Peter A. Loeb

Real Analysis

Birkhäuser

Birkhäuser

Peter A. Loeb

Real Analysis

Birkhäuser

Peter A. Loeb
University of Illinois
Urbana, IL, USA

ISBN 978-3-319-30742-8 ISBN 978-3-319-30744-2 (eBook)
DOI 10.1007/978-3-319-30744-2

Library of Congress Control Number: 2016933667

Mathematics Subject Classification (2010): MSC 28-01

Printed on acid-free paper

This book is published under the trade name Birkhäuser
The registered company is Springer International Publishing AG Switzerland (www.birkhauser-science.com)

Dedicated to the memory of my teachers:
Lloyd B. Williams, Robert C. James, and
Halsey L. Royden

Preface

This book presents the work of many mathematicians. It is the product of the two-semester course I have taught at the University of Illinois since 1968, a span of over 1% of recorded western history. The development of the book's diverse topics has often been improved with input from current research, including my own and joint research with coauthors. In particular, the most difficult material of the first semester, appearing in Chapter 5, has been considerably simplified and shortened by application of work with potential theorist Jürgen Bliedtner.

As is common practice for a book on measure theory, the first five chapters are devoted to measures on the real numbers. This of course includes the generalization of the Riemann integral using Lebesgue measure. With essentially no more work for the student, however, it also includes more general measures on $\mathbb{R}$. The generalization is important for many applications, such as in statistics and probability theory. It is also helpful in developing the connection between integration and differentiation. That topic is considerably simplified in Chapter 5 with the use of a local maximal function and a simple, optimal covering theorem.

From there, the book continues with an introduction to general measure, metric, and normed spaces. This includes the Baire Category Theorem and classical L^p spaces. It is the material needed as a foundation for Chapter 8 on Hilbert spaces, Fourier series, and the proof of the important Radon-Nikodým Derivative Theorem.

Chapter 9 gives a parallel treatment of spaces with a distance function, i.e., metric spaces, and more general topological spaces. Open balls are used in a metric space to determine how close a point x is to a point z: The more balls centered at z that contain x, the closer x is to z. More general topological spaces replace open balls with sets not defined by a distance function. The chapter includes a development of the infinite product of such spaces and an elementary proof that if the spaces are compact, so is the product.

Chapter 10 is devoted to the construction of general measures including measures on product spaces. Chapter 11 develops properties of infinite-dimensional vector spaces with a norm, i.e., a function specifying the distance to $\mathbf{0}$. The chapter includes further information about classical L^p spaces and measures associated with linear maps on spaces of continuous functions.

The Axiom of Choice is assumed when needed throughout the book. It allows a point to be chosen from each set in an infinite collection of nonempty sets even when no rule can be given. The equivalence of several forms of the axiom is proved in the book's first appendix. The second appendix simplifies what is needed to show that a Radon-Nikodým derivative is the result of a limit process. That appendix also provides two powerful covering theorems for finite-dimensional spaces.

The third appendix introduces the reader to the rigorous use of infinitely large and infinitely small numbers in subjects such as calculus and measure theory. In calculus, for example, the treatment of the chain rule and applications of the integral can inform instruction in an ordinary calculus course. The measure spaces developed in the appendix have had a number of important applications over several decades. With the work of Yeneng Sun, for example, there is now a rigorous way of treating a continuum of independent random variables and traders in an economy.

Answers are provided at the end of the book for many of the problems; these are marked with an "A". By putting the period outside the quotes, I have just used the British rule allowing context to be considered for such punctuation. I have read that the American rule was set by typesetters to protect delicate type for commas and periods. Typesetters, however, have not always understood the needs of mathematicians. A famous example is the proof sheet returned to an author with a minute speck that magnification showed to be an epsilon. The text read "Let epsilon be as small as possible."

I am indebted to Erik Talvila for his help and advice in writing this book. I also thank Agus Soenjaya, Derek Jung, and Sepideh Rezvani for their helpful suggestions and careful reading of the manuscript. Finally, Birkhäuser–Springer editor Benjamin Levitt is due great thanks for his help and guidance.

Champaign-Urbana, IL, USA Peter A. Loeb
January 2016

Contents

1 Set Theory and Numbers . . . 1
- 1.1 Terms and Notation . . . 1
- 1.2 Indexed Families . . . 3
- 1.3 Algebras and σ-Algebras of Sets . . . 4
- 1.4 Orderings . . . 6
- 1.5 Sequences in $\mathbb{R}$ and $\overline{\mathbb{R}}$ . . . 7
- 1.6 Lim sup and Lim inf . . . 9
- 1.7 Open and Closed Sets . . . 10
- 1.8 Closures . . . 12
- 1.9 Directed Sets and Nets . . . 13
- 1.10 Compactness . . . 14
- 1.11 Continuous Functions . . . 16
- 1.12 Problems . . . 19

2 Measure on the Real Line . . . 25
- 2.1 Introduction . . . 25
- 2.2 Lebesgue Outer Measure . . . 26
- 2.3 General Outer Measures . . . 28
- 2.4 Measure from Outer Measure . . . 31
- 2.5 Approximation of Measurable Sets . . . 35
- 2.6 LimSup and LimInf of a Sequence of Sets . . . 37
- 2.7 The Existence of a Non-measurable Set . . . 38
- 2.8 Cantor Set . . . 39
- 2.9 Problems . . . 40

3 Measurable Functions . . . 45
- 3.1 Definition . . . 45
- 3.2 Limits and Special Functions . . . 48
- 3.3 Approximations and Theorems of Lusin and Egoroff . . . 51
- 3.4 Problems . . . 55

4 Integration . . . 57
4.1 The Riemann Integral . . . 57
4.2 The Integral of Simple Functions . . . 60
4.3 The Integral of Bounded Measurable Functions . . . 62
4.4 The Integral of Nonnegative Measurable Functions . . . 65
4.5 The Integral of Measurable Functions . . . 68
4.6 Generalization of Fatou's Lemma . . . 70
4.7 Improper Riemann Integral . . . 71
4.8 Convergence in Measure . . . 72
4.9 Problems . . . 73

5 Differentiation and Integration . . . 79
5.1 Introduction . . . 79
5.2 A Covering Theorem for $\mathbb{R}$ . . . 80
5.3 A Local Maximal Function . . . 81
5.4 Differentiation . . . 82
5.5 Functions of Bounded Variation . . . 84
5.6 Absolute Continuity . . . 86
5.7 Problems . . . 90

6 General Measure Spaces . . . 95
6.1 Introduction . . . 95
6.2 Integration . . . 96
6.3 Signed Measures . . . 102
6.4 Convexity and Jensen's Inequality . . . 105
6.5 Problems . . . 107

7 Introduction to Metric and Normed Spaces . . . 109
7.1 Metric Spaces . . . 109
7.2 Baire Category . . . 112
7.2.1 Application to Differentiable Functions . . . 114
7.3 Normed Spaces . . . 114
7.4 Classical Normed Spaces . . . 115
7.5 Linear Functionals . . . 119
7.6 Problems . . . 121

8 Hilbert Spaces . . . 127
8.1 Basic Definitions . . . 127
8.2 Basic Inequalities . . . 128
8.3 Convex Sets, Orthogonality, and Bounded Linear Functionals . . . 129
8.4 Radon-Nikodým Theorem . . . 132
8.5 Orthonormal Families and Fourier coefficients . . . 135
8.6 Separability . . . 138
8.7 Unordered Sums and ℓ^2 Spaces . . . 139
8.8 Fourier Series . . . 140

8.9 Trigonometric Series . . . 142
8.10 Problems . . . 143

9 Topological Spaces . . . 147
9.1 Neighborhoods . . . 147
9.2 Metric and Topological Notions . . . 150
9.3 Continuous Mappings . . . 152
9.4 Sequences and Nets . . . 152
9.5 Subspaces . . . 154
9.6 Connectedness . . . 155
9.7 Compact Spaces . . . 156
9.8 Properties of Topologies . . . 160
9.9 Product Spaces, Metrization, and Compactification . . . 163
9.10 Ascoli-Arzelá Theorem . . . 169
9.11 Stone-Weierstrass Theorem . . . 172
9.11.1 Vector Lattices . . . 172
9.11.2 Algebras . . . 173
9.12 Problems . . . 175

10 Measure Construction . . . 179
10.1 Measures from Outer Measures . . . 179
10.2 The Carathéodory Extension Theorem . . . 180
10.3 Lebesgue Measure on Euclidean Space . . . 183
10.4 Product Measures . . . 185
10.5 Other Integrals . . . 189
10.6 Problems . . . 189

11 Banach Spaces . . . 191
11.1 Banach Spaces . . . 191
11.2 Return to Classical Normed Spaces . . . 193
11.3 Dual Space of L^p . . . 196
11.4 Hahn-Banach Theorem . . . 199
11.5 Imbedding Into a Second Dual . . . 202
11.6 Properties of Banach Spaces . . . 202
11.7 Weak and Weak* Topologies . . . 205
11.8 Functionals on Continuous Functions . . . 206
11.9 Problems . . . 218

A Appendix on the Axiom of Choice . . . 221

B Appendix on Limit and Covering Theorems . . . 225
B.1 A General Limit Theorem . . . 225
B.2 Besicovitch and Morse Covering Theorems . . . 228
B.3 The Best Constant and Proof of Besicovitch's Theorem . . . 230
B.4 Proof of Morse Covering Theorem . . . 232

C Appendix on Infinitesimal Analysis and Measure Theory ... 235
C.1 Basic Nonstandard Analysis ... 235
C.2 A Simple Extension of the Real Numbers ... 236
C.3 The Transfer Principle ... 238
C.4 Using the Transfer Principle ... 240
C.5 Properties of ${}^*\mathbb{R}$ ... 241
C.6 The Nonstandard Natural Numbers and Hyperfinite Sets ... 243
C.7 Sequences ... 244
C.8 Open Sets in the Reals ... 245
C.9 Limits and Continuity ... 245
C.10 Differentiation ... 246
C.11 Riemann Integration ... 247
C.12 Metric and Topological Spaces ... 249
C.13 Measure and Probability Theory ... 251

D Answers ... 255

References ... 267

Index ... 271

Chapter 1
Set Theory and Numbers

1.1 Terms and Notation

Set notation should be familiar to the reader. Recall $x \in A$ means that x is an element of A; the negation is $x \notin A$. Notation for every member of a set A belonging to a set B is $A \subseteq B$ or $B \supseteq A$. We say that A is a **subset** of B or B is a **superset** of A. If A is a proper subset of B, that is, $A \subseteq B$ but $A \neq B$, then one can write $A \subset B$. To prove that two sets A and B are equal, it is necessary to show that each is contained in the other; that is, $A \subseteq B$ and $B \subseteq A$. Familiar sets are the empty set $\varnothing$, which contains no elements and is a subset of every set, the natural numbers $\mathbb{N} = \{1, 2, 3, \cdots\}$, the integers $\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$.

In our discussion, we will often use **quantifiers**. The universal quantifier $\forall$ is read as "for all" or "for any." The existential quantifier $\exists$ is read as "there exists."

An **interval** in $\mathbb{R}$ is a subset I of $\mathbb{R}$ such that if $c \in I$ and $d \in I$, and $c < d$, then all points $x \in \mathbb{R}$ with $c < x < d$ are in I. Here is the notation and definition for **open intervals** in $\mathbb{R}$: $(a,b) := \{x \in \mathbb{R} : a < x < b\}$; $(a, +\infty) := \{x \in \mathbb{R} : a < x\}$; and $(-\infty, b) := \{x \in \mathbb{R} : x < b\}$. The **closed intervals** $[a,b]$, $[a, +\infty)$, and $(-\infty, b]$ contain all real endpoints. The half-open intervals $(a,b]$ and $[a,b)$ have two real endpoints, but contain only one of them. A singleton set $\{a\}$ is a **degenerate interval**. We will be dealing with subsets of the real line or a more general set X. These can be quite different than intervals. Indeed, many sets contain no interval at all. Two examples are the rational numbers and the irrational numbers.

The basic operations on sets are **union** $A \cup B$, i.e., the collection of elements in either A or B, and **intersection** $A \cap B$, i.e., the collection of elements in both A and B. The **Cartesian product** of A and B is the collection of ordered pairs $A \times B := \{(a,b) : a \in A \text{ and } b \in B\}$. The **power set** of a set A is the collection of all subsets of A, written $\mathscr{P}(A)$. Note that for each set A we have $A \in \mathscr{P}(A)$ and $\varnothing \in \mathscr{P}(A)$. The **complement** of B with respect to A is $A \setminus B := \{x \in A : x \notin B\}$. If the set A is understood, then we write $\widetilde{B}$ or $\complement B$. For example, when working in the context of

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_1

the real numbers, $\widetilde{\mathbb{Q}} = \mathbb{R} \setminus \mathbb{Q}$ denotes the irrational numbers. A collection of sets is called **pairwise disjoint,** or just **disjoint**, if any two distinct members of the family have an empty intersection.

If A and B are subsets of a set X, then the **symmetric difference** of A and B, denoted by $A \,\Delta\, B$, is the set $(A \setminus B) \cup (B \setminus A)$. It is the set of points in the union of A and B but not in the intersection.

A **relation** is a subset of a Cartesian product $A \times B$. If $A = B$, the relation is said to be on A. An **equivalence relation** on A is a relation ρ with the **reflexive** property: $(x,x) \in \rho$ for each $x \in A$; the **symmetric** property: $(x,y) \in \rho \Rightarrow (y,x) \in \rho$; and the **transitive** property: $(x,y) \in \rho$ and $(y,z) \in \rho \Rightarrow (x,z) \in \rho$. For each $x \in A$ the **equivalence class** of x is written $[x]$; it is the set $\{y \in A : (x,y) \in \rho\}$. Note that $x \in [x]$ so none of the equivalence classes is empty. For each $x, y \in A$ either $[x] = [y]$ or $[x]$ and $[y]$ are disjoint. The set A is then partitioned into a union of disjoint equivalence classes. For example, the rational numbers $\mathbb{Q}$ is the collection of equivalence classes of ratios of integers with nonzero denominators.

A **function** $f : A \mapsto B$ is a relation in $A \times B$ such that for each $x \in A$ there is one and only one element $y \in B$ such that $(x,y) \in f$. That is, if $(x,y) \in f$ and $(x,z) \in f$, then $y = z$. One writes $y = f(x)$ and says y is the image of x under f. The set A is the set of points for which the function is defined; it is called the **domain** of f. The set B is the **target set** (or codomain) of f and is the set in which we expect an image. The **range** of f is the set of actual images. For example, if $f(x) = x^2$ on the real line, then $f : \mathbb{R} \mapsto \mathbb{R}$ has its target set equal to the real line, but the range is the nonnegative real line. For some algebraists, changing the target set changes the function even if the range remains the same.

We use the notation $f[S]$ to denote the image of a set S under a function f. That is,

$$f[S] := \{y : \exists x \in S \cap \mathrm{domain}(f) \ \text{ with } f(x) = y\}.$$

Note that if $f : A \mapsto B$, then $f[\cdot]$ is a function from $\mathscr{P}(A)$ into $\mathscr{P}(B)$. We use square brackets here because the set function derived from a point mapping f is not the same function as f. The notation f^{-1} denotes the inverse relation of the point mapping f. That is, f^{-1} is the relation on $\mathrm{range}(f) \times A$ given by $f^{-1} := \{(y,x) : (x,y) \in f\}$. Note that in general, f^{-1} is not a function since more than one element of the domain of f may map to the same y in the range. We use the notation $f^{-1}[S]$ with square brackets to denote the inverse image under a function f of a set S contained in the target set of f. That is,

$$f^{-1}[S] := \{x \in \mathrm{domain}(f) : \exists y \in S \text{ with } f(x) = y\}.$$

If $f : A \mapsto B$, then $f^{-1}[\cdot]$ is a function mapping $\mathscr{P}(B)$ into $\mathscr{P}(A)$.

We will write $f \equiv \alpha$ on A if $f(x) = \alpha$ for all points x in a subset A of the domain of f. We will write $f|A$ for the restriction of a function f to a subset A of its domain.

If $f : A \mapsto B$, then, in general, $f[A] \subseteq B$. If, however, the range of f is all of B, then f is a map **onto** B; we also call f a **surjection** or say that f is **surjective**. If for all $x, z \in A$, $f(x) = f(z)$ only if $x = z$, then f is a **one-to-one** map, also called an **injection**; we also say that f is **injective**. If f is both onto and one-to-one, then it is called a one-to-one correspondence or a **bijection**; we also say that f is **bijective**. If $f : A \mapsto B$ is a bijection, then f^{-1} is a bijection mapping B onto A. For example, if $x \mapsto x^2$ is restricted to the positive real numbers, then f^{-1} is the square root function. A real-valued function f is **bounded** if its range is contained $[-K, K]$ for some K in $\mathbb{N}$.

Recall that a finite set is one that is empty or that can be put into one-to-one correspondence with an initial segment of the natural numbers terminating with some $n \in \mathbb{N}$. As usual, non-finite sets are called infinite. A **countably infinite** set is one that can be put into one-to-one correspondence with the natural numbers. We will call such a correspondence an **enumeration**. In particular, the integers and rational numbers are countably infinite; as we shall see, the real numbers are not. A set that is not countable is called **uncountable**. Often for simplicity, we let "countable" mean finite or countably infinite. An example is when we state the following fact:

Theorem 1.1.1. *A countable union of countable sets is countable.*

Proof. Exercise 1.2.

To see how large the collection of subsets of an infinite set can be, let us consider the set of all subsets of the natural numbers $\mathbb{N} = \{1, 2, 3, \cdots\}$, i.e., the power set $\mathscr{P}(\mathbb{N})$. The ordinary language we use to describe something has only a countable number of symbols that we string together in finite sentences. This means that even for the natural numbers, we can only describe a countable number of subsets. Here is a proof that there are actually an uncountable number of such subsets: Assume we have an enumeration A_1, A_2, $\cdots$ of the distinct subsets of $\mathbb{N}$. The enumeration should include the empty set. Form B by putting m in B if and only if m is not in A_m. Then B has not been counted in the enumeration. The fact that $\mathscr{P}(\mathbb{N})$ is uncountable can be applied to the binary expansion of numbers in the interval $[0, 1]$, while taking into account duplicate expansions for some numbers, to show that $[0, 1]$, and therefore $\mathbb{R}$, is uncountable.

1.2 Indexed Families

Let I be a set such that for each $\alpha \in I$ we have a set A_α. Then I is called an **index set** for the family $\{A_\alpha\}_{\alpha \in I}$. Note that I may be a finite set, or the natural numbers, or even an uncountable set. We will often deal with arbitrarily indexed families of sets and of functions. For example, we have the following rules: If f maps a set X into a set Y, then for all $A_\alpha \subseteq X$,

$$f\left[\bigcup_\alpha A_\alpha\right] = \bigcup_\alpha f[A_\alpha], \quad f\left[\bigcap_\alpha A_\alpha\right] \subseteq \bigcap_\alpha f[A_\alpha].$$

Note that the containment goes only one way for intersection. For example, f may take the same constant value on every one of the sets A_α and these sets may be pairwise disjoint. For inverse images, we can say more. That is, if for each index α, B_α is a subset of the range of f, then

$$f^{-1}\left[\bigcup_\alpha B_\alpha\right] = \bigcup_\alpha f^{-1}[B_\alpha] \quad \text{and} \quad f^{-1}\left[\bigcap_\alpha B_\alpha\right] = \bigcap_\alpha f^{-1}[B_\alpha].$$

We also have $f^{-1}[\complement B] = \complement(f^{-1}[B])$ for all B contained in the range of f. Note that the first complement is with respect to the range and the second is with respect to the domain of f.

The distributive law for sets says that

$$B\bigcap\left[\bigcup_\alpha A_\alpha\right] = \bigcup_\alpha (B\cap A_\alpha) \quad \text{and} \quad B\bigcup\left[\bigcap_\alpha A_\alpha\right] = \bigcap_\alpha (B\cup A_\alpha).$$

Here is an important rule that we will use often; the proof is left to the reader.

Theorem 1.2.1 (De Morgan's Laws). *Given two or more subsets of a set X forming an indexed family of sets* $\{A_\alpha\}_{\alpha\in I}$,

$$\complement\left[\bigcup_{\alpha\in I} A_\alpha\right] = \bigcap_{\alpha\in I} \complement A_\alpha \quad \textit{and} \quad \complement\left[\bigcap_{\alpha\in I} A_\alpha\right] = \bigcup_{\alpha\in I} \complement A_\alpha.$$

We will often employ what is called the **Axiom of Choice** without explicitly noting its use. The Axiom of Choice states that if $\{S_\alpha : \alpha \in \mathscr{I}\}$ is a nonempty collection of nonempty sets, then there is a function $T : \mathscr{I} \mapsto \bigcup_{\alpha\in\mathscr{I}} S_\alpha$ such that for each $\alpha \in \mathscr{I}$, $T(\alpha) \in S_\alpha$. Equivalent statements are discussed in an appendix.

Bertrand Russell gives an example in terms of pairs of shoes and pairs of socks: Presented with a finite or even countably infinite set of pairs of shoes, one can always pick one shoe from each pair, e.g., the left shoe. Given a finite collection of pairs of socks, one can pick one sock from each pair, but what happens with an infinite collection of pairs of socks when the two socks in any pair are identical? The Axiom of Choice says that even without a rule, there is a set consisting of exactly one sock from each pair.

1.3 Algebras and σ-Algebras of Sets

When an operation on a collection of sets, functions, or numbers always produces another member of the collection, it is common to say that the collection is **closed** with respect to the operation. Because the term "closed" has another meaning in real analysis, we shall say the collection is **stable** with respect to the operation.

Definition 1.3.1. An **algebra** $\mathscr{A}$ of sets in a set X is a collection of subsets of X containing the set X as a member and stable with respect to the operations of complementation and forming finite unions.

Remark 1.3.1. It follows that an algebra contains the empty set $\varnothing$ and is stable with respect to forming finite intersections. If one does not specify that $X \in \mathscr{A}$, then one should require that $\mathscr{A}$ is nonempty so that then there is a set and its complement in $\mathscr{A}$, whence the union $X \in \mathscr{A}$.

Definition 1.3.2. A **σ-algebra** of sets in a set X is a collection of subsets of X containing X as a member and stable with respect to the operations of complementation and forming countably infinite unions.

Remark 1.3.2. A σ-algebra contains the empty set and by De Morgan's laws is also stable with respect to forming countable intersections. Since the empty set is in a σ-algebra, a σ-algebra is also an algebra. A σ-algebra is also called a **Borel field**, especially in the theory of probability.

An intersection of a nonempty collection of algebras or σ-algebras is again an algebra or, respectively, a σ-algebra. For each set X, the power set $\mathscr{P}(X)$ is an algebra and a σ-algebra. Therefore, for any collection $\mathscr{C}$ of sets in X there is a smallest algebra and also a smallest σ-algebra (denoted by $\sigma(\mathscr{C})$) that contains $\mathscr{C}$. It follows that if $\mathscr{B}$ is a σ-algebra of subsets of X and $\mathscr{B} \supseteq \mathscr{C}$, then $\mathscr{B} \supseteq \sigma(\mathscr{C})$.

The σ-algebra generated by a nontrivial collection $\mathscr{C}$ contains more than just countable unions and intersections of the sets in $\mathscr{C}$. For example, if $\mathscr{C}$ is the collection of open intervals in the real line, then $\sigma(\mathscr{C})$ contains countable unions, countable intersections of countable unions, countable unions of countable intersections of countable unions of sets in $\mathscr{C}$, etc.

Example 1.3.1. The algebra generated by intervals of the form $[\alpha, \beta)$, with α and β rational and $\alpha < \beta$, consists of finite unions of such intervals. The σ-algebra generated by this algebra is a much larger collection of sets. For example, for each real number r, the singleton set $\{r\}$ is in the σ-algebra but not in the algebra. Other interesting examples are the algebra and σ-algebra generated by the finite sets in $\mathbb{R}$.

Here is a result we will need for measure theory. A disjoint, or pairwise disjoint, sequence of sets is a sequence of sets with no point contained in any two of the sets.

Proposition 1.3.1. *If $\mathscr{A}$ is an algebra of sets and $\langle A_i : i \in \mathbb{N}\rangle$ is a sequence from $\mathscr{A}$, then there is a disjoint sequence $\langle B_i \rangle$ from $\mathscr{A}$ with the same union such that $B_i \subseteq A_i$ for each $i \in \mathbb{N}$.*

Proof. Set $B_1 = A_1$, and for $n > 1$, set $B_n = A_n \setminus \bigcup_{i<n} A_i$.

Proposition 1.3.2. *Let $\mathscr{C}$ be any collection of nonempty subsets of some set X, and $E \in \sigma(\mathscr{C})$, i.e., the smallest σ-algebra containing all the sets in $\mathscr{C}$. Then for some countable collection $\mathscr{C}_0 \subseteq \mathscr{C}$, $E \in \sigma(\mathscr{C}_0)$.*

Proof. Exercise 1.8(A).

1.4 Orderings

An ordering on a set is a relation on the set satisfying the properties listed below. The reader is familiar with the ordering $\leq$ on $\mathbb{R}$ and $\subseteq$ on the power set of a set.

Definition 1.4.1. A **partial ordering** on a nonempty set E is a relation $\leq$ on E such that

1) $\forall\, a \in E$, $a \leq a$, (**reflexive property**) and
2) $\forall\, a, b, c \in E$, $a \leq b$ and $b \leq c \Rightarrow a \leq c$. (**transitive property**)
We sometimes write $b \geq a$ for $a \leq b$.
We call a partial ordering **antisymmetric** if
3) $\forall\, a, b \in E$, $a \leq b$ and $b \leq a \Rightarrow a = b$.
For an antisymmetric ordering we write $a < b$ for $a \leq b$ but $a \neq b$.
A partial ordering $\leq$ is called a **total ordering** if
4) $\forall\, a, b \in E$, either $a \leq b$ or $b \leq a$.
A total ordering $\leq$ is called a **linear ordering** if it is antisymmetric.

We do not assume that a partial ordering is antisymmetric. There are many important instances of partial orderings that are not antisymmetric. One example is the ordering on Riemann sums, which we will discuss in defining the Riemann integral. An ordering can be put on a river and its tributaries with point b being further along than point a if it is downstream of a. If the river has any width, the ordering is not antisymmetric. Also note that if the points are on different tributaries, neither may be downstream of the other.

If E has a partial ordering, it is said to be a partially ordered set. We will distinguish between $z \in E$ being the **biggest** or **greatest element** of E and z being a **maximal element** of E. The first means that for all $y \in E$, y and z are comparable with $y \leq z$, and if we also have $z \leq y$, then $z = y$. A greatest element is unique. The second means that for each $y \in E$, y need not be related to z, but if $y \geq z$, then we also have $y \leq z$. For an antisymmetric ordering, this means that $y = z$. Otherwise, this may not be the case. Similar definitions hold for the **smallest** or **least element**, and **minimal element**.

Example 1.4.1. An example is given by the set of points (x, y) in the plane with $0 \leq y \leq 1$, and $(x_1, y_1) \leq (x_2, y_2)$ if and only if $y_1 \leq y_2$; points of the form $(x, 1)$ are maximal. There is no greatest element, and this total ordering is not antisymmetric.

Definition 1.4.2. Given a set E with a partial ordering $\leq$ and a subset S of E, $x \in E$ is an **upper bound** of S if for every $a \in S$, $a \leq x$. An upper bound of S may or may not be an element of S. An upper bound x is a **least upper bound** of S if it is the least of the upper bounds. Lower bounds and greatest lower bounds are similarly defined.

Example 1.4.2. Given a set X, the power set is partially ordered by $\subseteq$; by definition, this ordering is antisymmetric. The greatest element is X. Any collection $\mathscr{S}$ of subsets of X has a least upper bound, which is the union of the sets in $\mathscr{S}$.

Definition 1.4.3. A linear ordering on a set E is called a **well-ordering** if every nonempty subset of E contains a first (i.e., least) element.

Example 1.4.3. The set $\mathbb{N}$ of natural numbers with the usual ordering is a well-ordered set.

Throughout this book, we will use $\mathbb{R}^+$ to denote the nonnegative real numbers and $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$ to denote the extended real numbers. The linear ordering $\leq$ on $\mathbb{R}$ is extended to $\overline{\mathbb{R}}$ by setting $-\infty < x < +\infty$ for all $x \in \mathbb{R}$. When working with functions, we will use terminology such as "real-valued function" and "extended-real valued function." That is, we leave out the hyphen in the latter case since extending a function refers to augmenting the domain.

1.5 Sequences in $\mathbb{R}$ and $\overline{\mathbb{R}}$

A sequence is a function using the natural numbers for the domain. The image of $n \in \mathbb{N}$ is often denoted by x_n, a_n, f_n, etc. The sequence itself will usually be denoted here using angle brackets; for example, $\langle x_n : n \in \mathbb{N} \rangle$ or just $\langle x_n \rangle$. A real-valued sequence **converges** in $\mathbb{R}$ if there is a real number x such that, for each $\varepsilon > 0$ there is a $K \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ for all $n \geq K$ in $\mathbb{N}$. The number x is the unique limit of the sequence. We write $\lim_{n\to\infty} x_n = x$ and say $\langle x_n \rangle$ has limit x or converges to x. We may also write $x_n \to x$. Moreover, if x_n is increasing, that is, $x_n \leq x_{n+1}$ for each n, then we may write $x_n \nearrow x$; the notation $x_n \searrow x$ may be used if x_n is decreasing.

A sequence $\langle x_n \rangle$ in $\mathbb{R}$ is a **Cauchy sequence** if for each $\varepsilon > 0$ there is a $K \in \mathbb{N}$ such that for all $m \geq K$ and $n \geq K$ in $\mathbb{N}$, $|x_n - x_m| < \varepsilon$. An important property of the real numbers is that a real sequence converges if and only if it is a Cauchy sequence. The requirement of a limit for every Cauchy sequence is often used to define the real numbers $\mathbb{R}$ as an extension of the rational numbers $\mathbb{Q}$.

Theorem 1.5.1. *Any nonempty subset of $\mathbb{R}$ with an upper bound in $\mathbb{R}$ has a least upper bound in $\mathbb{R}$. Any decreasing sequence of finite closed intervals I_n in $\mathbb{R}$, i.e., $I_n \supseteq I_{n+1}$ for each n, has a nonempty intersection.*

Proof. Let A be a nonempty set with a finite upper bound. For each $n \in \mathbb{N}$, we may choose $a_n \in A$ and an upper bound b_n of the set A such that $b_n - a_n \leq 1/n$. It is left to the reader (Exercise 1.15) to show that the b_n's form a Cauchy sequence, and the limit is the least upper bound of A. For a decreasing sequence of closed intervals, the set of left endpoints has a least upper bound, which is in all of the intervals.

A sequence $\langle x_n \rangle$ in $\mathbb{R}$ has a **cluster point** x_0 if for each $\varepsilon > 0$ and each $K \in \mathbb{N}$ there is an $n \geq K$ in $\mathbb{N}$ for which $|x_n - x_0| < \varepsilon$. If a point $x_0 \in \mathbb{R}$ is the limit of a sequence, it is the unique cluster point of the sequence. A sequence may have many cluster points but no limit, such as the sequence $n \mapsto (-1)^n$; it may also have no cluster point in $\mathbb{R}$, such as the sequence $n \mapsto n$.

An equivalent condition for convergence of a real-valued sequence $\langle x_n \rangle$ to x is that for each open interval I that contains x there is a natural number K so that for all $n \geq K$ in $\mathbb{N}$ we have $x_n \in I$. We may say that the sequence is **eventually** in every open interval containing x. If x_0 is a cluster point of a sequence $\langle x_n \rangle$, then for any open interval I containing x_0 and any $K \in \mathbb{N}$ there is an $n \geq K$ with $x_n \in I$. We may say the sequence is **frequently** in every open interval containing x_0.

Since the extended real numbers form a linearly ordered set, the notion of upper bound and lower bound is the same as for any linear ordered set. A nonempty subset of $\mathbb{R}$ also has a least upper bound, denoted by **lub** or **sup** for **supremum**. If the set does not have a finite upper bound, then the least upper bound is $+\infty$. It also follows by multiplying members in a subset of $\mathbb{R}$ by -1 that a nonempty subset of the reals has a greatest lower bound, denoted by **glb** or **inf** for **infimum**. If the set does not have a finite lower bound, then the greatest lower bound is $-\infty$. In general, if $A \subseteq \mathbb{R}$ is not empty, then $\inf A \leq \sup A$, with equality holding if and only if A consists of only one point. However, since every real number is both an upper bound and a lower bound of the empty set, $\sup(\varnothing) = -\infty$ and $\inf(\varnothing) = +\infty$.

If $\langle x_n \rangle$ is a real-valued sequence, then $\lim_{n\to\infty} x_n = +\infty$ if for each $M > 0$ in $\mathbb{R}$ there is a $K \in \mathbb{N}$ such that $x_n > M$ for all $n \geq K$. In this case, the sequence diverges to plus infinity. A similarly definition gives $\lim_{n\to\infty} x_n = -\infty$. The point $+\infty$ is a cluster point of the sequence if x_n is frequently in each interval of type (M, ∞) for $M \in \mathbb{R}$. A similarly definition is used for a cluster point at $-\infty$. A **bounded sequence** is one for which the range is contained in a finite interval in $\mathbb{R}$.

Theorem 1.5.2. *A real-valued sequence has a cluster point in the extended real line* $\overline{R}$.

Proof. If neither $+\infty$ nor $-\infty$ is a cluster point, then for some $M \in \mathbb{N}$, there is a $K \in \mathbb{N}$ such that for all $n \geq K$, $x_n \in [-M, M]$ (Exercise 1.13). In this case, divide $[-M, M]$ into $2M$ closed intervals of length 1; i.e., $[-M, -M+1], \cdots, [M-1, M]$. The sequence cannot be eventually in the complement of each of these intervals. Let $I_1 = [m, m+1]$ be the first, counting from the left, of these closed intervals to which the sequence frequently returns. That is, for each $J \in \mathbb{N}$ there is an $n > J$ with $x_n \in I_1$. Divide I_1 in half, forming $[m, m+1/2]$ and $[m+1/2, m+1]$. Let I_2 be the first, counting from the left, of these two intervals to which the sequence frequently returns. Continue in this way, each time cutting the previous interval in half. At the n^{th} stage, let I_{n+1} be the first, counting from the left, of the two intervals formed from I_n to which the sequence frequently returns. Now for each $n \in \mathbb{N}$, let $I_n = [a_n, b_n]$. Then $a_n \leq a_{n+1} < b_{n+1} \leq b_n$, and $\lim(b_n - a_n) = 0$. Therefore, the a_n's form a Cauchy sequence that converges to a point x_0, which is also a cluster point of the original sequence. That is, given $\varepsilon > 0$, there is a $K \in \mathbb{N}$ such that for all $n \geq K$, $b_n - a_n < \varepsilon$, $a_n \leq x_0 \leq b_n$, and for infinitely many j's after K, $a_n \leq x_j \leq b_n$, whence $|x_j - x_0| \leq b_n - a_n < \varepsilon$.

Remark 1.5.1. If for $n \geq K$ the sequence is in a bounded interval, then since only a finite number of terms precede K, the sequence is bounded. The main part of Theorem 1.5.2 is called the Bolzano-Weierstrass theorem. We will give another proof when we take up compactness.

Definition 1.5.1. A **subsequence** of a sequence $\langle x_n : n \in \mathbb{N} \rangle$ is given by a function s from $\mathbb{N}$ into $\mathbb{N}$, such that for each $k \in \mathbb{N}$, $s(k+1) > s(k) \geq k$. The corresponding subsequence is the map $k \mapsto s(k) \mapsto x_{s(k)}$ from $\mathbb{N}$ into the range of the original sequence. The image of $k \in \mathbb{N}$ is often denoted by x_{n_k}.

Theorem 1.5.3. *If $x_0 \in \overline{\mathbb{R}}$ is a cluster point of a sequence $\langle x_n : n \in \mathbb{N} \rangle$, then a subsequence $\langle x_{n_k} : k \in \mathbb{N} \rangle$ converges to x_0.*

Proof. If $+\infty$ is a cluster point, then there is a function s from $\mathbb{N}$ into $\mathbb{N}$, such that for each $k \in \mathbb{N}$, $s(k+1) > s(k) \geq k$, and $x_{s(k)} \geq k$. The sequence $x_{s(k)}$ converges to $+\infty$. If $-\infty$ is a cluster point, then $+\infty$ is a cluster point of $-x_n$; a subsequence $-x_{r(k)}$ converges to $+\infty$, so the subsequence $x_{r(k)}$ converges to $-\infty$. If x_0 is a real-valued cluster point, then there is a map s from $\mathbb{N}$ into $\mathbb{N}$, such that for each $k \in \mathbb{N}$, $s(k+1) > s(k) \geq k$, and $\left|x_{s(k)} - x_0\right| < 1/k$.

Corollary 1.5.1. *If $\langle x_n \rangle$ is a sequence of real numbers, then a subsequence $\langle x_{n_i} \rangle$ converges to a real or extended real number.*

1.6 Lim sup and Lim inf

Definition 1.6.1. If a and b are real numbers, $a \vee b$ denotes the maximum of a and b, while $a \wedge b$ denotes the minimum of a and b. If $\langle x_n \rangle$ is a sequence in $\mathbb{R}$, $\bigvee_k x_k$ denotes the supremum, i.e., lub of the x_k's, and $\bigwedge_k x_k$ denotes the infimum, i.e., glb of the x_k's. The **limit superior** is the extended real number

$$\limsup x_n := \inf_{n \in \mathbb{N}} (\sup_{k \geq n} x_k) = \bigwedge_{n \in \mathbb{N}} \left(\bigvee_{k \geq n} x_k \right).$$

We may also use $\overline{\lim} x_n$ to denote this number. The **limit inferior** is the extended real number

$$\liminf x_n := \sup_{n \in \mathbb{N}} (\inf_{k \geq n} x_k) = \bigvee_{n \in \mathbb{N}} \left(\bigwedge_{k \geq n} x_k \right).$$

We may also use $\underline{\lim} x_n$ to denote this number.

Remark 1.6.1. Note that $n \mapsto \bigvee_{k \geq n} x_k$ is a decreasing sequence in $\overline{\mathbb{R}}$ and $n \mapsto \bigwedge_{k \geq n} x_k$ is an increasing sequence in $\overline{\mathbb{R}}$.

Proposition 1.6.1. *If $\langle x_n \rangle$ is a sequence in $\mathbb{R}$ and $L = \limsup x_n$ is in $\mathbb{R}$, then given any $\varepsilon > 0$, the sequence is eventually less than $L + \varepsilon$ and frequently greater than $L - \varepsilon$. If $\limsup x_n = +\infty$, then for all $M > 0$, the sequence is frequently greater than M. If $\limsup x_n = -\infty$, then $\lim x_n = -\infty$, whence for all $M > 0$, the sequence is eventually less than $-M$. Similar statements are true for $\liminf x_n = -\limsup(-x_n)$. It follows that $\limsup x_n$ is the largest cluster point of the sequence in $\overline{\mathbb{R}}$ and $\liminf x_n$ is the smallest cluster point of the sequence in $\overline{\mathbb{R}}$.*

Proof. The proof is left to the reader.

Corollary 1.6.1. *If $\langle x_n \rangle$ is a sequence in $\mathbb{R}$, then $\liminf x_n \leq \limsup x_n$, and the two are equal if and only if they are the limit of the sequence in the extended real numbers.*

Proposition 1.6.2. *Let $\langle x_n \rangle$ and $\langle y_n \rangle$ be real-valued sequences. Then provided we never add or subtract $+\infty$ from itself, we have*

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

Proof. To show the first inequality, we may assume the left side is not $-\infty$ since otherwise, the first inequality is trivial. If $\underline{\lim}\, x_n = +\infty$, then since, by assumption, $\underline{\lim}\, y_n \neq -\infty$, the y_n's are bounded below and eventually the x_n's take arbitrarily large values, so the sum has $\underline{\lim}$ equal to $+\infty$. A similar fact is true if $\underline{\lim}\, y_n = +\infty$. Assume both terms on the left of the inequality are finite. Given $\varepsilon > 0$, we fix a k such that for all $m \geq k$, $\underline{\lim}\, x_n - \varepsilon/2 < x_m$ and $\underline{\lim}\, y_n - \varepsilon/2 < y_m$. Now it is true that for all $m \geq k$, that is, eventually, $\underline{\lim}\, x_n + \underline{\lim}\, y_n - \varepsilon \leq x_m + y_m$. Next, we note that for an infinite number of m's greater than k, i.e., frequently, $x_m + y_m \leq \underline{\lim}\,(x_n + y_n) + \varepsilon$. It follows that $\underline{\lim}\, x_n + \underline{\lim}\, y_n \leq \underline{\lim}\,(x_n + y_n)$. To establish the second inequality, we use the fact that

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

whence

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

The rest is left to the reader.

