

Real Analysis

SECOND EDITION

H. L. ROYDEN

H. L. Royden STANFORD UNIVERSITY

Real Analysis

SECOND EDITION

The Macmillan Company

Collier-Macmillan Limited, London

Preface

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.

The preparation of a revised edition has given me the opportunity to expand points that students have found difficult and to improve a number of awkward proofs. Except for the addition of problems, little change has been made in the first seven chapters. The remaining chapters in Part Two have been enlarged by the addition of sections on products of topological spaces, the Stone-Čech compactification, and topological vector spaces. Part Three, on the other hand, has been considerably revised. The approach to the abstract Lebesgue integral in Chapter 11 has been altered slightly to make the development parallel to the one used in Chapter 13. The proof of the Radon-Nikodym theorem has been simplified, and a section on L^p spaces has been added. Chapter 12 contains new material on inner

measure and its role in extension and uniqueness theorems. Chapter 13 on the Daniell integral has been rearranged to correct several erroneous proofs that made tacit use of things not proved until later. A new chapter on measure theory in locally compact spaces has been added.

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 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

Contents

Prologue to the Student	1
1 Set Theory	5
1 <i>Introduction</i> , 5	
2 <i>Functions</i> , 8	
3 <i>Unions, intersections, and complements</i> , 11	
4 <i>Algebras of sets</i> , 16	
5 <i>The axiom of choice and infinite direct products</i> , 18	
6 <i>Countable sets</i> , 19	
7 <i>Relations and equivalences</i> , 22	
8 <i>Partial orderings and the maximal principle</i> , 23	
9 <i>Well ordering and the countable ordinals</i> , 24	

Part One □ Theory of Functions of a Real Variable

2 The Real Number System	29
1 <i>Axioms for the real numbers</i> , 29	
2 <i>The natural and rational numbers as subsets of \mathbf{R}</i> , 32	
3 <i>The extended real numbers</i> , 34	
4 <i>Sequences of real numbers</i> , 35	
5 <i>Open and closed sets of real numbers</i> , 38	
6 <i>Continuous functions</i> , 44	
7 <i>Borel sets</i> , 50	

3 Lebesgue Measure	52
1 <i>Introduction</i> , 52	
2 <i>Outer measure</i> , 54	
3 <i>Measurable sets and Lebesgue measure</i> , 56	
*4 <i>A nonmeasurable set</i> , 63	
5 <i>Measurable functions</i> , 65	
6 <i>Littlewood's three principles</i> , 71	
4 The Lebesgue Integral	73
1 <i>The Riemann integral</i> , 73	
2 <i>The Lebesgue integral of a bounded function over a set of finite measure</i> , 75	
3 <i>The integral of a nonnegative function</i> , 82	
4 <i>The general Lebesgue integral</i> , 86	
*5 <i>Convergence in measure</i> , 91	
5 Differentiation and Integration	94
1 <i>Differentiation of monotone functions</i> , 94	
2 <i>Functions of bounded variation</i> , 98	
3 <i>Differentiation of an integral</i> , 101	
4 <i>Absolute continuity</i> , 104	
*5 <i>Convex functions</i> , 108	
6 The Classical Banach Spaces	111
1 <i>The L^p spaces</i> , 111	
2 <i>The Hölder and Minkowski inequalities</i> , 112	
3 <i>Convergence and completeness</i> , 115	
4 <i>Bounded linear functionals on the L^p spaces</i> , 119	

Part Two □ Abstract Spaces

7 Metric Spaces	127
1 <i>Introduction</i> , 127	
2 <i>Open and closed sets</i> , 129	
3 <i>Continuous functions and homeomorphisms</i> , 131	
4 <i>Convergence and completeness</i> , 133	
5 <i>Uniform continuity and uniformity</i> , 135	
6 <i>Subspaces</i> , 137	
7 <i>Baire category</i> , 139	

* Indicates sections peripheral to the principal line of argument.

8 Topological Spaces	142
1 Fundamental notions,	142
2 Bases and countability,	145
3 The separation axioms and continuous real-valued functions,	147
4 Product spaces,	150
5 Connectedness,	150
*6 Absolute \mathbb{G}_δ 's,	154
*7 Nets,	155
9 Compact Spaces	157
1 Basic properties,	157
2 Countable compactness and the Bolzano-Weierstrass property,	159
3 Compact metric spaces,	163
4 Products of compact spaces,	165
5 Locally compact spaces,	168
*6 The Stone-Čech compactification,	170
7 The Stone-Weierstrass theorem,	171
*8 The Ascoli theorem,	177
10 Banach Spaces	181
1 Introduction,	181
2 Linear operators,	184
3 Linear functionals and the Hahn-Banach theorem,	186
4 The closed graph theorem,	193
*5 Topological vector spaces,	197
*6 Weak topologies,	200
*7 Convexity,	203
8 Hilbert space,	210

Part Three □ General Measure and Integration Theory

11 Measure and Integration	217
1 Measure spaces,	217
2 Measurable functions,	223
3 Integration,	225
*4 General convergence theorems,	231
5 Signed measures,	232
6 The Radon-Nikodym theorem,	238
7 The L^p spaces,	243

12	Measure and Outer Measure	250
1	<i>Outer measure and measurability</i> , 250	
2	<i>The extension theorem</i> , 253	
*3	<i>The Lebesgue-Stieltjes integral</i> , 261	
4	<i>Product measures</i> , 264	
*5	<i>Inner measure</i> , 274	
*6	<i>Extension by sets of measure zero</i> , 281	
*7	<i>Carathéodory outer measure</i> , 283	
13	The Daniell Integral	286
1	<i>Introduction</i> , 286	
2	<i>The extension theorem</i> , 288	
3	<i>Uniqueness</i> , 294	
4	<i>Measurability and measure</i> , 295	
14	Measure and Topology	301
1	<i>Baire sets and Borel sets</i> , 301	
2	<i>Positive linear functionals and Baire measures</i> , 304	
3	<i>Bounded linear functionals on $C(X)$</i> , 308	
*4	<i>The Borel extension of a measure</i> , 313	
15	Mappings of Measure Spaces	317
1	<i>Point mappings and set mappings</i> , 317	
2	<i>Measure algebras</i> , 319	
3	<i>Borel equivalences</i> , 324	
4	<i>Set mappings and point mappings on complete metric spaces</i> , 328	
5	<i>The isometries of L^p</i> , 331	
	Epilogue	335
	Bibliography	337
	Index of Symbols	339
	Subject Index	341

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 treatments 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 notions 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 in Chapter 1 some of the basic facts from set theory. Since the remainder of the book is full of applications of set theory, the student should become adept at set-theoretic arguments as he

progresses through the book. I recommend that he first read Chapter 1 lightly and then refer back to it as needed. The books by Halmos [9]¹ and Suppes [23] contain a more thorough treatment of set theory and can be profitably read by the student while he reads this book. Gleason [7] also contains useful material on set theory and has an excellent discussion of the role of set theory in characterizing abstract mathematical structure. The reader will also find there a good discussion and development of the real number system and its properties.

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 notion 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 negation of a complex statement. Thus

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

¹ Numbers in brackets refer to the Bibliography, p. 337.

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

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

For a thorough discussion of the formal use of logical symbolism the student should refer to Suppes [22].

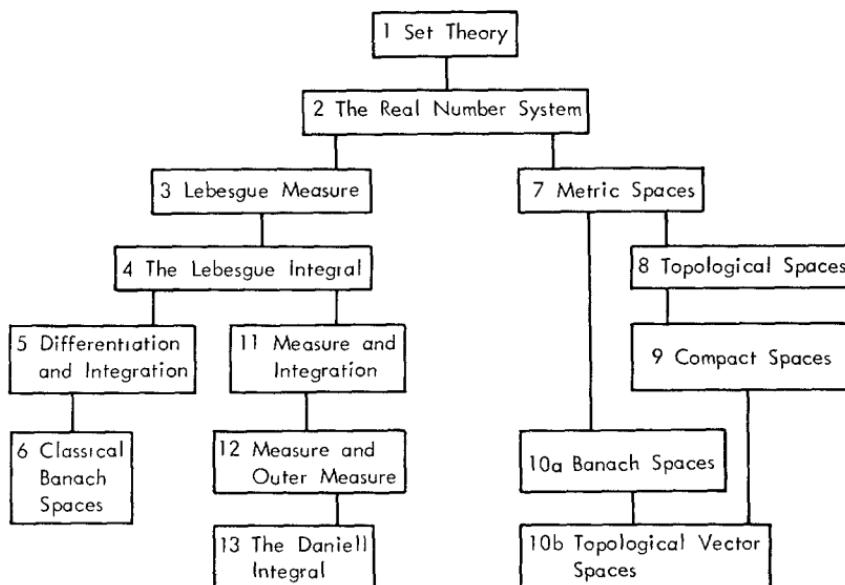
Statements and their proofs. Most of the principal statements (theorems, propositions, etc.) in mathematics have the standard form ‘if A , then B ’ or in symbols ‘ $A \Rightarrow B$ ’. The *contrapositive* of $A \Rightarrow B$ is the statement $(\neg B) \Rightarrow (\neg A)$. It is readily seen that 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 .

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 prob-

lens. 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 various preceding ones is indicated by the following chart (except for a few peripheral references). Chapters 14 and 15 depend on most of the preceding chapters. 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

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 shall 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 toward 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* [9], 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 more primitive notions of set theory.

For an axiomatic treatment, I recommend Suppes' book *Axiomatic Set Theory* [23] or the appendix to Kelley's book, *General Topology* [14].

The natural numbers (positive integers) play such an important role in this book that we introduce the special symbol \mathbf{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 \mathbf{N} , then $\{P(1) \& [P(n) \Rightarrow P(n + 1)]\} \Rightarrow (n)P(n)$. The well-ordering principle asserts that each nonempty subset of \mathbf{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 \in , and write ' $x \in A$ ' 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 $x \in A$ if and only if $x \in B$, then $A = B$. Suppose that each x in a set A is in the set B , that is, $x \in A \Rightarrow x \in B$; 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 \in and \subset , but we shall use it only in the latter context. We write ' $x \notin A$ ' to mean ' $(x \in A)$ ', that is, that x is not an element of A .

Since a set is determined by its elements, one of the commonest 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

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

and where the set X is understood we sometimes write¹

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

¹ The presence (explicit or implicit) of the qualifying set X is essential. Otherwise we are confronted with the Russell paradox (cf. Suppes [23], 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 $\{x: x \neq x\} = \emptyset$.
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 $\{n \in \mathbb{N}: P(n) \text{ false}\}.$]
5. Use mathematical induction to establish the well-ordering principle. [Given a set S of positive integers, let $P(n)$ be the proposition “If $n \in S$, 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 which 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 $x \in X$ 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 often express the fact that f is a function on X into Y by writing

$$f: X \rightarrow Y.$$

The set X is called the *domain* (or domain of definition) of f . The set of values taken by f , that is, the set $\{y \in Y: (\exists x)[y = f(x)]\}$, is called the *range* of f . The range of a function f will in general 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*. We shall not use it here.)

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

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

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

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 $(f = g) \Leftrightarrow (\forall x)[f(x) = g(x)]$, as well as axioms enabling us to construct functions.

More important than the notion of image of a 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 ; 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 $f: X \rightarrow Y$ 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 *bijection*.) In this case there is a function $g: Y \rightarrow X$ 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 $f: X \rightarrow Y$ and $g: Y \rightarrow Z$, we define a new function $h: X \rightarrow Z$ 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 $f: X \rightarrow Y$ and A is a subset of X , we can define a new function $g: A \rightarrow Y$ 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 $f|A$. 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 $\{i \in \mathbb{N} : i \leq n\}$. (We call such a set a *segment* of \mathbb{N} .) Similarly, an **infinite sequence** is a function whose domain is the set \mathbb{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 which is not finite is called infinite. (Many authors restrict the use of the word ‘countable’ to sets which 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.

One of the most useful ways 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

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

such that

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

From the uniqueness, it follows that $x_i^{(n)} = x_i^{(m)}$ for $i \leq n \leq m$. Thus if we define x_n to be $x_n^{(n)}$, we have $x_i^{(n)} = x_i$, 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 $x_{i+1} = f_i(x_1, \dots, x_i)$.

An important notion connected with that of a sequence is the notion of a subsequence. We say that a mapping g of \mathbb{N} into \mathbb{N} is **monotone** if $(i > j) \Rightarrow (g(i) > g(j))$. If f is an infinite sequence (that is, a function whose domain is \mathbb{N}), we say that h is an infinite **sub-**

sequence 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 $f: X \rightarrow Y$ 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 $g: Y \rightarrow X$ 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 $f: X \rightarrow Y$ be a mapping of X into Y . Show that f is onto if there is a mapping $g: Y \rightarrow X$ 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 20.)

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 $\wp(X)$ 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 which belong to both A and B . Thus

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

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 which 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:

$$A \cup \emptyset = A, \quad A \cap \emptyset = \emptyset$$

$$A \cup X = X, \quad A \cap X = A.$$

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

$$\tilde{\emptyset} = X, \quad \tilde{X} = \emptyset$$

$$\tilde{\tilde{A}} = A, \quad A \cup \tilde{A} = X, \quad A \cap \tilde{A} = \emptyset$$

$$A \subset B \Leftrightarrow \tilde{B} \subset \tilde{A}.$$

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 which are not in A . Thus

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

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

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

$$A \triangle B = (A \sim B) \cup (B \sim A).$$

The symmetric difference of two sets consists of all those points which 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 which belong to each member of \mathcal{C} . We denote this intersection by $\bigcap_{A \in \mathcal{C}} A$ 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 $\wp(X)$, that is, a mapping of \mathbf{N} (or a segment of \mathbf{N}) into $\wp(X)$. If $\langle A_i \rangle$ is an infinite sequence of subsets of X , we write $\bigcup_{i=1}^{\infty} A_i$ 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 $\bigcap_{i=1}^n B_i$ 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 Λ to X (or the set of subsets of X). If Λ 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_λ instead of $x(\lambda)$, and denote the indexed set itself by $\{x_\lambda\}$ or $\{x_\lambda : \lambda \in \Lambda\}$. We say that $\{x_\lambda\}$ is indexed by Λ . 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

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

and

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

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

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

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 ,

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

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

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 \mathcal{A} and \mathcal{B} are two collections of sets, then

$$\left[\bigcup \{A: A \in \mathcal{A}\} \right] \cap \left[\bigcup \{B: B \in \mathcal{B}\} \right] = \bigcup \{A \cap B: \langle A, B \rangle \in \mathcal{A} \times \mathcal{B}\}.$$
 16. a. Show $f[\bigcup A_\lambda] = \bigcup f[A_\lambda]$.
 b. Show $f[\bigcap A_\lambda] \subset \bigcap f[A_\lambda]$.
 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 \mathcal{A} of subsets of X is called an **algebra** of sets or a **Boolean algebra** if (i) $A \cup B$ is in \mathcal{A} whenever A and B are, and (ii) \tilde{A} is in \mathcal{A} whenever A is. It follows from De Morgan's laws that (iii) $A \cap B$ is in \mathcal{A} whenever A and B are. If a collection \mathcal{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, \dots, A_n are sets in \mathcal{A} then $A_1 \cup A_2 \cup \dots \cup A_n$ is again in \mathcal{A} . Similarly, $A_1 \cap A_2 \cap \dots \cap A_n$ is in \mathcal{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 \mathcal{A} which contains \mathcal{C} ; that is, there is an algebra \mathcal{A} containing \mathcal{C} and such that if \mathcal{B} is any algebra containing \mathcal{C} , then \mathcal{B} contains \mathcal{A} .*

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

2. Proposition: *Let \mathcal{A} be an algebra of subsets and $\langle A_i \rangle$ a sequence of sets in \mathcal{A} . Then there is a sequence $\langle B_i \rangle$ of sets in \mathcal{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_m and B_n be two such sets, and suppose $m < n$. Then $B_m \subset A_m \subset A_n$, 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

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

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 \mathfrak{A} is again in \mathfrak{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 σ -algebra which contains \mathcal{C} ; that is, there is a σ -algebra \mathfrak{A} containing \mathcal{C} such that if \mathfrak{B} is any σ -algebra containing \mathcal{C} then $\mathfrak{A} \subset \mathfrak{B}$.*

Problem

19. Prove Proposition 3.

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 Λ . We define the **direct product**

$$\bigtimes_{\lambda} X_{\lambda}$$

to be the collection of all sets $\{x_\lambda\}$ indexed by Λ 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_λ the λ^{th} coordinate of z .

If one of the X_λ 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_λ 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

20. Let $f: X \rightarrow Y$ be a mapping onto Y . Then there is a mapping $g: Y \rightarrow X$ 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 $\{x_1, \dots, x_n\}$ 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 \mathbf{N} of natural numbers is countable but not finite, any set which can be put in one-to-one correspondence with \mathbf{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 \mathbf{N} . To do this we define a function φ from \mathbf{N} into \mathbf{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 \mathbf{N} there is always a least such m . The correspondence $n \rightarrow x_{\varphi(n)}$ is a one-to-one correspondence between \mathbf{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 \mathbf{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 $\langle 2, 3, 5, 7, 11, \dots, p_k, \dots \rangle$ be the sequence of prime numbers. Then each n in \mathbf{N} has a unique factorization of the form³ $n = 2^{x_1} 3^{x_2} \cdots p_k^{x_k}$, where $x_i \in \mathbf{N}_0 = \mathbf{N} \cup \{0\}$ and $x_k > 0$. Let f be the function on \mathbf{N} which assigns to the natural number n the finite sequence $\langle x_1, \dots, x_k \rangle$ from \mathbf{N}_0 . Then S is a subset of the range of f . Hence S is countable by Proposition 4. ■

³ Except 1, we agree to write $1 = 2^0$.

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

21. Show that every subset of a finite set is finite.
22. Prove Proposition 6 by using Propositions 4 and 5. [Hint: The mapping

$$\begin{aligned}\langle p, q, 1 \rangle &\rightarrow p/q \\ \langle p, q, 2 \rangle &\rightarrow -p/q \\ \langle 1, 1, 3 \rangle &\rightarrow 0\end{aligned}$$

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

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

24. Let f be a function from a set X to the collection $\wp(X)$ of subsets of X . Then there is a set $E \subset X$ which is not in the range of f . (Let $E = \{x: x \notin f(x)\}$.)

25. 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 $x R y$) 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 $x R y$ 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 $\{(x, y) : x R y\}$. Since we consider two relations R and S to be the same if $(x R y) \Leftrightarrow (x S y)$, 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 $x R y$ and $y R z$ imply $x R z$ 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 $x R y$ implies $y R x$ for all x and y in X . It is said to be *reflexive* on X if for all $x \in X$ we have $x R x$.

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

⁴ Cf Suppes [23], p. 57, or Halmos [9], 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 [14], p. 260, where relations are not necessarily *sets* of ordered pairs.

by X/\equiv . The mapping $x \rightarrow E_x$ is called the natural mapping of X onto X/\equiv .

A *binary operation* on a 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 the reader may consult Birkhoff and MacLane [2], pp. 145ff.

Problems

26. Prove that $F + G$ as defined above depends only on F and G .
27. 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 $x R y$ and $y R x$ imply $x = y$ for all x and y in X . A relation $<$ 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 $\wp(X)$. A partial ordering $<$ 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 $x < y$ or $y < x$. Thus \leq linearly orders the set of real numbers, while \subset is not a linear ordering on $\wp(X)$.

If $<$ is a partial order on X and if $a < b$, 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 $a < x$. 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 $x < a$, 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 \prec 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 \prec 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 [23], Chapter 8, or Kelley [14], 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

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

29. Give an example of a partially ordered set which has a unique minimal element but no smallest element.

9 Well Ordering and the Countable Ordinals

A strict linear ordering \prec 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 $X = \mathbb{N}$ and \prec to mean less than, then \mathbb{N} is well ordered by \prec . 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 (cf. Suppes [23], Chapter 8, or Kelley [14], pp. 31–36).

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

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

- i. There is a last element Ω 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 Ω 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 Ω 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 ω 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

30. 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 non-empty subset of X has a least element, then $<$ is a linear ordering and consequently a well ordering.
31. 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^* .)

32. 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.

33. 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 \mathbf{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 \mathbf{R} of real numbers, the set P of positive real numbers, and the functions ‘+’ and ‘·’ on $\mathbf{R} \times \mathbf{R}$ to \mathbf{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 $x \in \mathbf{R}$ there is a $w \in \mathbf{R}$ such that $x + w = 0$.
- A5. $xy = yx$.
- A6. $(xy)z = x(yz)$.
- A7. $\exists 1 \in \mathbf{R}$ such that $1 \neq 0$ and $x \cdot 1 = x$ for all $x \in \mathbf{R}$.
- A8. For each x in \mathbf{R} different from 0 there is $w \in \mathbf{R}$ such that $xw = 1$.
- A9. $x(y + z) = xy + xz$.

Any set which 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 the real numbers are ordered. We could axiomatize the notion of a less than b , but it is somewhat more convenient to use the notion 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. $(x, y \in P) \Rightarrow x + y \in P$.
- B2. $(x, y \in P) \Rightarrow xy \in P$.
- B3. $(x \in P) \Rightarrow -x \notin P$.
- B4. $(x \in \mathbf{R}) \Rightarrow (x = 0) \text{ or } (x \in P) \text{ or } (-x \in P)$.

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 $y - x \in P$. 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 which 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 $\sup_{x \in S} x$ 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 $\inf_{x \in S} x$. Note that $\inf_{x \in S} x = -\sup_{x \in S} -x$.

Problems

1. Show that $1 \in P$.
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, and one is tempted to say that we define the real number 3 as $1 + 1 + 1$, and that in a "similar" fashion we can define real numbers corresponding to any natural number. Actually we may use the tools we have at hand to do this in a more precise fashion.

By the principle of recursive definition there is a function φ 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 φ is a one-to-one mapping of \mathbf{N} into \mathbf{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 φ 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 φ gives a one-to-one correspondence between the natural numbers and a subset of \mathbf{R} , and φ preserves sums, products, and the relation $<$. Strictly speaking, we should distinguish between the natural number n and its image $\varphi(n)$ under φ , but we shall not make the distinction here; we shall consider the set \mathbf{N} of natural numbers to be a subset of \mathbf{R} . By taking differences of natural numbers, we obtain the integers as a subset of \mathbf{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 reasons¹ 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: Let S be the set of integers k such that $k \leq x$. 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 It was first used by 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

$$\begin{aligned} x + \infty &= \infty, & x - \infty &= -\infty \\ x \cdot \infty &= \infty & \text{if } x > 0 \\ x \cdot -\infty &= -\infty & \text{if } x > 0 \end{aligned}$$

for all real numbers x , and set

$$\begin{aligned} \infty + \infty &= \infty, & -\infty - \infty &= -\infty \\ \infty \cdot (\pm \infty) &= \pm \infty, & -\infty \cdot (\pm \infty) &= \mp \infty. \end{aligned}$$

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 , and 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*.

4 Sequences of Real Numbers

By a sequence² $\langle x_n \rangle$ of real numbers we mean a function which 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 ϵ 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 10).

We extend this notion of limit of a sequence to include the value ∞ as follows: $\lim x_n = \infty$ if given Δ 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 $x_n \rightarrow l$. 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 ϵ of l . A weaker requirement is to have infinitely many terms of the sequence within ϵ 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 ∞ is a cluster point of $\langle x_n \rangle$ if, given Δ

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.

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 \limsup 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 ∞ is the limit superior of $\langle x_n \rangle$ if and only if, given Δ 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

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

Problems

6. Show that a sequence can have at most one limit.
7. Show that l is a cluster point of $\langle x_n \rangle$ if and only if there is a subsequence $\langle x_{n_j} \rangle_{j=1}^{\infty}$ which converges to l .
8. 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 which converges to a real number.
9. Show that a sequence $\langle x_n \rangle$ is convergent if and only if there is exactly one extended real number which is a cluster point of the sequence. Is this statement true if we omit the word ‘extended’?

10. 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 which 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.

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

12. 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 $x_k > l - \epsilon$.

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

14. 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 $l = \lim x_n$.

15. 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 $\infty - \infty$.

16. Prove that if $x_n \geq 0$ 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 $0 \cdot \infty$.

17. We shall say that a sequence (or series) $\langle x_\nu \rangle$ 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

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

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

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

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

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

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

- a. Show that $\sum_{x \in E} x < \infty$ 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 $\sum_{x \in E} x = \sum_{n=1}^{\infty} x_n$.

21. 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 $0 \leq a_n < p$ 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 $0 \leq a_n < p$, the series

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

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

If $p = 10$, this sequence is called the *decimal* expansion of x . For $p = 2$ it is called the *binary* expansion; and for $p = 3$, the *ternary* expansion.

22. Show that \mathbf{R} is uncountable. [Use Problem 1.23. 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 $\{x: a \leq x \leq b\}$. For closed intervals we take a and b finite. 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 δ to be the smaller of δ_1 and δ_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 $(-\frac{1}{n}, \frac{1}{n})$. 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 $x \in O$ 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 $w \in I_x$, say $x < w < b$, we have by the definition of b a number $y > w$ such that $(x, y) \subset O$, and so $w \in O$.

Moreover, $b \notin O$, for if $b \in O$, 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 $\{I_x\}$, $x \in O$. 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 (Lindelof): *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*

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

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

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

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 $x \in \bigcap \{F : F \in \mathcal{C}\}$. ■

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. ■

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 $\{O_1, \dots, O_n\}$ of sets in \mathcal{C} such that*

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

Proof: Let us first consider the case that 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} . The set E is bounded by b and so has a least upper bound c . Since $c \in [a, b]$, there is an $O \in \mathcal{C}$ which 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 $\{O_1, \dots, O_k\}$ of sets in \mathcal{C} which covers $[a, x]$. Consequently, the finite collection $\{O_1, \dots, O_k, O\}$ 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}^* which 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 $\{O_1, \dots, O_n, \tilde{F}\}$. Then $F \subset \tilde{F} \cup O_1 \cup \dots \cup O_n$. Since no point of F is contained in \tilde{F} , we have $F \subset O_1 \cup \dots \cup O_n$, and the collection $\{O_1, \dots, O_n\}$ 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

23. Is the set of rational numbers open or closed?
24. What are the sets of real numbers that are both open and closed?
25. Find two sets A and B such that $A \cap B = \emptyset$ and $\overline{A} \cap \overline{B} \neq \emptyset$.
26. 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 $x = \lim y_n$.
27. 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.

28. Show that $\overline{E} = E \cup E'$.
29. A set is called **isolated** if $E \cap E' = \emptyset$. Show that every isolated set of real numbers is countable.
30. A set D is called **dense** in \mathbf{R} if $\overline{D} = \mathbf{R}$. Show that the set of rational numbers is dense in \mathbf{R} .
31. Prove Propositions 12 and 13 using Propositions 5, 7, and 14.
32. Prove Propositions 5 and 7 using Propositions 12, 13, and 14.
33. 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° . Show that
- A is open if and only if $A = A^\circ$.
 - $A^\circ = \sim(\bar{A})$.
34. Derive Proposition 16 from the Heine-Borel Theorem using De Morgan's laws.
35. 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.
36. The **Cantor ternary set** C consists of all those real numbers in $[0, 1]$ which have ternary expansion (cf. Problem 21) $\langle a_n \rangle$ 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.
37. Show that the Cantor set can be put into a one-to-one correspondence with the interval $[0, 1]$.
38. 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 $\{I_{x_1}, \dots, I_{x_n}\}$ which covers F . Let $M = 1 + \max_{x \in F} [|f(x_1)|, \dots, |f(x_n)|]$. 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 $m = \sup_{x \in F} f(x)$. 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 $\{I_{x_1}, \dots, I_{x_n}\}$ which cover F . Set $a = \max_{x \in F} [f(x_1), \dots, f(x_n)]$. 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. ■

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

19. 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 $\{O_x : x \in F\}$ is an open covering of F . By the Heine-Borel Theorem there is a finite subcollection $\{O_{x_1}, \dots, O_{x_n}\}$ which covers F . Let $\delta = \frac{1}{2} \min(\delta_{x_1}, \dots, \delta_{x_n})$. Then δ 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,

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

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

Problems

39. 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.]

40. 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 $f^{-1}[O] = E \cap U$.

41. 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 .

42. Let f be that function defined by setting

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

At what points is f continuous?

43. 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 $|f|$.

44. 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).

45. A continuous function φ 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 φ is linear on each interval $[x_i, x_{i+1}]$. Let f be an arbitrary continuous function on $[a, b]$ and ϵ a positive number. Show that there is a polygonal function φ on $[a, b]$ with $|f(x) - \varphi(x)| < \epsilon$ for all $x \in [a, b]$.

46. Let x be a real number in $[0, 1]$ with the ternary expansion $\langle a_n \rangle$ (cf. Problem 21). 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 36), and that f maps the Cantor ternary set onto the interval $[0, 1]$. (This function is called the *Cantor ternary function*.)

47. *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

$$\overline{\lim}_{x \rightarrow y} f(x) = \inf_{\delta > 0} \sup_{0 < |x-y| < \delta} f(x)$$

$$\overline{\lim}_{x \rightarrow y+} f(x) = \inf_{\delta > 0} \sup_{0 < x-y < \delta} f(x)$$

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

- a. $\overline{\lim}_{x \rightarrow y} f(x) \leq A$ if and only if, given $\epsilon > 0$, there is a $\delta > 0$ such that for all x with $0 < |x - y| < \delta$ we have $f(x) \leq A + \epsilon$.
- b. $\overline{\lim}_{x \rightarrow y} f(x) \geq A$ if and only if, given $\epsilon > 0$ and $\delta > 0$, there is an x such that $0 < |x - y| < \delta$ and $f(x) \geq A - \epsilon$.
- c. $\underline{\lim}_{x \rightarrow y} f(x) \leq \overline{\lim}_{x \rightarrow y} f(x)$ with equality (for $\overline{\lim}_{x \rightarrow y} f \neq \pm\infty$) if and only if $\lim_{x \rightarrow y} f(x)$ exists.
- d. If $\overline{\lim}_{x \rightarrow y} f(x) = A$ and $\langle x_n \rangle$ is a sequence with $x_n \neq y$ such that $y = \lim_{x \rightarrow y} x_n$, then $\overline{\lim}_{x \rightarrow y} f(x_n) \leq A$.
- e. If $\overline{\lim}_{x \rightarrow y} f(x) = A$, then there is a sequence $\langle x_n \rangle$ with $x_n \neq y$ such that $y = \lim_{x \rightarrow y} x_n$ and $A = \lim_{x \rightarrow y} f(x_n)$.
- f. For a real number l we have $l = \lim_{x \rightarrow y} f(x)$ if and only if $l = \lim_{x \rightarrow y} f(x_n)$ for every sequence $\langle x_n \rangle$ with $x_n \neq y$ and $y = \lim_{x \rightarrow y} x_n$.

48. *Semicontinuous functions.* An extended real-valued function f is called *lower semicontinuous* at the point y if $f(y) \neq -\infty$ and $f(y) \leq \lim_{x \rightarrow y} f(x)$. Similarly, f is called *upper semicontinuous* at y if $f(y) \neq +\infty$

and $f(y) \geq \liminf_{x \rightarrow y} f(x)$. 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 $f(y) \leq f(x) + \epsilon$ 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 λ .

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 φ defined on an interval $[a, b]$ is called a **step function** if there is a partition $a = x_0 < x_1 < \dots < x_n = b$ such that for each i the function φ assumes only one value in the interval (x_i, x_{i+1}) . Show that a step function φ 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. Show that the step functions in (g) can be replaced by continuous functions.

i. Prove that a function f which 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 $f(y) \leq f(x)$ for all $x \in [a, b]$.

49. 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 $x \in [a, b]$, $g(x) \leq f(x) \leq h(x)$, 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 φ is any lower semicontinuous function such that $\varphi(x) \leq f(x)$ for all $x \in [a, b]$, then $\varphi(x) \leq g(x)$ 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 σ -algebras of sets which 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 σ -algebra which contains all of the open sets.*

Such a smallest σ -algebra exists by Proposition 1.3. It is also the smallest σ -algebra which contains all closed sets and the smallest σ -algebra which contains the open intervals.

A set which is a countable union of closed sets is called an \mathcal{F}_σ (\mathcal{F} for closed, σ for sum). Thus every countable set is an \mathcal{F}_σ , as is, of course, every closed set. A countable union of sets in \mathcal{F}_σ is again in \mathcal{F}_σ . Since

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

each open interval is an \mathcal{F}_σ , and hence each open set is an \mathcal{F}_σ .

We say that a set is a \mathcal{G}_δ if it is the intersection of a countable collection of open sets (\mathcal{G} for open, δ for *durchschnitt*). Thus the complement of an \mathcal{F}_σ is a \mathcal{G}_δ , and conversely.

The \mathcal{F}_σ and \mathcal{G}_δ are relatively simple types of Borel sets. We could also consider sets of type $\mathcal{F}_{\sigma\delta}$, which are the intersections of countable

collections of sets each of which is an \mathcal{F}_σ . Similarly, we can construct the classes $\mathcal{G}_{\delta\sigma}$, $\mathcal{F}_{\sigma\delta\sigma}$, etc. Thus the classes in the two sequences

$$\mathcal{F}_\sigma, \mathcal{F}_{\sigma\delta}, \mathcal{F}_{\sigma\delta\sigma}, \dots, \mathcal{G}_\delta, \mathcal{G}_{\delta\sigma}, \mathcal{G}_{\delta\sigma\delta}, \dots$$

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 [16], but we shall need only the properties that follow directly from the fact that they form the smallest σ -algebra containing the open and closed sets.

Problems

50. 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 $\{x: f(x) = \alpha\}$?

51. 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 \mathcal{G}_δ .

52. 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 $\mathcal{F}_{\sigma\delta}$.

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 which 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 which 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, $\mathfrak{M} = \wp(\mathbb{R})$;
- 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 σ -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 σ -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 σ -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 ∞ 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 which 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 which cover A , that is, collections for which $A \subset \bigcup I_n$, and for each such collection consider the sum of the lengths 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 ϵ , 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 ϵ , 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

$$(1) \quad \sum l(I_n) \geq b - a.$$

By the Heine-Borel theorem, any collection of open intervals covering $[a, b]$ contains a finite subcollection which 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 $\{I_n\}$ which cover $[a, b]$. Since a is contained in $\bigcup I_n$, there must be one of the I_n 's which 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

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.

interval (a_2, b_2) in the collection $\{I_n\}$ such that $b_1 \in (a_2, b_2)$; that is, $a_2 < b_1 < b_2$. Continuing in this fashion we obtain a sequence $(a_1, b_1), \dots, (a_k, b_k)$ 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

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

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 Δ , 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 Δ , $m^*I = \infty = l(I)$. ■

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 $\sum l(I_{n,i}) < m^*A_n + 2^{-n}\epsilon$. 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 ϵ was an arbitrary positive number,

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

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 $G \in \mathcal{G}_\delta$ 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 $m^*A \geq$

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

$m^*(A \cap E) + m^*(A \cap \tilde{E})$. 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.*

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, \dots, 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

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

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

10. Theorem: *The collection \mathfrak{M} of measurable sets is a σ -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

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

by the countable subadditivity of m^* . ■

11. Lemma: *The interval (a, ∞) 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 ϵ 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 σ -algebra, we have $(-\infty, a]$ measurable for each a since $(-\infty, a] = \sim(a, \infty)$.

Since $(-\infty, b) = \bigcup_{n=1}^{\infty} (-\infty, b - 1/n]$, 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 σ -algebra containing the open sets and must therefore contain the family \mathcal{G} of Borel sets, since \mathcal{G} is the smallest σ -algebra containing the open sets. [Note: The theorem also follows immediately from the fact that \mathfrak{M} is a σ -algebra containing each interval of the form (a, ∞) and the fact that \mathcal{G} is the smallest σ -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

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

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

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

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

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

The reverse inequality follows from countable subadditivity, and we have

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

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

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

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

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

Since $mE_1 < \infty$, we have

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

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 \mathcal{G}_δ with $E \subset G$, $m^*(G \sim E) = 0$;
- v. there is an F in \mathcal{F}_σ 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 $m^*(A \cap \bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} m^*(A \cap E_i)$.

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 2.36) 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

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

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 $P_i = P + r_i$. Then $P_0 = P$. Let $x \in P_i \cap P_j$. Then $x = p_i + r_i = p_j + r_j$ 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 σ -algebra containing the set P , then $m[0, 1)$ is either zero or infinite.*

The nonmeasurability of P with respect to any translation invariant countably 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 $E_i = E + r_i$. 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 non-measurable 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 which 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 α the set $\{x: f(x) > \alpha\}$ is measurable.
- ii. For each real number α the set $\{x: f(x) \geq \alpha\}$ is measurable.
- iii. For each real number α the set $\{x: f(x) < \alpha\}$ is measurable.
- iv. For each real number α the set $\{x: f(x) \leq \alpha\}$ is measurable.

These statements imply

- v. For each extended real number α 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 α 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 α 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

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

is measurable. ■

We will 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 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 $\sup \{f_1, \dots, f_n\}$, $\inf \{f_1, \dots, f_n\}$, $\sup_n f_n$, $\inf_n f_n$, $\overline{\lim} f_n$, and $\underline{\lim} f_n$ are all measurable.

Proof: If h is defined by $h(x) = \sup \{f_1(x), \dots, f_n(x)\}$, 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

⁵ French: *presque partout* (p.p.).

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

$$\begin{aligned}\{x: g(x) > \alpha\} &= \{x: f(x) > \alpha\} \cup \{x \in E: g(x) > \alpha\} \\ &\sim \{x \in E: g(x) \leq \alpha\}.\end{aligned}$$

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 α , 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*

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

except on a set of measure less than ϵ ; i.e. $m\{x: |f(x) - g(x)| \geq \epsilon\} < \epsilon$ and $m\{x: |f(x) - h(x)| \geq \epsilon\} < \epsilon$. 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** χ_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 χ_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 φ is called **simple** if it is measurable and assumes only a finite number of values. If φ is simple and has the values $\alpha_1, \dots, \alpha_n$ then $\varphi = \sum_{i=1}^n \alpha_i \chi_{A_i}$, where $A_i = \{x: \varphi(x) = \alpha_i\}$. 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 $\{x: f(x) > 0\} = E$, 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 $\{x: f(x) > \alpha\}$ 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

$$\begin{aligned}\chi_{A \cap B} &= \chi_A \cdot \chi_B \\ \chi_{A \cup B} &= \chi_A + \chi_B - \chi_A \cdot \chi_B \\ \chi_{\tilde{A}} &= 1 - \chi_A.\end{aligned}$$

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 α a fixed number, then $f + g$ is measurable if we define $f + g$ to be α whenever it is of the form $\infty - \infty$ or $-\infty + \infty$.

d. Let f and g be measurable extended real-valued functions which 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 $[a, b]$ which takes the values $\pm\infty$ only on a set of measure zero, and given $\epsilon > 0$, there is an M such that $|f| \leq M$ 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 φ such that $|f(x) - \varphi(x)| < \epsilon$ except where $|f(x)| \geq M$. If $m \leq f \leq M$, then we may take φ so that $m \leq \varphi \leq M$.

c. Given a simple function φ 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 $m \leq \varphi \leq M$, then we can take g so that $m \leq g \leq M$.

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 $m \leq g \leq M$, then we may take h so that $m \leq h \leq M$.

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 σ -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 α 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 σ -algebra \mathfrak{A} of sets?

28. Let f_1 be the Cantor ternary function (cf. Problem 2.46), 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 Problems 4.15 and 6.14. 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 measurable real-valued function such that for each x in E we have $f_n(x) \rightarrow f(x)$. Then, given $\epsilon > 0$ and $\delta > 0$, there is a measurable set $A \subset E$ with $mA < \delta$ and an integer N such that for all $x \notin A$ and all $n \geq N$,*

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

Proof: Let

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

and set

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

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 $f_n(x) \rightarrow f(x)$. Thus $\bigcap E_N = \emptyset$, and so, by Proposition 14, $\lim mE_N = 0$. Hence given $\delta > 0$, $\exists N$ so that

⁶ *Lectures on the Theory of Functions*, Oxford, 1944, p. 26.

$mE_N < \delta$; that is,

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

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

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

If, as in the hypothesis of the proposition, we have $f_n(x) \rightarrow f(x)$ 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 $mA = 0$ such that $f_n \rightarrow f$ pointwise on $E \sim B$, we say that $f_n \rightarrow f$ 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 which 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 which 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 φ 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 2.39.]

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 $\lim_{t \rightarrow 0} f_t(x) = 0$, 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

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

We then define the upper Riemann integral of f by

$$R\int_a^b f(x) dx = \inf S$$

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

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

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 ψ which has the form

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

for some subdivision of $[a, b]$ and some set of constants 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

$$R \int_a^b f(x) dx = \inf \int_a^b \psi(x) dx$$

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 $R \int_a^b f(x) dx = b - a$ and $R \underline{\int}_a^b f(x) dx = 0$.

