**25. Riesz Representation Theorem:** *Let $X$ be a compact Hausdorff space and $C(X)$ the space of continuous real-valued functions on $X$. Then to each bounded linear functional $F$ on $C(X)$ there corresponds a unique finite signed Baire measure $\nu$ on $X$ such that*

$$F(f) = \int f\,d\nu$$

*for each $f$ in $C(X)$. Moreover, $\|F\| = |\nu|(X)$.*

***Proof:*** Let $F = F_+ - F_-$ as in Proposition 24. Then by Theorem 23 there are finite Baire measures $\mu_1$ and $\mu_2$ such that

$$F_+(f) = \int f\,d\mu_1$$

and

$$F_-(f) = \int f\,d\mu_2.$$

If we set $\nu = \mu_1 - \mu_2$, then $\nu$ is a finite signed Baire measure, and

$$F(f) = \int f\,d\nu.$$

Now

$$\begin{aligned}|F(f)| &\leq \int |f|\,d|\nu|\\ &\leq \|f\|\,|\nu|(X).\end{aligned}$$

Hence $\|F\| \leq |\nu|(X)$. But

$$|\nu|(X) \leq \mu_1(X) + \mu_2(X)$$
$$= F_+(1) + F_-(1) = \|F\|.$$

Thus $\|F\| = |\nu|(X)$.

To show the uniqueness of $\nu$, we note that if, $\nu_1$ and $\nu_2$ were both finite signed Baire measures such that

$$\int f\, d\nu_i = F(f)$$

for $i = 1, 2$ and $f \varepsilon C(X)$, then $\lambda = \nu_1 - \nu_2$ would be a finite signed Baire measure such that

$$\int f\, d\lambda = 0$$

for all $f \varepsilon C(X)$. Let $\lambda = \lambda^+ - \lambda^-$ be the Jordan decomposition of $\lambda$. Then integration with respect to $\lambda^+$ gives the same positive linear functional on $C(X)$ as that given by $\lambda^-$, and so by Theorem 23 we must have $\lambda^+ = \lambda^-$. Hence $\lambda = 0$, and $\nu_1 = \nu_2$. ∎

**26. Corollary:** *Let $X$ be a compact Hausdorff space. Then the dual of $C(X)$ is (isometrically isomorphic to) the space of all finite signed Baire measures on $X$ with norm defined by $\|\nu\| = |\nu|(X)$.*

The fact that the space of finite signed Baire measures on $X$ is the dual of $C(X)$ enables us to conclude a number of things about this space. For example, it follows from Proposition 10.3 that the space of Baire measures is complete, and it follows from Theorem 10.17 that the set of Baire measures with $|\nu|(X) \leq 1$ is compact in the weak* topology. Some consequences of this are explored in the problems.

## Problems

**27.** Let $L$ and $F$ be as in Proposition 24. Show that if $G$ and $H$ are two positive linear functionals on $L$ such that $F = G - H$ and $G(1) + H(1) \leq \|F\|$, then $G = F^+$ and $H = F^-$. [Hint: Use the definition of $F^+$ to show that $G - F^+$ is a positive linear functional.]

**28.** Let $X$ be a compact Hausdorff space $\mathcal{F} = \{f_\alpha\}$ a family of continuous real-valued functions on $X$ and $\{c_\alpha\}$ a corresponding family of constants.

Suppose that for each finite set $\{f_{\alpha_1}, \ldots, f_{\alpha_n}\}$ there is a signed Baire measure $\nu$ with $|\nu|(X) \le 1$ such that

$$\int f_{\alpha_i} \, d\nu = c_{\alpha_i}.$$

Then there is a finite signed Baire measure $\nu$ with $|\nu|(X) \le 1$ such that for every $f_\alpha$,

$$\int f_\alpha \, d\nu = c_\alpha.$$

**29. a.** Let $X$ be a compact Hausdorff space and $g, f_1, \ldots, f_n$ continuous real-valued functions on $X$. Suppose that there is a signed Baire measure $\nu$ on $X$ with $|\nu|(X) \le 1$ such that for each $i$ we have $\int f_i \, d\nu = c_i$. Then there is a signed Baire measure $\mu$ on $X$ with $|\mu|(X) \le 1$ such that

$$\int f_i \, d\mu = c_i$$

and

$$\int g \, d\mu \le \int g \, d\lambda$$

for any signed Baire measure $\lambda$ with $|\lambda|(X) \le 1$ and such that $\int f_i \, d\lambda = c_i$.

**b.** Suppose that there is a Baire measure $\nu$ on $X$ with $\nu(X) = 1$ and $\int f_i \, d\nu = c_i$. Then there is a Baire measure $\mu$ on $X$ with $\mu(X) = 1$ and $\int f_i \, d\mu = c_i$ which minimizes $\int g \, d\mu$ among all Baire measure which satisfy these conditions.

**c.** Let $G$, $F_1, \ldots, F_n$ be continuous functions on $\mathbf{R}^m$ ($=$ Euclidean $m$-dimensional space), and let $f_1, \ldots, f_m$ be continuous functions on $X$. Show that if there is a Baire measure $\nu$ with $\nu(X) = 1$ such that

$$F_i(\textstyle\int f_1 \, d\nu, \ldots, \int f_m \, d\nu) = c_i,$$

then there is a Baire measure on $X$ which minimizes

$$G(\textstyle\int f_1 \, d\nu, \ldots, \int f_m \, d\nu)$$

under these restrictions.

**30.** Let $B$ be the Banach space of signed Baire measures on a compact Hausdorff space $X$. What are the extreme points of the unit ball of $B$?

**31.** *Alternate proof of the Stone–Weierstrass Theorem.* We can use the techniques of this section, together with results of Chapter 10, to give a proof of the Stone–Weierstrass Theorem which does not depend on Lemma 9.33. This proof is due to deBranges.

Let $\mathfrak{A}$ be an algebra of real-valued continuous functions on a compact space $X$ which separates points and contains the constants. Let $\mathfrak{A}^\perp$ be the set of signed Baire measures on $X$ such that $|\mu|(X) \leq 1$ and $\int f\, d\mu = 0$ for all $f \varepsilon \mathfrak{A}$.

**a.** Use the Hahn–Banach Theorem and Corollary 26 to show that if $\mathfrak{A}^\perp$ contains only the zero measure, then $\overline{\mathfrak{A}} = C(X)$.

**b.** Use the Krein–Milman Theorem and the compactness of the unit ball in $C^*(X)$ to show that if the zero measure is the only extreme point of $\mathfrak{A}^\perp$, then $\mathfrak{A}^\perp$ contains only the zero measure.

**c.** Let $\mu$ be an extreme point of $\mathfrak{A}^\perp$. If $f \varepsilon \mathfrak{A}$, $0 \leq f \leq 1$, the measures $\mu_1$ and $\mu_2$ given by $d\mu_1 = f\, d\mu$ and $d\mu_2 = (1 - f)\, d\mu$ are in $\mathfrak{A}^\perp$, with $\|\mu_1\| + \|\mu_2\| = \|\mu\|$, and $\mu_1 + \mu_2 = \mu$. Since $\mu$ is an extreme point, $\mu_1 = c\mu$ for some constant $c$.

**d.** Then $f - c \equiv 0$ on the support of $\mu$ (see Problem 24).

**e.** Since $\mathfrak{A}$ separates points, the support of $\mu$ can contain at most one point. Since $\int 1 d\mu = 0$ the support of $\mu$ is empty, and $\mu$ is the zero measure.

# 14 Invariant Measures

## 1 Homogeneous Spaces

Let $X$ be a locally compact Hausdorff space. A group $G$ of homeomorphisms of $X$ onto itself is said to be *transitive* on $X$ if, given any two elements $x$, $y$ of $X$, there is a homeomorphism $g \in G$ with $g(x) = y$. By a *homogeneous space* we mean a locally compact Hausdorff space and a transitive group $G$ of homeomorphisms on $X$. Some examples of homeogeneous spaces are: (1) the line $\mathbf{R}$ with the group of translations; (2) the space $\mathbf{R}^n$ with the group of translations; (3) the space $\mathbf{R}^n$ with the group of rigid motions, i.e. rotations followed by translation; (4) the line $\mathbf{R}$ with the group of all linear functions $g(x) = ax + b$, $a > 0$. If $G$ is a group of homeomorphisms on $X$, we often write $gx$ for $g(x)$ and $gE$ for $g[E]$.

A Baire or Borel measure $\mu$ on a homogeneous space is said to be an invariant measure if $\mu(gE) = \mu(E)$ for each measurable set $E$ and for each $g \in G$. In this chapter we shall follow the usage of Chapter 13 and require Borel measures on $X$ to be finite on compact sets.[1] We should like to know conditions on a homogeneous space $(X, G)$ which ensure the existence of an invariant Baire or Borel measure which is not identically zero.

Not every transitive group of homeomorphisms on a locally compact space admits a nonzero invariant measure. If we take the

[1] In the more general usage of Chapters 12 and 15 the Hausdorff $\alpha$-dimensional measures $m_\alpha$ on $\mathbf{R}^n$ are invariant Borel measures for the group $G$ of rigid motions on $\mathbf{R}^n$. They satisfy the criterion of being finite on all compact sets only if $\alpha = n$.

group $G$ on $\mathbf{R}$ to consist of all linear functions $g(x) = ax + b$ with $a > 0$, then any two intervals $[\alpha, \beta)$ and $[\gamma, \delta)$ are congruent under $G$. Thus an invariant measure must give the same measure to $[0, 2)$, $[0, 1)$, and $[1, 2)$. But the first interval is the disjoint union of the other two, and so $\mu([0, 2)) = 2\mu([0, 2))$. Thus $\mu([0, 2)) = 0$, since $[0, 2) \subset [0, 2]$, which is compact. Consequently, all intervals have measure zero, and it follows that $\mu(\mathbf{R}) = 0$.

It is fairly clear why the group $G$ does not admit an invariant measure: An open set, such as $(0, 1)$, has "arbitrarily large" translates under $G$. We need a condition on $G$ to prevent this from happening, and a convenient condition turns out to be the requirement that $G$ be topologically equicontinuous. We discuss this notion in some detail in the next section.

### Problems

**1.** Let $X$ be a homogeneous space under a group $G$ of homeomorphisms. Show that if an invariant Baire or Borel measure $\mu$ on $X$ has a positive value for some compact set, then $\mu(O) > 0$ for every open set $O$.

**2.** Let $(X, G)$ be a homogeneous space.

**a.** If $\nu$ is a Baire measure with $\nu(gK) = \nu(K)$, whenever $K$ is a compact $G_\delta$, then $\nu$ is invariant, i.e., $\nu(gE) = \nu(E)$ for any Baire set $E$.

**b.** If $\nu$ is an inner regular Borel measure or a quasi-regular Borel measure, then $\nu$ is invariant if $\nu(gK) = \nu(K)$ for each compact set $K$.

## 2 Topological Equicontinuity

In this section we discuss a concept, called topological equicontinuity, which generalizes the notion of equicontinuity. Let $\mathfrak{F}$ be a family of continuous mappings of a topological space $X$ into a topological space $Y$, and $x$, $y$ points in $X$ and $Y$. We say that $\mathfrak{F}$ is **topologically equicontinuous** at $x$ and $y$ if, given any open set $O$ containing $y$, there are neighborhoods $U$ of $x$ and $V$ of $y$ such that for every $f \in \mathfrak{F}$ we have $f[U] \subset O$ whenever $f[U] \cap V$ is not empty. We say that $\mathfrak{F}$ is topologically equicontinuous at $x$ if it is topologically equicontinuous at $x$ and $y$ for each $y \in Y$, and that it is topologically equicontinuous if it is topologically equicontinuous for each $x \in X$. If $Y$ is a metric space and the family $\mathfrak{F}$ is equicontinuous at $x$, then it is topologically equicontinuous at $x$. The following proposition expresses a useful property of an equicontinuous family:

**1. Proposition:** *Let $\mathcal{F}$ be a family of maps from $X$ to $Y$ which is topologically equicontinuous at $x \varepsilon X$. Then, given $K \subset O \subset Y$ with $K$ compact and $O$ open, there is a neighborhood $U$ of $x$ such that $f[U] \subset O$ whenever $f \varepsilon \mathcal{F}$ and $f[U]$ meets $K$.*

***Proof:*** For each $y \varepsilon K$ there are neighborhoods $U(y)$ of $x$ and $V(y)$ of $y$ such that $f[U(y)] \subset O$ whenever $f[U(y)]$ meets $V(y)$. Cover $K$ by a finite number $V_1, \ldots, V_n$ of the neighborhoods $V(y)$, and let $U_1, \ldots, U_n$ be the corresponding neighborhoods of $x$. Set $U = \bigcap U_j$. Then if $f[U]$ meets $K$, it must meet some $V_j$. Since $f[U] \subset f[U_j]$, $f[U_j]$ meets $V_j$, and so $f[U_j] \subset O$. Thus $f[U] \subset f[U_j] \subset O$. ∎

If we take the open set $O$ of this proposition to be the complement of a closed set disjoint from $K$, then the proposition has the following reformulation:

**2. Corollary:** *Let $F$ and $K$ be disjoint subsets of $Y$ with $F$ closed and $K$ compact. If $\mathcal{F}$ is a family of maps from $X$ to $Y$ that is topologically equicontinuous at $x$, then there is a neighborhood $U$ of $x$ such that $f[U]$ does not meet both $F$ and $K$ for any $f \varepsilon \mathcal{F}$.*

We now turn our attention to the case when $G$ is a transitive group of homeomorphisms of a topological space $X$ with itself.

**3. Proposition:** *Let $G$ be a transitive group of homeomorphisms on a topological space $X$. If $G$ is topologically equicontinuous at some $x_0$ and $y_0$, it is topologically equicontinuous (at each $x$ and $y$).*

***Proof:*** Let $\langle x, y\rangle$ be an arbitrary pair of points of $X$. Since $G$ is transitive, there are elements $g$ and $h$ in $G$ such that $x = gx_0$ and $y = hy_0$. Given any open set $O$ containing $y$, the set $h^{-1}[O]$ is an open set containing $y_0$. If $G$ is topologically equicontinuous at $x_0$, $y_0$, then there are neighborhoods $U_0$ of $x_0$ and $V_0$ of $y_0$ such that for each $f \varepsilon G$ we have $f[U_0] \subset h^{-1}[O]$ whenever $f[U_0]$ meets $V_0$. Set $U = g[V_0]$ and $V = h[V_0]$. Then $U$ and $V$ are neighborhoods of $x$ and $y$, respectively. Suppose that $f[U]$ meets $V$ for some $f \varepsilon G$, i.e., that there is a $u \varepsilon U$ such that $f(u) = v \varepsilon V$. Then $h^{-1}(v) \varepsilon V_0$, and

$g^{-1}(u) \in U_0$, and so $h^{-1} \circ f \circ g[U_0] \subset h^{-1}[O]$. This implies that $f[U] \subset O$, and we have shown the topological equicontinuity of $G$ at $x$ and $y$. ∎

We leave the proof of the following corollary to the reader.

**4. Corollary:** *Let $G$ be a transitive group of homeomorphisms on a topological space $X$, and $p$ a point of $X$. Then $G$ is topologically equicontinuous iff given any open set $O$ containing $p$, there is a neighborhood $U$ of $p$ such that for any $f \in G$ we have $f[U] \subset O$ whenever $f[U]$ meets $U$.*

The property of topological equicontinuity that we have introduced is slightly stronger than the property of even continuity due to Kelley and Morse (see [9], p. 234). Consequently, topological equicontinuity has all of the consequences Kelley and Morse derive from even continuity, including a general version of the Ascoli theorem. In the presence of enough compactness (in $Y$) the notions of even continuity, topological equicontinuity, and equicontinuity coincide (cf. [9], Theorem 7.23). I have found topological equicontinuity more convenient to work with than even continuity. One can give examples of families which are evenly continuous and not topologically equicontinuous, but I do not know whether or not it is possible to have a transitive group of homeomorphisms that is evenly continuous and not topologically equicontinuous.

## Problems

**3. a.** Let $\mathcal{F}$ be a family of maps from a topological space $X$ to a metric space $(Y, \rho)$. Recall that $\mathcal{F}$ is said to be equicontinuous at $x$, if given $\epsilon > 0$, there is a neighborhood $U$ of $x$ such that $\rho(f(x), f(x')) < \epsilon$ for all $f \in \mathcal{F}$ and all $x' \in U$. Show that if $\mathcal{F}$ is equicontinuous at $x$, then $\mathcal{F}$ is topologically equicontinuous at $x$.

**b.** Assume that $Y$ is compact. Show that topological equicontinuity at $x$ implies equicontinuity.

**4.** Let $X$ be a metric space and $G$ its group of isometries. Show that $G$ is equicontinuous and hence topologically equicontinuous.

**5.** A family $\mathcal{F}$ of maps of a topological space $X$ into a topological space $Y$ is said to be evenly continuous (at $x$ and $y$) if, given any open set $O$ containing $y$, there are neighborhoods $U$ of $x$ and $V$ of $y$ with $f[U] \subset O$ whenever $f(x) \in V$.

**a.** Show that topological equicontinuity implies even continuity.

**b.** Give an example of a family $\mathcal{F}$ of maps of $\mathbf{R}$ to $\mathbf{R}$ that is evenly continuous but not topologically equicontinuous.

**c.** Let $\mathcal{F}$ be a family of maps of $X$ into $Y$. Suppose that $x \varepsilon X$ has a neighborhood $W$ such that $\mathcal{F}[W]$ has compact closure. Then even continuity of $\mathcal{F}$ at $x$ implies topological equicontinuity at $x$. [Use Kelley [9], Theorems 6.30 and 7.23.]

**6.** Let $\mathcal{F}$ be a family of maps from a topological space $X$ to a uniform space $Y$. If $\mathcal{F}$ is equicontinuous at $x$, it is topologically equicontinuous at $x$ (cf. Kelley [9], p. 232).

**7. a.** Let $G$ be a topologically equicontinuous transitive group of homeomorphisms on a Hausdorff space $X$. Show that there is a uniform structure $\mathcal{V}$ for $X$ such that each $g \varepsilon G$ maps each element of $\mathcal{V}$ to itself; i.e., if $V \varepsilon \mathcal{V}$, then for each $g \varepsilon G$ we have $\{\langle gu, gv\rangle : \langle u, v\rangle \varepsilon V\} = V$. (Fix a point $p \varepsilon X$. For each neighborhood $U$ of $p$ let $V_U = \{\langle x, y\rangle : \exists g \varepsilon G$ with $gx \varepsilon U$ and $gy \varepsilon U\}$. Then $\{V_U\}$ is a base for the desired uniform structure.)

**b.** Suppose that the topology for $X$ has a countable base at $p$. Then $X$ is metrizable by a metric such that $G$ is a group of isometries.

## 3 The Existence of Invariant Measures

Let $G$ be a transitive topologically equicontinuous group of homeomorphisms on a locally compact Hausdorff space $X$. It is the purpose of this section to demonstrate the existence of a nonzero Borel measure which is invariant under $G$. By a nonzero Borel measure we mean here a Borel measure $\mu$ that is finite on each compact set and positive on some compact set. If a nonzero Borel measure is invariant, we have $\mu O > 0$ for each open set $O$.

If $E$ is a subset of $X$ and $g$ an element of $G$, we call $gE = g[E]$ the *translate* of $E$ by $g$. A set $F$ is said to be a translate of $E$ if there is some $g \varepsilon G$ such that $F = gE$. We sometimes express this by saying that $E$ and $F$ are *congruent* (under $G$). Thus an invariant measure is one that assigns the same value to each of a pair of congruent sets.

It is convenient to fix a point $p \varepsilon X$ as a base point. Let $\mathcal{U}$ be the collection of open sets $U$ with $p \varepsilon U$ and $\bar{U}$ compact. Since $X$ is locally compact, $\mathcal{U}$ is a base for the topology at $p$, and the collection of translates of sets in $\mathcal{U}$ is a base for the topology of $X$. Let $H$ be a compact set containing $p$ in its interior. Thus $H^\circ \varepsilon \mathcal{U}$. We shall use $H$ to normalize our constructions so that $\mu H = 1$.

If $E$ is any bounded set and $F$ any set with nonempty interior, we can always cover $E$ by a finite number of translates of $F$ (since we can cover $\bar{E}$ by a finite number of translates of $F^\circ$). We define the *covering number* $[E:F]$ of $E$ by $F$ to be the least number of translates of $F$ required to cover $E$. Thus $[E:F]$ is a nonnegative integer and positive if $E \neq \emptyset$. This covering number has the invariance property

$$[gE:F] = [E:F] = [E:gF].$$

If $A \subset B$, then $[A:F] \leq [B:F]$. If $A$ is a bounded set with nonempty interior, the covering numbers $[E:A]$ and $[A:F]$ are defined, and we have

$$[E:F] \leq [E:A] \cdot [A:F].$$

For we can cover $E$ by $[E:A]$ translates of $A$ and cover each translate of $A$ by $[A:F]$ translates of $F$. This gives a cover of $E$ by $[E:A] \cdot [A:F]$ translates of $F$.

If we fix $E$ and look at the covering numbers $[E:U]$ for the various sets in $\mathfrak{U}$, we find that $[E:U]$ gets larger as $\mathfrak{U}$ gets smaller. To control this we consider, for each $U$, the normalized covering ratio

$$\xi_U(E) = \frac{[E:U]}{[H:U]}.$$

Since

$$[E:U] \leq [E:H] \cdot [H:U],$$

we have

$$\xi_U(E) \leq [E:H].$$

Thus for a fixed $E$ the numbers $\xi_U(E)$ are bounded independently of $U$.

For a fixed $U$ the set function $\xi_U$ is nonnegative and invariant,

$$\xi_U[gE] = \xi_U[E].$$

It is also subadditive,

$$\xi_U(E \cup F) \leq \xi_U(E) + \xi_U(F).$$

We have $\xi_U(H) = 1$, and $\xi_U(E) \leq \xi_U(F)$ if $E \subset F$.

Let $K_1$ and $K_2$ be disjoint compact sets. Since $G$ is topologically equicontinuous at $p$, there is a $U \in \mathfrak{U}$ such that no translate of $U$

meets both $K_1$ and $K_2$ (Corollary 4). Hence any covering of $K_1 \cup K_2$ by translates of $U$ must be the disjoint union of coverings of $K_1$ and of $K_2$. Thus

$$\xi_U(K_1 \cup K_2) \geq \xi_U(K_1) + \xi_U(K_2).$$

This suggests that, if we could somehow take the limit of $\xi_U(E)$ as $U$ gets smaller and smaller, this limit would be an invariant set function which is additive on compact sets. One way of doing this is by generalizing the notion of Banach limits discussed in Problem 10.20.