1.7 Open and Closed Sets

Definition 1.7.1. A set $O \subseteq \mathbb{R}$ is **open** if for all $x \in O$, there is an open interval $(x - \varepsilon, x + \varepsilon) \subseteq O$. A set is **closed** if its complement in $\mathbb{R}$ is open.

Remark 1.7.1. Note that a set is open if and only if its complement is closed. Therefore, a set A is closed if and only if for each point z not in the set, there is an open interval $(z - \varepsilon, z + \varepsilon) \subseteq \complement A$.

Proposition 1.7.1. *The collection of open sets is stable under the operations of taking finite intersections and arbitrary unions. The collection of closed sets is stable under the operations of taking finite unions and arbitrary intersections.*

Proof. Let $G = \bigcap_{i=1}^{n} O_i$ where the O_i are open sets. Let $x \in G$. Then $x \in O_i$ for each i, and so there are $\delta_i > 0$ such that $(x - \delta_i, x + \delta_i) \subseteq O_i$. Set $\delta := \bigwedge_{i=1}^{n} \delta_i$. Then $\delta > 0$ and $(x - \delta, x + \delta) \subseteq O_i$ for each i, whence $(x - \delta, x + \delta) \subseteq G$. It follows that G is open. Now let $G = \cup O_\alpha$ for an arbitrary collection of open sets O_α. If $x \in G$, then

$x \in O_\beta$ for some index β, whence there is a $\delta > 0$ such that $(x-\delta, x+\delta) \subseteq O_\beta \subseteq G$. It follows that G is open. The results for closed sets follow from these facts for open sets and De Morgan's laws.

Remark 1.7.2. The sets $\mathbb{R}$ and $\varnothing$ are both open and closed in $\mathbb{R}$; these are the only sets both open and closed in $\mathbb{R}$ (Exercise 1.22).

Proposition 1.7.2. *Every nonempty open set in $\mathbb{R}$ is the union of a countable collection of disjoint open intervals.*

Proof. Once we obtain a collection of disjoint intervals, the collection must be finite or countably infinite because each interval contains a rational number not in any of the other intervals, and the rational numbers form a countable set. Given a nonempty open set O, to each $x \in O$ we associate the interval $I_x = (a_x, b_x)$ where $a_x = \inf\{z \in O : (z,x) \subseteq O\}$ and $b_x = \sup\{z \in O : (x,z) \subseteq O\}$. Of course, a_x may equal $-\infty$ and b_x may equal $+\infty$. Since O is open, I_x is not the empty set, and neither a_x nor b_x is in O. For each $x \in O$, $x \in I_x \subseteq O$, so $\cup_x I_x = O$. We now need only show that if $x < y$ in O and $I_x \cap I_y \neq \varnothing$, then $I_x = I_y$, since then the I_x's form the desired disjoint collection. Suppose there is a point $w \in I_x \cap I_y$. If $w \leq x$, then $x \in I_y$, and so

$$a_y = \inf\{z \in O : z < x \text{ and } (z,x) \subseteq O\} = a_x.$$

If $x < w$, then the closed interval $[x,w] \subseteq I_x \subseteq O$. It follows that $x \in I_y$, whence again, $a_x = a_y$. Similarly, $b_x = b_y$. Therefore, the collection $\{I_x : x \in O\}$ is the desired collection.

Theorem 1.7.1 (Lindelöf). *The union of a collection $\mathscr{C}$ of open subsets of the real line is equal to the union of some countable subcollection of $\mathscr{C}$.*

Proof. Let $\mathscr{D} = \{I_n : n \in \mathbb{N}\}$ be the collection of all open intervals with rational endpoints such that each I_n is contained in a set $O \in \mathscr{C}$. The union over the intervals I_n is equal to the union over the sets in $\mathscr{C}$. Replace each I_n with one of the sets $O_n \in \mathscr{C}$ such that $I_n \subseteq O_n$. (Here we are using the Axiom of Choice.) The collection $\{O_n\}$ (with no repetitions) is the desired collection.

The following corollary, due to Aldaz [2], works for the real line; it will be useful later. Recall that an interval is non-degenerate if it is not a singleton set. A point is in the interior of an interval if it is in the interval but is not an endpoint of the interval.

Corollary 1.7.1. *If $\mathscr{C}$ is a collection of non-degenerate intervals in $\mathbb{R}$, then the union of the intervals in $\mathscr{C}$ equals the union of some countable subcollection of $\mathscr{C}$.*

Proof. Let $\mathscr{A}$ be the collection of those intervals in $\mathscr{C}$ such that each $I \in \mathscr{A}$ contains its left endpoint, and that left endpoint is not in the interior of any interval in $\mathscr{C}$. If $I \in \mathscr{A}$ and a is its left endpoint, then for some $\varepsilon > 0$, the open interval $(a, a+\varepsilon) \subseteq I$, and no left endpoint of any member of $\mathscr{A}$ lies in $(a, a+\varepsilon)$. Moreover, $(a, a+\varepsilon)$ contains a rational number. It follows that the collection $\mathscr{A}_L$ of left endpoints of intervals in $\mathscr{A}$ is countable. Let $\mathscr{B}$ be the collection of those intervals in $\mathscr{C}$ such that each $I \in \mathscr{B}$ contains its right endpoint but that right endpoint is not in the interior of any interval in $\mathscr{C}$. By an argument similar to that for $\mathscr{A}$, the collection $\mathscr{B}_R$ of right endpoints of intervals in $\mathscr{B}$ also forms a countable set. A countable subcollection of $\mathscr{C}$ contains the points of $\mathscr{A}_L \cup \mathscr{B}_R$. The rest of the union of the intervals in $\mathscr{C}$ is covered by the intervals of $\mathscr{C}$ with any endpoints removed, so by Lindelöf's theorem, a countable subcollection has the same union. Since a union of two countable sets is countable, we are done.

1.8 Closures

Definition 1.8.1. A point $x \in \mathbb{R}$ is a **point of closure** of a set $A \subseteq \mathbb{R}$ if every open interval about x contains a point of A. A point $x \in \mathbb{R}$ is an **accumulation point** of a set $A \subseteq \mathbb{R}$ if every open interval about x contains a point of $A \setminus \{x\}$.

Note that a point of closure of a set A need not be in A, but every point in A is a point of closure of A. In particular, any point of closure of A that is not an element of A is an accumulation point of A. Note that a point of A may or may not be an accumulation point of A.

Example 1.8.1. Every natural number is a point of closure of $\mathbb{N}$, but there are no accumulation points of $\mathbb{N}$. The point 0 is the only accumulation point of the set $\{1/n : n \in \mathbb{N}\}$.

Definition 1.8.2. The closure $\overline{E}$ of a set E is the set E together with all of its accumulation points.

Remark 1.8.1. It is easy to see that $\overline{A \cup B} = \overline{A} \cup \overline{B}$ and if $A \subseteq B$ then $\overline{A} \subseteq \overline{B}$. Therefore, $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$, but the containment does not necessarily go the other way. For example, $A \cap B$ may be empty.

Proposition 1.8.1. *A set F is closed if and only if $F = \overline{F}$; that is, every accumulation point of F is a member of F.*

Proof. A set F is closed if and only if each point of $\mathbb{R} \setminus F$ is contained in an open interval that does not intersect F, if and only if all points of closure of F are contained in F, if and only if $F \subseteq \overline{F} \subseteq F$.

Proposition 1.8.2. *The closure of the closure is the closure; i.e., if $E \subseteq \mathbb{R}$, then $\overline{(\overline{E})} = \overline{E}$. It follows that the closure of a set is a closed set.*

Proof. If z is a point of closure of $\overline{E}$ and $I = (z - \varepsilon, z + \varepsilon)$ for some $\varepsilon > 0$, then there is a $y \in \overline{E} \cap I$. Since y is a point of closure of E, there is an $x \in E \cap I$. This means that z is a point of closure of E. Therefore, we cannot obtain a strictly larger set by going from $\overline{E}$ to $\overline{(\overline{E})}$.

Proposition 1.8.3. *The closure of a set E is the intersection of all closed sets that contain E.*

Proof. By Proposition 1.7.1, if F is the intersection of all closed sets that contain E, then F is closed. Since $\overline{E}$ is a closed set, $E \subseteq F \subseteq \overline{E} \subseteq \overline{F} = F$. Therefore, $\overline{E} = F$.

A notion analogous to the closure of a set is that of the interior of a set.

Definition 1.8.3. A point x is in the **interior** A° of a set A if for some open interval I_x centered at x, we have $x \in I_x \subseteq A$.

Remark 1.8.2. Since each $y \in I_x$ is also in A°, it follows that $A^\circ = \bigcup_{x \in A^\circ} I_x$ is an open subset of A. If y is in an open set $O \subseteq A$, then there is an open interval I_y centered at y with $y \in I_y \subseteq O \subseteq A$. It follows that $O \subseteq A^\circ$, so A° is the largest open subset of A; it is the union of all open subsets of A. Therefore, A is open if and only if $A = A^\circ$.

1.9 Directed Sets and Nets

In this section, we generalize sequences and sequential convergence.

Definition 1.9.1. A **directed set** is a set D supplied with a partially ordering $\leq$ having the additional property that for all a, b in D, there is a $c \in D$ with $a \leq c$ and $b \leq c$.

Remark 1.9.1. It may not be the case that a and b are comparable with respect to the ordering in D, but there is an element comparable to both and further along than both. Antisymmetry is not needed for the ordering $\leq$.

Definition 1.9.2. A **net** or **generalized sequence** in the real line is a mapping from a directed set into $\mathbb{R}$. A net $\langle x_\alpha : \alpha \in D \rangle$ converges to a point L if for any $\varepsilon > 0$, there is an $\alpha_0 \in D$ such that for all $\alpha \geq \alpha_0$ in D, $|L - x_\alpha| < \varepsilon$. That is, with respect to the ordering on D, the net is eventually in every open interval about L. Convergence in the extended real line is similarly defined.

Example 1.9.1. $D = \mathbb{N}$ with the usual ordering gives sequential convergence.

Example 1.9.2. A river with tributaries flowing downstream is a directed set with $a \leq b$ if b is downstream of a. The height above sea-level is a net using this directed set.

Example 1.9.3. Fix $c \in \mathbb{R}$, and let $D = \mathbb{R} \setminus \{c\}$. Set $a \preccurlyeq b$ in D if $|b - c| \leq |a - c|$. For a real-valued function f on $\mathbb{R}$, this gives the notion of $\lim_{x \to c} f(x)$. Note here that $c - 1 \preccurlyeq c + 1 \preccurlyeq c - 1$. That is, we do not have antisymmetry for $\preccurlyeq$.

Example 1.9.4. Fix an interval $[a,b]$ in $\mathbb{R}$, and let $\mathscr{P}$ be the set of finite partitions of $[a,b]$ used for Riemann sums. Each element of $\mathscr{P}$ consists of a finite collection of **partition points** $a = x_0 < x_1 < \cdots < x_n = b$ for some $n \in \mathbb{N}$, and also what are called **tags** $z_i \in [x_{i-1}, x_i]$ for $1 \leq i \leq n$. For a bounded function $f : [a,b] \mapsto \mathbb{R}$, the Riemann sum formed by these partition points and tags is $\sum_{i=1}^{n} f(z_i)(x_i - x_{i-1})$. Given two partitions P_1 and P_2, we have $P_1 \leq P_2$ if the partition points of P_1 are also partition points of P_2. For the Riemann integral, the tags are arbitrary within the required subintervals. This gives the general convergence notion for Riemann sums converging to the Riemann integral. Note here that we do not have, and absolutely do not want, antisymmetry for $\leq$.

As noted, nets generalize sequences and sequential convergence. We also want to extend the notion of convergence for a series of numbers. Given a series of nonnegative numbers, a sum exists if the partial sums have a finite upper bound. One can come arbitrarily close to the same sum by adding the numbers in a sufficiently large finite subset of the set of numbers in the series. Adding any more numbers from the series will not change the sum by much. That is the idea of an unordered sum.

Example 1.9.5. To obtain an **unordered sum** of nonnegative numbers forming a set E, the directed set is the collection $\mathscr{F}$ of finite subsets of E. Containment $\subseteq$ is the ordering on $\mathscr{F}$. This makes $\mathscr{F}$ a directed set since if A and B are finite subsets of E, then while neither may be a subset of the other, the union contains them both. If it exists, the unordered sum is a number S such that for any $\varepsilon > 0$, there is a finite subset A of E such that if S_B is the sum of the numbers in any finite set $B \supseteq A$, then $|S - S_B| < \varepsilon$. The proof of the following equivalence of unordered sums and series convergence is left to the reader.

Proposition 1.9.1. *Let E be a set of nonnegative real numbers. If the unordered sum S of the elements of E is a finite number, then there are at most a countable number of nonzero elements in E. If E itself is a countable set and $\langle x_n : n \in \mathbb{N} \rangle$ is an enumeration, then S is the limit of the partial sums of the series $\sum_{n=1}^{\infty} x_n$.*

1.10 Compactness

Definition 1.10.1. A covering of a set A by open sets is a collection of open sets $\{O_\alpha : \alpha \in \mathscr{I}\}$, which we may index by some set $\mathscr{I}$, such that each $x \in A$ is in some O_α. That is, $A \subseteq \cup_{\alpha \in \mathscr{I}} O_\alpha$. We speak of an **open covering** of A. A **finite subcovering** of such a covering is a finite subset $\{O_{\alpha_1}, \cdots, O_{\alpha_n}\}$ of the original covering that itself forms a covering of A.

Definition 1.10.2. A collection of sets has the **finite intersection property** if every finite subcollection has a nonempty intersection.

Theorem 1.10.1 (Heine-Borel). *A set $F \subset \mathbb{R}$ is closed and bounded if and only if every covering of F by open sets has a finite subcovering.*

Remark 1.10.1. In general, one calls a set F with this covering property a **compact** set. The theorem says a set is compact if and only if it is closed and bounded. The reader should be warned that for an infinite-dimensional space, compact sets are closed and bounded but there may be closed and bounded sets, such as the closed unit ball in an infinite-dimensional Hilbert space, that are not compact in terms of the open sets defined using the space's norm. Here is the proof of Theorem 1.10.1 for the space $\mathbb{R}$.

Proof. If $A \subseteq \mathbb{R}$ is not bounded, then the covering $\{(-n,n) : n \in \mathbb{N}\}$ has no finite subcovering. If A is not closed, then there is some accumulation point $x \notin A$. The covering $\{(-\infty, x-1/n) \cup (x+1/n, +\infty) : n \in \mathbb{N}\}$ of A has no finite subcovering.

For the converse, we may reduce the proof to the case that F is a finite closed interval $[a,b] \subset \mathbb{R}$. This follows from the fact that if we are given a closed subset $F \subseteq [a,b]$ and an open covering $\mathscr{C}$ of F, adjoining the open set $\mathbb{R} \setminus F$ to $\mathscr{C}$ gives an open covering of $[a,b]$. If we know this covering has a finite subcovering $\mathscr{D}$, then the open sets in $\mathscr{D}$ other than $\mathbb{R} \setminus F$ form a finite open covering of F.

It only remains to show that for any $a < b$ in $\mathbb{R}$, the interval $[a,b]$ is compact. Let $\mathscr{C}$ be an open covering of $[a,b]$. Each point $x \in [a,b]$ is contained in an open interval $I_x = (x - \delta_x, x + \delta_x)$ that in turn is contained in one of the open sets in $\mathscr{C}$. Let $\mathscr{I} = \{I_x : x \in [a,b]\}$. We now need only show that $\mathscr{I}$ can be reduced to a finite subcovering of $[a,b]$, since each open interval in that subcovering can be replaced with a larger open set from $\mathscr{C}$. (Notice that we have simplified the proof by reducing the generality in two steps.)

Let

$$S := \{x \in [a,b] : \exists \text{ a finite subcover from } \mathscr{I} \text{ of } [a,x]\}.$$

The set S is not empty since it contains all points of $I_a \cap [a,b]$. First, suppose $a < y < b$ and there is a point $x \leq y$ in $S \cap I_y$. Then adding I_y to a finite cover of $[a,x]$, we obtain a finite cover of $[a,z]$ for some $z > y$. That is, no point less than b can be an upper bound of S, so b is the least upper bound of S. But then, given $x < b$ in $I_b \cap S$, we may add I_b to a finite covering of $[a,x]$ to obtain a finite covering of $[a,b]$, so $b \in S$.

Corollary 1.10.1. *A collection $\mathscr{C}$ of closed sets in $\mathbb{R}$ with the finite intersection property has a nonempty intersection provided at least one of the sets is bounded.*

Proof. Let K be a bounded set in $\mathscr{C}$; then, K is compact. Let $\mathscr{O}$ be the collection of complements of the other sets in $\mathscr{C}$; these are all open. No finite subcollection of $\mathscr{O}$ can cover K since then the intersection of K with the closed complements of the sets in that subcollection would be empty. Therefore, since K is compact, $\mathscr{O}$ does not cover K. That is, there is a point in K that is in all of the other sets in $\mathscr{C}$.

Example 1.10.1. The collection $\{[n, +\infty) : n \in \mathbb{N}\}$ is a collection of closed subsets of $\mathbb{R}$ such that every finite subcollection has a nonempty intersection, but the collection has empty intersection.

Theorem 1.10.2 (Bolzano-Weierstrass). *Every bounded sequence has a cluster point, and therefore a subsequence converging to that cluster point.*

Proof. Fix $M \in \mathbb{N}$ such that the range of the sequence is in the interval $J = [-M, M]$. If $x \in J$ is not a cluster point of the sequence, then there is an open interval I_x about x such that for only a finite number of values $n \in \mathbb{N}$ is x_n in I_x. Since $\mathbb{N}$ is an infinite set, no finite subcollection of the intervals I_x can cover J. Since J is compact, the intervals I_x cannot cover J, so there must be a cluster point of the sequence in J.

1.11 Continuous Functions

We will speak of an **ε-neighborhood** of a point $x \in \mathbb{R}$ when we mean an open interval $(x-\varepsilon, x+\varepsilon)$ for $\varepsilon > 0$. Given two functions f and g on a common domain A, $(f \vee g)(x) := f(x) \vee g(x)$ and $(f \wedge g)(x) := f(x) \wedge g(x)$ for all $x \in A$. If $f : A \mapsto \mathbb{R}$, $g : B \mapsto \mathbb{R}$, and $g[B] \subseteq A$, then $(f \circ g)(x) := f(g(x))$ for each $x \in B$.

Definition 1.11.1. A real-valued function f defined on a set A is **continuous** at $x \in A$ if for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$f[(x-\delta, x+\delta) \cap A] \subseteq (f(x) - \varepsilon, f(x) + \varepsilon).$$

The function f is continuous on a set $B \subseteq A$ if it is continuous at each point of B. When f is continuous on its domain, we say that f is continuous. A continuous bijection $f : A \mapsto B$ is a **homeomorphism** if the function f^{-1} is continuous on B.

Note that a function can be continuous when restricted to a subset B of its domain A and not be continuous on A. For example, let $A = \mathbb{R}$, $B = \mathbb{N}$, and set $f(x) = 1$ for $x \in \mathbb{N}$ and 0 for $x \in \mathbb{R} \setminus \mathbb{N}$. Of course, if f is continuous at each point of B taking into account its values on A, then it is continuous when restricted to B.

Example 1.11.1. Any continuous real-valued bijection defined on a compact subset of $\mathbb{R}$ is a homeomorphism (Problem 1.36). A counterexample for the case of a noncompact domain is given by the function f with domain $[0,1) \cup [2,3]$ and values given by $f(x) = x$ on $[0,1)$ and $f(x) = x - 1$ on $[2,3]$. The inverse function f^{-1} has a discontinuity at 1.

Proposition 1.11.1. *The family of continuous real-valued functions on a set $A \subseteq \mathbb{R}$ is stable under addition, subtraction, pointwise multiplication, and also division at points where the denominator is not* 0. *If two functions f and g are continuous on A, so are $f \vee g$ and $f \wedge g$. If $f : A \mapsto \mathbb{R}$, $g : B \mapsto \mathbb{R}$, and $g[B] \subseteq A$, then $f \circ g$ is continuous.*

Proof. The proof is left to the reader.

Example 1.11.2. Constant functions and the function $x \mapsto x$ are continuous on any subset of $\mathbb{R}$. Therefore, a polynomial is continuous on any subset of $\mathbb{R}$.

Proposition 1.11.2. *A real-valued function f with domain $A \subseteq \mathbb{R}$ is continuous on A if and only if for every open set $O \subseteq \mathbb{R}$, $f^{-1}[O]$ is the intersection of A with an open set. It follows that if A is open, then f is continuous on A if and only if for each open set $O \subseteq \mathbb{R}$, $f^{-1}[O]$ is open.*

Proof. Assume f is continuous on A, and O is open in $\mathbb{R}$. Then for all $x \in f^{-1}[O]$, we may fix an ε-neighborhood of $f(x)$ contained in O and a corresponding δ-neighborhood U_x of x such that $U_x \cap A$ maps into the ε-neighborhood of $f(x)$ and therefore into O. That is, $f[U_x \cap A] \subseteq O$, whence $U_x \cap A \subseteq f^{-1}[O]$. Let $U = \bigcup_{x \in f^{-1}[O]} U_x$. By definition,

$$f^{-1}[O] \subseteq \bigcup_{x \in f^{-1}[0]} U_x \cap A = U \cap A \subseteq f^{-1}[O],$$

so $f^{-1}[O] = U \cap A$.

Now assume that for each open set $O \subseteq \mathbb{R}$, $f^{-1}[O]$ is the intersection of A with an open set. For each $x \in A$ and each $\varepsilon > 0$, $f^{-1}[(f(x) - \varepsilon, f(x) + \varepsilon)]$ has the form $U \cap A$ for some open U. For some $\delta > 0$, the open interval $(x - \delta, x + \delta) \subseteq U$. It follows that $f[(x - \delta, x + \delta) \cap A] \subseteq f[U \cap A] \subseteq (f(x) - \varepsilon, f(x) + \varepsilon)$.

Theorem 1.11.1. *The continuous image of a compact set is compact. That is, if $A \subset \mathbb{R}$ is compact and contained in the domain of f, then $f[A]$ is compact. Indeed, this is true if just the restriction of f to A is continuous.*

Proof. Let $\{O_\alpha : \alpha \in \mathscr{I}\}$ be an open covering of $f[A]$. For each α, let U_α be an open set such that $f^{-1}[O_\alpha] = U_\alpha \cap A$. The U_α's cover A, so there is a finite subcover $U_{\alpha_1}, \cdots, U_{\alpha_n}$. If $x \in A \cap U_{\alpha_i}$ for some i, then $f(x) \in O_{\alpha_i}$. It follows that the sets $O_{\alpha_1}, \cdots, O_{\alpha_n}$ cover $f[A]$.

Corollary 1.11.1. *A continuous real-valued function f takes a finite maximum and minimum value on any compact set A.*

Proof. Let F be the compact set $f[A]$. Since F is closed and bounded, the least upper bound and the greatest lower bound of F belong to F.

Theorem 1.11.2 (Intermediate Value Theorem). *The continuous image of an interval is an interval. That is, suppose f is continuous when restricted to an interval I in its domain. Also suppose there are points a and b in I with $a < b$ and $f(a) \neq f(b)$. Then for every point w strictly between $f(a)$ and $f(b)$, there is a point $z \in (a,b)$ with $f(z) = w$.*

Proof. We may assume that $f(a) < w < f(b)$; otherwise, we work with $-f$. Let $S = \{z \in [a,b] : f(x) \leq w \text{ for all } x \in [a,z]\}$, and let $c = \sup S$. Since $f(a) < w$, there is a $\gamma > 0$ such that $[a, a+\gamma) \subseteq S$. Therefore, $a < c$. If $f(c) < w$, then $c < b$ and there is an interval $(c - \delta, c + \delta) \subseteq (a,b)$ on which $f(x) < w$; but, this contradicts the definition of c. If $f(c) > w$, then for some $\delta > 0$, $f(x) > w$ for all $x \in (c - \delta, c]$. This again contradicts the definition of c. Therefore, $f(c) = w$.

Definition 1.11.2. A real-valued function f defined on a set $E \subseteq \mathbb{R}$ is **uniformly continuous** on E if for every $\varepsilon > 0$, there is a $\delta > 0$ such that for all points x and y in E with $|x - y| < \delta$, $|f(x) - f(y)| < \varepsilon$.

Example 1.11.3. The functions $f(x) = 1/x$ and $g(x) = x^2$ are not uniformly continuous on $(0,+\infty)$.

Theorem 1.11.3. *A real-valued function continuous on a compact set A is uniformly continuous on A.*

Proof. Fix $\varepsilon > 0$. For each $x \in A$, there is a $\delta_x > 0$ so that if $z \in A$ and $|z-x| < \delta_x$ then $|f(z) - f(x)| < \varepsilon/2$. For each $x \in A$, let $U_x = (x - \delta_x/2, x + \delta_x/2)$. Take a finite subcover $U_{x_1}, \cdots, U_{x_n}$ of the covering of A by the U_x's. Let δ be the smallest of the numbers $\delta_{x_1}/2, \cdots, \delta_{x_n}/2$. Now δ works for ε everywhere on A, for if x and y are in A and $|x-y| < \delta$, then for some i, $|x - x_i| < \delta_{x_i}/2$, and so $|y - x_i| \leq |y-x| + |x - x_i| < \delta_{x_i}$. It follows that

$$|f(x) - f(y)| \leq |f(x) - f(x_i)| + |f(y) - f(x_i)| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Remark 1.11.1. Theorem 1.11.3 is needed in rigorous calculus for the integration on an interval $[a,b]$ of a function that is known to be continuous. The theorem, however, is beyond the ability of most beginning calculus students. Here is an easier, equivalent fact: Fix a continuous function f on an interval $[a,b]$. For each $\Delta x > 0$, divide the interval $[a,b]$ into subintervals $[x_{i-1}, x_i]$, each of which has length Δx except for the last subinterval, which has length at most Δx. The following function of Δx

$$E_f(\Delta x) := \max_i \left(\max_{x_{i-1} \leq x \leq x_i} f(x) - \min_{x_{i-1} \leq x \leq x_i} f(x) \right)$$

has limit 0 as $\Delta x \to 0$. If f has a bounded derivative, this result, call it the maximum change theorem, follows from the mean-value theorem.

Definition 1.11.3. Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of real-valued functions, and let E be a set on which all are defined. Let f be a real-valued function also defined on E. The functions f_n **converge pointwise** to f on E if for every $x \in E$, $f_n(x) \to f(x)$. The functions f_n **converge uniformly** to f on E if for every $\varepsilon > 0$ there is an $m_\varepsilon \in \mathbb{N}$ such that for all $x \in E$ and all $n > m_\varepsilon$, $|f_n(x) - f(x)| < \varepsilon$.

Example 1.11.4. For each $n \in \mathbb{N}$, let $f_n(x) = x^n$. This sequence converges pointwise but not uniformly to 0 on $[0,1)$. The limit is 1 at 1. Clearly, the limit function is not continuous on $[0,1]$. A better result occurs when the convergence is uniform.

Theorem 1.11.4. *On any set $E \subseteq \mathbb{R}$, the uniform limit of continuous functions is continuous.*

Proof. Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of real-valued functions converging uniformly to f on E. Fix $\varepsilon > 0$, and choose n so that $|f_n(x) - f(x)| < \varepsilon/3$ for all $x \in E$. Fix any $x \in E$. Fix $\delta_x > 0$ so that if $z \in E$ and $|z - x| < \delta_x$, then $|f_n(z) - f_n(x)| < \varepsilon/3$, in which case,

$$|f(z) - f(x)| \leq |f(z) - f_n(z)| + |f_n(z) - f_n(x)| + |f_n(x) - f(x)| < \varepsilon.$$

Proposition 1.11.3. *Let F be a closed set in $\mathbb{R}$ and f a continuous function on F. The function f can be extended to all of $\mathbb{R}$ with a continuous function g so that* $\sup_{x\in F} f(x) = \sup_{x\in\mathbb{R}} g(x)$ *and* $\inf_{x\in F} f(x) = \inf_{x\in\mathbb{R}} g(x)$.

Proof. Exercise 1.40(A).

Proposition 1.11.4. *Let f be a continuous function on a closed and bounded interval $[a,b]$ with $b > a$. Given $\varepsilon > 0$, there is a* ***polygonal function*** *g, i.e., a continuous function formed by a finite number of line segments, that is uniformly within ε of f; that is,* $\max_{x\in[a,b]} |f(x) - g(x)| \le \varepsilon$.

Proof. Exercise 1.41(A).

Note that $\cap_{n\in\mathbb{N}}(-\frac{1}{n}, \frac{1}{n})$ is the singleton set $\{0\}$, which is not open. Moreover, $\cup_{n\in\mathbb{N}}\left[-1+\frac{1}{n}, 1-\frac{1}{n}\right]$ is the open interval $(-1,1)$, which is not closed. We will, however, have need to consider countable operations acting on open and closed sets even without stability.

Definition 1.11.4. A G_δ set is a set that is the countable intersection of open sets. An F_σ set is a set that is the countable union of closed sets. A $G_{\delta\sigma}$ set is a countable union of G_δ sets. An $F_{\sigma\delta}$ set is a countable intersection of F_σ sets, etc.

Proofs for the following propositions are left as exercises.

Proposition 1.11.5. *The set of points of continuity of a real-valued function defined on the real line $\mathbb{R}$ is a G_δ set.*

Proposition 1.11.6. *If $\langle f_n : n \in \mathbb{N}\rangle$ is a sequence of continuous functions defined on $\mathbb{R}$, then the set consisting of points where this sequence converges is an $F_{\sigma\delta}$ set.*

1.12 Problems

Problem 1.1. Show that for an equivalence relation ρ on a set A and points $x, y \in A$, either the equivalence class $[x]$ equals the equivalence class $[y]$ or $[x]$ and $[y]$ are disjoint.

Problem 1.2. Prove that a countable union of countable sets is countable.

Problem 1.3. Prove that the set of rational numbers $\mathbb{Q}$ is countable.

Problem 1.4. Prove De Morgan's laws for two sets A and B.

Problem 1.5. Let E be the set of functions from the natural numbers $\mathbb{N}$ into the two point set $\{0,1\}$. Show that E is not a countable set.

Problem 1.6. Show that every open set in $\mathbb{R}$ either has no elements or is uncountable. You may use the fact that $[0,1]$ is uncountable.

Problem 1.7. Show that an algebra $\mathscr{A}$ is a σ-algebra if and only if $\mathscr{A}$ is stable under countable increasing unions. That is, if $\langle E_i : i \in \mathbb{N} \rangle$ is a sequence in $\mathscr{A}$ such that for all i, $E_i \subseteq E_{i+1}$, then $\cup_{i=1}^{\infty} E_i$ is in $\mathscr{A}$.

Problem 1.8 (A). Prove Proposition 1.3.2. **Hint:** Set $\mathscr{A} := \sigma(\mathscr{C})$; let $\mathscr{A}'$ be the collection of all sets $E \in \mathscr{A}$ such that the result is true for E (that is, for some countable collection $\mathscr{C}_0 \subseteq \mathscr{C}$, $E \in \sigma(\mathscr{C}_0)$.) You want to show that $\mathscr{A}'$ is a σ-algebra and $\mathscr{C} \subseteq \mathscr{A}'$. It will then follow by definition that $\mathscr{A}' = \mathscr{A} = \sigma(\mathscr{C})$.

Problem 1.9. Recall that the Borel σ-algebra $\mathscr{B}$ of $\mathbb{R}$ is the smallest σ-algebra containing the open subsets of $\mathbb{R}$. Show that $\mathscr{B}$ is also the smallest σ-algebra containing all intervals of the form $[\alpha, \beta)$, with α and β rational and $\alpha < \beta$.

Problem 1.10. Show that the relation $\subseteq$ in the power set of $\mathbb{R}$ is an antisymmetric partial ordering but not a total ordering.

Problem 1.11. Prove that a convergent sequence in $\mathbb{R}$ is a Cauchy sequence.

Problem 1.12. Let $\langle x_n \rangle$ be a sequence in $\mathbb{R}$. Suppose that for every subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$, there is a further subsequence $\langle x_{n_{k_\ell}} \rangle$ of $\langle x_{n_k} \rangle$ such that $x_{n_{k_\ell}} \to x$ as $\ell \to \infty$. Show that $x_n \to x$ as $n \to \infty$.

Problem 1.13. Show that if a real-valued sequence $\langle x_n \rangle$ has neither $+\infty$ nor $-\infty$ as a cluster point, then for some $M \in \mathbb{N}$, there is a $K \in \mathbb{N}$ such that for all $n > K$, $x_n \in [-M, M]$.

Problem 1.14. a) Give an example of a sequence in $\mathbb{R}$ for which the set of cluster points consists of the natural numbers.

b) Can a sequence in $\mathbb{R}$ have uncountably many cluster points?

Problem 1.15. Let A be a nonempty set with a finite upper bound. For each $n \in \mathbb{N}$, you are given $a_n \in A$ and an upper bound b_n of the set A such that $b_n - a_n \leq 1/n$. Show that the b_n's form a Cauchy sequence, and the limit is the least upper bound of A.

Problem 1.16. Let $I_n = [a_n, b_n]$ be a decreasing sequence of finite closed intervals in $\mathbb{R}$. That is, for each $n \in \mathbb{N}$, $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$. Assume that $\lim_{n\to\infty}(b_n - a_n) = 0$. Use Cauchy sequences directly to show that there is a unique point x_0 in $\cap_n I_n$.

Problem 1.17. Prove Proposition 1.6.1.

Problem 1.18. Finish the proof of Proposition 1.6.2.

Problem 1.19. Construct a bounded real-valued sequence $\langle x_n \rangle$ and a bounded real-valued sequence $\langle y_n \rangle$ that satisfy:

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

In particular, equality does not hold in general in any of the inequalities in Proposition 1.6.2.

Problem 1.20. Let x be an accumulation point of a set $S \subseteq \mathbb{R}$. Show that every open interval centered at x contains infinitely many points of S.

Problem 1.21. Given $n \in \mathbb{N}$, find an example of an open set $E \subseteq \mathbb{R}$ with exactly n accumulation points in $\mathbb{R}$ that do not lie in E. Can you find such an example for some $n \in \mathbb{N}$ when E is assumed to be closed?

Problem 1.22. **a)** Let U and V be disjoint open sets in $\mathbb{R}$ with union equal to $\mathbb{R}$. Show that either U or V is empty.
b) Show that the only subsets of $\mathbb{R}$ that are both open and closed are the empty set and $\mathbb{R}$ itself.

Problem 1.23. Prove Proposition 1.9.1. **Hint:** If the unordered sum S of the elements of E is finite, how many elements of E can be larger than $m \in \mathbb{N}$?

Problem 1.24. Let $\langle x_n : n \in \mathbb{N} \rangle$ be a sequence on $\mathbb{R}$ such that $x_n \to x_0$. Show that the set $\bigcup_{n=0}^{\infty} \{x_n\}$ is compact.

Problem 1.25. Let $K \subset \mathbb{R}$ be compact, and fix $x \in K$. Let $\langle x_n \rangle$ be a sequence of points in K such that every subsequence of $\langle x_n \rangle$ that converges to some point in $\mathbb{R}$ actually converges to x. Prove that the sequence $\langle x_n \rangle$ converges to x. If K is not compact, is the conclusion still true?

Problem 1.26. Show that a sequence $\langle x_n : n \in \mathbb{N} \rangle$ in a compact set $K \subseteq \mathbb{R}$ has a cluster point in K by considering the intersection of the sets C_n, where for each n, C_n is the closure of the tail sequence $\{x_i : i \geq n\}$.

Problem 1.27. Let K be a compact subset of $\mathbb{R}$. Show that there is a countable dense subset D of K; that is, the closure $\overline{D} = K$.

Problem 1.28. Prove Proposition 1.11.1.

Problem 1.29. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a continuous function such that for each open set $O \subseteq \mathbb{R}$, $f[O]$ is open in $\mathbb{R}$. Show that f is either an increasing or a decreasing function. That is, either it is the case that for all $x < y$ in $\mathbb{R}$, $f(x) \leq f(y)$, or it is the case that for all $x < y$ in $\mathbb{R}$, $f(x) \geq f(y)$.

Problem 1.30. Let $K \subset \mathbb{R}$ be a compact set. Suppose that $T : K \mapsto K$ satisfies $|T(x) - T(y)| < |x - y|$ for all points x and y with $x \neq y$ in K. Show that there exists a unique $x_0 \in K$ such that $T(x_0) = x_0$. **Hint:** Show that the function $x \mapsto |x - T(x)|$ is continuous on K, and so achieves its minimum value.

Problem 1.31. Show that the function f given by $f(x) = 1/x$ is not uniformly continuous on $(0,1)$.

Problem 1.32. Let $f : [0,\infty) \mapsto \mathbb{R}$ be a continuous function such that $\lim_{x\to\infty} f(x)$ exists and is finite. Show that f is uniformly continuous on $[0,\infty)$.

Problem 1.33. Prove that on a set $E \subseteq \mathbb{R}$, the uniform limit of uniformly continuous functions is uniformly continuous.

Problem 1.34. Let $\langle f_n \rangle$ be a sequence of differentiable functions on $\mathbb{R}$ with uniformly bounded derivatives, i.e., $|f_n'(x)| \leq M$ for every $n \in \mathbb{N}$ and $x \in \mathbb{R}$. Suppose $f_n(x)$ converges pointwise to $f(x)$ for all $x \in \mathbb{R}$. Show that f is continuous. **Hint:** You may use the mean-value theorem.

Problem 1.35. Let K be a subset of $\mathbb{R}$ such that every continuous real-valued function on K is bounded. Show that K is compact.

Problem 1.36. **a)** Show that a continuous real-valued bijection $f : A \mapsto B$ is a homeomorphism if for any open set $U \subseteq \mathbb{R}$, $f[U \cap A] = V \cap B$ for some open set $V \subseteq \mathbb{R}$.
b) Show that any continuous bijection f mapping a compact domain $K \subseteq \mathbb{R}$ onto $S \subseteq \mathbb{R}$ is a homeomorphism. **Hint:** Given an open set $U \subseteq \mathbb{R}$, what can you say about $K \backslash U$?

Problem 1.37 (A). Let $K \subset \mathbb{R}$ be a compact set, and let $f : K \mapsto K$ have the property that $|f(x) - f(y)| = |x - y|$ for all $x, y \in K$. Show that f is a bijection.

Problem 1.38. Show that a continuous real-valued function f defined on an interval $I \subseteq \mathbb{R}$ such that the derivative f' exists and is bounded on I is uniformly continuous on I.

Problem 1.39. Recall that the uniform limit of continuous functions is continuous.

a) Show that if $\langle f_n : n \in \mathbb{N} \rangle$ is an increasing sequence of continuous functions converging pointwise to a continuous function f, i.e., $f_n(x) \nearrow f(x)$, on a **compact** set $K \subseteq \mathbb{R}$, then the convergence is uniform (This is **Dini's theorem**).
b) Show by example that the result need not be true if the limit f is not continuous.
c) Show by example that the result need not be true if the domain K is not compact.

Problem 1.40 (A).

(a) Let F be a closed set in $\mathbb{R}$ and f a continuous function on F. Show that one can extend f to all of $\mathbb{R}$ as a continuous function g so that $\sup_{x \in F} f(x) = \sup_{x \in \mathbb{R}} g(x)$ and $\inf_{x \in F} f(x) = \inf_{x \in \mathbb{R}} g(x)$.
(b) Give an example showing that the result is false if F is replaced with an open subset of $\mathbb{R}$.

Problem 1.41 (A). Let f be a continuous function on a closed and bounded interval $[a, b]$ with $b > a$. Given $\varepsilon > 0$, show that there is a polygonal function g such that $|f(x) - g(x)| < \varepsilon$ for all $x \in [a, b]$.

Problem 1.42 (A). Recall that a continuous bijection $f : A \mapsto B$ is a homeomorphism if the function f^{-1} is continuous on B.

(a) Show that if $A \subset \mathbb{R}$ is compact and $f : A \mapsto B$ is a real-valued homeomorphism, then f and f^{-1} are actually uniformly continuous.
(b) Give an example of a closed set $K \subseteq \mathbb{R}$ and a real-valued homeomorphism $f : K \mapsto B$ such that f is uniformly continuous, but f^{-1} is not uniformly continuous.

(c) Give an example of a bounded set $E \subset \mathbb{R}$ and a real-valued homeomorphism $f : E \mapsto B$ such that f^{-1} is uniformly continuous, but f is not uniformly continuous.

Problem 1.43 (A). Show that the set P_c of points of continuity of a real-valued function f defined on the real line $\mathbb{R}$ is a G_δ set.

Problem 1.44 (A). Given a sequence $\langle f_i \rangle$ of continuous functions defined on $\mathbb{R}$, show that the set C consisting of points where this sequence converges is an $F_{\sigma\delta}$.