- 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

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

The function χ_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 φ is not unique. However, we note that a function φ is simple if and only if it is measurable and assumes only a finite number of values. If φ is a simple function and $\{a_1, \dots, a_n\}$ the set of non-zero values of φ , then

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

where $A_i = \{x: \varphi(x) = a_i\}$. This representation for φ 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 φ vanishes outside a set of finite measure, we define the integral of φ by

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

when φ 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 $A_a = \{x: \varphi(x) = a\} = \bigcup_{a_i=a} E_i$. Hence $a mE_a = \sum_{a_i=a} a_i mE_i$ by the additivity of m , and so

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

2. Proposition: Let φ and ψ 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 which occur in the canonical representations of φ and ψ . Let A_0 and B_0 be the sets where φ and ψ are zero. Then the sets E_k obtained by taking all the intersections $A_i \cap B_j$ form a finite disjoint collection of measurable sets, and we may write

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

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

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. \blacksquare

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 φ and ψ 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*

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

for all simple functions φ and ψ , 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

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

Thus

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

and

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

whence

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

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

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

Then given n there are simple functions φ_n and ψ_n such that

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

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

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

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 Δ_ν is contained in the set $\{x: \varphi_n(x) < \psi_n(x) - 1/\nu\}$, and this latter set has measure less than ν/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 $\psi \geq f$.

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 which 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

$$R \underline{\int}_a^b f(x) dx \leq \sup_{\varphi \leq f} \int_a^b \varphi(x) dx \leq \inf_{\psi \geq f} \int_a^b \psi(x) dx \leq R \bar{\int}_a^b f(x) dx.$$

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 $f \leq g$ a.e., then

$$\int_E f \leq \int_E g.$$

Hence $|\int f| \leq \int |f|.$

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

$$AmE \leq \int_E f \leq BmE.$$

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 ψ is a simple function so is $a\psi$, and conversely (if $a \neq 0$). Hence for $a > 0$,

$$\int_E af = \inf_{\psi \geq f} \int_E a\psi = a \inf_{\psi \geq f} \int_E \psi = a \int_E f.$$

If $a < 0$,

$$\int_E af = \inf_{\varphi \leq f} \int_E a\varphi = a \sup_{\varphi \leq f} \int_E \varphi = a \inf_{\psi \geq f} \int_E \psi = a \int_E f,$$

using Proposition 3.

If ψ_1 is a simple function not less than f and ψ_2 a simple function not less than g , then $\psi_1 + \psi_2$ is a simple function not less than $f + g$. Hence

$$\int_E f + g \leq \int_E (\psi_1 + \psi_2) = \int_E \psi_1 + \int_E \psi_2.$$

Since the infimum of the right side is $\int f + \int g$, we have

$$\int_E f + g \leq \int_E f + \int_E g.$$

On the other hand, $\varphi_1 \leq f$ and $\varphi_2 \leq g$ imply $\varphi_1 + \varphi_2$ is a simple function not greater than $f + g$. Hence

$$\int_E f + g \geq \int_E (\varphi_1 + \varphi_2) = \int_E \varphi_1 + \int_E \varphi_2.$$

Since now the supremum of the right side is $\int f + \int g$, we have

$$\int_E f + g \geq \int_E f + \int_E g,$$

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 $\psi \geq f - g$, $\psi \geq 0$ a.e. From this it follows that

$$\int_E \psi \geq 0,$$

whence

$$\int_E f - g \geq 0.$$

Similarly,

$$\int_E f - g \leq 0,$$

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 15. The proposition is a special case of Theorem 15.

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

$$\int_E f_n \rightarrow \int_E f. \blacksquare$$

Problem

2. a. Let f be a bounded function on $[a, b]$, and let h be the upper envelope of f (cf. Problem 2.49). Then $R\int_a^b f = \int_a^b h$. (If $\varphi \geq f$ is a step function, then $\varphi \geq h$ except at a finite number of points and so $\int_a^b h \leq R\int_a^b f$. But there is a sequence $\langle \varphi_n \rangle$ of step functions such that $\varphi_n \downarrow h$. By Proposition 6 we have $\int_a^b h = \lim \int_a^b \varphi_n \geq R\int_a^b f$.)

- b. Use part (a) to prove *Lebesgue's Theorem*: 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.

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.

7. Proposition: *If f and g are nonnegative measurable functions, then:*

- i. $\int_E cf = c \int_E f, \quad c > 0.$
- 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 $h(x) + k(x) \leq f(x) + g(x)$, and so

$$\int_E h + \int_E k \leq \int_E f + g.$$

Taking suprema, we have

$$\int_E f + \int_E g \leq \int_E f + g.$$

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

$$h(x) \leq f(x)$$

and

$$k(x) \leq g(x),$$

while h and k are bounded by the bound for l and vanish where l vanishes. Hence $\int_E l = \int_E h + \int_E k \leq \int_E f + \int_E g$, and so

$$\int_E f + \int_E g \geq \int_E f + g. \blacksquare$$

8. Theorem (Fatou's Lemma): *If $\langle f_n \rangle$ is a sequence of nonnegative measurable functions and $f_n(x) \rightarrow f(x)$ almost everywhere on a set E , then*

$$\int_E f \leq \underline{\lim} \int_E f_n.$$

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 $h_n(x) \rightarrow h(x)$ 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

$$\int_E f \leq \underline{\lim} \int_E f_n. \blacksquare$$

9. Monotone Convergence Theorem: Let $\langle f_n \rangle$ be an increasing sequence of nonnegative measurable functions, and let $f = \lim f_n$. Then

$$\int f = \lim \int f_n.$$

Proof: By Theorem 8 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

$$\int f = \lim \int f_n. \blacksquare$$

10. Corollary: Let u_n be a sequence of nonnegative measurable functions, and let $f = \sum_{n=1}^{\infty} u_n$. Then

$$\int f = \sum_{n=1}^{\infty} \int u_n.$$

11. 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

$$\int_E f < \infty.$$

12. 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 7

$$\int_E f = \int_E (f - g) + \int_E g.$$

Since the left side is finite, the terms on the right must also be finite, and so g is integrable. ■

13. 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 $\delta < \frac{\epsilon}{2N}$. If $mA < \delta$, we have

$$\begin{aligned} \int_A f &= \int_A (f - f_N) + \int_A f_N \\ &< \int_E (f - f_N) + NmA \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \blacksquare \end{aligned}$$

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 9.

6. Let $\langle f_n \rangle$ be a sequence of nonnegative measurable functions which converges 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 sequence 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 functions on $(-\infty, \infty)$ such that $f_n \rightarrow f$ a.e., and suppose that $\int f_n \rightarrow \int f < \infty$. Then for each measurable set E we have $\int_E f_n \rightarrow \int_E f$.

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

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

14. 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 $f \leq g$ a.e., then $\int_E f \leq \int_E g$.

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 7. 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 $f^+ + f_2 = f^- + f_1$. By Proposition 7,

$$\int f^+ + \int f_2 = \int f^- + \int f_1,$$

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

$$\begin{aligned}\int_{A \cup B} f &= \int f \chi_{A \cup B} \\ &= \int f \chi_A + \int f \chi_B \\ &= \int_A f + \int_B f. \blacksquare\end{aligned}$$

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 $\int(f + g)$ is independent of the choice of values in these ambiguous cases.

15. Lebesgue Convergence Theorem: *Let g be integrable over E and let $\langle f_n \rangle$ be a sequence of measurable functions such that $|f_n| \leq g$ 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 8

$$\int_E (g - f) \leq \underline{\lim} \int_E (g - f_n).$$

Since $|f| \leq g$, f is integrable, and we have

$$\int_E g - \int_E f \leq \int_E g - \overline{\lim} \int_E f_n,$$

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. \blacksquare

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:

16. Theorem: Let $\langle g_n \rangle$ be a sequence of integrable functions which converge 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 which converge 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

$$\left| \int_E f \right| \leq \int_E |f|.$$

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 φ is a simple function, we have two definitions for $\int \varphi$, that on page 75 and that on page 87. Show that they are the same.

12. Let g be an integrable function on a set E and suppose $\langle f_n \rangle$ is a sequence of measurable functions such that $|f_n(x)| \leq g(x)$ a.e. on E . Then

$$\int_E \underline{\lim} f_n \leq \underline{\lim} \int_E f_n \leq \overline{\lim} \int_E f_n \leq \int_E \overline{\lim} f_n.$$

13. Let h be an integrable function and $\langle f_n \rangle$ a sequence of measurable functions with $f_n \geq -h$ and $\lim f_n = f$. Show that $\int f_n$ and $\int f$ have a meaning and that $\int f \leq \underline{\lim} \int f_n$.

14. Let $\langle f_n \rangle$ be a sequence of integrable functions such that $f_n \rightarrow f$ a.e. with f integrable. Then $\int |f - f_n| \rightarrow 0$ if and only if $\int |f_n| \rightarrow \int |f|$.

15. a. Let f be integrable over E . Then given $\epsilon > 0$, there is a simple function φ 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 ψ such that

$$\int_E |f - \psi| < \epsilon.$$

[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 $\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f(x) \cos nx dx = 0$. [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

$$\lim_{t \rightarrow 0} \int_{-\infty}^{\infty} |g(x)[f(x) - f(x+t)]| = 0.$$

[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 $Q = \{\langle x, t \rangle : 0 \leq x \leq 1, 0 \leq t \leq 1\}$ and which is a measurable function of x for each fixed value of t . Suppose that $\lim_{t \rightarrow 0} f(x, t) = f(x)$ and that for all t we have $|f(x, t)| \leq g(x)$, where g is an integrable function on $[0, 1]$. Then

$$\lim_{t \rightarrow 0} \int f(x, t) dx = \int f(x) dx.$$

[Problem 2.47(f) 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 $Q = \{\langle x, t \rangle : 0 \leq x \leq 1, 0 \leq t \leq 1\}$, 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 $\frac{\partial f}{\partial t}$ exist. Suppose that $\frac{\partial f}{\partial t}$ 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 $\int |f_n| \rightarrow 0$. 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 η the measure of the sets $\{x : |f_n(x)| > \eta\}$ 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

$$m\{x : |f(x) - f_n(x)| \geq \epsilon\} < \epsilon.$$

An example of a sequence $\langle f_n \rangle$ which 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

$$m\{x: |f_n(x)| > \epsilon\} \leq \frac{2}{n},$$

and so $f_n \rightarrow 0$ 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:

17. Proposition: *Let $\langle f_n \rangle$ be a sequence of measurable functions which converges in measure to f . Then there is a subsequence $\langle f_{n_k} \rangle$ which converges to f almost everywhere.*

Proof: Given ν , there is an integer n_ν such that for all $n \geq n_\nu$ we have

$$m\{x: |f_n(x) - f(x)| \geq 2^{-\nu}\} < 2^{-\nu}.$$

Let $E_\nu = \{x: |f_{n_\nu}(x) - f(x)| \geq 2^{-\nu}\}$. Then if $x \notin \bigcup_{\nu=k}^{\infty} E_\nu$, we have $|f_{n_\nu}(x) - f(x)| < 2^{-\nu}$ for $\nu \geq k$, and so $f_{n_\nu}(x) \rightarrow f(x)$. Hence $f_{n_\nu}(x) \rightarrow f(x)$ 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 $mA = 0$. ■

18. 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 which converges to f in measure, then each subsequence $\langle f_{n_k} \rangle$ converges to f in measure.

21. Deduce Proposition 18 from Proposition 17 using Problems 20 and 2.11.

22. Prove that a sequence $\langle f_n \rangle$ of measurable functions converges to f in measure if and only if every subsequence of $\langle f_n \rangle$ has in turn a subsequence which converges in measure to f .

23. 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 which converges to f a.e. on E .

24. Use Proposition 13 to prove directly that, if $f_n \rightarrow f$ in measure and if there is an integrable function g such that for all n we have $|f_n| \leq g$, then $\int |f_n - f| \rightarrow 0$.

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 $n_{v+1} > n_v$ so that $m\{x: |f_{n_v} - f_{n_{v+1}}| > 2^{-v}\} < 2^{-v}$. Then the series $\sum(f_{n_{v+1}} - f_{n_v})$ converges almost everywhere to a function g . Let $f = g + f_{n_1}$. Then $f_{n_v} \rightarrow f$ in measure, and one can show that consequently $f_n \rightarrow f$ in measure.]

5 Differentiation and Integration

In this chapter we shall 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 difficult, 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{s} be a collection of intervals. Then we say that \mathcal{s} covers a set E in the sense of Vitali, if for each $\epsilon > 0$ and any x in E there is an interval $I \in \mathcal{s}$ such that $x \in I$ and $l(I) < \epsilon$.

1. Lemma (Vitali): Let E be a set of finite outer measure and \mathfrak{s} a collection of intervals which cover E in the sense of Vitali. Then given $\epsilon > 0$ there is a finite disjoint collection $\{I_1, \dots, I_N\}$ of intervals in \mathfrak{s} 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 \mathfrak{s} is closed, for otherwise we replace each interval by its closure and observe that the set of endpoints of I_1, \dots, I_N has measure zero.

Let O be an open set of finite measure containing E . Since \mathfrak{s} is a Vitali covering of E , we may assume without loss of generality that each I of \mathfrak{s} is contained in O . We choose a sequence $\langle I_n \rangle$ of disjoint intervals of \mathfrak{s} by induction as follows: Let I_1 be any interval in \mathfrak{s} , and suppose I_1, \dots, I_n have already been chosen. Let k_n be the supremum of the lengths of the intervals of \mathfrak{s} which do not meet any of the intervals I_1, \dots, 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 \mathfrak{s} with $l(I_{n+1}) > \frac{1}{2}k_n$ and I_{n+1} disjoint from I_1, \dots, I_n .

Thus we have a sequence $\langle I_n \rangle$ of disjoint intervals of \mathfrak{s} , 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 \mathfrak{s} which contains x and whose length is so small that I does not meet any of the intervals I_1, \dots, 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 $l(I) \leq k_{n-1} \leq 2l(I_n)$. 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 $l(I) + \frac{1}{2}l(I_n) \leq \frac{5}{2}l(I_n)$. 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

$$m^*R \leq \sum_{N+1}^{\infty} l(J_n) = 5 \sum_{N+1}^{\infty} l(I_n) < \epsilon. \blacksquare$$

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:

$$D^+f(x) = \overline{\lim}_{h \rightarrow 0+} \frac{f(x+h) - f(x)}{h},$$

$$D^-f(x) = \overline{\lim}_{h \rightarrow 0+} \frac{f(x) - f(x-h)}{h},$$

$$D_+f(x) = \underline{\lim}_{h \rightarrow 0+} \frac{f(x+h) - f(x)}{h},$$

$$D_-f(x) = \underline{\lim}_{h \rightarrow 0+} \frac{f(x) - f(x-h)}{h}.$$

Clearly we have $D^+f(x) \geq D_+f(x)$ and $D^-f(x) \geq D_-f(x)$. If $D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x) \neq \pm\infty$, we say that f is differentiable at x and define $f'(x)$ to be the common value of the derivates at x .

2. 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

$$E_{u,v} = \{x: D^+f(x) > u > v > D_-f(x)\}$$

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 $\{I_1, \dots, I_N\}$ 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 $y \in A$ is the left endpoint of an arbitrarily small interval $(y, y + k)$ which is contained in some I_n and such that $f(y + k) - f(y) > uk$. Using Lemma 1 again, we can pick out a finite collection $\{J_1, \dots, J_M\}$ 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

$$\begin{aligned} \sum_{i=1}^M f(y_i + k_i) - f(y_i) &> u \sum_{i=1}^M k_i \\ &> u(s - 2\epsilon). \end{aligned}$$

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 ϵ , we have $vs \geq us$. But $u > v$, and so s must be zero.

This shows that

$$g(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

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 $g_n(x) \rightarrow g(x)$ 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

- 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 $D_-f(0)$.
- If the function f assumes its maximum at c , then $D^+f(c) \leq 0$ and $D_-f(c) \geq 0$.
- If f is continuous on $[a, b]$ and one of its derivates (say D^+) is everywhere nonnegative on $[a, b]$, then $f(b) \geq f(a)$. [Hint: First show this result for a function g for which $D^+g \geq \epsilon > 0$. Apply this to $g = f + \epsilon x$.]

2 Functions of Bounded Variation

Let f be a real-valued function defined on the interval $[a, b]$, and let $a = x_0 < x_1 < \dots < x_k = b$ 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 $f \in BV$.

3. 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 $[a, b]$

$$\begin{aligned} p &= n + f(b) - f(a) \\ &\leq N + f(b) - f(a), \end{aligned}$$

and taking suprema over all possible subdivisions, we obtain

$$P \leq N + f(b) - f(a).$$

Since $N \leq T < \infty$,

$$P - N \leq f(b) - f(a).$$

Similarly,

$$N - P \leq f(a) - f(b),$$

and so

$$P - N = f(b) - f(a).$$

Thus

$$T \geq p + n = p + p - \{f(b) - f(a)\} = 2p + N - P,$$

and

$$T \geq 2P + N - P = P + N.$$

Since $T \leq P + N$, we have $T = P + N$. ■

4. 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 3. 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

$$T_a^b(f) \leq g(b) + h(b) - g(a) - h(a). ■$$

5. Corollary: *If f is of bounded variation on $[a, b]$, then $f'(x)$ exists for almost all x in $[a, b]$.*

Problems

4. 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 $x \rightarrow c^-$ and also as $x \rightarrow c^+$. 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.

5. Show that if $a \leq c \leq b$, then $T_a^b = T_a^c + T_c^b$ and that hence $T_a^c \leq T_a^b$.

6. Show that $T_a^b(f + g) \leq T_a^b(f) + T_a^b(g)$, and $T_a^b(cf) = |c| T_a^b(f)$.
7. Let $\langle f_n \rangle$ be a sequence of functions on $[a, b]$ which converges at each point of $[a, b]$ to a function f . Then $T_a^b(f) \leq \underline{\lim} T_a^b(f_n)$.
8. a. Let f be defined by $f(0) = 0$ and $f(x) = x^2 \sin(1/x^2)$ for $x \neq 0$. 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 $x \neq 0$. Is g of bounded variation on $[-1, 1]$?

3 Differentiation of an Integral

In this section we shall 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.

6. **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 4.13. To show that F is of bounded variation, let $a = x_0 < x_1 < \dots < x_k = b$ be any subdivision of $[a, b]$. Then

$$\begin{aligned} \sum_{i=1}^k |F(x_i) - F(x_{i-1})| &= \sum_{i=1}^k \left| \int_{x_{i-1}}^{x_i} f(t) dt \right| \leq \sum_{i=1}^k \int_{x_{i-1}}^{x_i} |f(t)| dt \\ &= \int_a^b |f(t)| dt. \end{aligned}$$

Thus

$$T_a^b(F) \leq \int_a^b |f(t)| dt < \infty. \blacksquare$$

7. **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 4.11

$$\int_O f = \sum \int_{a_n}^{b_n} f.$$

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 $x \in [a, b]$ we have

$$\int_a^x f \neq 0.$$

Similarly for f negative on a set of positive measure, and the lemma follows by contraposition. ■

8. 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 6, 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

$$f_n(x) \rightarrow F'(x) \text{ a.e.,}$$

the bounded convergence theorem implies that

$$\begin{aligned} \int_a^c F'(x) dx &= \lim \int_a^c f_n(x) dx = \lim_{h \rightarrow 0} \frac{1}{h} \int_a^c (F(x + h) - F(x)) dx \\ &= \lim \left[\frac{1}{h} \int_c^{c+h} F(x) dx - \frac{1}{h} \int_a^{a+h} F(x) dx \right] \\ &= F(c) - F(a) = \int_a^c f(x) dx, \end{aligned}$$

since F is continuous. Hence

$$\int_a^c \{F'(x) - f(x)\} dx = 0$$

for all $c \in [a, b]$, and so

$$F'(x) = f(x) \text{ a.e.}$$

by Lemma 7. ■

9. Theorem: Let f be an integrable function on $[a, b]$, and suppose

$$F(x) = F(a) + \int_a^x f(t) dt.$$

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 8

$$\frac{d}{dx} \int_a^x f_n = f_n(x) \quad \text{a.e.},$$

and so

$$\begin{aligned} F'(x) &= \frac{d}{dx} G_n + \frac{d}{dx} \int_a^x f_n \\ &\geq f_n(x) \quad \text{a.e.} \end{aligned}$$

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 2 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 4.13 that every indefinite integral is absolutely continuous.

10. Lemma: *If f is absolutely continuous on $[a, b]$, then it is of bounded variation on $[a, b]$.*

Proof: Let δ be the δ in the definition of absolute continuity which 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 δ , where K is the largest integer less than $1 + (b - a)/\delta$. Thus for any subdivision we have $t \leq K$ and so $T \leq K$. ■

11. Corollary: *If f is absolutely continuous, then f has a derivative almost everywhere.*

12. 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 ϵ and η 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 δ , where δ is the positive number corresponding to ϵ in the definition of the absolute continuity of f . If we label the x_k so that $x_k \leq x_{k+1}$, we have

$$y_0 = a \leq x_1 < y_1 \leq x_2 < \cdots \leq y_n \leq c = x_{n+1}$$

and

$$\sum_{k=0}^n |x_{k+1} - y_k| < \delta.$$

Now

$$\begin{aligned}\sum_{k=1}^n |f(y_k) - f(x_k)| &\leq \eta \sum (y_k - x_k) \\ &< \eta(c - a)\end{aligned}$$

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

$$\begin{aligned}|f(c) - f(a)| &= \left| \sum_{k=0}^n [f(x_{k+1}) - f(y_k)] + \sum_{k=1}^n [f(y_k) - f(x_k)] \right| \\ &\leq \epsilon + \eta(b - a).\end{aligned}$$

Since ϵ and η are arbitrary positive numbers, $f(c) - f(a) = 0$. ■

13. 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 4.13. 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 2, 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 9 that $f'(x) = F'(x) - G'(x) = 0$ a.e., and

so f is constant by Lemma 12. Thus

$$F(x) = \int_a^x F'(t) dt + F(a). \blacksquare$$

14. Corollary: *Every absolutely continuous function is the indefinite integral of its derivative.*

Problems

9. 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 $[0, 1]$?

10. Let f be absolutely continuous on $[a, b]$. Show that

$$T_a^b(f) = \int_a^b |f'| \quad \text{and} \quad P_a^b(f) = \int_a^b [f']^+.$$

11. The Cantor ternary function (Problem 2.46) is continuous and monotone but not absolutely continuous.

12. 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. Construct a function f which is monotone and continuous on $[0, 1]$ but which is not absolutely continuous on any subinterval of $[0, 1]$.

c. Show that there is a strictly increasing singular function on $[0, 1]$.

13. *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 $H = \{x: g'(x) \neq 0\}$. 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

$$\int_c^d f(y) dy = \int_a^b f(g(x))g'(x) dx.$$

14. Let g be an absolutely continuous monotone function on $[0, 1]$ and E a set of measure zero. Then $g[E]$ has measure zero.

15. 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 χ_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?

16. 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.

*5 Convex Functions

A function φ defined on an open interval (a, b) is said to be *convex* if for each $x, y \in (a, b)$ and each $\lambda, 0 \leq \lambda \leq 1$ we have

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y).$$

If we look at the graph of φ 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 φ . An important property of the chords of a convex function is given by the following lemma, whose proof is left to the reader.

15. Lemma: If φ is convex on (a, b) and if x, y, x', y' are points of (a, b) with $x \leq x' < y' \text{ and } x < y \leq y'$, then the chord over (x', y') has larger slope than the chord over (x, y) ; that is,

$$\frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(y') - \varphi(x')}{y' - x'}.$$

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 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.

16. Proposition: *If φ is convex on (a, b) , then φ is absolutely continuous on each closed subinterval of (a, b) . The right- and left-hand derivatives of φ 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 15 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 φ is absolutely continuous there.

If $x_0 \in (a, b)$, then $[\varphi(x) - \varphi(x_0)]/(x - x_0)$ is an increasing function of x by Lemma 15, and so the limits as x approaches x_0 from the right and from the left exist and are finite. Thus φ 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. ■

Let φ be a convex function on (a, b) and $x_0 \in (a, b)$. The line $y = m(x - x_0) + \varphi(x_0)$ through $(x_0, \varphi(x_0))$ is called a *supporting line* at x_0 if it always lies below the graph of φ , that is, if $\varphi(x) \geq m(x - x_0) + \varphi(x_0)$. It follows from Lemma 15 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:

17. Proposition (Jensen Inequality): *Let φ be a convex function on $(-\infty, \infty)$ and f an integrable function on $[0, 1]$. Then*

$$\int \varphi(f(t)) dt \geq \varphi \left[\int f(t) dt \right].$$

Proof: Let $\alpha = \int f(t) dt$, and let $y = m(x - \alpha) + \varphi(\alpha)$ be the equation of a supporting line at α . Then

$$\varphi(f(t)) \geq m(f(t) - \alpha) + \varphi(\alpha).$$

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 λ 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 μ on the line by setting $\mu(a, b] = m\{t: a < f(t) \leq b\}$, 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 φ .

An important application of the Jensen inequality is obtained by taking for our convex function φ the function \exp defined by $\exp x = e^x$. The inequality then becomes a generalization of the inequality between the arithmetic and geometric mean:

18. Corollary: *Let f be an integrable function on $[0, 1]$. Then*

$$\int \exp(f(t)) dt \geq \exp \left[\int f(t) dt \right].$$

Problems

17. 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.

18. When do we have equality in Corollary 18?

19. 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

$$\prod_{n=1}^{\infty} \xi_n^{\alpha_n} \leq \sum_{n=1}^{\infty} \alpha_n \xi_n.$$

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 αf 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 α and β . 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 α is a constant, then $\|\alpha f\| = |\alpha| \|f\|$. In the next section we shall derive two inequalities, the second of which states that $\|f + g\| \leq \|f\| + \|g\|$, if $p \geq 1$. Henceforth, we shall always assume $p \geq 1$. 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 $\|f\| = 0 \Leftrightarrow f \equiv 0$. Unfortunately, the norm for the L^p

spaces does 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^∞ 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^∞ is a linear space, and it becomes a normed linear space if we define

$$\|f\| = \|f\|_\infty = \text{ess sup } |f(t)|,$$

where $\text{ess sup } f(t)$ 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 $\lim_{p \rightarrow \infty} \|f\|_p = \|f\|_\infty$.
3. Prove that $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$.
4. If $f \in L^1$ and $g \in L^\infty$, then

$$\int |fg| \leq \|f\|_1 \cdot \|g\|_\infty.$$

2 The Hölder and Minkowski Inequalities

Before proving inequalities about the L^p norm, we first establish the following lemma, which is a generalization of the inequality between the arithmetic and geometric mean:

1. **Lemma:** Let α and β be nonnegative real numbers, and suppose $0 < \lambda < 1$. Then $\alpha^\lambda \beta^{1-\lambda} \leq \lambda\alpha + (1 - \lambda)\beta$ with equality only if $\alpha = \beta$.

¹ To be pedantic we should say the elements of L^p are not functions but rather equivalence classes of functions (cf. Problem 10.10).

Proof: Consider the function φ defined for nonnegative real numbers t by

$$\varphi(t) = (1 - \lambda) + \lambda t - t^\lambda.$$

Then

$$\varphi'(t) = \lambda(1 - t^{\lambda-1}).$$

Since $\lambda - 1 < 0$, we have $\varphi'(t) < 0$ for $t < 1$ and $\varphi'(t) > 0$ for $t > 1$. Thus for $t \neq 1$ we have $\varphi(t) < \varphi(1) = 0$. Hence

$$(1 - \lambda) + \lambda t \geq t^\lambda,$$

with equality only for $t = 1$. If $\beta \neq 0$, the lemma follows by substituting α/β for t , while, if $\beta = 0$, the lemma is trivial. ■

2. Hölder Inequality: If p and q are nonnegative extended real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$, and iff $f \in L^p$ and $g \in L^q$, then $f \cdot g \in L^1$ and

$$\int |fg| \leq \|f\|_p \cdot \|g\|_q.$$

Equality holds if and only if, for some nonzero constants α and β , 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 $1 < p < \infty$ and consequently $1 < q < \infty$. Let us first suppose that $\|f\|_p = \|g\|_q = 1$, and apply our lemma with $\alpha = |f(t)|^p$, $\beta = |g(t)|^q$, $\lambda = \frac{1}{p}$, $1 - \lambda = \frac{1}{q}$. Then

$$(1) \quad |f(t)g(t)| \leq \lambda|f(t)|^p + (1 - \lambda)|g(t)|^q,$$

and integrating both sides we have

$$(2) \quad \int |fg| \leq \lambda \int |f|^p + (1 - \lambda) \int |g|^q = 1.$$

The inequality is trivial if $\|f\| = 0$ or $\|g\| = 0$. Let f and g be any elements of L^p and L^q with $\|f\| \neq 0$ and $\|g\| \neq 0$. Then $f/\|f\|_p$ and $g/\|g\|_q$ both have norm 1. Substituting them in (2) gives

$$\frac{1}{\|f\|_p \|g\|_q} \int |fg| = \int \frac{|f|}{\|f\|_p} \cdot \frac{|g|}{\|g\|_q} \leq 1.$$

Equality in (1) can occur only if $|f(t)|^p = |g(t)|^q$, and equality can hold in (2) only if it holds almost everywhere in (1). Hence equality holds only if $\|g\|^q |f|^p = \|f\|^p |g|^q$ a.e. ■

3. Minkowski Inequality: If f and g are in L^p , then so is $f + g$ and

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof: The cases $p = 1$ and $p = \infty$ are straightforward and are left to the reader. Hence we assume $1 < p < \infty$.

Since $|f + g|^p \leq 2^p(|f|^p + |g|^p)$, we have $f + g$ in L^p . Also

$$\int |f + g|^p \leq \int |f + g|^{p-1} |f| + \int |f + g|^{p-1} |g|.$$

By the Hölder inequality,

$$\int |f + g|^{p-1} |f| \leq \|f\|_p \|(f + g)^{p-1}\|_q,$$

$$\int |f + g|^{p-1} |g| \leq \|g\|_p \|(f + g)^{p-1}\|_q,$$

and

$$\begin{aligned} \|(f + g)^{p-1}\|_q &= \left\{ \int |f + g|^{(p-1)q} \right\}^{1/q} \\ &= \{\|f + g\|_p\}^{p/q}, \end{aligned}$$

since $p = q(p - 1)$. Hence

$$\|f + g\|_p^p \leq (\|f\|_p + \|g\|_p)(\|f + g\|_p)^{p/q}$$

or

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \blacksquare$$

Problems

5. For $1 \leq p < \infty$, we denote by l^p the space of all sequences $\langle \xi_\nu \rangle_{\nu=1}^\infty$ such that $\sum_{\nu=1}^\infty |\xi_\nu|^p < \infty$. 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

$$\sum_{\nu=1}^\infty |\xi_\nu \eta_\nu| \leq \|\langle \xi_\nu \rangle\|_p \cdot \|\langle \eta_\nu \rangle\|_q,$$

where

$$(\|\langle \xi_\nu \rangle\|_p)^p = \sum_{\nu=1}^\infty |\xi_\nu|^p$$

and

$$\|\langle \eta_\nu \rangle\|_\infty = \sup |\eta_\nu|.$$

This is the Hölder inequality for sequences.

6. 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 $1 \leq p \leq \infty$.

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 $f_n \rightarrow f$.

Another way of formulating the convergence of f_n to f is by noting that $f_n \rightarrow f$ if $\|f_n - f\| \rightarrow 0$. Convergence in the space L^p , $1 \leq p < \infty$, 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 $\|f - f_n\|_p \rightarrow 0$. Convergence in L^∞ is nearly uniform convergence (Problem 8).

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 $n \geq N$ and all $m \geq N$ 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 $f_n \rightarrow f$. 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,

$$\left\| s - \sum_{i=1}^n f_i \right\| \rightarrow 0.$$

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.

4. 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

$$\|s_n - s_m\| = \left\| \sum_{i=m}^n f_i \right\| \leq \sum_{i=m}^n \|f_i\| \leq \sum_{i=N}^{\infty} \|f_i\| < \epsilon.$$

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 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 $f_{n_k} \rightarrow f$, 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 $f_n \rightarrow f$. ■

5. 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 the Fatou 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 $|s_n(x)| \leq g(x)$, we have $|s(x)| \leq g(x)$. Consequently, s is in L^p and we have

$$|s_n(x) - s(x)|^p \leq 2^p [g(x)]^p.$$

Since $2^p g^p$ is integrable and $|s_n(x) - s(x)|^p$ converges to 0 for almost all x , we have

$$\int |s_n - s|^p \rightarrow 0$$

by the Lebesgue convergence theorem. Thus $\|s_n - s\|^p \rightarrow 0$, whence $\|s_n - s\| \rightarrow 0$. Consequently, the series $\langle f_n \rangle$ has in L^p the sum s . ■

Problems

7. Prove that every convergent sequence is a Cauchy sequence.
8. Let $\langle f_n \rangle$ be a sequence of functions in L^∞ . Prove that $\langle f_n \rangle$ converges to f in L^∞ if and only if there is a set E of measure zero such that f_n converges to f uniformly on \tilde{E} .
9. Prove that L^∞ is complete.
10. Prove that l^p is complete ($1 \leq p < \infty$) (see Problem 5).
11. 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.
12. We denote by l^∞ the space of all bounded sequences of real numbers and define $\|\langle \xi_\nu \rangle\|_\infty = \sup |\xi_\nu|$. Show that l^∞ is a Banach space.
13. 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^∞ norm).
14. Let f be a function in L^p , $1 \leq p < \infty$. Show that given $\epsilon > 0$ there is a continuous function φ and a step function ψ such that $\|f - \varphi\| < \epsilon$ and $\|f - \psi\| < \epsilon$.
15. Let $\langle f_n \rangle$ be a sequence of functions in L^p , $1 \leq p < \infty$, 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 $\|f_n\| \rightarrow \|f\|$. (For $p = 1$ this is just Problem 4.14.)
16. 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 $\|f_n\| \leq M$ 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 $p = 1$?

17. Let $f_n \rightarrow f$ in L^p , $1 \leq p < \infty$, and let $\langle g_n \rangle$ be a sequence of measurable functions such that $|g_n| \leq M$, all n , and $g_n \rightarrow g$ a.e. Then $g_n f_n \rightarrow g f$ in L^p .

4 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 $|F(f)| \leq M \cdot \|f\|$ 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 $\|F\| \leq \|g\|_q$. In fact, we actually have $\|F\| = \|g\|_q$. To see this for the case $1 < p < \infty$, we set²

$$f = |g|^{q/p} \operatorname{sgn} g.$$

Then $|f|^p = |g|^q = fg$. Hence f is in L^p and $\|f\|_p = (\|g\|_q)^{q/p}$. Now

$$\begin{aligned} F(f) &= \int fg = \int |g|^q \\ &= (\|g\|_q)^q = \|g\|_q \|f\|_p, \end{aligned}$$

and so $\|F\|$ must be at least as great as $\|g\|_q$. We state this result as a proposition, the cases $p = 1$ and $p = \infty$ being left to the reader (see Problem 18).

6. Proposition: *Each function g in L^q defines a bounded linear functional F on L^p by*

$$F(f) = \int fg.$$

We have $\|F\| = \|g\|_q$.

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 $x < 0$.

The goal of the present section is to show that for $1 \leq p < \infty$ 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.

7. Lemma: *Let g be an integrable function on $[0, 1]$, and suppose that there is a constant M such that*

$$\left| \int fg \right| \leq M \|f\|_p$$

for all bounded measurable functions f . Then g is in L^q , and $\|g\|_q \leq M$.

Proof: Let us first assume $1 < p < \infty$. We define a sequence of bounded measurable functions by setting

$$g_n(x) = \begin{cases} g(x) & \text{if } |g(x)| \leq n \\ 0 & \text{if } |g(x)| > n, \end{cases}$$

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

$$(\|g_n\|_q)^q = \int f_n g \leq M \|f_n\|_p = M (\|g_n\|_q)^{q/p}.$$

Since $q - q/p = 1$,

$$\|g_n\|_q \leq M$$

and

$$\int |g_n|^q \leq M^q.$$

Since $|g_n|^q$ converges to $|g|^q$ almost everywhere, we have

$$\int |g|^q \leq \lim \int |g_n|^q \leq M^q$$

by the Fatou lemma. Thus $g \in L^q$, and $\|g\|_q \leq M$.

For the case $p = 1$, let $E = \{x: |g(x)| \geq M + \epsilon\}$, 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 $1 \leq p < \infty$:

8. 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*

$$F(f) = \int f g.$$

We have also $\|F\| = \|g\|_q$.

Proof: Let χ_s be the characteristic function of the interval $[0, s]$. We begin our investigation of F by observing what it does to χ_s . For each s the value of $F(\chi_s)$ is a real number $\Phi(s)$, and this defines a function Φ on $[0, 1]$. Now I maintain that Φ is absolutely continuous. For let $\{(s_i, s'_i)\}$ be any finite collection of nonoverlapping subintervals of $[0, 1]$ of total length less than δ . 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 Φ is less than ϵ over any finite collection of disjoint intervals of total length δ if we take $\delta = \epsilon^p / \|F\|^p$. Thus Φ is absolutely continuous.

By Theorem 5.13 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 ψ , 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 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 $\|f - \psi_n\|_p \rightarrow 0$. Since F is bounded, and

$$\begin{aligned} |F(f) - F(\psi_n)| &= |F(f - \psi_n)| \\ &\leq \|F\| \|f - \psi_n\|_p, \end{aligned}$$

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 $|F(f)| \leq \|F\| \|f\|_p$, we have g in L^q and $\|g\|_q \leq \|F\|$ by Lemma 7.

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 Problem 14 there is for each $\epsilon > 0$ a step function ψ such that $\|f - \psi\|_p < \epsilon$. Since ψ is bounded we have

$$F(\psi) = \int \psi g.$$

Hence

$$\begin{aligned} |F(f) - \int fg| &= |F(f) - F(\psi) + \int \psi g - \int fg| \\ &\leq |F(f - \psi)| + \left| \int (\psi - f)g \right| \\ &\leq \|F\| \|f - \psi\|_p + \|g\|_q \|f - \psi\|_p \\ &< (\|F\| + \|g\|_q) \epsilon. \end{aligned}$$

Since ϵ is an arbitrary number, we must have

$$F(f) = \int fg.$$

The equality $\|F\| = \|g\|_q$ follows from Proposition 6. ■

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 14.8, we give a representation for the bounded linear functionals on C . Unfortunately, the bounded linear functionals on L^∞ (and on l^∞) do not admit of a similar representation.

Problems

- 18. 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.]

- 19.** Find a representation for the bounded linear functionals on l^p , $1 \leq p < \infty$.

- 20.** Find a representation for the bounded linear functionals on c and on c_0 . (*Caveat:* These representations are different.)

- 21.** Show that the element g in L^q given by Theorem 8 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. These 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 ρ 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 ρ 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 $x = \langle x_1, \dots, x_n \rangle$ 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 if we set

$$\rho(x, y) = \|x - y\|,$$

the triangle inequality being equivalent to $\|x + y\| \leq \|x\| + \|y\|$.

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 ρ . For example, the set of all n -tuples of real numbers can also be made into a metric space by use of the metric ρ^* 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 τ 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 τ has all the properties required of a metric and that $\mathbf{R}^m \times \mathbf{R}^n = \mathbf{R}^{m+n}$.

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^∞ .

Problems

1. a. Show that the set of all n -tuples of real numbers becomes a metric space under each of the following metrics:

$$\rho^*(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|$$

$$\rho^+(x, y) = \max \{|x_1 - y_1|, \dots, |x_n - y_n|\}.$$

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 **spheroid** (or sphere) centered at x and having radius δ 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. *Pseudometrics.* A pair $\langle X, \rho \rangle$ is called a *pseudometric space* if ρ satisfies all the conditions of a metric except that $\rho(x, y) = 0$ need not imply $x = y$. 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^* .

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 sets 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 \overline{E} to denote the set of points of closure of E . Clearly $E \subset \overline{E}$.

2. Proposition: If $A \subset B$, then $\overline{A} \subset \overline{B}$. Also $(\overline{A \cup B}) = \overline{A} \cup \overline{B}$, and $(\overline{A \cap B}) \subset \overline{A} \cap \overline{B}$.

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

3. Proposition: The closure \overline{E} of any set E is closed; that is, $\overline{\overline{E}} = \overline{E}$.

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 $\overline{D} = X$.*

Since the set of rational numbers is a countable dense subset of \mathbf{R} , we see that \mathbf{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 spheroid at x with radius δ we mean the set

$$S_{x,\delta} = \{y: \rho(x, y) < \delta\}.$$

Let $\{O_i\}$ consist of those spheroids $S_{x,\delta}$ for which x is in D and δ 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 spheroid $S_{y,\delta}$ such that $S_{y,\delta} \subset O$. By taking δ even smaller we may assume δ 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 such that $x \in O_i \subset S$. Hence $x_i \in S$, and we see that $x \in \overline{D}$. ■

Problems

4. a. Show that C is a closed subset of L^∞ (see Chapter 6).
 b. Show that the set of all integrable functions which vanish for $0 \leq t < \frac{1}{2}$ 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 $\overline{E} = \bigcap_{E \subset F} F$ for closed sets F . The *interior* E° 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 $\sim(\overline{E}) = (\sim E)^\circ$.
6. a. Show that each spheroid is open.
 b. Show that the sets $\{x: \rho(x, y) \leq \delta\}$ are closed.
 c. Is the set in (b) always the closure of the spheroid $\{x: \rho(x, y) < \delta\}$?
7. Which of the spaces \mathbf{R}^n , C , L^∞ , 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 which associates to each $x \in X$ a unique $y \in Y$. 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 $y \in Y$ there is some $x \in X$ 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 $x \in X$.