Let $\Xi$ be the space of all bounded real-valued functions on $\mathfrak{U}$. We denote the value of $\xi \in \Xi$ at $U \in \mathfrak{U}$ by $\xi_U$. Then $\Xi$ becomes a linear space if we define

$$(\xi + \eta)_U = \xi_U + \eta_U$$

and

$$(\alpha\xi)_U = \alpha\xi_U.$$

We define $\overline{\lim}$ for an element of $\Xi$ by

$$\overline{\lim}\ \xi = \inf_{U \in \mathfrak{U}} \sup_{V \subset U} \xi_V.$$

Similarly,

$$\underline{\lim}\ \xi = \sup_{U \in \mathfrak{U}} \inf_{V \subset U} \xi_V.$$

It is easy to verify that

$$\overline{\lim}\ (\xi + \eta) \leq \overline{\lim}\ \xi + \overline{\lim}\ \eta$$

and

$$\overline{\lim}\ \alpha\xi = \alpha\ \overline{\lim}\ \xi$$

for $\alpha \geq 0$. Also,

$$\underline{\lim}\ \xi = -\overline{\lim}\ (-\xi).$$

If we now take $p(\xi) = \overline{\lim}\ \xi$, we see that $p$ is subadditive and positively homogeneous. By the Hahn–Banach Extension Theorem the zero functional on the zero subspace of $\Xi$ extends to a linear functional Lim defined on all of $\Xi$ with

$$\text{Lim}\ \xi \leq \overline{\lim}\ \xi.$$

We have

$$-\text{Lim}\ \xi = \text{Lim}\ (-\xi) \leq \overline{\lim}\ (-\xi) = -\underline{\lim}\ \xi,$$

whence

$$\underline{\lim}\ \xi \le \text{Lim}\ \xi \le \overline{\lim}\ \xi.$$

If $\xi$ is a constant sequence from some $U$ on, i.e., if $\xi_V = c$ for all $V \subset U$, then $\overline{\lim}\ \xi = c$ and $\underline{\lim}\ \xi = c$. Consequently,

$$\text{Lim}\ \xi = c,$$

if $\xi$ is ultimately constant. If for some $U$ we have $\xi_V = \eta_V$ for all $V \subset U$, we say that $\xi$ and $\eta$ are ultimately equal. In this case we have $\xi - \eta$ ultimately 0, and so

$$\text{Lim}\ \xi - \text{Lim}\ \eta = \text{Lim}\ (\xi - \eta) = 0,$$

whence

$$\text{Lim}\ \xi = \text{Lim}\ \eta.$$

We say that $\xi \ge \eta$ if $\xi_U \ge \eta_U$ for all $U$. If $\xi \ge 0$, then $\underline{\lim}\ \xi \ge 0$. Hence

$$\text{Lim}\ \xi \ge 0 \qquad \text{for } \xi \ge 0.$$

Thus if $\xi > \eta$, we have $\xi - \eta \ge 0$, and so

$$\text{Lim}\ \xi - \text{Lim}\ \eta = \text{Lim}\ (\xi - \eta) \ge 0.$$

Hence

$$\text{Lim}\ \xi \ge \text{Lim}\ \eta \qquad \text{for } \xi \ge \eta.$$

For each bounded set $E$, the covering ratios $\xi_U(E)$ give an element of $\Xi$. Define a function $\lambda$ on compact sets by setting

$$\lambda(K) = \text{Lim}\ \xi_U(K).$$

Since $\xi_U(gK) = \xi_U(K)$, we have

$$\lambda(gK) = \lambda(K)$$

for each $g \varepsilon G$. Suppose $K_1$ and $K_2$ are disjoint compact sets. Then there is a $U$ in $\mathcal{U}$ such that no translate of $U$ meets both $K_1$ and $K_2$. If $V \subset U$, no translate of $V$ can meet both $K_1$ and $K_2$, and so

$$\xi_V(K_1 \cup K_2) = \xi_V(K_1) + \xi_V(K_2)$$

for all $V \subset U$. Thus

$$\begin{aligned}\text{Lim}\ \xi_V(K_1 \cup K_2) &= \text{Lim}\ (\xi_V(K_1) + \xi_V(K_z)) \\ &= \text{Lim}\ \xi_V(K_1) + \text{Lim}\ \xi_V(K_2).\end{aligned}$$

Consequently,

$$\lambda(K_1 \cup K_2) = \lambda(K_1) + \lambda(K_2)$$

for any two disjoint compact sets. For our normalizing set $H$ we have $\xi_U(H) = 1$, whence $\lambda(H) = 1$. If $E \subset F$, then $\xi_U(E) \leq \xi_U(F)$ and so $\lambda E \leq \lambda F$.

We summarize our results in the following propositions:

**5. Proposition:** *Let $G$ be a transitive and topologically equicontinuous group of homeomorphisms on a locally compact Hausdorff space $X$. Let $H$ be a compact set with nonempty interior. Then there is a real-valued function $\lambda$, defined on all compact sets, such that*

i. $K_1 \subset K_2 \Rightarrow \lambda K_1 \subset \lambda K_2$.
ii. $\lambda(K_1 \cup K_2) = \lambda K_1 + \lambda K_2$ *if* $K_1 \cap K_2 = \emptyset$.
iii. $\lambda H = 1$.
iv. $\lambda(gK) = \lambda K$ *for all* $g \in G$.

The set function $\lambda$ is a content in the sense of Section 3 of Chapter 13. By Proposition 13.21 there is a quasi regular Borel measure $\bar{\mu}$ such that

$$\bar{\mu}O = \sup \{\lambda K\colon K \subset O,\ K \text{ compact}\}.$$

Since $\lambda(gK) = \lambda K$ for all $K$, we have $\bar{\mu}(gO) = \bar{\mu}O$ for all open sets. Since $\bar{\mu}$ is outer regular, $\bar{\mu}(gE) = \bar{\mu}E$ for all measurable sets $E$. If we define

$$\underline{\mu}E = \sup \{\bar{\mu}B\colon B \subset E,\ B \text{ a } \sigma\text{-bounded Borel set}\},$$

we obtain an inner regular Borel measure as in the proof of Theorem 13.22. This measure is clearly invariant. For the set $H$ we have, by Proposition 13.21,

$$\bar{\mu}H \geq \lambda H = 1.$$

Since $\underline{\mu}$ and $\bar{\mu}$ agree on $\sigma$-bounded sets,

$$\underline{\mu}H = \bar{\mu}H \geq 1.$$

Thus $\bar{\mu}$ and $\underline{\mu}$ are nontrivial Borel measures and thus positive on all open sets. We summarize these results in the following Theorem:

**6. Theorem:** *Let $G$ be a transitive topologically equicontinuous group of homomorphisms on a locally compact Hausdorff space $X$. Then there is a quasi regular Borel measure $\bar{\mu}$ on $X$ which is invariant* *under $G$, finite on compact sets, and positive on every open set. There is also an inner regular Borel measure $\underline{\mu}$ with these properties.*

It can be shown that the measures $\bar{\mu}$ and $\underline{\mu}$ are unique, but the proof in the general context of topologically equicontinuous groups is beyond the scope of this book. We will, however, establish unicity in several important cases in Section 6.

It should be noted that the only place we have used the topological equicontinuity of $G$ is to infer that, given disjoint compact sets $K_1$ and $K_2$, there is a neighborhood $U$ of $p$ none of whose translates meet both $K_1$ and $K_2$. This is slightly weaker than topological equicontinuity, which is equivalent (on locally compact spaces) to the statement that, given a compact set $K$ and a closed set $F$ that are disjoint, we can find a $U$ such that no translate of $U$ meets both $K$ and $F$. Thus we can still assert the existence of an invariant Borel measure if $G$ satisfies the weaker condition. I do not know whether or not the invariant measures are unique in this more general case.

## Problems

**8.** A group $G$ is said to be almost transitive if for each $x$ the set $\{gx: g \varepsilon G\}$ is dense in $X$. Show that Theorem 6 is still true if we only require $g$ to be almost transitive.

**9.** Let $\mu$ be an invariant measure on the homogeneous space $(X, G)$, and $f$ a $\mu$-integrable function on $X$. Then

$$\int_X f(gx)\, d\mu(x) = \int_X f(x)\, d\mu(x).$$

## 4 Topological Groups

Let $G$ be an abstract group. We usually write the group operation multiplicatively so that the product of $x$ and $y$ is written $xy$ and the inverse of $x$ is denoted by $x^{-1}$. We denote the identity elements by $e$. If $A$ and $B$ are subsets of $G$, we write $AB$ for the set $\{z: z = xy, x \varepsilon A, y \varepsilon B\}$ and $A^{-1}$ for the set $\{z: z = x^{-1}, x \varepsilon A\}$. We also simplify $\{x\}A$ to $xA$ and $A\{y\}$ to $Ay$.

A topology $\mathfrak{J}$ for $G$ is said to be an *invariant* topology for $G$ if multiplication is continuous in each factor (separately). If $\mathfrak{J}$ is an invariant topology on $G$, then $G$ can be regarded as a group of homomorphisms on itself in two different ways: For each $g \varepsilon G$ let $l_g$ be the mapping of $G$ into $G$ defined by

$$l_g(x) = gx.$$

Then $l_g$ is one-to-one onto and continuous, since multiplication is continuous in the second factor. The inverse map is $l_{g^{-1}}$, which is also continuous. Thus $l_g$ is a homeomorphism of $G$ onto $G$. We have

$$l_g l_h = l_{gh},$$

and the map $l$ which takes $g$ into $l_g$ is a homomorphism of $G$ onto a group $G_L$ of homeomorphisms of $G$ onto itself. Since different elements of $G$ give different homeomorphisms, $l$ is an isomorphism. The group $G_L$ is called the *group of left translations* of $G$. We sometimes identify $G_L$ with $G$ and speak of $G$ acting on $G$ by left translation. The group $G_L$ is transitive on $G$, since $y = (yx^{-1})x$.

We may also consider the maps $r_g$ defined by $r_g(x) = xg$. These again form a transitive group of homeomorphisms of $G$ onto $G$. Since

$$r_g r_h = r_{hg},$$

we see that the map $r$ is an anti-isomorphism of $G$ onto a group $G_R$ of homeomorphisms of $G$ onto itself. The group $G_R$ is called the group of *right translations* of $G$ onto itself. In the case of a commutative group $G$, $l_g = r_g$ and $G_R = G_L$.

It is important to note that an invariant topology for $G$ is determined by a neighborhood base at any one point. This and related facts are stated in the following lemmas, whose proofs are left to the reader.

**7. Lemma:** *Let $G$ be a group and $\mathfrak{J}$ a topology for $G$. Then $\mathfrak{J}$ is an invariant topology for $G$ iff, given a base $\mathfrak{U}$ for $\mathfrak{J}$ at $e$, the collections $g\mathfrak{U} = \{V: V = gU, U \varepsilon \mathfrak{U}\}$ and $\mathfrak{U}g = \{V: V = Ug, U \varepsilon \mathfrak{U}\}$ are both bases for $\mathfrak{J}$ at $g$.*

**8. Lemma:** *Let $(G, \mathfrak{J})$ and $(H, \mathfrak{S})$ be groups with invariant topologies. Then a homomorphism $\varphi: G \to H$ is continuous iff it is continuous at $e$.*

A topology $\mathfrak{J}$ for $G$ is said to be a **group topology** for $G$ if multiplication is jointly continuous and inversion is continuous, that is, if the map $\langle x, y\rangle \to xy$ is a continuous map from $G \times G$ to $G$ and the map

$x \to x^{-1}$ is continuous from $G$ to $G$. A group topology is an invariant topology. By a **topological group** we mean a group $G$ endowed with a *Hausdorff* group topology. The two conditions for $\mathfrak{J}$ to be a group topology can be combined into the single condition that the map $\langle x, y\rangle \to xy^{-1}$ be continuous from $G \times G$ to $G$.

**9. Lemma:** *Let $\mathfrak{J}$ be a group topology for $G$ and $O$ an open set containing $e$. Then there is an open set $U$ containing $e$ with $U = U^{-1}$ and $U \cdot U \cdot U \subset O$.*

***Proof:*** Since multiplication is jointly continuous at $e$, there are neighborhoods $W_1$ and $W_2$ of $e$ such that $W_1 W_2 \subset O$. The set $W = W_1 \cap W_2$ is open and satisfies $W \cdot W \subset O$. Repeating this process with $O$ replaced by $W$, there is an open set $V$ containing $e$ such that $V \cdot V \subset W$. Then

$$V \cdot V \cdot V \subset V \cdot V \cdot V \cdot V \subset W \cdot W \subset O.$$

Since $V^{-1}$ is open, the set $U = V \cap V^{-1}$, is an open set with the desired properties. ▮

The next two statements express useful properties of topological groups. We leave the proofs to the reader.

**10. Proposition:** *Let $G$ be a topological group. Then there is a base $\mathfrak{U}$ for the topology at $e$ satisfying:*

i. *Given $U \in \mathfrak{U}$, $\exists V \in \mathfrak{U}$ with $V \cdot V \subset U$.*
ii. *If $U \in \mathfrak{U}$, then $U^{-1} = U$.*
iii. *If $U, V \in \mathfrak{U}$, then $U \cap V \in \mathfrak{U}$.*
iv. *If $U \in \mathfrak{U}$ and $g \in G$, then $gUg^{-1} \in \mathfrak{U}$.*
v. *If $u \in U \in \mathfrak{U}$, then $\exists V \in \mathfrak{U}$ with $uV \subset U$ and $Vu \subset U$.*
vi. $\bigcap U = \{e\}$.

*Conversely, given any collection $\mathfrak{U}$ of subsets of an abstract group $G$ satisfying conditions (i) through (vi) is the base at $e$ for a group topology on $G$.*

**11. Lemma:** *Let $G$ be a topological group and $E$ any subset of $G$. Then*

$$\bar{E} = \bigcap_{U \in \mathfrak{U}} \{U \cdot E\} = \bigcap_{U \in \mathfrak{U}} \{E \cdot U\},$$

*where $\mathfrak{U}$ is a base for the neighborhoods at $e$.*

**12. Lemma:** *Let $G$ be a group and $U$ a set containing $e$. Then*

$$\{w\colon w \varepsilon gU,\ gU \text{ meets } U\} = U \cdot U^{-1} \cdot U,$$

*and*

$$\{w\colon w \varepsilon Ug,\ Ug \text{ meets } U\} = U \cdot U^{-1} \cdot U.$$

***Proof:*** We note that $gU$ meets $U$ if and only if there are $u_1$, $u_2$ in $U$ with $u_1 = gu_2$. Thus $\{g\colon gU \text{ meets } U\}$ is $U \cdot U^{-1}$, and so $\{w\colon w \varepsilon gU,\ gU \text{ meets } U\}$ is $U \cdot U^{-1} \cdot U$. Similarly for the second identity. ▮

**13. Proposition:** *Let $(G, \mathfrak{J})$ be a topological group. Then the group of left translations on $G$ is a transitive and topologically equicontinuous group of homeomorphisms on $G$. So is the group of right translations.*

***Proof:*** This is an immediate consequence of Lemma 12 and Proposition 3. ▮

A Borel measure $\mu$ on a topological group $G$ is said to **left invariant** if $\mu(gE) = \mu E$ for each $g \varepsilon G$ and each Borel set $E$. Thus $\mu$ is left invariant if it is invariant under the group of left translations. Similarly, a Borel measure $\nu$ is said to be right invariant if $\nu(Eg) = \nu E$, that is, if it is invariant under the group of right translations. Since each of these groups of translations is transitive and topologically equicontinuous, Theorem 6 gives us the following theorem:

**14. Theorem:** *Let $G$ be a locally compact topological group. Then there is an inner regular Borel measure $\underline{\mu}$ and a quasi regular Borel measure $\bar{\mu}$ which are left invariant and positive on open sets. Similarly, there are right invariant measures $\underline{\nu}$ and $\bar{\nu}$ which are inner regular and quasi regular, respectively, and positive on open sets.*

An invariant measure on a locally compact group is often called a **Haar measure**; left Haar measure if it is left invariant, and right Haar measure if it is right invariant. We usually want it to satisfy some regularity conditions and also allow it to be complete. For definiteness let us require it to be inner regular and complete. Thus a *left Haar measure* on a locally compact group $G$ is the completion of an inner regular left invariant Borel measure, and similarly for a right

Haar measure. We shall show in Section 6 that the left and right Haar measures are unique under these regularity conditions apart from a multiplicative constant, although left Haar measure is generally not equal to right Haar measure on noncommutative groups.

The following proposition about locally compact groups will be useful later. Problem 19 gives some hints for its proof.

**15. Proposition:** *Let G be a locally compact group. Then G has a $\sigma$-compact subgroup H which is both open and closed.*

We will find the following proposition and its corollary useful later. We leave the proof to the reader.

**16. Proposition:** *Let G be a locally compact group and $\varphi$ a continuous real-valued function on G with compact support. Then, given $\epsilon > 0$, there is a neighborhood U of e such that $|\varphi(ux) - \varphi(x)| < \epsilon$ and $|\varphi(xu) - \varphi(x)| < \epsilon$ for all $u \in U$ and all $x \in G$.*

This proposition says in effect that $\varphi$ is uniformly continuous.

**17. Corollary:** *Let G be a locally compact group, $\mu$ a Baire measure on G, and $\varphi$ a continuous function with compact support. Then the function $\Delta(x) = \int_G \varphi(yx)\, d\mu(y)$ is continuous. If $\mu$ is positive on open sets, then $\Delta(x) > 0$.*

## Problems

**10.** Prove Lemma 7.

**11.** Prove Lemma 8.

**12.** Let $G$ be a group with a topology $\mathfrak{J}$. Show that $\mathfrak{J}$ is a group topology if and only if the map $\langle x, y\rangle \to xy^{-1}$ is continuous.

**13.** Prove Proposition 10.

**14.** Prove Lemma 11.

**15.** Let $G$ be a group and $\mathcal{U}$ a collection of sets containing $e$. Then $\mathcal{U}$ is the base at $e$ for a group topology for $G$ if and only if the following conditions are satisfied:

i. Given $U \in \mathcal{U}$, $\exists V \in \mathcal{U}$ with $V \cdot V \subset U$.
ii. If $U \in \mathcal{U}$, $\exists V \in \mathcal{U}$ with $V \subset U^{-1}$.
iii. If $U, V \in \mathcal{U}$, $\exists W \in \mathcal{U}$ with $W \subset U \cap V$.

iv. If $U \in \mathcal{U}$, $g \in G$, $\exists V \in \mathcal{U}$ with $V \subset gUg^{-1}$.
v. If $u \in U \in \mathcal{U}$, then $\exists V \in \mathcal{U}$ with $uV \subset U$.

The topology will be Hausdorff iff $\bigcap U = \{e\}$.

**16.** Let $G$ be a locally compact topological group.

**a.** If $\mu$ is a left invariant Borel measure and $f$ a $\mu$ integrable function, then

$$\int_G f(yx)\, d\mu(x) = \int_G f(x)\, d\mu(x)$$

for each $y \in G$.

**b.** If $\mu$ is an inner regular Borel measure such that the conclusion of (a) holds for each continuous function $f$ with compact support, then $\mu$ is left invariant.

**c.** If $\nu$ is a right invariant Borel measure and $f$ is integrable with respect to $\nu$, then

$$\int_G f(xy)\, d\nu(x) = \int_G f(x)\, d\nu(x).$$

**17.** Let $\mu$ be a left invariant Borel measure, and define a Borel measure $\nu$ by $\nu[E] = \mu[E^{-1}]$. Then $\nu$ is a right invariant measure.

**18.** Let $G$ be a locally compact group.

**a.** The map that takes $x$ to $x^{-1}$ is a Borel isomorphism of $G$ to $G$.

**b.** The map that takes $\langle x, y\rangle$ to $xy$ is a Borel measurable map of $G \times G$ to $G$.

**c.** The map that takes $\langle x, y\rangle$ to $xy^{-1}$ is a Borel measurable map of $G \times G$ to $G$.

**19.** Let $G$ be a topological group.

**a.** If $K \subset G$ is compact, so are $gK$ and $Kg$ for each $g \in G$.

**b.** If $K$ is compact, so is $K^{-1}$.

**c.** If $K_1$ and $K_2$ are compact subsets of $G$, then so are $K_1K_2$ and $K_1K_2^{-1}$.

**d.** Statements $a$, $b$, $c$ remain true if 'compact' is replaced by 'bounded' or '$\sigma$-bounded'.

**e.** If $O$ is open and $E$ is an arbitrary subset, then $OE$ and $EO$ are open.

**20.** Prove Proposition 15.

**a.** Show that an open subgroup $H$ of a topological group is closed. [Hint: $\sim H = \bigcup_{g \notin H} gH$].

**b.** Let $U$ be a neighborhood of $e$ with $U = U^{-1}$ and $\bar{U}$ compact. Set $U_n = U_{n-1} \cdot U$, $U_1 = U$. Then $U_n$ is open and $\bar{U}_n \subset U_{n+1}$.

**c.** Let $H = \bigcup U_n$. Then $H$ is $\sigma$-compact and open.

**d.** Show that $H$ is a subgroup.

**21.** Prove Proposition 16.

**a.** Let $K$ be the support of $\varphi$. For each $x \in K$ there is a neighborhood $V_x$ of $e$ such that $|\varphi(ux) - \varphi(vx)| < \epsilon/2$ and $|\varphi(xu) - \varphi(xv)| < \epsilon/2$ for $u, v \in V_x$.

**b.** For each $V_x$ there is a neighborhood $U_x$ of $e$ with $U_x \cdot U_x \subset V_x$. Show that there are a finite number $x_1, \ldots, x_n$ such that the sets $\{x_i U_{x_i}\}$ and the sets $\{U_{x_i} x_i\}$ cover $K$.

**c.** The set $U = \bigcap U_{x_i}$ can be used for the proposition.

**22.** Prove Corollary 17.

## 5 Group Actions and Quotient Spaces

Let $G$ be a topological group and $X$ a topological space. By a **left action** of $G$ on $X$ we mean a continuous map $\varphi: G \times X \to X$ such that

$$\varphi(e, x) = x$$

and

$$\varphi(g, \varphi(h, x)) = \varphi(gh, x).$$

We often write $gx$ for $\varphi(g, x)$. With this notation the conditions on $\varphi$ become

$$ex = x,$$

$$g(hx) = (gh)x.$$

We sometimes say *action* for left action and speak of $G$ *acting on* $X$. An action is said to be **open** if $O \cdot x$ is an open subset of $X$ for each open set $O$ in $G$ and each $x \in X$. Thus the action $\varphi$ is open iff for each $x \in X$ the map $\varphi_x: G \to X$ defined by $\varphi_x(g) = gx$ is an open map.

An action of $G$ on $X$ is said to be transitive if for each $x$ and $y$ in $X$ there is a $g \in G$ with $y = gx$. A transitive action is open if the map $\varphi_x$ is open for one $x \in X$. In fact, it will be open if there is an $x \in X$ such that $Ux$ is open for each $U$ in a base $\mathcal{U}$ at $e$ for the topology of $G$. In the case of an open action, we speak of $G$ acting openly on $X$.

An action is said to be *proper* if the map $\varphi_x$ is a proper map for each $x \in X$, i.e. if $\varphi_x^{-1}[K]$ is a compact subset of $G$ for each compact set $K \subset X$. In the case of a proper action we speak of $G$ acting properly on $X$. The following proposition, whose proof we leave to the reader, is sometimes useful:

**18. Proposition:** *If a locally compact group $G$ acts transitively and properly on a locally compact space, the action is open.*

Let $G$ be a topological group and $F$ a closed subgroup. For $g \in G$ the set $gF$ is called a (left) coset of $F$. Note that two cosets of $F$ are either disjoint or identical. Let $H$ be the collection of all cosets of $F$ and let $\varphi: G \to H$ be the map that takes each $g$ to the coset $gF$. Define a set $O$ in $H$ to be open if the union of the cosets in $O$ is an open set of $G$. This defines a Hausdorff topology for $H$ called the quotient topology. The map $\varphi$ is a continuous open map and is a left action of $G$ on $H$. The space $H$ with this topology is called the (left) quotient of $G$ by $F$ and is denoted by $G/F$. The mapping $\varphi$ is called the natural projection of $G$ onto $G/F$. The quotient $G/F$ can be furnished with a group structure so that $\varphi$ is a homomorphism if and only if $F$ is a normal subgroup.

If $G$ is locally compact, so is $G/F$, and the map $\varphi$ is proper if and only if $F$ is compact.

Let $\psi$ be a transitive action of the topological group $G$ on $X$, and let $H$ be the isotropy subgroup of a point $p \in X$. Then $H$ is closed, and there is a one-to-one continuous map $\theta: G/H \to X$ such that $\psi = \theta \circ \varphi$. The map $\theta$ is a homeomorphism iff $\psi$ is open.

Since we have defined left actions, we also define right actions. A continuous map $\varphi$ of $G \times X$ into $X$ is called a **right action** if

$$\varphi(e, x) = x$$

and

$$\varphi(g, \varphi(h, x)) = \varphi(hg, x).$$

A right action is often written $\varphi(g, x) = xg$. The conditions then become

$$xe = x$$

and

$$(xh)g = x(hg).$$

The set of right translations of a topological group gives us a right action of the group $G$ on itself.

Let $G^*$ be the group $G$ with the group operation $*$ defined by $x * y = yx$. Then $G$ and $G^*$ are anti-isomorphic, and a right action by $G$ is a left action by $G^*$. Thus all statements about left actions can be carried over to right actions.

## Problems

**23.** Let $\varphi$ be a transitive action of $G$ on $X$. Then $\varphi$ is an open action iff for some $p \in X$ and some base $\mathcal{U}$ for the topology of $G$ at $e$ we have $p$ in the interior of $\varphi_p[U]$ for each $U \in \mathcal{U}$.

**24.** Let $\varphi$ be an action of $G$ on $X$ and $p \in X$. The *isotropy* subgroup $H_p$ of $G$ at $p$ is $H_p = \{g \in G: gp = p\}$.

**a.** Show that the isotropy subgroup is in fact a subgroup and that it is closed.

**b.** Let $E$ be a subset of $G$. Then

$$\varphi_p^{-1}[\varphi_p[E]] = E \cdot H_p.$$

**c.** If $q$ is another point of $X$ and $q = gp$, then the isotropy group at $q$ is $H_q = g[H_p]g^{-1}$.

**25.** Prove Proposition 18.

## 6 Unicity of Invariant Measures

The purpose of this section is to show that, in a number of important cases, there is only one invariant Baire measure, apart from multiplication by a positive constant. It would take us too far afield to give unicity results in the general setting of Section 1, but we will show that the left invariant and the right invariant Baire measures on a locally compact group are unique, although they may differ from each other. We also treat the case of a locally compact group acting properly on a locally compact space. In these cases we can deduce unicity for invariant Borel measures satisfying a given regularity condition.