Chapter 2
Measure on the Real Line

2.1 Introduction

There are many examples of functions that associate a nonnegative real number or $+\infty$ with a set. There is, for example, the number of members forming the set. Given a finite probability experiment, probabilities are associated with outcomes. Riemann integration associates with each finite interval in the real line, the length of that interval. These are all examples of a "finitely additive measure." Recall that an **algebra** $\mathscr{A}$ of subsets of a set X is a collection that contains the set X together with the complement in X of each of its members; it is also stable under the operation of taking finite unions and, therefore, finite intersections. Also recall that a collection of sets is pairwise disjoint if for any two sets A and B in the collection, $A \cap B = \varnothing$.

Definition 2.1.1. A **finitely additive measure** m is a function from an algebra $\mathscr{A}$ of subsets of a set X into the extended nonnegative real line, $\mathbb{R} \cup \{+\infty\}$, such that $m(\varnothing) = 0$ and for any finite collection $\{A_i : i = 1, 2, \ldots, n\}$ of pairwise disjoint sets in $\mathscr{A}$,

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

Such a function m is **countably additive** if for any pairwise disjoint sequence $\{A_i : i \in \mathbb{N}\}$ in $\mathscr{A}$ with union also in $\mathscr{A}$,

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

Remark 2.1.1. If the summation condition for countable additivity holds and $m(\varnothing) = 0$, then the summation condition for finite additivity also holds.

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_2

Definition 2.1.2. Given a set $A \subseteq \mathbb{R}$ and $r \in \mathbb{R}$, the **translate** of A by r, denoted by $A+r$, is the set $\{a+r : a \in A\}$. A finitely additive measure m defined on an algebra $\mathscr{A}$ of subsets of $\mathbb{R}$ is **translation invariant** if for each $A \in \mathscr{A}$ and each $r \in \mathbb{R}$, $A+r$ is in $\mathscr{A}$ and

$$m(A+r) = m(A).$$

The translation invariant, finitely additive measure m that associates to each subinterval of $\mathbb{R}$ the length of that interval is defined on the algebra consisting of all finite unions of subintervals of $\mathbb{R}$. Any such union can be written as a finite union of pairwise disjoint intervals. The sum of the lengths of those intervals is the value, independent of the decomposition, that is taken by m. We want to extend Riemann integration. We need, therefore, to extend the function m to a larger class of sets. We would like the extension to be countably additive and translation invariant. It turns out that an extension with these properties cannot be defined for all subsets of $\mathbb{R}$. There is, however, an important translation invariant extension that is defined for all subsets of $\mathbb{R}$.

2.2 Lebesgue Outer Measure

For each interval $I \subseteq \mathbb{R}$, we write $l(I)$ for the length of I. For example, if $I = (a,b)$, then $l(I) = b-a$. If I is an infinitely long interval, then $l(I) = +\infty$. Given a set $A \subseteq \mathbb{R}$, we let $\mathscr{C}(A)$ denote the family of all collections of open intervals such that the intervals in the collection cover A. That is, $\mathscr{I}$ is a member of $\mathscr{C}(A)$ if and only if $\mathscr{I}$ is a set of open intervals in $\mathbb{R}$ and the union of the intervals in $\mathscr{I}$ contains the set A. By $\sum_{I\in\mathscr{I}} l(I)$ we mean the unordered sum of the length of the intervals in $\mathscr{I}$. Recall that this is the supremum of the sums obtained by adding the length of intervals in finite subsets of $\mathscr{I}$. If $\mathscr{I}$ is an uncountable collection of intervals, then by the Lindelöf theorem, a finite or countably infinite subfamily of $\mathscr{I}$ also covers A and has a sum of lengths that is no greater than the sum for the whole family. Therefore, in applying the following definition, we usually consider just finite and countably infinite families of open intervals that cover A. Every enumeration of a countably infinite family of intervals will produce the same sum of lengths, which is the usual limit of partial sums.

Definition 2.2.1 (Lebesgue outer measure). For each subset $A \subseteq \mathbb{R}$, the **Lebesgue outer measure**, denoted by $\lambda^*(A)$, is obtained as follows:

$$\lambda^*(A) = \inf_{\mathscr{I}\in\mathscr{C}(\mathscr{A})} \left(\sum_{I\in\mathscr{I}} l(I) \right).$$

Lebesgue outer measure is defined on the power set of $\mathbb{R}$, that is, the algebra comprised of all subsets of $\mathbb{R}$. We will show that Lebesgue outer measure is translation invariant and extends the notion of interval length. To obtain finite additivity, however, we will need to restrict λ^* to a proper subfamily of the algebra of all subsets of $\mathbb{R}$.

Proposition 2.2.1. *For each* $A \subseteq \mathbb{R}$, $\lambda^*(A) \geq 0$, $\lambda^*(\varnothing) = 0$, $\lambda^*(\mathbb{R}) = +\infty$, *and if* $A \subseteq B \subseteq \mathbb{R}$, *then* $\lambda^*(A) \leq \lambda^*(B)$.

Proof. Since every open interval contains the empty set, $\lambda^*(\varnothing) = 0$. The rest is clear from the definition.

Theorem 2.2.1. *Lebesgue outer measure is translation invariant. That is, for any* $A \subseteq \mathbb{R}$ *and each* $r \in \mathbb{R}$, $\lambda^*(A + r) = \lambda^*(A)$.

Proof. Exercise 2.4(A).

Definition 2.2.2. Given a closed and bounded interval $[a,b]$ with $a < b$, let $\mathscr{BP}[a,b]$ be the sequence of **bisection partitions** $\langle P_n : n \in \mathbb{N}\rangle$ of $[a,b]$. That is, P_1 is the pair $\{[a, a + \frac{b-a}{2}], [a + \frac{b-a}{2}, b]\}$, and for each $n \in \mathbb{N}$, P_{n+1} is the set of closed intervals obtained by cutting each interval in P_n in half, thus forming closed intervals of length $(b-a)/2^{n+1}$.

Proposition 2.2.2. *Fix an interval* $[a,b]$ *and a finite collection of open intervals* $\mathscr{I} = \{(a_k, b_k) : k = 1, \cdots, k_0\}$ *covering* $[a,b]$. *There is a* $j \in \mathbb{N}$ *such that every interval in the bisection partition* $P_j \in \mathscr{BP}[a,b]$ *is contained in at least one of the open intervals* (a_k, b_k) *from* $\mathscr{I}$.

Proof. Since a is contained in an open interval from $\mathscr{I}$, there is a first $m \in \mathbb{N}$ such that $[a, a + \frac{b-a}{2^m}]$ is contained in an open interval from $\mathscr{I}$. For any $n < m$, let $x_n = a$. For each $n > m$, let x_n be the largest right endpoint of the intervals in P_n such that each of the intervals in P_n below x_n is contained in an open interval from $\mathscr{I}$. The increasing sequence $\langle x_n\rangle$ has a limit x_0 in $[a,b]$. Since x_0 is contained in an open interval from $\mathscr{I}$, that limit is b, and $b = x_j$ for some $j \in \mathbb{N}$.

Theorem 2.2.2. *The Lebesgue outer measure of an interval is its length.*

Proof. For any $x \in \mathbb{R}$, $\lambda^*([x,x]) = \lambda^*(\{x\}) = 0$. Now assume the interval is $[a,b]$ with $a < b$. For each $\varepsilon > 0$, $[a,b] \subset (a - \varepsilon, b + \varepsilon)$, so $\lambda^*([a,b]) \leq b - a + 2\varepsilon$, and since ε is arbitrary, $\lambda^*([a,b]) \leq l([a,b])$. Note that this proof works for any finite interval. To show the reverse inequality, we must show that whatever the finite or countably infinite covering of $[a,b]$ by open intervals, the sum of their lengths is no less than $b - a$. Fix such a covering, and let $\{(a_k, b_k) : k = 1, \cdots, n_0\}$ be a finite subcovering. We need only show that $\sum_{k=1}^{n_0} b_k - a_k \geq b - a$. By Proposition 2.2.2, we may fix a bisection partition P_n of $[a,b]$ so that each member of P_n, which is a subinterval of $[a,b]$ of length $(b-a)/2^n$, is contained in at least one of the open intervals (a_k, b_k). For each $k \leq n_0$, $b_k - a_k$ is greater than the sum of the lengths of the closed intervals from P_n that are contained in (a_k, b_k). Since $b - a$ is the sum of the lengths of the intervals in P_n and every one of those intervals is in at least one of the intervals (a_k, b_k), it follows that $\sum_{k=1}^{n_0} b_k - a_k \geq b - a$.

We have already shown that for an arbitrary, not necessarily closed, finite interval I of positive length, the Lebesgue outer measure of I is less than or equal to the length of I. On the other hand, the length is less than or equal to the outer measure since there are closed intervals $J_n \subset I$ with $l(J_n) \uparrow l(I)$. That is, for each

$\varepsilon > 0$, there is an $n \in \mathbb{N}$ so that $l(I) - \varepsilon \leq l(J_n) = \lambda^*(J_n) \leq \lambda^*(I)$. It follows that $\lambda^*(I) = l(I)$. Finally, an infinite interval contains arbitrarily large closed subintervals, so the outer measure of an infinite interval is $+\infty$.

Theorem 2.2.3. *Lebesgue outer measure is finitely and countably subadditive. That is, for any finite or infinite sequence $\langle A_n \rangle$ of subsets of $\mathbb{R}$,*

$$\lambda^*\left(\bigcup_n A_n\right) \leq \sum_n \lambda^*(A_n).$$

Proof. If for some n we have $\lambda^*(A_n) = +\infty$, then the inequality is clear. If not, we fix $\varepsilon > 0$ and for each n find a countable family of intervals covering A_n with the sum of the length of those intervals less than $\lambda^*(A_n) + \varepsilon/2^n$. The union of these families of intervals forms a countable interval covering of $\cup A_n$, and the sum of the lengths is less than $\sum_n \lambda^*(A_n) + \varepsilon$. Since ε is arbitrary, the result follows.

Corollary 2.2.1. *A countable set has Lebesgue outer measure* 0.

Corollary 2.2.2. *Any interval of positive length is uncountable.*

Example 2.2.1. The set of integers has Lebesgue outer measure 0, and the set of rational numbers has Lebesgue outer measure 0.

Recall that a G_δ set is a set that is the countable intersection of open sets. An F_σ set is a set that is the countable union of closed sets. A $G_{\delta\sigma}$ set is a countable union of G_δ sets. An $F_{\sigma\delta}$ set is a countable intersection of F_σ sets, etc.

Proposition 2.2.3. *Given $A \subseteq \mathbb{R}$ and $\varepsilon > 0$, there is an open set O with $A \subseteq O$ and $\lambda^*(O) \leq \lambda^*(A) + \varepsilon$. Moreover, there is a G_δ set $S \supseteq A$ with $\lambda^*(S) = \lambda^*(A)$.*

Proof. The first and second part are clear if $\lambda^*(A) = +\infty$. For example, let $O = S = \mathbb{R}$. Otherwise, take open intervals that cover A with total length at most ε, and let O be the union. For the second part, let O_n be an open set given in the first part that works when $\varepsilon = 1/n$. Now the desired set is $S = \bigcap_n O_n$.

To obtain Lebesgue measure, we will restrict λ^* to a family of sets on which it is finitely additive. The restriction will then, in fact, be countably additive. We will call the reduced family of sets "the Lebesgue measurable sets", and the restriction of λ^* will be Lebesgue measure λ.

2.3 General Outer Measures

Lebesgue outer measure generalizes the length of finite open intervals. The length of a finite interval is the change on the interval of the function $F(x) = x$. More general outer measures are constructed using the changes of more general increasing functions. Such functions will have discontinuities at points where the limit from the right and the limit from the left are not equal. Our more general outer measures will be formed using increasing functions, called "integrators", that are continuous from the right.

Definition 2.3.1. An increasing real-valued function F is an **integrator** if for each x in the domain of F, $F(x) = \lim_{y \to x+} F(y)$.

We are only interested in the changes of an integrator, so when we restrict work to a finite interval in $\mathbb{R}$ on which the integrator is bounded below, we may add a constant so that the integrator is nonnegative. The integral that results from general integrators relates to what is called "the Riemann–Stieltjes integral" in the same way that the Lebesgue integral relates to the Riemann integral. This generalization is very important in probability theory. It will cost us essentially nothing to work with results for which a more general integrator can be used. This generalization of the approach to Lebesgue integration also simplifies later material on measure differentiation. It will be clear which results hold only for Lebesgue outer measure and the corresponding Lebesgue measure.

As noted, the construction of Lebesgue outer measure employs the change of the integrator $F(x) = x$ on open intervals. For a general integrator F, however, we use the change $F(b) - F(a)$ on intervals of the form $(a,b]$. In this way, the value of any jump of F is associated with the interval on which it occurs. If we already have a measure taking only finite values, then we may set $F(x)$ equal to the measure of $(-\infty, x]$. If F is only defined on a finite interval $[a,b]$, then we can extend F with the value $F(a)$ to points below a and $F(b)$ to points above b. Then the change of F will be 0 on any interval that does not intersect $[a,b]$. If an integrator is continuous, such as the integrator $F(x) = x$ for Lebesgue outer measure, then the same outer measure is obtained using open intervals or intervals of the form $(a,b]$ (Exercise 2.13). We have shown in Corollary 1.7.1 that any collection of intervals of the form $(a,b]$ has a finite or countably infinite subcollection with the same union.

Definition 2.3.2. Let F be an integrator, that is, an increasing real-valued function, continuous from the right at each point of $\mathbb{R}$. For each subset $A \subseteq \mathbb{R}$, let $m^*(A)$ be defined in a way similar to Lebesgue outer measure, but using finite intervals of the form $(a,b]$ and the change $F(b) - F(a)$.

When we used length, we used compactness and open coverings to show that the outer measure of an interval is its length. The analogous result for a general integrator F is still true.

Proposition 2.3.1. *Let F be an integrator on $\mathbb{R}$. Then $m^*(\varnothing) = 0$. If $A \subseteq B \subseteq \mathbb{R}$, then $m^*(A) \le m^*(B)$, and for any interval $(a,b]$, $m^*((a,b]) = F(b) - F(a) < +\infty$.*

Proof. Every interval $(a,b]$ contains the empty set, and for every $\varepsilon > 0$ there is such an interval for which $F(b) - F(a) < \varepsilon$ (Problem 2.14). Therefore, $m^*(\varnothing) = 0$. It is clear that a more general outer measure is still an increasing function; that is, the bigger the set, the bigger the outer measure. It is also clear that $m^*((a,b]) \le F(b) - F(a)$

since $(a,b]$ covers itself. To show the reverse inequality for F, we fix $\varepsilon > 0$. Fix a countable covering of $(a,b]$ by intervals of the form $(c_n,d_n]$. Since F is continuous from the right, we may replace each interval $(c_n,d_n]$ with an open interval (c_n,e_n) where $e_n > d_n$, but $F(e_n) - F(c_n) \leq F(d_n) - F(c_n) + \varepsilon/2^{n+1}$. Fix δ with $0 < \delta < b-a$ and $F(a+\delta) < F(a) + \varepsilon/2$. The intervals (c_n,e_n) form an open interval covering of $[a+\delta,b]$, and so we may assume it is a finite covering of that interval. By Proposition 2.2.2, we may fix a bisection partition P_n of $[a+\delta,b]$ so that each member of P_n, which is a subinterval of $[a+\delta,b]$ of length $(b-a-\delta)/2^n$, is contained in at least one of the open intervals (c_k,e_k). For each of the intervals (c_k,e_k), $F(e_k) - F(c_k)$ is greater than or equal to the sum of the changes of F on the closed intervals of P_n contained in (c_k,e_k). Moreover, $F(b) - F(a+\delta)$ is equal to the sum of the changes of F on the intervals of P_n. Since each interval of P_n is contained in at least one of the intervals (c_k,e_k), it follows that

$$\begin{aligned}&\sum_{k=1}^{\infty}[F(d_k)-F(c_k)]+\frac{\varepsilon}{2}\\ &\geq \sum_{k=1}^{n}F(e_k)-F(c_k) \geq F(b)-F(a+\delta) \geq F(b)-F(a)-\frac{\varepsilon}{2}.\end{aligned}$$

Since ε is arbitrary, the result follows.

Remark 2.3.1. It is no longer necessarily true for a more general integrator F that points have 0 outer measure. If F jumps at a point x, then the outer measure of $\{x\}$ is the size of the jump.

Example 2.3.1. If we defined m^* using open intervals, then it would no longer be always true that the outer measure of an open interval would equal the change of the integrator at the endpoints of the interval. For example, if $F(x) = 0$ for $x < 1$ and $F(x) = 1$ for $x \geq 1$, then the change of F for $(0,1)$ is 1, but the outer measure using countable coverings by small open intervals would be 0.

Remark 2.3.2. In what follows, results and proofs that hold for general integrators will be stated using m^* and m for the corresponding outer measure and measure. We will use λ^* and λ when the result is special for the Lebesgue case. Since most results use only the common properties of outer measures, in only a few instances, such as Proposition 2.3.1 above, is there a difference in wording of proofs for the Lebesgue and the general case. For the next result, already established for the Lebesgue case, one can also use the fact that there is an integrator $F(x) = x$ for the Lebesgue case.

Theorem 2.3.1. *Outer measure is finitely and countably subadditive. That is, for any sequence $\langle A_n : n \in \mathbb{N}\rangle$ of subsets of $\mathbb{R}$, where some sets may be empty,*

$$m^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} m^*(A_n).$$

Proof. If for some n, $m^*(A_n) = +\infty$, then the inequality is clear. If not, fix $\varepsilon > 0$, and for each $n \in \mathbb{N}$ find a countable family $\mathscr{F}_n$ of appropriate intervals covering A_n

such that the sum of the changes of the integrator F is less than $m^*(A_n) + \varepsilon/2^n$. The union $\cup_{n\in\mathbb{N}}\mathscr{F}_n$ is a countable interval covering of $\cup_{n\in\mathbb{N}}A_n$ such that the sum of the changes of F is less than $\sum_{n\in\mathbb{N}} m^*(A_n) + \varepsilon$. Since ε is arbitrary, the result follows.

2.4 Measure from Outer Measure

As we shall see, Lebesgue outer measure is not even finitely additive on the family of all subsets of $\mathbb{R}$. There is, however, a finite additivity condition that yields not just finite additivity, but also countable additivity on an appropriate family of subsets of $\mathbb{R}$. It is a condition, due to Carathéodory, that is applicable to all outer measures.

Definition 2.4.1 (Carathéodory). A set $E \subseteq \mathbb{R}$ is called **measurable** if for all subsets $A \subseteq \mathbb{R}$,

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

We denote the family of measurable sets by $\mathscr{M}$. If the outer measure extends length, we may say "Lebesgue measurable."

The idea is that a set E is in $\mathscr{M}$ if and only if E splits any set in an additive fashion. Since outer measure is subadditive, we always have

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

We also have the reverse inequality if $m^*(A) = +\infty$. Therefore, to show E is measurable, we need only show that for any set $A \subseteq \mathbb{R}$ with $m^*(A) < +\infty$,

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

Proposition 2.4.1. *Any set of outer measure* 0 *is measurable.*

Proof. If $m^*(E) = 0$, then for any $A \subseteq \mathbb{R}$,

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

since $m^*(A \cap E) \leq m^*(E) = 0$.

Lemma 2.4.1. *The family $\mathscr{M}$ of measurable sets is an algebra of sets.*

Proof. By symmetry, a set E is in $\mathscr{M}$ if and only if the complement $\widetilde{E} = \complement E$ is in $\mathscr{M}$. Moreover, $\mathbb{R}$ and $\varnothing$ are clearly measurable. We need to show that $\mathscr{M}$ is stable under the operation of taking finite unions. For this we need only consider two measurable sets E_1 and E_2. Fix $A \subseteq \mathbb{R}$. We will use the fact that since E_1 and E_2 are measurable,

$$\begin{aligned} m^*(A) &= m^*(A \cap E_1) + m^*(A \cap \widetilde{E}_1), \\ m^*(A \cap \widetilde{E}_1) &= m^*(A \cap \widetilde{E}_1 \cap E_2) + m^*(A \cap \widetilde{E}_1 \cap \widetilde{E}_2). \end{aligned}$$

We will also use the following consequence of subadditivity:

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

Now,

$$\begin{aligned}m^*(A) &\le m^*(A\cap[E_1\cup E_2])+m^*(A\cap\complement[E_1\cup E_2])\\ &= m^*(A\cap[E_1\cup E_2])+m^*(A\cap[\widetilde{E}_1\cap\widetilde{E}_2])\\ &\le m^*(A\cap E_1)+m^*(A\cap E_2\cap\widetilde{E}_1)+m^*(A\cap[\widetilde{E}_1\cap\widetilde{E}_2])\\ &= m^*(A\cap E_1)+m^*(A\cap\widetilde{E}_1)=m^*(A).\end{aligned}$$

Therefore, $E_1\cup E_2\in\mathscr{M}$.

Lemma 2.4.2. *For any finite, pairwise disjoint sequence of measurable sets E_i, $1\le i\le n$, and any $A\subseteq\mathbb{R}$,*

$$m^*(A\cap[\cup_1^n E_i])=\sum_1^n m^*(A\cap E_i).$$

Proof. The proof is by induction. The equality is clear for $n=1$. Assuming it holds for $n-1$, that is,

$$\sum_{i=1}^{n-1}m^*(A\cap E_i)=m^*(A\cap[\cup_1^{n-1}E_i])=m^*(A\cap[\cup_1^n E_i]\cap\widetilde{E}_n),$$

we also have

$$m^*(A\cap E_n)=m^*(A\cap[\cup_1^n E_i]\cap E_n).$$

Therefore, equality holds for n since E_n is measurable and

$$\begin{aligned}\sum_1^n m^*(A\cap E_i) &= m^*(A\cap[\cup_1^n E_i]\cap\widetilde{E}_n)+m^*(A\cap[\cup_1^n E_i]\cap E_n)\\ &= m^*(A\cap[\cup_1^n E_i]).\end{aligned}$$

Recall that an algebra of sets is called a σ-algebra if it is stable with respect to the operation of taking countable unions.

Definition 2.4.2. A nonnegative function μ defined on a σ-algebra $\mathscr{A}$ is a **measure** on $\mathscr{A}$ if $\mu(\varnothing)=0$ and μ is **countably additive**; that is, given a countable, pairwise disjoint sequence $\langle A_n : n\in\mathbb{N}\rangle$ of sets in $\mathscr{A}$, where some sets may be empty,

$$\mu\left(\bigcup_{n\in\mathbb{N}}A_n\right)=\sum_{n\in\mathbb{N}}\mu(A_n).$$

The pair $(\mathscr{A},\mu)$ is called a **measure space**.

Theorem 2.4.1. *The family $\mathscr{M}$ is a σ-algebra containing all sets of outer measure* 0*, and the restriction of m^* to $\mathscr{M}$ is a measure on $\mathscr{M}$.*

Proof. We have already noted that $\mathscr{M}$ contains all sets of outer measure 0. Let B_i, $i \in \mathbb{N}$, be a countable family of sets in $\mathscr{M}$, and let E be the union. We must show that $E \in \mathscr{M}$. Since $\mathscr{M}$ is an algebra, it follows from Proposition 1.3.1 that we may replace each set B_i with a subset $E_i \in \mathscr{M}$ so that the E_i's are pairwise disjoint but have the same union E. For each finite n, let $F_n = \bigcup_{i=1}^n E_i$. Then, because $\mathscr{M}$ is an algebra and $\widetilde{F_n} \supseteq \widetilde{E}$, for each $A \subseteq \mathbb{R}$ we have

$$\begin{aligned} m^*(A) &= m^*(A \cap F_n) + m^*(A \cap \widetilde{F_n}) \\ &= \sum_{i=1}^n m^*(A \cap E_i) + m^*(A \cap \widetilde{F_n}) \\ &\geq \sum_{i=1}^n m^*(A \cap E_i) + m^*(A \cap \widetilde{E}). \end{aligned}$$

Since this is true for all $n \in \mathbb{N}$, we have by subadditivity

$$\begin{aligned} m^*(A) &\geq \sum_{i=1}^\infty m^*(A \cap E_i) + m^*(A \cap \widetilde{E}) \qquad (2.4.1) \\ &\geq m^*(A \cap E) + m^*(A \cap \widetilde{E}). \end{aligned}$$

Therefore, $E \in \mathscr{M}$. It now follows from Inequality (2.4.1) applied to any pairwise disjoint sequence $\langle E_i : i \in \mathbb{N} \rangle$ in $\mathscr{M}$ and the set $A = E = \cup_{i=1}^\infty E_i$ that the restriction of m^* to $\mathscr{M}$ is countably additive.

Definition 2.4.3. Lebesgue measure λ is λ^* restricted to the σ-algebra $\mathscr{M}$ of sets measurable with respect to λ^*. For a general outer measure m^*, including Lebesgue outer measure, we let m denote the measure obtained by restricting m^* to the corresponding collection of measurable sets.

Recall that the intersection of all σ-algebras in $\mathbb{R}$ containing a family of sets $\mathscr{S}$ is again a σ-algebra; it is the smallest σ-algebra in $\mathbb{R}$ containing the family $\mathscr{S}$.

Definition 2.4.4. The family of **Borel sets** in $\mathbb{R}$ is the smallest σ-algebra containing the open subsets of $\mathbb{R}$.

We have seen that every open subset of $\mathbb{R}$ is a finite or countably infinite union of pairwise disjoint open intervals. To show, therefore, that a σ-algebra on $\mathbb{R}$, such as $\mathscr{M}$, contains the Borel sets, it is enough to show that it contains every open interval. Indeed, this need only be shown for certain open intervals.

Lemma 2.4.3. *The interval $I = (a, +\infty)$ is measurable.*

Proof. Fix $A \subseteq \mathbb{R}$ with $m^*(A) < +\infty$. Let $A_1 = A \setminus I = \{x \in A : x \leq a\}$ and $A_2 = A \cap I = \{x \in A : x > a\}$. We must show that $m^*(A) \geq m^*(A_1) + m^*(A_2)$. Fix $\varepsilon > 0$.

For the Lebesgue case, find a countable family of open intervals I_n that cover A with total length less than $\lambda^*(A)+\varepsilon$. Let $I'_n = I_n \setminus I$ and $I''_n = I_n \cap I$. For each n, I'_n is either empty or it is an interval, and I''_n is either empty or an interval. Moreover, the nonempty intervals I'_n cover A_1, so by subadditivity their total length, which is the same as their total outer measure, is greater than or equal to $\lambda^*(A_1)$. Similarly, the nonempty intervals I''_n cover A_2, so their total length, which is the same as their total outer measure, is greater than or equal to $\lambda^*(A_2)$. Note that for each n, the length of I_n is the length of I'_n added to the length of I''_n. Therefore,

$$\begin{aligned}\lambda^*(A_1)+\lambda^*(A_2) &\le \lambda^*(\cup_n I'_n)+\lambda^*(\cup_n I''_n)\\ &\le \sum \lambda^*(I'_n)+\sum \lambda^*(I''_n) = \sum l(I_n) \le \lambda^*(A)+\varepsilon,\end{aligned}$$

and since ε is arbitrary, the result is established for the Lebesgue case.

For more general outer measures, we modify the above proof using the fact that if an interval $(\alpha,\beta]$ is cut by an interval $(a,+\infty)$, that is, if $\alpha < a < \beta$, then $(\alpha,\beta]$ will be cut into two intervals of the same kind: $(\alpha,a]$ and $(a,\beta]$. In this case, the sum of the changes on the two intervals of an integrator F will be the total change on $(\alpha,\beta]$.

Proposition 2.4.2. *The family of measurable sets $\mathscr{M}$ contains the Borel sets. In particular, $\mathscr{M}$ contains every open set and every closed set.*

Proof. We have shown that every open interval of the form $(a,+\infty)$ is in $\mathscr{M}$. Therefore, intervals of the form $(-\infty,a]$ are in $\mathscr{M}$. Since $(-\infty,a) = \bigcup_{n\in\mathbb{N}}(-\infty, a-\frac{1}{n}] \in \mathscr{M}$, and for each $a,b \in \mathbb{R}$, $(a,b) = (-\infty,b)\cap(a,+\infty)$, every open interval is in $\mathscr{M}$. Thus every open set is in $\mathscr{M}$. Since $\mathscr{M}$ is a σ-algebra containing the open sets, $\mathscr{M}$ contains the smallest σ-algebra containing the open sets, namely, the Borel sets.

Remark 2.4.1. The collection $\mathscr{M}$ of measurable sets changes with changes in the integrator, but $\mathscr{M}$ always contains the Borel sets. The collection of sets of measure 0 will, in general, be different. For example, suppose an integrator F is constant on the interval $I=(0,1)$. Then every subset of I will be measurable and have m-measure 0. As shown in Problem 2.32, however, there are non-Lebesgue measurable subsets of I.

Proposition 2.4.3. *If E and F are measurable sets such that $F\subseteq E$ and F has finite measure, then $m(E\setminus F)=m(E)-m(F)$.*

Proof. This follows from the fact that $m(E\setminus F)+m(F)=m(E)$.

Definition 2.4.5. We will use the notation $E_n \nearrow E$ to indicate a sequence of sets such that $E_n \subseteq E_{n+1}$ for all n and $\cup_n E_n = E$. Similarly, $E_n \searrow E$ indicates a sequence of sets such that $E_n \supseteq E_{n+1}$ for all n and $\cap_n E_n = E$.

Proposition 2.4.4. *Let $\langle E_n : n\in\mathbb{N}\rangle$ be a sequence of measurable sets. If $E_n \nearrow E$, then $m(E)=\lim m(E_n)$. If $E_n \searrow E$, and for some k, $m(E_k)$ is finite, then $m(E) = \lim m(E_n)$.*

Proof. Fix a sequence $E_1 \subseteq E_2 \subseteq \cdots$ with union E, and set $E_0 = \varnothing$. Form the disjoint sequence $F_k = E_k \setminus E_{k-1}$ in $\mathscr{M}$ with union E. Now, for each n, the set E_n is the disjoint union $\bigcup_{k=1}^n F_k$. Moreover, $E = \bigcup_{k=1}^\infty F_k$, and so $m(E) = \sum_{k=1}^\infty m(F_k)$. The last equality means $m(E) = \lim_n \sum_{k=1}^n m(F_k) = \lim_n m(E_n)$.

Now assume that $E_n \searrow E$ and for some k, which we may assume is 1, $m(E_1) < +\infty$. Let $H_n = E_1 \setminus E_n$ and $H = E_1 \setminus E$. Then $H_n \nearrow H$, so

$$m(H_n) = m(E_1) - m(E_n) \nearrow m(E_1) - m(E) = m(H),$$

whence $m(E_n) - m(E_1) \searrow m(E) - m(E_1)$. Since $m(E_1) < +\infty$, it follows that $m(E_n) \searrow m(E)$.

Example 2.4.2. An example showing that the finiteness condition cannot be dropped is given by Lebesgue measure and the sequence $[n, +\infty) \searrow \varnothing$.

2.5 Approximation of Measurable Sets

Results for measurable sets are often obtained using results for a smaller class of approximating sets. In this section we have examples of such approximations.

Lemma 2.5.1. *If $E \in \mathscr{M}$ and $m(E) < +\infty$, then for any $\varepsilon > 0$, there is an open set $O \supseteq E$ with $m(O \setminus E) < \varepsilon$.*

Proof. If $E = \varnothing$, set $O = \varnothing$. Otherwise, for Lebesgue measure $\lambda = m$, we take a covering of E by a countable number of open intervals so that the sum of their lengths is less than $\lambda(E) + \varepsilon$. The open set O is the union of the intervals. By subadditivity, $\lambda(E) \le \lambda(O) < \lambda(E) + \varepsilon$. Since $O = (O \setminus E) \cup E$, $\lambda(O \setminus E) = \lambda(O) - \lambda(E) < \varepsilon$.

For a general integrator F, we can take a covering by intervals of the form $(a_n, b_n]$ such that the sum of the changes in F is smaller than $m(E) + \varepsilon/2$. Since F is continuous from the right, we may replace each interval $(a_n, b_n]$ with an open interval (a_n, c_n) where $c_n > b_n$, but $F(c_n) - F(a_n) \le F(b_n) - F(a_n) + \varepsilon/2^{n+1}$. Now by Proposition 2.3.1,

$$m((a_n, c_n)) \le m((a_n, c_n]) = F(c_n) - F(a_n) \le F(b_n) - F(a_n) + \varepsilon/2^{n+1}.$$

Let $O = \cup_{n=1}^\infty (a_n, c_n)$. Then $O \supseteq E$, and by subadditivity,

$$m(O) \le \sum_{n=1}^\infty m((a_n, c_n)) \le \sum_{n=1}^\infty \left(F(b_n) - F(a_n) + \frac{\varepsilon}{2^{n+1}} \right) < m(E) + \varepsilon,$$

whence $m(O \setminus E) = m(O) - m(E) < \varepsilon$.

Theorem 2.5.1. *Fix $E \subseteq \mathbb{R}$. Then the following are equivalent:*

1) *$E \in \mathscr{M}$.*
2) *$\forall \varepsilon > 0$, $\exists$ an open set $O \supseteq E$ with $m^*(O \setminus E) < \varepsilon$.*

3) *$\forall \varepsilon > 0$, $\exists$ a closed set $F \subseteq E$ with $m^*(E \setminus F) < \varepsilon$.*
4) *$\exists$ a G_δ set G with $E \subseteq G$ such that $m^*(G \setminus E) = 0$.*
5) *$\exists$ an F_σ set S with $S \subseteq E$ such that $m^*(E \setminus S) = 0$.*
6) *$\exists$ a G_δ set G and a set A of outer measure 0 such that $E = G \backslash A = G \cap \widetilde{A}$.*
7) *$\exists$ an F_σ set S and a set A of outer measure 0 such that $E = S \cup A$.*

Proof. ($0 \Rightarrow 1$) Assume E is measurable. Let $I_1 = [-1,1]$ and $E_1 = E \cap I_1$. For each integer $n > 1$, let $I_n = [-n, -n+1) \cup (n-1, n]$ and $E_n = E \cap I_n$. Given $\varepsilon > 0$ and $n \in \mathbb{N}$, there is by Lemma 2.5.1 an open set $O_n \supseteq E_n$ such that

$$m(O_n \setminus E_n) < \varepsilon / 2^n.$$

Now, $O := \cup_n O_n$ contains E, and since

$$O \setminus E = O \cap \widetilde{E} = \cup_n (O_n \cap \widetilde{E}) = \cup_n (O_n \setminus E) \subseteq \cup_n (O_n \setminus E_n),$$

$m^*(O \setminus E) < \varepsilon$ by subadditivity.

($1 \Rightarrow 3$) By taking the intersection over a countable sequence of open sets O_n given by Condition 1 with $\varepsilon_n = 1/n$, we find a G_δ set $G \supseteq E$ with $m^*(G \setminus E) = 0$.

($3 \Rightarrow 5$) Given a G_δ set $G \supseteq E$ with $m^*(G \setminus E) = 0$, we set $A = G \setminus E$. Then $E = G \backslash A = G \cap \widetilde{A}$.

($5 \Rightarrow 0$) Any set $E \subseteq \mathbb{R}$ for which there is a G_δ set $G \supseteq E$ such that $A := G \setminus E$ has outer measure 0 is measurable since $E = G \cap \widetilde{A}$.

We have shown that measurability, Condition 1, Condition 3, and Condition 5 are equivalent. It follows that the following are equivalent statements with respect to an arbitrary set $E \subseteq \mathbb{R}$:

i) E is measurable.
ii) $\mathbb{R} \setminus E$ is measurable.
iii) $\forall \varepsilon > 0$, $\exists$ an open $O \supseteq \mathbb{R} \setminus E$, whence $\mathbb{R} \setminus O \subseteq E$, such that

$$m^*(O \setminus (\mathbb{R} \setminus E)) = m^*(O \cap E) = m^*(E \setminus (\mathbb{R} \setminus O)) < \varepsilon.$$

iv) $\forall \varepsilon > 0$, $\exists$ a closed $F \subseteq E$ with $m^*(E \setminus F) < \varepsilon$.
v) $\exists$ an F_σ set $S \subseteq E$ such that $m^*(E \setminus S) = 0$.
vi) $\exists$ an F_σ set S and a set A of measure 0 such that $E = S \cup A$.

Thus, measurability, Condition 2, Condition 4, and Condition 6 are equivalent.

Corollary 2.5.1. *A set $E \subseteq \mathbb{R}$ is measurable if and only if E is a Borel set, in fact an F_σ set, to which a set of outer measure 0 has been adjoined.*

We will see that very nice properties hold for sets of finite measure from which appropriate sets of small measure have been removed. The following is an example of such a result.

Corollary 2.5.2. *Given a measurable set $A \subseteq \mathbb{R}$ with $m(A) < +\infty$, and given $\varepsilon > 0$, there is a compact set $K \subseteq A$ with $m(A \setminus K) < \varepsilon$.*

Proof. Since A is measurable, there is a closed subset F of A with $m(A \setminus F) < \varepsilon/2$. Since the sequence

$$F \cap [-n, n] \nearrow F,$$

and $m(F) < +\infty$, there is an n_0 such that $m(F \setminus [-n_0, n_0]) < \varepsilon/2$. The desired compact set is $F \cap [-n_0, n_0]$.

Proposition 2.5.1. *If $A \notin \mathscr{M}$, then there is a G_δ set S containing A such that $m^*(S \cap A) + m^*(S \cap \widetilde{A}) \neq m(S)$. Therefore, there is no collection larger than $\mathscr{M}$ on which the restriction of m^* is even finitely additive.*

Proof. Exercise 2.24.

2.6 LimSup and LimInf of a Sequence of Sets

Recall that for a sequence $\langle x_n : n \in \mathbb{N} \rangle$ in $\mathbb{R}$, $\limsup x_n := \inf_{n \in \mathbb{N}}(\sup_{k \geq n} x_k) = \bigwedge_{n \in \mathbb{N}} \left(\bigvee_{k \geq n} x_k \right)$, and $\liminf x_n := \sup_{n \in \mathbb{N}}(\inf_{k \geq n} x_k) = \bigvee_{n \in \mathbb{N}} \left(\bigwedge_{k \geq n} x_k \right)$. Here are analogous operations on sets.

Definition 2.6.1. Let $\langle A_n : n \in \mathbb{N} \rangle$ be an infinite sequence of subsets of a set X.

$$\limsup A_n := \bigcap_{n \in \mathbb{N}} \left(\bigcup_{k \geq n} A_k \right)$$

$$\liminf A_n := \bigcup_{n \in \mathbb{N}} \left(\bigcap_{k \geq n} A_k \right).$$

Theorem 2.6.1. *Let $\langle A_n : n \in \mathbb{N} \rangle$ be an infinite sequence of subsets of a set X. Then* $\limsup A_n$ *is the set of points in an infinite number of the sets A_n, while* $\liminf A_n$ *is the set of points in all but a finite number of the sets A_n.*

Proof. Exercise 2.25.

Theorem 2.6.2 (Borel-Cantelli Lemma). *Let $\langle E_n : n \in \mathbb{N} \rangle$ be an infinite sequence of measurable subsets of $\mathbb{R}$ such that $\sum_{n=1}^{\infty} m(E_n) < +\infty$. Then* $\limsup E_n$ *is a set of measure* 0. *That is, outside of a set of measure* 0, *all points are in at most a finite number of the sets E_n.*

Proof. Let $S_k = \cup_{n=k}^{\infty} E_n$. Since $m(S_1) = m\left(\cup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} m(E_n) < +\infty$ and $S_k \searrow \limsup E_n$,

$$m(\limsup E_n) = \lim_{k \to \infty} m(S_k) \leq \lim_{k \to \infty} \sum_{n=k}^{\infty} m(E_n) = 0.$$

2.7 The Existence of a Non-measurable Set

In this section and the next, we work just with Lebesgue outer measure and Lebesgue measure. Using the Axiom of Choice (see the Appendix), we will show that there are subsets of $[0,1]$ that are not Lebesgue measurable. Robert Solovay [47] showed in 1970 that there exist models of set theory in which the Axiom of Choice does not hold and every subset of the real line is Lebesgue measurable. We will say "measurable" when we mean Lebesgue measurable.

For the construction of a non-measurable set, we work with $[0,1)$ and addition modulo 1. That is, for $x,\ y \in [0,1)$ we set $x +' y = x + y$ if $x + y < 1$, and we set $x +' y = x + y - 1$ if $x + y \geq 1$. By associating 0 with 1, one can think of $[0,1)$ with addition modulo 1 as the circle of circumference 1 centered at the origin in the plane. The operation $+'$ corresponds to rotation or addition of angles. It is easy, therefore, to see that the operation $+'$ is commutative and associative.

Lemma 2.7.1. *Given $y \in [0,1)$ and $E \subseteq [0,1)$, $\lambda^*(E +' y) = \lambda^*(E)$. If E is Lebesgue measurable, then so is $E +' y$.*

Proof. Set $E_1 = E \cap [0,\ 1-y)$ and $E_2 = E \cap [1-y, 1)$. If E is measurable, so is $E +' y = (E_1 + y) \cup (E_2 + y - 1)$. If E is any subset of $[0,1)$, then since $[0,\ 1-y)$ is measurable and Lebesgue outer measure is translation invariant,

$$\begin{aligned}\lambda^*(E) &= \lambda^*(E_1) + \lambda^*(E_2) = \lambda^*(E_1 + y) + \lambda^*(E_2 + y - 1)\\ &\geq \lambda^*((E_1 + y) \cup (E_2 + y - 1)) = \lambda^*(E +' y)\\ &\geq \lambda^*((E +' y) +' (1-y)) = \lambda^*(E).\end{aligned}$$

The last equality follows since if $x \in E$ and $x + y < 1$, then $(x +' y) +' (1-y) = x$, and the same is true if $x + y \geq 1$.