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 which 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 $h: X \rightarrow Y$ 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 which 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 which arises naturally in this connection is that of equivalent metrics: Two metrics ρ and σ 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

$$\rho(x, E) = \inf_{y \in E} \rho(x, y).$$

a. Show that for a fixed E the function f given by $f(x) = \rho(x, E)$ is continuous.

b. Show that $\{x: \rho(x, E) = 0\} = \overline{E}$.

10. a. Prove that two metrics on a set X are equivalent if and only if given $x \in X$ and $\epsilon > 0$ there is a $\delta > 0$ such that for all $y \in X$

$$\rho(x, y) < \delta \Rightarrow \sigma(x, y) < \epsilon$$

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:

$$\rho(x, y) = [(x_1 - y_1)^2 + \dots + (x_n - y_n)^2]^{1/2}$$

$$\rho^*(x, y) = |x_1 - y_1| + \dots + |x_n - y_n|$$

$$\rho^+(x, y) = \max \{|x_1 - y_1|, \dots, |x_n - y_n|\}.$$

c. Find a metric for the set of n -tuples which is not equivalent to these.

11. Show that if ρ 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 .

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 spheroid about x contains all but a finite number of terms of the sequence.

We often write $x = \lim x_n$ or $x_n \rightarrow x$ to mean that x is the limit of the sequence $\langle x_n \rangle$. If we have merely the weaker condition that each spheroid 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 $\langle x_n \rangle$ which 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 which 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 $x \in \overline{E}$, while, if $x \in \overline{E}$, there is a sequence from E which 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 under a continuous function of a Cauchy sequence 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 homeomorphisms, as is the property of being complete. Properties which are preserved under uniform homeomorphisms are called uniform properties. In addition to the properties of being a Cauchy sequence and of completeness we have also 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 ρ and σ 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 σ and ρ 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 $\sigma(x, y) < \delta \Rightarrow \rho(x, y) < \epsilon$.

We conclude this section by stating the following useful extension theorem for uniformly continuous mappings. Its proof is outlined in Problem 20.

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 \overline{E} ; that is, there is a unique continuous mapping g from \overline{E} into Y such that $g(x) = f(x)$ for $x \in E$. Moreover, g is uniformly continuous.*

Problems

19. Prove Proposition 10.

20. Prove Proposition 11 by the following steps:

a. If $\langle x_n \rangle$ is a sequence from E which 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 which agrees with f on E , then $h \equiv g$.

21. a. Show that the metrics in Problem 10b are uniformly equivalent.

b. Find a metric for the set of n -tuples of real numbers which 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 ρ .

22. a. Boundedness is a metric but not a uniform property (cf. Problem 21c).

b. A metric space X is said to be totally bounded if given $\epsilon > 0$ there are a finite number of spheroids of radius ϵ which 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 $[0, \infty)$.]

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 ρ 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^∞ .

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 $\overline{E} \cap S$, where \overline{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 $\overline{E} \cap S$. If A is closed in S we must have $A = S \cap \overline{A}$, while on the other hand if F is closed in X , then the closure in S of $S \cap F$ is

$$S \cap \overline{(S \cap F)} \subset S \cap (\overline{S} \cap \overline{F}) \subset S \cap F,$$

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 which 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.

Problem

23. Prove Proposition 14. [Hint: Problem 14 is helpful.]

7 Baire Category

In this section we shall go more deeply into certain aspects of metric spaces. We begin by defining a set E to be **nowhere dense** if $\sim\bar{E}$ is dense. This is equivalent to stating that \bar{E} contains no spheroid. For example, the integers are nowhere dense in \mathbf{R} , and the Cantor ternary set is nowhere dense in $[0, 1]$. A set E is said to be of **first category** (or meager) if it is the union of a countable collection of nowhere dense sets. A set which is not of first category is said to be of **second category**, and the complement of a set of first category is called **residual**. We shall show that a complete metric space is of second category when considered as a subset of itself. We begin with the following theorem:

15. Theorem: *Let X be a complete metric space and $\{O_n\}$ a countable collection of dense open subsets of X . Then $\bigcap O_n$ is not empty.*

Proof: Let x_1 be a point of O_1 and S_1 a spheroid (of radius r_1) which is centered at x_1 and contained in O_1 . 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 spheroid 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 spheroids 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 spheres. Then for $n, m \geq N$ we have $x_n \in S_N$ and $x_m \in S_N$. Hence $\rho(x_n, x_m) \leq 2r_N$, and $\langle x_n \rangle$ is a Cauchy sequence, since $r_n \rightarrow 0$. By the completeness of X there is a point x such that $x_n \rightarrow x$. Since $x_n \in S_{N+1}$ for $n > N$, we have $x \in \bar{S}_{N+1} \subset S_N \subset O_N$. Hence $x \in \bigcap O_n$. ■

16. Corollary (Baire Category Theorem): *A complete metric space is not the union of a countable collection of nowhere dense sets.*

Proof: Let $\{E_n\}$ be a countable collection of nowhere dense sets. Then $O_n = \sim\bar{E}_n$ is a dense open set, and so there is a point $x \in \bigcap O_n$. But this means $x \notin \bigcup E_n$. ■

As an application of the Baire Category Theorem we establish the following theorem, which is known as the *uniform boundedness principle*:

17. Theorem: Let \mathfrak{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 $|f(x)| \leq M_x$ for all $f \in \mathfrak{F}$. Then there is a nonempty open set $O \subset X$ and a constant M such that $|f(x)| \leq M$ for all $f \in \mathfrak{F}$ and all $x \in O$.

Proof: For each integer m , let $E_{m,f} = \{x : |f(x)| \leq m\}$, and set $E_m = \bigcap_{f \in \mathfrak{F}} E_{m,f}$. 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 $|f(x)| \leq m$ for all $f \in \mathfrak{F}$; 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 spheroid O . But for every $x \in O$ we have $|f(x)| \leq m$ for all $f \in \mathfrak{F}$. ■

Problems

24. 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 spheroid contained in $O \sim E$.

25. 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.

26. If X is a complete metric space and E is a set of first category in E , then $\sim E$ is dense in X .

27. 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 $[0, 1]$ which has measure 1.

28. 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.

29. 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 \mathcal{F}_σ .

c. The set of rational numbers in $[0, 1]$ is not a \mathcal{G}_δ .

d. Is there a real-valued function on $[0, 1]$ which is continuous on the rationals and discontinuous on the irrationals?

30. Let C be the space of continuous functions on $[0, 1]$ and set $F_n = \{f: (\exists x_0) \text{ with } |f(x) - f(x_0)| \leq n(x - x_0) \text{ for all } x, x_0 \leq x < 1\}$.

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 φ , and φ can be approximated to within $\epsilon/2$ by a polygonal function ψ 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 derivative on the right at at least one point of $[0, 1]$ is a set of first category in C .

d. There is a nowhere differentiable continuous function on $[0, 1]$.

31. Let $\langle f_n \rangle$ be a sequence of continuous functions from a metric space X to a metric space Y such that for each x in X the sequence $\langle f_n(x) \rangle$ converges to a point $f(x)$ in Y . Then f defines a mapping of X into Y . Show that there is a set E of first category in X such that f is continuous at each point $x \in X \sim E$. Hence if X is complete, f is continuous at a dense set of points in X . [Hint: Let $F_{n,m} = \{x: \sigma(f_n(x), f_k(x)) < 1/m \text{ for all } k \geq n\}$. Let $F_{n,m}^\circ$ be the interior of $F_{n,m}$. Set $E = \bigcup_{n,m} [F_{n,m} \sim F_{n,m}^\circ]$.]

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 $\langle X, \mathfrak{J} \rangle$ is a nonempty set X of points together with a family \mathfrak{J} of subsets (which we shall call open) possessing the following properties:

- i. $X \in \mathfrak{J}$, $\emptyset \in \mathfrak{J}$;
- ii. $O_1 \in \mathfrak{J}$ and $O_2 \in \mathfrak{J}$ imply $O_1 \cap O_2 \in \mathfrak{J}$;
- iii. $O_\alpha \in \mathfrak{J}$ implies $\bigcup_\alpha O_\alpha \in \mathfrak{J}$.

The family \mathfrak{J} 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 to some metric space is called *metrizable*, and the metric ρ 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.*

If A is a subset of a topological space $\langle X, \mathfrak{J} \rangle$, we can define a topology s for A by taking for s those subsets B of A for which there is a set $O \in \mathfrak{J}$ such that $B = A \cap O$. We call s the topology inherited from \mathfrak{J} and call the topological space $\langle A, s \rangle$ a subspace of $\langle X, \mathfrak{J} \rangle$. 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 $x_n \in O$ 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 $x_n \in O$. Thus if $\langle x_n \rangle$ has a subsequence which converges 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 which 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 $\langle X, \mathfrak{J} \rangle$ into a topological space $\langle Y, \mathfrak{s} \rangle$ is said to be **continuous** if the inverse image of every open set is open, that is, if $O \in \mathfrak{s} \Rightarrow f^{-1}[O] \in \mathfrak{J}$.

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 .

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{J} and \mathfrak{s} are two topologies for the same set X . Then \mathfrak{s} is said to be stronger than \mathfrak{J} if $\mathfrak{s} \supset \mathfrak{J}$. Thus \mathfrak{s} is stronger than \mathfrak{J} if and only if the identity mapping of (X, \mathfrak{J}) into (X, \mathfrak{s}) 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.

If \mathfrak{s} and \mathfrak{J} are two topologies for a set X , then $\mathfrak{s} \cap \mathfrak{J}$ is also a topology. In fact, if $\{\mathfrak{J}_\alpha\}$ is any collection of topologies, then $\bigcap_\alpha \mathfrak{J}_\alpha$ 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.

2. 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{J} which 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. We say that x is a *point of closure* of a set E if for every $O \in \mathfrak{J}$ with $x \in O$ we have $O \cap E \neq \emptyset$. Prove that the set \overline{E} of points of closure of E is the intersection of all closed sets containing E .
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 $\langle f(x_n) \rangle$.

2 Bases and Countability

A collection \mathfrak{B} of open subsets of a topological space X is called a **base** for the topology \mathfrak{J} of X if for each open set O in X and each $x \in O$ there is a set $B \in \mathfrak{B}$ such that $x \in B \subset O$. A collection \mathfrak{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 $B \in \mathfrak{B}_x$ such that $x \in B \subset O$. Thus a collection \mathfrak{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 spheroids form a base, and the spheroids centered at x form a base at x .

If \mathfrak{B} is a base for the topology \mathfrak{J} , then $O \in \mathfrak{J}$ if and only if for each $x \in O$ there is a $B \in \mathfrak{B}$ with $x \in B \subset O$. The “only if” part of

the statement follows from the definition of a base, while if for each x in O we have $x \in B \subset O$, 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:

3. 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 $x \in B_1 \cap B_2$ then there is a $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.*

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 set $\mathfrak{I} = \{O : (x \in O)(\exists B \in \mathcal{B})(x \in B \subset O)\}$, then $\emptyset \in \mathfrak{I}$, and the union of sets in \mathfrak{I} will be in \mathfrak{I} , and the first condition on \mathcal{B} implies that $X \in \mathfrak{I}$. To show that the intersection of two sets O_1 and O_2 in \mathfrak{I} is itself in \mathfrak{I} , let $x \in O_1 \cap O_2$. Then there are sets B_1 and B_2 in \mathcal{B} such that $x \in B_1 \subset O_1$ and $x \in B_2 \subset O_2$. Let B_3 be a set in \mathcal{B} with $x \in B_3 \subset B_1 \cap B_2$. Then $x \in B_3 \subset O_1 \cap O_2$, and it follows that $O_1 \cap O_2$ is 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 spheroids 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

8. a. Let \mathcal{B} be a base for the topological space (X, \mathfrak{J}) . Then $x \in \overline{E}$ if and only if for every $B \in \mathcal{B}$ with $x \in B$, there is a $y \in B \cap E$.

b. Let X satisfy the first axiom of countability. Then $x \in \overline{E}$ if and only if there is a sequence from E which 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 which converges to x .

9. A mapping f of a topological space X into a topological space Y is said to be continuous at x if for any open set O containing $f(x)$ there is an open set U containing x such that $f[U] \subset O$.

a. Show that f is continuous if and only if it is continuous at each point of X .

b. Let \mathcal{G}_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 $C \in \mathcal{C}_y$ there is a $B \in \mathcal{G}_x$ such that $B \subset f^{-1}[C]$.

10. Let \mathcal{C} be any collection of subsets of X . Let \mathfrak{G} consist of X and all finite intersections of sets in \mathcal{C} . Show that \mathfrak{G} is a base for the weakest topology which contains \mathcal{C} .

11. Let X be an uncountable set of points, and let \mathfrak{J} consist of the empty set and all subsets of X whose complement is finite. Show that \mathfrak{J} is a topology for X and that the space $\langle X, \mathfrak{J} \rangle$ does not satisfy the first axiom of countability.

12. Let X be the set of real numbers, and let \mathfrak{G} be the set of all intervals of the form $[a, b)$. Show that \mathfrak{G} is the base of a topology \mathfrak{J} for X . (This topology is called the *half-open interval topology*.) Show that $\langle X, \mathfrak{J} \rangle$ satisfies the first but not the second axiom of countability and that the rationals are dense in X . Is $\langle X, \mathfrak{J} \rangle$ metrizable?

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:

- T₁. *Given two distinct points x and y , there is an open set which contains y but not x .*
- T₂. *Given two distinct points x and y , there are disjoint open sets O_1 and O_2 such that $x \in O_1$ and $y \in O_2$.*
- T₃. *In addition to T₁, 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 $F \subset O_2$.*
- T₄. *In addition to T₁, 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 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 $T_4 \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1$.

4. 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 18 and 19).

5. 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 .*

6. 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 $x \in A$.*

If X is any set of points and \mathfrak{F} is any collection of real-valued functions on X , there is always a weakest topology on X such that every function in \mathfrak{F} is continuous, for let $C = \{E: E = f^{-1}[O], f \in \mathfrak{F} \text{ and } O \text{ an open subset of } \mathbf{R}\}$ and apply Proposition 3. This topology is called the **weak topology** generated (or induced) by \mathfrak{F} . If the functions in \mathfrak{F} are all continuous in some topology \mathfrak{J} , then the weak topology generated by \mathfrak{F} is in general weaker than \mathfrak{J} . For these topologies to coincide, there must be enough continuous real-valued functions. One condition which guarantees this is the following: Given a closed set F and a point $x \notin F$ there is a function $f \in \mathfrak{F}$ with $f(x) = 1$ and $f \equiv 0$ on F . If this condition is satisfied when \mathfrak{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 $T_{3\frac{1}{2}}$.

The following theorem characterizes separable metric spaces. It can be proved using the concepts of the next section (see Problem 30).

7. Urysohn Metrization Theorem: *Every normal topological space satisfying the second axiom of countability is metrizable.*

Problems

13. 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 $O_2 = \{x: \rho(x, F_2) < \rho(x, F_1)\}$.]
14. Let X consist of the numbers $[0, 1]$ and an element $0'$, and take as a basis for a topology the sets (α, β) , $[0, \beta)$, $(\alpha, 1]$, and $\{0'\} \cup (0, \beta)$. Show that these are a base for a topology on X which is T_1 but not Hausdorff.
15. 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.
16. 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 $(f \vee g)(x) = \max f(x), g(x)$.)
17. 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.
18. 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 $\overline{U} \subset O$.
- b. Let F be a closed subset of a normal space contained in an open set O . By repeating the result in (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 $\overline{U}_r \subset U_s$ 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, $0 \leq f \leq 1$, with $f \equiv 0$ on F and $f \equiv 1$ on \bar{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 $0 \leq f \leq 1$, $f \equiv 0$ on A and $f \equiv 1$ on B .

19. Prove Tietze's Extension Theorem by using the following steps:

a. Let $h = f/(1 + |f|)$. Then $|h| < 1$.

b. Let $B = \{x: h(x) \leq -\frac{1}{3}\}$, $C = \{x: h(x) \geq \frac{1}{3}\}$. 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 $|h_1(x)| \leq \frac{1}{3}$ 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 $|h_n(x)| < \frac{2^{n-1}}{3^n}$ for all $x \in X$ and $|h(x) - \sum_{i=1}^n h_i(x)| < \frac{2^n}{3^n}$ for all $x \in A$.

d. The sequence $\langle h_n \rangle$ is uniformly summable to a continuous function k on X , $|k| < 1$ and $k = h$ on A . Set $g = k/(1 - |k|)$.

20. Let \mathfrak{F} be a family of real-valued functions on a set X . Show that a base for the weak topology on X generated by \mathfrak{F} is given by the sets of the form $\{x: |f_i(x) - f_i(y)| < \epsilon, \text{ for some } \epsilon > 0, y \in X, \text{ and some finite set } f_1, \dots, f_n \text{ of functions in } \mathfrak{F}\}$. Show that this topology is Hausdorff if and only if given any pair $\{x, y\}$ of distinct points in X there is an $f \in \mathfrak{F}$ with $f(x) \neq f(y)$.

21. Let \mathfrak{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 \mathfrak{F} is \mathfrak{J} if for each closed set F and each $x \notin F$ there is an $f \in \mathfrak{F}$ with $f(x) = 1$ and $f \equiv 0$ on F .

22. Show that every completely regular space is regular.

23. Prove that every subset of a Hausdorff space is Hausdorff.

4 Product Spaces

If $\langle X, \mathfrak{J} \rangle$ and $\langle Y, \mathfrak{S} \rangle$ 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 $O_1 \in \mathfrak{J}$ and $O_2 \in \mathfrak{S}$. 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 $\langle X_\alpha, \mathfrak{J}_\alpha \rangle$ is any indexed family of topological spaces, we

define the product topology on $\prod_{\alpha} X_{\alpha}$ by taking as a base all sets of the form $\prod_{\alpha} O_{\alpha}$ where $O_{\alpha} \in \mathfrak{J}_{\alpha}$ and $O_{\alpha} = X_{\alpha}$ except for a finite number of α . If the X_{α} are all the same space X and indexed by an index set A , we write X^A for $\prod_{\alpha} X_{\alpha}$.

If $\langle X_{\alpha}, \mathfrak{J}_{\alpha} \rangle$ is a collection of topological spaces and Y their product, we define for each α a mapping π_{α} (called a *projection*) of Y into X_{α} by letting $\pi_{\alpha}(x)$ be the α -th coordinate of x . Each π_{α} is continuous, and the product topology in Y is the weakest topology such that each π_{α} 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^{ω} (or X^N) for a countable product. If we denote the discrete space with two elements by $\mathbf{2}$, then $\mathbf{2}^{\omega}$ is homeomorphic to the Cantor set. If we use ω to denote not only a countable set but also a countable set with discrete topology, then ω^{ω} 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^{ω} is metrizable and is called the *Hilbert cube*. Let X be any set and \mathfrak{F} a family of functions f on X with $0 \leq f \leq 1$ and such that for any two distinct points x and y in X there is an $f \in \mathfrak{F}$ with $f(x) \neq f(y)$. Then \mathfrak{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 \mathfrak{F} into I which takes f into $f(x)$ is simply the projection π_x restricted to the image of \mathfrak{F} . The topology that \mathfrak{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 $I^{\mathfrak{F}}$ by letting each x correspond to the element whose f -th coordinate is $f(x)$. The topology of X as a subspace of $I^{\mathfrak{F}}$ is the weak topology generated by \mathfrak{F} . If X is a topological space and each f in \mathfrak{F} is continuous, then the mapping of X onto its image in $I^{\mathfrak{F}}$ is continuous, and, if \mathfrak{F} has the property that for each closed subset $F \subset X$ and each $x \notin F$ there is an $f \in \mathfrak{F}$ with $f(x) = 1$ and $f \equiv 0$ on F , then X is homeomorphic with its image in $I^{\mathfrak{F}}$.

Problems

24. Prove that the direct product of Hausdorff spaces is a Hausdorff space.

25. 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.

26. 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 α in A .

27. 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 ρ . Define a metric σ on X^A by $\sigma(x, y) = \sum_{\alpha \in A} 2^{-n} \rho(x_\alpha, y_\alpha)$.]

28. Show that each projection π_α is continuous and that the product topology on X^A is the weakest topology such that each π_α is continuous.

29. Show that 2^ω is homeomorphic to the Cantor ternary set.

30. 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 (a).

c. Prove the Urysohn Metrization Theorem.

5 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 which 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 , both meeting E such that $E \subset O_1 \cup O_2$ and $E \cap O_1 \cap O_2 = \emptyset$.

8. 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. ■

9. 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. ■

10. Proposition: A subset E of \mathbf{R} is connected if and only if it is either an interval or a single point.

Problems

31. Let A be a connected subset of a topological space, and suppose $A \subset B \subset \bar{A}$. Then B is connected.

32. a. Let E be a connected subset of \mathbf{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 \mathbf{R} is connected. [Let $I = (a, b)$ and let O be a subset of I which is both open and closed in I . Show that $\sup \{y: (x, y) \subset O\} = b$, and use Problem 31 to take care of nonopen intervals.]

33. 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

$$X = \{\langle x, y \rangle: x = 0, -1 \leq y \leq 1\} \cup \{\langle x, y \rangle: y = \sin 1/x, 0 < x \leq 1\}.$$

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 in G which can be connected to x by a polygonal arc. Then H is open and closed in G .]

34. Let X be a topological space and $\{E_\alpha\}$ a collection of connected subsets of X each of which contains a fixed point x . Then $\bigcup E_\alpha$ is connected.

35. Show that the direct product of connected topological spaces is itself connected.

*6 Absolute \mathbb{G}_δ 's

The purpose of this section is to characterize those topological spaces which are homeomorphic to complete metric spaces: They are the metrizable spaces whose dense homeomorphic images in every Hausdorff space are \mathbb{G}_δ 's. We state this property as a theorem. Some suggestions for its proof are given in Problems 37 and 38 and an application is given by Problem 39. This characterization is utilized in Chapter 15 to show that in a sense the complete metric spaces are absolutely measurable.

11. Theorem: *Let X be a complete metric space and φ a homeomorphism of X onto a dense subset E of a Hausdorff space F . Then E is a \mathbb{G}_δ . Conversely, if E is a \mathbb{G}_δ in a complete metric space, then E is homeomorphic to a complete metric space.*

Problems

36. Show that if the Hausdorff space F of Theorem 11 is metrizable, then we do not need to require that E be dense.

37. Let E be a dense subset of a Hausdorff space F and g a homeomorphism of E onto a complete metric space X . Then E is a \mathbb{G}_δ . (For each positive integer n , let O_n be the subset consisting of all $y \in F$ such that some open set containing y has an image under g whose diameter is less than $1/n$. Then O_n is an open set, and the homeomorphism g can be extended to a continuous mapping h of $G = \bigcap_{n=1}^{\infty} O_n$ into X , since X is complete. But $g^{-1} \circ h$ must be the identity, whence $E = G$.)

38. Let (Y, σ) be a complete metric space, and let $E = \bigcap O_n$ be a \mathcal{G}_δ . For each n construct a continuous real-valued function φ_n on Y with $0 \leq \varphi_n \leq 1$ such that $O_n = \{x: \varphi_n(x) > 0\}$. For x and y in E let

$$\begin{aligned}\rho(x, y) &= \sigma(x, y) + \sum 2^{-n} |\varphi_n(x) - \varphi_n(y)| \\ &\quad \times [| \varphi_n(x) - \varphi_n(y) | + |\varphi_n(x)\varphi_n(y)|]^{-1}.\end{aligned}$$

Then ρ is a metric for E which is equivalent to σ and makes E a complete metric space.

39. Let E be a dense \mathcal{G}_δ in $[0, 1]$, and suppose that \tilde{E} is also dense. Then E is homeomorphic to the product space ω^ω . [Hint: Show that $E = \bigcap O_n$, where $O_{n+1} \subset O_n$, each component of O_n containing infinitely many components of O_{n+1} , and each component of O_n having a diameter in the complete metric for E which tends to zero. Then the natural correspondence between elements of ω^ω and decreasing sequence $\langle I_n \rangle$ of components of O_n is a homeomorphism onto E .]

*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_α for the value of the net at α and $\langle x_\alpha \rangle$ for the net itself. A point $x \in X$ is said to be the *limit* of a net $\langle x_\alpha \rangle$ if for each open set O containing x there is an $\alpha_0 \in A$ such that $x_\alpha \in O$ for all $\alpha > \alpha_0$. A point x is called a *cluster point* of $\langle x_\alpha \rangle$ if given O containing x and given $\alpha \in A$ there is a $\beta > \alpha$ such that $x_\beta \in O$. For sequences these notions coincide with our earlier notions of limit and cluster point.

12. 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 $O_1 < O_2$ if $O_1 \supset O_2$. Since x is a point of closure of E , for each $O \in A$ 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 $x_{O'} \in O$ for all $O' > O$. ■

Problems

40. 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 which cannot be separated and let the directed system be the collection of all pairs $\langle A, B \rangle$ of open sets with $x \in A, y \in B$. Choose $x_{\langle A, B \rangle}$ to be in $A \cap B$ and show that both x and y are limits of this net.]

41. Prove that a function f from X to Y is continuous at x if and only if for each net $\langle x_\alpha \rangle$ which converges to x the net $\langle f(x_\alpha) \rangle$ converges to $f(x)$.

42. 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 $F < G$ meaning $F \subset G$. For each $F \in A$, let $y_F = \sum_{x \in F} f(x)$. 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, $\lim y_F = \sum_{n=1}^{\infty} f(x_n)$.

43. 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_β converges to the corresponding coordinate of x .

9 Compact Spaces

1 Basic Properties

Many of the important properties of the interval $[0, 1]$ follow from the Heine-Borel theorem. We introduce a class of topological 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 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 there is 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 topo-

logical 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 \mathfrak{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 \mathfrak{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 \mathfrak{U} an open covering for F . Then $\mathfrak{U} \cup \{\tilde{F}\}$ is an open covering for X and so must have a finite subcovering $\{\tilde{F}, O_1, \dots, O_N\}$. Then the sets O_1, O_2, \dots, O_N cover F , and so \mathfrak{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 $\{O_{x_1}, O_{x_2}, \dots, O_{x_n}\}$ 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 which maps the compact set K onto a topological space Y . If \mathfrak{U} is an open covering for Y , then the collection of sets $f^{-1}[O]$ for all $O \in \mathfrak{U}$ is an open covering of K . By the compactness of K , there are a finite number O_1, \dots, O_n of sets of \mathfrak{U} such that the sets $f^{-1}[O_i]$ cover K . Since f is onto, the sets O_1, \dots, 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. ■

Problems

1. a. Prove that a compact Hausdorff space is regular.
b. Prove that a compact Hausdorff space is normal.
2. 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.
3. a. Prove that if (X, \mathfrak{J}) is a compact space then (X, \mathfrak{J}_1) is compact for all \mathfrak{J}_1 weaker than \mathfrak{J} .
b. Show that if (X, \mathfrak{J}) is a Hausdorff space then (X, \mathfrak{J}_2) is a Hausdorff space for all \mathfrak{J}_2 stronger than \mathfrak{J} .
c. Show that if (X, \mathfrak{J}) is a compact Hausdorff space then any weaker topology is not Hausdorff and any stronger topology is not compact.

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. Since every space satisfying the second axiom of countability has the property that every open

covering of it has a countable subcovering, 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 $\{\overline{B}_n\}$ is a countable collection of closed sets with the finite intersection property, and so there is a point x which belongs to $\bigcap \overline{B}_n$. The point x is a cluster point of the sequence, since given N and any open set O containing x we have $x \in \overline{B}_N$, and so there must be an $x_n \in O$ with $n \geq N$. ■

A somewhat old-fashioned concept which resembles the Bolzano-Weierstrass property is that of sequential compactness. A space X is said to be **sequentially compact** if every infinite sequence from X has a convergent subsequence. The relationship of sequential compactness to countable compactness is given by the following proposition.

Problem 6 gives an example of a space which is sequentially compact but not compact, and Problem 27 an example of a compact space which is not sequentially compact. Problem 7 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 8.8c. ■

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 α 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 $\{O_1, \dots, O_N\}$. 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 semi-continuous 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 $f_n(x) \rightarrow 0$ for each x , we have $X \subset \bigcup O_n$. By the countable compactness of X , there are a finite number of open sets $\{O_1, \dots, O_N\}$ 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

4. 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.

5. 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.

6. Let X be the set of ordinals less than the first uncountable ordinal, and let \mathcal{G} be the collection of sets of the form $\{x: x < a\}$, $\{x: a < x < b\}$, and $\{x: a < x\}$.

a. Show that \mathcal{G} 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 $\underline{f(x)} = f(x_0)$ for all $x \geq x_0$. [Hint: Show that the set of x for which $f(x) < \overline{\lim} f$ is countable.]
7. Let Y be the set of ordinals less than or equal to the first uncountable ordinal Ω , and let \mathcal{G} be the collection of sets of the form $\{x: x < a\}$, $\{x: a < x < b\}$, and $\{x: a < x\}$.
- Show that \mathcal{G} is a base for a topology for X .
 - Show that X is compact but neither separable nor first countable.

3 Compact Metric Spaces

In this section we investigate special properties of compactness that are true in metric spaces. We shall show that in a metric space the notions of compactness, countable compactness, and sequential compactness coincide.

12. Lemma: *Let X be a countably compact metric space. Then given $\epsilon > 0$ there are a finite number of points x_1, \dots, x_N of X such that for each $x \in X$ there is an x_k with $\rho(x, x_k) < \epsilon$.*

Proof: Suppose that no such finite collection of points exists. Then we can choose an infinite sequence $\langle x_n \rangle$ of points in X so that $\rho(x_n, x_m) \geq \epsilon$ for $m \neq n$. Since each spheroid of radius $\epsilon/3$ can contain at most one term of this sequence, the Bolzano-Weierstrass property does not hold, and X is not countably compact. ■

13. Proposition: *A countably compact metric space is separable.*

Proof: For each positive integer n , let F_n be a finite set of points such that for each $x \in X$ there is a $y \in F_n$ with $\rho(x, y) < 1/n$. Then $\bigcup_{n=1}^{\infty} F_n$ is a countable dense subset of X . ■

14. Corollary: *For a metric space the notions of compactness, countable compactness, and sequential compactness are equivalent.*

Proof: Since a metric space satisfies the first axiom of countability, the notions of countable compactness and sequential compactness

coincide by Proposition 8. Compactness trivially implies countable compactness. If X is countably compact, it must satisfy the second axiom of countability by Proposition 13, and so be compact. ■

A metric space X is said to be **totally bounded** if for each $\epsilon > 0$ there is a finite collection of spheroids of radius ϵ which covers X . Every compact space is totally bounded, and every subset of a totally bounded space is totally bounded. The following proposition characterizes compactness in terms of completeness and total boundedness:

15. 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 which 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 $S_1 \cap \dots \cap S_k$ contains infinitely many terms of the sequence. Since there are infinitely many terms of the sequence in $S_1 \cap \dots \cap S_k$, we may choose n_k so that $n_k > n_{k-1}$ and $x_{n_k} \in S_1 \cap \dots \cap S_k$. Then $\langle x_{n_k} \rangle$ is a subsequence of $\langle x_n \rangle$, and it must be a Cauchy sequence, since $\rho(x_{n_k}, x_{n_l}) \leq 2/N$ for $k, l \geq N$. Since X is complete, this subsequence converges. ■

16. 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 $\sigma(f(x), f(y)) < \epsilon/2$. 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 $\{O_{x_1}, \dots, O_{x_n}\}$. Let $\delta = \frac{1}{2} \min \{\delta_{x_1}, \dots, \delta_{x_n}\}$. 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, $\rho(z, x_i) \leq \rho(z, y) + \rho(y, x_i) < \frac{1}{2}\delta_{x_i} + \delta \leq \delta_{x_i}$. 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

- 8.** 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., iff there is a $\delta > 0$ such that $\rho(x, y) > \delta$ for all $x \in F$ and all $y \in K$. [Consider the function $\rho(x, F) = \inf_{y \in F} \rho(x, y)$.]
- 9.** Let X be a compact metric space and \mathcal{U} an open covering of X . Then there is a $\delta > 0$ such that every spheroid of radius δ is contained in some element of \mathcal{U} .

4 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.

17. Lemma: *Let \mathcal{Q} be a collection of subsets of a set X , and suppose that \mathcal{Q} has the finite intersection property. Then there is a collection $\mathcal{G} \supset \mathcal{Q}$ such that \mathcal{G} has the finite intersection property and is maximal with respect to this property; that is, no collection properly containing \mathcal{G} has the finite intersection property.*

Proof: Consider the family of all collections containing \mathcal{Q} 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{G} be the union of the

collections in \mathfrak{F} . If B_1, \dots, B_n are in \mathfrak{G} , then each B_i belongs to some $\mathfrak{C}_i \in \mathfrak{F}$. Since \mathfrak{F} is linearly ordered by inclusion, one of the collections \mathfrak{C}_k contains the others, and so all B_i belong to \mathfrak{C}_k . Since \mathfrak{C}_k has the finite intersection property, $\bigcap B_i \neq \emptyset$. Thus \mathfrak{G} has the finite intersection property. If $\mathfrak{G}' \supset \mathfrak{G}$ and \mathfrak{G}' has the finite intersection property, then \mathfrak{G}' contains every \mathfrak{C} in \mathfrak{F} and so must belong to \mathfrak{F} by the maximality of \mathfrak{F} . Thus \mathfrak{G} is a union of collections one of which is \mathfrak{G}' , and so $\mathfrak{G}' \subset \mathfrak{G}$. This shows that \mathfrak{G} is maximal with respect to the finite intersection property. ■

18. Lemma: *Let \mathfrak{G} be a collection of subsets of X which is maximal with respect to the finite intersection property. Then each intersection of a finite number of sets in \mathfrak{G} is again in \mathfrak{G} , and each set which meets each set in \mathfrak{G} is itself in \mathfrak{G} .*

Proof: Let \mathfrak{G}' be the collection of all sets which are finite intersections of sets in \mathfrak{G} . Then \mathfrak{G}' is a collection having the finite intersection property and containing \mathfrak{G} . Thus $\mathfrak{G}' = \mathfrak{G}$ by the maximality of \mathfrak{G} .

Suppose that a set C meets each element of \mathfrak{G} . Since \mathfrak{G} contains each finite intersection of sets in \mathfrak{G} , it follows that $\mathfrak{G} \cup \{C\}$ has the finite intersection property. By maximality $\mathfrak{G} \cup \{C\} = \mathfrak{G}$, and so $C \in \mathfrak{G}$. ■

19. Theorem (Tychonoff): *Let $\langle X_\alpha \rangle$ be an indexed family of compact topological spaces. Then the product space $\prod_\alpha X_\alpha$ is compact in the product topology.*

Proof: Let π_α be the mapping of X to X_α which assigns to each $x \in X$ its α -th coordinate. Then the sets obtained by taking finite intersections of sets of the form $\pi_\alpha^{-1}[O_\alpha]$ with O_α open in X_α form a base \mathfrak{N} for the topology of X .

Let \mathfrak{A} be any collection of closed subsets of X with the finite intersection property, and let \mathfrak{G} be a collection of (not necessarily closed) sets which contains \mathfrak{A} and is maximal with respect to the finite intersection property. Let \mathfrak{G}_α be the collection of subsets of X_α of the form $\pi_\alpha(B)$ with $B \in \mathfrak{G}$. Then \mathfrak{G}_α has the finite intersection property, and by the compactness of X_α it is possible to choose a point x_α belonging to $\bigcap_{\mathfrak{G}} \overline{[\pi_\alpha(B)]}$, that is, a point x_α which is a point of

closure of each set $\pi_\alpha[B]$. Let x be that point in X whose α -th coordinate is x_α .

Consider a set S which is of the form $\pi_\alpha^{-1}[O_\alpha]$ for some α and some open set O_α in X_α with $x_\alpha \in O_\alpha$. Since x_α is a point of closure of $\pi_\alpha[B]$ for each B in \mathfrak{B} , the set S must intersect each B in \mathfrak{B} . By Lemma 18 we must have $S \in \mathfrak{B}$. Each set containing x in the base \mathfrak{N} for the topology of X is a finite intersection of sets of this form and so must be in \mathfrak{B} by Lemma 18. Let F be a closed set in \mathfrak{B} . Then F meets each $N \in \mathfrak{N}$ 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 \mathfrak{C} , and \mathfrak{C} has nonempty intersection. ■

Problems

10. Each closed and bounded set in \mathbf{R}^n is compact.
11. 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 \mathfrak{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 \mathfrak{U} . Use the compactness of X to show that $X \times [0, t]$ can also be covered by a finite number of sets in \mathfrak{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 \mathfrak{U} .]
12. 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.]
13. 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 $Q = \prod_{f \in \mathcal{F}} I_f$. Then the mapping g of X into Q which takes x into the point whose f -th coordinate is $f(x)$ is one-to-one into Q and continuous.]
14. 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 ϵ and look at the inverse images of these intervals.]

5 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 \overline{O} is compact. Thus X is locally compact iff the collection of open sets with compact closures form 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.

If X is a locally compact Hausdorff space, we can form a new space X^* by adding to X a single point ω 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 ω is often referred to as the point at infinity in X^* .

Some useful properties of compact subsets in a locally compact Hausdorff space are given by the following propositions, whose proofs are left to the reader (see Problems 16, 17, and 18).

20. Proposition: *Let K be a compact subset of a locally compact Hausdorff space X . Then there is an open set O containing K with \overline{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 \mathcal{G}_δ , we may take $f < 1$ in \tilde{K} .*

21. Proposition: *Let K be a compact subset of a locally compact Hausdorff space and $\{O_\alpha\}$ an open covering for K . Then there are a finite number of nonnegative continuous functions f_1, \dots, f_n on X , each f_i vanishing outside a compact set and outside some O_α , such that $f_1 + f_2 + \dots + f_n$ is identically 1 on K .*

Problems

15. a. 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.

16. 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 \bar{O} is compact. [Hint: For each point $x \in K$ there is an O_x containing x with \bar{O}_x compact. Let O be the union of a finite number of these O_x which cover K .]

17. Let X be a locally compact Hausdorff space and K a compact set. Then there is a continuous real-valued function on X which is identically 1 on K and for which the set $O = \{x: f(x) \neq 0\}$ has compact closure. [Hint: Use Problem 16 and the Urysohn Lemma.] Prove Proposition 20.

18. Prove Proposition 21. [For each x in K there is a nonnegative continuous function g which is positive at x and vanishes identically outside some O_α . Choose a finite number g_1, \dots, g_n of such functions so that $g = g_1 + \dots + g_n$ is positive on K . Let h be a continuous function on X which is equal to $1/g$ on K , and set $f_i = hg_i$.]

19. Show that the Alexandroff one-point compactification of \mathbf{R}^n is homeomorphic to the boundary of a sphere in \mathbf{R}^{n+1} .

20. Show that the one-point compactification of the space X in Problem 6 is the space Y in Problem 7.

21. 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 ω is continuous.

22. A continuous mapping f from a topological space X to a topological space Y is called **proper** if the inverse image of every compact set is compact. 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.

23. 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.

24. 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

$$V_\epsilon = \{l: |m(l) - m_0| < \epsilon, b(l) = b_0\} \\ \cup \{\langle x, y \rangle: |(y - b_0)/x \pm m_0| < \epsilon, x < \epsilon\}.$$

- 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.

*6 The Stone-Čech Compactification

Let X be a completely regular topological space and \mathfrak{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 8.30 that X is homeomorphic with a set $E \subset I^{\mathfrak{F}}$. Let $F = \overline{E}$. Since $I^{\mathfrak{F}}$ 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.

22. 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 open 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 φ of $\beta(X)$ onto Y such that $\varphi(x) = x$ for all $x \in X$.*

Problems

25. Prove Proposition 22:

- a. If f is a bounded continuous real-valued function on X with $|f| \leq 1$, then f is the restriction to X of π_f , and π_f is continuous on $\beta(X)$.
- b. Use the fact that Y is a subset of I^G , where 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 φ of Z with $\beta(X)$ such that $\varphi(x) = x$ for all $x \in X$.

26. Let X and Y be the spaces in Problems 6 and 7. Show that $\beta(X) = Y$.

27. 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.

7 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 $C(X)$ which 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:

23. Proposition: Let L be a lattice of continuous real-valued functions on a compact space X , and suppose that the function h defined by

$$h(x) = \inf_{f \in L} f(x)$$

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 $y \in O_x$. 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 $\{O_{x_1}, \dots, O_{x_n}\}$, which cover X . Let $g = f_{x_1} \wedge f_{x_2} \wedge \dots \wedge f_{x_n}$. Then $g \in L$, and given y in X we may choose i so that $y \in O_{x_i}$, whence

$$g(y) - h(y) \leq f_{x_i}(y) - h(y) < \epsilon. \blacksquare$$

24. 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 $f \in L$ with $f(x) \neq f(y)$.
- ii. If $f \in L$, 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 $g \in L$ such that for all $x \in X$

$$0 \leq g(x) - h(x) < \epsilon.$$

Before proving the proposition, we first establish two lemmas.