Let $\mu$ and $\nu$ be Baire measures on a locally compact Hausdorff space $X$ with

$$\int_X \varphi \, d\mu = \int_X \varphi \, d\nu$$

for each $\varphi \in C_c(X)$, then

$$\mu E = \nu E$$

for all $\sigma$-bounded Baire sets $E$. If $\lambda$ is a Baire measure on the locally compact space $Y$, then all Baire subsets of $X \times Y$ are measurable with respect to the product measure $\mu \times \lambda$. Thus every function $\varphi$ in $C_c(X \times Y)$ is integrable with respect to $X \times Y$. We shall be concerned principally with the case $G \times G$, where $G$ is a locally compact group. If $\varphi$ and $\psi \in C_c(G)$, then functions such as $h(x, y) = \varphi(x)\psi(xy^{-1})$ belong to $C_c(G \times G)$ and are thus integrable with respect to $\mu \times \lambda$.

Let $G$ be a compact group, and let $\mu$ be a left invariant Haar measure on $G$ and $\nu$ a right Haar measure. Normalize them by requiring $\mu G = \nu G = 1$. If $\varphi$ is any continuous real-valued function on $G$, we have

$$\begin{aligned}
\int_G \varphi(x)\, d\mu(x) &= \int_G \varphi(yx)\, d\mu(x) \\
&= \int_G \left[ \int_G \varphi(yx)\, d\mu(x) \right] d\nu(y) \\
&= \int_G \left[ \int_G \varphi(yx)\, d\nu(y) \right] d\mu(x) \\
&= \int_G \left[ \int_G \varphi(y)\, d\nu(y) \right] d\mu(x) \\
&= \int_G \varphi(y)\, d\nu(y).
\end{aligned}$$

Here we have made use of the left invariance of $\mu$, Fubini's Theorem applied to $\varphi(yx)$, and the right invariance of $\nu$. Since

$$\int \varphi\, d\mu = \int \varphi\, d\nu$$

for each function $\varphi \in C(G)$, we have $\mu = \nu$. Thus every normalized left invariant Baire measure is equal to every normalized right invariant Baire measure. Consequently, any two normalized left invariant Baire measures on $G$ must be equal, and so must any two right invariant Baire measures. We have thus established the following theorem:

**19. Theorem:** *Let $G$ be a compact group. Then there is a unique left invariant Baire measure $\mu$ with $\mu G = 1$. It is also the unique right invariant Baire measure with $\mu G = 1$.*

If a Borel measure on a compact space is inner regular, it is regular and determined by its values on the compact $G_\delta$'s. This gives us the following corollary:

**20. Corollary:** *Let $G$ be a compact group. There is a unique regular left invariant Borel measure $\mu$ on $G$ which satisfies $\mu G = 1$. It is also the unique regular right invariant Borel measure satisfying $\mu G = 1$.*

In the proof of unicity for a compact group $G$, the compactness of $G$ was used to ensure the integrability of $\varphi(yx)$ on $G \times G$. The proof of unicity in the locally compact case is a little more subtle.

Let $\nu$ be a nonzero right invariant Baire measure on $G$ and $\psi$ a fixed nonnegative function in $C_c(G)$. Multiplying $\psi$ by a suitable constant we may assume that

$$\int \psi(y^{-1})\, d\nu(y) = 1.$$

Define a real-valued function $\Delta$ on $G$ by setting

$$\Delta(x) = \int \psi(y^{-1}x)\, d\nu(y). \tag{1}$$

Then $\Delta$ is continuous by Proposition 16. Since $\nu$ is positive on each open set and $\psi(y^{-1}x)$ is positive on some open set of $y$'s, the function $\Delta(x)$ is positive. We note that $\Delta(e) = \int \psi(y^{-1})\, d\nu(y) = 1$. Let $\Gamma(x) = [\Delta(x)]^{-1}$.

Let $\mu$ be any nonzero left invariant Baire measure. Multiplying by a suitable positive constant, we may assume that $\mu$ is normalized so that $\int \psi(x)\, d\mu(x) = 1$. For any $\varphi \in C_c(G)$ we have

$$\begin{aligned}
\int \varphi(x)\, d\mu(x) &= \int \varphi(x)\Gamma(x)\Delta(x)\, d\mu(x) \\
&= \int \left[ \int \varphi(x)\Gamma(x)\psi(y^{-1}x)\, d\nu(y) \right] d\mu(x) \\
&= \int \left[ \int \varphi(x)\Gamma(x)\psi(y^{-1}x)\, d\mu(x) \right] d\nu(y),
\end{aligned}$$

by Fubini, which is applicable since the continuous function $\varphi(x)\Gamma(x)\psi(y^{-1}x)$ has compact support in $G \times G$. From the left invariance of $\mu$ we have

$$\int \varphi(x)\Gamma(x)\psi(y^{-1}x)\, d\mu(x) = \int \varphi(yx)\Gamma(yx)\psi(x)\, d\mu(x).$$

Hence

$$\begin{aligned}\int \varphi(x)\, d\mu(x) &= \int \left[\int \varphi(yx)\Gamma(yx)\psi(x)\, d\mu(x)\right] d\nu(y)\\ &= \int \left[\int \varphi(yx)\Gamma(yx)\, d\nu(y)\right] \psi(x)\, d\mu(x),\end{aligned}$$

by Fubini. The right invariance of $\nu$ implies that

$$\int \varphi(yx)\Gamma(yx)\, d\nu(y) = \int \varphi(y)\Gamma(y)\, d\nu(y).$$

Hence

$$\begin{aligned}\int \varphi(x)\, d\mu(x) &= \int \varphi(y)\Gamma(y)\, d\nu(y) \cdot \int \psi(x)\, d\mu(x)\\ &= \int \varphi(y)\Gamma(y)\, d\nu(y).\end{aligned}$$

This holds for all $\varphi \in C_c(G)$, and so $d\mu = \Gamma\, d\nu$, i.e., $\mu$ is absolutely continuous with respect to $\nu$, and its Radon–Nikodym derivative is $\Gamma$. Since $\Gamma$ was defined independently of $\mu$, any two normalized left invariant Baire measures must be the same. Repeating the argument with left and right interchanged (or applying this result to the group $G^*$), we see that any two right invariant Baire measures are the same, apart from a multiplicative constant. We thus have the following theorem:

**21. Theorem:** *Any two left invariant Baire measures on a locally compact group are the same apart from a multiplicative constant. Similarly, any two right invariant Baire measures are the same up to a multiplicative constant.*

By regularity we have the following corollary:

**22. Corollary:** *Any two inner regular left invariant Borel measures on a locally compact group are the same, apart from a multiplicative constant, and similarly for right invariant measures.*

We also have the existence and unicity of quasi-regular left invariant Borel measures. We note, however, that the inner regular and quasi-regular invariant Borel measures will generally be different when $G$ is not $\sigma$-compact. The following example is typical.

Let $X$ be the set $\mathbf{R}^2$ with the topology in which a set is open if and only if its intersection with each horizontal line is open. Then $X$ is the direct union of the horizontal lines in $\mathbf{R}^2$, each horizontal line having its usual topology. Thus $X$ is locally compact. It becomes a commutative group if we define addition to be the usual vector addition in $\mathbf{R}^2$. A set $E \subset X$ is a Borel set iff its intersection with each horizontal line is a Borel set. It is $\sigma$-bounded iff it meets only a countable number of horizontal lines. Since $X$ is a commutative group, left and right invariance are the same. The inner regular invariant Borel measure $\underline{\mu}$ is defined on a Borel set $E$ by

$$\underline{\mu}E = \sum_a m(E \cap L_a),$$

where $L_a$ is the horizontal line $\{\langle x, y\rangle : y = a\}$, and $m$ is Lebesgue measure on $\mathbf{R}$. The quasi-regular invariant Borel measure $\bar{\mu}$ is defined by

$$\bar{\mu}E = \infty$$

if $E$ meets uncountably many horizontal lines, and by $\bar{\mu}E = \underline{\mu}E$ otherwise. The principal difference between $\bar{\mu}$ and $\underline{\mu}$ is their treatment of sets which meet uncountably many lines but whose intersection with each line is a set of Lebesgue measure zero. If $E$ is such a set, then $\underline{\mu}E = 0$ and $\bar{\mu}E = \infty$.

The function $\Delta$ defined by (1) is called the modular function of $G$. At first sight it seems that $\Delta$ is dependent on the choice of $\psi$ and $\nu$. Since every left invariant Borel measure has the form $d\mu = c\Gamma\, d\nu$, where $\Gamma = \Delta^{-1}$, we see that different choices of $\psi$ can give us $\Delta$'s which differ by at most a constant factor. Since $\Delta(e) = 1$, they must give us the same $\Delta$. If we make use of the unicity of right invariant Baire measures, we see that each right invariant measure $\nu$ must have the form $d\nu = c\Delta\, d\mu$. Since $\Delta(e) = 1$, $\Delta$ must be independent of the choice of $\nu$. If $\psi_1$ is another nonnegative function in $C_c(G)$, then

$$\int \psi_1(y^{-1}x)\, d\nu(y) = \Delta(x) \cdot \int \psi_1(y^{-1})\, d\nu(y).$$

From this it follows that, if $f$ is any $\nu$ integrable function, then

$$\int f(y^{-1}x)\, d\nu(y) = \Delta(x) \int f(y^{-1})\, d\nu(y).$$

We state these and other facts about $\Delta$ in the following proposition:

**23. Proposition:** *There is a positive continuous function $\Delta$ on $G$ with $\Delta(xy) = \Delta(x)\Delta(y)$ such that every left invariant Baire measure $\mu$ and every right invariant measure $\nu$ are related by $d\nu = c\Delta\, d\mu$ for some positive constant $c$. If $f$ is any function integrable with respect to $\nu$, we have*

i. $$\int f(y^{-1}x)\, d\nu(y) = \Delta(x) \int f(y^{-1})\, d\nu(y).$$

ii. $$\int f(xy)\, d\nu(y) = \Delta(x^{-1}) \int f(y)\, d\nu(y).$$

*If $f$ is integrable with respect to $\mu$, we have*

iii. $$\int f(xy)\, d\mu(x) = \Delta(y) \int f(x)\, d\mu(x).$$

iv. $$\int f(yx^{-1})\, d\mu(x) = \Delta(y^{-1}) \int f(x^{-1})\, d\mu(x).$$

We observe that there is a two-sided invariant Baire measure on $G$ if and only if $\Delta(x) = 1$ for all $x \in G$. A group with this property is said to be *unimodular*. Thus all compact groups and all commutative groups are unimodular. The group of all invertible $n \times n$ matrices is also unimodular (see Problem 33). In Problem 32 we give an example of a locally compact group whose left and right invariant measures are different.

The interplay between existence and uniqueness for invariant measures can sometimes be used to derive additional properties of an invariant measure. As an example, we establish the following proposition:

**24. Proposition:** *Lebesgue measure on $\mathbf{R}^n$ is invariant under the group of rigid motions on $\mathbf{R}^n$.*

***Proof:*** Lebesgue measure is translation invariant, that is, invariant on the commutative locally compact group $T$ which $\mathbf{R}^n$ becomes when we take the group operation to be vector addition. By Theorem 21, Lebesgue measure is the only invariant Baire measure on this group; that is, any translation invariant Baire measure on $\mathbf{R}^n$ must be a constant multiple of Lebesgue measure.

The group of rigid motions of $\mathbf{R}^n$ is a transitive group of isometries, and hence equicontinuous. By Theorem 6 there is a Baire measure $\mu$ on $\mathbf{R}^n$ which is invariant under this group of rigid motions. Since $\mu$ is invariant *a fortiori* under the group of translations, it must be a constant multiple of Lebesgue measure. Therefore, Lebesgue measure is invariant under the full group of rigid motions. ▮

We conclude our study of unicity with the following proposition:

**25. Theorem:** *Let $G$ be a locally compact group acting transitively and properly on a locally compact Hausdorff space $X$. Then there is a unique Baire measure on $X$ invariant under the action of $G$.*

***Proof:*** Let $p$ be a point of $X$, which we take as a base point, and let $\lambda$ be a left invariant Baire measure on $G$. Let $\pi: G \to X$ be the map that takes $g$ into $gp$. Then $\pi$ is a proper map.

Since the inverse image of a compact set under a proper map is compact and the inverse image of a $G_\delta$ under any continuous map is a $G_\delta$, we see that $\pi^{-1}[E]$ is a $\sigma$-bounded Baire set whenever $E$ is. Thus we may define a Baire measure $\mu$ on $X$ by setting

$$\mu E = \lambda(\pi^{-1}[E]). \tag{2}$$

Since $\pi^{-1}[gE] = g\pi^{-1}[E]$ and $\lambda$ is left invariant, we see that $\mu$ is invariant under $G$.

We next show that every $G$-invariant Baire measure $\mu$ on $X$ is given by (2) for some left invariant Baire measure $\lambda$ on $G$. Let

$$K = \{k \in G: k(p) = p\} = \pi^{-1}[p].$$

Since $\pi$ is proper, $K$ is a compact subgroup of $G$. Let $\nu$ be the invariant Baire measure on $K$ and $\varphi$ a function in $C_c(G)$. Define

$$\psi(g) = \int_K \varphi(gk)\, d\nu(k).$$

For $h \in K$ we have

$$\begin{aligned}\psi(gh) &= \int_K \varphi(ghk)\, d\nu(k)\\ &= \int_K \varphi(gk)\, d\nu(k)\\ &= \psi(g).\end{aligned}$$

If $gp = x$ and $g_1 p = x$, then $g_1 = gh$ with $h \in K$. Hence $gp = g_1 p$ implies that $\psi(g) = \psi(g_1)$.

If we set $\psi^*(x) = \psi(g)$ whenever $gp = x$, then $\psi^*(x)$ is well defined independently of our choice of $g$ from those with $gp = x$. It is not difficult to show that supp $\psi^*$ is $\pi[\text{supp } \psi]$ which is contained in $\pi[(\text{supp } \psi) \cdot K]$. Since the product of two compact sets is compact and $\pi$ takes compact sets into compacts sets, we must have supp $\psi^*$ compact. Define

$$I(\varphi) = \int \psi^*(x)\, d\mu(x)$$

for each $\varphi \in C_c(G)$. Then $I$ is a positive linear functional on $C_c(G)$, and there is a unique Baire measure $\lambda$ on $G$ such that

$$I(\varphi) = \int_G \varphi\, d\lambda.$$

For $f \in G$ let $\varphi_f(g) = \varphi(fg)$. Then

$$\begin{aligned}\psi_f(g) &= \int_K \varphi_f(gk)\, d\nu(k)\\ &= \int_K \varphi(fgk)\, d\nu(k)\\ &= \psi(fg).\end{aligned}$$

From this we see that

$$\psi_f^*(x) = \psi^*(fx).$$

Hence

$$\begin{aligned} I(\varphi_f) &= \int \psi_f^*(x)\, d\mu(x) \\ &= \int \psi^*(fx)\, d\mu(x) \\ &= \int \psi^*(x)\, d\mu(x) \\ &= I(\varphi). \end{aligned}$$

Consequently,

$$\int \varphi(fg)\, d\lambda(g) = \int \varphi(g)\, d\lambda(g),$$

and so $\lambda$ is a left invariant Baire measure on $G$. If $\varphi^*$ is any function in $C_c(X)$, let $\varphi$ be defined on $G$ by

$$\varphi(g) = \varphi^*(gp).$$

Then

$$I(\varphi) = \int \varphi^*(x)\, d\mu(x).$$

This implies that

$$\mu E = \lambda(\pi^{-1}[E]).$$

Since $\lambda$ is unique apart from a multiplicative constant, the invariant Baire measure $\mu$ must also be unique apart from a multiplicative constant. ▮

## Problems

**26.** Let $G$ be a locally compact group and $\varphi$, $\psi$ functions in $C_c(G)$.

**a.** Let $S$ and $T$ be the maps of $G \times G$ onto itself given by $S(\langle x, y\rangle) = \langle x, yx\rangle$ and $T(\langle x, y\rangle) = \langle x, xy^{-1}\rangle$. Show that $S$ and $T$ are homeomorphisms of $G \times G$ onto itself.

**b.** The functions $\varphi(x)$, $\psi(yx)$, and $\psi(xy^{-1})$ are all continuous functions on $G \times G$.

**c.** The function $\varphi(x)\psi(xy^{-1})$ has compact support on $G \times G$. Describe its support in terms of the supports of $\varphi$ and $\psi$.

**27.** Prove Proposition 23. [Hint: To deduce (ii) from (i), replace $f$ by the function $f^*$ defined by $f^*(z) = f(z^{-1})$. To show (iii), write

$$\int f(xy)\, d\mu(x) = \Delta(y) \int f(xy)\Delta[(xy)^{-1}]\Delta(x)\, d\mu(x),$$

using the multiplicative property of $\Delta$. Use the right invariance of the measure $\Delta(x)\, d\mu(x)$.]

**28. a.** Show that the function $\psi^*$ used in the proof of Theorem 25 is a continuous function on $X$. [Hint: First show that the function $\psi$ is a continuous function on $G$.]

**b.** Show that $I(\varphi) = \int \varphi^*(x)\, d\mu(x)$ implies that $\mu E = \lambda(\pi^{-1}[E])$.

**29.** Show that a group is unimodular iff every function which is integrable with respect to its left invariant measure is integrable with respect to its right invariant measure. [Hint: This condition implies that $\Delta$ is bounded.]

**30.** Let $G$ be a locally compact metric group, that is, one whose underlying topological space is a locally compact metric space. A metric $\rho$ for $G$ is said to be a two-sided invariant metric if

$$\rho(gx, gy) = \rho(x, y) = \rho(xg, yg)$$

for all $x, y, g \in G$. Show that if a locally compact metric group $G$ has a two-sided invariant metric, then $G$ is unimodular. [Hint: The group of isometries on $G$ is equicontinuous.]

**31.** Let $G$ be the group $SO(3)$ consisting of all proper rotations in $\mathbf{R}^3$, or equivalently, of all $3 \times 3$ orthogonal matrices with determinant one.

**a.** Let $R_\varphi$ denote a rotation through the angle $\varphi$ about the positive $z$-axis, and $S_\theta$ a notation through the angle $\theta$ about the positive $y$-axis. Show that every $O \in G$ can be expressed as a product

$$O = R_\varphi S_\theta R_\psi,$$

where $0 \leq \varphi \leq 2\pi$, $0 \leq \theta \leq \pi$, and $0 \leq \psi < 2\pi$.

**b.** Show that this representation is unique unless $\theta = 0$ or $\pi$, that is, unless $O$ is a rotation about the $z$-axis or a rotation about the $z$ axis followed by a reflection on the $y$-axis.

**c.** Let $H$ be the subgroup of rotations about the $z$-axis. Show that $X = G/H$ is naturally homeomorphic to the unit sphere in $\mathbf{R}^3$ with the group element $R_\varphi S_\theta R_\psi$ going to the point with spherical coordinates $(1, \theta, \varphi)$.

**d.** Describe the action of $G$ on $X$ geometrically and deduce that the

$G$-invariant measure on $X$ is the spherical area given by

$$\mu E = \int_E \sin \theta \, d\theta \, d\varphi.$$

**e.** Use Theorem 25 to show that the invariant measure $\lambda$ on $G$ is given by

$$\lambda E = \int_E \sin \theta \, d\theta \, d\varphi \, d\psi.$$

## 7 Groups of Diffeomorphisms

In this section we consider the existence and properties of invariant measures for the case when $X$ is a differentiable manifold and $G$ is a transitive group of diffeomorphisms. If $G$ itself can be given the structure of a differentiable manifold so that the maps $\langle x, y\rangle \to xy$ and $x \to x^{-1}$ are differentiable maps on $G \times G$ and $G$, we say that $G$ is a *Lie group*. Examples of Lie groups are given by those subgroups of the invertible $n \times n$ real or complex matrices which satisfy certain algebraic conditions: e.g. the group of all invertible matrices, the group of matrices of determinant one, the group of orthogonal matrices, etc.

Historically, invariant integration was first introduced for Lie groups by A. Hurwitz [19] in 1897. The existence of invariant measures for arbitrary locally compact separable metric groups was first shown by A. Haar [18] in 1933, and extended to general locally compact groups by A. Weil [27] in 1940. The treatment of invariant measures given in Section 3 was inspired by the elegant treatment of the separable metric case by S. Banach ([13], Appendix II).

Although the approach of Hurwitz only works in the case of groups of diffeomorphisms, it is the simplest approach and has the advantage of leading to specific formulae for the invariant measure.

Let $X$ be a differentiable manifold and $\mu$ a Baire measure on $X$. We say that $\mu$ is a *differential measure* or a smooth measure if for each coordinate chart $U$ with differentiable coordinates $x_1, \ldots, x_n$ there is a positive continuous function $\Phi_U(x)$ such that for each measurable $E \subset U$ we have

$$\mu E = \int_E \Phi_U(x) \, dx_1 \, dx_2 \cdots dx_n, \tag{3}$$

where we have written $dx_1 \cdots dx_n$ for $n$-dimensional Lebesgue measure. If $U$ overlaps another coordinate chart $V$ with coordinates $y_1, \ldots, y_n$, then we must have

$$\Phi_V(y)\left|\frac{\partial y}{\partial x}\right| = \Phi_U(x)$$

in $U \cap V$, where we have written

$$\frac{\partial y}{\partial x} = \det\left[\frac{\partial y_j}{\partial x_k}\right].$$

Such an assignment of functions $\Phi_U$ to coordinate charts is often called a volume element.

Suppose now that $G$ is a transitive group of diffeomorphisms on $X$. Let $U$ and $V$ be coordinate neighborhoods with $g[U] \subset V$ for some $g \varepsilon G$. In order for our volume form to give a measure invariant on sets $E \subset U$, we must have

$$\Phi_V(gx)\left|\frac{\partial g}{\partial x}\right| = \Phi_U(x). \tag{4}$$

If we fix a base point $p$ in $U$ and multiply by a suitable constant so that $\Phi(p) = 1$, then any invariant volume form must be given at $q = gp$ by

$$\Phi_V(q) = \Phi_V(gp) = \left|\frac{\partial g}{\partial x}\right|^{-1} \tag{5}$$

in order for (4) to hold. This shows us that an invariant volume form, if it exists, is unique. The formula (5) can be used to define the value of an invariant volume form at $q$, provided we get the same value for $\Phi_V(q)$ no matter which element $g \varepsilon G$ satisfying $gp = q$ we use. Thus we get a well defined value for $\Phi_V(q)$ if and only if

$$\left|\frac{\partial g}{\partial x}\right| = \left|\frac{\partial h}{\partial x}\right| \tag{6}$$

for any two elements $g, h \varepsilon G$ with $q = gp$ and $q = hp$. But $g$ and $h$ map $p$ to the same point if and only if $h^{-1}gp = p$, that is, if $h^{-1}g = k$, an element in the isotropy subgroup at $p$. By the chain rule

$$\left|\frac{\partial k}{\partial x}\right| = \left|\frac{\partial h}{\partial x}\right|^{-1}\left|\frac{\partial g}{\partial x}\right|.$$

Hence (6) is satisfied for all $g$, $h$ with $gp = hp$ if and only if $|\partial k/\partial x| = 1$ for all $k \varepsilon G$ with $kp = p$. This gives us the following theorem:

**26. Theorem:** *Let $G$ be a transitive group of diffeomorphisms on a differentiable manifold $X$ and $p \varepsilon X$. Then there is a differentiable measure invariant under $G$ if and only if*

$$\left|\frac{\partial k}{\partial x}\right| = 1 \qquad \textit{at } p$$

*for each element $k \varepsilon G$ for which $kp = p$. In this case the measure is unique, apart from a multiplicative constant, and is given by* (5).

To see how this works in practice, let us take $X$ to be the right half-plane $\{\langle x_1, x_2\rangle \varepsilon \mathbf{R}^2 : x_1 > 0\}$ and $G$ the group of matrices of the form

$$g = \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix}$$

with $u > 0$ and $gx$ given by

$$\begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Then $G$ is simply transitive; that is, given $x$ and $y$ there is a unique $g$ such that $gx = y$. Let $p = \langle 1, 0\rangle$ and take our volume form there to be $dx_1\, dx_2$. If

$$gx = \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

then

$$\frac{\partial g}{\partial x} = \det \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix} = u.$$

When $gp = y$, we have $\langle u, v\rangle = \langle y_1, y_2\rangle$. Thus

$$\Phi(y) = \left|\frac{\partial g}{\partial x}\right|^{-1} = y_1^{-1},$$

and the invariant measure on $X$ is given by

$$\mu E = \int_E y_1^{-1}\, dy_1\, dy_2 .$$

## Problems

**32.** Let $G$ be the group of $2 \times 2$ matrices of the form

$$\begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix}$$

with $u > 0$ and with the topology inherited by considering $G$ as the right-half $u,v$-plane.