We now define an equivalence relation $\sim$ in $[0,1)$ by setting $x \sim y$ if x and y differ by a rational number. By the Axiom of Choice, there is a set $P \subseteq [0,1)$ containing exactly one element from each equivalence class. Let $\langle r_i : i \in \mathbb{N} \cup \{0\} \rangle$ be an enumeration of the rational numbers in $[0,1)$ with $r_0 = 0$. Let $P_i = P +' r_i$, so $P_0 = P$. If $i \neq j$, then $P_i \cap P_j = \varnothing$. To see this, assume $x \in P_i \cap P_j$. Then for elements p_i and p_j in P, we have

$$x = p_i +' r_i = p_j +' r_j.$$

It follows that $\left|p_i - p_j\right|$ is a rational number, i.e., $p_i \sim p_j$. Since P contains only one element from each equivalence class, $p_i = p_j$ and so $r_i = r_j$. That is, $P_i = P_j$. On the other hand, for each $x \in [0,1)$, x is in some equivalence class, so for some $p \in P$ and some r_i, $x = p +' r_i$. Therefore, the collection $\{P_i\}$ is a countable, pairwise disjoint collection of sets with union $[0,1)$.

Proposition 2.7.1. *The set P is not Lebesgue measurable.*

Proof. Assume that P is measurable. Then $\lambda(P)$ is defined, and by Lemma 2.7.1, $\lambda(P) = \lambda(P_i)$ for all i, whence $\lambda([0,1)) = \sum \lambda(P_i) = \sum \lambda(P)$. Since the sum is finite, $\lambda(P) = 0$. But then $\lambda([0,1)) = 0$. Since this is not true, we conclude that P is not measurable.

We have actually shown that the following is true.

Proposition 2.7.2. *If μ is a σ-additive,* ***translation invariant*** *measure defined on a σ-algebra containing P, then $\mu([0,1))$ is either* 0 *or* $+\infty$.

2.8 Cantor Set

The **Cantor set** C, also called the Cantor ternary set, is the closed subset of $[0,1]$ obtained by removing the following open set:

$$\left(\tfrac{1}{3},\tfrac{2}{3}\right) \cup \left[\left(\tfrac{1}{9},\tfrac{2}{9}\right) \cup \left(\tfrac{7}{9},\tfrac{8}{9}\right)\right] \cup \left[\left(\tfrac{1}{27},\tfrac{2}{27}\right) \cup \left(\tfrac{7}{27},\tfrac{8}{27}\right) \cup \left(\tfrac{19}{27},\tfrac{20}{27}\right) \cup \left(\tfrac{25}{27},\tfrac{26}{27}\right)\right] \cdots$$

That is, remove the open middle third from $[0,1]$, and at each successive step, remove the open middle third of each of the remaining closed intervals. It consists of all numbers in $[0,1]$ that have a ternary expansion (i.e., an expansion base 3) that does not use the digit 1. If there are two ternary expansions of a point in C, one of them satisfies this property. The set C is the intersection of closed subsets of $[0,1]$ such that no finite subcollection of these closed sets has an empty intersection. Since $[0,1]$ is compact, the intersection C of all of the closed subsets is nonempty.

The set C is, in fact, uncountable. To see this, assume $\langle c_n : n \in \mathbb{N} \rangle$ is a sequence of points in C. Let F_1 be the closed interval remaining after removing $\left(\frac{1}{3},\frac{2}{3}\right)$ from $[0,1]$ that does not contain c_1. At stage $n-1$, we have a closed interval that does not contain the points $c_1, c_2, \cdots, c_{n-1}$. We remove the middle third, and let F_n be the one of the two remaining closed intervals that does not contain c_n. For each $n \in \mathbb{N}$, $\cap_{i=1}^{n} F_i \neq \varnothing$. Therefore, the set $\cap_{i=1}^{\infty} F_i$ is a nonempty subset of C, and it contains no point of the enumeration. This shows that we cannot exhaust C with an enumeration; that is, C is uncountable. Working Exercise 2.33, one shows that the Lebesgue measure of the removed open set is 1, so $\lambda(C) = 0$. Note that C is an example of an uncountable set of measure 0.

One can form a **generalized Cantor set** with positive measure by scaling each of the removed intervals by α where $0 < \alpha < 1$ and removing the scaled intervals from the centers of the intervals left in the previous stage of the construction. The removed set is an open set O of measure α, and the complement F has measure $1-\alpha$. A generalized Cantor set with positive measure is also called a **fat Cantor set**.

For the Cantor set and any generalized Cantor set, the removed open set O is dense in $[0,1]$; that is, its closure is $[0,1]$. To see this, note that for any $x \in C$, there is a point y_1 removed at the first stage so that $|x - y_1| \leq 1/2$. Similarly, at the n^{th} stage, there is a point y_n removed at that stage such that $|x - y_n| \leq 1/2^n$. The extreme case would be realized if we removed first the singleton set $\{1/2\}$, then the set

$\{1/4, 3/4\}$, etc. It would still be true that this removed set (no longer open) would be dense in $[0,1]$. Note that if we take the union of the generalized Cantor sets for each $\alpha = 1/n$, we get an F_σ set with total measure 1.

Along with the Cantor set, there is a continuous increasing function g called the **Cantor-Lebesgue function** mapping $[0,1]$ onto $[0,1]$ taking all of its increase on the Cantor set, that is, on a set of measure 0. The function g is identically equal to $1/2$ on the removed middle third; it is identically equal to $1/4$ and $3/4$, respectively, on the next two removed open intervals, etc. The value at points of the Cantor set is the limit of values on the removed open intervals.

2.9 Problems

Problem 2.1. Let $\mathscr{B}$ be the collection of all subsets $A \subseteq \mathbb{R}$ such that either A or $\mathbb{R} \setminus A$ is finite or countably infinite. For each $A \in \mathscr{B}$, let $\mu(A) = 0$ if A is finite or countably infinite, and let $\mu(A) = 1$ otherwise. Show that $\mathscr{B}$ is a σ-algebra and μ is a measure on $\mathscr{B}$; that is, μ is a nonnegative, countably additive function on $\mathscr{B}$ with $\mu(\emptyset) = 0$.

Problem 2.2. Recall that a measure on a set E is a mapping μ from a σ-algebra $\mathscr{A}$ of subsets of E into $[0, +\infty]$ such that $\mu(\emptyset) = 0$ and μ is countably additive, whence μ is also finitely additive. Show that such a **general measure** is subadditive. That is, the measure of the union of a countable number of not necessarily disjoint sets in $\mathscr{A}$ is less than or equal to the sum of the measures of the sets forming the union.

Problem 2.3. Let ν be a finitely additive measure on a σ-algebra $\mathscr{A}$ of sets in a set X.

a) Suppose that for any sequence $\langle E_n \rangle$ of sets in $\mathscr{A}$, if $E_n \nearrow E$, then $\nu(E) = \lim_n \nu(E_n)$. Show that in fact ν is countably additive.
b) Suppose that for any sequence $\langle E_n \rangle$ of sets in $\mathscr{A}$, if $E_n \searrow \emptyset$, then $\lim_n \nu(E_n) = 0$. Show that in fact ν is countably additive.

Problem 2.4 (A). Prove Proposition 2.2.1.

Problem 2.5 (A). Recall that the set A consisting of the rationals between 0 and 1 is countable, and so it has Lebesgue outer measure 0. Show that any *finite* collection of open intervals covering A has total length ≥ 1.

Problem 2.6. Fix nonempty sets A and $B \subseteq \mathbb{R}$ such that

$$d(A,B) := \inf\{|x-y| : x \in A, y \in B\} = a > 0.$$

Show that Lebesgue outer measure $\lambda^*(A \cup B) = \lambda^*(A) + \lambda^*(B)$. **Hint:** Show that for any $\varepsilon > 0$, there is a countable covering of $A \cup B$ by open intervals I_k, each having length strictly less than a, such that $\sum_{k=1}^{\infty} \ell(I_k) \leq \lambda^*(A \cup B) + \varepsilon$.

Problem 2.7. Let $E \subseteq \mathbb{R}$ have finite Lebesgue outer measure. Show that E is Lebesgue measurable if and only if for any open, bounded interval (a,b) we have $b-a = \lambda^*((a,b)\cap E) + \lambda^*((a,b)\setminus E)$.

Problem 2.8. Suppose $A \subseteq \mathbb{R}$ is a Lebesgue measurable set with $\lambda(A) > 0$. Show that for any δ with $0 < \delta < 1$, there is a bounded interval $I_\delta = [a,b]$, with $a < b$, such that $\lambda(A \cap I_\delta) \geq \delta \cdot \lambda(I_\delta)$. That is, A occupies a large part of I_δ.

Problem 2.9. Suppose that $A \subseteq [0,1]$ is a Lebesgue measurable set with $\lambda(A) = 1$. Show that A is dense in $[0,1]$; that is, the closure $\overline{A} = [0,1]$.

Problem 2.10. For this problem, let $\mathscr{M}$ be the Lebesgue measurable sets in $[0,1]$, and let ν be a nonnegative, real-valued function on $\mathscr{M}$ such that for disjoint sets A and B in $\mathscr{M}$, $\nu(A\cup B) = \nu(A) + \nu(B)$. Also assume that for any $\varepsilon > 0$ there is a $\delta > 0$ so that if $A \in \mathscr{M}$ and its Lebesgue measure $\lambda(A) < \delta$, then $\nu(A) < \varepsilon$. Prove that ν is a measure. **Hint:** If $\langle A_i \rangle$ is a sequence of pairwise disjoint sets in $\mathscr{M}$ with union A, what can you say about $A \setminus \cup_{i=1}^{n} A_i$?

Problem 2.11. Let $\mathscr{M}$ be the collection of Lebesgue measurable sets in $\mathbb{R}$, and let λ be Lebesgue measure on $\mathbb{R}$. Let f be a real-valued function defined on $\mathbb{R}$. Let $\mathscr{A}$ be the collection of subsets of $\mathbb{R}$ with inverse image in $\mathscr{M}$. That is, $\mathscr{A} := \left\{S \subseteq \mathbb{R} : f^{-1}[S] \in \mathscr{M}\right\}$. Show that $\mathscr{A}$ is a σ-algebra of sets in $\mathbb{R}$. Then for each $S \in \mathscr{A}$, let $\mu(S) := \lambda\left(f^{-1}[S]\right)$. Show that μ is a measure on $\mathscr{A}$; that is, μ is countably additive with $\mu(\varnothing) = 0$.

Problem 2.12. Let f be an increasing function on $[0,1]$; that is, for $x < y$, $f(x) \leq f(y)$. The jump of f at a point x is $\lim_{y\to x+} f(y) - \lim_{y\to x-} f(y)$, with the obvious modification at endpoints of $[0,1]$. Show that if the jump of f is 0 at every point of $[0,1]$, then f is continuous on $[0,1]$.

Problem 2.13. Show that if an integrator is continuous, such as the integrator $F(x) = x$ for Lebesgue outer measure, then the same outer measure is obtained using open intervals and intervals of the form $(a,b]$.

Problem 2.14. Let F be an integrator on $\mathbb{R}$. Show that for any $\varepsilon > 0$, there is an interval $(a,b]$ such that $F(b) - F(a) < \varepsilon$.

Problem 2.15. Consider the integrator F on $\mathbb{R}$ given by $F(x) = 0$ for $x < 0$, and $F(x) = x^2$ for $x \geq 0$. Let m^* be the outer measure generated by the integrator F.

a) Given $M \in \mathbb{N}$, suppose $\langle I_n \rangle$ is a sequence of intervals contained in $[0,M]$. Show that if $l(I_n) \to 0$, then $m^*(I_n) \to 0$.
b) Construct a sequence of intervals $\langle J_n \rangle$ contained in $\mathbb{R}$ such that $l(J_n) \to 0$, but $m^*(J_n) \to \infty$.
c) Construct a sequence of intervals $\langle K_n \rangle$ contained in $\mathbb{R}$ such that $l(K_n) \to 0$, but $m^*(K_n) = 1$ for all n.

Problem 2.16. Show that an outer measure is translation invariant if and only if the integrator is $F(x) = cx + d$ for some constants $c \geq 0$, and d.

Problem 2.17. Prove or disprove: All subsets of $\mathbb{R}$ having 0 Lebesgue measure also have 0 measure with respect to the measure generated by any continuous, increasing integrator.

Problem 2.18. Let $F : (0,+\infty) \mapsto \mathbb{R}$ be given by setting $F(x) = 0$ for $x < 1$, and $F(x) = n$ for $n \in \mathbb{N}$ and $n \leq x < n+1$. Let m^* be the outer measure generated by the integrator F.

a) For each set $A \subseteq (0,+\infty)$, what is the value of $m^*(A)$?
b) Prove that every subset of $(0,\infty)$ is measurable with respect to m^*.
c) Give an example of a Lebesgue measurable set $E \subseteq (0,+\infty)$ such that $m(E) = \infty$, but the Lebesgue measure $\lambda(E) = 0$.
d) Give an example of a Lebesgue measurable set $F \subseteq (0,+\infty)$ such that $\lambda(F) = \infty$, but $m(F) = 0$.

Problem 2.19. Give an example or disprove the following statement: There exists an integrator $F : \mathbb{R} \mapsto \mathbb{R}$ such that for some set A of strictly positive Lebesgue measure, the outer measure m^* generated by F has value $m^*(\{x\}) > 0$ for each point $x \in A$.

Problem 2.20. Suppose A is a measurable subset of $\mathbb{R}$ such that $m(A \cap (a,b)) \leq \frac{1}{2}(b-a)$ for any $a, b \in \mathbb{R}$, where $a < b$. Show that $m(A) = 0$.

Problem 2.21. a) The **Heaviside step function** is $H = \chi_{[0,\infty)}$. That is, $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x \geq 1$. Show that the resulting outer measure is in fact a measure on the σ-algebra consisting of all subsets of $\mathbb{R}$. It is called a **Dirac measure** or **unit mass** at 0, and denoted by δ_0. Show that for each set $E \subseteq \mathbb{R}$, we have $\delta_0(E) = 1$ if $0 \in E$ and $\delta_0(E) = 0$ if $0 \notin E$. A similar unit mass δ_a can exist at any point $a \in \mathbb{R}$.
b) Define an integrator F such that the corresponding measure on $\mathbb{R}$ is Lebesgue measure to which is added a unit mass at 0, at 1, and at 2.

Problem 2.22 (A). Prove the following result, which is valid for Lebesgue measure, and show that it is not valid for general measures: If $E \in \mathscr{M}$, then $\forall r \in \mathbb{R}$, $E + r \in \mathscr{M}$.

Problem 2.23. Let $\langle \mu_n \rangle$ be a sequence of finite measures on a σ-algebra $\mathscr{A}$ of subsets of $\mathbb{R}$; that is, $\mu_n(\mathbb{R}) < \infty$ for all $n \in \mathbb{N}$. Let $\langle a_n \rangle$ be a sequence of positive numbers such that $\sum_{n=1}^{\infty} a_n \mu_n(\mathbb{R}) < \infty$. Let $\mu(A) = \sum_{n=1}^{\infty} a_n \mu_n(A)$ for each $A \in \mathscr{A}$. Show that μ is a finite measure on $\mathscr{A}$.

Problem 2.24 (A). Prove Proposition 2.5.1.

Problem 2.25. Prove Theorem 2.6.1.

Problem 2.26. Let μ be a measure defined on the Borel subsets of $J := [-1, 1]$ such that $\mu(J) = 17$. Assume that any Borel set of Lebesgue measure 0 in J is a set of μ-measure 0. Show that for any $\varepsilon > 0$ there is a $\delta > 0$ such that if E is a Borel set in J and $\lambda(E) < \delta$, then $\mu(E) < \varepsilon$. **Hint:** Suppose there is a sequence of Borel sets E_n contained in J with $\lambda(E_n) < 2^{-n}$ and yet $\mu(E_n) \geq \varepsilon$ for each n. Let $E = \limsup_n E_n$. What is $\lambda(E)$? What is $\mu(E)$?

Problem 2.27. Let f be a real-valued, continuous function defined on $\mathbb{R}$. Show that for each Borel set $E \subseteq \mathbb{R}$, $f^{-1}[E]$ is a Borel set.

Problem 2.28. Let m^* be the outer measure on $\mathbb{R}$ generated by an integrator F.

a) Show that for any $E \subseteq \mathbb{R}$, there is a Borel set B with $E \subseteq B$ and $m(B) = m^*(E)$.
b) Let $\langle E_n \rangle$ be a sequence of sets in $\mathbb{R}$ and E a subset of $\mathbb{R}$ such that $E_n \nearrow E$. Show that $\lim_n m^*(E_n) = m^*(E)$. **Hint:** For each n, let $B_n \supseteq E_n$ be the Borel set given by Part a. Let $C_n = \bigcap_{k=n}^{\infty} B_k$.

Problem 2.29. Let m be a measure on $\mathbb{R}$ generated by an integrator F. Let $\langle A_n \rangle$ be a sequence of measurable subsets of $\mathbb{R}$. Show that $m(\liminf A_n) \leq \liminf m(A_n)$. Assume that m is a finite measure, and show that $m(\limsup A_n) \geq \limsup m(A_n)$.

Problem 2.30. Let m be a measure on $\mathbb{R}$ generated by an integrator F. Let K be a compact set such that $m(K) < +\infty$. For each $x \in K$, let $B_1(x)$ be the interval $(x-1, x+1)$, and define $f : K \mapsto \mathbb{R}$ by setting $f(x) = m(B_1(x))$. Show that for some $x_0 \in K$, $f(x_0) = \alpha := \inf_{x \in K} f(x)$. **Hint:** Show that there is a convergent sequence $\langle x_n \rangle$ in K such that $f(x_n) \searrow \alpha$, and use Problem 2.29.

Problem 2.31 (A). Show that if E is a Lebesgue measurable subset of the non-measurable set P constructed in Section 2.7, then $\lambda(E) = 0$.

Problem 2.32 (A). Show that if A is any set with Lebesgue outer measure $\lambda^*(A) > 0$, then there is a non-measurable set $E \subseteq A$.

Problem 2.33. Show that the Cantor set has Lebesgue measure 0.

Problem 2.34 (A). Use a generalized Cantor set of positive Lebesgue measure to show there is an open subset of $[0,1]$ having a boundary (i.e., the closure of the set from which the open set has been removed) such that the boundary has positive measure.

Problem 2.35. A nonempty set S is **perfect** if it is closed and each element of S is an accumulation point of S. Prove that the Cantor set is perfect and has no interior points.

Problem 2.36. How is the Cantor set changed if closed middle third intervals are removed at each step?

Problem 2.37. Show the Cantor-Lebesgue function g is continuous on $[0,1]$ and has derivative g' equal to 0 outside of a set of Lebesgue measure 0 in $[0,1]$. **Hint:** How much does g increase on the part of the Cantor set C between two successive open intervals that have been removed at the k^{th} step of the removal process?

Chapter 3
Measurable Functions

3.1 Definition

This chapter lays the groundwork for integration applied to a large class of real-valued functions. First we note a useful fact that is independent of the real line.

Theorem 3.1.1. *Fix a σ-algebra $\mathscr{A}$ in a set X and a function f with domain $A \in \mathscr{A}$ and range in a set Y. The collection $\mathscr{B}$ of all sets $B \subseteq Y$ such that $f^{-1}[B] \in \mathscr{A}$ contains $\varnothing$ and is stable with respect to complementation and the operation of taking countable unions. That is, $\mathscr{B}$ is a σ-algebra in Y.*

Proof. The result follows from the fact that $f^{-1}[Y \setminus B] = A \setminus f^{-1}[B]$, and if the sequence $\langle B_n : n \in \mathbb{N} \rangle$ is in $\mathscr{B}$, then

$$f^{-1}\left[\bigcup_{n=1}^{\infty} B_n\right] = \bigcup_{n=1}^{\infty} f^{-1}[B_n].$$

We will use Lebesgue measure and other measures on the real line to extend Riemann integration. The extended integral will apply to functions having an appropriate structure in terms of the family of measurable sets. The definitions and results in this chapter hold for any integrator on $\mathbb{R}$. The resulting general measure is denoted by m, but when m is just Lebesgue measure, we write λ. We denote the class of measurable sets by $\mathscr{M}$. This is the class of sets measurable with respect to a given outer measure. If that outer measure is specifically Lebesgue outer measure, then we say "Lebesgue measurable."

It is convenient to follow the convention of probability theory and write $\{S(f)\}$ instead of $\{x \in A : S(f)(x)\}$ for a function f with domain A that is understood, and a property S involving f. For example, $\{\sin > 0\}$ denotes the set $\{x \in \mathbb{R} : \sin(x) > 0\}$.

As previously noted, an extended-real valued function is one taking values in the set $\mathbb{R} \cup \{+\infty, -\infty\}$. The hyphen indicates it is $\mathbb{R}$ that is extended to a larger set;

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_3

the function is not extended to a larger domain. In working with extended-real valued functions, we exclude certain combinations. We do not allow the addition of $+\infty$ to $-\infty$. Similarly, we do not allow multiplication of 0 with either infinity. One reason for the latter prohibition is that a product of sequences $x_n \searrow 0$ and $y_n \nearrow +\infty$ can have any nonnegative limiting result, depending on the choice of sequences. When the operation is allowed, we have $+\infty + a = +\infty$, $-\infty + a = -\infty$, and $+\infty \cdot a = +\infty$ if $a > 0$, $+\infty \cdot a = -\infty$ if $a < 0$, $-\infty \cdot a = -\infty$ if $a > 0$, and $-\infty \cdot a = +\infty$ if $a < 0$.

Definition 3.1.1. An extended-real valued function f with measurable domain A is a **measurable function** if for every $\alpha \in \mathbb{R}$, the set $\{f > \alpha\}$ is in $\mathscr{M}$. If the class $\mathscr{M}$ consists of the Lebesgue measurable sets, we say that f is a **Lebesgue measurable function**.

Definition 3.1.2. A set A is **dense** in a set B if $A \subseteq B$ and the closure of A contains the set B.

Note that if A is dense in B, the closure of A may be larger than B. For applications of the following result, we note that $\mathbb{R}$ itself is dense in $\mathbb{R}$, and the rational numbers are dense in $\mathbb{R}$. Any dense subset D of $\mathbb{R}$ contains a countable dense subset of $\mathbb{R}$ since for each rational number r and each $n \in \mathbb{N}$, there is a point $s \in D$ with $|r - s| < 1/n$, so (using the Axiom of Choice) we can choose one point $s \in D$ for each pair (r, n).

Proposition 3.1.1. *For an extended-real valued function f with measurable domain A, the following are equivalent:*

1) f is measurable.
2) $\forall \alpha$ in a dense subset of $\mathbb{R}$, $\{f > \alpha\} \in \mathscr{M}$.
3) $\forall \alpha$ in a dense subset of $\mathbb{R}$, $\{f \geq \alpha\} \in \mathscr{M}$.
4) $\forall \alpha$ in a dense subset of $\mathbb{R}$, $\{f < \alpha\} \in \mathscr{M}$.
5) $\forall \alpha$ in a dense subset of $\mathbb{R}$, $\{f \leq \alpha\} \in \mathscr{M}$.

Proof. Let D be a dense subset of $\mathbb{R}$. We may assume that D is countable. The result is a consequence of the following equalities, which hold for any $\alpha \in \mathbb{R}$:

$$\{f < \alpha\} = A \setminus \{f \geq \alpha\}, \quad \{f > \alpha\} = A \setminus \{f \leq \alpha\}$$

$$\{f \geq \alpha\} = \bigcap_{\substack{\gamma \in D \\ \gamma < \alpha}} \{f > \gamma\}, \quad \{f < \alpha\} = \bigcup_{\substack{\gamma \in D \\ \gamma < \alpha}} \{f < \gamma\}$$

$$\{f \leq \alpha\} = \bigcap_{\substack{\gamma \in D \\ \gamma > \alpha}} \{f < \gamma\}, \quad \{f > \alpha\} = \bigcup_{\substack{\gamma \in D \\ \gamma > \alpha}} \{f > \gamma\}.$$

Corollary 3.1.1. *If an extended-real valued function f is measurable, then for any $\alpha \in \mathbb{R} \cup \{+\infty, -\infty\}$, the set $\{f = \alpha\}$ is measurable.*

Remark 3.1.1. Even for a function f that can take the value $+\infty$ or $-\infty$, measurability only depends on values α in $D \subseteq \mathbb{R}$.

We will need sets such as $\{f \geq \alpha\}$ to be measurable in order to define an integral. We will say that f is **measurable on** B if B is measurable and the restriction of f to B is measurable. Note that measurability of a function involves only measurable sets; it does not involve a measure.

Proposition 3.1.2. *The restriction of a measurable function f with measurable domain A to a measurable subset $B \subset A$ is measurable on B. Conversely, if A is the union of a finite or countably infinite number of measurable sets on which f is measurable, then f is measurable on A.*

Proof. Exercise 3.3.

In general integration theory, one speaks about a function with measurable domain A that is measurable with respect to σ-algebras $\mathscr{A}$ and $\mathscr{B}$. That is, the inverse image of each set $B \in \mathscr{B}$ is in $\mathscr{A}$. The usual definition of measurability of an extended-real valued function uses, as is the case here, the inverse images of semi-infinite open intervals in the extended real line. That definition, however, is equivalent to the following definition in terms of the inverse image of Borel sets.

Theorem 3.1.2. *An extended-real valued function f with measurable domain A is measurable if and only if the inverse image of every Borel set in $\mathbb{R}$ is measurable and also $f^{-1}[+\infty]$ and $f^{-1}[-\infty]$ are measurable.*

Proof. We have seen that if f is measurable, then the inverse image of every semi-infinite interval in the extended real line is measurable. Also, $f^{-1}[+\infty] = \bigcap_{n\in\mathbb{N}} f^{-1}[(n,+\infty]]$ and $f^{-1}[-\infty] = \bigcap_{n\in\mathbb{N}} f^{-1}[[-\infty,-n)]$ are measurable. It now follows that the inverse image of every finite open interval is measurable, and therefore the inverse image of every open subset of the real line is measurable. Since the family of Borel sets is the smallest σ-algebra containing all open sets, it follows from Theorem 3.1.1 that the inverse image of every Borel subset of the real line is measurable. The converse is clear.

Proposition 3.1.3. *A continuous real-valued function is measurable on any measurable subset B of its domain.*

Proof. If f is continuous, then $\{x \in B : f(x) > \alpha\}$ is the intersection of B with an open set.

Definition 3.1.3. If A is a measurable subset of $\mathbb{R}$, $M(A)$ denotes the collection of measurable real-valued functions with domain A.

Recall that for functions f and g, the functions $f \vee g$ and $f \wedge g$ are defined pointwise by setting $(f \vee g)(x) := \max(f(x), g(x))$ and $(f \wedge g)(x) := \min(f(x), g(x))$. When adding or multiplying measurable functions, we will often set an arbitrary value for the sum or product on the set where the original operation is not defined. Usually that value is 0, and the set where this happens will have measure 0.

Theorem 3.1.3. *If A is a measurable set, $M(A)$ forms a vector space over $\mathbb{R}$, and $M(A)$ is stable with respect to pointwise multiplication and the operations $\vee$ and $\wedge$. Given the collection of measurable extended-real valued functions on A, for each of the following operations, there is a measurable subset that depends on the functions involved and is the set where the operation is defined; moreover, the operation yields a measurable result on that subset. The operations are: Pointwise multiplication, multiplication by any real number, pointwise addition, and the operations $\vee$ and $\wedge$.*

Proof. Fix f, g in $M(A)$ and c and α in $\mathbb{R}$. If $c = 0$, cf is constant. Otherwise,

$$\{cf > \alpha\} = \{f > \alpha/c\} \text{ if } c > 0 \text{ and } \{cf > \alpha\} = \{f < \alpha/c\} \text{ if } c < 0.$$

In either case, $cf \in M(A)$. If $f(x) + g(x) < \alpha$, then since the set $\mathbb{Q}$ of rational numbers is dense in $\mathbb{R}$, there is an $r \in \mathbb{Q}$ with $f(x) < r < \alpha - g(x)$, whence $g(x) < \alpha - r$. It follows that

$$\{f + g < \alpha\} = \bigcup_{r \in \mathbb{Q}} [\{f < r\} \cap \{g < \alpha - r\}] \in \mathscr{M}.$$

Therefore, $M(A)$ is a vector space. To see that $fg \in M(A)$, we note that the function $f^2 \in M(A)$ since if $\alpha < 0$, $A = \{f^2 > \alpha\}$ and for $\beta \geq 0$ and $\alpha = \beta^2$,

$$\{f^2 > \alpha\} = \{f < -\beta\} \cup \{f > \beta\} \in \mathscr{M}.$$

Therefore,

$$fg = (1/2)[(f + g)^2 - f^2 - g^2] \in M(A).$$

Since

$$\{f \vee g > \alpha\} = \{f > \alpha\} \cup \{g > \alpha\}, \; \{f \wedge g > \alpha\} = \{f > \alpha\} \cap \{g > \alpha\},$$

we have $f \vee g$ and $f \wedge g \in M(A)$. The result for extended-real valued functions is left as Exercise 3.4(A).

3.2 Limits and Special Functions

Recall that for a sequence $\langle f_n : n \in \mathbb{N} \rangle$, the value of the function $\sup_n f_n$ at x is $\sup_n f_n(x)$; a similar definition holds for $\inf_n f_n$. Also, $\limsup_n f_n := \inf_n(\sup_{k \geq n} f_k)$, and $\liminf_n f_n := \sup_n(\inf_{k \geq n} f_k)$.

Theorem 3.2.1. *If $\langle f_n : n \in \mathbb{N} \rangle$ is a sequence of measurable extended-real valued functions on a measurable set A, then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, and $\liminf_n f_n$ are also measurable on A.*

Proof. For any $\alpha \in \mathbb{R}$,

$$\{\sup_n f_n > \alpha\} = \cup_n \{f_n > \alpha\}, \quad \{\inf_n f_n < \alpha\} = \cup_n \{f_n < \alpha\}.$$

The rest is clear.

Definition 3.2.1. A measure space (see Definition 2.4.2) is **complete** if every subset of a set of measure 0 is measurable. In this case, the measure is also called complete. If μ is a non-complete measure on a σ-algebra $\mathscr{A}$, then the family of sets

$$\{A \cup B : A \in \mathscr{A}, B \subseteq C \text{ for some } C \in \mathscr{A} \text{ with } \mu(C) = 0\}$$

is a σ-algebra on which the extension of μ is a complete measure. The extension of μ takes the value $\mu(A)$ for all sets $A \cup B$ with $A \in \mathscr{A}$ and $B \subseteq C$ for some $C \in \mathscr{A}$ with $\mu(C) = 0$. The enlarged σ-algebra together with the extension of μ is called the **completion** of the measure space (A, μ).

Proposition 3.2.1. *The completion of a measure space is a complete measure space.*

Proof. Exercise 3.10.

In the case of a complete measure, the value of the measure on each subset of a set of measure 0 is 0. The measures we have defined using integrators are all complete. One may, however, want to consider incomplete measures such as Lebesgue measure restricted to the Borel sets. When dealing with several measures at the same time, one often cannot complete all of them, since a set of measure 0 for one measure may not be measurable for another.

When the measure m is understood, we will say that something is true **almost everywhere** (a.e.) if it is true in the complement of a set of measure 0. For example, $f = g$ a.e. on A if there is a subset B of A with $m(B) = 0$ such that $f(x) = g(x)$ for every $x \in A \setminus B$. A sequence $\langle f_n : n \in \mathbb{N} \rangle$ converges to f a.e. on A if there is a subset $B \subseteq A$ of measure 0 such that $f_n(x) \to f(x)$ for each $x \in A \setminus B$. We don't know what happens on B. This definition is useful in dealing with measures that are not complete. For example, a function f may be identically 0 except for a non-Borel subset B of a Borel set of measure 0. For integration, the value taken by f on B is not important, and so it is useful to say that $f = 0$ almost everywhere.

Proposition 3.2.2. *For a complete measure, such as Lebesgue measure, if f is measurable on A and $f = g$ almost everywhere on A, then g is measurable on A.*

Proof. Let B be the set of measure 0 outside of which $f = g$. Then g is measurable on $A \setminus B$, and since any subset of B is measurable, g is measurable on B.

Proposition 3.2.3. *If m is a complete measure and $\langle f_n : n \in \mathbb{N} \rangle$ is a sequence of measurable functions on a measurable set A such that $f_n \to f$ a.e. on A, then f is measurable on A.*

Proof. The result follows from the equality $f = \limsup_n f_n = \liminf_n f_n$ a.e.

Definition 3.2.2. A **step function** is a real-valued function g defined on an interval $[a,b]$ such that for some finite set $\{x_i : 0 \le i \le n\}$ with $a = x_0 < \cdots < x_n = b$, g is constant on each of the open intervals (x_{i-1}, x_i).

Definition 3.2.3. A **characteristic function** is a function that takes only the values 0 and 1. The set on which it takes the value 1 is the **associated set** A, and the function is called the **characteristic function** of A. We will write χ_A for this function. Another common notation for the function is 1_A. The term **indicator function** is also used.

Clearly, $\chi_\varnothing$ is the constant 0, while the characteristic function of the set in which one is working is the constant 1. A characteristic function is measurable if and only if the associated set is a measurable set (all with respect to some fixed σ-algebra). It is easy to see that,

$$\chi_{A\cap B} = \chi_A \cdot \chi_B, \quad \chi_{A\cup B} = \chi_A + \chi_B - \chi_A \cdot \chi_B, \quad \chi_{\widetilde{A}} = 1 - \chi_A.$$

Definition 3.2.4. A **simple function** is a measurable function with range equal to a finite subset of $\mathbb{R}$.

Any finite linear combination of measurable characteristic functions is a simple function. Such a representation is not unique. For example, the characteristic function of the union of two disjoint sets is the sum of their characteristic functions. Conversely, if $\alpha_1, \cdots, \alpha_n$ are the distinct nonzero values in the range of a simple function φ that is not identically equal to 0, then $\varphi = \sum_{i=1}^n \alpha_i \cdot \chi_{\{\varphi=\alpha_i\}}$. This is the simplest such combination that gives φ. A step function is a finite linear combination of characteristic functions of intervals. Some intervals may be degenerate intervals of the form $[c] := \{c\}$.

Recall that the family of Borel sets in $\mathbb{R}$ is the smallest σ-algebra containing the open subsets of $\mathbb{R}$.

Definition 3.2.5. A real-valued function is **Borel measurable** if the inverse image of each open subset of $\mathbb{R}$ is a Borel set.

Proposition 3.2.4. *The Borel measurable real-valued functions defined on a fixed Borel subset of $\mathbb{R}$ form a vector space over $\mathbb{R}$; that vector space contains the continuous functions and is stable with respect to pointwise multiplication and the operations $\vee$ and $\wedge$.*

Proof. Exercise 3.12.

Proposition 3.2.5. *Let f be a measurable real-valued function with Borel measurable range, and let h be a Borel measurable real-valued function with Borel measurable range. If g is a Borel measurable real-valued function defined on the range of f, then $g \circ f$ is measurable. If g is a Borel measurable real-valued function defined on the range of h, then $g \circ h$ is Borel measurable.*

Proof. We know that for any open set O, $f^{-1}[O]$ is a measurable set, $h^{-1}[O]$ is a Borel set, and $g^{-1}[O]$ is a Borel set. By Theorem 3.1.1, the collection of sets with measurable inverse images forms a σ -algebra. It follows that for any Borel set B, $f^{-1}[B]$ is a measurable set, $h^{-1}[B]$ is a Borel set, and $g^{-1}[B]$ is a Borel set. The rest is clear.

Example 3.2.1. The following example shows that there are Lebesgue measurable sets that are not Borel sets. Let f_1 be the Cantor-Lebesgue function on $[0,1]$. Recall that f_1 is an increasing continuous function mapping $[0,1]$ onto $[0,1]$. It is constant on each of the intervals that is removed to form the Cantor set C. For example, on $(1/3,2/3)$, f_1 takes the value $1/2$. Let f be defined on $[0,1]$ by setting $f(x) = f_1(x)+x$ for all $x \in [0,1]$. Since f_1 is an increasing continuous function and $x \mapsto x$ is a strictly increasing continuous function, f is strictly increasing and continuous; the range is $[0,2]$. Since f is a continuous bijection of the compact set $[0,1]$ onto $[0,2]$, it is a homeomorphism. (See Problem 1.36.) Each of the open intervals (a,b) removed from $[0,1]$ to form C has image $(f_1(a)+a, f_1(a)+b)$, so $f[[0,1]\setminus C]$ has the same Lebesgue measure in $[0,2]$ as has the set $[0,1]\setminus C$ in $[0,1]$. Therefore, $\lambda(f[C]) = 1$. By Problem 2.32, there is a non-Lebesgue measurable set $A \subset f[C]$. The function $g := f^{-1}$ is a homeomorphism of $[0,2]$ onto $[0,1]$, and $g[A]$ is a Lebesgue measurable subset of C since $\lambda(C) = 0$. While $g[A]$ is a Lebesgue measurable set, it is not a Borel set since g is continuous, and therefore Borel measurable, but $g^{-1}[g[A]] = f[g[A]] = A$ is not even Lebesgue measurable. Also note that the restriction of the continuous function f to $g[A]$ does not have a measurable range.

3.3 Approximations and Theorems of Lusin and Egoroff

In this section, we show that a set of finite measure has nice properties once a set of small measure, appropriate for the property, is removed. This heuristic principle, i.e., "sets of finite measure are nearly good", is essentially due to Littlewood [25]. We start with an operation used to indicate the difference of two sets. Recall that for two sets A and B, the **symmetric difference** is $A\Delta B := (A\setminus B)\cup(B\setminus A)$. We apply the symmetric difference to obtain an approximation for a measurable set of finite measure in $\mathbb{R}$.

Theorem 3.3.1. *Let A be a set of finite measure in $\mathbb{R}$. Given $\delta > 0$, there is a compact subset K of A for which $m(A\setminus K) < \delta$. Given $\varepsilon > 0$, there is a finite collection of disjoint open intervals I_i such that the measure of the symmetric difference $m(A\Delta(\cup_i I_i)) < \varepsilon$.*

Proof. We may assume $m(A) > 0$. As $n \to \infty$, $A\cap[-n,n] \nearrow A$, so we may choose a positive integer n so that $m(A\setminus[-n,n]) < \delta/2$. By Theorem 2.5.1, there is a closed, and therefore compact, set $K \subseteq A\cap[-n,n]$ so that $m((A\cap[-n,n])\setminus K) < \delta/2$. Therefore, $K \subseteq A$ and $m(A\setminus K) < \delta$. Choose a compact set $K \subseteq A$ with

$m(A \setminus K) < \varepsilon/2$. We may cover K with a bounded open set O so that $m(O \setminus K) < \varepsilon/2$. Since O is the countable union of pairwise disjoint, finite open intervals, those intervals form an open cover of K. We may discard all but a finite number of them and still cover K with the union $\cup_i I_i$. This is the desired approximation since

$$A\Delta(\cup_i I_i) = (A \setminus (\cup_i I_i)) \cup ((\cup_i I_i) \setminus A) \subseteq (A \setminus K) \cup (O \setminus K).$$

Proposition 3.3.1. *Let f be a measurable extended-real valued function that is finite almost everywhere on its domain A. If $m(A) < +\infty$, then for each $\varepsilon > 0$, there is a measurable subset $B \subseteq A$ with $m(B) < \varepsilon$ such that for some $M \in \mathbb{N}$, $|f(x)| \leq M$ for all $x \in A \setminus B$.*

Proof. Fix $\varepsilon > 0$. If f is already bounded, set $B = \varnothing$. Otherwise, for each $n \in \mathbb{N}$, set $E_n = \{|f| > n\}$, and let $E = \{f = +\infty\} \cup \{f = -\infty\}$. Since $m(E_1) < +\infty$ and $E_n \searrow E$, while $m(E) = 0$, there is an $n \in \mathbb{N}$ with $m(E_n) < \varepsilon$. Set $B = E_n$ and $M = n$.

Example 3.3.1. The function given by $f(x) = x$ is finite everywhere on $\mathbb{R}$, but it is unbounded on sets of infinite Lebesgue measure.

It follows from Proposition 3.3.1 that, if we are given an unbounded measurable function f that is finite almost everywhere on an interval $[a,b]$, then for any $\delta > 0$ and some $M \in \mathbb{N}$, we may apply our next result to the function $-M \vee f \wedge M$, which equals f outside of a set of measure less than δ.