25. Lemma: Let L be a family of real-valued functions on a compact space X which satisfies properties (i) and (ii) of Proposition 24. Then given any two real numbers a and b and any pair x and y of distinct points of X , there is an $f \in L$ 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 $f(x) = a, f(y) = b$. ■

26. Lemma: Let L be as in Proposition 24, 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 25 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 $\{O_{x_1}, \dots, O_{x_n}\}$ which cover F . Let $f = f_{x_1} \vee \dots \vee f_{x_n}$. 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 24: 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 24 will follow from Proposition 23 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 η . 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 26 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 η was an arbitrary positive number, we have $g(p) = \inf f(p)$, $f \in L'$. ■

27. 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 $s \in [-1, 1]$. ■

28. Theorem (Stone-Weierstrass): *Let X be a compact space and A an algebra of continuous real-valued functions on X which separates the points of X and which 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 which 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 24 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 27. 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 $|f| \in \bar{A}$. 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

$$f \vee g = \frac{1}{2}(f + g) + \frac{1}{2}|f - g|$$

and

$$f \wedge g = \frac{1}{2}(f + g) - \frac{1}{2}|f - g|.$$

Thus \bar{A} is a lattice and must be $C(X)$ by Proposition 24. ■

29. 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 28 applies. ■

Problems

28. Let f be a continuous periodic real-valued function on \mathbf{R} with period 2π ; that is, $f(x + 2\pi) = f(x)$. Show that, given $\epsilon > 0$, there is a finite Fourier series φ , 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.]

29. 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 $\overline{A} = C(X)$ or there is a point $p \in X$ and $\overline{A} = \{f : f \in C(X), f(p) = 0\}$.

30. Let \mathfrak{F} be a family of continuous real-valued functions on a compact Hausdorff space X , and suppose that \mathfrak{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 \mathfrak{F} .

31. 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 φ 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 φ is continuous.

d. If X is compact, then so is \tilde{X} , and the functions \tilde{f} in (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 φ of X onto \tilde{X} such that A is the set of all functions f of the form $\tilde{f} \circ \varphi$ with $\tilde{f} \in C(\tilde{X})$.

32. 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, \dots, g_n on X and h_1, \dots, h_n on Y such that for each $\langle x, y \rangle \in X \times Y$ we have

$$|f(x, y) - \sum_{i=1}^n g_i(x)h_i(y)| < \epsilon.$$

33. The Weierstrass theorem, which states that a continuous function on a cube in \mathbf{R}^n can be uniformly approximated by a polynomial, can be proved directly by giving an integral formula for the approximating

polynomials.¹ Show that this special case implies the general Stone-Weierstrass Theorem by showing that the functions of norm 1 in the algebra \mathcal{A} give a mapping of X into the infinite-dimensional cube $\mathbf{X} \{I_f : f \in \mathcal{A}, \|f\| = 1\}$. Use the Tietze Extension Theorem and Problem 14 to show that each continuous function 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.

34. 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 $f(1) \equiv f(-1) \pmod{2}$. 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 φ be the polynomial defined by $\varphi(x) = x + x(1 - 2x)(1 - x)$. Then φ is a monotone increasing function whose fixed points are 0, $\frac{1}{2}$, and 1.

b. Choose $\epsilon > 0$. Then some iterate φ_n of φ 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 α , $0 < \alpha < 1$, and any $\epsilon > 0$, then there is a polynomial ψ 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 β be any real number. Then for each $\epsilon > 0$, βP can be uniformly approximated to within ϵ 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.

35. 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 which 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

¹ See, for example, R. Courant and D. Hilbert, *Mathematische Physik*, Bd. I, pp. 55–57, Springer, Berlin, 1931.

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 .

36. a. The statement of Problem 34 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 35b.]

b. Show that we cannot take I in Problem 34 to be $[-2, 2]$. [Hint: If P is a polynomial with integral coefficients,

$$\frac{1}{\pi} \int_{-2}^2 P(x)(4 - x^2)^{1/2} dx \quad \text{is an integer.}]$$

*8 The Ascoli Theorem

It is often useful in analysis to know conditions under which a sequence of functions has a subsequence which is convergent in some sense. The following notion plays a central role in such questions: A family \mathfrak{F} of functions from a topological 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.

30. Lemma: Let $\langle f_n \rangle$ be a sequence of mappings of a countable set D into a topological space Y such that for each $x \in D$ the closure of the set $\{f_n(x): 0 \leq n < \infty\}$ is sequentially compact. Then there is a subsequence $\langle f_{n_k} \rangle$ which converges for each x in D .

Proof: Let $D = \{x_k\}$. By the sequential compactness of the closure of $\{f_n(x_1): 0 \leq n < \infty\}$ 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 $\langle f_{nn} \rangle_{n=1}^\infty$ a subsequence of $\langle f_{jn} \rangle$, and so $\langle f_{nn}(x_j) \rangle$ converges. ■

31. Lemma: Let $\langle f_n \rangle$ be an equicontinuous sequence of mappings from a topological 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 . ■

32. Lemma: Let K be a compact topological space and $\langle f_n \rangle$ an equicontinuous sequence of functions to a metric space Y which converge 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 $\{O_{x_1}, \dots, O_{x_k}\}$ 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

$$\begin{aligned}\sigma[f_n(y), f(y)] &\leq \sigma[f_n(y), f_n(x_i)] + \sigma[f_n(x_i), f(x_i)] + \sigma[f(x_i), f(y)] \\ &< \epsilon\end{aligned}$$

for $n \geq N$. Thus $\langle f_n \rangle$ converges to f uniformly on K . ■

These three lemmas taken together imply the following theorem:

33. Theorem (Ascoli): Let \mathfrak{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 \mathfrak{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 $\langle f_{n_k} \rangle$ which converges pointwise to a continuous function f , and the convergence is uniform on each compact subset of X .

34. Corollary: Let \mathfrak{F} be an equicontinuous family of real-valued functions on a separable space X . Then each sequence $\langle f_n \rangle$ in \mathfrak{F} which is bounded at each point (of a dense subset) has a subsequence $\langle f_{n_k} \rangle$ which converges pointwise to a continuous function, the convergence being uniform on each compact subset of X .

Problems

37. The set of all functions from X to Y is precisely the product of X copies of Y and is often denoted by Y^X . The product topology in Y^X is called the *topology of pointwise convergence*.

a. Show that a sequence $\langle f_n \rangle$ of mappings of X into Y converges to f in the topology of pointwise convergence if and only if $\langle f_n(x) \rangle$ converges to $f(x)$ for each x in X .

b. Let \mathfrak{F} be an equicontinuous family of functions from a topological space X to a metric space Y . Then the closure $\bar{\mathfrak{F}}$ of \mathfrak{F} in the topology of pointwise convergence is also equicontinuous.

38. Let X be a space which is either locally compact, or Hausdorff and satisfies the first axiom of countability. 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. [Hint: Use Problem 23.]

39. 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 \overline{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

$$\sigma_n^*(f, g) = \sup_{\overline{O}_n} \frac{\sigma[f(x), g(x)]}{1 + \sigma[f(x), g(x)]}.$$

c. The sequence $\langle f_n \rangle$ converges to f uniformly on compact subsets of X if and only if $\sigma^*(f, f_n) \rightarrow 0$.

40. Let X and Y be metric spaces. A sequence $\langle f_n \rangle$ of mappings from X to Y is said to converge continuously to f at x if for each sequence $\langle x_n \rangle$ such that $x = \lim x_n$ we have $f(x) = \lim f_n(x_n)$. We say $\langle f_n \rangle$ converges continuously to f if it converges continuously at each x in X .

a. Show that if each f_n is continuous on X and if $\langle f_n \rangle$ converges continuously to f at each $x \in X$, then f is continuous.

b. Show that $\langle f_n \rangle$ converges continuously to f if and only if $\langle f_n \rangle$ converges to f uniformly on each compact subset of X .

41. A real-valued function f on $[0, 1]$ is said to be Hölder continuous of order α if there is a constant C such that $|f(x) - f(y)| \leq C|x - y|^\alpha$.

Define $\|f\|_\alpha = \max |f(x)| + \sup \frac{|f(x) - f(y)|}{|x - y|^\alpha}$.

Show that for $0 < \alpha \leq 1$, the set of functions with $\|f\|_\alpha \leq 1$ is a compact subset of $C[0, 1]$.

10 Banach Spaces

1 Introduction

We are going to study a class of spaces which 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 which satisfy the following conditions:

- i. $x + y = y + x.$
- ii. $(x + y) + z = x + (y + z).$
- iii. There is a vector θ 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 θ defined in (iii) is unique, for if θ' 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. $\|x + y\| \leq \|x\| + \|y\|.$
- iii. $\|\alpha x\| = |\alpha| \|x\|.$

A normed vector space becomes a metric space if we define a metric ρ 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 6.4, 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 λA is the set of all elements of the form λx 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 $x_n \rightarrow x$, then $\|x_n\| \rightarrow \|x\|$.
2. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are called equivalent if there is a positive constant K such that $K^{-1}\|x\|_1 \leq \|x\|_2 \leq K\|x\|_1$. 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 λ .
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, \dots, x_n such that $M = \{x_1, \dots, x_n\}$. 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 θ ; that is, $S = \{x: \|x\| < 1\}$. Prove that S is open and that
- $$\overline{S} = \{x: \|x\| \leq 1\}.$$
- We call S the open unit sphere (or ball) and \overline{S} the unit sphere (or ball).
10. A nonnegative real-valued function $\|\cdot\|$ 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 φ of X onto X' which 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 φ ? 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 φ of X onto X' has kernel \overline{M} . Prove that φ takes open sets into open sets. The space X' is usually denoted by X/\overline{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 α_1, α_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

$$\|A\| = \sup_{\substack{x \in X \\ x \neq 0}} \frac{\|Ax\|}{\|x\|}.$$

Since $A(\alpha x) = \alpha Ax$, 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 boundedness 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 which 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 , 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 \mathcal{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 \mathcal{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 θ . Thus $\|\cdot\|$ satisfies all the requirements of a norm and we have only to show that if Y is complete so is \mathcal{B} .

Let $\langle A_n \rangle$ be a Cauchy sequence from \mathcal{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

$$\begin{aligned} \|A_n x - Ax\| &= \lim_{m \rightarrow \infty} \|A_n x - A_m x\| \\ &\leq \overline{\lim}_{m \rightarrow \infty} \|A_n - A_m\| \|x\| \\ &\leq \epsilon \|x\| \end{aligned}$$

for all $n \geq N$. Thus for $n \geq N$,

$$\|A_n - A\| = \sup_{\|x\|=1} \|(A_n - A)x\| \leq \epsilon.$$

Thus $A_n \rightarrow A$ and \mathfrak{G} is complete. ■

Problems

- 13.** Show that if $A_n \rightarrow A$ and $x_n \rightarrow x$, then $A_n x_n \rightarrow Ax$.
- 14.** The *kernel* of an operator A is the set $\{x: Ax = 0\}$. 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 φ 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 which 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 from 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 $p(x + y) \leq p(x) + p(y)$ and $p(\alpha x) = \alpha p(x)$ for each $\alpha \geq 0$. Suppose that f is a linear functional defined on a subspace S and that $f(s) \leq p(s)$ for all s in S . Then there is a linear functional F defined on X such that $F(x) \leq p(x)$ 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 $g(x) \leq p(x)$ whenever $g(x)$ is defined. This set is partially ordered by setting $g_1 < g_2$ 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 $\{g_\alpha\}$ which contains the given functional f . We define a functional F on the union of the domains of the g_α by setting $F(x) = g_\alpha(x)$ if x is in the domain of g_α . Since $\{g_\alpha\}$ is linearly ordered, this definition is independent of the g_α 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_α and $y \in$ domain g_β for some α, β . By the linear ordering of $\{g_\alpha\}$, we have either $g_\alpha < g_\beta$ or $g_\beta < g_\alpha$, say the former. Then x and y are in the domain of g_β , and so $\lambda x + \mu y$ is in the domain of g_β 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 , $g_\alpha < F < G$ implies that G must belong to $\{g_\alpha\}$ by the maximality of $\{g_\alpha\}$. Hence $G < F$, 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 which is defined on a proper subspace T of X and satisfies $g(t) \leq p(t)$ 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

$$g(t_1) + g(t_2) = g(t_1 + t_2) \leq p(t_1 + t_2) \leq p(t_1 - y) + p(t_2 + y).$$

Hence

$$-p(t_1 - y) + g(t_1) \leq p(t_2 + y) - g(t_2),$$

and so

$$\sup_{t \in T} [-p(t - y) + g(t)] \leq \inf_{t \in T} [p(t + y) - g(t)].$$

Define $h(y) = \alpha$, where α is a real number such that

$$\sup [-p(t - y) + g(t)] \leq \alpha \leq \inf [p(t + y) - g(t)].$$

We must now show that

$$h(\lambda y + t) = \lambda\alpha + g(t) \leq p(\lambda y + t).$$

If $\lambda > 0$, then

$$\begin{aligned} \lambda\alpha + g(t) &= \lambda[\alpha + g(t/\lambda)] \\ &\leq \lambda[\{p(t/\lambda + y) - g(t/\lambda)\} + g(t/\lambda)] \\ &= \lambda p(t/\lambda + y) = p(t + \lambda y). \end{aligned}$$

If $\lambda = -\mu < 0$, then

$$\begin{aligned} -\alpha\mu + g(t) &= \mu(-\alpha + g(t/\mu)) \\ &\leq \mu(\{p(t/\mu - y) - g(t/\mu)\} + g(t/\mu)) \\ &= \mu p(t/\mu - y) = p(t - \mu y). \end{aligned}$$

Thus $h(\lambda y + t) \leq p(\lambda y + t)$ for all λ , 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 $p(Ax) \leq p(x)$ 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 $F(x) \leq p(x)$ and $F(Ax) = F(x)$ for all x in X .

Proof: Define a function q on X by setting

$$q(x) = \inf \frac{1}{n} p(A_1x + \cdots + A_nx),$$

where the inf is taken over all finite sequences $\langle A_1, \dots, A_n \rangle$ from G . We clearly have $q(x) \leq p(x)$ and $q(\alpha x) = \alpha q(x)$ for $\alpha \geq 0$. For any x and y in X and any $\epsilon > 0$, we can choose $\langle A_1, \dots, A_n \rangle$ and $\langle B_1, \dots, B_m \rangle$, so that

$$\frac{1}{n} p(A_1x + \cdots + A_nx) < q(x) + \epsilon$$

and

$$\frac{1}{m} p(B_1y + \cdots + B_my) < q(y) + \epsilon.$$

Then

$$\begin{aligned} q(x+y) &\leq \frac{1}{nm} p\left(\sum_{i=1}^n \sum_{j=1}^m A_i B_j (x+y)\right) \\ &\leq \frac{1}{nm} p\left(\sum_{j=1}^m B_j \left(\sum_{i=1}^n A_i x\right)\right) + \frac{1}{nm} p\left(\sum_{i=1}^n A_i \left(\sum_{j=1}^m B_j y\right)\right) \\ &\leq \frac{1}{n} p\left(\sum_{i=1}^n A_i x\right) + \frac{1}{m} p\left(\sum_{j=1}^m B_j y\right) \\ &< q(x) + q(y) + 2\epsilon. \end{aligned}$$

Since ϵ is arbitrary, $q(x+y) \leq q(x) + q(y)$.

For s in S ,

$$f(s) = \frac{1}{n} f(A_1s + \cdots + A_ns) \leq \frac{1}{n} p(A_1s + \cdots + A_ns).$$

Hence $f(s) \leq q(s)$, and we may apply Theorem 4 with p replaced by q to obtain an extension F of f to all X such that $F(x) \leq q(x) \leq p(x)$. It remains only to show that $F(Ax) = F(x)$. Now

$$\begin{aligned} q(x - Ax) &\leq \frac{1}{n} p((x - Ax) + A(x - Ax) + \cdots + A^n(x - Ax)) \\ &= \frac{1}{n} p(x - A^{n+1}x) \leq \frac{1}{n} [p(x) + p(-x)]. \end{aligned}$$

Thus $q(x - Ax) \leq 0$. Since

$$F(x) - F(Ax) = F(x - Ax) \leq q(x - Ax) \leq 0,$$

we have $F(x) \leq F(Ax)$, 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 δ , 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 $\|f\| \leq 1$. 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 $[0, 1]$ which are bounded by 1, have $x_n(0) = 1$, and are such that $x_n(t) \rightarrow 0$ for all $t \neq 0$. Then for each $y \in L^1$, $\int x_n y \rightarrow 0$, while $F(x_n) = 1$.

If we consider the dual X^{**} of X^* , then to each x in X there corresponds an element φx 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 φ is clearly a linear mapping, φ is an isometric isomorphism of X onto some linear subspace $\varphi[X]$ of X^{**} . The mapping φ 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^∞ which 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^∞ 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 which 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 we see that $h(s) = -g(is)$.

Since $g(s) \leq |f(s)| \leq p(s)$, we can extend g to a functional G on X which 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 $F(ix) = G(ix) - iG(-x) = i[G(x) - iG(ix)]$, we have F linear in the complex sense. For any x , choose ω to be a complex number such that $|\omega| = 1$ and $\omega F(x) = |F(x)|$. Then $|F(x)| = \omega F(x) = F(\omega x) = G(\omega x) \leq p(\omega x) = p(x)$. ■

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 $\inf_{t \in T} \|y - t\| = \sup \{f(y): \|f\| = 1, f(t) = 0 \text{ all } t \in T\}$.

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^∞ be the space of all bounded sequences. Use Proposition 5 to show that there is a linear functional F on l^∞ 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 μ defined for all bounded sets of \mathbf{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 μA 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 functional $y \in X^{***}$ 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 $S^0 = \{y \in X^*: y(s) = 0 \text{ for all } s \in S\}$. If T is a subspace of X^* , we define $T^0 = \{x \in X: t(x) = 0 \text{ for all } t \in T\}$.

a. Show that S^0 is a closed linear subspace of X^* .

b. Show that $S^{00} = \overline{S}$.

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 sphere $S = \{x: \|x\| \leq 1\}$ 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 η 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

$$\left\{ y: \|y\| < \frac{\eta}{2} \right\} \subset A(S_0). \blacksquare$$

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 which is contained in $A[O]$, and so $A[O]$ is open. \blacksquare

11. Proposition: *Let X be a linear vector space which is complete in each of the norms $\|\cdot\|$ and $\|\cdot\|\|$, and suppose there is a constant C such that*

$$\|x\| \leq C\|\cdot\|$$

for all $x \in X$. Then the norms are equivalent. That is to say, there is a second constant C' such that

$$\|\cdot\| \leq C'\|x\|$$

for all $x \in X$.

Proof: The identity map of $(X, \|\cdot\|)$ onto $(X, \|\cdot\|)$ is a one-to-one continuous linear transformation and so must be an isomorphism by Proposition 10. Therefore the inverse mapping must be bounded. \blacksquare

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 which 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 $\|\cdot\|$. For if $\|x_p - x_q\| \rightarrow 0$, then $\|x_p - x_q\| \rightarrow 0$ and $\|Ax_p - Ax_q\| \rightarrow 0$. Hence by the completeness of X and Y we have $x \in X$ and $y \in Y$ such that $\|x_p - x\| \rightarrow 0$ and $\|Ax_p - y\| \rightarrow 0$. By the hypothesis of our theorem $y = Ax$. Hence $\|x_p - x\| \rightarrow 0$, and X is complete with respect to $\|\cdot\|$. 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 $\|Tx\| \leq M_x$ for all T in \mathfrak{F} . Then the operators in \mathfrak{F} are uniformly bounded; that is, there is a constant M such that $\|T\| \leq M$ 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 7.17 an open set O in X on which these functions are uniformly bounded. Thus there is a constant M' such that $\|Tx\| \leq M'$

for all $x \in O$. Let y be a point in O . Since O is open, there is a sphere $S = \{x: \|x - y\| < \delta\}$ of some radius δ centered at y and contained in O . If $\|z\| \leq \delta$, then $Tz = T(y + z) - Ty$ with $y + z$ in $S \subset O$. Hence $\|Tz\| \leq \|T(y + z)\| + \|Ty\| \leq M' + M_y$. Consequently $\|T\| \leq \frac{M' + M_y}{\delta}$ 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 $C[0, 1]$ which 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 $\|f\|_2 \leq \|f\|_\infty$ and $\|f\|_\infty \leq M\|f\|_2$.

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.

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{T} 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{S} is a translation invariant topology on X and \mathfrak{B} a base for \mathfrak{S} at θ , then the sets of the form $x + U$, $U \in \mathfrak{B}$, form a base for \mathfrak{S} at x . Thus it suffices to give a base at θ in order to determine a translation invariant topology. A base at θ is often called a *local base*. The following proposition gives conditions on a set \mathfrak{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 \mathfrak{B} at θ which satisfies the following:*

- i. *If $U, V \in \mathfrak{B}$, then there is a $W \in \mathfrak{B}$ with $W \subset U \cap V$.*
- ii. *If $U \in \mathfrak{B}$ and $x \in U$, there is a $V \in \mathfrak{B}$ such that $x + V \subset U$.*
- iii. *If $U \in \mathfrak{B}$, there is a $V \in \mathfrak{B}$ such that $V + V \subset U$.*
- iv. *If $U \in \mathfrak{B}$ and $x \in X$, there is an $\alpha \in \mathbf{R}$ such that $x \in \alpha U$.*
- v. *If $U \in \mathfrak{B}$ and $0 < |\alpha| \leq 1$, then $\alpha U \subset U$ and $\alpha U \in \mathfrak{B}$.*

Conversely, given a collection \mathfrak{B} of subsets containing θ and satisfying the above conditions, there is a topology for X making X a topological vector space and having \mathfrak{B} as a base at θ . The topology will be Hausdorff if and only if

$$\text{vi. } \bigcap \{U \in \mathfrak{B}\} = \{\theta\}.$$

The proof of the proposition is left to the reader. We note that if X is a normed linear space we may take \mathfrak{B} to be the set of spheres about θ , 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 con-

tinuous at one point. A linear mapping φ of X onto Y is called a (topological) *isomorphism* if φ and φ^{-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 which makes it into a topological vector space is the usual one.

15. Proposition (Tychonoff): *Let X be a finite-dimensional 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{G} of subsets containing θ is a base at θ 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 $\langle 0, 0 \rangle$) from $\mathbf{R} \times X$ to X , then (iv).
 - d. If X is a topological vector space, then the family \mathcal{G} of all open sets U which contain θ and such that $\alpha U \subset U$ for all α with $|\alpha| < 1$ is a local base for the topology and satisfies (v). [If O is any open set containing θ , the continuity of multiplication implies that there is an open set V containing θ 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 α with $|\alpha| < 1$.]
 - e. If \mathcal{G} 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 $\langle \alpha, 0 \rangle$, and use (iii).]
 - f. If X is T_1 , then (vi) holds. If (vi) and (iii) hold, then X is Hausdorff.
- 31. 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 φ 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 φ 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 φ 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 $X \sim \varphi[B]$ open.

c. There is an open set U of X containing θ such that $\alpha U \subset U$ for each $|\alpha| < 1$ and $U \subset X \sim \varphi[B]$.

d. The set U in (c) is contained in $\varphi[B]$, and so φ^{-1} is continuous.

34. Show that a finite-dimensional subspace M of a 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 topological 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 φ is the natural mapping of X on X/M and B is continuous by Problem 35.]

37. Prove that every locally compact topological vector space X is finite-dimensional. [Hint: Let V be a neighborhood of θ with \bar{V} compact and $\alpha V \subset V$ for each α , $|\alpha| < 1$. If we cover \bar{V} by a finite number of translates $x_1 + \frac{1}{3}V, \dots, x_n + \frac{1}{3}V$, then x_1, \dots, x_n are a basis for X .]

*6 Weak Topologies

If X is any vector space and \mathfrak{F} a collection of linear functionals on X , we define the *weak topology* generated by \mathfrak{F} to be the weakest topology such that each f in \mathfrak{F} is continuous (cf. Problem 8.20). This topology is easily seen to be translation invariant, and a base at θ for this topology is given by the sets $\{x: |f_i(x)| < \epsilon, i = 1, \dots, n\}$,

where $\epsilon > 0$ and $\{f_1, \dots, f_n\}$ is a finite subset of \mathfrak{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 $f(x_n) \rightarrow f(x)$ for each $f \in \mathfrak{F}$.

If X is a normed vector space and the functionals in \mathfrak{F} are all continuous (that is, if $\mathfrak{F} \subset X^*$), then the weak topology generated by \mathfrak{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 M is strongly closed and 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 which vanishes on M and does not vanish at x . Now $\{y : f(y) \neq 0\}$ is an open set in the weak topology which 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 φ 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 θ for the weak* topology is given by sets of the form $\{f: |f(x_i)| < \epsilon, i = 1, \dots, n\}$ where $\{x_1, \dots, x_n\}$ 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 sphere $S^* = \{f: \|f\| \leq 1\}$ of X^* is compact in the weak* topology.*

Proof: If $f \in S^*$, then $|f(x)| \leq \|x\|$, and so $f(x) \in [-\|x\|, \|x\|]$. Let $I_x = [-\|x\|, \|x\|]$. Then each $f \in S^*$ corresponds to a point in $P = \bigcup_{x \in X} I_x$, 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 $|g(x)| \leq \|x\|$ for $g \in S^*$, and evaluation at x is a continuous function on P , we have $|f(x)| \leq \|x\|$. Let x, y , and z be three points of X such that $z = \alpha x + \beta y$. For each $\epsilon > 0$ the set $N = \{g \in P: |g(x) - f(x)| < \epsilon, |g(y) - f(y)| < \epsilon, \text{ and } |g(z) - f(z)| < \epsilon\}$ 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 $g(z) = \alpha g(x) + \beta g(y)$. Hence it follows that $|f(z) - \alpha f(x) - \beta f(y)| < \epsilon(1 + |\alpha| + |\beta|)$. This inequality holding for each positive ϵ , 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 $x_n \rightarrow x$ 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 $\xi_{m,n} \rightarrow \xi_m$.
- c. Let $\langle x_n \rangle$ be a sequence in $L^p[0, 1]$, $1 \leq p < \infty$. 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 6.16).

d. 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 $x_n \rightarrow 0$ in the weak topology.

e. Let $\langle x_n \rangle$ be as in (d), 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 which converge in the strong topology.]

f. The set F in (e) has θ as a weak closure point. However, there is no sequence $\langle z_n \rangle$ from F which converges weakly to zero.

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 $f_n \rightarrow f$ weakly in L^2 , then $f_n(y) \rightarrow f(y)$ for each $y \in [0, 1]$.

b. If $f_n \rightarrow f$ weakly in L^2 , then $\|f_n\|_\infty$ is bounded, and hence $f_n \rightarrow f$ 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 $0 \leq \lambda \leq 1$. The set $\{z: z = \lambda x + (1 - \lambda)y \text{ for } 0 \leq \lambda \leq 1\}$ 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 $K_1 \cap K_2$, λK_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 λ with $|\lambda| < \epsilon$. Let K be a convex set which contains θ as an internal point. Then we define the support function p of K (with respect to θ) 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 θ as an internal point, then the support function p has the following properties:

- i. $p(\lambda x) = \lambda p(x)$ for $\lambda \geq 0$.
- ii. $p(x + y) \leq p(x) + p(y)$.
- iii. $\{x : p(x) < 1\} \subset K \subset \{x : p(x) \leq 1\}$.

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 $p(x + y) \leq \lambda + \mu$, and taking infima over all admissible λ and μ we obtain $p(x + y) \leq p(x) + p(y)$. ■

Two convex sets K_1 and K_2 are said to be **separated** by a linear functional f if there is a real number α such that $f(x) \leq \alpha$ on K_1 and $f(x) \geq \alpha$ 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 which 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 θ 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 θ). Then $p(-x_0) \geq 1$. 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 $f(s) \leq p(s)$ 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 $f(x) \leq p(x)$ for all x . Thus if $x \in K$, we have $f(x) \leq 1$.

Let $x \in K_1$ and $y \in K_2$. Then $x - y - x_0 \in K$, and we have

$$f(x) - f(y) - f(x_0) = f(x - y - x_0) \leq 1.$$

Since $f(x_0) = -1$, we have $f(x) \leq f(y)$. This being true for each $x \in K_1$ and each y in K_2 , we have $\sup_{x \in K_1} f(x) \leq \inf_{y \in K_2} f(y)$. 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 which makes X into a locally convex topological vector space:*

- i. *If $N \in \mathfrak{N}$, each point of N is internal.*
- ii. *If N_1 and N_2 are 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 α 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 θ which satisfies these conditions.

It follows from this proposition that the weak topology on a vector space X which 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 θ is not a point of closure of F , there is a convex open set N which contains θ but does not meet F . Let $O = N \cap (-N)$. Then O is an open convex set containing θ and disjoint from F , and $-O = O$. Since θ 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 θ , and so f is continuous at θ . Since f is linear and continuous at θ , 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 θ is an internal point of O , we can choose $\lambda > 0$ so that λx is in O . Then

$$0 < \lambda f(x) = f(\lambda x) \leq \alpha. \blacksquare$$

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 topological 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 s of nonempty supporting sets with S in s . Since K is compact, the intersection T of all members of 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 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 continu-

ous 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 which 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 convex 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 which has an internal point that is not an interior point.

45. Let K be a convex set containing θ , 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 θ such that for each $x \in X$

there is an $\alpha > 0$ with $\alpha x \in V$. Then \mathcal{G} 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 $S = \{x: \|x\| \leq 1\}$.

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 $\sigma(x + y) \leq \sigma(x) + \sigma(y)$. Hence if we define $\rho(x, y) = \sigma(x - y)$, ρ is a metric for X .

b. In this metric $x_n \rightarrow x$ if and only if $x_n \rightarrow x$ 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 $x_n \rightarrow x$ and $\lambda_n \rightarrow \lambda$ then $\lambda_n x_n \rightarrow \lambda x$. 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 which 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 properties¹:

- 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

$$(x, y) \leq \|x\| \cdot \|y\|,$$

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 S in H is

¹We have defined here a real 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'): $(x, y) = \overline{(y, x)}$.

called an *orthogonal system* if any two different elements φ and ψ of \mathcal{S} are orthogonal, that is, $(\varphi, \psi) = 0$. An orthogonal system \mathcal{S} is called *orthonormal* if $\|\varphi\| = 1$ for each φ 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 have²

$$\begin{aligned} 0 \leq \left\| x - \sum_{\nu=1}^n a_\nu \varphi_\nu \right\|^2 &= \|x\|^2 - 2 \sum_{\nu=1}^n a_\nu (x, \varphi_\nu) + \sum_{\nu=1}^n \sum_{\mu=1}^n a_\nu a_\mu (\varphi_\nu, \varphi_\mu) \\ &= \|x\|^2 - \sum_{\nu=1}^n a_\nu^2. \end{aligned}$$

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

$$z_n = \sum_{\nu=1}^n a_\nu \varphi_\nu$$

is a Cauchy sequence, since for $m \geq n$

$$z_m - z_n = \sum_{\nu=n+1}^m a_\nu \varphi_\nu,$$

and we have

$$\|z_m - z_n\|^2 = \sum_{\nu=n+1}^m a_\nu^2,$$

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 $N \leq n \leq \infty$.

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 $y = \sum_{\nu=1}^{\infty} a_\nu \varphi_\nu$, which 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 ν , then $z = 0$. 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 ν and $z \neq 0$, 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 orthonormal

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 $\dim H = \aleph_0$.)

An isomorphism Φ 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 space is isomorphic to l^2 . Since $L^2[0, \pi]$ is separable and $\{\cos vt\}$ 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 $y \in H$ such that $f(x) = (x, y)$ for all x . Moreover, $\|f\| = \|y\|$.*

Proof: (We consider only the case of separable H . For nonseparable 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

$$\begin{aligned}\sum_{\nu=1}^n b_\nu^2 &= f\left(\sum_{\nu=1}^n b_\nu \varphi_\nu\right) \leq \|f\| \cdot \left\| \sum_{\nu=1}^n b_\nu \varphi_\nu \right\| \\ &\leq \|f\| \left[\sum_{\nu=1}^n b_\nu^2 \right]^{1/2}.\end{aligned}$$

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 $\sum_{\nu=1}^n a_\nu \varphi_\nu \rightarrow x$, 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 $\|f\| \leq \|y\|$. ■

Problems

- 50.** Show that the inner product is continuous; that is, if $x_n \rightarrow x$ and $y_n \rightarrow y$, then $(x_n, y_n) \rightarrow (x, y)$.
- 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. Problems 6.14 and 9.28).
- 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 n of elements in a complete orthonormal system in H is the smallest cardinal n such that there is a dense subset of H with 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 $\langle f_1, \dots, f_n \rangle$ 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 σ -algebra \mathfrak{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 μ 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, \mathfrak{B}) consisting of a set X and a σ -algebra \mathfrak{B} of subsets of X . A subset A of X is called measurable (or measurable with respect to \mathfrak{B}) if $A \in \mathfrak{B}$.

Definition: By a measure μ on a measurable space (X, \mathfrak{B}) we mean a nonnegative set function defined for all sets of \mathfrak{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, \mathfrak{B}, μ) we mean a measurable space (X, \mathfrak{B}) together with a measure μ defined on \mathfrak{B} .

This second property of μ is often referred to by saying that μ is countably additive. We also have that μ 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 \mathfrak{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 $(\mathbf{R}, \mathfrak{B}, m)$, where \mathfrak{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, \mathfrak{B} the family of those subsets which are either countable or the complement of a countable set. Then \mathfrak{B} is a σ -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 $A \in \mathfrak{B}$, $B \in \mathfrak{B}$, and $A \subset B$, then

$$\mu A \leq \mu B.$$

Proof: Since

$$B = A \cup [B \sim A]$$

is a disjoint union, we have

$$\mu B = \mu A + \mu(B \sim A) \geq \mu A. \blacksquare$$

2. Proposition: If $E_i \in \mathfrak{B}$, $\mu E_1 < \infty$ and $E_i \supset E_{i+1}$, then

$$\mu \left(\bigcap_{i=1}^{\infty} E_i \right) = \lim_{n \rightarrow \infty} \mu E_n.$$

¹ A set function μ 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.

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

$$\begin{aligned} \mu E_1 &= \mu E + \sum_{i=1}^{\infty} (\mu E_i - \mu E_{i+1}) \\ &= \mu E + \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} (\mu E_i - \mu E_{i+1}) \\ &= \mu E + \mu E_1 - \lim \mu E_n, \end{aligned}$$

whence the proposition follows. ■

3. Proposition: If $E_i \in \mathfrak{B}$, then

$$\mu \left(\bigcup_{i=1}^{\infty} E_i \right) \leq \sum_{i=1}^{\infty} \mu E_i.$$

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

$$\mu G_n \leq \mu E_n,$$

while

$$\mu \left(\bigcup E_i \right) = \sum_{n=1}^{\infty} \mu G_n \leq \sum_{n=1}^{\infty} \mu E_n. \blacksquare$$

A measure μ is called **finite** if $\mu(X) < \infty$. It is called **σ -finite** if there is a sequence $\langle X_n \rangle$ of sets in \mathfrak{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 σ -finite measure. The counting measure on an uncountable set is a measure which is not σ -finite. Other examples are given in Problem 46.

A set E is said to be of **finite measure** if $E \in \mathfrak{B}$ 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 σ -finite measure is itself of σ -finite measure, and the union of a countable collection of sets of σ -finite measure is again of σ -finite measure. If μ is σ -finite, then every measurable set is of σ -finite measure.

Roughly speaking, nearly all the familiar properties of Lebesgue measure and Lebesgue integration hold for arbitrary σ -finite measures, and many treatments of abstract measure theory limit themselves to σ -finite measures. However, many parts of the general theory do not require the assumption of σ -finiteness, and it seems undesirable to have a development that is unnecessarily restrictive. A notion weaker than that of σ -finiteness is semifiniteness. A measure μ is said to be **semifinite** if each measurable set of infinite measure contains measurable sets of arbitrary large finite measure. Thus every σ -finite measure is semifinite, while the measure which assigns 0 to countable subsets of an uncountable set X and ∞ to the uncountable sets is not semifinite. Measures which 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, take us out of the class of semifinite measures.

A measure space (X, \mathfrak{B}, μ) is said to be **complete** if \mathfrak{B} contains all subsets of sets of measure zero, that is, if $B \in \mathfrak{B}$, $\mu B = 0$, and $A \subset B$ imply $A \in \mathfrak{B}$. Thus Lebesgue measure is complete, while Lebesgue measure restricted to the σ -algebra of Borel sets is not complete. The following proposition, whose proof is left to the reader (Problem

2), shows that each measure space can be completed by the addition of subsets of sets of measure zero. The measure space $(X, \mathcal{G}_0, \mu_0)$ given in the proposition is called the *completion* of (X, \mathcal{G}, μ) .

4. Proposition: *If (X, \mathcal{G}, μ) is a measure space, then we can find a complete measure space $(X, \mathcal{G}_0, \mu_0)$ such that*

- i. $\mathcal{G} \subset \mathcal{G}_0$.
- ii. $E \in \mathcal{G} \Rightarrow \mu E = \mu_0 E$.
- iii. $E \in \mathcal{G}_0 \Leftrightarrow E = A \cup B$ where $B \in \mathcal{G}$ and $A \subset C$, $C \in \mathcal{G}$, $\mu C = 0$.

If (X, \mathcal{G}, μ) is a measure space, we say that a subset E of X is locally measurable if $E \cap B \in \mathcal{G}$ for each $B \in \mathcal{G}$ with $\mu B < \infty$. The collection \mathcal{C} of all locally measurable sets is a σ -algebra containing \mathcal{G} . The measure μ is called *saturated* if every locally measurable set is measurable, i.e., is in \mathcal{G} . Every σ -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 σ -algebra, for we simply take the σ -algebra \mathcal{C} of all locally measurable sets, but the measure μ 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

$$\mu \left(\bigcup_{k=1}^{\infty} A_k \right) = \lim_{n \rightarrow \infty} \mu \left(\bigcup_{k=1}^n A_k \right).$$

2. Let $\{(X_\alpha, \mathcal{G}_\alpha, \mu_\alpha)\}$ 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 $X = \bigcup X_\alpha$, $\mathcal{G} = \{B: (\alpha)[B \cap X_\alpha \in \mathcal{G}_\alpha]\}$, and defining $\mu(B) = \sum \mu_\alpha(B \cap X_\alpha)$.

a. Show that \mathcal{G} is a σ -algebra.

b. Show that μ is a measure.

c. Show that μ is σ -finite if and only if all but a countable number of the μ_α are zero and the remainder are σ -finite.

3. a. Show that $\mu(E_1 \Delta E_2) = 0$ implies $\mu E_1 = \mu E_2$ provided that E_1 and $E_2 \in \mathcal{G}$.

b. Show that if μ is complete, then $E_1 \in \mathcal{G}$ and $\mu(E_1 \Delta E_2) = 0$ imply $E_2 \in \mathcal{G}$.

4. Let (X, \mathcal{B}, μ) be a measure space and $Y \in \mathcal{B}$. Let \mathcal{B}_Y consist of those sets of \mathcal{B} which 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 μ_Y is called the **restriction** of μ to Y .

5. Let (X, \mathcal{B}) be a measurable space.

a. If μ and ν are measures defined on \mathcal{B} , then the set function λ defined on \mathcal{B} by $\lambda E = \mu E + \nu E$ is also a measure. We denote λ by $\mu + \nu$.

b. If μ and ν are measures on \mathcal{B} and $\mu \geq \nu$, then there is a measure λ on \mathcal{B} such that $\mu = \nu + \lambda$.

c. If ν is σ -finite, the measure λ in (b) is unique.

d. Show that in general the measure λ need not be unique but that there is always a smallest such λ .

6. a. Show that each σ -finite measure is semifinite.

b. Show that each measure μ is the sum $\mu_1 + \mu_2$ of a semifinite measure μ_1 and a measure μ_2 which assumes only the values 0 and ∞ . The measure μ_1 is always unique, and there is a smallest μ_2 .

7. Prove Proposition 4. [First show that the family \mathcal{B}_0 defined by (iii) is a σ -algebra. If $E \in \mathcal{B}_0$, show that μA 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 μ_0 and show μ_0 is a measure.]

8. a. Show that each σ -finite measure is saturated.

b. Show that the collection of locally measurable sets is a σ -algebra.

c. Let (X, \mathcal{B}, μ) be a measure space and \mathcal{C} the σ -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 μ is semifinite and $E \in \mathcal{C}$, set $\underline{\mu}E = \sup \{\mu B : B \in \mathcal{B}, B \subset E\}$. Show that $(X, \mathcal{C}, \underline{\mu})$ is a saturated measure space and that $\underline{\mu}$ is an extension of μ .

e. Use Problem 5d to show that, even if μ is not semifinite, there is a smallest extension $\underline{\mu}$ of μ to \mathcal{C} , and that $(X, \mathcal{C}, \underline{\mu})$ is saturated.

f. Give an example to show that $\bar{\mu}$ and $\underline{\mu}$ may be different.

9. **σ -Rings and σ -Algebras.** Some authors prefer to define a measure on a σ -ring \mathfrak{R} , that is, a collection of subsets of X such that if A and B are in \mathfrak{R} so is $A \sim B$, and if $\langle A_n \rangle$ is a sequence from \mathfrak{R} then $\bigcup A_n$ is in \mathfrak{R} . Thus a σ -ring \mathfrak{R} is a σ -algebra if and only if $X \in \mathfrak{R}$. Some of the relations between σ -rings and σ -algebras are given below:

a. Let \mathfrak{R} be a σ -ring which is not a σ -algebra, let \mathcal{B} be the smallest σ -algebra containing \mathfrak{R} , and set $\mathfrak{R}' = \{E : \tilde{E} \in \mathfrak{R}\}$. Then $\mathcal{B} = \mathfrak{R} \cup \mathfrak{R}'$ and $\mathfrak{R} \cap \mathfrak{R}' = \emptyset$.