**a.** Let $G$ act differentially on $\mathbf{R}^1$ by setting $gx = ux + v$. What is the isotropy group of $G$ at $x = 1$? Use Theorem 26 to deduce that there is no differential measure invariant on $\mathbf{R}^1$ under this action of $G$.

**b.** Show that the left invariant measure on $G$ is given by

$$\mu E = \int_E u^{-1}\, du\, dv.$$

**c.** Show that the right invariant measure on $G$ is given by

$$\mu E = \int_E u^{-2}\, du\, dv.$$

**d.** What is the modular function $\Delta$ on $G$.

**33.** Let $G$ be the group of all $n \times n$ invertible real matrices $[x_{ij}]$. Show that $G$ is unimodular and that its invariant measure is given by

$$\mu E = \int_E |\det [x]|^{-n}\, dx_{11}\, dx_{12} \cdots dx_{nn}.$$

**34.** Let $G$ be a group of diffeomorphisms of $X$ into itself and $H$ the isotropy subgroup of $G$ at $p \varepsilon X$.

**a.** Show that the map $h \to |\partial h/\partial x|_p$ is a homorphism of $H$ into the positive reals.

**b.** Show that, if this homorphism is bounded, then $|\partial h/\partial x| = 1$ for all $h \varepsilon H$.

**c.** If the isotropy subgroup $H$ at $p$ is compact, then there is a differential metric invariant under $G$.

**35.** Give an example of a transitive group of diffeomorphisms on $\mathbf{R}^2$ which is not topologically equicontinuous but which nevertheless admits a nonzero invariant differential metric.

# 15 Mappings of Measure Spaces

## 1 Point Mappings and Set Mappings

Let $X$ and $Y$ be any two spaces and $\varphi$ a mapping of $X$ into $Y$. Associated with $\varphi$ are several mappings of objects associated with $Y$ into corresponding objects associated with $X$. For example, the set mapping $\Phi$ defined by $\Phi(E) = \varphi^{-1}[E]$ is a mapping of the subsets of $Y$ into the subsets of $X$. This mapping preserves unions, intersections, and complements. It is called the set mapping induced by or adjoint to $\varphi$. We refer to $\varphi$ as a point mapping. If $\langle X, \mathfrak{A} \rangle$ and $\langle Y, \mathfrak{B} \rangle$ are measurable spaces, the point mapping $\varphi$ of $X$ into $Y$ is called **measurable** if $\varphi^{-1}[E] \in \mathfrak{A}$ for each $E \in \mathfrak{B}$. Thus $\varphi$ is measurable iff $\Phi$ maps $\mathfrak{B}$ into $\mathfrak{A}$.

If $f$ is a real-valued function on $\langle X, \mathfrak{A} \rangle$, then $f$ is a mapping of $X$ into $\mathbf{R}$, and $f$ is measurable with respect to $\mathfrak{A}$ if and only if $f$ is a measurable mapping of $\langle X, \mathfrak{A} \rangle$ into $\langle \mathbf{R}, \mathfrak{B} \rangle$ where $\mathfrak{B}$ is the $\sigma$-algebra of Borel sets. A real-valued function of a real variable is Lebesgue measurable iff it is a measurable mapping of $\langle \mathbf{R}, \mathfrak{M} \rangle$ into $\langle \mathbf{R}, \mathfrak{B} \rangle$, where $\mathfrak{M}$ is the class of Lebesgue measurable sets. It is Borel measurable iff it is a measurable mapping of $\langle \mathbf{R}, \mathfrak{B} \rangle$ into $\langle \mathbf{R}, \mathfrak{B} \rangle$.

Also associated with $\varphi$ is the mapping $\varphi^*$ of the space of real-valued functions on $Y$ into the space of real-valued functions on $X$ defined by $\varphi^*(f) = f \circ \varphi$. The mapping $\varphi^*$ is often called the adjoint of $\varphi$, and it preserves sums, products, maxima, etc. If $\varphi$ is measurable, then $\varphi^*$ takes measurable functions into measurable functions.

Let $\mathcal{A}$ be an algebra of subsets of $X$ and $\mathcal{B}$ an algebra of subsets of $Y$. Then a map $\Phi$ of $\mathcal{B}$ into $\mathcal{A}$ such that $\Phi(Y) = X$, $\Phi(\tilde{E}) = \sim\Phi(E)$, and $\Phi(A \cup B) = \Phi(A) \cup \Phi(B)$ is called a (lattice) **homomorphism.** If $\mathcal{A}$ and $\mathcal{B}$ are $\sigma$-algebras and $\Phi$ has the property that

$$\Phi\left(\bigcup_{i=1}^{\infty} E_i\right) = \bigcup_{i=1}^{\infty} \Phi(E_i),$$

then $\Phi$ is called a **σ-homomorphism.** Every set mapping induced by a point mapping of $X$ into $Y$ is a $\sigma$-homomorphism, but we can have $\sigma$-homomorphisms which are not induced by any point mapping (Problem 2).

Let $\langle X, \mathcal{A}\rangle$ and $\langle Y, \mathcal{B}\rangle$ be measurable spaces and $\Phi$ a $\sigma$-homomorphism of $\mathcal{B}$ into $\mathcal{A}$. Then $\Phi$ induces a mapping $\Phi^*$ of measures on $\langle X, \mathcal{A}\rangle$ into measures on $\langle Y, \mathcal{B}\rangle$ if we define $\Phi^*\mu$ by $(\Phi^*\mu)(E) = \mu(\Phi(E))$. The following proposition may be thought of as a change-of-variable formula.

1. **Proposition:** *Let $\varphi$ be a measurable point mapping of the measure space $\langle X, \mathcal{A}, \mu\rangle$ into the measurable space $\langle Y, \mathcal{B}\rangle$. Let $\Phi$ be the induced set mapping of $\mathcal{B}$ into $\mathcal{A}$. Then for each nonnegative measurable function $f$ on $Y$ we have*

$$\int_Y f\, d\Phi^*\mu = \int_X (f \circ \varphi)\, d\mu.$$

***Proof:*** The proposition is clearly true if $f$ is a characteristic function. From this it follows for $f$ a simple function. Since $\int f$ is the supremum of the integrals of all nonnegative simple functions less than $f$, the proposition follows. ▌

## Problems

1. Let $\mathcal{B}$ be a family of subsets of $Y$ which contains $Y$, $\emptyset$, and each set $\{y\}$ consisting of a single element. Then each mapping $\Phi$ of $\mathcal{B}$ into the subsets of $X$ which preserves finite intersections and arbitrary unions and for which $\Phi(Y) = X$ and $\Phi(\emptyset) = \emptyset$ is induced by a point mapping. [The sets $E_y = \Phi(\{y\})$ are disjoint and their union is $X$. Let $\varphi(x) = y$ for $x$ in $E_y$.]

2. Let $X = Y = [0, 1]$, and let $\mathcal{A}$ be the collection of all subsets of $[0, 1]$ which are either countable or the complements of countable sets. Then $\mathcal{A}$ is a $\sigma$-algebra. Let $\mathcal{B} = \{Y, \emptyset\}$. For $E \in \mathcal{A}$, define $\Phi(E) = \emptyset$ if $E$ is countable,

$\Phi(E) = Y$ if $E$ is the complement of a countable set. Then $\Phi$ is a $\sigma$-homomorphism and it is not induced by any point mapping of $Y$ into $X$.

3. Show that the adjoint mapping $\varphi^*$ can be extended to map the extended real-valued functions on $Y$ into the space of such functions on $X$. Generalize Proposition 1 to this case.

4. Let $\langle X, \mathfrak{A}\rangle$, $\langle Y, \mathfrak{B}\rangle$, and $\langle Z, \mathfrak{C}\rangle$ be measure spaces and $\varphi: X \to Y$ and $\psi: Y \to Z$ measurable mappings. Show that $\psi \circ \varphi$ is a measurable mapping of $\langle X, \mathfrak{A}\rangle$ into $(Z, \mathfrak{C}\rangle$. What has this to do with the fact that, if $f$ is a Lebesgue measurable and $g$ a Borel measurable real-valued function, then $g \circ f$ is Lebesgue measurable, but $f \circ g$ need not be?

5. Prove that $\Phi^*\mu$ is a measure.

6. **a.** Let $\varphi$ be a measurable point mapping of $\langle X, \mathfrak{A}, \mu\rangle$ into $\langle Y, \mathfrak{B}, \nu\rangle$ and $\Phi$ the induced set mapping of $\mathfrak{B}$ into $\mathfrak{A}$. Suppose that $\Phi^*\mu$ is absolutely continuous with respect to $\nu$ and that $\nu$ is a finite (or $\sigma$-finite) measure. Define $[d\mu/d\nu]$ to be the Radon–Nikodym derivative of $\Phi^*\mu$ with respect to $\nu$. Then for each nonnegative measurable function $f$ on $Y$ we have

$$\int_X (f \circ \varphi)\, d\mu = \int_Y f\left[\frac{d\mu}{d\nu}\right] d\nu.$$

**b.** Let $f$ be a nonnegative measurable function on $[0, 1]$ and $g$ a monotone absolutely continuous function on $[0, 1]$ with $g(0) = 0$, $g(1) = 1$. Then

$$\int_0^1 f[g(t)]g'(t)\, dt = \int_0^1 f(t)\, dt.$$

7. **a.** Let $\langle X, \mathfrak{A}\rangle$ and $\langle Y, \mathfrak{B}\rangle$ be measurable spaces and $\Phi$ a $\sigma$-homomorphism from $\mathfrak{B}$ to $\mathfrak{A}$. Show that there is a unique linear mapping $T_\Phi$ of the measurable real-valued functions on $Y$ into the measurable real-valued functions on $X$, which takes nonnegative functions into nonnegative functions, such that for characteristic functions $\chi_E$ we have

$$T_\Phi(\chi_E) = \chi_{\Phi(E)}.$$

**b.** Let $\mu$ be a measure on $\langle X, \mathfrak{A}\rangle$ and $f$ a nonnegative measurable function on $Y$. Then

$$\int_X T_\Phi(f)\, d\mu = \int_Y f\, d\Phi^*(\mu).$$

## 2 Boolean $\sigma$-Algebras

A *Boolean algebra* is a set of elements on which there are defined two binary operations $\vee$ and $\wedge$ and a unary operation $'$ which satisfy the following rules:

i. $A \vee A = A$.
i′. $A \wedge A = A$.
ii. $A \vee B = B \vee A$.
ii′. $A \wedge B = B \wedge A$.
iii. $(A \vee B) \vee C = A \vee (B \vee C)$.
iii′. $(A \wedge B) \wedge C = A \wedge (B \wedge C)$.
iv. $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$.
iv′. $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$.
v. $(A \wedge B)' = A' \vee B'$.
vi. $(A')' = A$.
vii. $\exists 0$ such that $A \wedge 0 = 0$ and $A \vee 0 = A$.
viii. $A' \wedge A = 0$.

One example of a Boolean algebra is an algebra $\mathfrak{A}$ of subsets of some set $X$ with $\vee$, $\wedge$, $'$ interpreted to mean $\cup$, $\cap$, $\sim$. We shall sometimes call $\vee$, $\wedge$, $'$, 'union', 'intersection', and 'complementation'.

A Boolean algebra $\mathfrak{A}$ becomes partially ordered if we define $A \leq B$ to mean $A \wedge B = A$. Then 0 is the smallest element, while $X = 0'$ is the largest element. Moreover, $A \vee B$ is the smallest element of $\mathfrak{A}$ which is larger than both $A$ and $B$.

A Boolean algebra $\mathfrak{A}$ is called a *Boolean $\sigma$-algebra* if for each sequence $\langle A_n \rangle$ of elements of $\mathfrak{A}$ there is a smallest element $B$ such that $A_n \leq B$ for all $n$. This element $B$ is denoted by $\bigvee_{n=1}^{\infty} A_n$. In a Boolean $\sigma$-algebra the element $C = \left(\bigvee_{n=1}^{\infty} A_n'\right)'$ is the largest element such that $C \leq A_n$ for all $n$. We write $C = \bigwedge_{n=1}^{\infty} A_n$. One example of a Boolean $\sigma$-algebra is given by a $\sigma$-algebra of subsets of a set $X$. Another is the following: Let $\langle X, \mathfrak{A}, \mu \rangle$ be a measure space and $\mathfrak{N}$ the family of sets of measure zero. Two sets of $\mathfrak{A}$ are said to be equivalent modulo $\mathfrak{N}$ if their symmetric difference is in $\mathfrak{N}$. Finite or countable unions and intersections of equivalent sets are equivalent, and so the classes of equivalent sets form a Boolean $\sigma$-algebra. We denote it by $\mathfrak{A}/\mathfrak{N}$.

Every Boolean algebra is isomorphic to an algebra of subsets of a suitable space $X$, but not every Boolean $\sigma$-algebra is isomorphic to a $\sigma$-algebra of subsets of some space. The situation we encounter most often is that of the last example, where the Boolean $\sigma$-algebra is $\mathfrak{A}/\mathfrak{N}$. In fact, the only properties we need of $\mathfrak{N}$ are (i) if $A \in \mathfrak{N}$ and

$B \in \mathfrak{A}$ with $B \subseteq A$, then $B \in \mathfrak{N}$, and (ii) if $A_n \in \mathfrak{N}$, then $\bigvee_{n=1}^{\infty} A_n \in \mathfrak{N}$. A subset of a Boolean $\sigma$-algebra $\mathfrak{A}$ with these properties is called a **σ-ideal**, and we may define the Boolean $\sigma$-algebra $\mathfrak{A}/\mathfrak{N}$ of equivalence classes of $\mathfrak{A}$ mod $\mathfrak{N}$.

A triple $(X, \mathfrak{A}, \mathfrak{N})$ consisting of a set $X$, a $\sigma$-algebra $\mathfrak{A}$ of subsets of $X$, and a $\sigma$-ideal $\mathfrak{N}$ of $\mathfrak{A}$ will be called a *measurable space* with *null sets*. Then $\mathfrak{A}/\mathfrak{N}$ gives us a typical Boolean $\sigma$-algebra. It is particularly important for the Boolean $\sigma$-algebras of the form $\mathfrak{A}/\mathfrak{N}$, to distinguish between points (elements of $X$), sets (elements of $\mathfrak{A}$), and elements of $\mathfrak{A}/\mathfrak{N}$ (sets modulo null sets).

A mapping $\Phi\colon \mathfrak{A} \to \mathfrak{B}$ between two Boolean $\sigma$-algebras is said to be a $\sigma$-homomorphism if $\Phi(A') = \Phi(A)'$, $\Phi(A_1 \vee A_2) = \Phi(A_1) \vee \Phi(A_2)$, and $\Phi(\bigvee A_i) = \bigvee \Phi(A_i)$. Thus there is a natural $\sigma$-homomorphism from $\mathfrak{A}$ to $\mathfrak{A}/\mathfrak{N}$ defined by taking $\Phi(A)$ to be the equivalence class in $\mathfrak{A}/\mathfrak{N}$ containing $\mathfrak{A}$.

Let $(X, \mathfrak{A}, \mathfrak{N})$ be a measurable space with null sets and $f$, $g$ two measurable functions on $X$. We say that $f = g$ almost everywhere $[\mathfrak{N}]$ if there is a set $N \in \mathfrak{N}$ such that $f(x) = g(x)$ whenever $x \notin N$. In many contexts we do not distinguish between functions that are equal a.e. This leads us to consider the space $M(X)$ of real-valued measurable functions on $X$ modulo equality a.e. We often follow our earlier informal practice of speaking of a function $f$ in $M(X)$ when we mean the equivalence class $[f]$ to which $f$ belongs. We denote by $M^*(X)$ the space of extended real-valued functions modulo equality a.e.

If $f$ and $g$ are two measurable functions that are equal a.e., then the sets $\{x: f(x) < \alpha\}$ and $\{x: g(x) < \alpha\}$ differ by a set $N \in \mathfrak{N}$. Thus each element $[f]$ of $M^*(X)$ determines a family $\{B_\alpha: \alpha \in \mathbf{R}\}$ of elements of the Boolean $\sigma$-algebra $\mathfrak{A}/\mathfrak{N}$ by letting $B_\alpha$ be the equivalence class containing $\{x: f(x) < \alpha\}$ for some $f$ in $[f]$. We call the family $\{B_\alpha\}$ of elements of $\mathfrak{A}/\mathfrak{N}$ the **soma**[1] of the function $f$. The soma of a function satisfies

i. $B_\alpha < B_\beta$ for $\alpha < \beta$.
ii. $B_\beta = \bigvee B_{\alpha_i}$ whenever $\alpha_i < \beta$ and $\beta = \lim \alpha_i$.

Any collection $\{B_\alpha\}$ of elements of a Boolean $\sigma$-algebra that is ordered by $\mathbf{R}$ and satisfies (i) and (ii) will be called an *abstract soma.*

[1] From the Greek (το σῶμα) for "body". Our usage differs from that of Carathéodory [17], who calls $\{B_\alpha\}$ a "soma scale" and each $B_\alpha$ a "soma".

The following Proposition states that each soma in a Boolean $\sigma$-algebra $\mathfrak{A}/\mathfrak{N}$ is the soma of some element of $M^*(X)$.

**2. Proposition:** *Let $(X, \mathfrak{A}, \mathfrak{N})$ be a measure algebra with null sets and $\{B_\alpha\}$ a soma from $\mathfrak{A}/\mathfrak{N}$. Then there is a unique element $[f]$ in $M^*(X)$ such that $\{B_\alpha\}$ is the soma of $[f]$. The element $[f]$ will be real-valued, i.e., belong to $M(X)$, iff*

iii. $\bigvee_{n=1}^{\infty} B_n = X$ *and* $\bigwedge_{n=-\infty}^{\infty} B_n = \emptyset$.

***Proof:*** The existence of an $f$ whose soma is $\{B_\alpha\}$ follows from Proposition 11.10. To establish uniqueness, let $g$ be any measurable function whose soma is $\{B_\alpha\}$. Then for each pair $\langle \alpha, \beta \rangle$ of rational numbers with $\alpha < \beta$ the set

$$\{x: g(x) < \alpha < \beta \leq f(x)\} = \{x: g(x) < \alpha\} \sim \{x: f(x) < \beta\},$$

and so must be in $\mathfrak{N}$. Hence the set

$$\{x: g(x) < f(x)\} = \bigcup_{\langle \alpha, \beta \rangle} \{x: g(x) < \alpha < \beta \leq f(x)\}$$

must also be in $\mathfrak{N}$. Similarly, the set $\{x: g(x) > f(x)\}$ must be in $\mathfrak{N}$, and $f = g$ a.e. ▮

A useful corollary is the following proposition:

**3. Proposition (Sikorski):** *Let $\langle X, \mathfrak{A}, \mathfrak{N} \rangle$ be a measurable space with null sets, and let $\Phi$ be a $\sigma$-homomorphism of the family $\mathfrak{B}$ of Borel sets in $[0, 1]$ into the $\sigma$-algebra $\mathfrak{A}/\mathfrak{N}$ with $\Phi([0, 1]) = X$. Then there is a measurable mapping of $X$ into $[0, 1]$ such that for each $B \in \mathfrak{B}$ we have $\varphi^{-1}[B]$ in the equivalence class $\Phi(B)$. If $\psi$ is any other point mapping with this property, then $\psi = \varphi$ except on a subset in $\mathfrak{N}$.*

***Proof:*** Let $A_\alpha = \Phi([0, \alpha))$. Then $\{A_\alpha\}$ is a soma in $\mathfrak{A}/\mathfrak{N}$, and so there is a function $\varphi \in M(X)$ whose soma is $\{A_\alpha\}$. Since $A_0 = \emptyset$ and $A_\alpha = X$ for $\alpha > 1$, we have $0 \leq \varphi \leq 1$, and $\varphi$ maps $X$ into $[0, 1]$. Let $\Psi: \mathfrak{B} \to \mathfrak{A}/\mathfrak{N}$ be the $\sigma$-homomorphism defined by $\Psi(B) = \varphi^{-1}[B]$. Then

$$\Psi([0, \alpha)) = A_\alpha = \Phi([0, \alpha)).$$

Thus $\Psi$ and $\Phi$ agree on the half-open intervals. Since they are $\sigma$-homomorphisms, they agree on a $\sigma$-algebra. Hence $\Phi = \Psi$ for all Borel sets. ▮

## Problems

**8.** Prove that in a Boolean $\sigma$-algebra we have

$$B \wedge \left( \bigvee_{n=1}^{\infty} A_n \right) = \bigvee_{n=1}^{\infty} (B \wedge A_n).$$

**9.** Let $\mathfrak{A}$ be a Boolean $\sigma$-algebra and $\mathfrak{N}$ a Boolean $\sigma$-ideal. Show that if $A \,\Delta\, B \in \mathfrak{N}$, then $A' \,\Delta\, B' \in \mathfrak{N}$, and that if $(A_n \,\Delta\, B_n) \in \mathfrak{N}$, then we have $(\bigvee A_n) \,\Delta\, (\bigvee B_n) \in \mathfrak{N}$.

**10.** Let $\mathfrak{A}$ and $\mathfrak{B}$ be Boolean $\sigma$-algebras and $\Phi\colon \mathfrak{A} \to \mathfrak{B}$ a $\sigma$-homomorphism. Let $\mathfrak{N} = \{N \in \mathfrak{A}\colon \Phi(N) = 0\}$. Then $\mathfrak{N}$ is a $\sigma$-ideal of $\mathfrak{A}$.

## 3 Measure Algebras

By a **measure algebra**, we mean a Boolean $\sigma$-algebra $\mathfrak{A}$ together with a nonnegative real-valued function $\mu$ defined on $\mathfrak{A}$ such that $\mu(A) = 0$ if and only if $A = 0$ and

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

if $A_i \wedge A_j = 0$ for $i \neq j$. We call $\mu$ a measure on $\mathfrak{A}$. If $\langle X, \mathfrak{B}, \mu \rangle$ is a finite measure space and $\mathfrak{N}$ is the collection of sets of measure zero, then we obtain a measure algebra if we consider $\mu$ on the Boolean $\sigma$-algebra $\mathfrak{A} = \mathfrak{B}/\mathfrak{N}$, that is, if we fail to distinguish between sets of $\mathfrak{B}$ which differ by a set of measure zero.

In a Boolean algebra we define the symmetric difference $A \,\Delta\, B$ of two elements by $A \,\Delta\, B = (A \wedge B') \vee (A' \wedge B)$. A measure algebra becomes a metric space if we define $\rho(A, B) = \mu(A \,\Delta\, B)$. This metric space is always complete, and the mappings $A \to A'$, $\langle A, B \rangle \to A \vee B$, and $\langle A, B \rangle \to A \wedge B$ are continuous. A measure algebra is called **separable** if it is separable as a metric space.

A mapping $\Phi$ of a measure algebra $\langle \mathfrak{A}, \mu \rangle$ into a measure algebra $\langle \mathfrak{B}, \nu \rangle$ is called an **isomorphism into** if $\Phi(A') = [\Phi(A)]'$, $\Phi(A_1 \vee A_2) = \Phi(A_1) \vee \Phi(A_2)$ and $\mu(A) = \nu(\Phi(A))$. The mapping $\Phi$ is called an isomorphism if it is onto. If the measure algebras are considered as metric spaces, an isomorphism is an isometry which preserves complements and finite unions. It follows that an isomorphism also preserves countable unions and intersections (Problem 10).

An element $A \neq 0$ in a measure algebra $\mathfrak{A}$ is called an **atom** if $B \leq A$ can occur only for $B = A$ and $B = 0$. Let $\mathfrak{M}$ be the class of measurable subsets of [0, 1], $\mathfrak{N}$ the class of subsets of measure zero, and $m$ Lebesgue measure. Then $\langle \mathfrak{M}/\mathfrak{N}, m\rangle$ is a separable measure algebra without atoms. The following theorem asserts that, apart from isomorphism, it is the only such measure algebra.

**4. Theorem (Carathéodory):** *Let $\langle \mathfrak{A}, \mu\rangle$ be a separable measure algebra with $\mu(X) = 1$. Then there is an isomorphism $\Phi$ of $\langle \mathfrak{A}, \mu\rangle$ into the measure algebra $\langle \mathfrak{M}/\mathfrak{N}, m\rangle$ induced by Lebesgue measure $m$ on* [0, 1]. *The isomorphism $\Phi$ is onto iff $\mathfrak{A}$ has no atoms.*

***Proof:*** Since $\langle \mathfrak{A}, \mu\rangle$ is separable, there is a sequence $\langle A_n\rangle$ of elements which are dense in $\mathfrak{A}$. Let $\mathfrak{A}_n$ be the Boolean algebra obtained by taking all unions of intersections of the sets $A_1, \ldots, A_n$ and their complements, and let $\mathfrak{A}_\infty = \bigcup_{n=1}^{\infty} \mathfrak{A}_n$. Then $\mathfrak{A}_\infty$ is again a Boolean algebra. For, given any $A$ and $B$ in $\mathfrak{A}_\infty$, they belong to $\mathfrak{A}_n$ and $\mathfrak{A}_m$, respectively. But if $m \le n$, we have $\mathfrak{A}_m \subset \mathfrak{A}_n$, and so $A'$, $A \vee B$, and $A \wedge B$ are all contained in $\mathfrak{A}_n \subset \mathfrak{A}_\infty$.