Theorem 3.3.2. *Let f be a bounded measurable function on a closed, non-degenerate interval $[a,b]$. Let $s = \inf_{x\in[a,b]} f(x)$ and $S = \sup_{x\in[a,b]} f(x)$. Fix $\varepsilon > 0$.*

a) *There is a simple function φ defined on $[a,b]$ such that $s \leq \varphi$ and $f(x) - \varepsilon \leq \varphi(x) \leq f(x)$ for all $x \in [a,b]$, whence $|f - \varphi| \leq \varepsilon$ on $[a,b]$.*
b) *If $m(\{x\}) = 0$ for each singleton set $\{x\} \subset [a,b]$, then there is a subset $B_1 \subseteq [a,b]$ with $m(B_1) < \varepsilon/2$ and a step function g defined on $[a,b]$ such that $s \leq g \leq S$ and $g(x) = \varphi(x)$ for all $x \in [a,b] \setminus B_1$.*
c) *If $m(\{x\}) = 0$ for each singleton set $\{x\} \subset [a,b]$, then there is a subset $B_2 \subseteq [a,b]$ with $m(B_2) < \varepsilon/2$ and a continuous function h defined on $[a,b]$ such that $s \leq h \leq S$ and $h(x) = g(x)$ for all $x \in [a,b] \setminus B_2$. In this case, $|f(x) - g(x)| \leq \varepsilon$ and $|f(x) - h(x)| \leq \varepsilon$ for all $x \in [a,b] \setminus (B_1 \cup B_2)$.*

Proof. **a)** Partition $[s,S]$ with a finite number of points $s = y_0 < y_1 < \cdots < y_k = S$, so that for each i, $y_i - y_{i-1} < \varepsilon$. Of course, $f^{-1}[[y_{i-1}, y_i)]$ may be empty for some values of i. Let

$$\varphi = \left(\sum_{i=1}^{k-1} y_{i-1} \cdot \chi_{f^{-1}[[y_{i-1},y_i)]}\right) + y_{k-1} \cdot \chi_{f^{-1}[[y_{k-1},y_k]]}.$$

Now φ is a simple function with $s \leq \varphi$, and $f(x) - \varepsilon \leq \varphi(x) \leq f(x)$ for all $x \in [a,b]$, whence $|f - \varphi| \leq \varepsilon$ on $[a,b]$.

b) Let $\alpha_1, \cdots, \alpha_n$ be the n distinct values taken by φ, and let these be taken on n pairwise disjoint, measurable subsets $A_1, \cdots, A_n$ of $[a,b]$. By Theorem 3.3.1, for each set A_k, $1 \leq k \leq n$, there is a finite, pairwise disjoint collection of open intervals $I_1^k, \cdots, I_{l_k}^k$, with each contained in (a,b), such that $m(A_k \Delta (\cup_i I_i^k)) < \varepsilon/(2n)$. Set $B_1 := \cup_{k=1}^n (A_k \Delta (\cup_i I_i^k))$, and note that $m(B_1) < \varepsilon/2$. Let $\mathscr{J}$ be the collection of all of the intervals involved; that is, $\mathscr{J} = \cup_{k=1}^n \left\{ I_1^k, \cdots, I_{l_k}^k \right\}$. Let P be the finite collection of endpoints of the intervals in $\mathscr{J}$. Add the points of the null set P to B_1. For each $I \in \mathscr{J}$, if $I \cap P \neq \varnothing$, replace I with the open intervals in $I \diagdown P$. Removing duplication, this yields a finite collection $\mathscr{I}$ of pairwise disjoint open intervals such that each interval in $\mathscr{I}$ is contained in at least one interval of $\mathscr{J}$. Consider an interval $J \in \mathscr{I}$ such that for $p \neq q$ and some i_0 and j_0, $J \subseteq I_{i_0}^p \cap I_{j_0}^q$. We now show that since $A_p \cap A_q = \varnothing$, $J \subseteq B_1$. That is, fix $x \in J$. If $x \notin A_p$, then since $x \in I_{i_0}^p$, $x \in A_p \Delta (\cup_i I_i^p) \subseteq B_1$. If $x \in A_p$, then $x \notin A_q$, but x is also in $I_{j_0}^q$, so $x \in A_q \Delta \left(\cup_j I_j^q \right) \subseteq B_1$. In either case, $x \in B_1$. Thus, $J \subseteq B_1$. Discard from the collection $\mathscr{I}$ all such intervals contained in B_1. Each of the remaining intervals in $\mathscr{I}$ corresponds to a unique A_k; set the value of g equal to the appropriate α_k for each such interval. At all other points of $[a,b]$, set g equal to $(s+S)/2$. The function g is a step function such that $s \leq g \leq S$ and $g(x) = \varphi(x)$ for all $x \in [a,b] \setminus B_1$.

c) Given the step function g formed in Part b), center an open interval at each member of the finite collection of points consisting of the endpoints of $[a,b]$ together with the points inside $[a,b]$ where g changes values. The intervals should be pairwise disjoint forming a set B_2 of total length $< \varepsilon/2$. It follows that $[a,b] \setminus B_2$ is the disjoint union of closed intervals on each of which g is constant. Use linear interpolation to obtain a continuous function h on $[a,b]$ such that $s \leq h \leq S$, and $g(x) = h(x)$ for all $x \in [a,b] \setminus B_2$. It is now the case that $\varphi(x) = h(x)$ for all $x \in [a,b] \setminus (B_1 \cup B_2)$.

We have shown that if f is a measurable real-valued function on an interval $[a,b]$ where points have 0 measure, then outside of a set of small measure we may uniformly approximate f with a continuous function h. The values of h, however, are only near the values of f. An important result due to Lusin [36] states that for a measurable real-valued function f on a set A of finite measure, there is a compact subset K of A having most of the measure of A such that the values taken by f on K are equal to the values taken by a continuous real-valued function g defined on the real line. In this sense, f is "nearly" continuous on A; that is, f deviates from the continuous function g on A only on the set of small measure $A \setminus K$. Recall that by Proposition 1.11.3, once it is shown that the restriction of f to a compact subset K of $\mathbb{R}$ is continuous, there is a continuous function g defined on the whole real line such that $g = f$ on K. Moreover, $\sup_{\mathbb{R}} g = \max_K f$, and $\inf_{\mathbb{R}} g = \min_K f$.

Theorem 3.3.3 (Lusin). *Fix a measurable set $A \subseteq \mathbb{R}$ with $m(A) < +\infty$, and let f be a real-valued measurable function with domain A. For any $\varepsilon > 0$, there is a compact set $K \subseteq A$ with $m(A \setminus K) < \varepsilon$ such that the restriction of f to K is continuous.*

Proof. Let $\langle V_n : n \in \mathbb{N}\rangle$ be an enumeration of the open intervals with rational endpoints in $\mathbb{R}$. By Theorem 3.3.1, we may fix compact sets $K_n \subseteq f^{-1}[V_n]$ and $K'_n \subseteq A \setminus f^{-1}[V_n]$ for each n so that $m(A \setminus (K_n \cup K'_n)) < \varepsilon/2^n$. Now, for the compact set $K := \bigcap_{n\in\mathbb{N}}(K_n \cup K'_n)$, $m(A \setminus K) < \varepsilon$. Given $x \in K$ and an open interval I containing $f(x)$, for some $n \in \mathbb{N}$, $f(x) \in V_n \subseteq I$. Now $x \in O := \complement K'_n$, and

$$f[O \cap K] \subseteq f\left[\complement K'_n \cap \left(K_n \cup K'_n\right)\right] = f[K_n] \subseteq V_n.$$

Remark 3.3.1. This simple proof of Lusin's theorem was first published by the text's author and Erik Talvila in 2004 [34]. Lusin's theorem holds in quite general settings, where it is usually stated just for a Borel measurable function f. The domain of f should have the property that sets of finite measure can be approximated from the inside by compact sets, and the target set or range of f should have a countable collection of open sets V_n such that for each open set O and each $y \in O$, there is an n with $y \in V_n \subseteq O$. (Later, we will call this property the second axiom of countability.)

Lusin's theorem is often established as a corollary of the following approximation theorem of Egoroff [18]. That important theorem states that on a set of finite measure, almost everywhere convergence of measurable functions to a finite limit is actually uniform convergence off of a set of small measure. That is, almost everywhere convergence on a set of finite measure is "nearly" the same as uniform convergence.

Lemma 3.3.1. *On a set $A \subseteq \mathbb{R}$ of finite measure, let $\langle f_n : n \in \mathbb{N}\rangle$ be a sequence of measurable functions converging a.e. to a function f. Suppose that f is finite a.e. on A. Then for any $\delta > 0$, there is an $N \in \mathbb{N}$ and a measurable $B \subseteq A$ with $m(B) < \delta$ such that*

$$\forall x \in A \setminus B, \quad \forall n \geq N, \quad |f_n(x) - f(x)| < \delta.$$

Proof. Let D be the set where either f is not finite-valued or the convergence fails. Since $m(D) = 0$, we may set each f_n and f equal to 0 on D and work with the modified functions without loss of generality. Fix $\delta > 0$, and let $S_n = \{|f_n - f| \geq \delta\}$. Now, $\limsup\ S_n = \varnothing$, so $\lim_{k\to\infty} m(\cup_{n\geq k} S_n) = 0$. Choose $N \in \mathbb{N}$ so that $m(\cup_{n\geq N} S_n) < \delta$, and let $B = \cup_{n\geq N} S_n$.

Theorem 3.3.4 (Egoroff). *On a set $A \subseteq \mathbb{R}$ of finite measure, let $\langle f_n : n \in \mathbb{N}\rangle$ be a sequence of measurable functions converging a.e. to a function f. Suppose f is finite a.e. on A. For any $\varepsilon > 0$, there is a measurable set $B \subseteq A$ with $m(B) < \varepsilon$ such that f_n converges uniformly to f on $A \setminus B$.*

Proof. Fix $\varepsilon > 0$. For each $k \in \mathbb{N}$, it follows from Lemma 3.3.1 with $\delta = \varepsilon/2^k$ that there is an $N_k \in \mathbb{N}$ and a measurable set $B_k \subset A$ with $m(B_k) < \varepsilon/2^k$ such that

$$\forall n \geq N_k, \ |f_n - f| < \varepsilon/2^k \text{ on } A \setminus B_k.$$

Let $B = \cup_k B_k$, so $m(B) < \varepsilon$. The functions f_n converge uniformly to f on $A \setminus B$ since for all $n \geq N_k$, $|f_n - f| < \varepsilon/2^k$ on $A \setminus B_k \supseteq A \setminus B$.

3.4 Problems

Problem 3.1. Prove Corollary 3.1.1.

Problem 3.2. Let $f : \mathbb{R} \mapsto \mathbb{R}$ and $g : \mathbb{R} \mapsto \mathbb{R}$ be measurable functions and $g(x) \neq 0$ at any $x \in \mathbb{R}$. Show that the function f/g is measurable.

Problem 3.3. Prove Proposition 3.1.2.

Problem 3.4 (A). Finish the proof of Proposition 3.1.3 for extended-real valued functions.

Problem 3.5. Let $\langle f_n \rangle$ be a sequence of real-valued measurable functions on $\mathbb{R}$. Show that the set $\{x \in \mathbb{R} : \lim_{n\to\infty} f_n(x)$ exists and is finite$\}$ is measurable.

Problem 3.6. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a differentiable function. Show that the derivative f' is a measurable function.

Problem 3.7. Prove or give a counterexample: The supremum of an uncountable family of measurable functions is always measurable.

Problem 3.8. Given $f : [0,1] \mapsto \mathbb{R}$, suppose the set $\{x \in [0,1] : f(x) = r\}$ is measurable for every $r \in \mathbb{R}$. Does it then follows that f is measurable?

Problem 3.9. Suppose that $f : [0,1] \mapsto \mathbb{R}$ is a function with the property that for any $\varepsilon > 0$, there is a continuous function $f_\varepsilon : [0,1] \mapsto \mathbb{R}$ such that $f = f_\varepsilon$ on A_ε, where $m([0,1] \setminus A_\varepsilon) < \varepsilon$. Show that f is measurable.

Problem 3.10. Prove Proposition 3.2.1. **Hint:** Given subsets A, B, and C of a set X with $B \subseteq C$,

$$X \setminus (A \cup B) = (X \setminus (A \cup C)) \cup (C \setminus (A \cup B)).$$

Problem 3.11. Let f be a real-valued function defined on $\mathbb{R}$ such that for each $\alpha \in \mathbb{R}$, $f^{-1}[(\alpha, +\infty)]$ is a Borel set.

a) Show that for each open subset O of $\mathbb{R}$, $f^{-1}[O]$ is a Borel set.
b) Show that for each Borel set $E \subseteq \mathbb{R}$, $f^{-1}[E]$ is a Borel set.
c) Show that if f is actually continuous on $\mathbb{R}$, then for each Borel set $E \subseteq \mathbb{R}$, $f^{-1}[E]$ is a Borel set.

Problem 3.12. Prove Proposition 3.2.4.

Problem 3.13. Let $\{I_\alpha : \alpha \in A\}$ be an uncountable collection of open intervals in the real line such that the measure of the union, $m(\cup_{\alpha\in A} I_\alpha)$, is a finite number $r > 0$. Given an arbitrary $\varepsilon > 0$, show that there is a finite subcollection $\{I_1, I_2, \cdots, I_n\}$ of the collection $\{I_\alpha : \alpha \in A\}$ such that $\sum_{i=1}^n m(I_i) > r - \varepsilon$.

Problem 3.14. **a)** Show that there does not exist a simple function $\varphi : \mathbb{R} \mapsto \mathbb{R}$ such that $|x^2 - \varphi(x)| \leq 1$ for all $x \in \mathbb{R}$.

b) Prove or give a counterexample: For every Lebesgue measurable set $E \subset \mathbb{R}$ of finite measure, there exists a simple function $\psi : \mathbb{R} \mapsto \mathbb{R}$ such that $|x^2 - \psi(x)| \leq 1$ for all $x \in E$.

Problem 3.15. a) Show that if f is a measurable real-valued function with measurable range and g a continuous real-valued function defined on the real line, then $g \circ f$ is measurable.

b) Show that a continuous function with measurable range followed by a measurable function need not be measurable. **Hint:** See Example 3.2.1.

Problem 3.16. Show that an increasing real-valued function on the interval $[0,1]$ can have only a finite or countably infinite number of jumps.

Problem 3.17. Define $f : (0,1) \longmapsto \mathbb{R}$ as follows: For each $k \in \mathbb{N}$, set $f(x) = \left(\frac{1}{k} - x\right)^{-1}$ for all $x \in \left[\frac{1}{k+1}, \frac{1}{k}\right)$. For example, for $x \in [\frac{1}{2}, 1)$, $f(x) = \frac{1}{1-x}$. For each $n \in \mathbb{N}$, set $f_n(x) := \frac{1}{n} f(x)$ for all $x \in (0,1)$. Note that f_n converges pointwise to 0, but not uniformly to 0 on $(0,1)$.

a) Show that f_n is a measurable function on $(0,1)$ for each $n \in \mathbb{N}$.

b) Fix $\varepsilon > 0$. Construct a Lebesgue measurable set E such that $\lambda(E) < \varepsilon$ and f_n converges uniformly to 0 on $(0,1) \setminus E$.

Problem 3.18 (A). Given an increasing real-valued function f on an interval I, show that f is measurable. **Hint:** First consider the strictly increasing function for some $n \in \mathbb{N}$, $x \longmapsto f(x) + x/n$.

Problem 3.19. Let f be a continuous real-valued function on $\mathbb{R}$. Show that if A is an F_σ subset of $\mathbb{R}$, then $f[A]$ is an F_σ set.

Problem 3.20. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a **Lipschitz function**; that is, there is an $M > 0$ such that $|f(x) - f(y)| \leq M \cdot |x - y|$ for all $x, y \in \mathbb{R}$. Show that for any Lebesgue measurable set E, $f[E]$ is a Lebesgue measurable set. **Hint:** Recall Corollary 2.5.1 and Problem 3.19.

Problem 3.21. Let $E \subseteq \mathbb{R}$ be a measurable set of finite measure, and let f be a real-valued measurable function on E. Show that f is the a.e. limit of a sequence of continuous functions.

Problem 3.22. Let f be a real-valued function with domain $\mathbb{R}$ such that the inverse image of every closed subset of $\mathbb{R}$ is an open subset of $\mathbb{R}$. Show that for some value $a \in \mathbb{R}$, $f(x) \equiv a$ on $\mathbb{R}$. **Hint:** Recall Problem 1.22.

Chapter 4
Integration

4.1 The Riemann Integral

In this chapter, we extend the operation of taking the Riemann integral to more general integrals on a large class of functions. A bounded function will always mean a bounded real-valued function.

The Riemann integral uses the length of intervals, that is, the change on intervals of the integrator $F(x) = x$; the corresponding measure is Lebesgue measure λ. The integral we will construct using Lebesgue measure will be called the Lebesgue integral to distinguish it from integrals obtained from other measures. As before, we use m to denote a measure constructed from a general integrator. For the Lebesgue integral and the more general integral, the construction methods are essentially the same. In the literature, therefore, the general integral is often called a Lebesgue integral. When there is an emphasis on the integrator, however, the general integral is also called a Lebesgue–Stieltjes integral. We begin here by recalling the construction of the Riemann integral.

Definition 4.1.1. A **partition** P of a closed and bounded interval $[a,b]$ is a finite set of points $a = x_0 < x_1 < \cdots < x_n = b$. The value of n depends on the partition P. Let f be a bounded function on $[a,b]$. Given a partition P of $[a,b]$, for $1 \leq i \leq n$, set $m_i = \inf_{[x_{i-1},x_i]} f$ and $M_i = \sup_{[x_{i-1},x_i]} f$. The **lower and upper Riemann sums** of f are the finite sums

$$s(P,f) = s(P) := \sum_{i=1}^{n} m_i(x_i - x_{i-1}), \text{ and } S(P,f) = S(P) := \sum_{i=1}^{n} M_i(x_i - x_{i-1}),$$

respectively.

Proposition 4.1.1. *Let f be a bounded function on $[a,b]$. If P_1 and P_2 are partitions of $[a,b]$, then $s(P_1) \leq s(P_1 \cup P_2) \leq S(P_1 \cup P_2) \leq S(P_2)$. It follows that every lower sum for f is smaller than any upper sum.*

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_4

Proof. Given an interval $[x_{i-1}, x_i] \subseteq [a,b]$ and a point c with $x_{i-1} < c < x_i$,

$$\inf_{[x_{i-1},x_i]} f \leq \inf_{[x_{i-1},c]} f \leq \sup_{[x_{i-1},c]} f \leq \sup_{[x_{i-1},x_i]} f.$$

A similar inequality holds for $[c, x_i]$. The rest is clear.

Definition 4.1.2. Let $\mathscr{P}$ denote the set of all partitions of $[a,b]$. Let f be a real-valued function on $[a,b]$. The **lower and upper Riemann integrals** of f on $[a,b]$ are given by

$$R\underline{\int_a^b} f(x)\,dx := \sup_{P\in\mathscr{P}} s(P,f), \text{ and } R\overline{\int_a^b} f(x)\,dx := \inf_{P\in\mathscr{P}} S(P,f),$$

respectively. The function f is **Riemann integrable** if the lower and upper Riemann integrals are equal. In this case, the **Riemann integral** of f is equal to the common value and is denoted by $R\int_a^b f(x)dx$.

The definition of upper and lower sums uses intervals overlapping at endpoints. To illuminate the relationship of the Riemann integral with Lebesgue measure and the Lebesgue integral, we now show that the Riemann integral can be obtained using step functions formed from pairwise disjoint intervals.

Definition 4.1.3. A partition P of a closed and bounded interval $[a,b]$ is a **partition for a step function** ψ if ψ is constant on the open intervals between partition points. Such a partition P is **minimal** for a step function ψ on $[a,b]$ if it is a partition for ψ and for each partition point x_i in (a,b), it is not the case that $\lim_{x\to x_i-}\psi(x) = \psi(x_i) = \lim_{x\to x_i+}\psi(x)$. That is, ψ changes values in crossing x_i; the function ψ may take the same value on the right and on the left of x_i, but then it must take a different value at x_i.

Note that a minimal partition of $[a,b]$ for a step function ψ is unique. Moreover, the definition leaves the endpoint a in the partition P even if $\lim_{x\to a+}\psi(x) = \psi(a)$. Similarly, the endpoint b is in P.

Definition 4.1.4. Given a step function ψ on $[a,b]$ and a minimal partition for ψ consisting of points $a = x_0 < x_1 < \cdots < x_n = b$, set $I_1 = [x_0]$, $I_2 = (x_0, x_1)$, $I_3 = [x_1]$, $I_4 = (x_1, x_2)$, etc. Let c_j be the value of ψ on I_j. This may be 0. The sum $\psi = \sum c_j \cdot \chi_{I_j}$ is the **minimal representation** of the step function ψ.

A step function is, of course, also a simple function. A simple function, however, may associate into a single set any two intervals where the same value is taken. Recall that for an interval I, the Lebesgue measure $\lambda(I)$ is the length $l(I)$.

Definition 4.1.5. If ψ is a step function with minimal representation $\sum c_j\chi_{I_j}$, then the **integral** with respect to Lebesgue measure λ is

$$\int \psi\, d\lambda := \sum c_j\lambda(I_j) = \sum c_j l(I_j).$$

Proposition 4.1.2. *If ψ is a step function on $[a,b]$ and $\{J_k\}$ is any finite, pairwise disjoint collection of subintervals of $[a,b]$ such that $\psi = \sum a_k \cdot \chi_{J_k}$, then*

$$\int \psi \, d\lambda = \sum a_k \lambda(J_k) = \sum a_k l(J_k).$$

Proof. Without changing the value of the sum $\sum a_k \lambda(J_k)$, we may assume that any endpoint of an interval J_k is a singleton interval $[x_i]$ in the collection. Since ψ is constant on each J_k, each J_k is contained in one of the intervals for the minimal partition. We may also assume, without changing the value of the sum $\sum a_k \lambda(J_k)$, that $[a,b] = \cup_k J_k$. That is, we adjoin intervals where ψ is 0. If I is one of the intervals formed by the minimal partition, then $\lambda(I) = \sum_{J_k \subseteq I} \lambda(J_k)$, and the value of ψ on I is the same as its value on each of the intervals J_k contained in I. The rest is clear.

Definition 4.1.6. A function g dominates a function f on a common domain A if $g(x) \geq f(x)$ for all $x \in A$.

Theorem 4.1.1. *Given a bounded function f on an interval $[a,b]$, the upper Riemann integral of f is the infimum of the integrals of all step functions that dominate f. Similarly, the lower Riemann integral of f is the supremum of the integrals of all step functions that are dominated by f.*

Proof. Fix $M > \sup_{[a,b]} |f|$. Fix a partition P consisting of points $a = x_0 < x_1 < \cdots < x_n = b$. As before, for each i, $M_i = \sup_{[x_{i-1},x_i]} f$, but we set $\psi = M_i$ only on (x_{i-1}, x_i), and we set $\psi = M$ on $\{x_{i-1}\}$ and $\{x_i\}$. Then $\psi \geq f$ and $\int \psi \, d\lambda = S(P,f)$. That is, every upper sum is equal to the integral of a step function that dominates f. It follows that the infimum of the integrals of all step functions that dominate f is less than or equal to the upper Riemann integral. To show it cannot be strictly less than the upper Riemann integral, let $\psi = \sum c_j \chi_{I_j}$ be a step function given in minimal form with domain $[a,b]$ such that $\psi \geq f$. For any $\varepsilon > 0$, there is a partition P of $[a,b]$ such that the endpoints of the partition intervals I_j for ψ are contained in intervals formed by P of total length at most ε/M; the contribution to the upper sum of f over these intervals is no more than ε. The step function ψ is constant and dominates f on each of the remaining closed intervals $[x_{i-1}, x_i]$. It follows that the upper Riemann sum $S(P,f) \leq \int \psi \, d\lambda + \varepsilon$, and since ε is arbitrary, the result is established for the upper Riemann sum. Applying that result to $-f$, we obtain the desired result for the lower Riemann sum.

The problem of characterizing those functions for which the Riemann integral exists first occurs in calculus. The answer can now be given in terms of Lebesgue measure λ.

Theorem 4.1.2. *A bounded function f on an interval $[a,b]$ is Riemann integrable if and only if the set of points of discontinuity of f in $[a,b]$ has Lebesgue measure* 0.

Proof. Let $\mathscr{U}_s(f)$ be the set of all step functions $\psi \geq f$ on $[a,b]$, and let $\mathscr{V}_s(f)$ be the set of all step functions $\varphi \leq f$ on $[a,b]$. For each $n \in \mathbb{N}$, let

$$B_n = \{x \in [a,b] : \forall \delta > 0, \ \exists y \in [a,b] \text{ with } |y-x| < \delta \text{ and } |f(y)-f(x)| \geq 1/n\}.$$

The set of points of discontinuity of f is $B = \cup_n B_n$. We assume first that f is Riemann integrable and show that for each $n \in \mathbb{N}$, $\lambda(B_n) = 0$, whence $\lambda(B) = 0$. Fix $\varepsilon > 0$. By Theorem 4.1.1 and the assumption that f is Riemann integrable, there is a $\psi \in \mathscr{U}_s$ and a $\varphi \in \mathscr{V}_s$ with $\int (\psi - \varphi)\ d\lambda < \varepsilon$. Let S be the finite set of partition points, including a and b, associated with the minimal representation of the step function $\psi - \varphi$. The set $[a,b] \setminus S$ consists of a finite number of disjoint open intervals $I_1, \cdots, I_k$ on each of which $\psi - \varphi$ is constant. Given $n \in \mathbb{N}$, let O_n be the union of those intervals for which that constant is at least $1/n$. The set $B_n \subseteq O_n \cup S$. Since $\frac{1}{n} \cdot \lambda(O_n) \leq \int (\psi - \varphi)\ d\lambda < \varepsilon$, the measure $\lambda(B_n) \leq \lambda(S) + \lambda(O_n) < n\varepsilon$. Since ε is arbitrary, $\lambda(B_n) = 0$.

Now assume that $\lambda(B) = 0$. Fix $M > \sup_{[a,b]} |f|$, and again fix $\varepsilon > 0$. There is a countable collection $\mathscr{J}$ of open intervals for which the union O contains B and $\lambda(O) < \varepsilon/(4M)$. For each $x \in [a,b]$ that is not in B, there is an open interval $I_x = (x - \delta_x, x + \delta_x)$ such that for every $y \in I_x \cap [a,b]$, $|f(y) - f(x)| < \frac{\varepsilon}{4(b-a)}$, whence for every y and z in $I_x \cap [a,b]$, $|f(y) - f(z)| < \frac{\varepsilon}{2(b-a)}$. A finite subcollection of these open intervals together with a finite number of intervals from $\mathscr{J}$ forms a covering of the compact set $[a,b]$. Let P consist of the endpoints in $[a,b]$ of the intervals in this finite covering together with the points a and b. The set P is a partition of $[a,b]$ such that each of the open intervals J_i between adjacent partition points is either contained in O or is contained in I_x for some $x \notin B$, or both. On each such interval J_i, let ψ take the value $\sup_{J_i} f$ and φ take the value $\inf_{J_i} f$. On the points of P, let $\psi = M$ and $\varphi = -M$. Then

$$\int (\psi - \varphi)\ d\lambda \leq 2M \cdot \lambda(O) + \frac{\varepsilon}{2(b-a)}(b-a) < \varepsilon.$$

Since $\psi \in \mathscr{U}_s(f)$, $\varphi \in \mathscr{V}_s(f)$, and ε is arbitrary, it follows from Theorem 4.1.1 that f is Riemann integrable.

4.2 The Integral of Simple Functions

While step functions are appropriate for the development of the Riemann integral, for a more general integral, we replace step functions with simple functions. Recall that a simple function is a measurable function with range equal to a finite subset of $\mathbb{R}$. By extending such a function with the value 0, we may assume that it is defined on all of $\mathbb{R}$. Most of the results that follow hold for a measure m generated by an arbitrary integrator, in which case, "measure", "measurable", and "almost everywhere" (abbreviated "a.e.") refer to the measure m. Of course, Lebesgue measure λ is a special case.

Definition 4.2.1. A **simple function with finite measure support** is a simple function that is 0 outside a set of finite measure. Such a function ψ that is not identically equal to 0 is presented in **canonical form** as the finite sum $\sum a_i \chi_{A_i}$ where the A_i's are measurable with finite measure and pairwise disjoint, and the a_i's are distinct and not equal to 0. The canonical representation of the function 0 is 0.

Definition 4.2.2. The **integral** $\int \psi$ of a simple function ψ with finite measure support and canonical form $\sum a_i \chi_{A_i}$ is the sum $\sum a_i \cdot m(A_i)$. The integral of 0 is 0. Notation that stresses the measure is $\int \psi \, d\lambda$ and $\int \psi \, \lambda(dx)$ for the integral with respect to Lebesgue measure and $\int \psi \, dF$ and $\int \psi \, dm$, as well as $\int \psi \, m(dx)$, for the integral with respect to a more general measure m. Given a measurable set E, $\int_E \psi$ denotes the integral $\int \psi \cdot \chi_E$.

Definition 4.2.3. Given a nonempty measurable set E and a finite number of measurable subsets E_i of E, the **partition refinement** of E determined by the sets E_i is the finite collection of nonempty, measurable, pairwise disjoint subsets $\{A_j : 1 \le j \le k\}$ of E such that $E = \cup_j A_j$, and each E_i is the union of the A_j's that have nonempty intersection with E_i.

Note that such a partition refinement is obtained by taking all the nonempty intersections of the sets E_i, $E \setminus E_i$ with the sets E_j, $E \setminus E_j$ for $i \ne j$.

Proposition 4.2.1. *If φ is a finite linear combination $\sum_{i=1}^n \alpha_i \chi_{E_i}$ of characteristic functions of nonempty, measurable sets E_i with $m(E_i) < +\infty$ for each i, then φ is a simple function with finite measure support, and $\int \varphi = \sum_{i=1}^n \alpha_i \cdot m(E_i)$.*

Proof. Let $E = \cup_i E_i$, and note that $m(E) < +\infty$. The function φ is measurable and takes only a finite number of values, so φ is a simple function with finite measure support. Let $\{A_j,\ 1 \le j \le k\}$ be the partition refinement of E determined by the sets E_i. Now $\varphi = \sum_{j=1}^k c_j \chi_{A_j}$, where for each j, $c_j = \sum_{E_i \supseteq A_j} \alpha_i$. It follows that

$$\sum_{i=1}^n \alpha_i \cdot m(E_i) = \sum_{i=1}^n \alpha_i \cdot \left(\sum_{A_j \subseteq E_i} m(A_j) \right) = \sum_{i=1}^n \sum_{A_j \subseteq E_i} \alpha_i \cdot m(A_j)$$
$$= \sum_{j=1}^k \left(\sum_{E_i \supseteq A_j} \alpha_i \right) \cdot m(A_j) = \sum_{j=1}^k c_j \cdot m(A_j).$$

Given the representation $\varphi = \sum_{j=1}^k c_j \chi_{A_j}$, we may drop any term with $c_j = 0$ without changing the function or the value of the sum $\sum_{j=1}^k c_j \cdot m(A_j)$. Also, we can combine, by taking a union, all of the remaining sets A_j with the same value c_j into one set without changing the function or the value of the sum $\sum_{j=1}^k c_j \chi_{A_j}$. The result is the canonical form for φ and its integral.

Proposition 4.2.2. *Let φ and ψ be simple functions with finite measure support. Then for any $\alpha, \beta \in \mathbb{R}$, $\int (\alpha\varphi + \beta\psi) = \alpha \int \varphi + \beta \int \psi$. If $\psi \ge \varphi$ a.e., then $\int \psi \ge \int \varphi$.*

Proof. Let $E = \{|\varphi| + |\psi| > 0\}$. If $E = \varnothing$, the result is clear; in any case, $m(E) < +\infty$. Take the partition refinement $\{A_j : 1 \le j \le k\}$ of E determined by the sets $E_i \subseteq E$ on which φ takes distinct values (including 0) and the sets $F_k \subseteq E$ on which ψ takes distinct values (including 0). Then φ and ψ have representations $\varphi = \sum_j c_j \chi_{A_j}$ and $\psi = \sum_j d_j \chi_{A_j}$. Therefore,

$$\begin{aligned}\int(\alpha\varphi+\beta\psi) &= \sum_j(\alpha c_j+\beta d_j)m(A_j) = \alpha\sum_j c_j m(A_j)+\beta\sum d_j m(A_j)\\ &= \alpha\int\varphi+\beta\int\psi.\end{aligned}$$

If $\varphi \geq \psi$ a.e., we may change their values on a set of measure 0 without changing the integrals so that $\varphi \geq \psi$ on E. Now for each j, $c_j \geq d_j$, so $\int\varphi \geq \int\psi$.

Corollary 4.2.1. *Given φ and ψ as in the proposition, if $b \leq \varphi \leq B$ on $E = \{\varphi \neq 0\}$, then $b \cdot m(E) \leq \int\varphi \leq B \cdot m(E)$. If $0 \leq \psi - \varphi \leq \varepsilon$ on $E = \{|\psi| + |\varphi| \neq 0\}$, then $0 \leq \int\psi - \int\varphi \leq \varepsilon \cdot m(E)$.*

Proof. Clear.

4.3 The Integral of Bounded Measurable Functions

In this section, we use the results for simple functions to define the integral for bounded, measurable functions defined on sets of finite measure. As before, the measure m need not be Lebesgue measure. In what follows, we use terminology that is common in the literature, and say that a function **vanishes** on a set if it is identically equal to 0 on the set. As before, we call a measurable set of measure 0 a **null set**.

Definition 4.3.1. Given a measurable set E of finite measure and a bounded measurable function f on E, let $\mathscr{U}(f)$, or just $\mathscr{U}$, denote the set of all simple functions ψ that vanish on $\mathbb{R}\setminus E$ such that $\psi \geq f$ on E. Let $\mathscr{V}(f)$, or just $\mathscr{V}$, denote the set of all simple functions φ that vanish on $\mathbb{R}\setminus E$ such that $\varphi \leq f$ on E.

Proposition 4.3.1. *Let E be a nonempty measurable set of finite measure and f a bounded function on E. If f is measurable on E, then there is an increasing sequence $\langle\varphi_n : n \in \mathbb{N}\rangle$ in $\mathscr{V}(f)$ and a decreasing sequence $\langle\psi_n : n \in \mathbb{N}\rangle$ in $\mathscr{U}(f)$ with both converging uniformly to f on E. It follows that*

$$\inf_{\psi\in\mathscr{U}}\int\psi = \sup_{\varphi\in\mathscr{V}}\int\varphi. \tag{4.3.1}$$

Conversely, given that m is a complete measure, if Equation (4.3.1) holds, then f is measurable on E.

Proof. For each $\psi \in \mathscr{U}$ and $\varphi \in \mathscr{V}$, $\psi \geq f \geq \varphi$, so $\int \psi - \int \varphi \geq 0$, and

$$\inf_{\psi \in \mathscr{U}} \int \psi - \sup_{\varphi \in \mathscr{V}} \int \varphi \geq 0.$$

Assume that f is measurable on E. Let $s = \inf_E f$ and $S = \sup_E f$; if $s = S$, the result is clear. Otherwise, fix $n \in \mathbb{N}$. Partition $[s, S]$ with points $s = y_0 < y_1 < \cdots < y_k = S$ so that $y_i - y_{i-1} < 1/n$ for each i. As in Theorem 3.3.2, let

$$\varphi_n = \left(\sum_{i=1}^{k-1} y_{i-1} \cdot \chi_{f^{-1}[[y_{i-1}, y_i))]} \right) + y_{k-1} \cdot \chi_{f^{-1}[[y_{k-1}, S]]}.$$

Then φ_n is a simple function with $s \leq \varphi_n \leq f$ on E and $f - \varphi_n < 1/n$ on E. Similarly, or applying this result to $-f$, there is a simple function ψ_n with $f \leq \psi_n \leq S$ on E and $\psi_n - f < 1/n$ on E. Therefore, $\psi_n - \varphi_n \leq 2/n$ on E. Replacing φ_n with $\varphi_1 \vee \varphi_2 \vee \cdots \vee \varphi_n$ and ψ_n with $\psi_1 \wedge \psi_2 \wedge \cdots \wedge \psi_n$, we obtain increasing and decreasing sequences of simple functions converging uniformly to f. It now follows that for any $\varepsilon > 0$, there is a $\psi \in \mathscr{U}$ and a $\varphi \in \mathscr{V}$ with $\psi - \varphi < \varepsilon$ on E, whence $\int \psi - \int \varphi < \varepsilon \cdot m(E)$. Therefore, Equation (4.3.1) holds.

Now assume that $\inf_{\psi \in \mathscr{U}} \int \psi = \sup_{\varphi \in \mathscr{V}} \int \varphi = \alpha$. Again using the operations $\wedge$ and $\vee$, we may find a decreasing sequence $\langle \psi_n \rangle$ in $\mathscr{U}$ and an increasing sequence $\langle \varphi_n \rangle$ in $\mathscr{V}$ with $\int \psi_n \to \alpha$ and $\int \varphi_n \to \alpha$. Let $\psi = \inf \ \psi_n$ and $\varphi = \sup \ \varphi_n$. Then $\varphi \leq f \leq \psi$, and ψ and φ are measurable. Since the measure is complete, we need only show that $\varphi = f = \psi$ a.e. Fix $k \in \mathbb{N}$, and let $B_k = \{\psi - \varphi \geq 1/k\}$. For every $n \in \mathbb{N}$, $(1/k) \cdot m(B_k) \leq \int \psi_n - \int \varphi_n$, so $m(B_k) = 0$. Since $\{\psi - \varphi > 0\} = \cup_k B_k$, it follows that $\psi = f = \varphi$ a.e., so f is measurable.

The last statement of the proposition may not be true if the measure is not complete. For example, if B is a non-measurable subset of a null set $A \subseteq E$, then $f = \chi_B$ is not measurable, but Equation (4.3.1) still holds.

Definition 4.3.2. Let f be a bounded measurable function defined on a measurable set E of finite measure. The **integral** of f is the common value $\inf_{\psi \in \mathscr{U}} \int \psi = \sup_{\varphi \in \mathscr{V}} \int \varphi$. The integral is denoted by $\int_E f$, or if E is understood, just $\int f$. Again, notation that stresses the measure is $\int f \, d\lambda$ and $\int f \, \lambda(dx)$ for the integral with respect to Lebesgue measure, called the **Lebesgue integral**; for the integral with respect to a more general measure m, it is $\int f \, dF$ and $\int f \, dm$, as well as $\int f \, m(dx)$. If A is a measurable subset of the domain of f, we write $\int_A f$ for $\int f \cdot \chi_A$.

It is easy to see that this definition of the integral gives the same value as the previous definition of the integral when applied to simple functions with finite measure support. It follows from the definitions of $\mathscr{U}(f)$ and $\mathscr{V}(f)$ that the integral of f on the empty set is 0. It is also easy to show (Problem 4.3) that for a measurable subset $A \subset E$, $\int_A f = \int f \cdot \chi_A$ has the same value as the integral obtained by restricting f to A. We next show that the Lebesgue integral is indeed an extension of the Riemann integral.

Theorem 4.3.1. *If f is a bounded Riemann integrable function on $[a,b]$, then f is measurable with respect to Lebesgue measure λ, and the Riemann integral of f and the Lebesgue integral of f are equal.*

Proof. The step functions on $[a,b]$ that dominate f are in $\mathscr{U}(f)$, and the step functions that are dominated by f are in $\mathscr{V}(f)$. Therefore,

$$R\underline{\int} f \leq \sup_{\varphi\in\mathscr{V}} \int \varphi \, d\lambda \leq \inf_{\psi\in\mathscr{U}} \int \psi \, d\lambda \leq R\overline{\int} f.$$

By assumption, all terms of the inequality are equal, so the result follows from the resulting equality and Proposition 4.3.1.

Recall that for a nonempty measurable set $E \subseteq \mathbb{R}$, we write $M(E)$ for the space of real-valued measurable functions on E. By Theorem 3.1.3, it is a vector space over $\mathbb{R}$. We will use $M_B(E)$ to denote the vector subspace of bounded measurable functions on E.