- b. If μ is a measure on \mathfrak{B} , define $\bar{\mu}$ on \mathfrak{B} by $\bar{\mu}E = \mu E$ if $E \in \mathfrak{B}$ and $\bar{\mu}E = \infty$ if $E \in \mathfrak{B}'$. Then $\bar{\mu}$ is a measure on \mathfrak{B} .
- c. Define $\underline{\mu}$ on \mathfrak{B} by $\underline{\mu}E = \mu A$ if $E \in \mathfrak{B}$ and

$$\underline{\mu}E = \sup \{\mu A : A \subset E, A \in \mathfrak{B}\}$$

if $E \in \mathfrak{B}'$. Then $\underline{\mu}$ is also a measure on \mathfrak{B} .

- d. More generally, for any nonnegative extended real number β define μ_β on \mathfrak{B} by $\mu_\beta E = \mu E$ if $E \in \mathfrak{B}$ and $\mu_\beta E = \underline{\mu}E + \beta$ if $E \in \mathfrak{B}'$. Then μ_β is a measure on \mathfrak{B} .

- e. If ν is any measure on \mathfrak{B} which agrees with μ on \mathfrak{B} , 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 α .
- ii. $\{x : f(x) \leq \alpha\} \in \mathfrak{B}$ for each α .
- iii. $\{x : f(x) > \alpha\} \in \mathfrak{B}$ for each α .
- iv. $\{x : f(x) \geq \alpha\} \in \mathfrak{B}$ for each α .

(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 Theorems 3.19 and 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 σ -finite measure space, then we may choose the functions φ_n so that each vanishes outside a set of finite measure.*

8. Proposition: *If μ 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 α . The following two lemmas will be useful later. They state that, given a countable collection $\{B_\alpha\}$ of measurable sets which increase with α , 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 α in a countable set D of real numbers there is assigned a set $B_\alpha \in \mathfrak{G}$ such that $B_\alpha \subset B_\beta$ for $\alpha < \beta$. Then there is a measurable extended real-valued function f on X such that $f \leq \alpha$ on B_α and $f \geq \alpha$ on $X \sim B_\alpha$.*

Proof: For each $x \in X$, let $f(x) = \inf \{\alpha: x \in B_\alpha\}$, where as usual $\inf \emptyset = \infty$. If $x \in B_\alpha$, $f(x) \leq \alpha$. If $x \notin B_\alpha$, $x \notin B_\beta$ for each $\beta < \alpha$ and so $f(x) \geq \alpha$. Thus we have only to show that f is measurable. But for any real α , $\{x: f(x) < \alpha\} = \bigcup_{\beta < \alpha} B_\beta$. For if $f(x) < \alpha$, then $x \in B_\beta$ for some $\beta < \alpha$ by the definition of f . On the other hand, if $x \in B_\beta$ for some $\beta < \alpha$, then $f(x) \leq \beta < \alpha$. ■

10. Lemma: *Suppose that for each α in a countable set D of real numbers there is assigned a set $B_\alpha \in \mathfrak{G}$ 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_α and $f \geq \alpha$ a.e. on $X \sim B_\alpha$.*

Proof: Let $C = \bigcup_{\alpha < \beta} B_\alpha \sim B_\beta$. Then $\mu C = 0$. Let $B'_\alpha = B_\alpha \cup C$. If $\alpha < \beta$, we have $B'_\alpha \sim B'_\beta = (B_\alpha \sim B_\beta) \sim C = \emptyset$. Thus by Lemma 9 there is a measurable f such that $f \leq \alpha$ on B'_α and $f \geq \alpha$ on $X \sim B'_\alpha$. Thus $f \leq \alpha$ on B_α and $f \geq \alpha$ on $X \sim B_\alpha$ except for $x \in C$. ■

Problems

10. Prove Proposition 7. [For each pair $\langle n, k \rangle$ of integers let

$$E_{n,k} = \{x: k2^{-n} \leq f(x) < (k+1)2^{-n}\}, \text{ and set } \varphi_n = 2^{-n} \sum_{k=0}^{2^{2n}} k \chi_{E_{n,k}}.$$

11. Prove Proposition 8, and show that it is false if the word ‘complete’ is omitted.

12. Let (X, \mathcal{B}, μ) 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} .

13. Show that the requirement that D be countable in Lemmas 9 and 10 is unnecessary. [Replace an uncountable D by a countable subset E which is dense in D .]

14. Show that if D is dense in \mathbf{R} , then the function f in Lemma 9 is unique and the function f in Lemma 10 is unique to within equivalence; that is, any other function satisfying the conclusion of the lemma is equal to f almost everywhere.

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 $\{B_\beta\}$.

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 φ 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 φ which we use. If a and b are positive numbers and φ and ψ nonnegative simple functions, then $\int a\varphi + b\psi = a \int \varphi + b \int \psi$.

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 general measures, 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}, μ) . Then $\int f d\mu$ is the supremum of the integrals $\int \varphi d\mu$ as φ ranges over all simple functions with $0 \leq \varphi \leq f$.

It follows immediately from this definition that $f < g$ implies $\int f \leq \int g$ and that $\int cf = c \int f$. 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 non-negative measurable functions which converge almost everywhere on a set E to a function f . Then

$$\int_E f \leq \underline{\lim} \int_E f_n.$$

Proof: Without loss of generality we may assume that $f_n(x) \rightarrow f(x)$ for each $x \in E$. From the definition of $\int f$ it suffices to show that, if φ is any nonnegative simple function with $\varphi \leq f$, then $\int_E \varphi \leq \underline{\lim} \int_E f_n$.

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 $A_n = \{x \in E : f_k(x) > a \text{ for all } k \geq n\}$. Then $\langle A_n \rangle$ is an increasing sequence of measurable sets whose union contains A , since $\varphi \leq \lim f_n$. Thus $\lim \mu A_n = \infty$. Since $\int_E f_n \geq a \mu A_n$, we have $\lim \int_E f_n = \infty = \int_E \varphi$.

If $\int_E \varphi < \infty$, then there is a measurable set $A \subset E$ with $\mu A < \infty$ such that φ vanishes identically on $E \sim A$. Let M be the maximum of φ , let ϵ be a given positive number, and set

$$A_n = \{x \in E : f_k(x) > (1 - \epsilon)\varphi(x) \text{ for all } k \geq n\}.$$

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 $k \geq n$

$$\begin{aligned} \int_E f_k &\geq \int_{A_k} f_k \geq (1 - \epsilon) \int_{A_k} \varphi \\ &\geq (1 - \epsilon) \int_E \varphi - \int_{A \sim A_k} \varphi \\ &\geq \int_E \varphi - \epsilon \left[\int_E \varphi + M \right]. \end{aligned}$$

Hence

$$\underline{\lim} \int_E f_n \geq \int_E \varphi - \epsilon \left[\int_E \varphi + M \right].$$

Since ϵ is arbitrary,

$$\underline{\lim} \int_E f_n \geq \int_E \varphi. \blacksquare$$

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

$$\int f \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n \leq \int f. \blacksquare$$

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

$$\begin{aligned} \int af + bg &= \lim \int a\varphi_n + b\psi_n \\ &= \lim (a \int \varphi_n + b \int \psi_n) \\ &= a \int f + b \int g. \end{aligned}$$

Clearly $\int f \geq 0$. If $\int f = 0$, let $A_n = \{x: f(x) \geq 1/n\}$. Then $f \geq \frac{1}{n} \chi_{A_n}$, 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. \blacksquare

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 μ) 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 $|h| \leq |f|$ and h is measurable then h is integrable;
- iii. if $f \geq g$ a.e., then $\int f \geq \int g.$

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*

$$|f_n(x)| \leq g(x)$$

and such that almost everywhere on E

$$f_n(x) \rightarrow f(x).$$

Then

$$\int_E f = \lim \int_E f_n.$$

Proof: Apply Fatou's Lemma to the sequences $g + f_n$ and $g - f_n$. ■

Problems

16. Prove Proposition 15.

17. Suppose that μ is not complete, but that we define a bounded function f to be integrable over a set E of finite measure if

$$\sup_{\varphi \leq f} \int_E \varphi d\mu = \inf_{\psi \geq f} \int_E \psi d\mu$$

for all simple functions φ and ψ . Show that f is integrable if and only if it is measurable with respect to the completion of μ .

18. Let f be an integrable function on the measure space (X, \mathcal{B}, μ) . Show that given $\epsilon > 0$, there is a $\delta > 0$ such that for each measurable set E with $\mu E < \delta$ we have

$$\left| \int_E f \right| < \epsilon.$$

19. Let $\langle f_n \rangle$ be a sequence of measurable functions. We say that f_n converges in measure to f if given $\epsilon > 0$, there is an N such that

$$\mu\{x: |f_n(x) - f(x)| > \epsilon\} < \epsilon$$

for $n \geq N$. We say that $\langle f_n \rangle$ is a Cauchy sequence in measure if given $\epsilon > 0$, there is an N such that

$$\mu\{x: |f_n(x) - f_m(x)| > \epsilon\} < \epsilon$$

for $m, n \geq N$.

Show that a necessary and sufficient condition that a sequence $\langle f_n \rangle$ of measurable functions converge in measure to some function f is that it be a Cauchy sequence in measure. If $\langle f_n \rangle$ converges to f in measure, then there is a subsequence $\langle f_{n_\nu} \rangle$ which converges to f almost everywhere, and hence almost everywhere convergence may be replaced by convergence in measure in the convergence theorems.

20. Show that if f is integrable then the set $\{x: f(x) \neq 0\}$ is of σ -finite measure.

21. a. Let (X, \mathcal{B}, μ) be a measure space and g a nonnegative measurable function on X . Set $\nu E = \int_E g d\mu$. Show that ν 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 that g is simple and then use the Monotone Convergence Theorem.]

22. A function f on a measure space (X, \mathcal{B}, μ) 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 σ -algebra of locally measurable sets.

- b. If we define integration for nonnegative locally measurable functions f by taking $\int f$ to be the supremum of $\int \varphi$ as φ ranges over all simple functions less than f , then $\int f = \int f d\underline{\mu}$, where $\underline{\mu}$ is the extension of μ 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 μ . We can generalize by allowing the measure to vary also. Let (X, \mathcal{G}) be a measurable space and $\langle \mu_n \rangle$ a sequence of set functions defined on \mathcal{G} . We say that μ_n converges setwise to the set function μ if for each $E \in \mathcal{G}$ 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{G}) be a measurable space, $\langle \mu_n \rangle$ a sequence of measures which converge setwise to a measure μ , and $\langle f_n \rangle$ a sequence of nonnegative measurable functions which converge pointwise to the function f . Then*

$$\int f d\mu \leq \underline{\lim} \int f_n d\mu_n.$$

Proof: Setwise convergence of μ_n to μ implies that

$$\int \varphi d\mu = \lim \int \varphi d\mu_n$$

for any simple function φ . From the definition of $\int f d\mu$ it suffices to prove that $\int \varphi d\mu \leq \underline{\lim} \int f_n d\mu_n$ for any simple function $\varphi \leq f$.

Suppose $\int \varphi d\mu < \infty$. Then φ vanishes outside a set E of finite measure. Let ϵ be a positive number, and set

$$E_n = \{x: f_k(x) \geq (1 - \epsilon)\varphi(x) \text{ for all } k \geq n\}.$$

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 $n \geq m$ so that $\mu_k(E \sim E_m) < \epsilon$ for $k \geq n$. Since $E \sim E_k \subset E \sim E_m$, we have $\mu_k(E \sim E_k) < \epsilon$ for $k \geq n$. Thus

$$\begin{aligned} \int f_k d\mu_k &\geq \int_{E_k} f_k d\mu_k \geq (1 - \epsilon) \int_{E_k} \varphi d\mu_k \\ &\geq (1 - \epsilon) \int_E \varphi d\mu_k - \int_{E \sim E_k} \varphi d\mu_k \\ &\geq (1 - \epsilon) \int_E \varphi d\mu_k - M\epsilon, \end{aligned}$$

where M is the maximum of φ . Thus

$$\underline{\lim} \int f_k d\mu_k \geq \int_E \varphi d\mu - \epsilon \left[M + \int \varphi d\mu \right].$$

Since ϵ was arbitrary,

$$\int_E \varphi d\mu \leq \underline{\lim} \int f_k d\mu_k.$$

The case when $\int \varphi d\mu = \infty$ is similarly handled. ■

18. Proposition: Let (X, \mathfrak{B}) be a measurable space and $\langle \mu_n \rangle$ a sequence of measures on \mathfrak{B} which converge setwise to a measure μ . Let $\langle f_n \rangle$ and $\langle g_n \rangle$ be two sequences of measurable functions which converge pointwise to f and g . Suppose $|f_n| \leq g_n$ 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 the sequences $g_n + f_n$ and $g_n - f_n$. ■

Problems

23. Let (X, \mathfrak{B}) be a measurable space and $\langle \mu_n \rangle$ a sequence of measures on \mathfrak{B} such that for each $E \in \mathfrak{B}$, $\mu_{n+1}E \geq \mu_n E$. Let $\mu E = \lim \mu_n E$. Then μ is a measure on \mathfrak{B} .

24. Give an example of a decreasing sequence $\langle \mu_n \rangle$ of measures on a measurable space such that the set function μ defined by $\mu E = \lim \mu_n E$ is not a measure.

25. Let (X, \mathfrak{B}) be a measurable space and $\langle \mu_n \rangle$ a sequence of measures on \mathfrak{B} which converge setwise to a set function μ . If $\mu X < \infty$, then μ is a measure.

5 Signed Measures

In this section we consider some of the possibilities which may arise if a measure is allowed to take on both positive and negative values.

We first note that if μ_1 and μ_2 are two measures defined on the same measurable space (X, \mathfrak{B}) , then we may define a new measure μ_3 on (X, \mathfrak{B}) by setting

$$\mu_3(E) = c_1\mu_1(E) + c_2\mu_2(E) \quad c_1, c_2 \geq 0.$$

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 ν 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 ν is not defined when $\mu_1 E = \mu_2 E = \infty$. For this reason we should have either μ_1 or μ_2 finite. With these considerations in mind we make the following definition:

Definition: By a signed measure on the measurable space (X, \mathfrak{B}) we mean an extended real-valued set function ν defined for the sets of \mathfrak{B} and satisfying the following conditions:

- i. ν assumes at most one of the values $+\infty, -\infty$.
- ii. $\nu(\emptyset) = 0$.
- iii. $\nu \left(\bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} \nu E_i$ 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 ν 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 ν 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 ν measure. A set which is both positive and negative with respect to ν is called a **null set**. A measurable set is a null set if and only if every measurable subset of it has ν 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 which 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, 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 νE is finite.

Thus $\sum \frac{1}{n_k}$ converges, and we have $n_k \rightarrow \infty$. 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 $n_k \rightarrow \infty$, 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 ϵ 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 ν 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 ν . Let λ be the supremum of νA over all sets A which are positive with respect to ν . Since the empty set is positive, $\lambda \geq 0$. Let $\langle A_i \rangle$ be a sequence of positive sets such that

$$\lambda = \lim_{i \rightarrow \infty} \nu A_i,$$

and set

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

By Lemma 18 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 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 ν and B negative is called a **Hahn decomposition** for ν . 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 ν , then we may define two measures ν^+ and ν^- with $\nu = \nu^+ - \nu^-$ by setting

$$\nu^+(E) = \nu(E \cap A)$$

and

$$\nu^-(E) = -\nu(E \cap B).$$

Two measures ν_1 and ν_2 on (X, \mathfrak{G}) 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 ν^+ and ν^- 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 ν be a signed measure on the measurable space (X, \mathfrak{G}) . Then there are two mutually singular measures ν^+ and ν^- on (X, \mathfrak{G}) such that $\nu = \nu^+ - \nu^-$. Moreover, there is only one such pair of mutually singular measures.*

The decomposition of ν given by the proposition is called the **Jordan decomposition** of ν . The measures ν^+ and ν^- are called the positive and negative parts (or variations) of ν . Since ν assumes at most one of the values $+\infty$ and $-\infty$, either ν^+ or ν^- must be finite. If they are both finite, we call ν a finite signed measure.

The measure $|\nu|$ defined by

$$|\nu|(E) = \nu^+ E + \nu^- E$$

is called the **absolute value** or **total variation** of ν . A set E is positive for ν if $\nu^- E = 0$. It is a null set if $|\nu|(E) = 0$.

Problems

26. 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.

27. Show that there is only one pair of mutually singular measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$. [Hint: Show that any such pair determines a Hahn decomposition and apply the results of Problem 26b.]

28. Show that if E is any measurable set, then

$$-\nu^-E \leq \nu E \leq \nu^+E$$

and

$$|\nu E| \leq |\nu|(E).$$

29. Show that if ν_1 and ν_2 are any two finite signed measures, then so is $\alpha\nu_1 + \beta\nu_2$, where α and β 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 .

30. We define integration with respect to a signed measure ν 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

$$\int_E f d\nu = |\nu|(E).$$

31. a. Let μ and ν be finite signed measures. Show that there is a signed measure $\mu \wedge \nu$ which is smaller than μ and ν but larger than any other signed measure which is smaller than μ and ν . [Hint: $\mu \wedge \nu = \frac{1}{2}(\mu + \nu - |\mu - \nu|)$.]

b. Show that there is a measure $\mu \vee \nu$ which is larger than μ and ν but smaller than any other measure which is larger than μ and ν . Also $\mu \vee \nu + \mu \wedge \nu = \mu + \nu$.

c. If μ and ν 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 μ and ν are two measures defined on (X, \mathcal{B}) , we say that μ and ν 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 ν and μ , we sometimes say that ν is singular with respect to μ . The notion antithetical to singularity is absolute continuity. A measure ν is said to be **absolutely continuous** with respect to the measure μ if $\nu A = 0$ for each set A for which $\mu A = 0$. We use the symbolism $\nu \ll \mu$ for ν absolutely continuous with respect to μ .

In the case of signed measures μ and ν , 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 μ or almost everywhere with respect to ν , etc. These are 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 μ 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 ν is a set function defined on \mathcal{B} , and it follows from Corollary 14 that ν is countably additive and hence a measure. The measure ν will be finite if and only if f is integrable. Since the integral over a set of μ -measure zero is zero, we have ν absolutely continuous with respect to μ . The next theorem shows that, subject to σ -finiteness restrictions, every absolutely continuous measure ν is obtained in this fashion.

23. Theorem (Radon-Nikodym): *Let (X, \mathcal{B}, μ) be a σ -finite measure space, and let ν be a measure defined on \mathcal{B} which is absolutely continuous with respect to μ . 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 σ -finite case is not difficult and is left to the reader. Thus we shall assume that μ is finite. Then $\nu - \alpha\mu$ is a signed measure for each rational number α . Let (A_α, B_α) 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 $(\nu - \beta\mu)(B_\alpha \sim B_\beta) \geq 0$. If $\beta > \alpha$, these imply $\mu(B_\alpha \sim B_\beta) = 0$, and so by Lemma 10 there is a measurable function f such that for each rational α we have $f \geq \alpha$ a.e. on A_α and $f \leq \alpha$ a.e. on B_α . Since $B_0 = \emptyset$, we may take f to be nonnegative.

Let E be an arbitrary set in \mathfrak{S} , and set

$$E_k = E \cap \left(\bigcup_{n=k}^{\infty} B_n \right), \quad E_\infty = E \sim \bigcup_{n=0}^{\infty} B_n.$$

Then $E = E_\infty \cup \bigcup_{k=0}^{\infty} E_k$, and this is a disjoint union. Thus

$$\nu E = \nu E_\infty + \sum_{k=0}^{\infty} \nu E_k.$$

Since $E_k \subset B_k \cap A_k$, 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_∞ 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 α . If $\mu E_\infty = 0$, we must 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

$$\nu E - \frac{1}{N} \mu E \leq \int_E f d\mu \leq \nu E + \frac{1}{N} \mu E.$$

Since μE is finite and N arbitrary, we must have

$$\nu E = \int_E f d\mu. \blacksquare$$

The function f given by Theorem 23 is called the **Radon-Nikodym derivative** of ν with respect to μ . It is sometimes denoted by $\left[\frac{d\nu}{d\mu} \right]$.

24. Proposition (Lebesgue Decomposition): *Let (X, \mathcal{G}, μ) be a σ -finite measure space and ν a σ -finite measure defined on \mathcal{G} . Then we can find a measure ν_0 which is singular with respect to μ and a measure ν_1 which is absolutely continuous with respect to μ such that $\nu = \nu_0 + \nu_1$. The measures ν_0 and ν_1 are unique.*

Proof: Since μ and ν are σ -finite measures, so is the measure $\lambda = \mu + \nu$. Since both μ and ν are absolutely continuous with respect to λ , the Radon-Nikodym Theorem asserts the existence of nonnegative measurable functions f and g such that for each $E \in \mathcal{G}$

$$\mu E = \int_E f d\lambda, \quad \nu E = \int_E g d\lambda.$$

Let $A = \{x: f(x) > 0\}$ and $B = \{x: f(x) = 0\}$. Then X is the disjoint union of A and B , while $\mu B = 0$. If we define ν_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 μ -measure zero. Then

$$0 = \mu E = \int_E f d\lambda,$$

and $f = 0$ a.e. [λ] 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. \blacksquare

Problems

32. a. Show that the Radon-Nikodym Theorem for a finite measure μ implies the theorem for a σ -finite measure μ . [Hint: Decompose X into a countable union of sets X_i of finite μ -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.

33. Radon-Nikodym derivatives. Show that the Radon-Nikodym derivative $\left[\frac{d\nu}{d\mu} \right]$ 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

$$\left[\frac{d\nu}{d\mu} \right] = \left[\frac{d\mu}{d\nu} \right]^{-1}.$$

34. a. Show that if ν is a signed measure such that $\nu \perp \mu$ and $\nu \ll \mu$, then $\nu = 0$.

b. Show that if ν_1 and ν_2 are singular with respect to μ , then so is $c_1\nu_1 + c_2\nu_2$.

c. Show that if ν_1 and ν_2 are absolutely continuous with respect to μ so is $c_1\nu_1 + c_2\nu_2$.

d. Prove the uniqueness assertion in the Lebesgue decomposition.

35. Extend the Radon-Nikodym Theorem to the case of signed measures.

36. Complex measures. A set function ν which assigns a complex number $\nu(E)$ to each E in a σ -algebra \mathcal{G} is called a complex measure if

$\nu\emptyset = 0$ and for each countable disjoint union $\bigcup E_i$ of sets in \mathfrak{G} 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 may be expressed as $\nu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$ where μ_1, μ_2, μ_3 , and μ_4 are finite measures.

b. Show that for each complex measure ν there is a measure μ and a complex-valued measurable function φ with $|\varphi| = 1$ such that for each set E in \mathfrak{G} ,

$$\nu E = \int_E \varphi d\mu.$$

[Hint: Apply the Radon-Nikodym Theorem to the measures μ_i with respect to the measure $\mu_1 + \mu_2 + \mu_3 + \mu_4$.]

c. Show that the measure μ in (b) is unique and that φ is uniquely determined to within sets of μ measure zero.

d. The measure μ in (b) is called the *total variation* or *absolute value* of ν and is sometimes denoted by $|\nu|$. Show that the results of Problem 30 hold for complex measures.

e. If ν is a complex measure with $|\nu|(X) = 1$ and $\nu(X) = 1$, then ν is a positive real measure.

37. *Alternate proof of the Radon-Nikodym Theorem.* We can give a proof of the Radon-Nikodym Theorem which 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 μ and ν be finite measures on a measurable space (X, \mathfrak{G}) 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 $0 \leq g \leq 1$, and

$$\begin{aligned}\mu(E) &= \int_E g d\lambda \\ \nu(E) &= \int_E (1 - g) d\lambda.\end{aligned}$$

c. If $\nu \ll \mu$, then $\lambda \ll \mu$, and $g = 0$ only on a set of μ -measure zero. In this case

$$\lambda(E) = \int_E g^{-1} d\mu.$$

[Hint: Consider Problem 21.]

- d. If $\nu \ll \mu$, then $(1 - g)g^{-1}$ is integrable with respect to μ and

$$\nu(E) = \int_E (1 - g)g^{-1} d\mu.$$

38. Use the following example to show that the hypothesis in the Radon-Nikodym Theorem that μ is σ -finite cannot be omitted. Let $X = [0, 1]$, \mathfrak{B} the class of Lebesgue measurable subsets of $[0, 1]$, and take ν to be Lebesgue measure and μ to be the counting measure on \mathfrak{B} . Then ν is finite and absolutely continuous with respect to μ , but there is no function f such that $\nu E = \int_E f d\mu$ for all $E \in \mathfrak{B}$. At what point does the proof of Theorem 23 break down for this example?

39. Decomposable measures. Let (X, \mathfrak{B}, μ) be a measure space. A collection $\{X_\alpha\}$ of disjoint measurable subsets of X is called a **decomposition** for μ if $\mu X_\alpha < \infty$ for each α and if $\mu E = 0$ for every measurable set E such that $\mu(E \cap X_\alpha) = 0$ for all α . A measure μ is called **decomposable** if it has a decomposition.

a. If $\{X_\alpha\}$ is a decomposition for μ , 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 2.20.

b. If $\{X_\alpha\}$ is a decomposition for a complete measure μ , then f is locally measurable if and only if the restriction of f to X_α is measurable for each α . 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 ν be absolutely continuous with respect to μ , and suppose that there is a collection $\{X_\alpha\}$ which is a decomposition for both μ and ν . 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 ν , we merely assume that if $E \in \mathfrak{B}$ and $\nu(E \cap X_\alpha) = 0$ for all α then $\nu E = 0$.

7 The L^p Spaces

If (X, \mathfrak{B}, μ) is a 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 $1 \leq p < \infty$ we set

$$\|f\|_p = \left\{ \int |f|^p d\mu \right\}^{\frac{1}{p}},$$

and for $p = \infty$ we set

$$\|f\|_\infty = \text{ess sup } |f|.$$

Note that the space $L^\infty(\mu)$ depends on the choice of μ 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 $1 \leq p \leq \infty$ the spaces $L^p(\mu)$ are Banach spaces, and if $f \in L^p(\mu)$, $g \in L^q(\mu)$, with $\frac{1}{p} + \frac{1}{q} = 1$, then $fg \in L^1(\mu)$ and*

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

The following proposition, whose proof is left to the reader, is a version of Littlewood's second principle:

26. Proposition: *Let $f \in L^p(\mu)$, $1 \leq p < \infty$. Then, given $\epsilon > 0$, there is a simple function φ 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 μ is σ -finite, each linear functional on $L^1(\mu)$ is of this form. We begin by establishing a lemma.

27. Lemma: Let (X, \mathfrak{B}, μ) 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 φ . 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

$$\varphi_n = (\psi_n)^{\frac{1}{p}} \operatorname{sgn} g.$$

Then φ_n is a simple function, and

$$\|\varphi_n\|_p = \left\{ \int \psi_n \, d\mu \right\}^{\frac{1}{p}}.$$

Since $\varphi_n g \geq |\varphi_n| |\psi_n|^{\frac{1}{q}} = |\psi_n|^{\frac{1}{p} + \frac{1}{q}} = \psi_n$, we have

$$\begin{aligned} \int \psi_n \, d\mu &\leq \int \varphi_n g \, d\mu \\ &\leq M \|\varphi_n\|_p \\ &\leq M \left\{ \int \psi_n \, d\mu \right\}^{\frac{1}{p}}. \end{aligned}$$

Since $1 - \frac{1}{p} = \frac{1}{q}$,

$$\left\{ \int \psi_n \, d\mu \right\}^{\frac{1}{q}} \leq M$$

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 ($1 \leq p < \infty$) which 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,

$$\left\| f - \sum_{i=1}^n f_i \right\|_p \rightarrow 0$$

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 $1 \leq p < \infty$ and μ a σ -finite measure. Then there is a unique element g in L^q , where $\frac{1}{q} + \frac{1}{p} = 1$, such that

$$F(f) = \int f g \, d\mu.$$

We have also $\|F\| = \|g\|_q$.

Proof: Let us first consider the case of a finite measure μ . Then every bounded measurable function is in $L^p(\mu)$. Define a set function ν 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 ν is a signed measure, and so 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 ν is always finite, g is integrable.

If φ 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 σ -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 which vanishes outside X_n such that

$$F(f) = \int fg_n \, d\mu$$

for all $f \in L^p$ which 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

$$\begin{aligned} \int |g|^q \, d\mu &= \lim \int |g_n|^q \, d\mu \\ &\leq \|F\|^q, \end{aligned}$$

and $g \in L^q$.

If $f \in L^p$, let $f_n = f$ on X_n , and $f_n = 0$ otherwise. Then $f_n \rightarrow f$ pointwise and in L^p . Since $|fg|$ is integrable and $|f_n g| \leq |fg|$, the Lebesgue Convergence Theorem gives

$$\begin{aligned}\int fg \, d\mu &= \lim \int f_n g \, d\mu \\ &= \lim \int f_n g_n \, d\mu \\ &= \lim F(f_n) \\ &= F(f). \blacksquare\end{aligned}$$

If $p = 1$, the requirement that μ be σ -finite is necessary. Some extensions and counterexamples are given in Problems 45 and 46. If $p > 1$, then σ -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 previous theorem that if E is any measurable set of σ -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 $f \in L^p$ which 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 σ -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 σ -finite measure such that $\lambda(E_n)$ tends to the maximum value m of λ . Then $H = \bigcup E_n$ is a set of σ -finite measure, and by the monotonicity of λ 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 σ -finite measure containing H , then $g_E = g_H$ a.e. on H . But $\int |g_E|^q = \lambda(E) \leq \lambda(H) = \int |g_H|^q$, 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 σ -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 previous theorem. ■

Problems

- 40. Prove Proposition 26.
- 41. Prove Lemma 27 for the case $p = 1$.
- 42. Prove Lemma 28.
- 43. For $g \in L^q$, let F be the linear functional on L^p defined by

$$F(f) = \int fg d\mu.$$

Show that $\|F\| = \|g\|_q$.

- 44. a. Let μ 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 μ is the counting measure on a not necessarily countable set X .
- 45. Let (X, \mathcal{G}, μ) be a decomposable measure space (cf. Problem 39).
 - 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 μ given in Problem 8e.

- 46. Let A and B be uncountable sets with different numbers of elements, and let $X = A \times B$. A set of the form $\{(x, y): x = a\}$ is called a vertical line and a set of the form $\{(x, y): y = b\}$ a horizontal line. Let \mathcal{G} 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{G} is a σ -algebra. Let μE be the number of horizontal and vertical lines L for which $\tilde{E} \cap L$ is countable and νE the number of horizontal lines with $\tilde{E} \cap L$ countable. Then μ and ν are measures on \mathcal{G} , 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 σ -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 which 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 μ^* we mean an 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

$$\text{iii'}. \quad E = \bigcup_{i=1}^{\infty} E_i, \quad E_i \text{ disjoint} \Rightarrow \mu^*E \leq \sum_{i=1}^{\infty} \mu^*E_i.$$

The outer measure μ^* 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 μ^* if for every set A we have

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap \tilde{E}).$$

Since μ^* is subadditive, it is only necessary to show that

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap \tilde{E})$$

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 μ^*A finite.

1. Theorem: *The class \mathfrak{B} of μ^* -measurable sets is a σ -algebra. If $\bar{\mu}$ is μ^* restricted to \mathfrak{B} , then $\bar{\mu}$ is a complete measure on \mathfrak{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

$$\mu^*(A \cap [E_1 \cup E_2]) \leq \mu^*(A \cap E_2) + \mu^*(A \cap \tilde{E}_2 \cap E_1)$$

by subadditivity, and so

$$\mu^*A \geq \mu^*(A \cap [E_1 \cup E_2]) + \mu^*(A \cap \tilde{E}_1 \cap \tilde{E}_2).$$

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 \otimes 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 \otimes is a σ -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 μ^* . 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. Show that $\bar{\mu}$ is saturated.
3. 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 μ 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 σ -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 be a measure defined on a σ -algebra \mathfrak{B} containing \mathfrak{A} . We shall do this by using

the measure on the algebra to construct an outer measure μ^* and show that the measure $\bar{\mu}$ induced by μ^* is the desired extension of μ . The process by which we construct μ^* from μ is analogous to that by which we constructed Lebesgue outer measure from the lengths of intervals: We define

$$\mu^*E = \inf \sum_{i=1}^{\infty} \mu A_i \quad (1)$$

where $\langle A_i \rangle$ ranges over all sequences from \mathcal{A} such that $E \subset \bigcup_{i=1}^{\infty} A_i$. We first establish some lemmas concerning μ^* .

2. Lemma: If $A \in \mathcal{A}$ and if $\langle A_i \rangle$ is any sequence of sets in \mathcal{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 $B_n \in \mathcal{A}$ and $B_n \subset A_n$. But A is the disjoint union of the sequence $\langle B_n \rangle$, and so by countable additivity

$$\mu A = \sum_{n=1}^{\infty} \mu B_n \leq \sum_{n=1}^{\infty} \mu A_n. \blacksquare$$

3. Corollary: If $A \in \mathcal{A}$, $\mu^*A = \mu A$.

4. Lemma: The set function μ^* is an outer measure.

Proof: Since μ^* 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 \mathcal{A} such that $E_i \subset \bigcup_{j=1}^{\infty} A_{ij}$ and

$$\sum_{j=1}^{\infty} \mu A_{ij} < \mu^*E_i + \epsilon/2^i.$$

Then

$$\mu^*E \leq \sum_{ij} \mu A_{ij} < \sum_{i=1}^{\infty} \mu^*E_i + \epsilon.$$

Since ϵ was an arbitrary positive number,

$$\mu^*E \leq \sum_{i=1}^{\infty} \mu^*E_i,$$

and μ^* is subadditive. ■

5. Lemma: If $A \in \alpha$ then A is measurable with respect to μ^* .

Proof: Let E be an arbitrary set of finite outer measure and ϵ a positive number. Then there is a sequence $\langle A_i \rangle$ from α such that $E \subset \bigcup A_i$ and

$$\sum \mu A_i < \mu^*E + \epsilon.$$

By the additivity of μ on α we have

$$\mu(A_i) = \mu(A_i \cap E) + \mu(A_i \cap \tilde{E}).$$

Hence

$$\begin{aligned} \mu^*E + \epsilon &> \sum_{i=1}^{\infty} \mu(A_i \cap E) + \sum_{i=1}^{\infty} \mu(A_i \cap \tilde{E}) \\ &> \mu^*(E \cap A) + \mu^*(E \cap \tilde{A}), \end{aligned}$$

since

$$E \cap A \subset \bigcup (A_i \cap A)$$

and

$$E \cap \tilde{A} \subset \bigcup (A_i \cap \tilde{A}).$$

Since ϵ was an arbitrary positive number,

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

and A is measurable. ■

The outer measure μ^* which we have defined is called the outer measure induced by μ . For a given algebra α of sets we use α_σ to denote those sets which are countable unions of sets of α and use $\alpha_{\sigma\delta}$ to denote those sets which are countable intersections of sets in α_σ .

6. Proposition: Let μ be a measure on an algebra \mathfrak{A} , μ^* the outer measure induced by μ , 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 μ^* 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}_σ such that $E \subset A_n$ and $\mu^*A_n < \mu^*E + \frac{1}{n}$. Let $B = \bigcap A_n$. Then $B \in \mathfrak{A}_{\sigma\delta}$ and $E \subset B$. Since $B \subset A_n$, $\mu^*B \leq \mu^*A_n < \mu^*E + \frac{1}{n}$. Since n is arbitrary, $\mu^*B \leq \mu^*E$. But $E \subset B$, and so $\mu^*B \geq \mu^*E$ by monotonicity. Hence $\mu^*B = \mu^*E$. ■

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 σ -finite case. It can be considered a generalization of the first principle of Littlewood.

7. Proposition: Let μ be a σ -finite measure on an algebra \mathfrak{A} , and let μ^* be the outer measure generated by μ . A set E is μ^* measurable if and only if E is the proper difference $A - B$ 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 σ -algebra, while each set of μ^* -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 μX_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 \mathcal{A}_σ 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 $A_n \in \mathcal{A}_\sigma$, the set $A = \bigcap_{n=1}^{\infty} A_n$ is in $\mathcal{A}_{\sigma\delta}$, 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

$$\bar{\mu}(A \sim E) = 0. \blacksquare$$

We summarize the results of this section in the following theorem.

8. Theorem (Carathéodory): *Let μ be a measure on an algebra \mathcal{A} , and μ^* the outer measure induced by μ . Then the restriction $\bar{\mu}$ of μ^* to the μ^* -measurable sets is an extension of μ to a σ -algebra containing \mathcal{A} . If μ is finite (or σ -finite) so is $\bar{\mu}$. If μ is σ -finite, then $\bar{\mu}$ is the only measure on the smallest σ -algebra containing \mathcal{A} which is an extension of μ .*

Proof: The fact that $\bar{\mu}$ is an extension of μ from \mathcal{A} to be a measure on a σ -algebra containing \mathcal{A} follows directly from Corollary 3, Lemma 5, and Theorem 1, and it is readily verified that $\bar{\mu}$ is finite or σ -finite whenever μ is.

To show the unicity of $\bar{\mu}$ when μ is σ -finite, we let \mathfrak{G} be the smallest σ -algebra containing α and $\tilde{\mu}$ some measure on \mathfrak{G} which agrees with μ on α .

Since each set in α_σ can be expressed as a disjoint countable union of sets in α , the measure $\tilde{\mu}$ must agree with $\bar{\mu}$ on α_σ . Let B be any set in \mathfrak{G} with finite outer measure. Then by Proposition 6 there is an A in α_σ 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 ϵ is an arbitrary positive number, we have

$$\tilde{\mu}B \leq \mu^*B$$

for each $B \in \mathfrak{G}$.

Since the class of sets measurable with respect to μ^* is a σ -algebra containing α , each B in \mathfrak{G} must be measurable. If B is measurable and A is in α_σ with $B \subset A$ and $\mu^*A \leq \mu^*B + \epsilon$, then

$$\mu^*A = \mu^*B + \mu^*(A \sim B),$$

and so

$$\mu^*(A \sim B) \leq \epsilon,$$

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 ϵ is arbitrary, we have

$$\mu^*B \leq \tilde{\mu}B$$

and so

$$\mu^*B = \tilde{\mu}B.$$

If μ is a σ -finite measure, let $\{X_i\}$ be a countable disjoint collection of sets in α with $X = \bigcup X_i$ and μX_i finite. If B is any set in \mathfrak{G} , then

$$B = \bigcup (X_i \cap B)$$

and this is a countable disjoint union of sets in \mathfrak{A} , 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

$$\bar{\mu}(X_i \cap B) = \tilde{\mu}(X_i \cap B). \blacksquare$$

This extension procedure not only extends μ to a measure on the smallest σ -algebra \mathfrak{G} containing \mathfrak{A} , but also completes and saturates the measure. If μ is σ -finite, the extension to \mathfrak{G} is already saturated, and the extension to the μ^* -measurable sets is merely the completion of μ . If μ is not σ -finite, then the extension to μ^* -measurable sets also saturates $\bar{\mu}$. It should be observed that in this case the extension of μ to \mathfrak{G} need not be unique (Problem 4), although any extension $\tilde{\mu}$ must agree with $\bar{\mu}$ for each set B of \mathfrak{G} 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 5 and 6.

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 \mathfrak{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 μ is a set function defined on \mathcal{C} , it is natural to attempt to define a finitely additive set function on \mathfrak{A} by setting

$$\mu A = \sum_{i=1}^n \mu E_i$$

whenever A is the disjoint union of the sets E_i in \mathcal{C} . Since a set A in \mathfrak{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 μA . The following proposition gives conditions under which this procedure can be carried out and will give a measure on the algebra \mathfrak{A} .

9. Proposition: Let \mathcal{C} be a semialgebra of sets and μ a nonnegative set function defined on \mathcal{C} with $\mu\emptyset = 0$ (if $\emptyset \in \mathcal{C}$). Then μ has a unique extension to a measure on the algebra \mathfrak{A} generated by \mathcal{C} if the following conditions are satisfied:

- If a set C in \mathcal{C} is the union of a finite disjoint collection $\{C_i\}$ of sets in \mathcal{C} , then $\mu C = \sum \mu C_i$.
- 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

4. 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 μ to the smallest σ -algebra containing \mathfrak{A} is not unique.

5. 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: $\mu C_i = \sum_j \mu(C_i \cap D_j)$.]

b. Condition (ii) implies that μ is countably additive on \mathfrak{A} [for finite additivity and monotonicity already imply the reverse inequality].

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}_σ 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}_σ .

7. Let μ be a finite measure on an algebra \mathfrak{A} , and μ^* 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 μ^* on X and form the induced measure $\bar{\mu}$ on the μ^* -measurable sets, we can use $\bar{\mu}$ to induce an outer measure μ^+ .

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 μ^* -measurable set $A \supset E$ with $\mu^* A = \mu^* E$.

c. An outer measure which satisfies the criterion in (b) for each set E is called a **regular** outer measure. Show that every outer measure induced by a measure on an algebra is regular.

d. Let X be a set consisting of two points. Construct an outer measure on X which is not regular.