We shall define by induction a mapping $\Phi$ of $\mathfrak{A}_\infty$ into the algebra $\mathfrak{I}$ of all finite unions of half open subintervals of [0, 1). The algebra $\mathfrak{A}_1$ consists of the four sets $0, A_1, A_1', X$, and $\mu A_1 + \mu A_1' = \mu X = 1$. Let $\Phi(A_1) = [0, \mu A_1)$, $\Phi(A_1') = [\mu A_1, 1)$, $\Phi(0) = \emptyset$, $\Phi(X) = [0, 1)$. Then $\Phi$ preserves unions, intersections, complements, and measure. Suppose now that $\Phi$ has been defined on $\mathfrak{A}_{n-1}$ so that it maps $\mathfrak{A}_{n-1}$ onto the algebra generated by the half open intervals $[0, x_1)$, $[x_1, x_2)$, $\ldots, [x_k, 1)$ and that $\Phi$ is measure preserving and preserves unions and complements. We wish to extend the mapping $\Phi$ to $\mathfrak{A}_n$. Let $B_0, \ldots, B_k$ be the sets in $\mathfrak{A}_{n-1}$ which are mapped onto the intervals $[0, x_1), \ldots, [x_k, 1)$. Then $\mathfrak{A}_{n-1}$ consists of all finite unions of the sets $B_0, \ldots, B_k$, and $\mathfrak{A}_n$ consists of all finite unions of $A_n \wedge B_0, \ldots, A_n \wedge B_k, A_n' \wedge B_0, \ldots, A_n' \wedge B_k$. For those intersections which are 0, set $\Phi$ equal to $\emptyset$. For those which are not 0, let $\Phi(A_n \wedge B_j) = [x_j, x_j + \mu(A_n \wedge B_j))$ and $\Phi(A_n' \wedge B_j) = [x_j + \mu(A_n \wedge B_j), x_{j+1})$. Since $\mu(A_n \wedge B_j) + \mu(A_n' \wedge B_j) = \mu(B_j) = x_{j+1} - x_j$, we see that these are properly defined intervals, $\Phi$ is measure preserving, and $\Phi(A_n \wedge B_j) \cup \Phi(A_n' \wedge B_j) = [x_j, x_{j+1}) = \Phi(B_j)$. From this it follows that we can extend $\Phi$ to all of $\mathfrak{A}_n$ so that it preserves unions, complements, and measures.

Thus we have defined by induction the mapping $\Phi$ from $\mathfrak{A}_\infty$ to $\mathfrak{M}/\mathfrak{N}$ so that it is measure preserving. Hence it is an isometry. Since $\mathfrak{A}_\infty$ is dense in $\mathfrak{A}$ and the metric space $\mathfrak{M}/\mathfrak{N}$ is complete, we can extend $\Phi$ to be an isometry from $\mathfrak{A}$ into $\mathfrak{M}/\mathfrak{N}$.

To see that $\Phi$ preserves complements, let $E$ be any element in $\mathfrak{A}$ and choose $A \in \mathfrak{A}_\infty$ so that $\mu(E \,\Delta\, A) < \epsilon$. Then $A' \in \mathfrak{A}_\infty$, and $\mu(E' \,\Delta\, A') = \mu(E \,\Delta\, A) < \epsilon$. Since $\Phi$ is an isometry and $\Phi(A') = {\sim}\Phi(A)$, we have $m(\Phi(E') \,\Delta\, \widetilde{\Phi(A)}) < \epsilon$ and $m(\widetilde{\Phi(E)} \,\Delta\, \widetilde{\Phi(A)}) = m(\Phi(E) \,\Delta\, \Phi(A)) < \epsilon$. Hence $m(\Phi(E') \,\Delta\, \widetilde{\Phi(E)}) < 2\epsilon$ for all $\epsilon > 0$. Thus $\Phi(E') = {\sim}\Phi(E)$ in the algebra $\mathfrak{M}/\mathfrak{N}$. A similar argument shows $\Phi(E \vee F) = \Phi(E) \cup \Phi(F)$.

Since $\Phi$ is an isometry, it is one-to-one into. To identify the range of $\Phi$, let $E$ be the set of endpoints of those intervals which were used in defining the mapping $\Phi$ on the algebras $\mathfrak{A}_n$. Then $\Phi$ maps $\mathfrak{A}_\infty$ onto the algebra of finite unions of half-open intervals with endpoints in $E$. Suppose that $\bar{E}$ is not all of $[0, 1]$, and let $I$ be one of the open intervals of which $[0, 1) \sim \bar{E}$ is composed, that is, an interval contained in ${\sim}\bar{E}$ whose endpoints lie in $\bar{E}$. Since the endpoints of $I$ lie in $\bar{E}$, $I$ is a limit of intervals in $\Phi[\mathfrak{A}_\infty]$, and so lies in $\Phi[\mathfrak{A}]$. Let $A \in \mathfrak{A}$ be such that $\Phi(A) = I$, and let $B$ be an element of $\mathfrak{A}$ with $B \leq A$. Then $\Phi(B) \subset I$. Since $B$ can be approximated by elements in $\mathfrak{A}_\infty$, $\Phi(B)$ can be approximated by elements in $\Phi[\mathfrak{A}_\infty]$. But these latter are sets which either contain $I$ or do not meet $I$. Hence $\Phi(B) = I$ or $\Phi(B) = \emptyset$, and we have $B = A$ or $B = \emptyset$, since $\Phi$ is one-to-one. Consequently, $A$ is an atom.

We have thus proved that if $\mathfrak{A}$ has no atoms, then $\bar{E} = [0, 1]$. But if $\bar{E} = [0, 1]$, then every half-open interval is in $\Phi[\mathfrak{A}]$, and hence $\Phi[\mathfrak{A}]$ contains every Borel set. Since every measurable subset of $[0, 1]$ is the union of a Borel set and a set of measure zero, we have in this case $\Phi[\mathfrak{A}] = \mathfrak{M}/\mathfrak{N}$. ▮

This theorem states that, if $\langle X, \mathfrak{B}, \mu\rangle$ is a separable measure space without atoms for which $\mu(X) = 1$, then the corresponding measure algebra is isomorphic to the measure algebra induced by Lebesgue measure on $[0, 1]$. It does not assert the existence of a point mapping between $[0, 1]$ and $X$, or even of a set mapping of $\mathfrak{B}$ into the measurable sets of $[0, 1]$, but only of a correspondence of sets of $\mathfrak{B}$ modulo null sets with measurable sets modulo sets of measure zero. Thus if we have a set $B \in \mathfrak{B}$, we do not assign to it a particular measurable set but only an equivalence class of measurable sets in $[0, 1]$. In the next sections we shall develop criteria for asserting the existence of point mappings which induce given mappings of measure algebras.

## Problems

**11. a.** Let $\langle \mathfrak{A}, \mu \rangle$ be a measure algebra and $\langle A_n \rangle$ a sequence of elements such that $A_n \wedge A_m = 0$ for $n \neq m$. Then $\bigvee_{n=1}^{\infty} A_n = \lim_{k \to \infty} \bigvee_{n=1}^{k} A_n$. (Here lim means limit in the metric space defined by $\mu$.)

**b.** Let $\langle \mathfrak{A}, \mu \rangle$ be a measure algebra and $\langle A_n \rangle$ any sequence of elements in $\mathfrak{A}$. Then $\bigvee_{n=1}^{\infty} A_n = \lim_{k \to \infty} \bigvee_{n=1}^{k} A_n$.

**c.** Show that if $\Phi$ is an isomorphism of a measure algebra $\langle \mathfrak{A}, \mu \rangle$ into a measure algebra $\langle \mathfrak{B}, \nu \rangle$, then

$$\Phi\left( \bigvee_{n=1}^{\infty} A_n \right) = \bigvee_{n=1}^{\infty} \Phi(A_n);$$

i.e., $\Phi$ is a $\sigma$-homomorphism.

**12.** Prove that a measure algebra is complete as a metric space. [If $\langle A_n \rangle$ is a Cauchy sequence, we may assume that $\mu(A_n \,\Delta\, A_m) < 2^{-N}$ for $n, m \geq N$. Then if $B_n = \bigvee_{\nu=n}^{\infty} A_\nu$, we have $\mu(A_n \,\Delta\, B_n) < 2^{-n+1}$. Now $\bigwedge_{n=1}^{\infty} B_n = \lim B_n = \lim A_n$.]

**13.** Show that in a measure algebra the operations $'$, $\wedge$, and $\vee$ are continuous.

**14.** Show that a measure algebra (as we have defined it with $\mu(X) < \infty$) can have only a countable number of atoms. Hence any complete separable measure algebra is isomorphic either to an interval (with Lebesgue measure), to a measure space consisting of a countable number of atoms (discrete measure space), or to a measure space which is the union of the preceding two.

**15.** Discuss measure algebras in which we allow $\mu$ to be an extended real-valued function.

## 4 Borel Equivalences

If $\langle X, \mathfrak{A} \rangle$ and $\langle Y, \mathfrak{B} \rangle$ are measurable spaces, we may ask for conditions under which they are equivalent in the sense that there is a one-to-one mapping $\varphi$ of $X$ onto $Y$ such that $\varphi$ and $\varphi^{-1}$ are measurable, that is, such that $\varphi[A] \in \mathfrak{B}$ for each $A \in \mathfrak{A}$ and $\varphi^{-1}[B] \in \mathfrak{A}$ for each $B \in \mathfrak{B}$. This is a difficult problem in general. In this section we shall show that this is always the case when $X$ and $Y$ are uncountable complete separable metric spaces.

By the class of Borel sets on a metric space $X$ we mean, as usual, the smallest $\sigma$-algebra on $X$ containing the open sets (or equivalently, containing the closed sets). Let $(X, \mathfrak{A})$ and $(Y, \mathfrak{B})$ be two metric spaces with their families of Borel sets. A mapping $f: X \to Y$ is called a Borel mapping or a Borel measurable map if $f^{-1}[B]$ is a Borel set in $X$ for each Borel set in $Y$. In order to show that $f$ is a Borel map it suffices to show that $f^{-1}[O]$ is a Borel set in $X$ for each open set $O$ in $Y$. Thus continuous maps are Borel maps. A one-to-one Borel mapping of $X$ onto $Y$ whose inverse $f^{-1}$ is also a Borel map is called a **Borel equivalence** or a Borel isomorphism. Our present goal is to show that every uncountable complete separable metric space is Borel equivalent to $[0, 1]$.

We can be a little more precise if we define the order or class of a Borel map and of a Borel isomorphism. We say that a Borel map $f$ is a map of Baire class 0 if it is continuous and of Baire class[2] 1 if $f^{-1}[O]$ is an $F_\sigma$ for each open set $O \subset Y$. A Borel equivalence is said to be of class $(\alpha, \beta)$ if $f$ is of class $\alpha$ and $f^{-1}$ is of class $\beta$. Note that the composition (in either order) of a continuous map with one of class 1 is a Borel map of class 1. We shall show that the Borel equivalence of an uncountable complete separable metric space with $[0, 1]$ may be taken to be of class $(1, 1)$.

We begin with some preliminary observations and lemmas. For the purposes of this section let us agree to call a subset $E$ of a metric space *thick* if each open set which meets $E$ meets $E$ in an infinite set, i.e., if every nonempty relatively open subset of $E$ is infinite. Let us say that $E$ is *very thick* if every nonempty relatively open subset of $E$ is uncountable. Note that the definitions of thick and very thick are absolute: They depend only on $E$ and not on $X$ or the way $E$ is embedded in $X$.

A point $x \in E$ is said to be *isolated* if there is an open set $O$ such that $O \cap E = \{x\}$. It is readily seen that a set $E$ is thick if and only if it has no isolated points. If $E$ is thick, so is $\bar{E}$. If $E$ is thick and $O$ is open, then $E \cap O$ is thick.

**5. Lemma:** *Every separable metric space $X$ is the disjoint union of a very thick set $E$ and a countable (open) set $C$.*

***Proof:*** Let $C$ be the set of those $x \in X$ that are contained in a countable open set. Since $X$ is separable, the Lindelöf property

[2] See Kuratowski [11], p. 373, for the definition of the classes of higher order.

holds, and $C$ may be covered by a countable collection of open sets each of which is countable. Thus $C$ is a countable open set. The set $E = X \sim C$ has the property that each open set which contains a point of $E$ contains uncountably many points of $X$ and hence of $E$. Therefore, $E$ is very thick. ▮

**6. Lemma:** *Let $F$ be a nonempty thick closed subset. Then $F = F_1 \cup G$ with $F_1 \cap G = \emptyset$, where $F_1$ and $G$ are nonempty thick sets, $F_1$ closed and $G$ the intersection of $F$ with an open set.*

***Proof:*** Let $x_0$ and $x_1$ be two points of $F$, and let $O$ be any open set containing $x_1$ with $x_0 \notin \bar{O}$. Then $O \cap F$ is thick and so is $F_1 = \overline{O \cap F}$. Set $G = F \cap \tilde{F}_1$. Then $G$ is thick. Since $x_0 \in F_1$, the sets $F_1$ and $G$ are nonempty. ▮

**7. Lemma:** *Let $X$ be a nonempty thick separable metric space. Then $X = \bigcup G_i$, where $\langle G_i \rangle$ is a disjoint infinite sequence of nonempty thick sets of diameter less than one with each $G_i$ the intersection of an open set with a closed set.*

***Proof:*** Since $X$ is separable, it can be covered by a countable collection $\{O_j\}$ of open sets of diameter less than one. Set

$$G_k = \bar{O}_k \sim \left( \bigcup_{j=1}^{k-1} \bar{O}_j \right).$$

Then $G_k$ is thick, the intersection of a closed set with an open set, and has diameter less than one. Let $\langle G_i \rangle$ be the sequence obtained by deleting those $G_k$ which are empty. Then $X = \bigcup G_i$, and the $G_i$ are disjoint and nonempty with $G_1 = \bar{O}_1$ closed. If the sequence is only a finite sequence, we can apply Lemma 6 repeatedly to $G_1$ to obtain the desired infinite sequence. ▮

As usual, we denote by $\mathbf{N}$ the set of natural numbers with the discrete topology, or equivalently with the metric $\rho(n, m) = 1$ if $n \neq m$. Then the direct product $\mathbf{N}^\omega$ of a countable number of copies of $\mathbf{N}$ is a metrizable space which is homeomorphic to the space $\mathcal{I}$ of irrational points in an interval (see Problem 9.45 or Problem 17). The space $\mathbf{N}^\omega$ consists of all infinite sequences $\langle m_j \rangle$ of integers, and a base for its topology is given by the family of sets of the form

$$N_{i_1\cdots i_n} = \{\langle m_j\rangle \in \mathbf{N}^\omega\colon m_j = i_j \text{ for } j \le n\}.$$

**8. Proposition:** *Let $X$ be a complete separable metric space without isolated points, and let $\mathscr{I}$ be the set of irrational points in the unit interval. Then there is a one-to-one continuous map $\varphi$ of $\mathscr{I}$ onto $X$ such that $\varphi[O]$ is an $F_\sigma$ in $X$ for each open subset $O$ of $\mathscr{I}$.*

***Proof:*** It suffices to construct a one-to-one map $\varphi$ of $\mathbf{N}^\omega$ onto $X$ with $\varphi[O]$ an $F_\sigma$ for each open $O$ of $\mathbf{N}^\omega$.

Since $X$ has no isolated points, it is thick. By Lemma 7

$$X = \bigcup G_i,$$

where $\langle G_i\rangle$ is an infinite disjoint sequence of nonempty thick sets, each of which is the intersection of an open set with a closed set. Thus each $\langle G_i\rangle$ is both a $G_\delta$ and an $F_\sigma$. Moreover, $\rho(x, x') < 1$ for $x$, $x' \in G_i$, where $\rho = \rho_0$ is the metric for $X$.

Since $G_i$ is a $G_\delta$ and $X$ is complete, there is a metric $\rho_1$ for $G_i$ which is equivalent on $G_i$ to $\rho_0$ and makes $G_i$ into a complete metric space (Proposition 7.33). Define such a metric for each $G_j$ and amalgamate them into an extended real-valued metric for $X$ by setting $\rho_1(x_i, x_j) = \infty$ when $x_i \in G_i$, $x_j \in G_j$, $i \ne j$. Let us also take $\rho_1 \ge 2\rho_0$.

Since each $G_i$ is a thick complete separable metric space, we may repeat this process to obtain

$$G_i = \bigcup_{j=1}^{\infty} G_{ij},$$

where $\langle G_{ij}\rangle_{j=1}^{\infty}$ is an infinite disjoint sequence of nonempty thick sets, each the intersection of an open and a closed set (in $X$), and $\rho_1(x, x') < 1$ for $x$, $x' \in G_{ij}$. Choose $\rho_2$ equivalent to $\rho$ on $G_{ij}$ with $\rho_2 \ge 2\rho_1$ and $(G_{ij}, \rho_2)$ a complete metric space.

By induction we associate with each finite sequence $s_n = \langle i_1, \ldots, i_n\rangle$, a set $G_{i_1\cdots i_n}$, which is the intersection of an open and a closed set, has diameter less than 1 in the metric $\rho_{n-1}$ and is complete with respect to a metric $\rho_n \ge 2\rho_{n-1}$. Two different sequences of length $n$ correspond to disjoint sets. Also,

$$G_{i_1\cdots i_{n-1}} = \bigcup_{j=1}^{\infty} G_{i_1\cdots i_{n-1}j}.$$

Let $s = \langle i_j \rangle \in \mathbf{N}^\omega$, and set $s_n = \langle i_1, \ldots, i_n \rangle$. Since the $\rho$ diameter of $G_{s_n}$ is less than $2^{-n}$, the intersection

$$\bigcap_{n=1}^{\infty} G_{s_n}$$

can have at most one point.

For each $n$ choose an element $x_n \in G_{s_n}$. Then $x_l \in G_{s_l} \subset G_{s_n}$ for $l \geq n$, and so $\rho(x_l, x_m) < 2^{-n}$ for $l, m \geq n$. Hence $\langle x_n \rangle$ is a Cauchy sequence, and by the completeness of $(X, \rho)$ it converges to an element $x_s \in X$. We also have

$$\rho_k(x_l, x_m) < 2^{k-n}$$

for $l, m \geq n$. Hence $\langle x_j \rangle$ is a Cauchy sequence in $(G_{s_k}, \rho_k)$. Consequently, $\langle x_k \rangle$ converges to some point in $G_{s_k}$, which must be $x_s$, since $\rho$ and $\rho_k$ are equivalent on $G_k$, and $\langle x_j \rangle$ can have at most one limit in $(X, \rho)$. Thus $x_s \in \bigcap G_{s_n}$.

Since $\bigcap G_{s_n}$ can have at most one point, we see that it has exactly one. Thus the correspondence $\psi(s) = x_s$ defines a one-to-one map of $\mathbf{N}^\omega$ into $X$. It is readily seen to be onto, and

$$\psi[N_{i_1 \cdots i_n}] = G_{i_1 \cdots i_n}.$$

Thus each $s \in \mathbf{N}^\omega$ has a neighborhood, namely $N_{s_n}$, whose image under $\psi$ has diameter less than $2^{-n}$. Consequently, $\psi$ is continuous.

Every open set $O \subset \mathbf{N}^\omega$ is a countable union of sets $N_{i_1 \cdots i_n}$, and so $\Psi[O]$ is a countable union of sets of the form $G_{i_1 \cdots i_n}$. Since each of the latter is an $F_\sigma$, $\Psi[O]$ is an $F_\sigma$. ∎

**9. Proposition:** *Let $Z$ be the direct union of the irrationals $\mathscr{I}$ and a countably infinite discrete set $D$. Any uncountable complete separable metric space $X$ is the image of $Z$ by a one-to-one continuous map $\varphi$ with $\varphi[O]$ an $F_\sigma$ for each open set $O \subset Z$.*

***Proof:*** We may write $X = E \cup C$, where $E$ is a very thick set and $C$ is countable. Removing additional points from $E$, if necessary, we may assume that $C$ is countably infinite. Since $C$ is countable, $E$ will still be very thick. Since $X$ is uncountable, $E$ is not empty.

Now $C$ is an $F_\sigma$, and so $E$ is a $G_\delta$. Hence $E$ can be remetrized to be a complete thick metric space. Let $\varphi_0$ be the map of $\mathscr{I}$ onto $E$ given by Proposition 8. Take $\varphi(z) = \varphi_0(z)$ for $z \in E$, and take $\varphi$ on $D$ to be any one-to-one map of $D$ onto $C$. Since $Z$ is the direct union of $\mathscr{I}$ and $D$, and $\varphi$ is continuous on each summand, $\varphi$ must be continuous on $Z$. Each open set $O \subset Z$ is the union of an open subset of $\mathscr{I}$

and a subset of $D$. Thus $\varphi[O]$ is the union of two sets each of which is an $F_\sigma$, and so is an $F_\sigma$ itself. ▌

The mapping $\varphi$ of this proposition is a Borel equivalence of class (0, 1) and its inverse is a Borel equivalence of class (1, 0), Since the composition of a Borel equivalence of class (1, 0) with one of class (0, 1) is a Borel equivalence of class (1, 1), we have the following corollary:

**10. Theorem (Kuratowski):** *Let* $(X, \rho)$ *and* $(Y, \sigma)$ *be two uncountable complete separable metric spaces. Then there is a Borel equivalence between them of class* (1, 1).

In particular, every uncountable complete separable metric space is Borel equivalent to [0, 1].

## Problems

16. **a.** Show that the closure of a thick set is thick.

    **b.** Show that the intersection of a thick set and an open set is thick.

    **c.** Show that a set is thick if and only if it has no isolated points.

17. Show that every countable complete metric space has isolated points.

18. A metric space is said to be *zero-dimensional* if there is a base for its topology consisting of sets that are both open and closed.

    **a.** Show that every subset of a zero-dimensional space is zero-dimensional.

    **b.** Let $E$ be a subset of an interval $I \subset \mathbf{R}$ whose complement is dense in $I$. Then $E$ is zero-dimensional.

    **c.** If $X$ is zero-dimensional, then the sets $G_i$ in Lemma 7 can be taken to be open (and closed).

    **d.** Let $E \subset \mathbf{R}$ be a $G_\delta$ without isolated points such that $\tilde{E}$ is dense in $\mathbf{R}$. Then $E$ is homeomorphic to $\mathbf{N}^\omega$. [Hint: Make use of (c) in proving Proposition 8.]

## 5 Borel Measures on Complete Separable Metric Spaces

Let $X$ be a metric space. By a Borel measure on $X$ we mean a measure defined on the Borel sets of $X$ or, since we prefer to deal with complete measure, the completion of such a measure. We begin with a generalization of Littlewood's first principle, that states a form of regularity for finite measures.

**11. Proposition:** *Let $\mu$ be a finite Borel measure on a metric space $X$. Then for any Borel set $E$ we have*

$$\mu E = \inf \{\mu O \colon E \subset O,\ O \text{ open}\}$$

*and*

$$\mu E = \sup \{\mu F \colon F \subset E,\ F \text{ closed}\}.$$

***Proof:*** The family $\mathcal{R}$ of sets $E$ that satisfy the conclusions of the proposition is a $\sigma$-algebra. If $\mathcal{R}$ contains all closed sets, it must, therefore, include all Borel sets.

Every closed set automatically satisfies the second of these conditions. Thus we need only show that for each closed set $F$ we have

$$\mu F = \inf \{\mu O \colon O \supset F,\ O \text{ open}\}.$$

Since $F \subset O$, $\mu F \leq \mu O$, and so $\mu F$ is at most equal to the infimum. On the other hand, each closed subset of a metric space is a $G_\delta$. Hence there is a sequence $O_i$ of open sets with

$$F = \bigcap O_i.$$

Since $mO_1 < \infty$, we have $\mu F = \lim \mu O_i \geq \inf \{\mu O \colon O \supset F, O \text{ open}\}$. ▮

Two measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are said to be *isomorphic* if there is a one-to-one map $\varphi$ of $X$ onto $Y$ such that for all $A \in \mathcal{A}$ we have $\varphi[A] \in \mathcal{B}$ and $\nu(\varphi[A]) = \mu A$, and for all $B \in \mathcal{B}$ we have $\varphi^{-1}[B] \in \mathcal{A}$, and hence $\mu(\varphi^{-1}[B]) = \nu B$. For brevity we often say that the measure $\mu$ and $\nu$ are isomorphic. If two measures are isomophic, so are their completions. As a consequence of the Borel equivalence of any uncountable complete separable metric space with $[0, 1]$, we have the following proposition:

**12. Proposition:** *Every Borel measure on an uncountable complete separable metric space is isomorphic to a Borel measure on* $[0, 1]$.