Proposition 4.3.2. *Let E be a measurable set of finite measure. The map $f \mapsto \int_E f$ is a* ***positive linear functional*** *on $M_B(E)$. That is, the integral is a mapping from $M_B(E)$ into the real numbers that sends nonnegative functions to nonnegative real numbers. Moreover, given f and g in $M_B(E)$ and numbers α, $\beta \in \mathbb{R}$, $\int(\alpha f+\beta g) = \alpha\int f+\beta\int g$. It now follows that if $f \leq g$ on E, then $0 \leq \int(g-f)$, whence $\int f \leq \int g$.*

Proof. Fix f and g in $M_B(E)$, and fix α, β in $\mathbb{R}$. Then for $\alpha > 0$,

$$\alpha\cdot\int f = \alpha\left(\inf_{\psi\in\mathscr{U}(f)}\int\psi\right) = \inf_{\psi\in\mathscr{U}(f)}\int\alpha\psi = \inf_{\psi\in\mathscr{U}(\alpha f)}\int\psi = \int\alpha f.$$

If $\alpha = 0$, $\alpha\int f = \int\alpha f = 0$. If $\alpha < 0$, then

$$\begin{aligned}\int\alpha f &= \int(-\alpha)(-f) = -\alpha\int(-f)\\ &= -\alpha\left(\sup_{\varphi\in\mathscr{V}(-f)}\int\varphi\right) = -\alpha\cdot\left(-\inf_{\psi\in\mathscr{U}(f)}\int\psi\right) = \alpha\int f.\end{aligned}$$

If $\psi_1 \in \mathscr{U}(f)$ and $\psi_2 \in \mathscr{U}(g)$, then $\psi_1+\psi_2 \in \mathscr{U}(f+g)$, so it follows from Proposition 4.2.2 that

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

Since this is true for any $\psi_1 \in \mathscr{U}(f)$ and $\psi_2 \in \mathscr{U}(g)$, we have $\int(f+g) \leq \int f+\int g$. Similarly, if $\varphi_1 \in \mathscr{V}(f)$ and $\varphi_2 \in \mathscr{V}(g)$, then $\varphi_1+\varphi_2 \in \mathscr{V}(f+g)$, so $\int(f+g) \geq \int\varphi_1+\int\varphi_2$, whence $\int(f+g) \geq \int f+\int g$, and so $\int(f+g) = \int f+\int g$. If $g \geq 0$ on E, then $0 \in \mathscr{V}(g)$, so $0 \leq \int g$.

Corollary 4.3.1. *Fix f and g in $M_B(E)$.*

a) *Changing the value of f on a null set $A \subset E$ does not change the value of $\int f$.*
b) *If $f \leq g$ a.e. on E, then $\int f \leq \int g$.*
c) *If $f = g$ a.e. on E, then $\int f = \int g$.*
d) $|\int f| \leq \int |f|$.
e) *If $s \leq f \leq S$ on E, then $s \cdot m(E) \leq \int_E f \leq S \cdot m(E)$.*
f) *If A and B are disjoint measurable subsets of E, then $\int_{A \cup B} f = \int_A f + \int_B f$.*

Proof. Exercise 4.4.

Many theorems concerning the integral have the form "The integral of the limit is the limit of the integral." Most basic of these is the following consequence of Egoroff's theorem.

Theorem 4.3.2 (Bounded Convergence). *Suppose E is a measurable set of finite measure and $\langle f_n : n \in \mathbb{N} \rangle$ is a uniformly bounded sequence of measurable functions converging a.e. to f on E, then $\int f = \lim_{n\to\infty} \int f_n$.*

Proof. By assumption, there is a constant $M > 0$ such that $|f_n| \leq M$ for every $n \in \mathbb{N}$. Replacing the limit f with $-M \vee f \wedge M$ changes f only on a null set. Therefore, we may also assume that $|f| \leq M$. Fix $\varepsilon > 0$. By Egoroff's theorem, there is a measurable set $B \subseteq E$ with $m(B) < \varepsilon/(3M)$ such that f_n converges uniformly to f on $E \setminus B$. Now for every $n \in \mathbb{N}$,

$$\left| \int_E f_n - \int_E f \right| \leq \int_E |f_n - f| \leq \int_{E\setminus B} |f_n - f| + \int_B |f_n| + \int_B |f| .$$

The result follows from the inequalities $\int_B |f_n| < \varepsilon/3$, $\int_B |f| < \varepsilon/3$, and

$$\int_{E\setminus B} |f_n - f| \leq \sup_{E\setminus B} |f_n - f| \cdot m(E) .$$

Example 4.3.1. An unbounded sequence for which the result fails is the mapping $n \mapsto n \cdot \chi_{(0,1/n]}$. It is an exercise (4.7) to show that the theorem fails for the Riemann integral.

4.4 The Integral of Nonnegative Measurable Functions

Extending the integral to unbounded functions defined on sets of infinite measure raises the problem of dealing with infinities. That is, $+\infty$ and $-\infty$ should not appear together in the calculation of an integral. For the purpose of integration, therefore, we will decompose any measurable function f into the difference

$(f \vee 0) - (-f \vee 0)$, and define the integral for each of the nonnegative functions in the decomposition. We note again that the product of 0 with either $+\infty$ or $-\infty$ is undefined.

For the rest of this chapter, it will be understood that a measurable function takes extended-real values and is defined on some nonempty measurable subset of $\mathbb{R}$. It is easy to see that the following definition agrees with the definition of the integral we have previously given when both definitions apply.

Definition 4.4.1. Let f be a nonnegative measurable function defined on $E \subseteq \mathbb{R}$. Let $\mathscr{W}(f)$, or just $\mathscr{W}$, denote the class of all bounded measurable functions h defined on E such that h vanishes off of a set of finite measure and $0 \le h \le f$. Then the integral $\int_E f := \sup_{h \in \mathscr{W}} \int_E h \le +\infty$. If A is a measurable subset of E and $g = f$ on A, while $g = 0$ on $E \setminus A$, then $\int_A f := \int_E g$. In particular, $\int_\varnothing f = 0$.

Proposition 4.4.1. *Let f and g be nonnegative measurable functions, both defined on $E \subseteq \mathbb{R}$, and fix $c > 0$ in $\mathbb{R}$. Then $\int cf = c \int f$, and $\int (f+g) = \int f + \int g$. Changing the value of f on a null set does not change the value of its integral. If $g \ge f$ a.e., then $\int g \ge \int f$, and if $g = f$ a.e., then $\int g = \int f$.*

Proof. Since $\mathscr{W}(cf) = \{ch : h \in \mathscr{W}(f)\}$, $\int cf = c \int f$. If $h \in \mathscr{W}(f)$ and $k \in \mathscr{W}(g)$, then $h + k \in \mathscr{W}(f+g)$, so $\int (f+g) \ge \int h + \int k$, whence $\int (f+g) \ge \int f + \int g$. Now fix $q \in \mathscr{W}(f+g)$. Let $h = q \wedge f$, so $h \in \mathscr{W}(f)$. Note that $h \le q$, and set $k = q - h$, so $0 \le k \le q$. If at x we have $q(x) < f(x)$, then $h(x) = q(x)$ and $k(x) = 0 \le g(x)$. If at x we have $q(x) \ge f(x)$, then $f(x) < +\infty$ and

$$k(x) = q(x) - h(x) = q(x) - f(x) \le (f+g)(x) - f(x) = g(x).$$

It follows that $k \in \mathscr{W}(g)$. Therefore, $\int f + \int g \ge \int h + \int k = \int q$. Since q is arbitrary in $\mathscr{W}(f+g)$, $\int f + \int g \ge \int (f+g)$, whence the inequality is in fact equality. It is an exercise (4.14) to prove that the integral of f is not changed if the value of f is changed on a set of measure 0. If $g \ge f$ a.e., we may assume that $g \ge f$ everywhere without changing the integrals. Since $\mathscr{W}(g) \supseteq \mathscr{W}(f)$, $\int g \ge \int f$. It follows that if $g = f$ a.e. then $\int g = \int f$.

Remark 4.4.1. If $f \ge 0$ and $\int f < +\infty$, then f is finite outside of a null set. Therefore, $0 \cdot f(x) = 0$ a.e., and of course $0 \cdot \int f = 0$.

The program of comparing limits of functions and limits of integrals continues with the next result called "Fatou's Lemma." Despite being called a lemma, the result is basic and very important in working with integration. Recall that for a sequence of functions f_n, $\liminf_n f = \underline{\lim} f_n = \vee_n (\wedge_{k \ge n} f_k)$.

Theorem 4.4.1 (Fatou's Lemma). *Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of nonnegative measurable functions on $E \subseteq \mathbb{R}$. Then*

$$\int_E \underline{\lim}\, f_n \le \underline{\lim} \int_E f_n.$$

Proof. Fix $h \in \mathscr{W}(\underline{\lim} f_n)$. For each $n \in \mathbb{N}$, set $h_n = h \wedge f_n$. For each $x \in E$, $h_n(x) \leq h(x) \leq \underline{\lim} f_n(x)$. Moreover, $h_n(x) \to h(x)$ since $\underline{\lim} f_n(x)$ is the smallest cluster point of the sequence $\langle f_n(x) \rangle$, and so for any $\varepsilon > 0$, there are at most a finite number n's with

$$h_n(x) = f_n(x) \leq h(x) - \varepsilon \leq \underline{\lim} f_n(x) - \varepsilon.$$

Now h is bounded, and it vanishes off of a set of finite measure. Also, $0 \leq h_n \leq h$, and $0 \leq h_n \leq f_n$ for each $n \in \mathbb{N}$. It follows from the Bounded Convergence Theorem that

$$\int h = \lim \int h_n \leq \underline{\lim} \int f_n.$$

Therefore,

$$\int \underline{\lim} f_n := \sup_{h \in \mathscr{W}(\underline{\lim} f_n)} \int h \leq \underline{\lim} \int f_n.$$

Remark 4.4.2. If $f_n \to f$ a.e. on E, then $f = \underline{\lim} f_n$ a.e. and $\int_E f = \int_E \underline{\lim} f_n \leq \underline{\lim} \int_E f_n$. Even here, the inequality is not necessarily equality. An example using Lebesgue measure is provided by the functions $f_n = \chi_{[n,+\infty)}$ on $\mathbb{R}$. This is a good example to keep in mind to remember which way the inequality goes in Fatou's Lemma.

Theorem 4.4.2 (Monotone Convergence Theorem). *Let $\langle f_n : n \in \mathbb{N} \rangle$ be an increasing sequence of nonnegative measurable functions on a fixed set E; that is, for each $n \in \mathbb{N}$, $f_n \leq f_{n+1}$. Let f denote the extended-real valued limit. Then*

$$\int f = \sup_{n \in \mathbb{N}} \int f_n = \lim_{n \to \infty} \int f_n.$$

Proof. For each $n \in \mathbb{N}$, $f_n \leq \sup_{n \in \mathbb{N}} f_n = \underline{\lim} f_n$, so $\int f_n \leq \int \underline{\lim} f_n$. Therefore, $\overline{\lim} \int f_n \leq \int \underline{\lim} f_n$. The result now follows from Fatou's Lemma since

$$\overline{\lim} \int f_n \leq \int \underline{\lim} f_n = \int f \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n.$$

Corollary 4.4.1. *If $\langle u_n : n \in \mathbb{N} \rangle$ is a sequence of nonnegative measurable functions on a fixed set E, then $\int \left(\sum_{n \in \mathbb{N}} u_n \right) = \sum_{n \in \mathbb{N}} \int u_n$.*

Proof. The result follows by setting $f_n = \sum_{i=1}^{n} u_i$ for each $n \in \mathbb{N}$.

Corollary 4.4.2. *If f is a nonnegative measurable function and $\langle E_i : i \in \mathbb{N} \rangle$ is a pairwise disjoint sequence of measurable sets in the domain of f, then $\int_{\cup E_i} f = \sum_{i=1}^{\infty} \int_{E_i} f$.*

Proof. For each $i \in \mathbb{N}$, let $A_i = E_i \cap f^{-1}[+\infty]$. If for some i, $m(A_i) > 0$, then both sides of the desired equality have the value $+\infty$. If $m(A_i) = 0$ for all i, then without changing any of the integrals for f, we may set $f = 0$ on $\cup_i A_i$. The result then follows from Corollary 4.4.1 with $u_n = f \cdot \chi_{E_n}$.

Corollary 4.4.3. *If f is a nonnegative measurable function on a measurable set $E \subseteq \mathbb{R}$, then the map $A \mapsto \int_A f$ is an extended-real valued measure, called the measure generated by f. It is defined for all measurable subsets of E. In particular, if A and B are measurable subsets of E with $A \subseteq B$, then $\int_A f \leq \int_B f$.*

Definition 4.4.2. Given a nonnegative measurable function f, we say that f is **integrable** on a measurable set $E \subseteq \mathbb{R}$ if $\int_E f < +\infty$. If E is the domain of f, we simply say that f is integrable.

Proposition 4.4.2. *If f is integrable on E and g is measurable with $0 \leq g \leq f$ on E, then g is integrable on E. Moreover, $\int_E g \leq \int_E f$ and*

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

Proof. This follows since $\int_E g + \int_E (f-g) = \int_E f < +\infty$.

Next we show that the measure generated by a nonnegative integrable function f has a continuity property with respect to the measure m. The property is obvious if f is bounded.

Proposition 4.4.3. *Let $f \geq 0$ be integrable on $E \subseteq \mathbb{R}$, and let ν be the finite measure on the measurable subsets of E generated by f (see Corollary 4.4.3). For any $\varepsilon > 0$, there is a $\delta > 0$ such that if A is a measurable subset of E with $m(A) < \delta$, then $\nu(A) = \int_A f < \varepsilon$.*

Proof. Let $f_n = f \wedge n$. By the Monotone Convergence Theorem, since $f_n \nearrow f$, the difference of the integrals $(\int_E f - \int_E f_n) \searrow 0$. We may choose N so large in $\mathbb{N}$ that the difference is at most $\varepsilon/2$. Fix $\delta = \varepsilon/(2N)$. For any measurable set $A \subseteq E$ with $m(A) < \delta$,

$$\nu(A) = \int_A f = \int_A (f - f_N) + \int_A f_N \leq \frac{\varepsilon}{2} + m(A) \cdot N < \varepsilon.$$

4.5 The Integral of Measurable Functions

We will use the notation $f^+ := f \vee 0$, and $f^- := -f \vee 0$, so $f = f^+ - f^-$, and $|f| = f^+ + f^-$. The following definition and results are given in terms of an arbitrary integrator and corresponding measure m.

Definition 4.5.1. A measurable function f is integrable on a measurable set $E \subseteq \mathbb{R}$ if f^+ and f^- are integrable on E. In this case, set $\int f := \int f^+ - \int f^-$. If only one of the integrals of f^+ and f^- is finite, the function f is not integrable, but the integral of f is the difference of the two integrals, and is therefore either $+\infty$ or $-\infty$. A function f is called "integrable" only if it is measurable.

Proposition 4.5.1. *A measurable function f is integrable on a measurable set $E \subseteq \mathbb{R}$ if and only if $|f|$ is integrable on E. In this case, f is finite almost everywhere on E.*

Proof. Exercise 4.6.

Example 4.5.1. Suppose E is a measurable set of finite measure. Let $B \subseteq E$ be nonmeasurable with positive outer measure. If $f = \chi_B - \chi_{E\setminus B}$, then $|f| = \chi_E$ is integrable, but f is not measurable.

Integrability and the corresponding integral have been defined for a measurable function f in terms of f^+ and f^-. Of course, $f = f^+ - f^-$ is not the only way that a function f may be presented as a difference of two functions. The following proof demonstrates a useful general technique for working with two equivalent differences.

Proposition 4.5.2. *Let g and h be nonnegative measurable functions on $E \subseteq \mathbb{R}$ such that $g^{-1}[+\infty] \cap h^{-1}[+\infty] = \varnothing$. Let $f = g - h$ a.e. If g and h are integrable, then f is integrable and $\int f = \int g - \int h$. If $\int g = +\infty$ and h is integrable, then $\int f = +\infty$. If g is integrable and $\int h = +\infty$, then $\int f = -\infty$.*

Proof. If g and h are both integrable, we may assume, without changing the values of the integrals, that g and h are finite everywhere on E, and $f = g - h$ everywhere on E. Now $f^+ = g - (g \wedge h)$ and $f^- = h - (g \wedge h)$, so f^+ and f^- are integrable. Since $g - h = f = f^+ - f^-$, we have $g + f^- = h + f^+$, so $\int g + \int f^- = \int h + \int f^+$, whence $\int g - \int h = \int f^+ - \int f^- = \int f$. If $\int g = +\infty$ and h is integrable, we may assume that h is finite everywhere. Moreover, $g \wedge h$ and $f^- = h - (g \wedge h)$ are integrable. Now $f^+ = g - (g \wedge h)$. Suppose f^+ is integrable; then, we may assume that f^+ is finite everywhere, whence g is finite everywhere, and $g = f^+ + (g \wedge h)$ is integrable. Since g is not integrable, it follows that $\int f^+ = +\infty$. Applying this result to $-f$ finishes the proof.

Proposition 4.5.3. *The family of real-valued functions integrable on $E \subseteq \mathbb{R}$ is a vector space, and the mapping $f \mapsto \int f$ is a positive linear functional on that space. In particular, this means that if $f \geq 0$ a.e. on E, then $\int f \geq 0$.*

Proof. Assume f and g are real-valued integrable functions. If $c \geq 0$, then $cf = cf^+ - cf^-$, so cf is integrable and $\int cf = c\int f^+ - c\int f^- = c\int f$. If $c < 0$, then $cf = (-c(-f))$ is integrable and $\int cf = -c\int -f = (-c)(-\int f) = c\int f$. By Proposition 4.5.2, the sum $f + g = (f^+ + g^+) - (f^- + g^-)$ is integrable and

$$\int (f+g) = \int f^+ - \int f^- + \int g^+ - \int g^- = \int f + \int g.$$

If $f \geq 0$ a.e., then $\int f = \int f^+ \geq 0$.

Proposition 4.5.4. *If $g \geq 0$ is integrable on $E \subseteq \mathbb{R}$ and f is measurable with $|f| \leq g$ a.e., then f is integrable on E.*

Proof. We may assume $|f| = f^+ + f^- \leq g$ at all points. The rest is clear.

Remark 4.5.1. The space of integrable functions on $E \subseteq \mathbb{R}$ is not quite a vector space because an integrable function may be infinite on a set of measure 0. Such functions can be changed on a set of measure 0 to become real-valued functions without changing the value of the integral. In that sense, the integral is a positive linear functional on the integrable functions on E. Moreover, if f and g are integrable and $f \leq g$ a.e., then $\int (g-f) \geq 0$, so $\int g \geq \int f$.

4.6 Generalization of Fatou's Lemma

In its simplest form, Fatou's Lemma is applicable only to sequences of nonnegative functions. It may, however, be applied to functions that take both positive and negative values by adding on a positive integrable function that lifts the range of the combination to the nonnegative real line. With such domination by an integrable function, we have the following theorem and important corollary.

Theorem 4.6.1 (First General Fatou Lemma). *Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of measurable functions on a measurable set E and $g \geq 0$ an integrable function on E such that for any $n \in \mathbb{N}$, $|f_n| \leq g$ a.e. Then*

$$\int \underline{\lim}\, f_n \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n \leq \int \overline{\lim}\, f_n.$$

Proof. This is a corollary of the Second General Fatou Lemma stated below.

Corollary 4.6.1 (Lebesgue Dominated Convergence). *Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of measurable functions on a measurable set E and $g \geq 0$ an integrable function on E such that for each $n \in \mathbb{N}$, $|f_n| \leq g$ a.e. If $f = \lim f_n$ a.e., then $\int f = \lim \int f_n$.*

Proof. When $f = \underline{\lim}\, f_n = \overline{\lim}\, f_n$ a.e., the inequality in Theorem 4.6.1 becomes equality.

As noted, Theorem 4.6.1 is an immediate consequence of an even more general result. That result uses the fact that if $a_n \to a$ and b_n is another sequence, then $\underline{\lim}\,(a_n + b_n) = \lim\,(a_n) + \underline{\lim}\,(b_n)$ and a similar equality holds for $\overline{\lim}\,(a_n + b_n)$.

Theorem 4.6.2 (Second General Fatou Lemma). *Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of measurable functions on a measurable set E, and let $\langle g_n : n \in \mathbb{N} \rangle$ be a sequence of integrable functions on E with $|f_n| \leq g_n$ a.e. for each $n \in \mathbb{N}$. Assume that the sequence g_n converges a.e. to an integrable function g. Also assume that $\int g = \lim_{n\to\infty} \int g_n$. Then*

$$\int \underline{\lim}\, f_n \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n \leq \int \overline{\lim}\, f_n.$$

Proof. We may, without loss of generality, assume finiteness, domination, and convergence at all points $x \in E$. We then have

$$-g = -\lim g_n \leq \underline{\lim}\, f_n \leq \overline{\lim}\, f_n \leq \lim g_n = g,$$

so $\underline{\lim}\, f_n$ and $\overline{\lim}\, f_n$ are integrable. Moreover, for each $n \in \mathbb{N}$, $f_n + g_n \geq 0$, so

$$\begin{aligned}\int g + \int \underline{\lim}\, f_n &= \int (g + \underline{\lim}\, f_n) = \int (\lim\, g_n + \underline{\lim}\, f_n) = \int \underline{\lim}(g_n + f_n) \\ &\leq \underline{\lim} \int (g_n + f_n) = \underline{\lim}(\int g_n + \int f_n) = \lim \int g_n + \underline{\lim} \int f_n \\ &= \int g + \underline{\lim} \int f_n.\end{aligned}$$

It follows, even when the f_n's take both positive and negative values, that $\int \underline{\lim}\, f_n \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n$. The inequality $\overline{\lim} \int f_n \leq \int \overline{\lim}\, f_n$ is an immediate consequence of the result for $\underline{\lim}\, f_n$ applied to the sequence $-f_n$. That is,

$$-\int \overline{\lim}\, f_n = \int \underline{\lim}(-f_n) \leq \underline{\lim}(-\int f_n) = -\overline{\lim} \int f_n.$$

Remark 4.6.1. The first General Fatou Lemma is a corollary of the second by setting $g_n = g$ for each $n \in \mathbb{N}$. Without the assumption $\int g_n \to \int g$ in the second Fatou Lemma, there are such counterexamples as $f_n = g_n = \chi_{[n,n+1]}$ and $f = g = 0$, or $f_n = g_n = n\chi_{(0,1/n]}$ and $f = g = 0$. A more general Dominated Convergence Theorem is also an immediate corollary of Theorem 4.6.2.

4.7 Improper Riemann Integral

The improper Riemann integral of a function f that is continuous Lebesgue a.e. on the interval $[0, +\infty)$ and bounded on bounded subintervals is $\lim_{B \to +\infty} \int_0^B f(x)dx$. If the interval is $(-\infty, +\infty)$, then the improper Riemann integral of f is given by $\lim_{\substack{A \to -\infty \\ B \to +\infty}} \int_A^B f(x)dx$. The values of A and B must tend independently to their respective limits. For example, consider the integral of $f(x) = x$ on the real line. Similarly, if f has a problem at 0 (we use the number 0 to illustrate) and f is appropriate on $(0, +\infty)$, then the improper Riemann integral of f is $\lim_{\substack{a \to 0+ \\ B \to +\infty}} \int_a^B f(x)dx$. Again, the values of a and B must tend independently to their respective limits.

Because of cancellation, the improper Riemann integral may exist as a finite number without f being Lebesgue integrable. For example, for the function

$$f(x) := \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \cdot \chi_{[n-1,n)},$$

the improper Riemann integral on $[0,+\infty)$ looks like the sum of the alternating harmonic series, while the integrals of f^+ and f^- look like the sums of the even and odd terms of the harmonic series. The proof of the following result is Exercise 4.31.

Theorem 4.7.1. *Assume f is both Riemann integrable and Lebesgue integrable on an interval I over which the Riemann integral $R\int_I f$ is improper. Then $R\int_I f = \int_I f$.*

4.8 Convergence in Measure

Again in this section, we work with a measure m derived from a general integrator. There is a notion of convergence that has many applications but is weaker than convergence almost everywhere. It is especially important in probability theory.

Definition 4.8.1. A sequence $\langle f_n : n \in \mathbb{N}\rangle$ of measurable functions converges to 0 in measure if for any $\varepsilon > 0$, $m\{|f_n| > \varepsilon\} \to 0$. In general, f_n converges to a function f in measure means that f is measurable and finite almost everywhere, and $|f_n - f|$ converges to 0 in measure.

Remark 4.8.1. One can also define the convergence of f_n to f in measure with the condition that for any $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$\forall n \geq N,\ \ m\{|f_n - f| > \varepsilon\} < \varepsilon.$$

Note that the values taken by $|f_n - f|$ that are greater than ε have no influence on the measure of the set $\{|f_n - f| > \varepsilon\}$.

By Egoroff's Theorem (3.3.4), if $f_n \to f$ a.e. on a set E of finite measure, then $f_n \to f$ in measure. For a set of infinite measure this is not true. For example, the sequence $n \mapsto \chi_{[n,n+1]}$ converges to 0 pointwise, but not in measure. On the other hand, convergence in measure, even on a set of finite measure, does not imply convergence almost everywhere. For example, let

$$f_1 = \chi_{[0,1]},\ \ f_2 = \chi_{[0,1/2]},\ \ f_3 = \chi_{[1/2,1]},\ \ f_4 = \chi_{[0,1/4]},\ \ f_5 = \chi_{[1/4,1/2]},\ \text{etc.}$$

Then f_n converges to 0 in measure, but for every $x \in [0,1]$, $\overline{\lim} f_n(x) = 1$ while $\underline{\lim} f_n(x) = 0$; that is, we have convergence at no point. We do, however, have almost everywhere convergence for a subsequence.

Theorem 4.8.1. *Let $\langle f_n : n \in \mathbb{N}\rangle$ be a sequence of measurable functions converging in measure to f on a measurable set E. Then there is a subsequence $\langle f_{n_k} : k \in \mathbb{N}\rangle$ that converges to f almost everywhere on E.*

Proof. Set $n_0 = 0$. For each $k \in \mathbb{N}$, we can choose $g_k = \left|f_{n_k} - f\right|$ so that $n_k > n_{k-1}$ and $m\{g_k > 1/2^k\} < 1/2^k$. We need only show that $g_k \to 0$ a.e. Let $A_k = \{g_k > 1/2^k\}$, and mirror the Borel-Cantelli Lemma (2.6.2). That is, for

$$A = \overline{\lim} A_k = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k,$$

we have $A = \{x \in E : x \text{ is in an infinite number of the } A_k\}$. Now for each $n \in \mathbb{N}$,

$$m(A) \leq m(\cup_{k=n}^{\infty} A_k) \leq \sum_{k=n}^{\infty} 1/2^k = 2/2^n,$$

and so $m(A) = 0$. If $x \notin A$, then for some $n \in \mathbb{N}$, $x \notin \cup_{k=n}^{\infty} A_k$, and so for all $k \geq n$, $0 \leq g_k(x) \leq 1/2^k$. It follows that $g_k(x) \to 0$ for all $x \notin A$.

Remark 4.8.2. For a set E of finite measure, there is a distance function that measures how far apart two functions f and g are in measure. It is given by the mapping $(f, g) \mapsto \int_E |f - g| \wedge 1$. We will develop such notions of distance later when we discuss metric spaces.

4.9 Problems

Problem 4.1. Using just Riemann sums and properties of continuous functions, show that any continuous real-valued function f on $[a, b]$ is Riemann integrable on $[a, b]$.

Problem 4.2. Let ψ be the step function on $[0, 7]$ given by the sum

$$\psi = 2 \cdot \chi_{[0,3]} + 3 \cdot \chi_{[2,4]} + 5 \cdot \chi_{[6,7]}.$$

a) Find the minimal representation of the step function ψ.
b) The function ψ is a simple function. Find its canonical form as a simple function.

Problem 4.3 (A). Given a bounded measurable function f on a set E of finite measure, show that for any measurable subset A of E, $\int_A f := \int f \cdot \chi_A$ has the same value as the integral obtained by restricting f to A.

Problem 4.4. Prove Corollary 4.3.1.

Problem 4.5. Let $f = \chi_{\{0\}}$; that is, $f(0) = 1$, and $f(x) = 0$ for $x \neq 0$. Let $g = \chi_{\{0,1\}}$. Recall that for each $x \in \mathbb{R}$, $g \circ f(x) = g(f(x))$. Show that $g \circ f$ is not Lebesgue integrable.

Problem 4.6. Prove Proposition 4.5.1.

Problem 4.7. a) Use upper and lower Riemann sums to show that the characteristic function of the rational numbers in $[0, 1]$ is not Riemann integrable. **b)** Show that the Bounded Convergence Theorem does not hold for the Riemann integral.

Problem 4.8. Suppose that $f : (0, \infty) \mapsto (0, \infty)$ is Lebesgue integrable. Show that there exists a sequence $\langle x_k \rangle$ increasing to $+\infty$ such that $\lim_{k \to \infty} x_k \cdot f(x_k) = 0$. **Hint:** It suffices to show that $\underline{\lim}_{x \to +\infty} x \cdot f(x) = 0$.

Problem 4.9. Recall the generalized Cantor sets of positive measure constructed in Section 2.8. Let $E \subset [0,1]$ be such a set with Lebesgue measure $1/2$.

a) Use the construction of E to form a decreasing sequence $\langle \varphi_n \rangle$ of step functions on $[0,1]$ such that $\lim_{n\to\infty} \varphi_n(x) = \chi_E(x)$ for all $x \in [0,1]$. Show that

$$\lim_{n\to\infty} R\int_0^1 \varphi_n(x)\,dx = \lambda(E).$$

b) Show that χ_E is not Riemann integrable on $[0,1]$.

Problem 4.10. Prove or give a counterexample: If f_n is a sequence of **negative-valued**, Lebesgue measurable functions on $[0,+\infty)$, then $\limsup \int f_n \le \int \limsup f_n$.

Problem 4.11 (A). Prove the following proposition: If f is measurable but not integrable on $\mathbb{R}$, and $\int f = +\infty$, and if h is integrable on $\mathbb{R}$, then $\int (f+h) = \int f + \int h$.

Problem 4.12. Let m be the measure generated by an integrator F. Fix a nonnegative measurable function f defined on a measurable set $E \subseteq \mathbb{R}$. Recall Definition 4.4.1: $\int_E f\,dm := \sup_{h\in\mathscr{W}(f)} \int_E h\,dm$. Show that

$$\int_E f\,dm = \sup_{0\le\varphi\le f,\ \varphi \text{ simple}} \int \varphi\,dm,$$

where the supremum runs over all nonnegative simple functions dominated by f. **Hint:** Prove the result for any $h \in \mathscr{W}$.

Problem 4.13. Let $f(x) = e^{-x^2}$ on $\mathbb{R}$. It is well-known that $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$. Show that $\int_{\mathbb{R}} \varphi\,d\lambda = +\infty$ for any simple function φ dominating f on $\mathbb{R}$. Compare with Problem 4.12.

Problem 4.14. Let f be a nonnegative measurable function on $E \subseteq \mathbb{R}$. Prove that the integral of f is not changed if the value of f is changed on a null set even if f is identically equal to $+\infty$ on that null set. **Hint:** Suppose the value of f is changed to 0 on a null set.

Problem 4.15. Simplify the proof of Theorem 4.6.2 to give a direct proof of Theorem 4.6.1.

Problem 4.16. a) Prove the following **inequality of Chebyshev**: If $f \ge 0$ is measurable, then for any $\alpha > 0$, $m\{f \ge \alpha\} \le \frac{1}{\alpha}\int f$.
b) Prove the following result: If $f \ge 0$ is integrable and $\int f = 0$, then $f = 0$ a.e.

Problem 4.17. Let $\langle f_n : n \in \mathbb{N}\rangle$ be a sequence of real-valued integrable functions on X. Suppose that for some integrable function f, we have $\int_X |f_n(t) - f(t)| \le 2^{-n}$ for all $n \in \mathbb{N}$. Show that $f_n \to f$ a.e. **Hint:** $\sum_{k=1}^{\infty} |f_{k+1}(t) - f_k(t)| < \infty$ a.e., whence $f_n(t) = f_1(t) + \sum_{k=2}^{n} (f_k(t) - f_{k-1}(t))$ converges a.e. Moreover, $f_n \to f$ in measure.

Problem 4.18. Let m be a measure on $\mathbb{R}$. Show that a measurable function $f : \mathbb{R} \mapsto [0,\infty)$ is integrable if the series $\sum_{n=0}^{\infty} m(\{x : f(x) \ge n\})$ converges.

Problem 4.19. Fix a function $f \geq 0$ that is integrable on $\mathbb{R}$. Set $F(x) = \int_{(-\infty,x]} f$. Clearly F is an increasing function of x. Show that F is **uniformly** continuous on $\mathbb{R}$.

Problem 4.20. Suppose $\langle f_n \rangle$ is a sequence of nonnegative measurable functions that converge pointwise to f, but not necessarily monotonically, i.e., we don't necessarily have $f_n(x) \leq f_{n+1}(x)$ or $f_n(x) \geq f_{n+1}(x)$ for all n and x. Suppose, however, that $f_n \leq f \; \forall n$. Show that $\lim \int f_n$ exists and equals $\int f$.

Problem 4.21. a) Show that the Monotone Convergence Theorem may not hold for a decreasing sequence.
b) Give a sufficient condition on the first function in the decreasing sequence for the result to hold.

Problem 4.22 (A). Prove the following result: Suppose that $\langle f_n \rangle$ is a sequence of nonnegative measurable functions on $\mathbb{R}$ such that $f_n \to f$ a.e. and $\int_{\mathbb{R}} f_n \to \int_{\mathbb{R}} f < +\infty$. Then for each measurable set E, $\int_E f_n \to \int_E f$.

Problem 4.23. Let $f : \mathbb{R} \mapsto [0,\infty)$ be a Lebesgue integrable function and $g : \mathbb{R} \mapsto [a,b]$ be a measurable function. Show that there is $c \in [a,b]$ such that $\int fg = c \int f$.

Problem 4.24. Show that for any integer $n \geq 2$, the unbounded function $x^{-1/n}$ is Lebesgue integrable on $(0,1]$.

Problem 4.25. Let $\langle f_n \rangle$ be a sequence of **nonnegative**, Lebesgue integrable functions on $[0,1]$, such that for Lebesgue measure λ, $\lim_{n\to\infty} \int_{[0,1]} f_n \, d\lambda = 0$. Must the sequence f_n converge to 0 in measure on $[0,1]$? Briefly justify your answer.

Problem 4.26. Let f be an integrable function on $[0,1]$.

a) Show that if $f > 0$ on a set $F \subseteq [0,1]$ of positive measure, then $\int_F f(x)\, m(dx) > 0$.
b) Suppose for each $x \in [0,1]$, $m(\{x\}) = 0$, and for each $y \in [0,1]$, $\int_0^y f(x)\, m(dx) = 0$. Show that $f(x) = 0$ for m-a.e. $x \in [0,1]$.

Problem 4.27. Let $\langle f_i \rangle$ be a sequence of Lebesgue measurable functions on $[0,1]$. Show that if f_i converges to f in (Lebesgue) measure on $[0,1]$, then f is Lebesgue measurable.

Problem 4.28. Suppose $\langle f_n \rangle$ is a sequence of nonnegative measurable functions converging to f in measure. Show that $\int f \leq \underline{\lim} \int f_n$.

Problem 4.29. a) Let $\langle f_n \rangle$ be a sequence of measurable real-valued functions on a set $E \subseteq \mathbb{R}$ of finite measure. Let f be a measurable real-valued function on E. Show that $f_n \to f$ in measure if and only if every subsequence $\langle f_{n_k} \rangle$ has a further subsequence that converges to f pointwise almost everywhere on E.
b) Now also suppose that $g_n \to g$ in measure on E. Show that $f_n g_n \to fg$ in measure.

Problem 4.30. In the setting of Problem 2.18, let f be a nonnegative, real-valued function on $(0,\infty)$. Show that $\int_{(0,\infty)} f \, dm = \sum_{n\in\mathbb{N}} f(n)$.

Problem 4.31 (A). Prove Theorem 4.7.1.

Problem 4.32 (A). Use the Second General Fatou Lemma (4.6.2) to prove the following result: Let $\langle f_n \rangle$ be a sequence of integrable functions with $f_n \to f$ a.e., and assume f is integrable. Then $\int |f_n - f| \to 0$ if and only if $\int |f_n| \to \int |f|$.

Problem 4.33. Let $f_n : [0,1] \mapsto [0,\infty)$ be a sequence of Lebesgue integrable functions on $[0,1]$ converging Lebesgue a.e. to f. Show that $\lim_{n\to\infty} \int_0^1 f_n e^{-f_n} d\lambda = \int_0^1 f e^{-f} d\lambda$. **Hint:** Where is the maximum value of $x \mapsto xe^{-x}$?

Problem 4.34. Evaluate $\lim_{n\to\infty} \int_0^\infty \left(\frac{\sin x}{x}\right)^n dx$.

Problem 4.35. Let $f : [0,1] \mapsto \mathbb{R}$ be a continuous function. Using Lebesgue measure λ, show that $\lim_{n\to\infty} \int_0^1 x^n f(x)\, \lambda(dx) = 0$.

Problem 4.36. Let $\langle f_n \rangle$ be a sequence of measurable functions with $f_n : \mathbb{R} \mapsto [0,\infty)$ for each $n \in \mathbb{N}$ and $f_n \to f$ a.e. Suppose $\lim_{n\to\infty} \int f_n(x) = 0$. Show that $f(x) = 0$ a.e.

Problem 4.37. Suppose that $f : [0,1] \mapsto \mathbb{R}$ and $h : [0,1] \mapsto \mathbb{R}$ are measurable functions on $[0,1]$ such that $\int_0^1 f = \int_0^1 h$. Show that either $f = h$ a.e. on $[0,1]$ or there exists a measurable set $A \subseteq [0,1]$ such that $\int_A h < \int_A f$.

Problem 4.38. Let f be a nonnegative Lebesgue measurable function on $X \subseteq \mathbb{R}$ such that $\int_X f^2 d\lambda < \infty$. Show that $\lim_{b\to\infty} b^2 \cdot \lambda(\{f \geq b\}) = 0$. **Hint:** For each $n \in \mathbb{N} \cup \{0\}$, let $E_n = \{x \in X : n \leq f < n+1\}$. Note that $\int_X f^2 d\lambda \geq \sum_{n=0}^\infty n^2 \lambda(E_n)$.

Problem 4.39. a) Show that the First General Fatou Lemma holds for a uniformly bounded sequence of measurable functions defined on a fixed set of finite measure.

b) Give an example of a sequence $\langle f_n : n \in \mathbb{N} \rangle$ of measurable functions on $[0,2]$ taking the values 1 and -1 such that for all $n \in \mathbb{N}$,

$$\int \underline{\lim}\, f_n \, d\lambda < \underline{\lim} \int f_n \, d\lambda < \overline{\lim} \int f_n \, d\lambda < \int \overline{\lim}\, f_n \, d\lambda.$$

This shows that in general, strict inequality may hold for each of the inequalities in the First General Fatou Lemma. **Hint:** The sequence can be chosen so that $f_n = f_{n+4}$ for all $n \in \mathbb{N}$.

Problem 4.40. a) Using Lebesgue measure, prove the following version of the **Riemann-Lebesgue Lemma**, which is important in Fourier analysis: If f is integrable on a finite closed interval $[a,b]$, then taking the limit as $n \to \infty$,

$$\int_{[a,b]} f(x) \cdot \cos(nx) \lambda(dx) \to 0, \quad \text{and} \quad \int_{[a,b]} f(x) \cdot \sin(nx) \lambda(dx) \to 0.$$

Hint: Approximate f with a step function.

b) Show that if f is integrable with respect to Lebesgue measure on the real line $\mathbb{R}$, then $\int_{\mathbb{R}} f(x) \cdot \cos(nx) \lambda(dx) \to 0$.

Problem 4.41. For each of the following statements, state whether it is true or false and quote a theorem or give an example as a reason for your answer:

a) There is a sequence of simple functions f_n taking only the values 0 and 1 defined on the interval $[0,1]$ such that for each $x \in [0,1]$, the sequence $f_n(x)$ converges to 0, but the sequence $\int_0^1 f_n(x)dx$ does not converge to 0.
b) There is a sequence of simple functions f_n taking only the values 0 and 1 defined on the interval $[0,1]$ such that for each $x \in [0,1]$, the sequence $f_n(x)$ **does not converge**, but the sequence $\int_0^1 f_n(x)dx$ converges to 0.
c) There is a sequence of simple functions f_n taking only the values 0 and 1 defined on the interval $[0,+\infty)$ such that for each $x \in [0,+\infty)$, the sequence $f_n(x)$ converges to 0, but the sequence $\int_{[0,+\infty)} f_n(x)dx$ does not converge to 0.
d) There is a sequence of simple functions f_n taking only the values 0 and 1 defined on the interval $[0,+\infty)$ such that the sequence f_n converges to 0 in measure, but the sequence $\int_{[0,+\infty)} f_n(x)dx$ does not converge to 0.

Problem 4.42. Let $\langle f_n \rangle$, $\langle g_n \rangle$, and $\langle h_n \rangle$ be sequences of real-valued μ-integrable functions converging μ-a.e. on $\mathbb{R}$ to μ-integrable functions f, g, and h, respectively, where $f_n \le g_n \le h_n$ for all $n \in \mathbb{N}$. Assume that $\lim_{n\to\infty} \int f_n(x)d\mu = \int f(x)d\mu$ and $\lim_{n\to\infty} \int h_n(x)d\mu = \int h(x)d\mu$. Show that $\lim_{n\to\infty} \int g_n(x)d\mu = \int g(x)d\mu$.