*3 The Lebesgue-Stieltjes Integral

Let X be the set of real numbers and \mathcal{B} the class of all Borel sets. A measure μ defined on \mathcal{B} and finite for bounded sets is called a *Borel measure* (on the real line). To each finite Borel measure we associate a function F by setting

$$F(x) = \mu(-\infty, x].$$

The function F is called the *cumulative distribution function* of μ 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 $\left(a, b + \frac{1}{n}\right]$, Proposition 11.2 implies that

$$\mu(a, b] = \lim_{n \rightarrow \infty} \mu\left(a, b + \frac{1}{n}\right],$$

and so

$$F(b) = \lim_{n \rightarrow \infty} F\left(b + \frac{1}{n}\right) = F(b+).$$

Thus a cumulative distribution function is continuous on the right. Similarly,

$$\begin{aligned} \mu\{b\} &= \lim_{n \rightarrow \infty} \mu\left(b - \frac{1}{n}, b\right] \\ &= \lim_{n \rightarrow \infty} F(b) - F\left(b - \frac{1}{n}\right) \\ &= F(b) - F(b-). \end{aligned}$$

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

$$\lim_{n \rightarrow -\infty} F(n) = 0,$$

and hence

$$\lim_{x \rightarrow -\infty} F(x) = 0,$$

because of the monotonicity of F . We summarize these properties in the following lemma:

10. Lemma: If μ is a finite Borel measure on the real line, then its cumulative distribution function F is a monotone increasing bounded function which is continuous on the right. Moreover, $\lim_{x \rightarrow -\infty} F(x) = 0$.

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 Borel measure μ such that

$$\mu(a, b] = F(b) - F(a) \quad (2)$$

for all intervals of the form $(a, b]$, where we define $F(\infty) = \lim_{x \rightarrow \infty} F(x)$ and $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$. We begin with the following lemma, whose proof is left to the reader (Problem 9):

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 $(a, b]$ and set $\mu(a, b] = F(b) - F(a)$, then μ is easily seen to satisfy condition (i) of Proposition 9, and since Lemma 11 is precisely the second condition, we see that μ admits a unique extension to a measure on the algebra generated by \mathcal{C} . By Theorem 8 this μ can be extended to a σ -algebra containing \mathcal{C} . Since the class \mathfrak{B} of Borel sets is the smallest σ -algebra containing \mathcal{C} , we have an extension of μ to a Borel measure. The measure μ is σ -finite, since X is the union of the intervals $(n, n + 1]$ and each has finite measure. Thus the extension of μ to \mathfrak{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 Borel measure μ 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 Borel measure provided $F(-\infty) = 0$.

If φ 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 φ with respect to F to be

$$\int \varphi dF = \int \varphi d\mu$$

where μ is the Borel measure having F as its cumulative distribution function. If φ is both positive and negative, we say that it is integrable with respect to F if it is integrable with respect to μ .

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 10), and we define the Lebesgue-Stieltjes integral of φ 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 11c), while the Riemann-Stieltjes integral can exist when F is not of bounded variation, say when F is continuous and φ is of bounded variation.

Problems

9. 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.]

10. Let F be a monotone increasing function, and define

$$F^*(x) = \lim_{y \rightarrow x+} F(y).$$

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 $F^* = G^*$.

11. a. Show that each bounded function F of bounded variation gives rise to a finite signed Borel measure ν such that

$$\nu(a, b] = F(b+) - F(a+).$$

b. Compare Theorem 5.4 with the Jordan decomposition of ν .

c. Extend the definition of the Lebesgue-Stieltjes integral $\int \varphi dF$ to functions F of bounded variation.

d. Show that if $|\varphi| \leq M$ and if the total variation of F is T , then $|\int \varphi dF| \leq MT$.

12. a. Let F be the cumulative distribution function of the Borel measure ν , 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.

13. Let F be a continuous increasing function on $[a, b]$ with $F(a) = c$, $F(b) = d$, and let φ 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 12a to take care of the case when φ is a characteristic function and generalize first to simple φ and then to general φ .]

14. a. Show that a measure μ is absolutely continuous with respect to Lebesgue measure if and only if its cumulative distribution function is absolutely continuous.

b. If μ 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

$$\int f dF = \int fF' dx.$$

4 Product Measures

Let (X, \mathcal{A}, μ) and (Y, \mathcal{B}, ν) 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 $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we call $A \times B$ a

measurable rectangle. The collection \mathcal{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 (x, y) 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 ν 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

$$\sum \nu B_i \cdot \mu A_i = \nu B \cdot \mu A. \blacksquare$$

The lemma implies that λ 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 λ to be a complete measure on a σ -algebra \mathcal{S} containing \mathcal{R} . This extended measure is called the product measure of μ and ν and is denoted by $\mu \times \nu$. If μ and ν are finite (or σ -finite), so is $\mu \times \nu$. If X and Y are the real line and μ and ν 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 $\mathfrak{R}_{\sigma\delta}$. Then E_x is a measurable subset of Y .

Proof: The lemma is trivially true if E is in the class \mathfrak{R} of measurable rectangles. We next show it to be true for E in \mathfrak{R}_σ . 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 χ_{E_x} must also be measurable, whence E_x is measurable.

Suppose now that $E = \bigcap_{i=1}^{\infty} E_i$ with $E_i \in \mathfrak{R}_\sigma$. Then

$$\begin{aligned}\chi_{E_x}(y) &= \chi_E(x, y) \\ &= \inf_i \chi_{E_i}(x, y) \\ &= \inf_i \chi_{(E_i)_x}(y),\end{aligned}$$

and we see that χ_{E_x} is measurable. Thus E_x is measurable for any $E \in \mathfrak{R}_{\sigma\delta}$. ■

16. Lemma: Let E be a set in $\mathfrak{R}_{\sigma\delta}$ 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}_σ 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}_σ 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 $g_i(x) = \nu[(E_i)_x]$. Then $g(x) = \lim g_i(x)$, and so g is measurable. 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

$$g_i \rightarrow g \text{ a.e.,}$$

and so g is measurable. Since $0 \leq g_i \leq g_1$, 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 ν 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*

$$g(x) = \nu(E_x)$$

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

$$g(x) = \nu E_x = \nu F_x \text{ a.e.,}$$

and so g is a measurable function by Lemma 16. Again by Lemma 16

$$\begin{aligned}\int g \, d\mu &= \mu \times \nu(F) \\ &= \mu \times \nu(E). \blacksquare\end{aligned}$$

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}, μ) and (Y, \mathcal{B}, ν) 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. $\int_X \left[\int_Y f d\nu \right] d\mu = \int_{X \times Y} f d(\mu \times \nu) = \int_Y \left[\int_X f d\mu \right] d\nu$.

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 φ_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

$$\begin{aligned} \int_X \left[\int_Y f d\nu \right] d\mu &= \lim \int_X \left[\int_Y \varphi_n d\nu \right] d\mu \\ &= \lim \int_{X \times Y} \varphi_n d(\mu \times \nu) \\ &= \int_{X \times Y} f d(\mu \times \nu). \blacksquare \end{aligned}$$

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 17). In the case when μ and ν are σ -finite, the integrability of f can be determined by iterated integration using the following theorem:

20. Theorem (Tonelli): *Let (X, \mathcal{A}, μ) and (Y, \mathcal{B}, ν) be two σ -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 μ and ν are σ -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 σ -algebras on X and Y , then the smallest σ -algebra containing the measurable rectangles is denoted by $\mathcal{A} \times \mathcal{B}$. Thus the product measure is defined on a σ -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 μ and ν are both σ -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 doesn’t 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 μ to show that $\int f(x, y) d\nu(y)$ is measurable, for if μ 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 μ is not complete (provided f is integrable), but the proof of this is surprisingly intricate. For a proof see Halmos [8], 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 σ -finiteness and nonnegativity from the Tonelli Theorem. Problem 22 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

15. 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 ∞ if E is an infinite set. (This measure is called the counting measure.) State the Fubini and Tonelli Theorems explicitly for this case.

16. Let (X, \mathcal{A}, μ) be any σ -finite measure space and Y the set of positive integers with ν the counting measure (Problem 15). Then Theorem 20 and Corollary 11.14 state the same conclusion. However, Corollary 11.14 is

valid even if μ is not σ -finite, and hence the Tonelli Theorem is true without σ -finiteness if (Y, \mathcal{G}, ν) is this special measure space.

17. 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.

18. 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 μ and ν are σ -finite.)

19. Show that Tonelli's Theorem is still true if, instead of assuming μ and ν to be σ -finite, we merely assume that $\{\langle x, y \rangle : f(x, y) \neq 0\}$ is a set of σ -finite measure.

20. 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

$$f(x, y) = \begin{cases} 2 - 2^{-x} & \text{if } x = y \\ -2 + 2^{-x} & \text{if } x = y + 1 \\ 0 & \text{otherwise.} \end{cases}$$

21. The following example shows that we cannot remove the hypothesis that f be integrable from the Fubini Theorem or that μ and ν are σ -finite from the Tonelli Theorem: Let $X = Y$ be the interval $[0, 1]$, with $\mathcal{A} = \mathcal{G}$ the class of Borel sets. Let μ be Lebesgue measure and ν 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.

22. 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 Ω . Let $\mathcal{A} = \mathcal{G}$ be the σ -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.

23. Show that if (X, \mathcal{G}, μ) and (Y, \mathcal{B}, ν) are two σ -finite measure spaces, then $\mu \times \nu$ is the only measure on $\mathcal{G} \times \mathcal{B}$ which assigns the value $\mu A \nu B$ to each measurable rectangle $A \times B$. Show that a measure on $\mathcal{G} \times \mathcal{B}$ with this property need not be unique, if we do not have σ -finiteness.

24. a. Show that if $E \in \mathcal{G} \times \mathcal{B}$, then $E_x \in \mathcal{G}$ for each x .

b. If f is measurable with respect to $\mathcal{G} \times \mathcal{B}$, then f_x is measurable with respect to \mathcal{G} for each x .

25. 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 \mathcal{G}_δ , 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 φ given by $\varphi(y) = f(x - y)g(y)$ is integrable. If we denote its integral by $h(x)$, then h is integrable and

$$\int |h| \leq \int |f| \int |g|.$$

26. 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

$$\widehat{f * g} = \hat{f}\hat{g}.$$

27. Let f be a nonnegative integrable function on $(-\infty, \infty)$, and let m_2 be two-dimensional Lebesgue measure on \mathbf{R}^2 . Then

$$m_2\{\langle x, y \rangle : 0 \leq y \leq f(x)\} = m_2\{\langle x, y \rangle : 0 < y < f(x)\} = \int f(x) dx.$$

Let $\varphi(t) = m\{x: f(x) \geq t\}$. Then φ is a decreasing function and

$$\int_0^\infty \varphi(t) dt = \int f(x) dx.$$

*5 Inner Measure

Let μ be a measure on an algebra \mathfrak{A} and μ^* the induced outer measure. Then μ^*E may be thought of as the largest possible measure for E compatible with μ . We can also define an inner measure μ_* which assigns to a given set E the smallest measure compatible with μ :

Definition: Let μ be a measure on an algebra \mathfrak{A} and μ^* the induced outer measure. We define the inner measure μ_* induced by μ 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. A set E of finite outer measure was defined to be measurable if $\mu^*E = \mu_*E$, and the measurability of sets of infinite outer measure was defined in terms of intersections with sets of finite measure. Even in the case of a finite measure μ this procedure is more cumbersome than the elegant approach due to Carathéodory, which we have followed in this chapter. Apart from this historical importance, inner measure is useful for the extension of μ 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 μ to a σ -algebra containing \mathfrak{A} . The purpose of this section is to do this and to develop the basic properties of inner measure.

21. 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 $E \in \alpha$, take $A = E$ in the definition of μ_* . Then $\mu_*E \geq \mu E = \mu^*E$. ■

22. 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 $A \in \alpha$ 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 μ_*E more easily.

23. Lemma: Let A and B be two sets in α 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 $E \subset A$. ■

24. Corollary: If $A \in \alpha$, 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 $\mu^*(A \sim F) < \infty$. ■

The following lemma expresses the additivity of μ_* 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 3 and 32e).

25. 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: Replacing E by $E \cap \bigcup_{i=1}^{\infty} A_i$, we may suppose that $E \subset \bigcup_{i=1}^{\infty} A_i$. For any $B \in \mathfrak{A}$ with $\mu^*(B \sim E) < \infty$, we have

$$\begin{aligned} \mu B - \mu^*(B \sim E) &= \sum_{i=1}^{\infty} \mu(A_i \cap B) - \mu^*(A_i \cap B \cap \tilde{E}) \\ &= \sum_{i=1}^{\infty} \mu_*(A_i \cap B \cap E) \\ &\leq \sum_{i=1}^{\infty} \mu_*(A_i \cap E). \end{aligned}$$

Taking the supremum over all such B gives

$$\mu_* E \leq \sum \mu_*(A_i \cap E).$$

Let B_i be any set in \mathfrak{A} with $\mu^*(B_i \sim (E \cap A_i)) < \infty$.

Then

$$\begin{aligned} \mu B_i - \mu^*(B_i \sim (E \cap A_i)) &= \mu_*(B_i \cap A_i \cap E) \\ &= \mu(B'_i) - \mu^*(B'_i \cap \tilde{E}), \end{aligned}$$

where $B'_i = B_i \cap A_i$. Since the sets B'_i are disjoint and measurable,

$$\begin{aligned} \sum_{i=1}^n \mu B_i - \mu^*(B_i \sim (E \cap A_i)) &= \mu \left[\bigcup_{i=1}^n B'_i \right] - \mu^* \left[\left(\bigcup_{i=1}^n B'_i \right) \cap \tilde{E} \right] \\ &\leq \mu_* E. \end{aligned}$$

Taking suprema over all choices of $B_i \in \mathfrak{A}$ with

$$\mu^*(B_i \sim (E \cap A_i)) < \infty,$$

we have

$$\sum_{i=1}^n \mu_*(E \cap A_i) \leq \mu_* E.$$

Thus

$$\sum_{i=1}^{\infty} \mu_*(E \cap A_i) \leq \mu_* E. \blacksquare$$

26. Theorem: Let μ be a measure on an algebra \mathfrak{A} of subsets of X and E any subset of X . If \mathfrak{G} is the algebra generated by \mathfrak{A} and E and if $\tilde{\mu}$ is any extension of μ to \mathfrak{G} , then

$$\mu_* E \leq \tilde{\mu} E \leq \mu^* E.$$

Moreover, there are extensions $\bar{\mu}$ and $\underline{\mu}$ of μ to \mathfrak{G} (and hence also to the σ -algebra generated by \mathfrak{G}) 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{G} 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{G} and containing \mathfrak{A} and E . For each $B \in \mathfrak{G}$ 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{G} , and it follows from Corollary 24 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} .

Let $B_i = (A_i \cap E) \cup (A'_i \cap \tilde{E})$ be a disjoint sequence of sets whose union B is also in \mathfrak{B} . Then $B = (A \cap E) \cup (A' \cap \tilde{E})$, where $A = \bigcup A_i$ and $A' = \bigcup A'_i$. If $i \neq j$, $B_i \cap B_j = \emptyset$, and hence $A_i \cap A_j \cap E = B_i \cap B_j \cap E = \emptyset$ and $A'_i \cap A'_j \cap \tilde{E} = B_i \cap B_j \cap \tilde{E} = \emptyset$. Let $C_n = A_n \cap \tilde{A}_1 \cap \dots \cap \tilde{A}_{n-1}$, and $C'_n = A'_n \cap \tilde{A}'_1 \cap \dots \cap \tilde{A}'_{n-1}$. Then C_i and C'_i are in \mathfrak{A} , and $C_i \cap C_j = \emptyset$, $C'_i \cap C'_j = \emptyset$ for $i \neq j$. Also $C_n \cap E = A_n \cap \tilde{A}_1 \cap \dots \cap \tilde{A}_{n-1} \cap E = A_n \cap E$, since we have $A_n \cap \tilde{A}_j \cap E = (A_n \cap \tilde{A}_j \cap E) \cup (A_n \cap A_j \cap E) = A_n \cap E$. Similarly, $C'_n \cap \tilde{E} = A'_n \cap \tilde{E}$, and so $B_n = (C_n \cap E) \cup (C'_n \cap \tilde{E})$. Now $B \cap E = \bigcup (C_n \cap E)$ and $B \cap \tilde{E} = \bigcup (C'_n \cap \tilde{E})$. Since C_n and C'_n are disjoint sequences of sets in \mathfrak{A} , we have

$$\mu^*(B \cap E) = \sum \mu^*(C_n \cap E) = \sum \mu^*(B_n \cap E)$$

and

$$\mu^*(B \cap \tilde{E}) = \sum \mu^*(C'_n \cap \tilde{E}) = \sum \mu^*(B_n \cap \tilde{E})$$

by Problem 3, and

$$\mu_*(B \cap E) = \sum \mu_*(C_n \cap E) = \sum \mu_*(B_n \cap E)$$

and

$$\mu_*(B \cap \tilde{E}) = \sum \mu_*(C'_n \cap \tilde{E}) = \sum \mu_*(B_n \cap \tilde{E}).$$

Thus $\bar{\mu}B = \sum \bar{\mu}B_n$ and $\underline{\mu}B = \sum \underline{\mu}B_n$, showing that $\bar{\mu}$ and $\underline{\mu}$ are countably additive and hence measures on \mathfrak{B} . ■

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.

27. 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 $\mu^*(A_n \sim E) < \infty$ and $\mu A_n - \mu^*(A_n \sim E) > \mu_*E - \frac{1}{n}$. 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, $\bar{\mu}H_n = \mu A_n - \bar{\mu}G_n > \mu_*E - \frac{1}{n}$. Thus the proposition is proved if we take $H = \bigcup H_n$. ■

28. 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 27 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.

Suppose that E is a measurable set of finite outer measure. Then E is contained in the union of a disjoint sequence $\langle A_n \rangle$ of sets of α of finite measure. By Proposition 25 and the measurability of E we have

$$\begin{aligned}\mu_*E &= \sum \mu_*(E \cap A_n) \\ &= \sum \mu A_n - \mu^*(A_n \cap \tilde{E}) \\ &= \sum \bar{\mu}(A_n \cap E) \\ &= \bar{\mu}E = \mu^*E.\blacksquare\end{aligned}$$

29. 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 μ_*E or μ_*F is infinite, the first inequality follows from the monotonicity of μ_* . If μ_*E and μ_*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 any set in α 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. ■

Problems

28. a. If $\mu X < \infty$, then $\mu_*E = \mu X - \mu^*(\tilde{E})$.

b. If \mathcal{Q} is a σ -algebra, then

$$\mu^*E = \inf \{\mu A : E \subset A, A \in \mathcal{Q}\}$$

and

$$\mu_*E = \sup \{\mu A : A \subset E, A \in \mathcal{Q}\}.$$

c. If μ is Lebesgue measure on \mathbf{R} , then $\mu_*E = \sup \{\mu F : F \subset E, F \text{ closed}\}$.

29. 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 σ -finite outer measure, then E has a measurable kernel and a measurable cover.

30. Let μ be a measure on an algebra \mathcal{Q} and E a set with $\mu^*E < \infty$. If β is any real number with $\mu_*E \leq \beta \leq \mu^*E$, then there is an extension $\tilde{\mu}$ of μ to a σ -algebra containing \mathcal{Q} and E such that $\tilde{\mu}E = \beta$.

31. a. If μ is a semifinite measure on an algebra \mathcal{Q} and if \mathfrak{G} is the smallest σ -algebra containing \mathcal{Q} , then μ_* restricted to \mathfrak{G} is a semifinite measure and is the smallest extension of μ to \mathfrak{G} .

b. There may be more than one semifinite extension to \mathfrak{G} (for example, let $X = \mathbf{N} \cup \{\omega\}$ and \mathcal{Q} the algebra generated by the finite subsets of \mathbf{N} with μ the counting measure).

32. Let \mathcal{Q} 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{G} of Borel sets is the smallest σ -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 μ_* to \mathfrak{G} is not a measure. Hence there is no smallest extension of μ to \mathfrak{G} .

d. The counting measure on \mathfrak{G} is an extension of μ to \mathfrak{G} .

e. Show that Lemma 25 is no longer true if we replace “sets in \mathfrak{A} ” by “measurable sets”.

33. If μ^* 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 $\mu_*E = \mu^*X - \mu^*(\bar{E})$. 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 μ_* .

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 29 fail.

34. 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 μA be the sum of the lengths of the intervals of which A is composed. Then μ 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}_δ .

c. Assuming that there is no finite nonatomic measure defined on $\mathcal{P}(\mathbf{R})$, show that each extension of μ to a measure on a σ -algebra must assign to E the value 0 or ∞ .

*6 Extension by Sets of Measure Zero

The results of Section 2 allow us to extend a measure μ on an algebra \mathfrak{A} to a σ -algebra containing \mathfrak{A} and those of Section 5 provide for the extension from \mathfrak{A} to a σ -algebra containing \mathfrak{A} and one additional set. It is sometimes useful to be able to extend to a σ -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.

30. Proposition: Let μ be a measure on a σ -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 μ to the smallest σ -algebra \mathfrak{G} 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 $\mathfrak{G} = \{B : B = A \triangle M, A \in \mathfrak{A}, M \in \mathfrak{M}\}$ is a σ -algebra containing \mathfrak{A} and \mathfrak{M} , and since each σ -algebra containing \mathfrak{A} and \mathfrak{M} contains \mathfrak{G} , \mathfrak{G} is the smallest σ -algebra containing \mathfrak{A} and \mathfrak{M} .

If $B = A_1 \triangle M_1 = A_2 \triangle M_2$, then $A_1 \triangle A_2 = M_1 \triangle M_2$, and so $\mu(A_1 \triangle A_2) = 0$. Thus $\mu A_1 = \mu A_2$, and, if we define $\bar{\mu}B$ to be μA_1 , then $\bar{\mu}$ is well defined on \mathfrak{G} and is an extension of μ . It remains only to show that $\bar{\mu}$ is countably additive.

Let $B = \bigcup B_i$, $B_i \cap B_j = \emptyset$. If $B_i = A_i \triangle M_i$, then $A_i \triangle A_j \in \mathfrak{M}$. 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 $A_n \triangle A'_n \in \mathfrak{M}$. Thus $B_i = A'_i \triangle M'_i$, and $B = A \triangle M$, 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 $A \in \mathfrak{A}$ with $A \subset M$ simply states that $\mu_*M = 0$. Thus the proposition states that we can extend the domain of μ 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 σ -algebra generated by \mathfrak{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.

Problem

35. Let \mathfrak{A} be a σ -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

$$\mathfrak{G} = \{B : B = A \triangle M, A \in \mathfrak{A} \text{ and } M \in \mathfrak{M}\}$$

is a σ -algebra.

36. Show that Proposition 30 need not be true if α is only an algebra of sets.

*7 Carathéodory Outer Measure

Let X be a set of points and Γ a set of real-valued functions on X . It is often of interest to know conditions under which an outer measure μ^* will have the property that every function in Γ 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 φ* if there are numbers a and b with $a > b$ such that φ is greater than a on one and less than b on the other. An outer measure μ^* is called a **Carathéodory outer measure** (with respect to Γ) if it satisfies the following axiom:

- iv. If A and B are two sets which are separated by some function in Γ , then $\mu^*(A \cup B) = \mu^*A + \mu^*B$.

31. Proposition: If μ^* is a Carathéodory outer measure, then every function in Γ is μ^* -measurable.

Proof: Given the real number a and the function $\varphi \in \Gamma$, we must show that the set

$$E = \{x: \varphi(x) > a\}$$

is μ^* -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

$$B_n = \left\{x: (x \in B) \& \left(\varphi(x) > a + \frac{1}{n}\right)\right\}.$$

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 φ separates R_n and B_{n-2} and hence

separates R_{2k} and $\bigcup_{j=1}^{k-1} R_{2j}$, since this latter set is contained in B_{2k-2} . Consequently,

$$\begin{aligned}\mu^* \left[\bigcup_{j=1}^k R_{2j} \right] &= \mu^* R_{2k} + \mu^* \left[\bigcup_{j=1}^{k-1} R_{2j} \right] \\ &= \sum_{j=1}^k \mu^* R_{2j},\end{aligned}$$

by induction. Since

$$\bigcup_{j=1}^k R_{2j} \subset B \subset A,$$

we have

$$\sum_{j=1}^k \mu^* R_{2j} \leq \mu^* A,$$

and so the series $\sum_{j=1}^{\infty} \mu^* R_{2j}$ converges. Similarly, the series

$$\sum_{j=1}^{\infty} \mu^* R_{2j+1}$$

converges, and 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 μ^* ,

$$\begin{aligned}\mu^* B &\leq \mu^* B_n + \sum_{k=n+1}^{\infty} \mu^* R_k \\ &< \mu^* B_n + \epsilon\end{aligned}$$

or

$$\mu^* B_n > \mu^* B - \epsilon.$$

Now

$$\begin{aligned}\mu^*A &\geq \mu^*(B_n \cup C) \\ &= \mu^*B_n + \mu^*C\end{aligned}$$

since φ separates B_n and C . Consequently,

$$\mu^*A \geq \mu^*B + \mu^*C - \epsilon.$$

Since ϵ is an arbitrary positive quantity,

$$\mu^*A \geq \mu^*B + \mu^*C. \blacksquare$$

32. Proposition: Let (X, ρ) be a metric space, and let μ^* 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 μ^* .

Problem

37. Prove Proposition 32. [Let Γ be the set of functions φ of the form $\varphi(x) = \rho(x, E)$. Show that μ^* satisfies (iv) with respect to Γ , and note that for a closed set F we have $F = \{x: \rho(x, F) \leq 0\}$.]

13 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 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 $f \rightarrow f^+ = f \vee 0$. 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, for $f^+ = \frac{1}{2}(f + |f|)$.

A linear functional I on L is said to be positive¹ if $I(\varphi) \geq 0$ for each nonnegative function φ 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 which 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 φ is a function in L such that² $\varphi \leq \lim \varphi_n$, then $I(\varphi) \leq \lim I(\varphi_n)$.

D''. If $\langle u_n \rangle$ is a sequence of nonnegative functions in L , and if φ 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 μ 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 μ . A third example is given by taking L to be the class of all functions integrable with respect to a measure μ .

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 αf 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 $h = f + g$. This

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.

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.

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 μ be a measure on an algebra \mathfrak{Q} , L the class of simple functions on \mathfrak{Q} , and I integration with respect to μ . 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. Let L be a vector lattice of extended real-valued functions on X and I a positive linear functional on L , as defined in this section. Let $f \in L$, and let $E = \{x: f(x) = \infty\}$. Show that any extended real-valued function h which vanishes outside E is in L and $I(h) = 0$.

2 The Extension Theorem

We begin by considering the class L_u consisting of all those extended real-valued functions 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 α and β 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 ∞). 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:

1. 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 α and β 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 $\varphi_n \vee \psi_n \uparrow f \vee g$.

2. 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 non-negative and $\varphi_n \uparrow f$ with $\varphi_n \in L$. By replacing φ_n by $\varphi_n \vee 0$, we may assume that each φ_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

$$\begin{aligned} I(f) &= \lim I(\varphi_n) \\ &= \lim I\left(\sum_{\nu=1}^n \psi_\nu\right) \\ &= \lim \sum_{\nu=1}^n I(\psi_\nu) \\ &= \sum_{\nu=1}^{\infty} I(\psi_\nu). \blacksquare \end{aligned}$$

3. 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 non-negative functions in L and so must be in L_u . Also

$$\begin{aligned} I(f) &= \sum_{v,n} I(\psi_{n,v}) \\ &= \sum_{n=1}^{\infty} I(f_n). \blacksquare \end{aligned}$$

For an arbitrary function f on X we define the upper integral $\bar{I}(f)$ by setting

$$\bar{I}(f) = \inf_{\substack{g \geq f \\ g \in L_u}} I(g),$$

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 4 follow directly from the definition of \bar{I} .

4. Lemma: Let $h = f + g$. Then $\bar{I}(h) \leq \bar{I}(f) + \bar{I}(g)$, provided the right-hand side is defined. If $c \geq 0$, $\bar{I}(cf) = c\bar{I}(f)$. If $f \leq g$, then $\bar{I}(f) \leq \bar{I}(g)$ and $\underline{I}(f) \leq \underline{I}(g)$.

5. Lemma: We have $\underline{I}(f) \leq \bar{I}(f)$. If $f \in L_u$, then $\underline{I}(f) = \bar{I}(f) = I(f)$.

Proof: We have $0 = I(0) = I(f - f) \leq \bar{I}(f) + \bar{I}(-f)$. Hence $\underline{I}(f) = -\bar{I}(-f) \leq \bar{I}(f)$.

To prove the second statement, we note that if $f \in L_u$ then $\bar{I}(f) \leq I(f)$ by the definition of \bar{I} . If $g \in L_u$ and $f \leq g$, then $I(f) \leq I(g)$ and so $I(f) \leq \bar{I}(f)$, whence $\bar{I}(f) = I(f)$. If $\varphi \in L$, then $-\varphi \in L \subset L_u$, and so $\bar{I}(-\varphi) = I(-\varphi) = -I(\varphi)$, and hence $\underline{I}(\varphi) = I(\varphi)$. But each $f \in L_u$ is the limit of an increasing sequence $\langle \varphi_v \rangle$ of functions in L . Since $f \geq \varphi_v$, we have $\underline{I}(f) \geq \underline{I}(\varphi_v) = I(\varphi_v)$. Hence

$$\underline{I}(f) \geq \lim I(\varphi_v) = I(f).$$

Since $\underline{I}(f) \leq \bar{I}(f) = I(f)$, we have $\underline{I}(f) = I(f)$. \blacksquare

6. Lemma: Let $\langle f_v \rangle$ be a sequence of nonnegative functions, and let $f = \sum_{v=1}^{\infty} f_v$. Then $\bar{I}(f) \leq \sum_{v=1}^{\infty} \bar{I}(f_v)$.

Proof: If $\bar{I}(f_\nu) = \infty$ for some ν , we are done. If not, given $\epsilon > 0$, there is a function $g_\nu \in L_u$ such that $f_\nu \leq g_\nu$ and $I(g_\nu) \leq \bar{I}(f_\nu) + \epsilon \cdot 2^{-\nu}$. Since each g_ν is nonnegative, Lemma 3 implies that the function $g = \sum g_\nu$ is in L_u and that $I(g) = \sum I(g_\nu) \leq \sum \bar{I}(f_\nu) + \epsilon$. Since $g \geq f$, we have

$$\bar{I}(f) \leq \sum_{\nu=1}^{\infty} \bar{I}(f_\nu) + \epsilon,$$

and the lemma follows since ϵ was an arbitrary positive number. ■

We shall call a function f on X **integrable** with respect to I (or \underline{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:

7. 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 \underline{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 5 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 $I(g) < I(f) + \epsilon < \infty$ 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 $-I(h_1) \leq I(f^+) \leq \bar{I}(f^+) \leq I(g_1)$, we have $\bar{I}(f^+) - I(f^+) < 2\epsilon$, and so $\bar{I}(f^+) = I(f^+)$, since ϵ 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.

8. 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 6,

$$\begin{aligned}\bar{I}(g) &\leq \sum_{n=1}^{\infty} I(f_{n+1} - f_n) \\ &= \sum_{n=1}^{\infty} I(f_{n+1}) - I(f_n) \\ &= \lim I(f_n) - I(f_1).\end{aligned}$$

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 $I(f) \geq I(f_n)$, and so

$$I(f) \geq \lim I(f_n).$$

Thus $I(f) = \bar{I}(f) = \lim I(f_n)$. ■

9. 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.

10. 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 $I(-g_n) \leq 0$, we must have $g \in L_1$ by Proposition 8.

To prove the rest of the proposition, let $h_n = \inf_{\nu \geq n} f_\nu$. Then $\langle h_n \rangle$ is a sequence of nonnegative functions in L_1 which increases to $\underline{\lim} f_\nu$.

Since $h_n \leq f_\nu$ for $n \leq \nu$, $\lim I(h_n) \leq \underline{\lim} I(f_\nu) < \infty$. Hence $\underline{\lim} f_\nu \in L_1$ and $I(\underline{\lim} f_\nu) \leq \underline{\lim} I(f_\nu)$ by Proposition 8. ■

11. 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 $|f_n| \leq g$. Then if $f = \lim f_n$, we have*

$$I(f) = \lim I(f_n).$$

Proof: The functions $f_n + g$ are nonnegative, and $I(f_n + g) \leq 2I(g)$. Hence by Proposition 10 we have $f + g$ in L_1 and

$$I(f + g) \leq \underline{\lim} I(f_n + g) = I(g) + \underline{\lim} I(f_n).$$

Hence

$$I(f) \leq \underline{\lim} I(f_n).$$

Since the functions $g - f_n$ are also nonnegative, we have

$$\begin{aligned} I(g - f) &\leq \underline{\lim} I(g - f_n) \\ &= I(g) - \overline{\lim} I(f_n). \end{aligned}$$

Hence

$$\overline{\lim} I(f_n) \leq I(f),$$

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 $g \in L_u$, $g \geq 0$, and $\varphi \in L$.
6. Prove Lemma 4.
7. Show that the ambiguous case in Lemma 4 does not cause difficulties in Lemma 5.

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 8 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 $h_n \in L_u$ 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 $\bar{I}(f) \leq I(g_n) \leq \bar{I}(f) + \frac{1}{n}$. 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:

12. Lemma: *If f is any function on X with $\bar{I}(f)$ finite, then there is a $g \in L_{ul}$ such that $f \leq g$ and $\bar{I}(f) = I(g)$.*

A function f on X is called a **null function** if $f \in L_1$ and $I(|f|) = 0$. If f is a null function and $|g| \leq f$, $0 \leq I(|g|) \leq \bar{I}(|g|) \leq I(f) = 0$. Hence $g \in L_1$, and g is a null function.

13. 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 12 asserts the existence of $g \in L_{ul}$ 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 12 there is a function $k \in L_{ul}$ with $|h| \leq k$ and $I(k) = I(|h|) = 0$. ■

14. 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 $f \in L$, then $\Lambda_1 \supset L_1$ and $I(f) = J(f)$ for all $f \in L_1$.

Proof: By applying Proposition 8 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 13 each function which is null with respect to I must also be null with respect to J . By the first part of Proposition 13 every function f in L_1 must be in Λ_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 .

15. 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 11. ■

16. Lemma: A nonnegative function f on X is measurable with respect to I if $\varphi \wedge f$ is in L_1 for each φ in L .

Proof: If $\varphi \in L$, then $\varphi \wedge f$ is in L_1 , and Proposition 8 implies that $g \wedge f \in L_1$ for $g \in L_u$ and $I(g) < \infty$. It follows from Proposition 11 that $g \wedge f \in L_1$ for all $g \in L_{ul}$. If h is any function in L_1 , Proposition 13 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 χ_A is measurable. We say that A is **integrable** if its characteristic function χ_A is integrable. Note that a measurable subset of an integrable set is itself integrable.

17. 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 \mathcal{A} of measurable sets is a σ -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

$$\chi_A = \lim (\chi_{A_1} \vee \dots \vee \chi_{A_n})$$

and the measurability of A follows from Lemma 15. 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. ■

18. Lemma: *If 1 is a measurable function and f a nonnegative integrable function, then for each real number α the set*

$$E = \{x: f(x) > \alpha\}$$

is measurable.

Proof: If α is negative, $E = X$ and is measurable, since $\chi_E = 1$ and 1 is measurable. Hence we assume $\alpha \geq 0$. 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 $x \in E$, and $g(x) = 0$ for $x \in \tilde{E}$. Let $\varphi_n = 1 \wedge (ng)$. Then $\varphi_n \in L_1$, and $\varphi_n \uparrow \chi_E$. Hence χ_E is measurable, and so E is measurable. ■

19. Lemma: *Let the function 1 be measurable, and define a set function μ on the class \mathfrak{A} of measurable sets by*

$$\mu E = I(\chi_E)$$

if χ_E is integrable, and $\mu E = \infty$ otherwise. Then μ 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 μ 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 8, since $\chi_E = \sum \chi_{E_i}$. In either case $\mu E = \sum \mu E_i$, and the measure μ is countably additive. ■

This measure μ 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 μ .

20. 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 σ -algebra \mathfrak{A} of subsets of X and a measure μ 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 μ . 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 16 that 1 is measurable. Lemma 17 then asserts that \mathfrak{A} is a σ -algebra, and Lemma 18 asserts that each non-negative 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 \mathcal{Q} .

Let μ be the measure given in Lemma 19, 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

$$\begin{aligned} I(\varphi_n) &= 2^{-n} \sum_{k=1}^{2^{2n}} I(\chi_{E_{k,n}}) \\ &= 2^{-n} \sum_{k=1}^{2^{2n}} \mu(E_{k,n}) \\ &= \int \varphi_n d\mu. \end{aligned}$$

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 μ . 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 μ and that

$$I(f) = \int f d\mu.$$

If f is a nonnegative function on X which is integrable with respect to μ , we construct $E_{k,n}$ and φ_n as before. Since $\int f d\mu < \infty$, each $E_{k,n}$ has finite measure, and so $\chi_{E_{k,n}}$ and hence φ_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 8. Thus each f which is integrable with respect to μ is also integrable with respect to I . ■

21. Proposition: *Let L be a vector lattice of functions on a set X , and suppose that $1 \in L$. Let \mathfrak{B} be the smallest σ -algebra of subsets of X such that each function in L is measurable with respect to \mathfrak{B} . Then for each Daniell integral I there is a unique measure μ on \mathfrak{B} such that for every $f \in L$*

$$I(f) = \int f d\mu.$$

Proof: The existence of μ is a special case of Theorem 20 and we have only to prove the uniqueness of μ on \mathfrak{B} . Let \mathfrak{A} be the σ -algebra of measurable sets given by Lemma 17. Lemma 18 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 χ_B are in L_1 for each B in \mathfrak{A} and hence for each B in \mathfrak{B} . The uniqueness of μ 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 Λ be the set of functions on X which are measurable with respect to \mathfrak{B} and integrable with respect to μ and set

$$J(f) = \int f d\mu$$

for $f \in \Lambda$, then Proposition 14 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 μ is uniquely determined on \mathfrak{B} by I . ■

We can still establish the uniqueness of the measure μ 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 μ need not be unique on \mathfrak{B} (Problem 11).

Problems

- 8.** Let μ be a measure on an algebra \mathcal{Q} of sets, and let L be the family consisting of those functions which are finite linear combinations of characteristic functions of sets in \mathcal{Q} with finite measure, and let I be integration with respect to μ . Discuss the extension of I to L_1 , and compare this process with the Carathéodory extension for μ .
- 9.** Prove directly from the definition that, if f_1 and f_2 are two non-negative measurable functions, then $f_1 + f_2$ is measurable.
- 10.** Prove that the conclusion of Proposition 21 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 μ is unique on sets of \mathcal{G} 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 ω . Then L is a vector lattice, and the smallest σ -algebra with respect to which each function in L is measurable is the family \mathcal{G} 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 μ constructed in Theorem 20? Show that there is another measure ν defined on \mathcal{G} such that for each f in L , $I(f) = \int f d\nu$.
- 12. a.** Define a measure μ 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 μ 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 μ is the unique measure such that $I(f) = \int f d\mu$ and $\mu(X) = \|I\|$.

14 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 which vanish outside a compact subset of X . If f is a real-valued function, we define the *support* of f to be 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 which have compact support. The class of **Baire sets** is defined to be the smallest σ -algebra \mathfrak{B} of subsets of X such that each function in $C_c(X)$ is measurable with respect to \mathfrak{B} . Thus \mathfrak{B} is the σ -algebra generated by the sets $\{x: f(x) \geq \alpha\}$ with $f \in C_c(X)$. If $\alpha > 0$, these sets are compact g_δ 's. It follows from Proposi-

tion 9.20 that each compact \mathcal{G}_δ is a Baire set. Consequently, \mathfrak{B} is the σ -algebra generated by the compact \mathcal{G}_δ 's.

If X is any topological space, the smallest σ -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 metric space, but there are compact spaces where the class of Borel sets is larger than the class of Baire sets (Problem 5).

The reader should be warned that not everyone uses the same definitions for Baire and Borel sets. When X is compact (or σ -compact), everyone agrees that the Baire sets are the smallest σ -algebra such that each f in $C(X)$ is measurable and that the Borel sets are the smallest σ -algebra containing the closed sets, but definitions differ when X is allowed to be locally compact. Some authors like to take the Baire sets to be the smallest σ -ring containing the compact \mathcal{G}_δ 's; others the smallest σ -algebra such that each f in $C(X)$ is measurable. Some authors prefer to take the class of Borel sets to be the smallest σ -algebra (or σ -ring) containing the compact sets. The definitions we have made here seem to me the most useful.

A set is called *σ -compact* if it is the union of a countable collection of compact sets. A set which is contained in a compact set is called *bounded*, and one which is contained in a σ -compact set is called *σ -bounded*. The following lemma is useful.

1. Lemma: *If B is a Baire set, then either B or \tilde{B} is σ -bounded.*

Proof: The collection of sets B such that either B or \tilde{B} is σ -bounded is a σ -algebra. Since it contains all compact sets, it must contain all Baire sets. ■

A measure μ defined on the σ -algebra of Baire sets is called a *Baire measure* if it is finite for each compact Baire set. A measure defined on the Borel sets is called a *Borel measure* if it is finite for each compact set. In the next two sections we shall show that each positive linear functional on $C_c(X)$ is given by integrating with respect to a suitable Baire measure and that for compact spaces X each bounded linear functional on $C(X)$ is given by a finite signed Baire measure. In Section 4 we show that each Baire measure can be extended to a Borel measure.