We shall obtain more precise results in the case of finite measures, but first we state some properties of Borel measures on separable metric spaces. We leave the proofs to the reader (Problems 21 and 22).

**13. Proposition:** *Let $X$ be a separable metric space and $\mu$ a Borel measure on $X$. Then there is a unique closed set $F \subset X$* such that *$\mu(X \sim F) = 0$ and $\mu(F \cap O) > 0$ for every open set $O$ for which $F \cap O \neq \emptyset$.*

The closed set $F$ given in the proposition is called the **support** of $\mu$. We say that a Borel measure is strictly positive on a Borel set $E$ if $\mu(E \cap O) > 0$ for every open set $O$ for which $E \cap O \neq \emptyset$. Thus the support of a measure is the largest closed set on which $\mu$ is strictly positive and the smallest closed set whose complement has zero measure.

A Borel set $A$ is called an *atom* for the measure $\mu$ if, given any Borel set $B \subset A$, we have either $\mu B = 0$ or $\mu(A \sim B) = 0$. The following is a useful property of atoms in separable spaces.

**14. Lemma:** *Let $\mu$ be a Borel measure on a separable metric space $X$, and $A$ an atom for $\mu$. Then there is a point $x \in A$ such that $\mu(A \sim \{x\}) = 0$.*

We call $\{x\}$ the support of the atom $A$. Recall that if $\mu$ is finite (or $\sigma$-finite) it can have only a countable number of atoms. Let $\mu$ be a finite (or $\sigma$-finite) Borel measure and $C = \{x_i\colon \{x_i\}$ the support of an atom of $\mu\}$. Then $C$ is a countable set, and the measure $\mu$ restricted to $X \sim C$ has no atoms. The measure $\mu_2$ defined by $\mu_2 E = \mu(E \cap C)$ is called the atomic part of $\mu$, and the measure $\mu_1$ defined by $\mu_1(E) = \mu(E \sim C)$ the atom-free part or the continuous part of $\mu$. Clearly, $\mu = \mu_1 + \mu_2$, and this decomposition is unique. We call $C$ the *carrier* of the measure $\mu_2$. Note that the support of $\mu_2$ is the set $\overline{C}$, which is generally larger than $C$. The set $X \sim C$ is a $G_\delta$. We summarize these facts in the following lemma:

**15. Lemma:** *Let $\mu$ be a finite (or $\sigma$-finite) Borel measure on a separable metric space $X$. Then $X$ is the disjoint union $X = X_1 \cup C$, where $C$ is countable, $X_1$ has no isolated points, and $\mu$ restricted to $X_1$ is without atoms.*

***Proof:*** Let $C$ be the union of the carrier of the atomic part of $\mu$ and the countable set with very thick complement in Lemma 5. ∎

By a **standard measure space** we shall mean one of the following:

i. An interval $[a, b] \subset \mathbf{R}$ with $m$ Lebesgue measure or $m$ the zero measure.

ii. A countable discrete set (of integers, say) and $m$ any finite measure.
iii. The disjoint union of a measure space of type (i) and one of type (ii).

**16. Theorem:** *Let $\mu$ be a finite Borel measure on a complete separable metric space $X$. Then $(X, \mathcal{B}, \mu)$ is isomorphic to a standard measure space. If $X$ is uncountable and $\mu$ has no atoms, then $(X, \mathcal{B}, \mu)$ is isomorphic to a standard space of type* (i).

***Proof:*** If $X$ is countable, $(X, \mathcal{B}, \mu)$ is isomorphic to a measure space of type (ii). Suppose that $X$ is uncountable and has atoms. Let $C$ be the carrier of the atomic part of $\mu$. Then $C$ can be put in one-to-one correspondence with a subset of the negative integers, and $\mu$ restricted to $C$ is isomorphic to a measure on this set.

Since $X \sim C$ is a $G_\delta$, it can be remetrized to be a complete metric space. Thus is suffices to show that if $\mu$ is a Borel measure without atoms on an uncountable complete separable metric space $X$, then $(X, \mathcal{B}, \mu)$ is isomorphic to Lebesgue measure on $[0, b]$, where $b = \mu X$ (which we assume to be positive).

By Proposition 12 the measure space $(X, \mathcal{B}, \mu)$ is isomorphic to a Borel measure $\bar{\mu}$ on $[0, 1]$. Let $F$ be the support of $\bar{\mu}$, and let us first consider the case when $F = [0, 1]$. Let $\varphi(x)$ be the cumulative distribution function of $\bar{\mu}$ defined by $\varphi(x) = \bar{\mu}[0, x]$. Since $\bar{\mu}$ has no atoms, $\varphi$ is continuous. The support of $\bar{\mu}$ is $[0, 1]$, and so $\varphi$ is strictly increasing. Thus $\varphi$ is a homeomorphism of $[0, 1]$ onto $[0, b]$, and it is readily seen that $\bar{\mu}E = m(\varphi[E])$ for any Borel set $E \subset [0, 1]$.

It remains only to consider the case when the support $F$ of $\bar{\mu}$ is properly contained in $[0, 1]$. Let $G = (0, 1) \cap \tilde{F}$. Then $G$ is an open set and so the disjoint union of a countable collection of open intervals. Let $\psi(x) = \bar{\mu}[0, x]$. Then $\varphi$ is a continuous map of $[0, 1]$ onto $[0, b]$, where $b = \mu X$, but $\psi$ maps each interval of $G$ onto a single point. Let $L$ and $R$ be the sets of left (or right) endpoints of intervals of $G$. Then $\psi[G] = \psi[L] = \psi[R]$.

Set $H = F \cap \tilde{R} \cap \tilde{L} \cap \tilde{D}$, where $D$ is a countable subset of $F$. Then $\varphi$ is a one-to-one continuous map of $F$ onto $[0, b] \sim \psi[R]$. Since $\psi$ takes compact sets of $F$ into compact sets, it takes relatively closed subsets of $H$ onto relatively closed subsets of $[0, b] \sim \psi[R \cup D]$. Hence it is a homeomorphism of $H$ onto $[0, b] \sim \psi[R \cup D]$. Let $N$ be a nonempty $G_\delta$ without isolated points contained in $\psi[H]$ and having $mN = 0$. By Lemma 6 we may express $N$ as the disjoint union $N = N_1 \cup N_2$ of two nonempty $G_\delta$'s without isolated points. Since $\psi^{-1}[N]$ is a $G_\delta$ without isolated points, there is a Borel equivalence $\varphi$ between $\psi^{-1}[N]$ and $N_1$, a Borel equivalence between $G$ and $N_2$, and a Borel equivalence between $L \cup R \cup D$ and $\psi[R \cup D]$. Putting these Borel equivalences with $\psi|(H \sim \psi^{-1}[N])$ gives a Borel equivalence $\varphi$ of $[0, 1]$ with $[0, b]$. For any Borel set $E$ in $[0, b]$,

$$\begin{aligned} mE &= m([E \cup \tilde{N}] \sim \psi[R \cup D]) \\ &= \mu(\psi^{-1}[E \cup \tilde{N}] \sim \psi[R \cup D]) \\ &= \mu(\psi^{-1}[E]). \quad \blacksquare \end{aligned}$$

In establishing this last theorem, we have not been concerned with the type of the Borel equivalence involved, but it is sometimes useful to have a version where our mapping from a "standard" space is a Borel equivalence of type (0, 1) as in Proposition 9. Since we now demand more of the Borel equivalence $\varphi$, we must allow more latitude in what is taken to be a standard space. For simplicity we consider only the case of nonatomic measures. this leads us to the following definition.

By an *atomless standard metric measure space* we mean a metric space $Z$ with a Borel measure $m$ on $Z$, where $(Z, m)$ is one of the following:

i. $Z = [a, b] \sim D_1$, where $[a, b] \subset \mathbf{R}$, $D_1$ is countable and $m$ is Lebesgue measure.
ii. $Z = (b, c] \sim D_2$, where $(b, c] \subset \mathbf{R}$, $D_2$ is countable, and $m$ is the zero measure.
iii. The disjoint union of a measure space of type (i) and one of type (ii).

Using Proposition 9 instead of its corollary, one can establish the following theorem. Details of the proof are left to the reader.

**17. Theorem:** *Let $\mu$ be a finite Borel measure on a complete separable metric space $X$. Then there is a continuous one-to-one map $\varphi: Z \to X$ from a standard metric measure space $Z$ onto $X$ such that $\varphi[O]$ is an $F_\sigma$ for each open set $O \subset Z$ and $\mu E = m(\varphi[E])$ for each Borel set $E \subset X$.*

We have to allow for the deletion from $Z$ of the countable sets $D_1$ and $D_2$; otherwise the thick part of $\varphi[Z]$ would be compact and connected by the continuity of $\varphi$. Also, $\varphi^{-1}[X \sim F]$ must be open when $F$ is the support of $\mu$. Thus the presence of a component of type (ii) is forced on us when the support of $\mu$ does not contain all of the thick part of $X$.

As a consequence of this theorem we have the following proposition, which is a strengthening of Proposition 11 in the case when $X$ is complete.

**18. Proposition:** *Let $X$ be a complete separable metric space and $\mu$ a Borel measure on $X$. Then for every Borel set $E$ with $\mu E < \infty$ we have*

$$\mu E = \sup \{\mu K \colon K \subset E, K \text{ compact}\}.$$

***Proof:*** Replacing $\mu$ by $\nu$, where $\nu A = \mu(A \cap E)$, we see that we need only to establish the proposition for finite measures. For simplicity, we suppose that $\mu$ has no atoms. Then there is a continuous map $\varphi$ of a standard metric measure space onto $X$ such that $\mu E = mA$, where $A = [a, b] \cap \varphi^{-1}[E]$. Proposition 3.15 says that given $\epsilon > 0$, there is a closed set $F \subset A$ such that

$$mF > mA - \epsilon.$$

Since $F$ is a closed and bounded subset of $\mathbf{R}$, it is compact, and so is $K = \varphi[F]$ by the continuity of $\varphi$. Thus $K \subset E$ and

$$\mu K = mF > mA - \epsilon = \mu E - \epsilon.$$

Since $\epsilon$ was arbitrary, the proposition follows. ∎

## Problems

**19.** Show that the collection $\mathfrak{R}$ of sets satisfying the conclusion of Proposition 11 is a $\sigma$-algebra. Hints:

**a.** A set $E$ satisfies the first condition if and only if $\tilde{E}$ satisfies the second.

**b.** If $\langle E_i \rangle$ is a sequence of sets, each of which satisfies the first condition, so does $\bigcap E_i$.

**c.** If $\langle E_i \rangle$ is a sequence of sets, each of which satisfies the second condition, so does $\bigcup E_i$.

**d.** The collection $\mathfrak{R}$ is is $\sigma$-algebra.

**20.** Show that if the measure spaces $(X, \mathfrak{A}, \mu)$ and $(Y, \mathfrak{B}, \nu)$ are isomorphic, so are their completions $(X, \mathfrak{A}_0, \bar{\mu})$ and $(Y, \mathfrak{B}_0, \bar{\nu})$.

**21.** Prove Proposition 13. [Hint: Let $U = \bigcup \{O: O \text{ open and } \mu O = 0\}$. Then $\mu U = 0$, $F = X \sim U$ is the desired set.]

**22.** Prove Lemma 14. [Hint: Define $\nu$ by $\nu E = \mu(E \cap A)$. Then the support of $\nu$ contains exactly one point.]

**23.** Prove Theorem 17.

**24.** Fill in the details of the proof of Proposition 18 when $\mu$ has atoms.

**25.** **a.** Show that Theorem 16 extends to the case of a $\sigma$-finite measure $\mu$ if we allow the semi-infinite interval $[0, \infty)$ in our definition of standard measure space.

**b.** Modify the definition of standard metric measure spaces so that Theorem 17 also encompasses the case of a $\sigma$-finite measure.

**c.** Show that the hypothesis "$\mu E < \infty$" in Proposition 18 may be weakened to "$E$ a set of $\sigma$-finite measure".

## 6 Set Mappings and Point Mappings on Complete Separable Metric Spaces

Let $(X, \mathcal{A}, \mathcal{N})$ and $(Y, \mathcal{B}, \mathcal{M})$ be measurable spaces with null sets with $X$ and $Y$ separable metric spaces and $\mathcal{A}$ and $\mathcal{B}$ their Borel sets. In this section we investigate the extent to which $\sigma$-homomorphisms and isomorphisms of the Boolean $\sigma$-algebras $\mathcal{A}/\mathcal{N}$ and $\mathcal{B}/\mathcal{M}$ are induced by point mappings. The following Proposition is an immediate consequence of Proposition 3 and Theorem 10.

**19. Proposition:** *Let $(X, \mathcal{A}, \mathcal{N})$ be a measurable space with null sets, $Y$ an uncountable complete separable metric space, and $\Phi$ a $\sigma$-homomorphism of the Borel sets $\mathcal{B}$ of $Y$ into $\mathcal{A}/\mathcal{N}$. Then there is a Borel measurable map $\varphi: X \to Y$ of $X$ onto $Y$ such that for each $B \in \mathcal{B}$ we have $\varphi^{-1}[B]$ in the equivalence class $\Phi(B)$. If $\psi$ is any other Borel map with this property, then $\varphi = \psi$* a.e. $[\mathcal{N}]$.

The following theorem is a generalization of a theorem of von Neumann, who assumed the unnecessary hypothesis that $\Phi$ be measure preserving with respect to suitable measures on $\mathcal{A}$ and $\mathcal{B}$.

**20. Theorem:** *Let $X$ and $Y$ be complete separable metric spaces, $\mathcal{A}$ and $\mathcal{B}$ their Borel sets, and let $\mathcal{M}$ and $\mathcal{N}$ be $\sigma$-ideals in $\mathcal{A}$ and $\mathcal{B}$. If $\Phi$ is a $\sigma$-isomorphism of $\mathcal{A}/\mathcal{M}$ onto $\mathcal{B}/\mathcal{N}$ then there are sets $X_0 \in \mathcal{M}$ and $Y_0 \in \mathcal{N}$ and a one-to-one mapping $\varphi$ of $Y \sim Y_0$ onto $X \sim X_0$ such that $\varphi$ and $\varphi^{-1}$ are measurable and $\Phi(A) = \varphi^{-1}[A]$ modulo $\mathcal{N}$.*

***Proof:*** By Proposition 19 there are Borel measurable mappings $\varphi: Y \to X$ and $\psi: X \to Y$ that induce $\Phi$ and $\Phi^{-1}$. Since $\psi \circ \varphi$ induces the identity $\sigma$-isomorphism on $\mathfrak{A}/\mathfrak{N}$ to itself, we must have $\psi \circ \varphi = \text{id}$ a.e. $[\mathfrak{N}]$ by the uniqueness assertion of Proposition 19. Hence there is a set $X_0 \in \mathfrak{M}$ such that $\varphi$ is a Borel equivalence of $X \sim X_0$ onto $\varphi[X \sim X_0]$. Since $\varphi[X \sim X_0]$ is in the equivalence class $\Phi^{-1}(X)$, which is the equivalence class containing $Y$, we have $Y_0 = Y \sim \varphi[X \sim X_0]$ an element of $\mathfrak{N}$. ∎

If the Borel sets $X_0$ and $Y_0$ have the same cardinality, it follows from a result of Kuratowski [11] that they are Borel equivalent, and hence the Borel equivalence $\varphi$ can be extended to a Borel equivalence of $X$ with $Y$. Simple examples (Problem 25) show that $X_0$ and $Y_0$ may have different cardinality, and so the presence of one or the other may be unavoidable. They can be eliminated, however, if $\Phi$ is an automorphism of $\mathfrak{A}/\mathfrak{N}$ onto itself, as the following theorem shows:

**21. Theorem:** *Let $X$ be a complete separable metric space, $\mathfrak{B}$ the family of Borel sets of $X$, and $\mathfrak{N}$ a $\sigma$-ideal of $\mathfrak{B}$. If $\Phi$ is any $\sigma$-isomorphism of $\mathfrak{B}/\mathfrak{N}$ onto itself, then there is a one-to-one mapping $\varphi$ of $X$ onto itself such that $\varphi$ and $\varphi^{-1}$ are Borel measurable and $\Phi(A) = \varphi^{-1}[A]$ modulo $\mathfrak{N}$.*

***Proof:*** By Theorem 20 there are sets $X_0$ and $Y_0$ in $\mathfrak{N}$ and a one-to-one Borel measurable mapping $\psi$ of $X \sim Y_0$ onto $X \sim X_0$. Let $Z_0 = X_0 \cup Y_0$, $Z_1 = \psi[Z_0] \cup \psi^{-1}[Z_0]$, $Z_{n+1} = \psi[Z_n] \cup \psi^{-1}[Z_n]$. Since $\psi$ and $\psi^{-1}$ take sets in $\mathfrak{N}$ into sets in $\mathfrak{N}$, we have $Z_n \in \mathfrak{N}$. Set $Z = \bigcup Z_n$. Then $Z \in \mathfrak{N}$, $\psi[Z] \subset Z$, and $\psi^{-1}[Z] \subset Z$. Thus $\psi$ and $\psi^{-1}$ are one-to-one mappings of $X \sim Z$ into $X \sim Z$, and since they are inverses of each other they must map $X \sim Z$ onto $X \sim Z$. Define $\varphi$ by letting $\varphi(x) = \psi(x)$ if $x \in X \sim Z$ and $\varphi(x) = x$ if $x \in Z$. Then $\varphi$ is a Borel equivalence of $X$ onto itself which agrees with $\psi$ except on the set $Z \in$ . ∎

We say that a metric space $X$ is topologically complete if it has an equivalent metric that makes it complete. The condition that $X$ be complete in the propositions and theorems of the last two sections can be replaced by the condition that $X$ is topologically complete, since the metric itself does not enter into their conclusions. The condition of topological completeness for metric spaces is equivalent to that of being an absolute $G_\delta$: Every $G_\delta$ in a complete metric space is topologically complete, and a topologically complete subset of any metric space is a $G_\delta$. Thus the condition that $X$ is topologically complete guarantees that the Borel subsets of $X$ are absolutely measurable, that is, that they are Borel subsets of any metric space in which they are embedded.

The assumption of completeness is essential for obtaining point mappings as in the theorems of this section. This is shown by the following example. There is a set $A \subset [0, 1]$ with $m^*A = 1$ and $m^*B = 1$, where $B = [0, 1] \sim A$ (see Halmos [5], p. 70). Let $\psi: \mathbf{R} \to \mathbf{R}$ be defined by $\psi(x) = -x$. For each set $E \subset \mathbf{R}$, let $h(E)$ be its measurable hull (see Section 12.6). Then $h(E)$ is determined modulo a set of measure zero and so $h(E)$ can be taken to be an element of $\mathcal{B}/\mathcal{M}_0$, where $\mathcal{B}$ is the class of Borel subsets of $[-1, 1]$, and $\mathcal{M}_0$ those of measure zero. Let $X = A \cup \psi[B]$, and take $\mathcal{A}$ to be the $\sigma$-algebra of subsets of $X$ of the form $E = X \cap F$, where $F$ is a Borel set, and take $\mathcal{N}$ to be the subset of $\mathcal{A}$ consisting of sets of measure zero. Then $h(X \cap F)$ is in the equivalence class of $F$, and so $h$ is a $\sigma$-isomorphism of $\mathcal{A}/\mathcal{N}$ onto $\mathcal{B}/\mathcal{M}_0$. If we define $\Phi$ by

$$\Phi(E) = X \cap \psi[h(E)],$$

then $\Phi$ is a $\sigma$-isomorphism of $\mathcal{A}/\mathcal{N}$ onto itself. Let $\Theta$ be the map of the Borel subsets of $[-1, 1]$ onto $\mathcal{A}/\mathcal{N}$ given by $\Theta(F) = [F \cap X]$. Then any map $\theta: X \to [-1, 1]$, which induces $\Phi \circ \Theta$ must be equal to $\psi$ a.e. But $X$ and $\psi[X]$ are disjoint except for 0. From this it follows that there is no point mapping $\varphi: X \to X$ which induces $\Phi$.

In this example we see that the associated point mapping ought to be $\psi$, but $\psi$ does not map $X$ into $X$, since $X$ has "too many gaps." Completion of $X$ under any suitable metric should fill in the gaps so that $\psi$ can map points of $X$ to the points added by completion.

## Problems

**26.** Show that the sets $X_0$ and $Y_0$ (or at least one of them) may be unavoidable in Theorem 20. [Let $X$ be the set of irrationals in [0, 1] and $\mathcal{M} = \{\emptyset\}$. Let $Y$ be [0, 1] and $\mathcal{N}$ consists of all subsets of rationals. Take $\Phi(E) = E$.]

**27.** Let $(X, \mathcal{A}, \mathcal{N})$ be the example described at the end of the section.

**a.** Show that for each Borel subset $F$ of $[-1, 1]$ we have $h(F \cap X) = [F]$, where $[F]$ is the class of Borel sets that differ from $F$ by a set of measure zero.

**b.** Show that $\Phi$ is a $\sigma$-isomorphism of $\mathfrak{A}/\mathfrak{N}$ onto itself.

**c.** Show that if $\theta: X \to [-1, 1]$ induces $\Phi \circ \Theta$, then $\theta$ is equal to $\psi$ a.e.

**d.** Show that there is no point mapping $\varphi: X \to X$ that induces $\Phi$.

## 7 The Isometries of $L^p$

We illustrate the use of the theorems in the preceding sections by deriving a characterization of the isometries of $L^p[0, 1]$ onto itself, that is, of those linear mappings $U$ of $L^p[0, 1]$ into itself such that $\|Uf\| = \|f\|$. We begin by establishing two inequalities, the first concerning real (or complex) numbers and the second elements in $L^p$.

**22. Lemma:** *Let $\xi$ and $\eta$ be real numbers. Then if $2 \le p < \infty$,*

$$|\xi + \eta|^p + |\xi - \eta|^p \ge 2(|\xi|^p + |\eta|^p),$$

*while if $0 < p \le 2$,*

$$|\xi + \eta|^p + |\xi - \eta|^p \le 2(|\xi|^p + |\eta|^p).$$

*If $p \ne 2$, equality can only hold if $\xi$ or $\eta$ is zero.*

***Proof:*** If $p = 2$, we have equality for all $\xi$ and $\eta$. If $2 < p < \infty$, then $1 \le p/2$, and applying the Hölder inequality with exponents $p/2$ and $p/(p-2)$ to $\alpha^2 + \beta^2$ we obtain

$$\alpha^2 + \beta^2 \le (\alpha^p + \beta^p)^{2/p}(1 + 1)^{(p-2)/p}$$

or

$$\alpha^p + \beta^p \ge 2^{(2-p)/2}(\alpha^2 + \beta^2)^{p/2}. \tag{1}$$

If $0 < p < 2$, replace $p$ by $4/p$. Then (1) becomes

$$\alpha^{4/p} + \beta^{4/p} \ge 2^{(p-2)/p}(\alpha^2 + \beta^2)^{2/p}.$$

If we replace $\alpha$ by $\alpha^{p/2}$ and $\beta$ by $\beta^{p/2}$, this becomes

$$\alpha^2 + \beta^2 \ge 2^{(p-2)/p}(\alpha^p + \beta^p)^{2/p},$$

or

$$\alpha^p + \beta^p \le 2^{(2-p)/2}(\alpha^2 + \beta^2)^{p/2}. \tag{2}$$

Since

$$0 \le \frac{\xi^2}{\xi^2 + \eta^2} \le 1,$$

we have

$$\frac{|\xi|^p}{(\xi^2+\eta^2)^{p/2}} \le \frac{\xi^2}{\xi^2+\eta^2} \qquad 2 < p \tag{3}$$

and

$$\frac{|\xi|^p}{(\xi^2+\eta^2)^{p/2}} \ge \frac{\xi^2}{\xi^2+\eta^2} \qquad p < 2. \tag{4}$$

Forming similar inequalities for $\eta$ and adding, we get

$$|\xi|^p + |\eta|^p \le (\xi^2+\eta^2)^{p/2} \qquad 2 < p \tag{5}$$

and

$$|\xi|^p + |\eta|^p \ge (\xi^2+\eta^2)^{p/2} \qquad p < 2. \tag{6}$$

We verify that equality can hold in (3) or (4) and hence in (5) or (6) only if $\xi = 0$ or $\eta = 0$.