Problem 4.43. Recall that the continuity property of the mapping $A \mapsto \int_A g$ for a nonnegative integrable g is the property that for any $\varepsilon > 0$, there is a $\delta > 0$ such that if $m(A) < \delta$, then $\int_A g < \varepsilon$. Use this property for such a function g to show that for a sequence $\langle f_n \rangle$ of measurable functions with $|f_n| \le g$ for all n, if f_n converges in measure to f, then $\int |f_n - f| \to 0$. **Hint:** First show that $|f_n - f| \le 2g$ a.e. You may then, after subtracting f, assume that for this problem, $0 \le f_n \le g$ for all n and $f_n \to 0$ in measure. You then want to show that $\int f_n \to 0$.

Problem 4.44 (A). Suppose $f(x,y)$ is defined on the unit square, i.e., the set in the xy-plane with $0 \le x \le 1$ and $0 \le y \le 1$. Also suppose $f(\cdot,y)$ is a measurable function of x for each y and that function $f(\cdot,y)$ is dominated in absolute value by a fixed integrable function g that does not change with y. Finally suppose that $\lim_{y\to 0+} f(x,y) = f(x)$. Show that $\lim_{y\to 0+} \int f(x,y)dx = \int f(x)dx$.

Chapter 5
Differentiation and Integration

5.1 Introduction

In this chapter, we consider under what conditions and to what extent integration and differentiation are inverse operations on a function. We apply new results obtained by the author of this text with J. Bliedner [11]. Those results use "local maximum functions"; they extend and simplify the usual techniques used for the material presented in this chapter.

As before, we use λ to denote Lebesgue measure and m to denote a general measure generated by an integrator F. We write m_F when emphasizing the integrator. Of course, for the integrator F given by $F(x) = x$, the measure $m_F = \lambda$.

Bounded intervals in the real line are the natural setting for differentiation, so we will work in a fixed, bounded, open interval $J = (-K, K)$ where K is a positive real number. We know that for any Borel set $A \subseteq J$, the Lebesgue measure $\lambda(A) \leq 2K$. A general integrator F is always real-valued. We may assume that F is continuous at K and $-K$ since an increasing function can have only a countable number of jumps. (See Problem 5.6.) Therefore, one can always assure continuity at the endpoints of J by slightly stretching or shrinking J. For any Borel set $A \subseteq J$, $m_F(A) \leq F(K) - F(-K)$. We may subtract $F(-K)$ from F to produce a nonnegative integrator on J that generates the same measure as F.

If we start with a finite Borel measure ν already defined on the closure of J, and for any $x \in J$ we set $F(x) = \nu([-K, x])$, then we have an increasing integrator that is continuous from the right since $\nu[-K, x] = \lim_{n \to \infty} \nu[-K, x + 1/n]$. That integrator will generate the measure ν. In this way, we can work with any finite measure defined at least on the Borel sets of J, and indeed on the completion (with respect to the measure) of the Borel sets in J (See Definition 3.2.1). Only finite measures will be considered on J.

We know that any set that is measurable with respect to a finite measure m is the union of an F_σ set and a subset of a Borel set of m-measure 0. Therefore, if we want to put an upper bound on the measure of sets, we need only do so for Borel sets. As before, we say that a property holds Lebesgue almost everywhere if it holds

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_5

except on a set of Lebesgue measure 0. A measure that is of particular interest is the measure λ_f formed using Lebesgue measure λ and a nonnegative function f that is Lebesgue integrable on J. That is, $\lambda_f(A) = \int_A f\, d\lambda$ for any Lebesgue measurable set $A \subseteq J$.

5.2 A Covering Theorem for $\mathbb{R}$

Among the most important tools in analysis are results known as covering theorems; they are used to show that the set where a desired property fails has measure 0. For the real line, the sets used to cover are intervals. Recall that there are four types of finite intervals: open, closed, open at the left and closed at the right, and closed at the left and open at the right. A degenerate interval is a closed interval $[a]$ consisting of only one point. In this section, we present the optimal covering theorem for the real line. The set covered here is the union of the covering intervals. The result is an extension by J. Aldaz [2] of a lemma of T. Radó [41]. We have modified the Aldaz result by using the constant 3 instead of his constant $2+\varepsilon$ for an arbitrary $\varepsilon > 0$. (See Problem 5.1.) For higher-dimensional spaces, one can use geometric results such as the two presented in this book's appendix on limit and covering theorems.

Theorem 5.2.1 (Radó-Aldaz). *Let m be a finite, Borel measure on $J = (-K, K)$. Given an arbitrary collection $\mathscr{I}$ of non-degenerate intervals, all contained in J, the set $\cup_{I\in\mathscr{I}} I$ is measurable, and there is a finite disjoint subset $\{I_1, \cdots, I_n\} \subseteq \mathscr{I}$ such that*

$$m(\cup_{I\in\mathscr{I}} I) \leq 3 \cdot \Sigma_{k=1}^n m(I_k).$$

Proof. By Corollary 1.7.1, we may assume that $\mathscr{I}$ itself is a countable collection given by the enumeration $\{I_n : n \in \mathbb{N}\}$, whence $\cup_{I\in\mathscr{I}} I = \cup_{n=1}^\infty I_n$ is measurable. Since all the intervals are contained in J,

$$m(\cup_{n=1}^\infty I_n) = \lim_{N\in\mathbb{N}} m(\cup_{n=1}^N I_n) \leq m(J) < +\infty.$$

We employ Radó's result after first choosing $N \in \mathbb{N}$ so that

$$\frac{3}{2} \cdot m(\cup_{n=1}^N I_n) \geq m(\cup_{n=1}^\infty I_n) = m(\cup_{I\in\mathscr{I}} I).$$

Given the enumeration of the N intervals, we discard the first interval if it is covered by the remaining intervals. Otherwise, we keep the first interval and consider the second. In either case, the union of the intervals we keep is the same as the union of the original N intervals. Continuing in this way, we may assume that each of the remaining intervals in our finite collection contains a representative point x not in any other interval of the collection. Now we give the finite set of representative points the ordering inherited from $\mathbb{R}$, and we reorder the remaining finite set of intervals so that they have the same ordering as their representative points. It follows that for any indices i, j, and k with $i < j < k$ we have $x_i < x_j < x_k$. Moreover, since

$x_j \notin I_i$, $I_i \subseteq (-\infty, x_j)$, and since $x_j \notin I_k$, $I_k \subseteq (x_j, +\infty)$. Thus, the intervals with even indices form a disjoint collection, as do the intervals with odd indices. Therefore, the desired subset of $\mathscr{I}$ is whichever of these two families has the greater total measure. For example, if that is the family with even indices, then

$$3 \cdot m(\cup_n I_{2n}) = \frac{3}{2} \cdot 2 \cdot m(\cup_n I_{2n}) \geq \frac{3}{2} \cdot m(\cup_n I_n) \geq m(\cup_{I \in \mathscr{I}} I).$$

5.3 A Local Maximal Function

In this section, we work with both a finite measure m on $J = (-K, K)$ and Lebesgue measure λ. We set $\mathscr{I}(x, r)$ equal to the collection of intervals $I \subseteq J$ containing x (perhaps as an endpoint) with strictly positive length $\lambda(I) \leq r$, and we set

$$M(m, r, x) := \sup_{I \in \mathscr{I}(x,r)} \frac{m(I)}{\lambda(I)}. \tag{5.3.1}$$

For example, if m is the measure λ_f generated by λ and a nonnegative, integrable function f,

$$M(\lambda_f, r, x) = \sup_{I \in \mathscr{I}(x,r)} \frac{1}{\lambda(I)} \int_I f \, d\lambda.$$

As r decreases, the collection $\mathscr{I}(x, r)$ gets smaller. Therefore, $M(m, r, x)$ decreases as r decreases, and so we may set

$$M(m, x) := \lim_{r \to 0+} M(m, r, x).$$

We call the function given by $x \mapsto M(m, x)$ a **local maximal function**. The word "local" distinguishes the function from the classical maximal function, defined by the supremum operation in Equation (5.3.1) with no upper bound on the interval length.

Proposition 5.3.1. *Given a set* $E \subseteq J$ *and* $\alpha > 0$*, let* $E_\alpha := \{x \in E : M(m, x) > \alpha\}$*. Then the Lebesgue outer measure*

$$\lambda^*(E_\alpha) \leq \frac{3}{\alpha} \cdot m(J).$$

Proof. Given $x \in E_\alpha$, there is an interval $I_x \in \mathscr{I}(x, r)$ for some $r \leq 1$ such that

$$\alpha \cdot \lambda(I_x) \leq m(I_x).$$

These intervals form a collection $\mathscr{I}$ that covers E_α, so by the Radó-Aldaz covering theorem 5.2.1, there is a finite disjoint subcollection $\{I_1, \cdots, I_n\} \subset \mathscr{I}$ such that

$$\lambda^*(E_\alpha) \leq \lambda(\cup_{I \in \mathscr{I}} I) \leq 3 \cdot \Sigma_{k=1}^n \lambda(I_k) \leq \frac{3}{\alpha} \Sigma_{k=1}^n m(I_k) \leq \frac{3}{\alpha} \cdot m(J).$$

This last result is similar to the one that is well-known for the classical maximal function. We next, however, indicate the advantage of using a local maximal function. We are implicitly using the fact that a set of measure 0 is contained in a Borel set of measure 0.

Theorem 5.3.1. *Let m be a finite measure on J, and let E be a Borel subset of J. If $m(E) = 0$, then $M(m,x) = 0$ for Lebesgue almost all $x \in E$.*

Proof. Fix $\alpha > 0$, and then fix $\varepsilon > 0$. Since $m(E) = 0$, there is an open set $U \supseteq E$ in J such that $m(U) < \varepsilon\alpha/3$. Let μ be the finite measure on J defined by setting $\mu(A) = m(A \cap U)$ for each Borel set $A \subseteq J$. Now in calculating the values of the local maximal function on E, we need only consider small intervals that fit inside the open set U. Therefore,

$$E_\alpha := \{x \in E : M(m,x) > \alpha\} = \{x \in E : M(\mu,x) > \alpha\}.$$

Therefore, $\lambda^*(E_\alpha) \leq \frac{3}{\alpha}\mu(J) = \frac{3}{\alpha}\mu(U) < \varepsilon$. Since ε is arbitrary, $\lambda^*(E_\alpha) = 0$. Since α is an arbitrary positive number, the result follows.

Corollary 5.3.1. *Let λ_f be the finite measure on J generated by Lebesgue measure λ and a nonnegative, integrable function f. Let E be a Borel subset of J. If $f(x) = 0$ for Lebesgue almost all points of E, then $M(\lambda_f,x) = 0$ for Lebesgue almost all $x \in E$.*

5.4 Differentiation

In this section, we establish various differentiation results. In all of its forms, differentiation is about limits of ratios; the denominator here is always interval length. Moreover, each ratio is bounded above by the supremum used to calculate the local maximal function. If the local maximal function is 0, the limit of the bounded ratios is 0.

First we have the Lebesgue Differentiation Theorem and its corollary extending a form of the Fundamental Theorem of Calculus. The theorem is obvious for continuous functions, and uses the fact (Lusin's theorem 3.3.3) that a measurable function on a set of finite measure is nearly continuous. Recall that the Radó-Aldaz covering theorem uses coverings with intervals that may contain the endpoint at one or both ends. Also recall that a Lebesgue integrable function f on a subset of $\mathbb{R}$ can be extended with the value 0 to all of $\mathbb{R}$. We will at times write $\int_a^c f\, d\lambda$ for the integral $\int_{[a,c]} f\, d\lambda$.

Theorem 5.4.1 (Lebesgue Differentiation). *Let f be a Lebesgue integrable function on $\mathbb{R}$. Then each of the following equalities holds Lebesgue almost everywhere on $\mathbb{R}$:*

$$\lim_{r\to 0+} \frac{1}{2r} \int_{[x-r,x+r]} f\,d\lambda = f(x),$$

$$\lim_{r\to 0+} \frac{1}{r} \int_{[x,x+r]} f\,d\lambda = f(x),$$

$$\lim_{r\to 0+} \frac{1}{r} \int_{[x-r,x]} f\,d\lambda = f(x).$$

Proof. We need only prove the result for $f \geq 0$ and x in our bounded, open interval $J = (-K,K)$. Without changing the value of any integral, we may assume that f takes only finite values. By Lusin's theorem, for each $n \in \mathbb{N}$, there is a compact set $C_n \subset J$ with $\lambda(J \setminus C_n) < 1/n$ such that $f|C_n$ is continuous on C_n. By Proposition 1.11.3, we may extend the function f restricted to C_n with a nonnegative, bounded, continuous function g that vanishes on C_n so that $h := f \cdot \chi_{C_n} + g$ is continuous on J. Since h is continuous on J, each of the limit results holds at any $x \in J$ with h replacing f. Now on J, the function

$$f = h - g + f \cdot \chi_{J\setminus C_n}.$$

Recall that λ_g is the measure on J defined by setting $\lambda_g(A) = \int_A g\, d\lambda$ for every Lebesgue measurable set $A \subseteq J$. Moreover, each of the desired limits using g instead of f is bounded above by the limit of the local maximal function $M(\lambda_g, x)$ at each $x \in J$. Since $\lambda_g(C_n) = 0$, it follows from Corollary 5.3.1 that each of those limits for g is 0 Lebesgue almost everywhere on C_n. Similarly each of the desired limits using $f \cdot \chi_{J\setminus C_n}$ instead of f is 0 Lebesgue almost everywhere on C_n. Therefore, the desired limit results hold Lebesgue almost everywhere on C_n for $f = h - g + f \cdot \chi_{J\setminus C_n}$. It now follows that they hold Lebesgue almost everywhere on $\cup_{n\in\mathbb{N}} C_n$, which is Lebesgue almost everywhere on J.

Corollary 5.4.1 (Extended Fundamental Theorem of Calculus). *Suppose that f is Lebesgue integrable on $[a,b]$, and for each $x \in [a,b]$, $F(x) = \int_a^x f\, d\lambda + C$ where C is a constant. Then $F'(x) = f(x)$ for Lebesgue almost all $x \in [a,b]$.*

Proof. For $\Delta x = r > 0$ and $x + \Delta x \leq b$,

$$\frac{F(x+\Delta x) - F(x)}{\Delta x} = \frac{1}{r} \int_{[x,x+r]} f\,d\lambda.$$

For $\Delta x = -r < 0$ and $x + \Delta x \geq a$,

$$\frac{F(x+\Delta x) - F(x)}{\Delta x} = \frac{1}{-r} \cdot - \int_{[x-r,x]} f\,d\lambda = \frac{1}{r} \int_{[x-r,x]} f\,d\lambda.$$

The value $F(x)$ in the corollary is often written $F(x) = \int_a^x f\, d\lambda + F(a)$. The corollary says that one can differentiate the indefinite integral of a Lebesgue integrable function and get back the integrand almost everywhere. We will seek conditions for when the process can be reversed so that integrating the derivative of a function restores the function except for an additive constant. We know this is impossible for the Cantor-Lebesgue function since that function's derivative is 0 Lebesgue almost everywhere.

Theorem 5.4.2 (Lebesgue Density). *If A is a Lebesgue measurable subset of $\mathbb{R}$, then*

$$\lim_{r\to 0+} \frac{\lambda\,(A\cap[x-r,x+r])}{2r} = 1 \;\; \textit{for } \lambda\textit{-a.e. } x\in A.$$

That is, λ-almost all points of A are "points of density" of A.

Proof. Exercise 5.3.

We have used Theorem 5.3.1 to show the existence Lebesgue almost everywhere of a derivative for any integrator of the form $F(x) = \int_{[-K,x]} f\,d\lambda$, where $f \geq 0$ is Lebesgue integrable. The same theorem also yields a result for general integrators.

Theorem 5.4.3. *Let F be an integrator on J, and let m be the corresponding measure. If E is a Borel subset of J with $m(E) = 0$, then F has a zero derivative Lebesgue almost everywhere on E.*

Proof. Since $m(E) = 0$, $m(\{a\}) = 0$ for each point $a \in E$. Given a point $a \in E$ and given $\Delta x > 0$,

$$\frac{F(a+\Delta x)-F(a)}{\Delta x} = \frac{m(a,a+\Delta x]}{\Delta x} = \frac{m[a,a+\Delta x]}{\Delta x},$$
$$\frac{F(a-\Delta x)-F(a)}{-\Delta x} = \frac{F(a)-F(a-\Delta x)}{\Delta x} = \frac{m(a-\Delta x,a]}{\Delta x}.$$

By Theorem 5.3.1, $M(m,a) = 0$ for Lebesgue almost all points a in E. Therefore, both of the above ratios have limit 0 Lebesgue almost everywhere on E as $\Delta x \to 0$, whence $F'(a)$ exists and is 0 for Lebesgue almost every a in E.

5.5 Functions of Bounded Variation

We have been using increasing functions to construct measures on $\mathbb{R}$. Finite signed measures take both positive and negative values. They correspond to differences of increasing functions. These are the functions of "bounded variation", which we now consider. We work with a fixed, finite interval $[a,b] \subset \mathbb{R}$. Recall that $\mathscr{P}$ denotes the set of all partitions of $[a,b]$ formed by points $a = x_0 < x_1 < \cdots < x_n = b$.

Definition 5.5.1. Let $\mathscr{I}\mathscr{P}$ denote the collection of sets of intervals associated with partitions of $[a,b]$. Each member Q of $\mathscr{I}\mathscr{P}$ is called an **interval partition** of $[a,b]$. It is a finite set of closed intervals $[x_{i-1},x_i]$, $1 \leq i \leq n$, overlapping only at the endpoints and having union $[a,b]$.

Definition 5.5.2. Given a real-valued function f on $[a,b]$, and a subinterval $I = [x_{i-1}, x_i]$ of $[a,b]$, we set $\Delta_I f = f(x_i) - f(x_{i-1})$. For each $Q \in \mathscr{I}\mathscr{P}$, we set $p_Q f = \Sigma_{I \in Q}[(\Delta_I f) \vee 0]$, $n_Q f = \Sigma_{I \in Q}[(-\Delta_I f) \vee 0]$, and $t_Q f = p_Q f + n_Q f = \Sigma_{I \in Q} |\Delta_I f|$. We call

$$P = P_a^b f := \sup_{Q \in \mathscr{I}\mathscr{P}} (p_Q f), \quad N = N_a^b f := \sup_{Q \in \mathscr{I}\mathscr{P}} (n_Q f), \quad \text{and} \quad T = T_a^b f := \sup_{Q \in \mathscr{I}\mathscr{P}} (t_Q f)$$

the **positive, negative**, and **total variation** of f, respectively, on the interval $[a,b]$. We say the function f is of **bounded variation** on $[a,b]$, and we write f is BV, or $f \in BV$, if $T_a^b f < +\infty$.

Note that all of the quantities discussed are nonnegative. Moreover, for each $Q \in \mathscr{I}\mathscr{P}$, $p_Q + n_Q = t_Q \le T$, so if T is finite, then so are P and N.

Definition 5.5.3. A **refinement of an interval partition** Q of $[a,b]$ is an interval partition of $[a,b]$ obtained by adding partition points to the partition points of Q.

Proposition 5.5.1. *Given the real-valued function f on $[a,b]$, if Q is a refinement of Q_0 in $\mathscr{I}\mathscr{P}$, then $p_{Q_0} \le p_Q$, $n_{Q_0} \le n_Q$, and $t_{Q_0} \le t_Q$.*

Proof. Exercise 5.5.

Proposition 5.5.2. *If f is BV on $[a,b]$, then $T = P + N$, and $f(b) - f(a) = P - N$.*

Proof. For each interval partition $Q \in \mathscr{I}\mathscr{P}$, $p_Q - n_Q = f(b) - f(a)$, so

$$p_Q = n_Q + f(b) - f(a) \le N + f(b) - f(a).$$

Therefore $P \le N + f(b) - f(a)$. Also,

$$n_Q = p_Q + f(a) - f(b) \le P + f(a) - f(b),$$

so $N \le P + f(a) - f(b)$, whence $N + f(b) - f(a) \le P$. Therefore, $P = N + f(b) - f(a)$. Since $f \in BV$, we may subtract and obtain $P - N = f(b) - f(a)$.

To show $T = P + N$, we note that for each $Q \in \mathscr{I}\mathscr{P}$,

$$t_Q = p_Q + n_Q = p_Q + (p_Q - p_Q) + n_Q = p_Q + p_Q + f(a) - f(b),$$

so $t_Q \le 2P + f(a) - f(b)$, whence $T \le 2P + f(a) - f(b)$. On the other hand, $2p_Q + f(a) - f(b) = t_Q \le T$, whence $2P + f(a) - f(b) \le T$. It follows that $T = 2P + f(a) - f(b) = 2P + (N - P) = P + N$.

Theorem 5.5.1. *A real-valued function f on $[a,b]$ is of bounded variation if and only if it is the difference $g - h$ of two increasing real-valued functions g and h on $[a,b]$. In this case,*

$$P_a^b f \le g(b) - g(a), \; N_a^b f \le h(b) - h(a), \; \text{so} \; T_a^b f \le g(b) - g(a) + h(b) - h(a). \tag{5.5.1}$$

Proof. Assume $f \in BV$. For each $x \in [a,b]$, $P_a^x f$ and $N_a^x f$ are increasing functions bounded above by $T_a^b f$, and $f(x) = [f(a) + P_a^x f] - N_a^x f$.

Conversely, if $f = g - h$, where g and h are increasing functions, then for any $Q \in \mathscr{I}\mathscr{P}$,

$$p_Q = \Sigma_{I\in Q}[(\Delta_I f) \vee 0] \leq \Sigma_{I\in Q}[(\Delta_I g \vee 0)] = \Sigma_{I\in Q}\Delta_I g = g(b) - g(a),$$

so $P_a^b f \leq g(b) - g(a)$. Similarly $N_a^b f \leq h(b) - h(a)$, so $T_a^b f \leq g(b) - g(a) + h(b) - h(a)$. It follows that f is BV on $[a,b]$.

Note that we do not necessarily have equality in Equation (5.5.1). For example, if $f(x) \equiv 0$ on $[0,1]$, then $f = g - h$, where $g(x) = x$ and $h(x) = x$ for all x in $[0,1]$.

Corollary 5.5.1. *Let m be a finite measure on $[a,b]$, and let f be a function integrable with respect to m on $[a,b]$. Then the function $x \mapsto G(x) := \int_a^x f\ dm = \int_{[a,x]} f\ dm$ is BV on $[a,b]$, and $T_a^b G \leq \int_{[a,b]} |f|\ dm$.*

Proof. Since $G(x) := \int_a^x f^+ dm - \int_a^x f^- dm$, G is BV, and

$$T_a^b G \leq \int_a^b f^+ dm + \int_a^b f^- dm = \int_{[a,b]} |f|\ dm.$$

5.6 Absolute Continuity

In this section, we consider which functions on an interval $[a,b]$ have the form $x \mapsto \int_a^x g\, d\lambda + C$ where g is Lebesgue integrable and C is a constant. By Proposition 4.4.3, such functions are uniformly continuous. Indeed, for any $\varepsilon > 0$, there is a $\delta > 0$ such that if A is a Lebesgue measurable subset of $[a,b]$ with $\lambda(A) < \delta$, then $\int_A |g|\ d\lambda < \varepsilon$. In particular, given a finite set of disjoint intervals of total Lebesgue measure (i.e., length) smaller than δ, the total measure obtained by integrating $|g|$ over those intervals is less than ε. We call functions with this property of the function $x \mapsto \int_a^x |g|\ d\lambda$ "absolutely continuous." We state below the formal definition of absolute continuity for our interval $J = (-K,K)$. One may replace J here with any non-degenerate interval or even the whole line $\mathbb{R}$.

First, let $\mathscr{S}_\delta(J)$ denote all finite sets of closed intervals $I \subseteq J$ with pairwise disjoint interiors and total length less than δ. That is, $S \in \mathscr{S}_\delta(J)$ if and only if $S = \{I_1, I_2, \cdots, I_n\}$ where each I_i is a closed interval in J, for $i \neq j$, $I_i \cap I_j$ is either empty or one point, and the sum of the length of the intervals in S is smaller than δ.

Definition 5.6.1. A real-valued function f defined on an interval J is said to be **absolutely continuous** on J if for any $\varepsilon > 0$, there is a $\delta > 0$ so that for any $S \in \mathscr{S}_\delta(J)$, the sum $\Sigma_{I\in S}|\Delta_I f| < \varepsilon$.

The functions absolutely continuous on an interval J are all uniformly continuous on J. It is an easy exercise (5.13) to show that they form a vector space over $\mathbb{R}$

containing the constants. A function absolutely continuous on an interval J is absolutely continuous on any subinterval of J. A function absolutely continuous on an interval $[a,b]$ can be extended with constant values to form an absolutely continuous function on $\mathbb{R}$. We shall see that although the Cantor-Lebesgue function is continuous on $[0,1]$, and therefore uniformly continuous on the compact set $[0,1]$, it is not absolutely continuous on $[0,1]$.

As noted, every function $x \mapsto \int_a^x g\, d\lambda + C$ formed from a Lebesgue integrable function g on an interval $[a,b]$ is absolutely continuous on $[a,b]$. We will show that all absolutely continuous functions on $[a,b]$ have this form. It will follow that they are functions of bounded variation on $[a,b]$, but we need that result first.

Proposition 5.6.1. *If a real-valued function f is absolutely continuous on an interval $[a,b]$, then f is BV on $[a,b]$, and the functions mapping x to $T_a^x f$, $P_a^x f$, and $N_a^x f$ are all absolutely continuous on $[a,b]$.*

Proof. Fix $\delta > 0$ that works for $\varepsilon = 1$ in terms of the absolute continuity of f. Then fix an $n \in \mathbb{N}$ such that $(b-a)/n < \delta$. Let Q_n be the interval partition of $[a,b]$ formed by the points $x_i = a + i \cdot (b-a)/n$ for $0 \leq i \leq n$. Given any interval I of the partition Q_n, by the choice of δ, any finite interval partitioning of I results in a sum of the absolute values of the change of f of no more than 1. Let Q_0 be an arbitrary interval partition of $[a,b]$, and let Q be the refinement obtained by adding the partition points of Q_n to the partition points of Q_0. Since $t_{Q_0} \leq t_Q \leq n$, the total variation $T_a^b f \leq n$, so f is BV on $[a,b]$.

To show $x \mapsto P_a^x f$ is absolutely continuous, fix $\varepsilon > 0$ and then a $\delta > 0$ that works for $\varepsilon/2$ in terms of the absolute continuity of f. Also fix a finite set of intervals $S_0 \in \mathscr{S}_\delta([a,b])$. We will show that $\Sigma_{I \in S_0} \Delta_I P_a^x f < \varepsilon$. Let Q be a partition of the entire interval $[a,b]$ such that $P_a^b f - p_Q f < \varepsilon/2$. We may assume, without loss of generality, that among the partition points of Q are the endpoints of the intervals of S_0. Let S be the set of intervals formed from the intervals of S_0 by cutting with the partition points of Q that are contained in the intervals of S_0. Clearly, $S \in \mathscr{S}_\delta([a,b])$. Now since $P_a^x f$ is a function for which the value increases as x increases,

$$\Sigma_{I \in S_0} \Delta_I P_a^x f = \Sigma_{I \in S} \Delta_I P_a^x f \leq \frac{\varepsilon}{2} + \Sigma_{I \in S}((\Delta_I f) \vee 0) \leq \frac{\varepsilon}{2} + \Sigma_{I \in S} |\Delta_I f| < \varepsilon.$$

This shows that $P_a^x f$ is absolutely continuous. Similarly, $N_a^x f = P_a^x(-f)$ is also absolutely continuous, and so the sum $T_a^x f$ is absolutely continuous on $[a,b]$.

Corollary 5.6.1. *Any absolutely continuous function f on an interval $[a,b]$ is the difference of two nonnegative, increasing, absolutely continuous functions on $[a,b]$.*

Proof. Since $f \in BV$ on $[a,b]$, for any $x \in [a,b]$, $f(x) - f(a) = P_a^x - N_a^x$.

Next we explore the relationship between the notion of absolute continuity of a function and what is called absolute continuity of a measure with respect to Lebesgue measure. We work again with our finite interval $J = (-K,K) \subset \mathbb{R}$. Recall that if m is a finite measure on J obtained from an integrator F, we may assume that F is continuous at $-K$ and K.

Definition 5.6.2. Let m be a finite measure on $J = (-K, K)$. The measure m is **absolutely continuous** with respect to Lebesgue measure λ if $m(E) = 0$ for every Lebesgue measurable set E with $\lambda(E) = 0$; in this case, we write $m << \lambda$. A finite measure ν on J is **singular** with respect to Lebesgue measure if there is a Lebesgue measurable subset $B \subset J$ with $\lambda(B) = 0$ such that $\nu(J \setminus B) = 0$. In this case, we write $\nu \perp \lambda$.

Remark 5.6.1. These definitions work for pairs of measures, but for this chapter we only work with absolute continuity and singularity with respect to Lebesgue measure. If $m << \lambda$, then every Lebesgue measurable set is measurable with respect to m (Exercise 5.16).

We show next that if m is a finite measure on $J = (-K, K)$, then absolute continuity of m with respect to Lebesgue measure is equivalent to the following condition on the Borel subsets E of J:

$$(*) \quad \forall \varepsilon > 0\, \exists \delta > 0 \quad \text{such that } \lambda(E) < \delta \Rightarrow m(E) < \varepsilon.$$

Note that by Proposition 4.4.3, if for each Lebesgue measurable set A, $m(A) = \int_A f\, d\lambda$ for a fixed, nonnegative, Lebesgue integrable function f, then Condition (*) holds.

Proposition 5.6.2. *If m is a finite measure on $J = (-K, K)$, then $m << \lambda$ if and only if Condition (*) holds for m.*

Proof. Clearly, if (*) holds and the Lebesgue measure of E is 0, then whatever the positive value of δ might be, $\lambda(E) < \delta$, so $m(E) < \varepsilon$ for every positive ε, whence $m(E) = 0$. That is, if Condition (*) holds, then $m << \lambda$. To establish the reverse implication, we assume that (*) is false. Then there is an $\varepsilon > 0$ and a sequence of Borel sets E_n contained in J with $\lambda(E_n) < 2^{-n}$ and yet $m(E_n) \geq \varepsilon$ for each n. Let $E := \limsup_n E_n$. Then $\lambda(E) = 0$ by the Borel-Cantelli Lemma, Theorem 2.6.2. However, for each k, $m(\cup_{n=k}^{\infty} E_n) \geq \varepsilon$, so since $m(J) < +\infty$, we have

$$m(E) = m(\cap_{k=1}^{\infty}(\cup_{n=k}^{\infty} E_n)) \geq \varepsilon.$$

Hence, when (*) is false, there is a Borel set of Lebesgue measure 0 in J with positive m-measure.

Example 5.6.1. It need not be true for an infinite measure that Condition (*) follows from the implication $\lambda(E) = 0 \Rightarrow m(E) = 0$. If, for example, we obtain m on the nonnegative real line by setting $m(E) = \int_E x\, d\lambda$ for each Lebesgue measurable set E, we still have $m << \lambda$, but Condition (*) does not hold.

Here is the connection between absolute continuity of a finite measure and that of its integrator.

Theorem 5.6.1. *If m is a finite measure obtained from an integrator F on $J = (-K, K)$, then $m << \lambda$ if and only if F is absolutely continuous on J.*

Proof. If $m << \lambda$ on J, then singleton sets have m-measure 0, and Condition (*) holds for m. The absolute continuity of F follows directly from Condition (*). Conversely, suppose F is absolutely continuous on J. As noted, we may assume that F is continuous at the endpoints of J. Moreover, singleton sets have m-measure 0. Given $\varepsilon > 0$, fix $\delta > 0$ so that 2δ works for $\varepsilon/2$ with respect to the absolute continuity of F. Fix $E \subseteq J$ with $\lambda(E) < \delta$. Fix an open set O with $E \subseteq O \subseteq J$ and $\lambda(O) < 2\delta$. The open set O is a countable union of disjoint open intervals $I_k = (a_k, b_k)$. For any $n \in \mathbb{N}$, $\Sigma_{k=1}^{n}(b_k - a_k) < 2\delta$. It now follows that Condition (*) holds for m since for sufficiently large $n \in \mathbb{N}$,

$$m(E) \leq m(\cup_{k=1}^{\infty} I_k) \leq 2 \cdot \Sigma_{k=1}^{n} m(I_k) = 2 \cdot \Sigma_{k=1}^{n} (F(b_k) - F(a_k)) < 2 \cdot \frac{\varepsilon}{2} = \varepsilon.$$

We now want to employ a special case of the **Radon-Nikodým Derivative Theorem**. It will be proved later using Hilbert space techniques.

Theorem 5.6.2 (Special Radon-Nikodým Theorem). *Let F be an increasing, absolutely continuous function on $J = (-K, K)$ with continuity at the endpoints of J, so F is also absolutely continuous on $[-K, K]$. Let m_F be the measure generated on J using F as an integrator, so $m_F << \lambda$. Then there is a nonnegative, Lebesgue integrable function f on J such that for any Lebesgue measurable set $A \subseteq J$, $m_F(A) = \int_A f \, d\lambda$. In particular, $F(x) - F(-K) = \int_{[-K,x]} f \, d\lambda$.*

We now have the following important result stating when a function can be recaptured from its derivative using integration.

Theorem 5.6.3. *Let F be a real-valued function on an interval $[a, b]$. If for some Lebesgue integrable function f on $[a, b]$, $F(x) = F(a) + \int_{[a,x]} f \, d\lambda$ for all $x \in [a, b]$, then F is absolutely continuous on $[a, b]$. On the other hand, if F is absolutely continuous on $[a, b]$, then an integrable derivative F' exists Lebesgue almost everywhere on $[a, b]$, and $F(x) = F(a) + \int_a^x F' \, d\lambda$ for all $x \in [a, b]$. That is, the indefinite integral of the derivative of an absolutely continuous function on $[a, b]$ is the function plus a constant.*

Proof. By Corollary 5.6.1, we may assume that F is increasing on $[a, b]$. By Theorem 5.6.2, there is a nonnegative, Lebesgue integrable function f on $[a, b]$ such that for all $x \in [a, b]$, $F(x) - F(a) = \int_{[a,x]} f \, d\lambda$. By Theorem 5.4.1, the derivative $F'(x)$ exists and equals $f(x)$ for Lebesgue almost all $x \in [a, b]$.

Remark 5.6.2. We know that the Cantor-Lebesgue function g has a derivative equal to 0 at Lebesgue almost all points of $[0, 1]$. Since we do **not** have $1 = g(1) = \int_0^1 g'(x) \, d\lambda$, we know that g is not absolutely continuous.

Finally, we want to show that any increasing real-valued function, and hence a *BV* function, has a derivative Lebesgue almost everywhere. If we start with an increasing function on an interval $[a, b]$, we can always extend it with constant values to form an increasing function on $\mathbb{R}$. Therefore, we will work with an increasing, real-valued function on the interval $[-K,K]$. To transform such a function to an integrator, we need the following result.

Proposition 5.6.3. *Let F be an increasing function on $J = (-K,K)$, continuous at $-K$ and K. Let G be another such increasing function on J such that for all $x \in J$,*

$$\lim_{y \to x-} F(y) \leq G(x) \leq \lim_{y \to x+} F(y).$$

Then the set of points $x \in J$ where the derivative $F'(x)$ exists equals the set of points $x \in J$ where $G'(x)$ exists. Moreover, $F'(x) = G'(x)$ at those points.

Proof. Exercise 5.17(A).

We also need the following result, which is proved later as an immediate consequence of the Radon-Nikodým Derivative Theorem.

Theorem 5.6.4 (Special Lebesgue Decomposition Theorem). *If m is a finite measure on J, then $m = \mu + \nu$ on J, where $\mu << \lambda$ and $\nu \perp \lambda$. The decomposition of m into an absolutely continuous measure and a singular measure is unique.*

Theorem 5.6.5. *An increasing real-valued function on an interval in the real line has a derivative Lebesgue almost everywhere on the interval, whence the same is true of a BV function.*

Proof. Let F be an increasing real-valued function on $J = (-K,K)$ that is continuous at the endpoints of J. By subtracting a constant and then changing F at at most a countable number of points (see Problem 5.6), we may assume that F is an integrator with $F(-K) = 0$. By Proposition 5.6.3, we need only establish the result for the changed function F. Now such an integrator F generates a finite measure. By Theorem 5.6.4, the measure generated by F is the sum of a measure $\mu << \lambda$ and a measure $\nu \perp \lambda$. By definition, there is a set B of Lebesgue measure 0 such that $\nu(J \setminus B) = 0$. It follows from Theorem 5.4.3 that the integrator F_ν for the measure ν has a zero derivative λ-a.e. on $J \setminus B$, but since $\lambda(B) = 0$, F_ν has a zero derivative λ-a.e. on J. On the other hand, μ is absolutely continuous with respect to λ. Therefore, by Theorem 5.6.3, its integrator F_μ has a derivative λ-a.e. on J. It follows that $F = F_\mu + F_\nu$ has a derivative λ-a.e. on J.

5.7 Problems

Problem 5.1. a) Prove the Radó-Aldaz covering theorem 5.2.1 with the constant 3 replaced with the constant $2 + \varepsilon$ for an arbitrary $\varepsilon > 0$.
b) Show that in Part a, the constant 3 cannot be replaced with just 2.

Problem 5.2. Give a counterexample for each of the following statements about a collection of intervals for which the hypotheses of Theorem 5.2.1 need not hold:

a) Given an arbitrary collection $\mathscr{I}$ of intervals (some possibly degenerate), all contained in $(-1,1)$, the set $\cup_{I\in\mathscr{I}} I$ is Lebesgue measurable.
b) Given an arbitrary collection $\mathscr{I}$ of non-degenerate intervals in $\mathbb{R}$ and Lebesgue measure λ, there is a finite disjoint subset $\{I_1,\cdots,I_n\} \subseteq \mathscr{I}$ such that

$$\lambda(\cup_{I\in\mathscr{I}} I) \leq 3\cdot\sum_{k=1}^{n}\lambda(I_k).$$

Problem 5.3. Prove Theorem 5.4.2. **Hint:** Work with χ_A.

Problem 5.4. For each Borel set $A\subseteq\mathbb{R}$, let $m(A) = \lambda(A\cap[0,1])$, where λ is Lebesgue measure. Find the set of all $x\in\mathbb{R}$ for which the local maximal function $M(m,x)\geq 1/2$.

Problem 5.5. Prove Proposition 5.5.1. **Hint:** What happens when a single point is added to the partition points forming an interval partition?

Problem 5.6. Show that a monotone (i.e., increasing or decreasing) real-valued function, and therefore a BV function, can have only a countable number of discontinuities. **Hint:** If f is increasing, and the total increase on an interval J is S, how many jumps of size 1 can f have in J?

Problem 5.7. Show that if f is BV on $[a,b]$, then right- and left-hand limits exist at all points of (a,b).

Problem 5.8. a) Show that if f is an increasing function on $[a,b]$, then $N_a^b f = 0$, and $P_a^b f = T_a^b f = f(b) - f(a)$.
b) Given a real-valued function f on $[a,c]$ and a point b with $a<b<c$, show that $P_a^b f + P_b^c f = P_a^c f$ with similar equalities for N and T.
c) Let $G(x) = \int_0^{4\pi}\sin x\,dx$. Show that $T_0^{4\pi} G = \int_0^{4\pi}|\sin x|\,dx$.
d) Let C be the "fat" Cantor set in $[0,1]$ such that $\lambda(C) = 1/2$. Let $f(x)=1$ for all $x\in C$ and $f(x) = -1$ for all $x\in[0,1]\setminus C$. Let $H(x) = \int_0^x f\,d\lambda$ for all $x\in[0,1]$. Is $T_0^1 H = \int_0^1 |f|\,d\lambda$? Explain.

Problem 5.9. Let $f(0) = 0$ and $f(x) = x^2\sin(1/x^2)$ for $x\neq 0$. Determine if f is a function of bounded variation on $[0,1]$. **Hint:** For each $n\in\mathbb{N}$, let $x_n = \sqrt{2/(n\pi)}$.

Problem 5.10. Let $g(0) = 0$ and $g(x) = x^2\sin(1/x)$ for $x\neq 0$. Determine if g is a function of bounded variation on $[0,1]$.

Problem 5.11. a) Show that if f is BV on $[a,b]$, then f is bounded on $[a,b]$.
b) Show that if f and g are BV on $[a,b]$, then fg is BV on $[a,b]$.

Problem 5.12 (A). Give an example of an increasing function on $[0,1]$ that is discontinuous at each rational number and continuous at each irrational number.

Problem 5.13. Show that the absolutely continuous functions on an interval J form a vector space over $\mathbb{R}$ containing the constants.

Problem 5.14. a) Assume f is BV on $[a,b]$, and suppose f is bounded below by some positive constant $c > 0$. Show that $1/f$ is also BV on $[a,b]$.

b) Let f be an absolutely continuous function on $[a,b]$ such that for all $x \in [a,b]$, $f(x) \neq 0$. Show that $1/f$ is also absolutely continuous on $[a,b]$.