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. Show that the collection of compact \mathcal{G}_δ 's is closed with respect to finite unions and intersections.
3.
 - a. Show that each compact set in a locally compact space is contained in an open σ -compact set O with \overline{O} compact.
 - b. Show that each σ -bounded set is contained in an open σ -compact set.
4. 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 ∞ .
 - c. If B is a Baire set in X^* , then $B \cap X$ is a Baire set in X .
5. 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^* ? Show that X^* has a compact subset which is not a Baire set.
 - e. Show that there is a Baire measure μ on X such that $\mu(X) = 1$ and $\int f d\mu = 0$ for each f in $C_0(X)$.
6. Let X and Y be two locally compact Hausdorff spaces.
 - a. Show that each $f \in C_c(X \times Y)$ is the limit of sums of the form $\sum_{i=1}^n \varphi_i(x)\psi_i(y)$. [Hint: The Stone-Weierstrass Theorem is useful.]
 - b. If \mathfrak{A} and \mathfrak{B} denote the Baire subsets of X and Y , respectively, then the class of Baire subsets of $X \times Y$ is $\mathfrak{A} \times \mathfrak{B}$.
 - c. Let X and Y each be the Alexandroff one-point compactification of a discrete uncountable set. Then the algebra of Borel subsets of $X \times Y$ is larger than the σ -algebra which is the product of the Borel sets of X and the Borel sets of Y .

2 Positive Linear Functionals and Baire Measures

In the present section we show that each positive linear functional on $C_c(X)$ is given by integration with respect to a Baire measure and that this measure is essentially unique. We begin with two lemmas about compact \mathcal{G}_δ 's and Baire measures.

2. Lemma: *If K is a compact \mathcal{G}_δ , there is a sequence $\langle \varphi_n \rangle$ of functions in $C_c(X)$ such that $\varphi_n \downarrow \chi_K$.*

Proof: By Proposition 9.20 there is a nonnegative function $\varphi \in C_c(X)$ with $\varphi \equiv 1$ on K and $0 \leq \varphi < 1$ on \tilde{K} . Let $\varphi_n = \varphi^n$. ■

3. Lemma: *Let μ_1 and μ_2 be Baire measures on a compact Hausdorff space X . If $\mu_1 K = \mu_2 K$ whenever K is a compact \mathcal{G}_δ , then $\mu_1 = \mu_2$.*

Proof: The semialgebra \mathcal{C} generated by the compact \mathcal{G}_δ 's is the collection of all sets of the form $C = K_1 \cap \tilde{K}_2$, where K_1 and K_2 are compact \mathcal{G}_δ 's with $K_2 \subset K_1$. Since $\mu_1 C = \mu_1 K_1 - \mu_1 K_2$, we see that μ_1 and μ_2 agree on \mathcal{C} . Propositions 12.9 and 12.8 imply that their extensions to the smallest σ -algebra containing \mathcal{C} are identical, since $\mu_1 X = \mu_2 X < \infty$. ■

4. Corollary: *Let μ_1 and μ_2 be Baire measures on a locally compact Hausdorff space such that $\mu_1 K = \mu_2 K$ whenever K is a compact \mathcal{G}_δ . Then $\mu_1 E = \mu_2 E$ for each σ -bounded Baire set E .*

This corollary shows that so far as the σ -bounded sets are concerned the value of a Baire measure is determined by its values on compact \mathcal{G}_δ 's. It is possible, unfortunately, for two Baire measures to agree on the σ -bounded Baire sets and yet differ on sets which are not σ -bounded.¹ (See Problem 10.)

5. Theorem: *Let X be a locally compact Hausdorff space and I a positive linear functional on $C_c(X)$. Then there is a Baire measure μ such that for each f in $C_c(X)$ we have*

$$I(f) = \int f d\mu.$$

If X is compact, the measure μ is unique. In general μ is uniquely determined on the σ -bounded Baire sets.

¹ This is one of the principal reasons many authors prefer to do measure on σ -rings rather than σ -algebras, since the smallest σ -ring containing the compact \mathcal{G}_δ 's contains only σ -bounded sets.

Proof: The space $C_c(X)$ is a vector lattice, and I will be a Daniell integral if it satisfies the Daniell condition D. Let $\langle \varphi_n \rangle$ be a sequence of functions in $C_c(X)$ which decrease to zero and let K be the support of φ_1 . Let ψ be a nonnegative function in $C_c(X)$ which is positive on K . Then for a given $\epsilon > 0$ the sets $F_n = \{x: \varphi_n(x) \geq \epsilon\psi(x)\}$ are a decreasing family of closed subsets of K whose intersection is empty. Thus for some N we have $F_N = \emptyset$, and $\varphi_n < \epsilon\psi$ for $n \geq N$. Thus $I(\varphi_n) \leq \epsilon I(\psi)$ for $n \geq N$, and $I(\varphi_n) \rightarrow 0$, since ϵ was arbitrary. The existence of a Baire measure μ so that $I(f) = \int f d\mu$ for each $f \in C_c(X)$ now follows from the Stone Theorem (13.20).

To see the uniqueness of μ , we note that, if K is a compact \mathcal{G}_δ , then χ_K is the limit of a decreasing sequence $\langle \varphi_n \rangle$ of functions in $C_c(X)$. Since φ_1 is integrable, the Lebesgue Convergence Theorem asserts that

$$\mu K = \lim \int \varphi_n d\mu = \lim I(\varphi_n).$$

Thus μK is uniquely determined by I . By Corollary 4 the measure μ is uniquely determined on the σ -bounded Baire sets and hence on all Baire sets if X is compact. ■

The following proposition is the analogue for Baire measures of Proposition 3.15.

6. Proposition: *Let μ be a Baire measure on a locally compact Hausdorff space X . Then given a σ -bounded Baire set E ,*

$$\mu E = \inf \{\mu O: E \subset O, O \text{ a } \sigma\text{-compact open set}\}$$

and

$$\mu E = \sup \{\mu K: K \subset E, K \text{ a compact } \mathcal{G}_\delta\}.$$

Proof: To prove the first statement, we note that, since E is σ -bounded, it is contained in a σ -compact open set O . If $\mu E = \infty$, then $\mu O = \infty$ and there is nothing to prove. Hence we shall assume $\mu E < \infty$.

If E is a compact \mathcal{G}_δ , then there is a function $\varphi \in C_c(X)$ which is 1 on E and $0 \leq \varphi < 1$ on \tilde{E} . Let $O_n = \{x: \varphi(x) > 1 - 1/n\}$. Then each O_n is a σ -compact open set, $O_n \supset O_{n+1}$, and $E = \bigcap O_n$.

Since \bar{O}_1 is compact, $\mu O_1 < \infty$ and $\mu E = \lim \mu O_n$. Thus for some O_n we have $\mu O_n < \mu E + \epsilon$.

Let $E = E_1 \cap \tilde{E}_2$ where E_1 and E_2 are compact \mathcal{G}_δ 's with $E_2 \subset E_1$, and let U be a σ -compact open set with \bar{U} compact such that $E_1 \subset U$ and $\mu U < \mu E_1 + \epsilon$. Set $O = U \cap \tilde{E}_2$. Then O is the intersection of two \mathfrak{F}_σ 's and so is an \mathfrak{F}_σ . Since O is contained in the compact set \bar{U} , O must be σ -compact. Since $O \sim E = U \sim E_1$, we have $\mu(O \sim E) = \mu(U \sim E_1) = \mu U - \mu E_1 < \epsilon$. Thus $\mu O < \mu E + \epsilon$, and the first statement in the proposition holds for sets E in the semi-algebra \mathcal{C} generated by the compact \mathcal{G}_δ 's.

The algebra \mathcal{A} generated by the compact \mathcal{G}_δ 's is the collection of all finite unions of sets from \mathcal{C} , and \mathcal{A}_σ consists of all countable unions of sets from \mathcal{C} . By Proposition 12.6 each set E in the σ -algebra generated by \mathcal{A} (that is, each Baire set E) with $\mu E < \infty$ is contained in an $A \in \mathcal{A}_\sigma$ with $\mu A < \mu E + \frac{\epsilon}{2}$. Let A be the disjoint union of the sequence $\langle C_i \rangle$ from \mathcal{C} , and choose σ -compact open sets O_i such that $C_i \subset O_i$ and $\mu O_i < \mu C_i + \frac{\epsilon}{2^{i+1}}$. Then $O = \bigcup O_i$ is a σ -compact open set, $A \subset O$, and $\mu(O \sim A) < \frac{\epsilon}{2}$. Thus $O \supset E$ and $\mu O < \mu E + \epsilon$. This proves the first statement of the proposition.

If E is a σ -bounded Baire set, then $E \subset \bigcup K_i$ with K_i compact. Since each compact set is contained in a compact \mathcal{G}_δ , we may assume

each K_i is a compact \mathcal{G}_δ and hence a Baire set. Thus $E_n = E \cap \bigcup_{i=1}^n K_i$

is a bounded Baire set, and E is the union of the increasing sequence $\langle E_n \rangle$ of bounded Baire sets. Hence $\mu E = \lim \mu E_n$, and so $\mu E = \sup \{\mu B : B \subset E, B \text{ a bounded Baire set}\}$.

Let B be a bounded Baire set and C a compact \mathcal{G}_δ containing B . Since $C \sim B$ is a bounded Baire set, we can find a σ -compact open set O containing $C \sim B$ such that $\mu O < \mu(C \sim B) + \epsilon$. Let $K = C \cap \tilde{O}$. Then $K \subset B$, and K is a \mathcal{G}_δ since it is the intersection of two \mathcal{G}_δ 's, and compact since it is a closed subset of a compact set. Since $C \subset K \cup O$, we have

$$\mu C \leq \mu O + \mu K < \mu(C \sim B) + \mu K + \epsilon.$$

Hence

$$\mu B < \mu K + \epsilon$$

and

$$\mu B = \sup \{\mu K : K \subset B, K \text{ a compact } \mathcal{G}_\delta\}. \blacksquare$$

A Baire measure μ is called **regular** if for each Baire set E we have $\mu E = \sup \mu K$ as K ranges over the compact \mathcal{G}_δ 's contained in E . Thus Proposition 6 states that each Baire measure on a σ -compact space is regular. Some further properties of regular Baire measures are given in Problem 10.

We have used throughout this section the fact that if a compact set is a \mathcal{G}_δ it is a Baire set. The converse is also true: If a compact set is a Baire set, it is a \mathcal{G}_δ . Since we do not need this here, we refer the interested reader to Halmos [8], p. 221, for a proof.

Problems

7. Let \mathcal{K} be a collection of subsets of X with the property that the union of any two sets in \mathcal{K} is in \mathcal{K} and the intersection of any two sets in \mathcal{K} is in \mathcal{K} . Show that the algebra generated by \mathcal{K} consists of the sets which are finite disjoint unions of sets of the form $K_1 \cap \tilde{K}_2$, where K_1 and K_2 are in \mathcal{K} and $K_2 \subset K_1$.

8. Show that each compact set in a locally compact Hausdorff space is contained in a compact \mathcal{G}_δ .

9. An alternate proof of Theorem 5 can be given using the results of Section 12.7 instead of those from Chapter 13:

a. Define an outer measure μ^* on X by setting $\mu^*E = \inf \sum I(\varphi_\nu)$ for all sequences $\langle \varphi_\nu \rangle$ from $C_c(X)$ with $\sum \varphi_\nu \geq 1$ on E . Show that μ^* is a Carathéodory outer measure with respect to $C_c(X)$ and that hence each Baire set is μ^* measurable.

b. Let μ be μ^* restricted to the Baire sets. If f is a nonnegative function in $C_c(X)$ and $f \geq 1$ on a Baire set E , then $\mu(E) \leq I(f)$. If $f \leq 1$ and the support of f is contained in a compact Baire set K , then $\mu K \geq I(f)$.

c. Show that for any Baire measure with the properties in (b) and any $f \in C_c(X)$ we have

$$I(f) = \int f d\mu.$$

10. a. Show that a finite Baire measure μ on a locally compact Hausdorff space X is regular iff $\mu X = \sup \{K: K \text{ a compact } \mathcal{G}_\delta\}$. [Problem 11.9 is useful.]

b. If μ is a regular Baire measure on X , then for each Baire set E we have $\mu E = \inf \{\mu O: O \supseteq E, O \text{ an open Baire set}\}$.

c. Show that the measure μ of Theorem 5 can be chosen so that it is regular, and that it is then unique. [See Problem 13.12.]

d. Let X be an uncountable set with the discrete topology. Construct a nonregular Baire measure on X .

e. Let μ be a Baire measure on X . Then there is a unique regular Baire measure on X which agrees with μ on the σ -bounded Baire sets.

11. Let μ 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 μ .

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 σ -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 σ -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 σ -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 σ -compact (see Problem 5e).

3 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 2 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 $|f| \leq 1$, then

$$|F(f)| \leq F(|f|) \leq F(1).$$

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. If L is a vector lattice of bounded real-valued functions on a set X , 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

$$|F(f)| \leq M\|f\|,$$

and we define as usual

$$\|F\| = \sup_{\|f\| \leq 1} |F(f)|.$$

7. 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

$$F_+(f) = \sup_{0 \leq \varphi \leq f} F(\varphi).$$

Then $F_+(f) \geq 0$, and $F_+(f) \geq F(f)$. Moreover, $F_+(cf) = cF_+(f)$ for $c \geq 0$. Let f and g be two nonnegative functions in L . If $0 \leq \varphi \leq f$ and $0 \leq \psi \leq g$, then $0 \leq \varphi + \psi \leq f + g$, and so

$$F_+(f + g) \geq F(\varphi + \psi) \geq F(\varphi) + F(\psi).$$

Taking suprema over all such φ and ψ , we obtain

$$F_+(f + g) \geq F_+(f) + F_+(g).$$

On the other hand, if $0 \leq \psi \leq f + g$, then $0 \leq \psi \wedge f \leq f$ and $0 \leq \psi - (\psi \wedge f) \leq g$, and so

$$\begin{aligned} F(\psi) &= F(\psi \wedge f) + F(\psi - [\psi \wedge f]) \\ &\leq F_+(f) + F_+(g). \end{aligned}$$

Taking the supremum over all such ψ , we get

$$F_+(f + g) \leq F_+(f) + F_+(g),$$

and so

$$F_+(f + g) = F_+(f) + F_+(g).$$

Let f be an arbitrary function in L , and let M and N be two non-negative 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 now defined on all of L , and we have easily $F_+(f + g) = F_+(f) + F_+(g)$ and $F_+(cf) = cF_+(f)$ for $c \geq 0$. Since $F_+(-f) + F_+(f) = F_+(0) = 0$, we have $F_+(-f) = -F_+(f)$, and F_+ is a linear functional on L .

Since $0 \leq F_+(f)$ and $F(f) \leq F_+(f)$ for $f \geq 0$, 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 φ 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 φ , we have

$$\begin{aligned} \|F\| &\geq 2F_+(1) - F(1) \\ &= F_+(1) + F_-(1). \end{aligned}$$

Hence $\|F\| = F_+(1) + F_-(1)$. ■

8. 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 ν 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 7. Then by Theorem 5 there are finite Baire measures μ_1 and μ_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 ν 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

$$\begin{aligned} |\nu|(X) &\leq \mu_1(X) + \mu_2(X) \\ &= F_+(1) + F_-(1) = \|F\|. \end{aligned}$$

Thus $\|F\| = |\nu|(X)$.

To show the uniqueness of ν , we note that if ν_1 and ν_2 were both finite signed Baire measures such that

$$\int f d\nu_i = F(f)$$

for $i = 1, 2$ and $f \in C(X)$, then $\lambda = \nu_1 - \nu_2$ would be a finite signed Baire measure such that

$$\int f d\lambda = 0$$

for all $f \in C(X)$. Let $\lambda = \lambda^+ - \lambda^-$ be the Jordan decomposition of λ . Then integration with respect to λ^+ gives the same positive linear functional on $C(X)$ as that given by λ^- , and so by Theorem 5 we must have $\lambda^+ = \lambda^-$. Hence $\lambda = 0$, and $\nu_1 = \nu_2$. ■

9. 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

12. Let L and F be as in Proposition 7. 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.]

13. Let X be a compact Hausdorff space $\mathfrak{F} = \{f_\alpha\}$ a family of continuous real-valued functions on X and $\{c_\alpha\}$ a corresponding family of constants. Suppose that for each finite set $\{f_{\alpha_1}, \dots, f_{\alpha_n}\}$ there is a signed Baire measure ν with $|\nu|(X) \leq 1$ such that

$$\int f_{\alpha_i} d\nu = c_{\alpha_i}.$$

Then there is a finite signed Baire measure ν with $|\nu|(X) \leq 1$ such that for every f_α ,

$$\int f_\alpha d\nu = c_\alpha.$$

14. a. Let X be a compact Hausdorff space and g, f_1, \dots, f_n continuous real-valued functions on X . Suppose that there is a signed Baire measure ν on X with $|\nu|(X) \leq 1$ such that for each i we have $\int f_i d\nu = c_i$. Then there is a signed Baire measure μ on X with $|\mu|(X) \leq 1$ such that

$$\int f_i d\mu = c_i$$

and

$$\int g d\mu \leq \int g d\lambda$$

for any signed Baire measure λ with $|\lambda|(X) \leq 1$ and such that $\int f_i d\lambda = c_i$.

b. Suppose that there is a Baire measure ν on X with $\nu(X) = 1$ and $\int f_i d\nu = c_i$. Then there is a Baire measure μ on X with $\mu(X) = 1$ and $\int f_i d\mu = c_i$ which minimizes $\int g d\mu$ among all Baire measures which satisfy these conditions.

c. Let G, F_1, \dots, F_n be continuous functions on \mathbf{R}^m (= Euclidean m -dimensional space), and let f_1, \dots, f_m be continuous functions on X . Show that if there is a Baire measure ν with $\nu(X) = 1$ such that

$$F_i \left(\int f_1 d\nu, \dots, \int f_m d\nu \right) = c_i,$$

then there is a Baire measure on X which minimizes

$$G \left(\int f_1 d\nu, \dots, \int f_m d\nu \right)$$

under these restrictions.

15. Let B be the Banach space of signed Baire measures on a compact Hausdorff space X . What are the extreme points of the unit sphere of B ?

16. *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 9.27. This proof is due to deBranges.

Let \mathcal{A} be an algebra of real-valued continuous functions on a compact space X which separates points and contains the constants. Let \mathcal{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 $f \in \mathcal{A}$.

a. Use the Hahn-Banach Theorem and Corollary 9 to show that if \mathcal{A}^\perp contains only the zero measure, then $\overline{\mathcal{A}} = C(X)$.

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 \mathcal{A}^\perp , then \mathcal{A}^\perp contains only the zero measure.

c. Let μ be an extreme point of \mathcal{A}^\perp . If $f \in \mathcal{A}$, $0 \leq f \leq 1$, the measures μ_1 and μ_2 given by $d\mu_1 = f d\mu$ and $d\mu_2 = (1 - f) d\mu$ are in \mathcal{A}^\perp , with $\|\mu_1\| + \|\mu_2\| = \|\mu\|$, and $\mu_1 + \mu_2 = \mu$. Since μ is an extreme point, $\mu_1 = c\mu$ for some constant c .

d. Then $f - c \equiv 0$ on the support of μ (see Problem 11d).

e. Since f separates points, the support of μ can contain at most one point. Since $\int 1 d\mu = 0$, the support of μ is empty, and μ is the zero measure.

*4 The Borel Extension of a Measure

Previously we have considered only Baire measures on a locally compact Hausdorff space, but it is sometimes convenient to extend a given Baire measure to be a Borel measure. In this section we show that, if the Baire measure is regular, then this can be done in such a way that the analogue of Proposition 6 holds and so that the Borel measure really lives on the Baire sets in the sense that each Borel set of finite measure differs from a Baire set by a set of measure zero. We begin with a lemma.

10. Lemma: Let K be a compact \mathcal{G}_δ , and suppose $K \subset \bigcup O_i$ with O_i open. Then $K = \bigcup_{i=1}^n K_i$ where the K_i are compact \mathcal{G}_δ 's and $K_i \subset O_i$.

Proof: By Proposition 9.21 there are continuous nonnegative functions f_i with support in O_i such that $f_1 + \cdots + f_n = 1$ on K . Let $C_i = \{x: f_i(x) \geq \frac{1}{n}\}$. Then C_i is a compact \mathcal{G}_δ contained in O_i and $K \subset \bigcup C_i$. Set $K_i = C_i \cap K$. ■

11. Theorem: Let μ be a regular Baire measure on a locally compact Hausdorff space X . Then there is a unique Borel measure $\bar{\mu}$ on X which extends μ and for which:

i. For each open set O ,

$$\bar{\mu}O = \sup \{\mu K: K \subset O, K \text{ a compact } \mathcal{G}_\delta\}.$$

ii. For each Borel set E ,

$$\bar{\mu}E = \inf \{\bar{\mu}O: E \subset O, O \text{ an open set}\}.$$

The measure $\bar{\mu}$ also satisfies:

iii. For each Borel set E with $\bar{\mu}E < \infty$,

$$\bar{\mu}E = \sup \{\bar{\mu}K: K \subset E, K \text{ compact}\}.$$

iv. For each Borel set E with $\bar{\mu}E < \infty$, there is a Baire set H and a Borel set N with $\bar{\mu}N = 0$ and $E = H \Delta N$.

Proof: Such a measure $\bar{\mu}$ is clearly unique, since (i) specifies its value on open sets and that together with (ii) determines its value for all Borel sets.

Define a set function μ^* on open sets by setting

$$\mu^*O = \sup \{\mu K: K \subset O, K \text{ a compact } \mathcal{G}_\delta\}.$$

Since μ is regular, it follows that, if O is an open Baire set, then $\mu^*O = \mu O$. Let $O = \bigcup_{i=1}^\infty O_i$, and let K be a compact \mathcal{G}_δ contained in O . Then $K = \bigcup K_i$ with $K_i \subset O_i$ by Lemma 10, and so $\mu K \leq \sum \mu K_i \leq \sum \mu^* O_i$. Taking the supremum over K , we have $\mu^*O \leq \sum \mu^* O_i$, and μ^* is countably subadditive on open sets.

Let \mathfrak{M} be the collection of subsets M of X with the property that for each $\epsilon > 0$ there is an open set O such that $M \subset O$ and $\mu^*O < \epsilon$. Any subset of a set in \mathfrak{M} is again in \mathfrak{M} , and the countable subadditivity of μ^* implies that the union of a countable collection of sets in \mathfrak{M} is in \mathfrak{M} . If E is a Baire set belonging to \mathfrak{M} , then $\mu E = \sup \{\mu K : K \subset E, K \text{ a compact } \mathcal{G}_\delta\}$. But for each $K \subset E$ and each $\epsilon > 0$ there is an open set $O \supset K$ with $\mu^*O < \epsilon$. Thus $\mu K < \epsilon$ by the definition of μ^*O . Since ϵ was arbitrary, $\mu K = 0$. Consequently, $\mu E = 0$, and the hypotheses of Proposition 12.30 are satisfied. Thus there is an extension $\bar{\mu}$ of μ to a σ -algebra \mathfrak{B} containing \mathfrak{M} and the Baire sets such that $\bar{\mu}M = 0$ for $M \in \mathfrak{M}$.

Each $B \in \mathfrak{B}$ is of the form $H \triangle M$ with H a Baire set, $M \in \mathfrak{M}$, and $\bar{\mu}B = \mu H$. Since M is contained in an open set of arbitrarily small measure, it follows from Problem 10 b. that if $\bar{\mu}B < \infty$, then there is an open set O with $B \subset O$ and $\mu^*O < \infty$.

Let O be an open set with $\mu^*O < \infty$. If K is a compact \mathcal{G}_δ contained in O , then $\mu^*(O \sim K) = \mu^*O - \mu K$. Let $\langle K_n \rangle$ a sequence of compact \mathcal{G}_δ 's in O such that $\mu^*O = \lim \mu K_n$. Then $H = \bigcup K_n$ is a Baire set contained in O , and, since $O \sim H \subset O \sim K_n$ and $\mu^*(O \sim K_n) = \mu^*O - \mu K_n \rightarrow 0$, we have $O \sim H \in \mathfrak{M}$. Thus $\bar{\mu}O = \mu H = \mu^*O$, and $O \in \mathfrak{B}$.

Let \mathfrak{A} be the σ -algebra consisting of all sets which are locally measurable with respect to $\bar{\mu}$, that is, of all sets E such that $E \cap B \in \mathfrak{B}$ for each $B \in \mathfrak{B}$ with $\bar{\mu}B < \infty$. Since each subset of \mathfrak{B} of finite measure is contained in an open set of finite measure and each open set of finite measure is in \mathfrak{B} , it follows that E is in \mathfrak{A} if and only if the intersection of E with each open set of finite measure is in \mathfrak{B} . Consequently, \mathfrak{A} contains all open sets and hence all Borel sets. Extend $\bar{\mu}$ to \mathfrak{A} by setting $\bar{\mu}E = \infty$ if $E \in \mathfrak{A} \sim \mathfrak{B}$.

If we restrict $\bar{\mu}$ to the Borel sets, we have a Borel measure which is an extension of μ and satisfies properties (i) and (iv) of the theorem. If E is a Borel set not in \mathfrak{B} , then every open set O containing E has $\bar{\mu}O = \infty$ and $\bar{\mu}E = \infty$. Thus (ii) holds for such E . Let E be a set in \mathfrak{B} . Then $E = H \triangle M$, where H is a Baire set and $M \in \mathfrak{M}$. Since H is contained in an open Baire set O_1 with $\mu O_1 \leq \mu H + \epsilon$ and M in an open set O_2 with $\mu O_2 < \epsilon$, we have $E \subset O_1 \cup O_2$ with $\mu(O_1 \cup O_2) \leq \mu E + 2\epsilon$. This proves (ii).

To prove (iii), let $\bar{\mu}E < \infty$. Then there is an open set $O \supset E$ with $\bar{\mu}(O \sim E) < \frac{\epsilon}{3}$. By (i) there is a compact set $C \subset O$ with

$\bar{\mu}(O \sim C) < \frac{\epsilon}{2}$ and by (ii) an open set $U \supset (O \sim E)$ with $\bar{\mu}U < \frac{\epsilon}{2}$.

Set $K = C \cap \tilde{U}$. Then $K \subset E$, and $\bar{\mu}K > \bar{\mu}C - \frac{\epsilon}{2} > \bar{\mu}O - \epsilon \geq \bar{\mu}E - \epsilon$. ■

Problem

17. a. Let f be a Borel measurable function on a locally compact Hausdorff space X and μ a σ -finite Borel measure on X satisfying conditions (i) and (ii) of Theorem 11. Then there is a Baire measurable function g such that $f = g$ a.e. [μ].

b. What if μ is not σ -finite?

c. If $f \geq 0$, can we always choose g with $0 \leq g \leq f$?

15 Mappings of Measure Spaces

1 Point Mappings and Set Mappings

Let X and Y be any two spaces and φ a mapping of X into Y . Associated with φ are several mappings of objects associated with Y into corresponding objects associated with X . For example, the set mapping Φ 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 φ . We refer to φ as a point mapping. If $\langle X, \mathfrak{A} \rangle$ and $\langle Y, \mathfrak{B} \rangle$ are measurable spaces, the point mapping φ of X into Y is called **measurable** if $\varphi^{-1}[E] \in \mathfrak{A}$ for each $E \in \mathfrak{B}$. Thus φ is measurable iff Φ 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 σ -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 φ is the mapping φ^* 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 φ^* is often called the adjoint of φ , and it preserves sums, products, maxima, etc. If φ is measurable, then φ^* takes measurable functions into measurable functions.

Let \mathfrak{A} be an algebra of subsets of X and \mathfrak{B} an algebra of subsets of Y . Then a mapping Φ of \mathfrak{B} into \mathfrak{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 \mathfrak{A} and \mathfrak{B} are σ -algebras and Φ has the property that $\Phi\left(\bigcup_{i=1}^{\infty} E_i\right) = \bigcup_{i=1}^{\infty} \Phi(E_i)$, then Φ is called a **σ -homomorphism**. Every set mapping induced by a point mapping of X into Y is a σ -homomorphism, but we can have σ -homomorphisms which are not induced by any point mapping (Problem 2).

Let $\langle X, \mathfrak{A} \rangle$ and $\langle Y, \mathfrak{B} \rangle$ be measurable spaces and Φ a σ -homomorphism of \mathfrak{B} into \mathfrak{A} . Then Φ induces a mapping Φ^* of measures on $\langle X, \mathfrak{A} \rangle$ into measures on $\langle Y, \mathfrak{B} \rangle$ 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 φ be a measurable point mapping of the measure space $\langle X, \mathfrak{A}, \mu \rangle$ into the measurable space $\langle Y, \mathfrak{B} \rangle$. Let Φ be the induced set mapping of \mathfrak{B} into \mathfrak{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 \mathfrak{B} be a family of subsets of Y which contains Y, \emptyset , and each set $\{y\}$ consisting of a single element. Then each mapping Φ of \mathfrak{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 \mathfrak{A} be the collection of all subsets of $[0, 1]$ which are either countable or the complements of countable sets. Then \mathfrak{A} is a σ -algebra. Let $\mathfrak{B} = \{Y, \emptyset\}$. For $E \in \mathfrak{A}$, define $\Phi(E) = \emptyset$ if E is countable, $\Phi(E) = Y$ if E is the complement of a countable set. Then Φ is a σ -homomorphism and it is not induced by any point mapping of Y into X .

3. Show that the adjoint mapping φ^* 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 $\varphi: X \rightarrow Y$ and $\psi: Y \rightarrow Z$ measurable mappings. Show that $\psi \circ \varphi$ is a measurable mapping of $\langle X, \mathfrak{A} \rangle$ into $\langle Z, \mathfrak{C} \rangle$. 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 φ be a measurable point mapping of $\langle X, \mathfrak{A}, \mu \rangle$ into $\langle Y, \mathfrak{B}, \nu \rangle$ and Φ the induced set mapping of \mathfrak{B} into \mathfrak{A} . Suppose $\Phi^*\mu$ is absolutely continuous with respect to ν and that ν is a finite (or σ -finite) measure.

Define $\left[\frac{d\mu}{d\nu} \right]$ to be the Radon-Nikodym derivative of $\Phi^*\mu$ with respect to ν . 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 Φ a σ -homomorphism from \mathfrak{B} to \mathfrak{A} . Show that there is a unique linear mapping T_Φ 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 χ_E we have

$$T_\Phi(\chi_E) = \chi_{\Phi(E)}.$$

b. Let μ 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 Measure 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 σ -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 σ -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 σ -algebra is given by a σ -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 σ -algebra. We denote it by $\mathfrak{A}/\mathfrak{N}$. 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 $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{N}$. A subset of a Boolean σ -algebra \mathfrak{A} with these properties is called a **σ -ideal**, and we may define the Boolean σ -algebra $\mathfrak{A}/\mathfrak{N}$ of equivalence classes of \mathfrak{A} mod \mathfrak{N} .

By a **measure algebra**, we mean a Boolean σ -algebra \mathfrak{A} together with a nonnegative real-valued function μ 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 μ 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 μ on the Boolean σ -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 $A \triangle B$ of two elements by $A \triangle B = (A \wedge B') \vee (A' \wedge B)$. A measure algebra becomes a metric space if we define $\rho(A, B) = \mu(A \triangle B)$. This metric space is always complete, and the mappings $A \rightarrow A'$, $\langle A, B \rangle \rightarrow A \vee B$, and $\langle A, B \rangle \rightarrow A \wedge B$ are continuous. A measure algebra is called **separable** if it is separable as a metric space.

A mapping Φ 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 Φ 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.

2. Theorem (Carathéodory): *Let $\langle \mathfrak{A}, \mu \rangle$ be a separable measure algebra with $\mu(X) = 1$. Then there is an isomorphism Φ 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 Φ 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, \dots, A_n and their complements, and let $\mathfrak{A}_{\infty} = \bigcup_{n=1}^{\infty} \mathfrak{A}_n$. Then \mathfrak{A}_{∞} is again a Boolean algebra. For, given any A and B in \mathfrak{A}_{∞} , they belong to \mathfrak{A}_n and \mathfrak{A}_m , respectively. But if $m \leq n$, 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 Φ of α_∞ into the algebra \mathfrak{A} of all finite unions of half open subintervals of $[0, 1)$. The algebra α_1 consists of the four sets $0, A_1, A'_1, X$, and we have $\mu A_1 + \mu A'_1 = \mu X = 1$. Let $\Phi(A_1) = [0, \mu A_1], \Phi(A'_1) = [\mu A_1, 1), \Phi(0) = \emptyset, \Phi(X) = [0, 1)$. Then Φ preserves unions, intersections, complements, and measure. Suppose now that Φ has been defined on α_{n-1} so that it maps α_{n-1} onto the algebra generated by the half open intervals $[0, x_1), [x_1, x_2), \dots, [x_k, 1)$ and that Φ is measure preserving and preserves unions and complements. We wish to extend the mapping Φ to α_n . Let B_0, \dots, B_k be the sets in α_{n-1} which are mapped onto the intervals $[0, x_1), \dots, [x_k, 1)$. Then α_{n-1} consists of all finite unions of the sets B_0, \dots, B_k , and α_n consists of all finite unions of $A_n \wedge B_0, \dots, A_n \wedge B_k, A'_n \wedge B_0, \dots, A'_n \wedge B_k$. For those intersections which are 0, set Φ 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, Φ 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 Φ to all of α_n so that it preserves unions, complements, and measures.

Thus we have defined by induction the mapping Φ from α_∞ to $\mathfrak{M}/\mathfrak{N}$ so that it is measure preserving. Hence it is an isometry. Since α_∞ is dense in α and the metric space $\mathfrak{M}/\mathfrak{N}$ is complete, we can extend Φ to be an isometry from α into $\mathfrak{M}/\mathfrak{N}$.

To see that Φ preserves complements, let E be any element in α and choose $A \in \alpha_\infty$ so that $\mu(E \Delta A) < \epsilon$. Then $A' \in \alpha_\infty$, and $\mu(E' \Delta A') = \mu(E \Delta A) < \epsilon$. Since Φ is an isometry and $\Phi(A') = \sim \Phi(A)$, we have $m(\Phi(E') \Delta \tilde{\Phi}(A)) < \epsilon$ and $m(\tilde{\Phi}(E) \Delta \tilde{\Phi}(A)) = m(\Phi(E) \Delta \Phi(A)) < \epsilon$. Hence $m(\Phi(E') \Delta \tilde{\Phi}(E)) < 2\epsilon$ 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 Φ is an isometry, it is one-to-one into. To identify the range of Φ , let E be the set of endpoints of those intervals which were used in defining the mapping Φ on the algebras α_n . Then Φ maps α_∞ onto the algebra of finite unions of half-open intervals with endpoints in E . Suppose that \overline{E} is not all of $[0, 1]$, and let I be one of the open intervals of which $[0, 1] \sim \overline{E}$ is composed, that is, an interval contained in $\sim \overline{E}$ whose endpoints lie in \overline{E} . Since the endpoints of I lie in \overline{E} , I is a limit of intervals in $\Phi[\alpha_\infty]$, and so lies in $\Phi[\alpha]$. 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}_∞ , $\Phi(B)$ 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 Φ is one-to-one. Consequently, A is an atom.

We have thus proved that if \mathfrak{A} has no atoms, then $\overline{E} = [0, 1]$. But if $\overline{E} = [0, 1]$, 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 $\Phi[\mathfrak{A}] = \mathfrak{M}/\pi$. ■

This theorem states that, if $\langle X, \mathcal{B}, \mu \rangle$ 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 \mathcal{B} into the measurable sets of $[0, 1]$, but only of a correspondence of sets of \mathcal{B} modulo null sets with measurable sets modulo sets of measure zero. Thus if we have a set $B \in \mathcal{B}$, 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

8. Prove that in a Boolean σ -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 σ -algebra and \mathfrak{N} a Boolean σ -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 $(\bigvee A_n) \Delta (\bigvee B_n) \in \mathfrak{N}$.

10. 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 $\bigvee_{n=1}^{\infty} A_n = \lim_{k \rightarrow \infty} \bigvee_{n=1}^k A_n$. (Here \lim means limit in the metric space defined by μ .)

- b. Let $\langle \mathfrak{A}, \mu \rangle$ be a measure algebra and $\langle A_n \rangle$ any sequence of elements in \mathfrak{A} . Then $\bigvee_{n=1}^{\infty} A_n = \lim_{k \rightarrow \infty} \bigvee_{n=1}^k A_n$.

c. Show that if Φ is an isomorphism of a measure algebra $\langle \mathfrak{A}, \mu \rangle$ into a measure algebra $\langle \mathfrak{B}, \nu \rangle$, then $\Phi\left(\bigvee_{n=1}^{\infty} A_n\right) = \bigvee_{n=1}^{\infty} \Phi(A_n)$.

11. 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 $\bigwedge_{n=1}^{\infty} B_n = \lim B_n = \lim A_n$.]

12. Show that in a measure algebra the operations $', \wedge$, and \vee are continuous.

13. 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.

14. Discuss measure algebras in which we allow μ to be an extended real-valued function.

3 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 φ of X onto Y such that φ and φ^{-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 $B \in \mathfrak{B}$. In the present section we shall always assume that X and Y are metric spaces and that \mathfrak{A} and \mathfrak{B} are the algebras of Borel sets. A one-to-one mapping of X onto Y which takes Borel sets into Borel sets and whose inverse takes Borel sets into Borel sets will be called a **Borel equivalence**. Thus any homeomorphism is a Borel equivalence, but we may have Borel equivalences which are not homeomorphisms. If X is Borel equivalent to Y and Y Borel equivalent to Z , then X is Borel equivalent to Z . We begin by establishing some lemmas:

3. Lemma: *Let X be a metric space and E a Borel subset of X . Then the Borel subsets of E are those Borel subsets of X which are contained in E .*

Proof: Since E is a Borel subset of X , $O \cap E$ is a Borel subset of X for each open subset O of X . Hence the Borel subsets of X which

are contained in E form a σ -algebra which contains the (relatively) open subsets of E and so must contain every Borel subset of E . On the other hand, the collection of those subsets of X each of which is the union of a Borel subset of E and a Borel subset of $X \sim E$ is a σ -algebra which contains the open subsets of X . Hence it must contain all Borel subsets of X , and we see that a Borel subset of X which is contained in E must be a Borel subset of E . ■

4. Lemma: *Let X and Y be metric spaces, and let X_0 and Y_0 be countably infinite subsets of X and Y . Then each Borel equivalence between $X \sim X_0$ and $Y \sim Y_0$ can be extended to a Borel equivalence between X and Y .*

Proof: Since X_0 and Y_0 are countably infinite, we can extend the Borel equivalence between $X \sim X_0$ and $Y \sim Y_0$ to be a one-to-one correspondence between X and Y . Since any countable subset of a metric space is an F_σ , the union and difference of a Borel set and a countable set are again Borel sets. Thus a subset of X [or Y] is a Borel set if and only if its intersection with $X \sim X_0$ [or $Y \sim Y_0$] is a Borel set. Hence the extended correspondence between X and Y is a Borel equivalence. ■

If X is a topological space and A any set, we use X^A to denote the product space $\prod_{\alpha \in A} X_\alpha$ where each $X_\alpha = X$. The set X^A is the set of all mappings of A into X . If X is metric and A is countable, then X^A is metrizable (cf. Problem 8.27). We shall use ω to denote the set of integers. The following lemma is an immediate consequence of the fact that $\omega \times \omega$ is countable.

5. Lemma: *If X is a topological space, X^ω and $(X^\omega)^\omega$ are homeomorphic.*

One of the simplest topological spaces is the space whose elements are 0 and 1 and whose topology is discrete. We often denote this space by 2 . The product space 2^ω is a compact metric space.