Assume $p > 2$, and replace $\alpha$ and $\beta$ in (1) by $|\xi + \eta|$ and $|\xi - \eta|$. Then

$$\begin{aligned} |\xi+\eta|^p + |\xi-\eta|^p &\ge 2^{(2-p)/2}(|\xi+\eta|^2 + |\xi-\eta|^2)^{p/2} \\ &= 2(\xi^2+\eta^2)^{p/2} \\ &\ge 2(|\xi|^p + |\eta|^p) \end{aligned}$$

by (5). This establishes the first inequality in the lemma, and the second follows similarly using (2) and (6). We see that in either case for $p \ne 2$ equality can occur in (3) and (4) only if $\xi = 0$ or $\eta = 0$. ∎

By integration we have the following lemma as a consequence of Lemma 22.

**23. Lemma:** *Let $1 \le p < \infty$, $p \ne 2$, and suppose that $f$ and $g$ are in $L^p$. Then*

$$\|f+g\|^p + \|f-g\|^p = 2(\|f\|^p + \|g\|^p)$$

*if and only if $f \cdot g = 0$ almost everywhere.*

**24. Theorem (Lamperti):** *Let $1 \le p < \infty$, $p \ne 2$, and let $U$ be a linear transformation of $L^p[0, 1]$ into itself such that $\|Uf\|_p = \|f\|_p$. Then there is a Borel measurable mapping $\varphi$ of $[0, 1]$ onto (almost all of) $[0, 1]$ and an $h \in L^p$ such that*

$$Uf = h \cdot (f \circ \varphi).$$

*The function h is uniquely determined (to within* a.e. *equivalence) and $\varphi$ is uniquely determined (to within* a.e. *equivalence) on the set where $h \neq 0$. For any Borel set E we have*

$$\int_{\varphi^{-1}[E]} |h|^p \, dt = \int_E dt.$$

***Proof:*** We define the carrier of a function $f$ to be the set $\{t : f(t) \neq 0\}$. If $f \varepsilon L^p$, then the carrier of $f$ is only defined modulo null sets. Thus for $f \varepsilon L^p$ the carrier of $f$ is an element in the $\sigma$-algebra $\mathcal{B}/\mathfrak{N}$ of Borel sets modulo sets of measure zero. We define a mapping $\Phi$ of the Borel sets of $[0, 1]$ into Borel sets modulo null sets by setting $\Phi(A)$ equal to the support of $U\chi_A$. If $A$ and $B$ are disjoint, we have by Lemma 23 that

$$\|\chi_A + \chi_B\|^p + \|\chi_A - \chi_B\|^p = 2(\|\chi_A\|^p + \|\chi_B\|^p).$$

Since $U$ is linear and preserves norms, we have, again using Lemma 23,

$$(U\chi_A) \cdot (U\chi_B) = 0 \text{ a.e.}$$

Thus $\Phi$ takes disjoint sets into disjoint sets. If $A$ and $B$ are disjoint, $\chi_{A \cup B} = \chi_A + \chi_B$, and so $U\chi_{A \cup B} = U\chi_A + U\chi_B$. Since the two functions on the right have disjoint carrier, we see that $\Phi(A \cup B) = \Phi(A) \cup \Phi(B)$ for disjoint sets and hence for any pair for Borel sets.

If $\Phi[[0, 1]] = E$, we have $\Phi(\tilde{A}) = E \sim \Phi(A)$. Thus $\Phi$ is a homomorphism of $\mathcal{B}$ into the algebra of Borel subsets of $E$. If $A = \bigcup_{i=1}^{\infty} A_i$ with the $A_i$ disjoint, we have

$$\chi_A = \lim \sum_{i=1}^{n} \chi_{A_i}$$

in $L^p$, and so by the continuity of $U$,

$$U\chi_A = \lim \sum_{i=1}^{n} U\chi_{A_i}.$$

Hence $\Phi(A) = \bigcup \Phi(A_i)$, and $\Phi$ is a $\sigma$-homomorphism.

By Proposition 3 there is a Borel-measurable mapping $\varphi$ of $E$ into $[0, 1]$ such that $\Phi(A) = \varphi^{-1}[A]$. Since $\Phi$ can only take sets of measure zero into sets of measure zero, the mapping $\varphi$ must be onto all of $[0, 1]$ except possibly for a set of measure zero. Extend $\varphi$ to be defined on all of $[0, 1]$ by setting $\varphi(t) = 0$ for $t \notin E$.

The function $1 = \chi_{[0,1]}$ is in $L^p$. Let $h = U(1)$. Then $h \in L^p$ and the carrier of $h$ is $E$. If $A$ is any Borel set in $[0, 1]$, we have $1 = \chi_A + \chi_{\tilde{A}}$. Hence $h = U\chi_A + U\chi_{\tilde{A}}$. But the functions on the right have disjoint carrier, and so $U\chi_A$ must equal $h$ on the carrier of $U\chi_A$. Thus $U\chi_A = h \cdot \chi_{\Phi(A)} = h \cdot (\chi_A \circ \varphi)$. Hence, if $\psi$ is any simple function, we must have $U\psi = h \cdot (\psi \circ \varphi)$. Since every function in $L^p$ can be approximated in norm by a simple function and $U$ is norm preserving, we have $Uf = h \cdot (f \circ \varphi)$ for all $f \in L^p$.

The remaining statements in the theorem now follow easily. ∎

## Problems

**28. a.** Show that if the linear transformation $U$ is onto $L^p[0, 1]$, then $X \sim E$ has measure zero.

**b.** Show in this case that $\varphi$ is essentially one-to-one (that is, one-to-one except on a set of measure zero), that $\varphi^{-1}$ is measurable, and that $\varphi$ and $\varphi^{-1}$ take sets of measure zero into sets of measure zero. [Hint: Apply Theorem 24 also to the transformation $U^{-1}$, and the uniqueness part of the theorem to $I = UU^{-1}$ and $I = U^{-1}U$.]

**c.** Show in this case that if we define a measure $\mu$ by $\mu A = m(\varphi[A])$, then $|h|^p = \left[\dfrac{d\mu}{dm}\right]$.

**29.** What can you say about the isometries of $L^p(X, \mu)$ for general finite measure spaces? (In general, you cannot get a point mapping but must be content with the set mapping $\Phi$.)

**30.** Show that the characterization given in Theorem 24 is false for $L^2[0, 1]$.

**31.** A Banach space $X$ is called *uniformly convex* if given $\epsilon > 0$ there is an $\eta < 1$ such that whenever $u$ and $v$ are elements of $X$ with $\|u\| = \|v\| = 1$ and $\|u - v\| \geq 2\epsilon$ we have $\|u + v\| \leq 2\eta$. Use the integrated form of Lemma 22 to show that $L^p$ is uniformly convex if $p \geq 2$ and that we can take $\eta = (1 - \epsilon^p)^{1/p}$.

# 16 The Daniell Integral

## 1 Introduction

It is sometimes convenient to introduce integration directly without using the concept of measure. This happens when we have an elementary integral $I$ defined on some class $L$ of 'elementary functions' and want to enlarge the class $L$ and extend $I$ so that it has all the properties of the abstract Lebesgue integral, including the convergence theorems. If we take, for example, our class $L$ to consist of the continuous real-valued functions on $(-\infty, \infty)$ each of which vanishes outside a finite interval and take $I$ to be the Riemann integral, we might expect to obtain the class of Lebesgue integrable functions and the Lebesgue integral by some sort of extension process. Such a process has been carried out by Daniell for this case and generalized by Marshall Stone, who also clarified the structure of the extended integral. It is the purpose of the present chapter to describe this extension procedure and to show its connection with measure theory.

Let $L$ be a family of real-valued functions on some set $X$, and suppose that $L$ is a vector lattice, that is, that whenever the functions $f$ and $g$ are in $L$ so also are the functions $\alpha f + \beta g$, $f \vee g$, and $f \wedge g$. Since $f \wedge g = f + g - (f \vee g)$ and $f \vee g = (f - g) \vee 0 + g$, we see that a vector space $L$ of functions is a vector lattice if for each $h$ in $L$ we have $h \vee 0$ in $L$. Thus a vector space of functions is a vector lattice if it is closed under the operation taking $f$ into $f^+ = f \vee 0$. Since $|f| = f^+ + (-f)^+$, each vector lattice contains the absolute value of each function in the lattice. Conversely, if $L$ is a vector space such that $|f|$ is in $L$ for each $f$ in $L$, then $L$ is a vector lattice, since $f^+ = \frac{1}{2}(f + |f|)$.

A linear functional $I$ on $L$ is said to be positive[1] if $I(\varphi) \geq 0$ for each nonnegative function $\varphi$ in $L$. If $I$ is positive and $\varphi \leq \psi$, then $I(\varphi) \leq I(\psi)$. A positive linear functional $I$ on $L$ is called a **Daniell functional** or a Daniell integral if the following condition is satisfied:

D. *If $\langle \varphi_n \rangle$ is a sequence of functions in $L$ that decrease to zero at each point, then* $\lim I(\varphi_n) = 0$.

This condition is clearly equivalent to each of the following conditions:

D′. *If $\langle \varphi_n \rangle$ is an increasing sequence of functions in $L$, and if $\varphi$ is a function in $L$ such that*[2] $\varphi \leq \lim \varphi_n$, *then* $I(\varphi) \leq \lim I(\varphi_n)$.

D″. *If $\langle u_n \rangle$ is a sequence of nonnegative functions in $L$, and if $\varphi$ is a function in $L$ such that $\varphi \leq \sum u_n$, then $I(\varphi) \leq \sum I(u_n)$.*

One example of a Daniell integral is obtained by taking $L$ to be the set of continuous functions on $(-\infty, \infty)$ each vanishing outside a finite interval and taking $I$ to be the Riemann integral. Another is obtained from a measure $\mu$ on an algebra $\mathfrak{A}$ of subsets of $X$ by letting $L$ be the class of simple functions and $I$ the natural integration with respect to $\mu$. A third example is given by taking $L$ to be the class of all functions integrable with respect to a measure $\mu$.

This last example does not quite satisfy the definition we have given, since integrable functions may be extended real-valued. We want to extend our definition to cover this case but must be careful about the fact that $f + g$ is not defined at points where it is of the form $\infty - \infty$ or $-\infty + \infty$. Thus we shall say that $L$ is a *vector lattice* of extended real-valued functions on $X$ provided that, whenever $f$ and $g$ are in $L$, the functions $f \vee g$, $f \wedge g$, and $\alpha f$ are in $L$ as well as each function $h$ such that $h(x) = f(x) + g(x)$ whenever the right side is defined. Thus we require all possible choices for the sum $f + g$ to be in $L$, and we shall write $h = f + g$ whenever $h$ is any possible choice. By a linear functional on $L$ we mean a mapping $I$ of $L$ into $\mathbf{R}$ such that $I(\alpha f) = \alpha I(f)$ and $I(h) = I(f) + I(g)$ whenever $h = f + g$.

[1] Properly speaking, we should call $I$ "nonnegative," but the use of "positive" in this connection seems to be standard.

[2] Here, of course, lim means pointwise limit.

This implies that if $h_1$ and $h_2$ are two functions with $h_1 = f + g$ and $h_2 = f + g$ then $I(h_1) = I(h_2)$. A positive linear functional $I$ on $L$ is called a *Daniell integral* if it satisfies condition D.

The fact that we allow all possible choices of $f(x) + g(x)$ at points $x$ where the sum is ambiguous has some consequences for the sets $P$ on which an element of a vector lattice of extended real-valued functions may be infinite. Thus we call a subset $P$ of $X$ a *polar set* for the lattice $L$ if there is a function $h \in L$ with $|h(x)| = \infty$ for all $x \in P$.

**1. Lemma:** *Let $P$ be a polar set, $f \in L$ and $g$ an extended real-valued function such that $g(x) = f(x)$ for $x \notin P$. Then $g \in L$ and $I(g) = I(f)$.*

***Proof:*** Let $h$ be an element of $L$ with $|h(x)| = \infty$ for $x \in P$. Then

$$h(x) - h(x) = g(x) - f(x)$$

wherever the left-hand side is defined, provided that we take $g(x) - f(x) = 0$ in all ambiguous cases. Thus $g - f$ belongs to $L$ and

$$I(g - f) = I(h) - I(h) = 0.$$

Hence

$$g(x) = f(x) + [g(x) - f(x)]$$

wherever the right-hand side is defined. Thus $g \in L$, and

$$I(g) = I(f) + I(g - f) = I(f). \quad \blacksquare$$

Using Lemma 1 to replace functions in $L$ with extended real values by functions with real values, we can establish the following proposition.

**2. Proposition:** *Let $L$ be a vector lattice of extended real-valued functions on $X$. Then conditions D, D′, and D″ are equivalent.*

In general the limits of functions in $L$ will not be in $L$, and the procedure of Daniell and Stone, which we are going to develop in this chapter, consists in extending $I$ to a class $L_1$ containing $L$ and closed under suitable limiting operations. In many ways this extension procedure parallels that of Carathéodory for measures, and in Section 3 we show that under mild restrictions this extended integral is in fact integration with respect to a suitable measure.

## Problems

1. Show that the conditions D, D′, and D″ are equivalent.

2. Let $\mu$ be a measure on an algebra $\mathcal{A}$, $L$ the class of simple functions on $\mathcal{A}$, and $I$ integration with respect to $\mu$. Show that $L$ is a vector lattice and that $I$ is a Daniell integral.

3. Let $L$ be the family of continuous real-valued functions on $(-\infty, \infty)$ each of which vanishes outside a finite interval, and let $I$ be Riemann integration on $L$. Show that $I$ is a Daniell integral. [Hint: The Dini Theorem is useful.]

4. Prove Proposition 2.

## 2 The Extension Theorem

We begin by considering the class $L_u$ consisting of all those extended real-valued function on $X$ each of which is a limit of a monotone increasing sequence of functions in $L$. It is clear that if $f$ and $g$ are in $L_u$, then so is the function $\alpha f + \beta g$ where $\alpha$ and $\beta$ are nonnegative constants. If $\langle \varphi_n \rangle$ is an increasing sequence of functions from $L$, then $\langle I(\varphi_n) \rangle$ must be an increasing sequence of real numbers and so must have a limit (which may be $\infty$). It is tempting to try to define $\lim I(\varphi_n)$ as the value of $I$ for the function $f$ which is the pointwise limit of the sequence $\langle \varphi_n \rangle$. In order to do this we need to know that this value depends only on the function $f$ and not on the choice of the increasing sequence $\langle \varphi_n \rangle$ whose limit is $f$. That this is so is guaranteed by the following lemma:

**3. Lemma:** *If $\langle \varphi_n \rangle$ and $\langle \psi_m \rangle$ are increasing sequences from $L$ and if $\lim \varphi_n \leq \lim \psi_m$, then $\lim I(\varphi_n) \leq \lim I(\psi_m)$.*

***Proof:*** For fixed $n$, we have $\varphi_n \leq \lim \varphi_n \leq \lim \psi_m$, and so $I(\varphi_n) \leq \lim I(\psi_m)$ by (D′). Hence $\lim I(\varphi_n) \leq \lim I(\psi_m)$. ∎

Thus we can extend the functional $I$ to be an extended real-valued functional on $L_u$ with the property that $I(f) \leq I(g)$ for $f \leq g$ and $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for positive constants $\alpha$ and $\beta$ and functions $f$ and $g$ in $L_u$. It is clear that $L_u$ is a lattice, for if $\varphi_n \uparrow f$ and $\psi_n \uparrow g$, then $\varphi_n \wedge \psi_n \uparrow f \wedge g$ and $\varphi_n \vee \psi_n \uparrow f \vee g$.

**4. Lemma:** *A nonnegative function $f$ belongs to $L_u$ if and only if there is a sequence $\langle \psi_\nu \rangle$ of nonnegative functions in $L$ such that $f = \sum_{\nu=1}^{\infty} \psi_\nu$. In this case $I(f) = \sum_{\nu=1}^{\infty} I(\psi_\nu)$.*

***Proof:*** The "if" part is trivial. On the other hand, let $f$ be nonnegative and $\varphi_n \uparrow f$ with $\varphi_n \,\varepsilon\, L$. By replacing $\varphi_n$ by $\varphi_n \vee 0$, we may assume that each $\varphi_n$ is nonnegative. Set $\psi_1 = \varphi_1$, $\psi_n = \varphi_n - \varphi_{n-1}$ for $n > 1$. Then $f = \sum_{\nu=1}^{\infty} \psi_\nu$, and

$$
\begin{aligned}
I(f) &= \lim I(\varphi_n) \\
&= \lim I\left(\sum_{\nu=1}^{n} \psi_\nu\right) \\
&= \lim \sum_{\nu=1}^{n} I(\psi_\nu) \\
&= \sum_{\nu=1}^{\infty} I(\psi_\nu). \quad \blacksquare
\end{aligned}
$$

**5. Lemma:** *Let $\langle f_n \rangle$ be a sequence of nonnegative functions in $L_u$. Then the function $f = \sum_{n=1}^{\infty} f_n$ is in $L_u$ and $I(f) = \sum_{n=1}^{\infty} I(f_n)$.*

***Proof:*** For each $n$ there is a sequence $\langle \psi_{n,\nu} \rangle$ of nonnegative functions in $L$ such that $f_n = \sum_{\nu=1}^{\infty} \psi_{n,\nu}$. Hence $f = \sum_{\nu,n} \psi_{n,\nu}$. Since the set of pairs of integers is countable, $f$ is the sum of a sequence of nonnegative functions in $L$ and so must be in $L_u$. Also

$$
\begin{aligned}
I(f) &= \sum_{\nu,n} I(\psi_{n,\nu}) \\
&= \sum_{n=1}^{\infty} I(f_n). \quad \blacksquare
\end{aligned}
$$

For an arbitrary function $f$ on $X$ we define the upper integral $\bar{I}(f)$ by setting

$$\bar{I}(f) = \inf_{\substack{g \geq f \\ g \,\varepsilon\, L_u}} I(g),$$

where we adopt the convention that the infimum of the empty set is $+\infty$. We define the lower integral $\underline{I}$ by setting $\underline{I}(f) = -\bar{I}(-f)$. Elementary properties of these upper and lower integrals are given by the following lemmas. The properties in Lemma 6 follow directly from the definition of $\bar{I}$.

**6. Lemma:** *Let $h = f + g$. Then $\bar{I}(h) \le \bar{I}(f) + \bar{I}(g)$, provided that the right-hand side is defined. If $c \ge 0$, $\bar{I}(cf) = c\bar{I}(f)$. If $f \le g$, then $\bar{I}(f) \le \bar{I}(g)$ and $\underline{I}(f) \le \underline{I}(g)$.*

**7. Lemma:** *We have $\underline{I}(f) \le \bar{I}(f)$. If $f \varepsilon L_u$, then $\underline{I}(f) = \bar{I}(f) = I(f)$.*

***Proof:*** We have $0 = I(0) = I(f - f) \le \bar{I}(f) + \bar{I}(-f)$. Hence $\underline{I}(f) = -\bar{I}(-f) \le \bar{I}(f)$.

To prove the second statement, we note that if $f \varepsilon L_u$ then $\bar{I}(f) \le I(f)$ by the definition of $\bar{I}$. If $g \varepsilon L_u$ and $f \le g$, then $I(f) \le I(g)$ and so $I(f) \le \bar{I}(f)$, whence $\bar{I}(f) = I(f)$. If $\varphi \varepsilon L$, then $-\varphi \varepsilon L \subset L_u$, and so $\bar{I}(-\varphi) = I(-\varphi) = -I(\varphi)$, and hence $\underline{I}(\varphi) = I(\varphi)$. But each $f \varepsilon L_u$ is the limit of an increasing sequence $\langle \varphi_\nu \rangle$ of functions in $L$. Since $f \ge \varphi_\nu$, we have $\underline{I}(f) \ge \underline{I}(\varphi_\nu) = I(\varphi_\nu)$. Hence

$$\underline{I}(f) \ge \lim I(\varphi_\nu) = I(f).$$

Since $\underline{I}(f) \le \bar{I}(f) = I(f)$, we have $\underline{I}(f) = I(f)$. ▌

**8. Lemma:** *Let $\langle f_\nu \rangle$ be a sequence of nonnegative functions, and let $f = \sum_{\nu=1}^{\infty} f_\nu$. Then $\bar{I}(f) \le \sum_{\nu=1}^{\infty} \bar{I}(f_\nu)$.*

***Proof:*** If $\bar{I}(f_\nu) = \infty$ for some $\nu$, we are done. If not, given $\epsilon > 0$, there is a function $g_\nu \varepsilon L_u$ such that $f_\nu \le g_\nu$ and

$$I(g_\nu) \le \bar{I}(f_\nu) + \epsilon \cdot 2^{-\nu}.$$

Since each $g_\nu$ is nonnegative, Lemma 5 implies that the function $g = \sum g_\nu$ is in $L_u$ and that $I(g) = \sum I(g_\nu) \le \sum \bar{I}(f_\nu) + \epsilon$. Since $g \ge f$, we have

$$\bar{I}(f) \le \sum_{\nu=1}^{\infty} \bar{I}(f_\nu) + \epsilon,$$

and the lemma follows since $\epsilon$ was an arbitrary positive number. ▌

We shall call a function $f$ on $X$ **integrable** with respect to $I$ (or $I$-integrable) if $\bar{I}(f) = \underline{I}(f)$ and this value is finite. We denote the class of functions integrable with respect to $I$ by $L_1$. For $f$ in $L_1$ we shall write $I(f)$ for $\bar{I}(f)$. Thus we have an extension of our original $I$ to all of $L_1$. Properties of $L_1$ and this extended $I$ are given by the following proposition:

**9. Proposition:** *The set $L_1$ is a vector lattice of functions containing $L$, and $I$ is a positive linear functional on $L_1$, which extends the functional $I$ on $L$.*

***Proof:*** If $f$ is in $L_1$, so is $cf$, since $\bar{I}(cf) = c\bar{I}(f) = c\underline{I}(f) = \underline{I}(cf)$ for $c \geq 0$, and $\bar{I}(cf) = c\underline{I}(f) = c\bar{I}(f) = \underline{I}(cf)$ for $c \leq 0$. If $f$ and $g$ are in $L_1$, then

$$\bar{I}(f + g) \leq I(f) + I(g)$$

and

$$-\underline{I}(f + g) = \bar{I}(-f - g) \leq -I(f) - I(g),$$

that is

$$\underline{I}(f + g) \geq I(f) + I(g).$$

Thus

$$\underline{I}(f + g) = \bar{I}(f + g) = I(f) + I(g),$$

and so $f + g$ is in $L_1$. Consequently, $L_1$ is a linear space, and $I$ is a linear functional on $L_1$. Lemma 7 implies that $L_1 \supset L$ and that our definition of $I$ on $L_1$ gives us a positive linear functional which agrees with our original $I$ on $L_1$.

To prove that $L_1$ is a lattice, it suffices to show that if $f \in L_1$, then $f^+ \in L_1$. Let $f \in L_1$. Then for each $\epsilon > 0$ there are functions $g$ and $h$ in $L_u$ such that $-h \leq f \leq g$, while also $I(g) < I(f) + \epsilon < \infty$ and $I(h) \leq -I(f) + \epsilon < \infty$. Since $g = (g \vee 0) + (g \wedge 0)$ and $(g \wedge 0) \in L_u$, we have $I(g \wedge 0) > -\infty$ and $I(g \vee 0) \leq I(g) - I(g \wedge 0) < \infty$. Thus the function $g_1 = g \vee 0$ is in $L_u$ and $I(g_1) < \infty$. Let $h_1 = h \wedge 0$. Then $h_1 \in L_u$, and $-h_1 \leq f^+ \leq g_1$. Since $g \geq -h$, $g_1 + h_1 \leq g + h$.

Consequently, $I(g_1) + I(h_1) \leq I(g) + I(h) < 2\epsilon$. Since $-I(h_1) \leq \underline{I}(f^+) \leq \bar{I}(f^+) \leq I(g_1)$, we have $\bar{I}(f^+) - \underline{I}(f^+) < 2\epsilon$, and so $\bar{I}(f^+) = \underline{I}(f^+)$, since $\epsilon$ was arbitrary. Since $0 \leq \bar{I}(f^+) \leq I(g_1) < \infty$, we have $f^+ \in L_1$. Thus $L_1$ is a lattice. ∎

The following proposition is the analogue for $L_1$ of the monotone convergence theorem. It also shows that $L_1$ and $I$ satisfy condition D′ and hence D.

**10. Proposition:** *Let $\langle f_n \rangle$ be an increasing sequence of functions in $L_1$, and let $f = \lim f_n$. Then $f \in L_1$ if and only if $\lim I(f_n) < \infty$. In this case $I(f) = \lim I(f_n)$.*

***Proof:*** Since $f \geq f_n$, $\bar{I}(f) \geq I(f_n)$. Thus if $\lim I(f_n) = \infty$, then $\bar{I}(f) = \infty$, and $f \notin L_1$.

Suppose $\lim I(f_n) < \infty$. Set $g = f - f_1$. Then $g \geq 0$, and $g = \sum_{n=1}^{\infty} (f_{n+1} - f_n)$. Hence by Lemma 8,

$$\begin{aligned}\bar{I}(g) &\leq \sum_{n=1}^{\infty} I(f_{n+1} - f_n)\\ &= \sum_{n=1}^{\infty} I(f_{n+1}) - I(f_n)\\ &= \lim I(f_n) - I(f_1).\end{aligned}$$

Thus

$$\begin{aligned}\bar{I}(f) &= \bar{I}(f_1 + g)\\ &\leq I(f_1) + \bar{I}(g) \leq \lim I(f_n).\end{aligned}$$

Since $f_n \leq f$, we have $\underline{I}(f) \geq I(f_n)$, and so

$$\underline{I}(f) \geq \lim I(f_n).$$

Thus $\underline{I}(f) = \bar{I}(f) = \lim I(f_n)$. ▮

**11. Corollary:** *The functional $I$ is a Daniell integral on the vector lattice $L_1$.*

The following two propositions are the analogues for the integral $I$ of Fatou's lemma and the Lebesgue convergence theorem.

**12. Proposition:** *Let $\langle f_\nu \rangle$ be a sequence of nonnegative functions in $L_1$. Then the function* $\inf f_\nu$ *is in $L_1$, and the function* $\underline{\lim} f_\nu$ *is in $L_1$ if* $\underline{\lim}\, I(f_\nu) < \infty$. *In this case*

$$I(\underline{\lim} f_\nu) \leq \underline{\lim}\, I(f_\nu).$$

***Proof:*** Let $g_n = f_1 \wedge f_2 \wedge \cdots \wedge f_n$. Then $\langle g_n \rangle$ is a sequence of nonnegative functions in $L_1$, which decrease to $g = \inf f_\nu$. Thus

$-g_n \uparrow -g$, and since $I(-g_n) \le 0$, we must have $g \in L_1$ by Proposition *10*.

To prove the rest of the proposition, let $h_n = \inf_{\nu \ge n} f_\nu$. Then $\langle h_n \rangle$ is a sequence of nonnegative functions in $L_1$ which increases to $\underline{\lim}\, f_\nu$. Since $h_n \le f_\nu$ for $n \le \nu$, $\lim I(h_n) \le \underline{\lim}\, I(f_\nu) < \infty$. Hence $\underline{\lim}\, f_\nu \in L_1$ and $I(\underline{\lim}\, f_\nu) \le \underline{\lim}\, I(f_\nu)$ by Proposition 10. ▮

**13. Proposition:** *Let $\langle f_n \rangle$ be a sequence of functions in $L_1$ and suppose that there is a function $g$ in $L_1$ such that for all $n$ we have $|f_n| \le g$. Then if $f = \lim f_n$, we have*

$$I(f) = \lim I(f_n).$$

***Proof:*** The functions $f_n + g$ are nonnegative, and $I(f_n + g) \le 2I(g)$. Hence by Proposition 12 we have $f + g$ in $L_1$ and

$$I(f+g) \le \underline{\lim}\, I(f_n + g) = I(g) + \underline{\lim}\, I(f_n).$$

Hence

$$I(f) \le \underline{\lim}\, I(f_n).$$

Since the functions $g - f_n$ are also nonnegative, we have

$$\begin{aligned} I(g-f) &\le \underline{\lim}\, I(g - f_n) \\ &= I(g) - \overline{\lim}\, I(f_n). \end{aligned}$$

Hence

$$\overline{\lim}\, I(f_n) \le I(f),$$

and so $\lim I(f_n)$ exists and is equal to $I(f)$. ▮

## Problems

5. A function $f$ belongs to $L_u$ if and only if $f = g + \varphi$ with $g \in L_u$, $g \ge 0$, and $\varphi \in L$.

6. Prove Lemma 6.

7. Show that the ambiguous case in Lemma 6 does not cause difficulties in Lemma 7.

## 3 Uniqueness

In the present section we show that the extension to $L_1$ of a Daniell integral $I$ on $L$ is unique. We begin by proving a proposition of some interest in its own right which describes the structure of functions in $L_1$. It is the analogue for $I$ of Proposition 12.7.

Let us denote by $L_{ul}$ the class of those functions on $X$ which are the limit of a decreasing sequence $\langle f_n \rangle$ of functions in $L_u$ with $I(f_n) < \infty$ and $\lim I(f_n) > -\infty$. It follows from Proposition 10 applied to $\langle -f_n \rangle$ that $L_{ul} \subset L_1$. If $f$ is any function on $X$ such that $\bar{I}(f)$ is finite, then given $n$ we can find $h_n \varepsilon L_u$ such that

$$f \leq h_n \quad \text{and} \quad I(h_n) \leq \bar{I}(f) + \frac{1}{n}.$$

Setting $g_n = h_1 \wedge h_2 \wedge \cdots \wedge h_n$, we have $f \leq g_n \leq h_n$, and so $\langle g_n \rangle$ is a decreasing sequence of functions in $L_u$ with $\bar{I}(f) \leq I(g_n) \leq \bar{I}(f) + 1/n$. Hence the function $g = \lim g_n$ is in $L_{ul}$, while $f \leq g$ and $\bar{I}(f) = I(g)$. We have thus established the following lemma:

**14. Lemma:** *If $f$ is any function on $X$ with $\bar{I}(f)$ finite, then there is a function $g \varepsilon L_{ul}$ such that $f \leq g$ and $\bar{I}(f) = I(g)$.*

A function $f$ on $X$ is called a **null function** if $f \varepsilon L_1$ and $I(|f|) = 0$. If $f$ is a null function and $|g| \leq f$, $0 \leq \underline{I}(|g|) \leq \bar{I}(|g|) \leq I(f) = 0$. Hence $g \varepsilon L_1$, and $g$ is a null function.

**15. Proposition:** *A function $f$ on $X$ is in $L_1$ if and only if $f$ is the difference $g - h$ of a function $g$ in $L_{ul}$ and a nonnegative null function $h$. A function $h$ is a null function if and only if there is a null function $k$ in $L_{ul}$ such that $|h| \leq k$.*

***Proof:*** If $f = g - h$, then $f$ is the difference of two functions in $L_1$ and so must itself be in $L_1$. If $|h| \leq k$ with $k$ null, then $h$ is a null function.

If $f$ is in $L_1$, then Lemma 14 asserts the existence of $g \varepsilon L_{ul}$ such that $f \leq g$ and $I(f) = I(g)$. Hence $h = g - f$ is a nonnegative function and $I(h) = 0$, making $h$ a null function. If $h$ is a null function, then by Lemma 14 there is a function $k \varepsilon L_{ul}$ with $|h| \leq k$ and $I(k) = I(|h|) = 0$. ▮

**16. Proposition:** *Let $I$ be a Daniell integral on a vector lattice $L$ of functions on $X$ and let $J$ be a Daniell integral on a vector lattice $\Lambda \supset L$. If $I(f) = J(f)$ for all $f \varepsilon L$, then $\Lambda_1 \supset L_1$ and $I(f) = J(f)$ for all $f \varepsilon L_1$.*

***Proof:*** By applying Proposition 10 twice we see that $L_{ul} \subset \Lambda_1$ and that $I(f) = J(f)$ for $f \in L_{ul}$. Hence by the second part of Proposition 15 each function which is null with respect to $I$ must also be null with respect to $J$. By the first part of Proposition 15 every function $f$ in $L_1$ must be in $\Lambda_1$, and $I(f) = J(f)$. ▮

## 4 Measurability and Measure

We say that a nonnegative function $f$ on $X$ is **measurable** (with respect to $I$) if $g \wedge f$ is in $L_1$ for each $g$ in $L_1$.

**17. Lemma:** *If $f$ and $g$ are nonnegative measurable functions, so are $f \wedge g$ and $f \vee g$. If $\langle f_n \rangle$ is a sequence of nonnegative measurable functions which converge pointwise to a function $f$, then $f$ is measurable.*

***Proof:*** If $f$ and $g$ are nonnegative measurable functions and $h$ is in $L_1$, then $h \wedge (f \wedge g) = (h \wedge f) \wedge (h \wedge g)$ and $h \wedge (f \vee g) = (h \wedge f) \vee (h \wedge g)$. Hence the measurability of $f \wedge g$ and $f \vee g$ follows from the fact that $L_1$ is a lattice. If $\langle f_n \rangle$ is a sequence of nonnegative measurable functions converging to $f$ and $g$ a function in $L_1$, then $\langle f_n \wedge g \rangle$ is a sequence of functions in $L_1$ converging to $f \wedge g$. Since $|f_n \wedge g| \leq |g|$, we have $f \wedge g$ in $L_1$ by Proposition 13. ▮

**18. Lemma:** *A nonnegative function $f$ on $X$ is measurable with respect to $I$ if $\varphi \wedge f$ is in $L_1$ for each $\varphi$ in $L$.*

***Proof:*** If $\varphi \in L$, then $\varphi \wedge f$ is in $L_1$, and Proposition 10 implies that $g \wedge f \in L_1$ for $g \in L_u$ and $I(g) < \infty$. It follows from Proposition 13 that $g \wedge f \in L_1$ for all $g \in L_{ul}$. If $h$ is any function in $L_1$, Proposition 15 tells us that $h = g - k$, where $g \in L_{ul}$ and $k$ is a nonnegative null function. Since $0 \leq g \wedge f - h \wedge f \leq k$, the function $h \wedge f$ differs from the integrable function $g \wedge f$ by a null function. Thus $h \wedge f$ is integrable, proving that $f$ is measurable. ▮

We say that a set $A$ in $X$ is **measurable** with respect to $I$ if its characteristic function $\chi_A$ is measurable. We say that $A$ is **integrable** if its characteristic function $\chi_A$ is integrable. Note that a measurable subset of an integrable set is itself integrable.

**19. Lemma:** *If $A$ and $B$ are measurable sets, so are the sets $A \cup B$, $A \cap B$, and $A \sim B$. If $\langle A_n \rangle$ is a sequence of measurable sets, then the* *sets $\bigcap A_n$ and $\bigcup A_n$ are measurable. If the function* 1 *is measurable, then the class $\mathfrak{A}$ of measurable sets is a $\sigma$-algebra.*

***Proof:*** The measurability of $A \cap B$ and $A \cup B$ follows from the fact that $\chi_{A \cap B} = \chi_A \wedge \chi_B$ and $\chi_{A \cup B} = \chi_A \vee \chi_B$. If $g$ is in $L_1$, we have $g \wedge \chi_{A \sim B} = g \wedge \chi_A - g \wedge \chi_{A \cap B} + g \wedge 0$, and the measurability of $A \sim B$ follows from that of $A$ and $B$. If $A = \bigcup A_n$, then

$$\chi_A = \lim (\chi_{A_1} \vee \cdots \vee \chi_{A_n})$$

and the measurability of $A$ follows from Lemma 17. A similar argument holds for $\bigcap A_n$. If 1 is a measurable function, then the set $X$ is a measurable set, and the complement of a measurable set is measurable. ∎

**20. Lemma:** *If* 1 *is a measurable function and $f$ a nonnegative integrable function, then for each real number $\alpha$ the set*

$$E = \{x : f(x) > \alpha\}$$

*is measurable.*

***Proof:*** If $\alpha$ is negative, $E = X$ and is measurable, since $\chi_E = 1$ and 1 is measurable. Hence we assume that $\alpha \geq 0$. If $\alpha = 0$, set $g = f$, while if $\alpha > 0$, let $g = (\alpha^{-1}f) - [(\alpha^{-1}f) \wedge 1]$. Since $g$ is the difference of two functions in $L_1$, $g$ is in $L_1$. In either case we have $g(x) > 0$ for $x \varepsilon E$, and $g(x) = 0$ for $x \varepsilon \tilde{E}$. Let $\varphi_n = 1 \wedge (ng)$. Then $\varphi_n \varepsilon L_1$, and $\varphi_n \uparrow \chi_E$. Hence $\chi_E$ is measurable, and so $E$ is measurable. ∎

**21. Lemma:** *Let the function* 1 *be measurable, and define a set function $\mu$ on the class $\mathfrak{A}$ of measurable sets by*

$$\mu E = I(\chi_E)$$

*if $\chi_E$ is integrable, and $\mu E = \infty$ otherwise. Then $\mu$ is a measure.*

***Proof:*** We have $\mu\emptyset = I(0) = 0$. If $A$ and $B$ are integrable sets with $A \subset B$, we have $\chi_A \leq \chi_B$, and so $\mu A \leq \mu B$. Thus $\mu$ is monotone for integrable sets and consequently for measurable sets.

Let $\langle E_i \rangle$ be a disjoint sequence of measurable sets, and let $E = \bigcup E_i$. If one of the $E_i$ is not integrable, then $E$ is not integrable, and

$$\mu E = \infty = \sum \mu E_i.$$

If each $E_i$ is integrable, $E$ will be integrable if and only if $\sum \mu E_i < \infty$ by Proposition 10, since $\chi_E = \sum \chi_{E_i}$. In either case $\mu E = \sum \mu E_i$, and the measure $\mu$ is countably additive. ▮

This measure $\mu$ has the property that the integrable sets are precisely the measurable sets of finite measure. The following theorem tells us that the Daniell integral $I$ on $L_1$ is equivalent to the integral with respect to this measure $\mu$.

**22. Theorem (Stone):** *Let $L$ be a vector lattice of functions on $X$ with the property that if $f \in L$ then $1 \wedge f \in L$, and let $I$ be a Daniell integral on $L$. Then there is a $\sigma$-algebra $\mathfrak{A}$ of subsets of $X$ and a measure $\mu$ on $\mathfrak{A}$ such that each function $f$ on $X$ is integrable with respect to $I$ if and only if it is integrable with respect to $\mu$. Moreover,*

$$I(f) = \int f\,d\mu.$$

***Proof:*** Let $\mathfrak{A}$ be the class of sets which are measurable with respect to $I$. It follows from Lemma 18 that 1 is measurable. Lemma 19 then asserts that $\mathfrak{A}$ is a $\sigma$-algebra, and Lemma 20 asserts that each nonnegative $I$-integrable function is measurable with respect to $\mathfrak{A}$. Since each $I$-integrable function is the difference of two nonnegative $I$-integrable functions, every $I$-integrable function must be measurable with respect to $\mathfrak{A}$.

Let $\mu$ be the measure given in Lemma 21, and let $f$ be a nonnegative function which is integrable with respect to $I$. For each pair $\langle k, n\rangle$ of positive integers let

$$E_{k,n} = \{x : f(x) > k2^{-n}\}.$$

Then $E_{k,n}$ is measurable, and since

$$\chi_{E_{k,n}} = \chi_{E_{k,n}} \wedge (k^{-1}2^n f),$$

we have $\chi_{E_{k,n}} \in L_1$, and $\mu(E_{k,n}) < \infty$. Set

$$\varphi_n = 2^{-n} \sum_{k=1}^{2^{2n}} \chi_{E_{k,n}}.$$

Then $\varphi_n \in L_1$ and $\varphi_n \uparrow f$. Hence $I(f) = \lim I(\varphi_n)$. But

$$I(\varphi_n) = 2^{-n} \sum_{k=1}^{2^{2n}} I(\chi_{E_{k,n}})$$

$$= 2^{-n} \sum_{k=1}^{2^{2n}} \mu(E_{k,n})$$

$$= \int \varphi_n \, d\mu.$$

Since

$$\int f \, d\mu = \lim \int \varphi_n \, d\mu$$

by the Monotone Convergence Theorem, we have

$$I(f) = \int f \, d\mu,$$

and $f$ is integrable with respect to $\mu$. Since an arbitrary $f$ which is $I$-integrable is the difference of two nonnegative $I$-integrable functions, it follows that such an $f$ must also be integrable with respect to $\mu$ and that

$$I(f) = \int f \, d\mu.$$

If $f$ is a nonnegative function on $X$ which is integrable with respect to $\mu$, we construct $E_{k,n}$ and $\varphi_n$ as before. Since $\int f \, d\mu < \infty$, each $E_{k,n}$ has finite measure, and so $\chi_{E_{k,n}}$ and hence $\varphi_n$ belong to $L_1$. Since $\varphi_n \uparrow f$ and $\lim I(\varphi_n) = \int f \, d\mu < \infty$, we have $f \in L_1$ by Proposition 10. Thus each $f$ which is integrable with respect to $\mu$ is also integrable with respect to $I$. ▮

**23. Proposition:** *Let $L$ be a vector lattice of functions on a set $X$, and suppose that $1 \in L$. Let $\mathcal{B}$ be the smallest $\sigma$-algebra of subsets of $X$ such that each function in $L$ is measurable with respect to $\mathcal{B}$. Then for each Daniell integral $I$ there is a unique measure $\mu$ on $\mathcal{B}$ such that for every $f \in L$*

$$I(f) = \int f \, d\mu.$$

***Proof:*** The existence of $\mu$ is a special case of Theorem 22 and we have only to prove the uniqueness of $\mu$ on $\mathcal{B}$. Let $\mathcal{A}$ be the $\sigma$-algebra of measurable sets given by Lemma 19. Lemma 20 asserts that each $f$ in $L$ is measurable with respect to $\mathfrak{A}$, and so we must have $\mathfrak{B} \subset \mathfrak{A}$. Since $1 \in L$, the functions $\chi_B$ are in $L_1$ for each $B$ in $\mathfrak{A}$ and hence for each $B$ in $\mathfrak{B}$. The uniqueness of $\mu$ on $\mathfrak{B}$ will be established if we can show that $\mu(B) = I(\chi_B)$ for each $B$ in $\mathfrak{B}$.

If we let $\Lambda$ be the set of functions on $X$ which are measurable with respect to $\mathfrak{B}$ and integrable with respect to $\mu$ and set

$$J(f) = \int f\, d\mu$$

for $f \in \Lambda$, then Proposition 16 implies that $J(f) = I(f)$ for $f \in L_1 \cap \Lambda$. But if $B \in \mathfrak{B}$, then $\chi_B \in L_1 \cap \Lambda$, and so

$$\begin{aligned} \mu B &= J(\chi_B) \\ &= I(\chi_B). \end{aligned}$$

Thus $\mu$ is uniquely determined on $\mathfrak{B}$ by $I$. ▮

We can still establish the uniqueness of the measure $\mu$ in this proposition if instead of assuming $1 \in L$, we make the weaker assumption that there is an everywhere positive function in $L_1$ (Problem 10). Without some such assumption the measure $\mu$ need not be unique on $\mathfrak{B}$ (Problem 11).

## Problems

**8.** Let $\mu$ be a measure on an algebra $\mathfrak{A}$ of sets, and let $L$ be the family consisting of those functions which are finite linear combinations of characteristic functions of sets in $\mathfrak{A}$ with finite measure, and let $I$ be integration with respect to $\mu$. Discuss the extension of $I$ to $L_1$, and compare this process with the Carathéodory extension for $\mu$.

**9.** Prove directly from the definition that, if $f_1$ and $f_2$ are two nonnegative measurable functions, then $f_1 + f_2$ is measurable.

**10.** Prove that the conclusion of Proposition 23 still holds when the hypothesis that $1 \in L$ is replaced by the hypothesis that there is an everywhere positive function $e$ in $L_1$. [Hint: $X = \bigcup_{n=1}^{\infty} \{x : e(x) > 1/n\}$, and the proof given can be modified to show that $\mu$ is unique on sets of $\mathfrak{B}$ which are contained in $\{x : e(x) > 1/n\}$.]

**11.** Let $X = (-\infty, \infty) \cup \{\omega\}$, and let $L$ consist of all functions on $X$ which are Lebesgue integrable on $(-\infty, \infty)$ and vanish at $\omega$. Then $L$ is a vector lattice, and the smallest $\sigma$-algebra with respect to which each function in $L$ is measurable is the family $\mathcal{B}$ consisting of all sets $B$ such that $B \cap (-\infty, \infty)$ is Lebesgue measurable. Let $I$ be defined on $L$ by $I(f) = \int f(x)\,dx$. Then $I$ is a Daniell integral on $L$. What is the measure $\mu$ constructed in Theorem 22? Show that there is another measure $\nu$ defined on $\mathcal{B}$ such that for each $f$ in $L$, $I(f) = \int f\,d\nu$.

12. a. Define a measure $\mu$ on the $I$-measurable sets by $\mu E = I(\chi_E)$ if $E$ integrable, and $I(\chi_E) = \sup \{I(\chi_A): A \subset E,\ A \text{ integrable}\}$ otherwise. Show that $\mu$ is a measure and is the smallest measure such that $I(f) = \int f\,d\mu$.

b. Show that for this measure $\mu(X) = \|I\| = \sup \{I(f): f \in L, f \leq 1\}$.

c. Show that if $\|I\| < \infty$, then this measure $\mu$ is the unique measure such that $I(f) = \int f\,d\mu$ and $\mu(X) = \|I\|$.

# Bibliography

## A. Textbooks and Standard References

[1] S. Banach, *Théorie des Opérations Linéaires*, 2nd ed., Reprint, New York, Chelsea Publishing Co.

[2] G. Birkhoff and S. MacLane, *Survey of Modern Algebra*, 4th ed., New York, Macmillan, 1977.

[3] M. M. Day, *Normed Linear Spaces*, 3rd ed. (Ergebnisse der Mathematik und ihrer Grenzgebiete, No. 21), Springer-Verlag, 1973.

[4] N. Dunford and J. T. Schwartz, *Linear Operators*, Part I, New York, Wiley-Interscience, 1958.

[5] P. R. Halmos, *Measure Theory*, Reprint, Springer-Verlag, 1974.

[6] P. R. Halmos, *Naive Set Theory*, Reprint, Springer-Verlag, 1974.

[7] T. Hawkins, *Lebesgue's Theory of Integration, Its Origins and Development*, Reprint, New York, Chelsea Publishing Co., 1975.

[8] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups*, Revised ed. (Colloquium Publications, Vol. 31), Providence, American Mathematical Society, 1957.

[9] J. L. Kelley, *General Topology*, Reprint, Springer-Verlag, 1975.

[10] J. L. Kelley, I. Namioka, et al., *Linear Topological Spaces*, New York, Van Nostrand, 1963.

[11] K. Kuratowski, *Topology*, Vol. 1, New York, Academic Press, 1966.

[12] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, 1986.

[13] S. Saks, *Theory of the Integral* (Monografje Matematyczne, Vol. 7), Warsaw, 1937.

[14] P. C. SUPPES, *Axiomatic Set Theory*, Reprint, New York, Dover Publishing Co., 1972.

## B. Other References

[15] N. BOURBAKI, *Topologie Général*, Paris, Hermann et Cie., 1971.

[16] C. CARATHÉODORY, *Vorlesungen über reelle Funktionen*, Reprint, New York, Chelsea Publishing Co., 1946.

[17] C. CARATHÉODORY, *Algebraic Theory of Measure and Integration*, Reprint, New York, Chelsea Publishing Co., 1963.

[18] A. HAAR, *Der Massbegriff in der Theorie der kontinuierlichen Gruppen*, Ann. of Math. (2) 34 (1933) 147–169.

[19] G. H. HARDY, J. E. LITTLEWOOD, and G. PÓLYA, *Inequalities*, New York, Cambridge University Press, 1934.

[20] A. HURWITZ, *Über die Erzeugung der Invarianten durch Integration*, Gött. Nachr. (1897) 71–90.

[21] J. E. LITTLEWOOD, *Lectures on the Theory of Functions*, Oxford, 1944.

[22] L. H. LOOMIS, *An Introduction to Abstract Harmonic Analysis*, New York, Van Nostrand, 1953.

[23] A. MARKOV, *On mean values and exterior densities*, Mat. Sbornik N.S. 4 (46) (1938) 165–191.

[24] F. RIESZ, *Sur les opérations fonctionnelles linéaires*, C.R. Acad. Sci. Paris 149 (1909) 974–977.

[25] F. RIESZ and B. NAGY, *Functional Analysis*, English ed., New York, Ungar, 1956.

[26] M. E. RUDIN, *A new proof that metric spaces are paracompact*, Proc. Amer. Math. Soc. **20** (1969) 603.

[27] A. WEIL, *L'intégration sur les groupes topologiques et ses applications*, Paris, Hermann et Cie., 1940.

# Index of Symbols

# Subject Index



**Note:** Numbers in parentheses are problem numbers.

D



E



F

## G



## H



## I



## J



## K



## L



## M

## U



## V



## W



## Z