6. Proposition: *The unit interval $[0, 1]$ is Borel equivalent to 2^ω .*

Proof: Let X_0 be that subset of 2^ω consisting of those elements which have only a finite number of coordinates equal to 0 and of

those which have only a finite number of coordinates equal to 1. Then X_0 is countably infinite. On $2^\omega \sim X_0$ define φ as that mapping which sends x into $\sum_{i=1}^{\infty} \xi_i 2^{-i}$, where ξ_i is the i^{th} coordinate of x . Then φ maps $2^\omega \sim X_0$ in a one-to-one manner onto those elements of $[0, 1]$ which are not of the form $p \cdot 2^{-n}$ with p and n integral, and φ is readily seen to be a homeomorphism between $2^\omega \sim X_0$ and these elements of $[0, 1]$. Since the set of elements of $[0, 1]$ of the form $p \cdot 2^{-n}$ is countably infinite, the proposition follows from Lemma 4. ■

7. Proposition: *Let I be the interval $[0, 1]$. Then I^ω and I are Borel equivalent.*

Proof: By Lemma 5 the spaces 2^ω and $(2^\omega)^\omega$ are homeomorphic. Thus by Proposition 6 we have I Borel equivalent to 2^ω , which is Borel equivalent to $(2^\omega)^\omega$, which in turn is Borel equivalent to I^ω . ■

8. Theorem: *Each complete separable metric space is Borel equivalent to a Borel subset of $[0, 1]$.*

Proof: Since $I = [0, 1]$ and I^ω are Borel equivalent, it suffices to show that each complete separable metric space $\langle X, \rho \rangle$ is homeomorphic to a Borel subset of I^ω . We may assume without loss of generality that ρ is bounded by one. Let $\langle r_n \rangle$ be a dense sequence of points, and let f be that mapping of X into I^ω which assigns to x the point whose i^{th} coordinate is $\rho(x, r_i)$. Then f is a one-to-one mapping of X onto a subset E of I^ω , for, if $x \neq y$, there is an r_i such that $\rho(x, r_i) < \rho(y, r_i)$. Since X and I^ω are metric, the mapping f will be continuous if $x_n \rightarrow x$ implies $f(x_n) \rightarrow f(x)$. But if $x_n \rightarrow x$, then $\rho(x_n, r_i) \rightarrow \rho(x, r_i)$ for each i , and so each coordinate of $f(x_n)$ converges to the corresponding coordinate of $f(x)$. Thus $f(x_n) \rightarrow f(x)$. (cf. Problem 8.26). The mapping f^{-1} will be a continuous mapping of E onto X if $f(x_n) \rightarrow f(x)$ implies $x_n \rightarrow x$. But if $f(x_n) \rightarrow f(x)$, then $\rho(x_n, r_i) \rightarrow \rho(x, r_i)$ for each i . Given $\epsilon > 0$, choose r_i so that $\rho(x, r_i) < \epsilon/3$ and choose N so that $|\rho(x_n, r_i) - \rho(x, r_i)| < \epsilon/3$ for $n \geq N$. Then, for $n \geq N$, we have $\rho(x_n, x) < 2\epsilon/3$, and so $\rho(x_n, x) < \epsilon$. Consequently, f is a homeomorphism of X onto E .

It remains only to show that E is a Borel set. But E is dense in $F = \overline{E}$. By Theorem 8.11 the completeness of X implies that E is a

\mathcal{G}_δ relative to F and hence a Borel set relative to F . Thus E is a Borel subset of I^ω . ■

Actually, a somewhat stronger statement is true: Each complete separable metric space is Borel equivalent to $[0, 1]$ (see Kuratowski [16], p. 227). We shall only make use of the weaker statement given in Theorem 8.

9. Theorem: *Let X be a complete separable metric space and μ a Borel measure on X such that $\mu X = 1$ and such that $\mu\{x\} = 0$ for each set $\{x\}$ consisting of a single point. Then there is a Borel set $X_0 \subset X$ with $\mu X_0 = 0$, and a subset Y_0 of $[0, 1]$ of Lebesgue measure zero, such that there is a Borel equivalence ψ of $X \sim X_0$ with $[0, 1] \sim Y_0$ having the property that $\mu(\psi^{-1}[A]) = mA$, where m is Lebesgue measure on $[0, 1]$.*

Proof: Let φ be a Borel equivalence of X with a Borel subset E of $[0, 1]$, and define the Borel measure ν on $[0, 1]$ by $\nu A = \mu(\varphi^{-1}[A])$. Let f be the cumulative distribution function of ν , that is, $f(x) = \nu[0, x]$. Then f is a monotone nondecreasing function on $[0, 1]$ which is continuous on the right. Since $f(x) - \lim_{t \rightarrow x^-} f(t) = \nu\{x\} = \mu\{\varphi^{-1}[x]\} = 0$, we see that f is continuous. Since $f(0) = 0$ and $f(1) = \nu[0, 1] = \mu X = 1$, f is a continuous mapping of $[0, 1]$ onto $[0, 1]$. For each $y \in [0, 1]$, the set $f^{-1}[\{y\}]$ is a closed set, and since f is nondecreasing it is either a closed interval or a single point. Let M be the set of y such that $f^{-1}[\{y\}]$ is a nondegenerate interval. Since these intervals are disjoint, M is countable. Now $mE = \nu(f^{-1}[E])$ by Problem 12.12, and so the set $N = f^{-1}[M]$ has ν -measure zero. Now f is a homeomorphism of $[0, 1] \sim N$ onto $[0, 1] \sim M$. Since N is a Borel set, the set $X_0 = \varphi^{-1}[N]$ is a Borel set, and $\mu X_0 = \nu N = 0$. Thus the mapping $\psi = f \circ \varphi$ is a Borel equivalence of $X \sim X_0$ with $\psi[X \sim X_0]$. Since φ is a Borel equivalence of X onto a Borel subset of $[0, 1]$, $\varphi[X \sim X_0]$ is a Borel subset of $[0, 1]$ contained in $[0, 1] \sim N$ and hence a Borel subset of $[0, 1] \sim N$. Thus $\psi[X \sim X_0] = f[\varphi[X \sim X_0]]$ is a Borel subset of $[0, 1] \sim M$ and hence of $[0, 1]$. Thus ψ is a Borel equivalence of $X \sim X_0$ with a Borel subset of $[0, 1]$. Since $mA = \mu(\psi^{-1}[A])$, we have $m(\psi[X \sim X_0]) = \mu(X \sim X_0) = 1$, and we see that the set $Y_0 = [0, 1] \sim \psi[X \sim X_0]$ is a Borel subset of Lebesgue measure zero. ■

Problems

15. Let X be a complete separable metric space and μ a Borel measure on X . Let E be a Borel subset of X with the property that if A is a Borel subset of E , then $\mu A = 0$ or $\mu(E \sim A) = 0$. Then there is a point x contained in E such that $\mu(E \sim \{x\}) = 0$. [Use Theorem 9.]

16. Let μ be a Borel measure on a space X . Then $\mu = \mu_0 + \mu_1$ where $\mu_1\{x\} = 0$ for each $x \in X$ and $\mu_0E = \sum_{x \in E} \mu\{x\}$.

4 Set Mappings and Point Mappings on Complete Metric Spaces

We say that $\langle X, \mathcal{A}, \mathfrak{N} \rangle$ is a measurable space with null sets if \mathcal{A} is a σ -algebra of subsets of X and \mathfrak{N} is a σ -ideal in \mathcal{A} . Each measurable point mapping φ of X into $[0, 1]$ induces a σ -homomorphism Φ of the Borel subsets of $[0, 1]$ into the σ -algebra \mathcal{A}/\mathfrak{N} by letting $\Phi(A)$ be the equivalence class containing $\varphi^{-1}[A]$. The following theorem states that, conversely, every σ -homomorphism of the Borel subsets of $[0, 1]$ into \mathcal{A}/\mathfrak{N} is induced in this way by a point mapping φ of X into $[0, 1]$. It is essentially Lemma 11.10 in another guise.

10. Theorem (Sikorski): *Let $\langle X, \mathcal{A}, \mathfrak{N} \rangle$ be a measurable space with null sets, and let Φ be a σ -homomorphism of the family \mathfrak{G} of Borel sets in $[0, 1]$ into the σ -algebra \mathcal{A}/\mathfrak{N} with $\Phi([0, 1]) = X$. Then there is a measurable mapping φ of X into $[0, 1]$ such that for each $B \in \mathfrak{G}$ we have $\varphi^{-1}[B]$ in the equivalence class $\Phi(B)$. If ψ is any other point mapping with this property, then $\psi = \varphi$ except on a subset in \mathfrak{N} .*

Proof: For each rational number α in $[0, 1]$ let A_α be a set in the equivalence class $\Phi([0, \alpha])$. We may take $A_1 = X$. If $\alpha < \beta$, then $\Phi([0, \alpha]) \leq \Phi([0, \beta])$, and so the set $E_{\alpha\beta} = A_\alpha \sim A_\beta$ is in \mathfrak{N} . Let $E = \bigcup_{\alpha < \beta} E_{\alpha\beta}$, where α and β run through the rationals. Since this is a countable union and \mathfrak{N} is a σ -ideal, we have $E \in \mathfrak{N}$. Set $B_\alpha = A_\alpha \cup E$. Then B_α is in the equivalence class $\Phi([0, \alpha])$, and $B_\alpha \subset B_\beta$ for $\alpha < \beta$.

For each $x \in X$, let $\varphi(x) = \inf \{\alpha: x \in B_\alpha\}$. Since $B_1 = X$, $\varphi(x)$ is defined for each x , and $0 \leq \varphi(x) \leq 1$. Thus φ is a mapping of X into $[0, 1]$. For each t we have $\{x: \varphi(x) \leq t\} = \bigcup_{\alpha \leq t} B_\alpha$.

Hence φ is measurable, and $\varphi^{-1}[[0, \alpha]] = B_\alpha$. If we let Ψ be the σ -homomorphism of \mathfrak{G} into $\mathfrak{A}/\mathfrak{N}$ induced by φ , then Ψ agrees with Φ on the closed intervals with rational endpoints. Since Ψ and Φ are σ -homomorphisms, the family of sets on which they agree is a σ -algebra and so must be the family \mathfrak{G} of Borel sets.

If ψ is any other mapping of X into $[0, 1]$ which also induces the σ -homomorphism Φ , then for each pair α and β of rational numbers the set $\{x: \varphi(x) \leq \alpha < \beta < \psi(x)\} = \varphi^{-1}[[0, \alpha]] \sim \psi^{-1}[[0, \beta]]$ and so must be in \mathfrak{N} . Thus we see that the set $\{x: \varphi(x) < \psi(x)\} = \bigcup_{\alpha, \beta} \{x: \varphi(x) \leq \alpha < \beta < \psi(x)\}$ is in \mathfrak{N} . Also $\{x: \varphi(x) > \psi(x)\} \in \mathfrak{N}$, and so $\{x: \varphi(x) \neq \psi(x)\} \in \mathfrak{N}$. ■

11. Theorem: Let $\langle X, \mathfrak{A}, \mathfrak{N} \rangle$ be a measurable space with null sets, Y a complete separable metric space, and Φ a σ -homomorphism of the Borel sets \mathfrak{G} of Y into $\mathfrak{A}/\mathfrak{N}$ with $\Phi(Y) = X$. Then there is a set $X_0 \in \mathfrak{N}$ and a point mapping φ of $X \sim X_0$ into Y such that for each $B \in \mathfrak{G}$ we have $\varphi^{-1}[B]$ in the equivalence class $\Phi(B)$.

Proof: By Theorem 8 there is a Borel equivalence χ of Y with a Borel subset E of $[0, 1]$. For each Borel subset B of $[0, 1]$ let $\Psi(B) = \Phi(\chi^{-1}[B \cap E])$. Then there is a mapping ψ of X into $[0, 1]$ which induces Ψ . Let $X_0 = \psi^{-1}[\tilde{E}]$. Since $\psi^{-1}[\tilde{E}] = \Phi(\emptyset)$, we have $X_0 \in \mathfrak{N}$. On $X \sim X_0$, the mapping $\chi^{-1} \circ \psi$ is defined and induces the σ -homomorphism Φ . ■

A σ -homomorphism Φ from a Boolean σ -algebra \mathfrak{A} to a Boolean σ -algebra \mathfrak{G} is called a **σ -isomorphism** if there is a σ -homomorphism Ψ from \mathfrak{G} to \mathfrak{A} such that $\Psi \circ \Phi$ and $\Phi \circ \Psi$ are the identity mappings on \mathfrak{A} and \mathfrak{G} . The following theorem is a generalization of a theorem of von Neumann, who assumed the unnecessary hypothesis that Φ be measure preserving with respect to suitable measures on \mathfrak{A} and \mathfrak{G} .

12. Theorem: Let X and Y be complete separable metric spaces, \mathfrak{A} and \mathfrak{G} their Borel sets, and let \mathfrak{M} and \mathfrak{N} be σ -ideals in \mathfrak{A} and \mathfrak{G} . If Φ is a σ -isomorphism of $\mathfrak{A}/\mathfrak{M}$ onto $\mathfrak{G}/\mathfrak{N}$ then there are sets $X_0 \in \mathfrak{N}$ and $Y_0 \in \mathfrak{M}$ and a one-to-one mapping φ of $Y \sim Y_0$ onto $X \sim X_0$ such that φ and φ^{-1} are measurable and $\Phi(A) = \varphi^{-1}[A]$ modulo \mathfrak{N} .

Proof: By Theorem 11 there is a set $Y_1 \in \mathfrak{M}$ and a mapping φ of $Y \sim Y_1$ into X such that $\Phi(A) = \varphi^{-1}[A]$ modulo \mathfrak{N} . Similarly,

there is a set $X_1 \in \mathfrak{M}$ and a mapping ψ of $X \sim X_1$ into Y such that $\Psi(B) = \psi^{-1}[B]$. Since $\Phi \circ \Psi$ is the identity on $\mathfrak{G}/\mathfrak{N}$, the mapping $\psi \circ \varphi$ can differ from the identity mapping on $Y \sim Y_1$ only on a subset $Y_2 \in \mathfrak{N}$. Let $Y_0 = Y_1 \cup Y_2$. Then φ is one-to-one on $Y \sim Y_0$, and $\varphi^{-1}[A] = \Phi(A)$ modulo \mathfrak{N} , $\varphi[B] = \Psi(B)$ modulo \mathfrak{M} . Thus $\varphi[Y \sim Y_0]$ differs from X by a set $X_0 \in \mathfrak{M}$. ■

Unfortunately, the exceptional sets X_0 and Y_0 in the preceding theorem may be unavoidable (see Problem 17), but they may be eliminated in special cases. One is given in Problem 18, another by the following theorem.

13. Theorem: *Let X be a complete separable metric space, \mathfrak{G} the family of Borel sets of X , and \mathfrak{N} a σ -ideal of \mathfrak{G} . If Φ is any σ -isomorphism of $\mathfrak{G}/\mathfrak{N}$ onto itself, then there is a one-to-one mapping φ of X onto itself such that φ and φ^{-1} are Borel measurable and $\Phi(A) = \varphi^{-1}[A]$ modulo \mathfrak{N} .*

Proof: By Theorem 12 there are sets X_0 and Y_0 in \mathfrak{N} and a one-to-one Borel measurable mapping ψ 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 ψ and ψ^{-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 ψ and ψ^{-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 φ by letting $\varphi(x) = \psi(x)$ if $x \in X \sim Z$ and $\varphi(x) = x$ if $x \in Z$. Then φ is a Borel equivalence of X onto itself which agrees with ψ except on the set $Z \in \mathfrak{N}$. ■

Problems

17. a. Show that the exceptional set X_0 in Theorem 11 may be unavoidable. [Let Y be a countable space with the discrete topology, X the unit interval with \mathfrak{N} the family of all sets which contain no rational numbers.]
- b. Show that the exceptional sets X_0 and Y_0 in Theorem 12 may be unavoidable.
18. Show that we can dispense with the exceptional sets X_0 and Y_0 in Theorem 12 if \mathfrak{M} and \mathfrak{N} are σ -ideals of $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ and each contains a set with as many elements as there are real numbers.

19. Let $X, Y, \mathfrak{A}, \mathfrak{B}, \mathfrak{M}, \mathfrak{N}$ be as in Theorem 12, Φ a σ -homomorphism of $\mathfrak{A}/\mathfrak{M}$ into $\mathfrak{B}/\mathfrak{N}$, and let φ be the mapping given by Theorem 11.

a. The mapping φ is one-to-one except on a set $Y_0 \in \mathfrak{N}$ if and only if there is a σ -homomorphism Ψ of $\mathfrak{B}/\mathfrak{N}$ into $\mathfrak{A}/\mathfrak{M}$ with $\Phi \circ \Psi$ the identity on $\mathfrak{B}/\mathfrak{N}$.

b. The mapping φ is onto all of X except for a set in \mathfrak{M} if and only if $\Phi(A) = 0$ implies $A = 0$.

5 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 .

14. Lemma: *Let ξ and η be real numbers. Then if $2 \leq p < \infty$,*

$$|\xi + \eta|^p + |\xi - \eta|^p \geq 2(|\xi|^p + |\eta|^p),$$

while if $0 < p \leq 2$,

$$|\xi + \eta|^p + |\xi - \eta|^p \leq 2(|\xi|^p + |\eta|^p).$$

If $p \neq 2$, equality can only hold if ξ or η is zero.

Proof: If $p = 2$, we have equality for all ξ and η . If $2 < p < \infty$, then $1 \leq p/2$, and applying the Hölder inequality with exponents $p/2$ and $p/(p - 2)$ to $\alpha^2 + \beta^2$ we obtain

$$\alpha^2 + \beta^2 \leq (\alpha^p + \beta^p)^{2/p}(1 + 1)^{(p-2)/p}$$

or

$$\alpha^p + \beta^p \geq 2^{(2-p)/2}(\alpha^2 + \beta^2)^{p/2}. \quad (1)$$

If $0 < p < 2$, replace p by $4/p$. Then (1) becomes

$$\alpha^{4/p} + \beta^{4/p} \geq 2^{(p-2)/p}(\alpha^2 + \beta^2)^{2/p}.$$

If we replace α by $\alpha^{p/2}$ and β by $\beta^{p/2}$, this becomes

$$\alpha^2 + \beta^2 \geq 2^{(p-2)/p}(\alpha^p + \beta^p)^{2/p},$$

or

$$\alpha^p + \beta^p \leq 2^{(2-p)/2}(\alpha^2 + \beta^2)^{p/2}. \quad (2)$$

Since

$$0 \leq \frac{\xi^2}{\xi^2 + \eta^2} \leq 1,$$

we have

$$\frac{|\xi|^p}{(\xi^2 + \eta^2)^{p/2}} \leq \frac{\xi^2}{\xi^2 + \eta^2} \quad 2 < p \quad (3)$$

and

$$\frac{|\xi|^p}{(\xi^2 + \eta^2)^{p/2}} \geq \frac{\xi^2}{\xi^2 + \eta^2} \quad p < 2. \quad (4)$$

Forming similar inequalities for η and adding, we get

$$|\xi|^p + |\eta|^p \leq (\xi^2 + \eta^2)^{p/2} \quad 2 < p \quad (5)$$

and

$$|\xi|^p + |\eta|^p \geq (\xi^2 + \eta^2)^{p/2} \quad p < 2. \quad (6)$$

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 α and β in (1) by $|\xi + \eta|$ and $|\xi - \eta|$. Then

$$\begin{aligned} |\xi + \eta|^p + |\xi - \eta|^p &\geq 2^{(2-p)/2}(|\xi + \eta|^2 + |\xi - \eta|^2)^{p/2} \\ &= 2(\xi^2 + \eta^2)^{p/2} \\ &\geq 2(|\xi|^p + |\eta|^p) \end{aligned}$$

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 $p \neq 2$ 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 14.

15. Lemma: Let $1 \leq p < \infty$, $p \neq 2$, 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.

16. Theorem (Lamperti): Let $1 \leq p < \infty$, $p \neq 2$, 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 φ 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 φ 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 support of a function f to be the set $\{t: f(t) \neq 0\}$. If $f \in L^p$, then the support of f is only defined modulo null sets. Thus for $f \in L^p$ the support of f is an element in the σ -algebra $\mathfrak{B}/\mathfrak{N}$ of Borel sets modulo sets of measure zero. We define a mapping Φ 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 15 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 15,

$$(U\chi_A) \cdot (U\chi_B) = 0 \text{ a.e.}$$

Thus Φ 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 support, we see that $\Phi(A \cup B) = \Phi(A) \cup \Phi(B)$ for disjoint sets and hence for any pair of Borel sets.

If $\Phi([0, 1]) = E$, we have $\Phi(\tilde{A}) = E \sim \Phi(A)$. Thus Φ is a homomorphism of \mathfrak{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 Φ is a σ -homomorphism.

By Theorem 10 there is a Borel-measurable mapping φ of E into $[0, 1]$ such that $\Phi(A) = \varphi^{-1}[A]$. Since Φ can only take sets of measure zero into sets of measure zero, the mapping φ must be onto all of $[0, 1]$ except possibly for a set of measure zero. Extend φ 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 support 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 support, and so $U\chi_A$ must equal h on the support of $U\chi_A$. Thus $U\chi_A = h \cdot \chi_{\Phi(A)} = h \cdot (\chi_A \circ \varphi)$. Hence, if ψ 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

20. 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 φ is essentially one-to-one (that is, one-to-one except on a set of measure zero), that φ^{-1} is measurable, and that φ and φ^{-1} take sets of measure zero into sets of measure zero. [Hint: Apply Theorem 16 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 μ by $\mu A = m(\varphi[A])$, then $|h|^p = \left[\frac{d\mu}{dm} \right]$.

21. What can you say about the isometries of $L^p(X, \mu)$ for general finite measure spaces? (Here you cannot get a point mapping but must be content with the set mapping Φ .)

22. Show that the characterization given in Theorem 16 is false for $L^2[0, 1]$.

23. 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 14 to show that L^p is uniformly convex if $p \geq 2$ and that we can take $\eta = (1 - \epsilon^p)^{\frac{1}{p}}$.

Epilogue

In the first part of the book we considered most of the central theory of functions of a real variable. This material falls roughly into three classes: sets of real numbers and their properties; measure and integration; and properties of real valued functions such as continuity and differentiability. The books of Saks [20] and Hobson [12] give an excellent account of this classical theory, particularly of those results which depend strongly on special properties of the real numbers. A very elegant presentation of this classical theory is given by Part I of Riesz-Nagy [18]. The books by Carathéodory [3] and de la Vallée Poussin [24] will give the reader some feeling for the earlier treatments of this subject. One topic from the classical theory of functions of a real variable which has been omitted is the theory of Fourier series, and a good reference here is Zygmund [25]. Another topic we have omitted is the classical theory of functions of several real variables. Although much of this theory is subsumed under the general theory of integration in Part Three, especially that dealing with Fubini's Theorem, there are many properties depending on the special structure of \mathbf{R}^n . These include much of the theory of partial derivatives, change of variables, and integration by parts. Some of this material is covered in Saks [20] and Edwards [6].

In Chapters 7, 8, and 9 we considered various aspects of general topology. For a more detailed treatment of this subject Kelley [14] is an excellent reference, and you will find there further references to the literature. In addition to the topics we have discussed briefly, there is a chapter on uniform spaces which generalize the metric spaces and the topology of linear spaces so that the notions of completeness and uniform continuity are meaningful. For deeper properties of metric spaces and Borel sets, Kuratowski [16] is good.

In Chapter 10 we discussed briefly some of the geometrical aspects of Banach spaces, and topological linear spaces were touched upon to the extent needed for the weak and weak* topologies in Banach spaces. The theorems (Hahn-Banach, closed graph, uniform boundedness, Alaoglu, Krein-Milman) are some of the most useful general tools at the disposal of the analyst. Useful theorems that we did not mention here are the fixed-point theorems of Schauder and Leray. A thorough catalogue of the “geometrical” properties of such spaces is given in Day [4], which also includes an excellent guide to the literature. The book by Banach [1] is very readable, although it suffers from the fact that general topology was not in existence at the time it was written and the weak topological notions are treated entirely in terms of sequential convergence. A thorough treatment of topological vector spaces can be found in Kelly-Namioka [15] or in Schaefer [21]. A central topic in modern analysis that we have not mentioned here is the theory of linear operators on a Banach or Hilbert space. Discussion of these from various points of view may be found in Dunford-Schwartz [5], Edwards [6], Hille-Phillips [11], Riesz-Nagy [18].

In Part Three we have given a reasonably comprehensive discussion of abstract measure and integration. The central theorems are the convergence theorems, the Radon-Nikodym theorem, the extension theorem for measures on an algebra, product measures and the Fubini-Tonelli theorems, the equivalence of abstract integration with that based on a measure, and the representation of bounded linear functionals on $C(X)$.

For a full discussion of measure theory based on σ -rings and of measure theory in locally compact spaces the reader should consult Halmos [8], which gives a very readable and thorough treatment of measure theory. For systematic treatment of the Daniell theory of integration in locally compact spaces, the book by Loomis [17] is a good reference.

An important branch of measure theory which we have not touched upon is the theory of Haar measure, which is the theory of measure invariant under a group of transformations on the measure space. A nice introduction is given in Banach's appendix to Saks [20], a full treatment in Halmos [8], and a treatment in terms of the Daniell integral in Loomis [17].

Another important aspect of measure theory is ergodic theory, which is the study of the properties of a measure preserving transformation and its iterates. The reader will find readable treatment in Hopf [12] and Halmos [10].

There are a number of applications of real analysis to the theory of functions of a complex variable, and the reader will find a nice treatment of many of them in Rudin [19].

Bibliography

- [1] S. BANACH, *Théorie des Opérations Linéaires* (Monografje Matematyczne, Vol. 1), Warsaw, 1932.
- [2] G. BIRKHOFF and S. MAC LANE, *A Survey of Modern Algebra* (3rd ed.), New York, Macmillan, 1965.
- [3] C. CARATHÉODORY, *Vorlesungen über reelle Funktionen*, Leipzig-Berlin, 1927.
- [4] M. M. DAY, *Normed Linear Spaces* (Ergebnisse der Mathematik und ihrer Grenzgebiete, No. 21), Berlin, Julius Springer, 1958.
- [5] N. DUNFORD and J. SCHWARTZ, *Linear Operators*, New York, Interscience, 1958.
- [6] R. E. EDWARDS, *Functional Analysis, Theory and Applications*, New York, Holt, Rinehart and Winston, 1965.
- [7] A. M. GLEASON, *Fundamentals of Abstract Analysis*, Reading, Mass., Addison-Wesley, 1966.
- [8] P. R. HALMOS, *Measure Theory*, New York, Van Nostrand, 1950.
- [9] P. R. HALMOS, *Naive Set Theory*, Princeton, Van Nostrand, 1960.
- [10] P. R. HALMOS, *Entropy in Ergodic Theory*, Chicago, Univ. of Chicago Press, 1959.
- [11] E. HILLE and R. PHILLIPS, *Functional Analysis and Semi-Groups* (American Mathematical Society Colloquium Publications, Vol. 31), Providence, 1957.
- [12] E. W. HOBSON, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series* (3rd ed.), Vol. 1, Cambridge, 1927.
- [13] E. HOPE, *Ergodentheorie* (Ergebnisse der Mathematik und ihrer Grenzgebiete, No. 5), Berlin, Julius Springer, 1937.
- [14] J. L. KELLEY, *General Topology*, New York, Van Nostrand, 1955.

- [15] J. L. KELLEY, I. NAMIOKA, et al., *Linear Topological Spaces*, Princeton, Van Nostrand, 1963.
- [16] C. KURATOWSKI, *Topologie I* (Monografje Matematyczne, Vol. 3), Warsaw, 1933.
- [17] L. H. LOOMIS, *An Introduction to Abstract Harmonic Analysis*, New York, Van Nostrand, 1953.
- [18] F. RIESZ and B. NAGY, *Functional Analysis* (English ed.), New York, Ungar, 1956.
- [19] W. RUDIN, *Real and Complex Analysis*, New York, McGraw-Hill, 1966.
- [20] S. SAKS, *Theory of the Integral* (Monografje Matematyczne, Vol. 7), Warsaw, 1937.
- [21] H. H. SCHAEFER, *Topological Vector Spaces*, New York, Macmillan, 1966.
- [22] P. C. SUPPES, *Introduction to Logic*, Princeton, Van Nostrand, 1957.
- [23] P. C. SUPPES, *Axiomatic Set Theory*, Princeton, Van Nostrand, 1960.
- [24] C. J. DE LA VALLÉE POUSSIN, *Intégrales de Lebesgue, Fonctions d'ensemble, Classe de Baire*, Paris, 1916.
- [25] A. ZYGMUND, *Trigonometrical Series* (Monografje Matematyczne, Vol. 5), Warsaw, 1935.

Index of Symbols

$A \ \& \ B$	A and B	<i>Page</i> 2
$A \vee B$	A or B	2
$\neg A$	not A	2
$A \Rightarrow B$	if A , then B	2
$A \Leftrightarrow B$	A if and only if B	2
(x)	for all x	2
$(\exists x)$	there is an x	2
■	Q.E.D.	4
\mathbb{N}	set of natural numbers	6, 33
$x \in A$	x is an element of A	6
$A \subset B$	A is a subset of B	6
$\{x \in X : P(x)\}$	set of x in X with $P(x)$	6
\emptyset	empty set	7
$\{x\}$	singleton	7
$\{x, y\}$	unordered pair	7
$\langle x, y \rangle$	ordered pair	7
$[a, b]$	closed interval	38
(a, b)	open interval	38
$X \times Y$	direct product of X and Y	7, 128, 150, 264
$\prod_{\lambda} X_{\lambda}$	direct product	18, 150
X^A	direct product	150
$g \circ f$	composition	9
$f A$	restriction of f to A	9

$\langle x_i \rangle_{i=1}^n$	finite sequence	<i>Page</i> 9
$\langle x_i \rangle_{i=1}^\infty$	infinite sequence	10
$\wp(X)$	set of subsets of P	11
$A \cap B$	intersection	11
$A \cup B$	union	11
\tilde{A}	complement of A	12
$A \triangle B$	symmetric difference	12
$\bigcap_{\substack{A \in \mathcal{C} \\ A: A \in \mathcal{C}}} \{A\}$	intersection of A in \mathcal{C}	13
\mathbf{R}	set of real numbers	29
$x \vee y$	$\max(x, y)$	32, 47, 237
$x \wedge y$	$\min(x, y)$	32, 47, 237
$\overline{\lim}$	limit superior	36
$\mathfrak{F}_\sigma \cup \mathfrak{G}_\delta$	special Borel sets	50
$l(I)$	length of I	52
\mathfrak{M}	class of Lebesgue measurable sets	60
χ_A	characteristic function of A	68
$C[0, 1]$	space of continuous functions on $[0, 1]$	118, 301
\mathbf{R}^n	Euclidean space	127
$S_{x, \delta}$	spheroid centered at x	129
\bar{E}	closure of E	41, 129
$\mu \perp \nu$	mutually singular measures	236
a.e. $[\mu]$	except for a set E , with $\mu E = 0$	238
$\nu \ll \mu$	ν absolutely continuous with respect to μ	238
$\left[\frac{d\nu}{d\mu} \right]$	Radon-Nikodym derivative	240
\mathfrak{Q}_σ	countable union from \mathfrak{Q}	255
$\mu \times \nu$	product measure	265

Subject Index

A

Absolute value of a measure, 236, 242 (36)
Absolutely continuous function, 104
measures, 238
Absolutely summable, 116, 118
Accumulation point, 43 (27)
Algebra
of functions, 171
of sets, 6
Almost everywhere, 67, 238
Annihilator, 193 (23)
Antisymmetric relation, 23
Archimedes, Axiom of, 33
Ascoli Theorem, 179
Atom, 321

B

Baire Category Theorem, 139
Baire measure, 302
regular, 307
Baire sets, 301
Banach limit, 192 (20)

Banach space, 116, 182
Base for a topology, 145
Bijective, 9
Binary expansion, 38 (21)
Bolzano-Weierstrass property, 160
Boolean algebra, 16, 319
Boolean- σ -algebra, 320
Borel equivalence, 324
Borel extension of a measure, 314
Borel field, 17
Borel measurability, 70 (26)
Borel measure, 261, 302
Borel set, 50 (302)
Bounded Convergence Theorem, 81
Bounded linear functional, 119
Bounded operator, 184
Bounded set, 302
Bounded variation, 99
Boundedness, total, 137 (22b), 164

C

Cantor ternary function, 48 (46), 70 (28), 107 (11)
Cantor ternary set, 44 (36), 63 (14)
Carathéodory extension theorem, 257

Note: Numbers in parentheses indicate a problem number.

Carathéodory outer measure, 283
 Category, 139
 Cauchy criterion, 35
 Cauchy sequence, 35, 115, 134, 136
 in measure, 93 (25)
 Change of variable, 107 (13)
 Characteristic function, 68, 75
 Choice
 Axiom of, 18
 function, 18
 Closed Graph Theorem, 196
 Closed set, 41, 129, 143
 Closure, point of, 40, 129, 145
 Cluster point, 35, 134, 143
 of a net, 155
 Collection, 5
 Compactification
 Alexandroff, 168
 Stone-Čech, 170
 Compactness, 157
 countable, 159
 local, 168
 metric, 163
 sequential, 160
 Complement, 12
 Complete measure, 220
 Complete orthonormal system, 212
 Complete space, 31, 115, 134
 Completely regular space, 148
 Complex measures, 241 (36)
 Composition, 9
 Conjugate space, 190
 Connected space, 152
 Continuity, absolute, 104, 238
 Continuous functions, 44, 132, 144
 Contraposition, 3
 Convergence, 46, 133, 143
 in mean, 115
 in measure, 91, 230 (19)
 pointwise, 46, 72, 115, 151, 179
 (37)
 setwise, 231
 uniform, 46, 162

Convergence Theorem
 Bounded, 81
 Lebesgue, 88, 229
 Monotone, 84, 227
 Convex function, 108
 Convex hull, 207
 Convex set, 203
 Correspondence, one-to-one, 19
 Countable compactness, 159
 Countable set, 9, 19
 Countability axioms, 148
 Counting measure, 53 (4)
 Cover, measurable, 280 (29)
 Covering, 42
 Cube, 151, 167 (13)
 Cumulative distribution function, 261

D

Daniell
 functional, 287
 integral, 287
 Decimal expansion, 38 (21)
 Decomposable measure, 243 (39),
 249 (45)
 Dense subset, 44 (30)
 Derivates, 96
 Dimension, 213, 214 (52d)
 Dini Theorem, 162
 Direct product, 7, 18, 128, 150, 264
 Directed system, 155
 Discrete topology, 143
 Disjoint, 13
 Domain, 8
 Dual Space, 190

E

Egoroff's Theorem, 72 (30)
 Empty set, 7
 Envelopes of a function, 49 (49)

Equicontinuous, 177

Equivalence

- of metrics, 132

- of norms, 182 (2), 195

Equivalence, uniform, 136

Equivalence relation, 22

Extended real numbers, 34

Extended real-valued function, 24

.Extreme point, 206

F

Family, 5

Fatou's Lemma, 83, 226, 231

Field, 30

Finite intersection property, 158

Finite measure, 221

Finite set, 9, 19

First axiom of countability, 146

First category, 139

First element, 23

Fubini Theorem, 269

Function, 8

- Cantor ternary, 70 (8)

- characteristic, 68, 75

- choice, 18

- continuous, 44, 132, 144

- absolutely continuous, 104

- semicontinuous, 48 (48), 161

- convex, 108

- measurable, 65, 223, 317

- set, 52, 217

- simple, 69, 75, 225

Functionals, bounded linear, 119

G

Generalized Cantor set, 63 (146)

Graph, 8, 196

H

Hahn-Banach Theorem, 187

Hahn decomposition, 236

Half-open interval topology, 147 (12)

Hausdorff Maximal Principle, 24

Hausdorff space, 148

Heine-Borel Theorem, 42, 157

Hilbert cube, 151

Hilbert space, 210

Holder Inequality, 113

Homeomorphism, 47 (44), 132, 144

- uniform, 136

Homomorphism, lattice, 318

I

Iff, 2

Induction, 6

Inf, 31

Interior, 44 (33), 131 (5)

Internal point, 204

Intersection, 11

Interval, 38

Intrinsic property, 138

Inverse image, 9

Isolated set, 44 (29)

Isometry, 132, 333

Isomorphism, 184, 199, 221

J

Jensen Inequality, 110

Jordan decomposition, 236

K

Kernel, measurable, 280 (29)

Kernel (of an operator), 186 (14)

Krein-Milman Theorem, 207

L

L^p Spaces, 111, 243
Lamperti Theorem, 333
Lattice, 171, 286, 309
 homomorphism, 318
Least element, 23
Lebesgue Convergence Theorem, 88, 229
Lebesgue decomposition, 240
Lebesgue integral, 79
Lebesgue measure, 60
Lebesgue-Stieltjes Integral, 261
Lebesgue's Theorem, 82 (2b)
Length, 52
Limit, 35, 115, 133, 143
 Banach, 192 (20)
 of a net, 155
 superior, 36, 48 (47)
Lindelof Theorem, 130
Linear functional, 119, 186, 246, 304, 310
Linear operator, 184
Linear ordering, 23
Linear space, 111, 181
Linear transformation, 184
Local base, 198
Locally compact spaces, 168
Locally convex space, 205
Locally measurable, 221
Lusin's Theorem, 72 (31)

M

Manifold, 169–170 (24)
Manifold, linear, 182
Measurable
 cover, 280 (29)
 function, 65, 223, 295
 kernel, 280 (29)
 mapping, 317
 set, 56, 251, 296
 space, 217

Measure, 53, 217

on an algebra, 253
Baire, 302
Borel, 261, 302
 complex, 241 (36)
 counting, 53
 inner, 274
Lebesgue, 60
 outer, 54, 250
 Carathéodory, 283

product, 296
 signed, 233

Measure algebra, 320
 separable, 321

Metric, 127

Metric space, 127

Metrizable space, 143

Minimal element, 23

Minkowski Inequality, 114

Monotone

 function, 47 (44), 96
 sequence, 10, 35

Monotone Convergence Theorem, 84, 227

Monotonicity, 53 (1)

N

n-tuple, 9

Natural number, 6

Negative variation, 99

Net, 155

Norm, 111, 181, 184

Normal space, 148

Nowhere dense, 139

Null function, 294

Null set, 233

O

One-to-one, 9

 correspondence, 19

Onto, 8, 19, 131

Open mapping, 193
 theorem, 195
 Open set, 38, 129, 142
 Operator, linear, 184
 Ordered field, 30
 Ordered pair, 7
 Ordering, 23
 Ordinal, 25
 Orthogonal, 210
 Orthonormal system, 211
 Outer measure, 54, 250
 Carathéodory, 283
 regular, 260 (8)

P

Partial ordering, 23
 Point of closure, 40, 129, 145 (2)
 Pointwise convergence, 46, 72, 115,
 151, 179 (37)
 Polygonal function, 47 (45)
 Positive set, 233
 Positive variation, 99
 Product

direct, 7, 18, 128, 150, 264
 measure, 265
 topology, 150

Projection, 151
 Proper map, 169 (22)
 Pseudometric, 129 (3)
 Pseudonorm, 183 (10)

Q

Quotient, 22, 183 (11), 320

R

Radon-Nikodym derivative, 240,
 241 (33)

Radon-Nikodym Theorem, 238,
 242 (37)
 Range, 8
 Real numbers, 29
 Recursive definition, 9
Reductio ad absurdum, 3
 Reflexive partial order, 24
 Reflexive relation, 22
 Reflexive space, 191
 Regular Baire measure, 307
 Regular outer measure, 260 (8)
 Regular space, 148
 Relation, 22
 Relative properties, 138
 Residual set, 139
 Restriction
 of a function, 9
 of a measure, 223 (4)
 Riemann integral, 74
 Riemann-Lebesgue Theorem, 90 (16)
 Riesz-Fischer Theorem, 117
 Riesz Representation Theorem, 121,
 246, 310

S

σ -algebra, 17, 320
 σ -bounded set, 302
 σ -compact, 302
 σ -finite measure, 220
 σ -homomorphism, 318
 σ -ideal, 320
 σ -isomorphism, 329
 σ -ring, 222 (9)
 Saturated measure, 221
 Schwarz inequality, 210
 Second axiom of countability, 146
 Semialgebra, 259
 Semicontinuous function, 48 (48),
 161, 162 (4)
 Semifinite measure, 220
 Separable measure algebra, 321

Separable space, 130
 Separated, linearly, 204
 Separation, 152
 Sequence, 9
 Sequentially compact, 160
 Set, 5
 Set function, 52, 217
 Set mapping, 317
 Setwise convergent, 231
 Signed measure, 233
 Simple function, 69, 75, 225
 Singleton, 7
 Singular function, 107 (12)
 Singular measures, 238
 Sphere, 129 (2)
 Spheroid, 129 (2)
 Step function, 49 (48f), 74
 Stone representation theorem, 297
 Stone-Weierstrass Theorem, 174,
 313 (16)
 Strict partial order, 24
 Strong topology, 201
 Stronger topology, 144
 Subadditivity, 53 (2)
 Subsequence (10)
 Subspace, 137, 143, 182
 Summable sequence, 37 (17), 116
 Sup, 31
 Support (of a measure), 308 (11)
 Surjective, 8
 Symmetric difference, 12
 Symmetric relation, 22
 System, directed, 155

T

T_1, T_2, T_3, T_4 , 147
 Ternary
 expansion, 38 (21)
 function, 48 (46), 107 (11)
 set, 44 (36), 63 (14)
 Tietze's Extension Theorem, 148

Tonelli Theorem, 270
 Topological property, 132
 Topological space, 142
 Topological vector space, 197
 Topology, 142
 Topology of pointwise convergence,
 151
 Total boundedness, 137, 164
 Total variation
 of a function, 99
 of a measure, 236, 242 (36)
 Transitive relation, 22
 Translation invariance, 52, 63, 198
 Trivial topology, 143
 Tychonoff Theorem, 166, 199

U

Uncountable set, 21 (23)
 Uniform boundedness principle,
 140, 196
 Uniform continuity, 135
 Uniform convergence, 46
 Uniform homeomorphism, 136
 Uniform properties, 136, 198
 Uniformly convex, 334 (23)
 Uniformly equivalent, 136
 Union, 11
 Unordered pair, 7
 Upper and lower envelopes of a
 function, 49 (49)
 Upper bound, 31
 Ursohn's Lemma, 148

V

Variation, total, 99
 Vector lattice, 286
 Vector space, 111, 181
 topological, 197
 Vitali covering, 94, 95

Subject Index**W**

Weak topology, 148, 150 (20), 200,
201

Weak* topology, 201
Weaker topology, 144
Weakest topology, 147 (10)
Well ordering, 24