Problem 5.15. Define $f : \mathbb{R} \mapsto \mathbb{R}$ by setting $f(x) = \cos(\pi x)$ for each $x \in \mathbb{R}$. For each $n \in \mathbb{N}$, prove that f is BV on $[0,2n]$, and calculate $P_0^{2n} f$, $N_0^{2n} f$, and $T_0^{2n} f$.

Problem 5.16. Show that if $m << \lambda$, then every Lebesgue measurable set is measurable with respect to m.

Problem 5.17 (A). Prove Theorem 5.6.3. **Hint:** Suppose x_0 is a point where a derivative $F'(x_0)$ exists, and x_n is a sequence of points where a jump may occur with $x_n \searrow x_0$. Since a countable set cannot form an open interval, there is a sequence of non-jump points $y_n \searrow x_0$ and a sequence of non-jump points $z_n \searrow x_0$ with $y_n \leq x_n \leq z_n$ for each n.

Problem 5.18. Suppose f is BV on $[a,b]$. Show that if the function $V(x) = T_a^x f$ is absolutely continuous on $[a,b]$, then f is absolutely continuous on $[a,b]$.

Problem 5.19 (A). Suppose g is a monotone (increasing or decreasing), absolutely continuous function on $[0,1]$ and E a set of Lebesgue measure 0 in $[0,1]$. Show that the Lebesgue measure $\lambda(g[E]) = 0$.

Problem 5.20. Let m be a finite measure on $[a,b]$ such that $m << \lambda$. Then there is a Lebesgue integrable function g on $[a,b]$ such that for every m-integrable function f, $\int f\, dm = \int fg\, d\lambda$. **Hint:** Recall Theorem 5.6.2.

Problem 5.21 (A).

a) Show that if f satisfies a Lipschitz condition, i.e., there is a constant $M > 0$ such that for all x and y in the domain of f,

$$|f(x) - f(y)| \leq M \cdot |x - y|$$

then f is absolutely continuous. Note that the constant M is called a Lipschitz constant, and the letter combination “sh” is rarely used in German.

b) Show that an absolutely continuous function f satisfies a Lipschitz condition if and only if the absolute value of the derivative, $|f'|$, is bounded (where it exists).

c) Assume $f : \mathbb{R} \mapsto \mathbb{R}$ satisfies a Lipschitz condition, and $g : [0,1] \mapsto \mathbb{R}$ is an absolutely continuous function. Recall that for all $x \in [0,1]$, $f \circ g(x) = f(g(x))$. Show that $f \circ g$ is an absolutely continuous function on $[0,1]$.

Problem 5.22. Let $f : [a,b] \mapsto \mathbb{R}$ be an absolutely continuous function. Show that f is a constant function if and only if $f'(x) = 0$ Lebesgue almost everywhere.

Problem 5.23. Let f be an absolutely continuous function for every interval $I \subseteq \mathbb{R}$. Suppose that f and f' are both Lebesgue integrable on $\mathbb{R}$. Show that $\int_{\mathbb{R}} f' d\lambda = 0$. **Hint:** What are the limits of f at $\pm\infty$?

Problem 5.24. Let $\langle f_n \rangle$ be a sequence of absolutely continuous functions on $[0,1]$ such that $f(x) = \sum_{n=1}^{\infty} f_n(x)$ for all $x \in [0,1]$ and $\sum_{n=1}^{\infty} |f_n'(x)|$ is Lebesgue integrable on $[0,1]$. Show that f is absolutely continuous on $[0,1]$.

Problem 5.25. For each bounded, Lebesgue measurable function f on $\mathbb{R}$ and each $x \in \mathbb{R}$, let $T(f,x) := \lim_{r \to 0+} \frac{1}{2r} \int_{[x-r,x+r]} f \, d\lambda$. Let C be the "fat" Cantor set in $[0,1]$ such that $\lambda(C) = 1/2$. Suppose f is a bounded Lebesgue measurable function of $\mathbb{R}$ such that the restriction of f to C is continuous at each point of C (that is, one ignores the values taken by f on $\mathbb{R} \setminus C$). Use Theorem 5.4.2 to show that $T(f,x) = f(x)$ at λ-almost all points of C. Don't forget that $T(f,x)$ makes use of the values of f off of C.

Chapter 6
General Measure Spaces

6.1 Introduction

In this chapter, we extend results obtained for the real line to more general spaces supplied with a measure. Recall that a σ-algebra in a set X is a collection of subsets of X; the collection contains X itself and is stable with respect to the operations of taking complements and countably infinite unions. It follows that a σ-algebra contains the empty set, and so it is also stable with respect to the operations of taking finite unions, and finite and countably infinite intersections. Also recall that the Borel subsets of the real line form the smallest σ-algebra containing the open subsets of $\mathbb{R}$.

Definition 6.1.1. A **measurable space** is a pair $(X,\mathscr{B})$ consisting of a set X and a σ-algebra $\mathscr{B}$ formed from subsets of X. Members of the collection $\mathscr{B}$ are called **measurable**. A **measure** μ on a measurable space $(X,\mathscr{B})$ is a function from $\mathscr{B}$ into the extended nonnegative real line, $[0,+\infty]$, such that $\mu(\varnothing)=0$ and for any pairwise disjoint sequence $\{A_i : i \in \mathbb{N}\}$ in $\mathscr{B}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

A **measure space** is a triple $(X,\mathscr{B},\mu)$, consisting of a measurable space $(X,\mathscr{B})$, and a measure μ on $(X,\mathscr{B})$. If $\mu(X)$ is finite, μ is a **finite measure**, and $(X,\mathscr{B},\mu)$ is a **finite measure space**. If X is the countable union of sets of finite μ-measure, μ is a **σ-finite measure**, and $(X,\mathscr{B},\mu)$ is a **σ-finite measure space**. A set $B \in \mathscr{B}$ is a **null set** if $\mu(B)=0$. A property holds **μ-almost everywhere**, or just a.e., if it holds outside of a null set. A measure is **complete** if any subset of a null set is measurable. A measure space is complete if the measure is complete.

Remark 6.1.1. In probability theory, a **probability measure** μ is a measure on a measurable space $(X,\mathscr{B})$ with $\mu(X)=1$. The measure space $(X,\mathscr{B},\mu)$ is called a **probability space**.

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_6

A measure space $(X, \mathscr{B}_0, \mu)$ that is not complete can be extended to a complete measure space as indicated in the construction for $\mathbb{R}$ in Definition 3.2.1. When we are working with only one measure, we will in general assume that the measure space is complete. Clearly, we must be more careful when more than one measure is involved in the discussion. Again, we will use the notation $E_n \nearrow E$ to indicate a sequence of sets such that $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$ and $\cup_n E_n = E$. Also, $E_n \searrow E$ indicates a sequence of sets such that $E_n \supseteq E_{n+1}$ for all n and $\cap_n E_n = E$. Moreover, we have the following limit result.

Proposition 6.1.1. *Let $(X, \mathscr{B}, \mu)$ be a measure space, and let $\langle E_n : n \in \mathbb{N} \rangle$ be a sequence of measurable sets. If $E_n \nearrow E$, then $\mu(E) = \lim \mu(E_n)$. If $E_n \searrow E$, and for some k $\mu(E_k) < +\infty$, then $\mu(E) = \lim \mu(E_n)$.*

Proof. Exercise 6.1.

6.2 Integration

We next extend results for measurable, extended-real valued functions and their corresponding integrals to measurable spaces that are not necessarily based on the real line. To avoid repetitive definitions, we treat real-valued functions as extended-real valued functions for which the infinite values are taken on the empty set.

Definition 6.2.1. Given a measurable space $(X, \mathscr{B})$, an extended-real valued function f defined on X is **measurable** if the inverse image with respect to f of each open subset of $\mathbb{R}$ is in $\mathscr{B}$ and both $f^{-1}[\{+\infty\}] \in \mathscr{B}$ and $f^{-1}[\{-\infty\}] \in \mathscr{B}$.

Remark 6.2.1. Since inverse images preserve complements, unions, and intersections, it is enough to know the measurability of $f^{-1}[\{+\infty\}]$, $f^{-1}[\{-\infty\}]$, and the inverse image of each interval of the form $(\alpha, +\infty)$ where α is a rational number. As is true for functions defined on $\mathbb{R}$, it then follows that each Borel set in the real line has measurable inverse image. In this chapter it will be understood that a measurable function takes its values in the real or extended real line.

Remark 6.2.2. In probability theory, a measurable real-valued function on a probability space is called a **random variable**. When terminology was set in the first half of the twentieth century, Joseph Doob wanted to call such a function a "chance variable", and William Feller disagreed, wanting to call it a "random variable"; so they tossed a coin, and Feller won.[1]

For what follows, we fix a complete measure space $(X, \mathscr{B}, \mu)$. The measure may not be finite. Results are stated for the whole space X; similar results are true for any measurable subset E of X. Either restrict the measure and all functions to E or replace all function values with 0 off of E. We first construct the integral for nonnegative functions.

[1] This history was verified by the author in a conversation with Joseph Doob.

Definition 6.2.2. A **simple function** is a measurable function with finite range in $\mathbb{R}$. A simple function that is not identically equal to 0 is presented in **canonical form** as the finite sum $\sum a_i \chi_{A_i}$, where the A_i's are measurable and pairwise disjoint, and the a_i's are distinct and not 0. The canonical representation of the function 0 is 0. The **integral** with respect to μ of a nonnegative, nonzero simple function having canonical form $\sum a_i \chi_{A_i}$ is the sum $\sum a_i \cdot \mu(A_i)$. The integral of 0 is 0. We write $\int \varphi$ to denote the integral of φ; notation that stresses the measure is $\int \varphi \, d\mu$ and $\int \varphi \, \mu(dx)$. Given a measurable set $E \subseteq X$, $\int_E \varphi$ denotes the integral $\int \varphi \cdot \chi_E$.

Definition 6.2.3. Given a nonempty measurable set $E \subseteq X$ and a finite number of nonempty measurable subsets E_i of E, the **partition refinement** of E determined by the sets E_i is the finite collection of nonempty, measurable, pairwise disjoint subsets $\{A_j : 1 \leq j \leq k\}$ of E such that $E = \cup_j A_j$, and each E_i is the union of the A_j's that have nonempty intersection with E_i.

Proposition 6.2.1. *If φ is a finite linear combination $\sum_{i=1}^n \alpha_i \chi_{E_i}$ of characteristic functions of nonempty, measurable sets E_i with each $\alpha_i > 0$, then φ is a nonnegative simple function, and $\int \varphi \, d\mu = \sum_{i=1}^n \alpha_i \cdot \mu(E_i)$.*

Proof. Let $E = \cup_i E_i$. Since φ is measurable and takes only a finite number of values, φ is a simple function. Let $\{A_j : 1 \leq j \leq k\}$ be the partition refinement of E determined by the sets E_i. Now $\varphi = \sum_{j=1}^k c_j \chi_{A_j}$, where for each j, $c_j = \sum_{i,\, A_j \subseteq E_i} \alpha_i > 0$. It follows that

$$\sum_{i=1}^n \alpha_i \cdot \mu(E_i) = \sum_{i=1}^n \alpha_i \cdot \left(\sum_{j, A_j \subseteq E_i} \mu(A_j) \right) = \sum_{i=1}^n \sum_{j, A_j \subseteq E_i} \alpha_i \cdot \mu(A_j)$$
$$= \sum_{j=1}^k \sum_{i, A_j \subseteq E_i} \alpha_i \cdot \mu(A_j) = \sum_{j=1}^k c_j \cdot \mu(A_j).$$

Given the representation $\varphi = \sum_{j=1}^k c_j \chi_{A_j}$, we may combine (by taking a union) all of the sets A_j with the same value c_j into one set without changing the function or the integral. This is now the canonical form for φ and its integral.

Proposition 6.2.2. *If φ and ψ are nonnegative, nonzero simple functions and $\alpha > 0$ is in $\mathbb{R}$, then $\int \alpha\varphi = \alpha \int \varphi$, and $\int (\varphi + \psi) = \int \varphi + \int \psi$. If $\psi \geq \varphi$ a.e., then $\int \psi \geq \int \varphi$.*

Proof. It is clear that $\int \alpha\varphi = \alpha \int \varphi$. Let $E = \{\varphi + \psi > 0\}$. Take the partition refinement $\{A_j : 1 \leq j \leq k\}$ of E determined by the sets $E_i \subseteq E$ on which φ takes distinct values (including 0) and the sets $F_k \subseteq E$ on which ψ takes distinct values (including 0). Then φ and ψ have representations $\varphi = \sum_j c_j \chi_{A_j}$ and $\psi = \sum_j d_j \chi_{A_j}$. On each set A_j, $c_j + d_j > 0$. Therefore,

$$\int (\varphi + \psi) = \sum_j (c_j + d_j)\mu(A_j) = \sum_{j,\, c_j \neq 0} c_j \mu(A_j) + \sum_{j,\, d_j \neq 0} d_j \mu(A_j) = \int \varphi + \int \psi.$$

If $\varphi \geq \psi$ a.e., we may change their values on a null set without changing the integrals, so that $\varphi \geq \psi$ on E. Now for each j, $c_j \geq d_j$, so $\int \varphi \geq \int \psi$.

Corollary 6.2.1. *If $0 < b \leq \varphi \leq B$ on $E = \{\varphi > 0\}$, then $b \cdot \mu(E) \leq \int \varphi \leq B \cdot \mu(E)$.*

Proof. Clear.

Proposition 6.2.3. *Any nonnegative, extended-real valued, measurable function f is the limit of an increasing sequence of simple functions.*

Proof. Exercise 6.4.

Definition 6.2.4. The **integral** of a nonnegative measurable function f is the supremum of the integrals of the nonnegative simple functions it dominates. We write $\int f$ to denote the integral of f, but notation that stresses the measure is $\int f\, d\mu$ and $\int f\, \mu(dx)$. The function f is called the **integrand** of the integral.

Remark 6.2.3. Clearly, this definition of the integral agrees with the definition for nonnegative simple functions when both definitions apply. Moreover, the following result is an easy application of the definition.

Proposition 6.2.4. *For nonnegative measurable functions, the integral is increasing. That is, $0 \leq f \leq g \Rightarrow \int f \leq \int g$. If $c > 0$ in $\mathbb{R}$, then $\int c \cdot f = c \cdot \int f$. Moreover, the value of the integral is independent of changes on sets of measure 0.*

Example 6.2.1. Given a nonempty set X, let the σ-algebra $\mathscr{B}$ consist of all subsets of X, and let μ be the measure on $(X, \mathscr{B})$ such that $\mu(\{x\}) = 1$ for each point $x \in X$. It follows that $\mu(A) = +\infty$ if A is not a finite set. The measure μ is called **counting measure** on X. Every extended-real valued function on X is measurable. The integral of a nonnegative function is the unordered sum of its values on X; that is, it is the supremum of sums over finite subsets of X. This is the sequential sum if X is countable.

We next prove the Fatou Lemma directly from the definition of the integral.

Theorem 6.2.1 (Fatou's Lemma). *Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of nonnegative measurable functions on X. Then*

$$\int \liminf\, f_n \leq \liminf \int f_n.$$

Proof. Fix a nonnegative simple function $\varphi \leq \liminf f_n$. We must show that $\int \varphi \leq \liminf \int f_n$. First, assume that $\int \varphi = +\infty$. Then for some set A of infinite measure and some strictly positive constant a, $\varphi > a$ on A. For each $k \in \mathbb{N}$ set $B_k = \{x \in A : f_k(x) > a\}$, and for each $n \in \mathbb{N}$ set $A_n = \bigcap_{k \geq n} B_k$. Now $\langle A_n \rangle$ is an increasing sequence of measurable sets with union equal to A since $\liminf\, f_n > a$ on A. Therefore, $\lim \mu(A_n) = \mu(A) = +\infty$. Since for all $k \geq n$, $\int f_k \geq a \cdot \mu(A_n) \to +\infty$, $\liminf \int f_n = +\infty$.

Now assume that $\int \varphi$ is finite. Since φ takes only a finite number of values, the set A where $\varphi(x) > 0$ has finite measure. The desired result holds for φ if it holds for $(1-\delta)\varphi$ for all small positive δ. Therefore, we may assume that $\varphi(x) < \liminf f_n(x)$ for all $x \in A$. Let M be the maximum value of φ. For each $k \in \mathbb{N}$, set $B_k = \{x \in A : f_k(x) > \varphi(x)\}$, and for each $n \in \mathbb{N}$, set $A_n = \bigcap_{k \geq n} B_k$. Then $A \setminus A_n$ decreases to the empty set, so the measure decreases to 0. Given $\varepsilon > 0$, there is an $n \in \mathbb{N}$ such that $\mu(A \setminus A_n) < \varepsilon$, whence for all $k \geq n$,

$$\int f_k \geq \int_{B_k} f_k \geq \int_{B_k} \varphi \geq \int_{A_n} \varphi = \int_A \varphi - \int_{A \setminus A_n} \varphi \geq \int \varphi - \varepsilon \cdot M.$$

Since $\varepsilon > 0$ is arbitrary, the result follows.

Remark 6.2.4. A good reminder of the direction of the inequality in Fatou's Lemma is the Lebesgue integral applied to characteristic functions for the sequence $n \mapsto [n, n+1]$.

Corollary 6.2.2. *If $f_n \to f$ a.e., and for all $n \in \mathbb{N}$, $0 \leq f_n \leq f$ a.e., then $\int f_n \to \int f$.*

Proof.
$$\int f = \int \liminf\, f_n \leq \liminf \int f_n \leq \limsup \int f_n \leq \int f.$$

Corollary 6.2.3 (Monotone Convergence Theorem). *Suppose $f_n \nearrow f$ a.e., that is, outside of a null set, $0 \leq f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$ and $\lim f_n = f$. Then $\int f = \lim \int f_n$.*

Proposition 6.2.5. *For nonnegative measurable functions, the map $f \mapsto \int f$ preserves addition, i.e., the integral of a sum is the sum of the integrals. Moreover, the map preserves multiplication by strictly positive real numbers. Also, $\int f = 0$ if and only if $f = 0$ a.e.*

Proof. Exercise 6.5.

Remark 6.2.5. If $\int f = +\infty$, then $0 \cdot \int f$ is not defined. If $\int f$ is finite, then f is finite a.e., $0 \cdot \int f = 0$, and $0 \cdot f(x)$ is defined and equals 0 a.e.

Definition 6.2.5. A nonnegative measurable function is called **integrable** if its integral is finite. A measurable function f is called integrable if $f^+ := f \vee 0$ and $f^- := -f \vee 0$ are integrable, and then the integral is given by $\int f = \int f^+ - \int f^-$. If just one of these integrals is finite, we still call the difference the integral of f. We say a function is **integrable on a set** E if replacing the function's values with 0 off of E yields an integrable function.

Proposition 6.2.6. *For integrable functions, the map $f \mapsto \int f$ is linear. Moreover, if f_1 and f_2 are nonnegative integrable functions taking only finite values such that $f = f_1 - f_2$, then $\int f = \int f_1 - \int f_2$.*

Proof. Assume f is integrable and $\alpha \in \mathbb{R}$. Without changing its integral, we may assume that f takes only finite values. If $\alpha \geq 0$, then

$$\int \alpha f = \int (\alpha f)^+ - \int (\alpha f)^- = \alpha \int f^+ - \alpha \int f^- = \alpha \int f.$$

If $\alpha < 0$, then

$$\begin{aligned} \int \alpha f &= \int (\alpha f)^+ - \int (\alpha f)^- \\ &= \int (-\alpha) f^- - \int (-\alpha) f^+ = \alpha \int f^+ - \alpha \int f^- = \alpha \int f. \end{aligned}$$

Suppose f_1 and f_2 are nonnegative integrable functions taking only finite values such that $f = f_1 - f_2$. To show that $\int f = \int f_1 - \int f_2$, we note that $f = f^+ - f^- = f_1 - f_2$, so $f^+ + f_2 = f_1 + f^-$. It follows that

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

whence

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

To finish the proof of linearity, suppose that g is also an integrable function taking only finite values. Then

$$f + g = (f^+ + g^+) - (f^- + g^-)$$

so

$$\int f + g = \int (f^+ + g^+) - \int (f^- + g^-) = \int f^+ + \int g^+ - \int f^- - \int g^- = \int f + \int g.$$

Proposition 6.2.7. *If f is measurable and $g \geq 0$ is integrable and $|f| \leq g$ a.e., then f is integrable.*

Proof. Since $f^+ \leq g$ and $f^- \leq g$ a.e., the result is clear.

Corollary 6.2.4. *A measurable function f is integrable if and only if $|f|$ is integrable.*

Remark 6.2.6. We must assume measurability. Otherwise, Lebesgue measure and the characteristic function of a non-measurable set in $[0,1]$ minus the characteristic function of its complement in $[0,1]$ provide a counterexample.

Proposition 6.2.8. *If f and g are integrable, and $f \leq g$ a.e., then $\int f \leq \int g$.*

Proof. Without changing the integrals, we may assume that both functions take only finite values. Since $g - f \geq 0$ a.e., $\int g - \int f = \int (g - f) \geq 0$.

Theorem 6.2.2 (Dominated Fatou Lemma). *Let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of measurable functions and $\langle g_n : n \in \mathbb{N} \rangle$ a sequence of integrable functions with $|f_n| \leq g_n$ a.e. for each $n \in \mathbb{N}$. Assume g_n converges to an integrable function g a.e. and $\int g_n \to \int g$. Then $\underline{\lim} f_n$ and $\overline{\lim} f_n$ are integrable, and*

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

Proof. By making changes on a null set, we may assume finiteness, domination, and convergence at all points. Since

$$-g = -\lim g_n \leq \underline{\lim}\, f_n \leq \overline{\lim} f_n \leq \lim g_n = g,$$

the functions $\underline{\lim} f_n$ and $\overline{\lim} f_n$ are integrable and take only finite values. Moreover,

$$\begin{aligned}\int g + \int \underline{\lim}\, f_n &= \int (g + \underline{\lim}\, f_n) = \int (\lim g_n + \underline{\lim}\, f_n) = \int \underline{\lim}(g_n + f_n)\\ &\leq \underline{\lim} \int (g_n + f_n) = \underline{\lim}\left(\int g_n + \int f_n\right) = \lim \int g_n + \underline{\lim} \int f_n\\ &= \int g + \underline{\lim} \int f_n.\end{aligned}$$

It follows that $\int \underline{\lim}\, f_n \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n$. To show that $\overline{\lim} \int f_n \leq \int \overline{\lim}\, f_n$, we apply this calculation to the sequence $-f_n$ to get

$$-\int \overline{\lim}\, f_n = \int \underline{\lim}(-f_n) \leq \underline{\lim}\left(-\int f_n\right) = -\overline{\lim} \int f_n.$$

Remark 6.2.7. We must assume that $\int g_n \to \int g$. Without this assumption, we have many counterexamples such as the Lebesgue integral applied to the sequence $f_n = g_n = \chi_{[n,n+1]}$ and $f = g = 0$, or $f_n = g_n = n\chi_{[0,1/n]}$ and $f = g = 0$. Of course, a simpler result holds, as in the following corollary, if a single function g can replace all of the functions g_n. See Problem 6.6.

Corollary 6.2.5 (Lebesgue Dominated Convergence). *Let $\langle f_n : n \in \mathbb{N}\rangle$ be a sequence of measurable functions with $f_n \to f$ a.e. Assume that $g \geq 0$ is integrable, and for all $n \in \mathbb{N}$ and almost every $x \in X$, $|f_n(x)| \leq g$. Then $\int f = \lim_n \int f_n$.*

Remark 6.2.8. The integral with respect to a probability measure of an integrable function f is a weighted average of the values taken by f. In probability theory, such an integral is called the **expectation** of f.

Remark 6.2.9. For some of our work, we will integrate complex-valued functions using an extended-real valued measure. Recall that a complex number z has the form $x + iy$ where x and y are real numbers and $i = \sqrt{-1}$. We associate the set $\mathbb{C}$ of complex numbers with the xy-plane by mapping $x + iy$ to the point (x, y). The number z has a complex conjugate $\overline{z} = x - iy$. The product $z \cdot \overline{z} = x^2 + y^2 = |z|^2$, where $|z|$ is the modulus of z, which is the distance of z from the origin. The modulus $|z| = 1$ if and only if for some value $\theta \in [0, 2\pi]$, $z = \cos\theta + i\sin\theta = e^{i\theta}$. While a complex number w has the above form in terms of a real part x and an imaginary part y, it also has the form $|w|\, e^{i\phi}$. Multiplying such a complex number w by $e^{i\theta}$ produces a value $|w|\, e^{i(\phi+\theta)}$, which is a point on the circle of radius $|w|$ rotated from w by the angle θ.

Definition 6.2.6. If h is a complex-valued function, $\overline{h}$ denotes the function taking the value $\overline{h}(x)$ for each x in the domain of h. If f and g are real-valued integrable functions and $h = f + ig$, then the integral of h is given by $\int h = \int f + i \int g$.

The notion of convergence in measure is important in our general setting.

Definition 6.2.7. Let $(X, \mathscr{B}, \mu)$ be a measure space. A sequence $\langle f_n : n \in \mathbb{N} \rangle$ of measurable functions converges to 0 in measure if for each $\varepsilon > 0$, $\mu(\{|f_n| > \varepsilon\}) \to 0$. In general, f_n converges to a measurable f in measure if $|f_n - f|$ converges to 0 in measure.

The latter definition is usually put as follows: for any $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that for all $n \geq N$, $\mu\{|f_n - f| \geq \varepsilon\} < \varepsilon$. Note that we do not take into account how much larger than ε the function f_n is when $|f_n| > \varepsilon$.

As noted in Problem 6.12, Egoroff's theorem holds not just for $\mathbb{R}$, but for this general setting as well. It follows that if $f_n \to f$ μ-a.e. on a set E of finite measure, then $f_n \to f$ in measure. For a set of infinite measure this is not true. For example, $\chi_{[n,n+1]}$ converges to 0 pointwise, but not in measure using Lebesgue measure. Moreover, convergence in measure, even on a set of finite measure, does not imply a.e. convergence. An example is given by Lebesgue measure and the sequence

$$f_1 = \chi_{[0,1]} \quad f_2 = \chi_{[0,1/2]} \quad f_3 = \chi_{[1/2,1]} \quad f_4 = \chi_{[0,1/4]} \text{ etc.}$$

Here, f_n converges to 0 in measure, but we have convergence at no point. In general, however, we do have the following result. The proof is essentially the same as for $\mathbb{R}$; see Theorem 4.8.1.

Theorem 6.2.3. *Suppose a sequence $\langle f_n \rangle$ of measurable functions converges to a measurable f in measure. Then a subsequence converges to f μ-a.e.*

6.3 Signed Measures

Fix a measurable space $(X, \mathscr{B})$. An extended-real valued function μ on $\mathscr{B}$ is called a **signed measure** if $\mu(\varnothing) = 0$ and μ is countably additive. That is, if $\langle A_n \rangle$ is a finite or countably infinite disjoint sequence in $\mathscr{B}$, then the measure of the union is the sum of the measures. For this to make sense, we must assume that one or both of the series $\sum_{n=1}^{\infty} (\mu(A_n) \vee 0)$, $\sum_{n=1}^{\infty} (-\mu(A_n) \vee 0)$ is finite. A signed measure, therefore, can take only one of the two infinite values in the extended real line. If, for example, $+\infty$ is the measure of any set, then $-\infty$ is not the measure of any set.

As before, a set is call **measurable** if it is in $\mathscr{B}$. We will reserve the term **measure** for a set function on $(X, \mathscr{B})$ taking only nonnegative values.

Definition 6.3.1. Given a signed measure μ on a measurable space $(X, \mathscr{B})$, a **positive set** (with respect to μ) is a measurable set, all of the measurable subsets of which have nonnegative measure. A **negative set** is a measurable set, all of the measurable subsets of which have nonpositive measure. A set that is both positive and negative is called a **null set**. A **Hahn decomposition** of X with respect to μ is a pair of disjoint sets A and B with union X such that A is a positive set and B is a negative set.

Note that a null set is a set of measure 0, but a set of measure 0 may not be a null set. Moving a null set from one set of a Hahn decomposition to the other set produces another Hahn decomposition. By definition any measurable subset of a positive set, negative set, or null set is, respectively, positive, negative, or null. Moreover, it is easy to see that a countable union of positive, negative, or null sets is, respectively, positive, negative, or null.

Proposition 6.3.1. *Let μ be a signed measure on $(X,\mathscr{B})$ and E a measurable set with $0 < \mu(E) < +\infty$. Then there is a positive set $A \subseteq E$ with $0 < \mu(A) < +\infty$.*

Proof. If there is a subset of E of measure ≤ -1, remove it and work with the complement in E. We can only do this a finite number of times since $\mu(E)$ is positive and finite. Therefore, if E is not itself a positive set, there is a subset E_1 of positive measure for which no further subset has measure ≤ -1. We work with E_1. Now if there is a subset of E_1 of measure $\leq -\frac{1}{2}$, remove it. You can do this only once. Now if there is a subset of what remains of measure $\leq -\frac{1}{4}$, remove it. You can do this only once. Continuing in this way, we remove a countable disjoint union of sets from E. The remainder A is a positive set since there is no subset of strictly negative measure. Moreover, $\mu(A)$ is finite and strictly positive since $A \subseteq E$ and the union of the removed sets has finite negative measure.

Corollary 6.3.1. *Let μ be a signed measure on $(X,\mathscr{B})$ that takes both strictly positive and strictly negative values on $\mathscr{B}$. If there is no measurable subset of X with infinite positive measure, then there is a finite upper bound to the measure of positive subsets of X. If there is no measurable subset of X with infinite negative measure, then there is a finite lower bound to the measure of negative subsets of X. Moreover, at least one of these two conditions must hold.*

Proof. Suppose every measurable subset of X with positive measure has finite measure. Then it follows from the proposition that any subset of X with strictly positive measure contains a positive set A with $0 < \mu(A) < +\infty$. If for each $n \in \mathbb{N}$ there would exist a positive set A_n with $\mu(A_n) \geq n$, then $\cup_n A_n$ would be a positive set with infinite positive measure, contradicting the assumption. Therefore, in this case there is a finite upper bound to the measure of positive subsets of X. Similarly, if there is no measurable subset of X with infinite negative measure, then there is a finite lower bound to the measure of negative subsets of X. By the definition of a signed measure, there cannot be a subset of X with infinite positive measure and a subset of X with infinite negative measure.

Theorem 6.3.1 (Hahn Decomposition). *Suppose μ is a signed measure on $(X,\mathscr{B})$ that takes both strictly positive and strictly negative values. Then there is a Hahn decomposition of X into a positive set A and negative set B. Given any other such Hahn decomposition, A_1 and B_1, the symmetric differences $A\Delta A_1$ and $B\Delta B_1$ are μ-null sets; that is, the Hahn decomposition is determined up to μ-null sets.*

Proof. We assume that there is no measurable subset of X with infinite positive measure; otherwise, we may work with $-\mu$. Set

$$\lambda := \sup\{\mu(A) : A \text{ is a } \mu\text{-positive subset of } X\}.$$

By Corollary 6.3.1, λ is finite. Let $\langle A_n : n \in \mathbb{N}\rangle$ be a sequence of positive sets with $\mu(A_n) \geq \lambda - 1/n$ for each n, and let $A = \cup_n A_n$. Then A is a positive set, and for each $n \in \mathbb{N}$,

$$\lambda - 1/n \leq \mu(A_n) \leq \mu(A) \leq \lambda,$$

whence $\mu(A) = \lambda$. It now follows that $B = X \setminus A$ is a negative set, for if not, there is a subset $C \subseteq B$ with finite strictly positive measure, and then by Proposition 6.3.1, a positive set $D \subseteq C$ with finite strictly positive measure, whence $A \cup D$ is a positive set with $\mu(A \cup D) > \lambda$. To show that a Hahn decomposition A, B is unique except for null sets, we assume that A_1, B_1 is another Hahn decomposition. Now

$$A \setminus A_1 \subseteq A, \;\; A \setminus A_1 \subseteq B_1, \quad A_1 \setminus A \subseteq A_1, \;\text{ and }\; A_1 \setminus A \subseteq B.$$

Thus, $A \setminus A_1$ is both a positive set and a negative set, and the same is true for $A_1 \setminus A$. That is, both are null sets, so the symmetric difference $A \Delta A_1 = (A \setminus A_1) \cup (A_1 \setminus A)$ is a null set. A similar proof shows that $B \Delta B_1$ is a null set.

Recall that the term "measure" refers to a function on $(X, \mathscr{B})$ taking only non-negative values.

Definition 6.3.2. Two measures μ and ν on a measurable space $(X, \mathscr{B})$ are called **mutually singular** measures, and we write $\mu \perp \nu$, if there are disjoint measurable sets A and B with $X = A \cup B$ and $\mu(B) = \nu(A) = 0$. A **Jordan decomposition** of a signed measure γ is given by a pair of mutually singular measures γ^+ and γ^- with $\gamma = \gamma^+ - \gamma^-$.

Theorem 6.3.2. *Every signed measure γ on $(X, \mathscr{B})$ has a unique Jordan decomposition obtained from a Hahn decomposition of X into a positive set A and a negative set B. The Jordan decomposition sets $\gamma^+(E) = \gamma(E \cap A)$ and $\gamma^-(E) = -\gamma(E \cap B)$ for all $E \in \mathscr{B}$.*

Proof. Clearly, what we have constructed is a Jordan decomposition. If $\gamma = \gamma_1 - \gamma_2$ is another Jordan decomposition and A_1, B_1 are the disjoint sets with $\gamma_1(B_1) = \gamma_2(A_1) = 0$, then A_1, B_1 is a Hahn decomposition for γ, so $A \Delta A_1$ is null. It follows that $\gamma^+(E) = \gamma_1(E)$ for all $E \in \mathscr{B}$. A similar statement is true for γ^-.

Definition 6.3.3. Given a Jordan decomposition $\gamma = \gamma^+ - \gamma^-$ of a signed measure γ on $(X, \mathscr{B})$, the absolute value or **total variation** of γ is the measure $|\gamma| := \gamma^+ + \gamma^-$. A signed measure γ is called **finite** if $|\gamma|(X)$ is finite.

Note that a set is γ-null if and only if it is a set of $|\gamma|$-measure 0. Examples of signed measures can be obtained from a measurable function f for which either f^+

or f^- or both are integrable with respect to a measure μ by setting $\gamma(E) = \int_E f \, d\mu$. In this case, a Hahn decomposition is the pair of sets $\{f \geq 0\}$, $\{f < 0\}$. Another Hahn decomposition is the pair $\{f > 0\}$, $\{f \leq 0\}$. The measure $|\gamma|$ is given by integrating $|f|$.

Definition 6.3.4. Let f be a measurable function and γ a signed measure on $(X, \mathscr{B})$. Then f is integrable with respect to γ if it is integrable with respect to $|\gamma|$, in which case $\int f \, d\gamma := \int f \, d\gamma^+ - \int f \, d\gamma^-$.

6.4 Convexity and Jensen's Inequality

Definition 6.4.1. A real-valued function φ defined on an open interval $J \subseteq \mathbb{R}$ is called **convex** if for any pair of points $x < y$ in J, the line segment joining $(x, \varphi(x))$ to $(y, \varphi(y))$ lies above the graph. That is, given $\alpha \geq 0$, and $\beta \geq 0$ with $\alpha + \beta = 1$,

$$\varphi(\alpha \cdot x + \beta \cdot y) \leq \alpha \cdot \varphi(x) + \beta \cdot \varphi(y).$$

The function φ is **strictly convex** on J if the above inequality is strict when $\alpha > 0$ and $\beta > 0$.

Remark 6.4.1. For an indication of the relationship between convexity and Jensen's Inequality, fix an open interval J containing points x and y with $x < y$. Let ν be a measure on the closed interval $[x,y]$ with $\nu(\{x\}) = \alpha$ and $\nu(\{y\}) = \beta$, while $\nu((x,y)) = 0$. This is a probability measure on $X := [x,y]$ because the total measure is 1. Now for the function f given by $f(x) = x$ and $f(y) = y$ together with a convex function φ on $[x,y]$, we have

$$\varphi(\int_X f \, d\nu) \leq \int_X \varphi \circ f \, d\nu.$$

Proposition 6.4.1. *A real valued function φ defined on an open interval $J \subseteq \mathbb{R}$ is convex if and only if for any triple $x < y < z$ in J,*

$$(*) \quad \frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(z) - \varphi(y)}{z - y}.$$

If φ is convex on J, then on any close interval $[a,b] \subset J$, φ satisfies a Lipschitz condition, and is therefore (Problem 5.21) absolutely continuous.

Proof. Let $y = \alpha \cdot x + \beta \cdot z$. If $(y, \varphi(y))$ is on the line from $(x, \varphi(x))$ to $(z, \varphi(z))$, then we have equality in (*). On the other hand, $(y, \varphi(y))$ is below the line if and only if the left side of (*) is less than the right side of (*). Suppose φ is convex on J, and let $[a,b]$ be a closed interval in J. Fix $c \in J$ and $d \in J$ with $c < a$ and $b < d$. Then for points $x < y$ in $[a,b]$,

$$\frac{\varphi(a) - \varphi(c)}{a - c} \leq \frac{\varphi(x) - \varphi(a)}{x - a} \leq \frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(b) - \varphi(y)}{b - y} \leq \frac{\varphi(d) - \varphi(b)}{d - b}.$$

The constant $M = \left|\frac{\varphi(a)-\varphi(c)}{a-c}\right| + \left|\frac{\varphi(d)-\varphi(b)}{d-b}\right|$ works as the desired Lipschitz constant.

Corollary 6.4.1. *If φ has an increasing derivative at all points of J, then φ is convex.*

Proof. This follows from the mean-value theorem of calculus.

Theorem 6.4.1 (Jensen's Inequality). *Let $(X, \mathscr{A}, \mu)$ be a probability space (i.e., $\mu(X) = 1$), and fix an integrable function f on X with range contained in an open interval $J \subseteq \mathbb{R}$. Let φ be a convex function on J. Then $\varphi(\int_X f\, d\mu) \leq \int_X \varphi \circ f\, d\mu$.*

Proof. Set $y := \int_X f d\mu$. Then $y \in J$ (Problem 6.19). Now, if

$$\beta = \sup_{x<y,\ x\in J} \frac{\varphi(y)-\varphi(x)}{y-x}$$

then for any $z > y$ in J, $\beta \leq \frac{\varphi(z)-\varphi(y)}{z-y}$. If $z < y$ in J, then by the definition of β, $\beta \geq \frac{\varphi(y)-\varphi(z)}{y-z} = \frac{\varphi(z)-\varphi(y)}{z-y}$. It follows for both positive and negative values of $z-y$ that $\varphi(z) \geq \varphi(y) + \beta \cdot (z-y)$ holds for all z in J. For any $t \in X$, set $z = f(t)$, so

$$\varphi(f(t)) \geq \varphi(y) + \beta(f(t)-y).$$

In particular, $\varphi(f(t))$ is bounded below by an integrable function. Since φ is absolutely continuous on any closed interval contained in J, it is the difference there of two increasing continuous functions, so $\varphi \circ f$ is measurable. Since $\mu(X) = 1$, when we integrate both sides of the inequality for $\varphi(f(t))$, we have the desired result

$$\varphi\left(\int_X f\, d\mu\right) = \varphi(y) + \beta \cdot 0 \leq \int_X \varphi \circ f\, d\mu.$$

Example 6.4.1. Let $\varphi(x) = e^x$ on $J = \mathbb{R}$. Let X be a set consisting of n points in $\mathbb{R}$ with each point having probability $1/n$. Let f be the function given by $x \mapsto x$ on X. Then

$$\exp\left(\frac{x_1 + \cdots + x_n}{n}\right) \leq \frac{e^{x_1} + \cdots + e^{x_n}}{n}.$$

Putting $y_i = e^{x_i}$, this gives the classical inequality between the arithmetic and geometric mean for positive numbers:

$$(y_1 \cdot y_2 \cdots y_n)^{\frac{1}{n}} \leq \frac{y_1 + y_2 + \cdots + y_n}{n}.$$

A generalization works with positive weights α_i such that $\Sigma \alpha_i = 1$. That is,

$$y_1^{\alpha_1} \cdots y_n^{\alpha_n} \leq \alpha_1 y_1 + \cdots + \alpha_n y_n.$$

For a countable number of points, we have

$$\Pi_{n=1}^{\infty} y_i^{\alpha_i} \leq \Sigma_{n=1}^{\infty} \alpha_i y_i.$$

6.5 Problems

Problem 6.1. Prove Proposition 6.1.1. **Hint:** See Proposition 2.4.4.

Problem 6.2. Let $\mathscr{M}$ be the collection of all subsets $A \subseteq \mathbb{R}$ such that either A or $\mathbb{R} \setminus A$ is finite or countably infinite. For each $A \in \mathscr{M}$, let $\mu(A) = 0$ if A is finite or countably infinite, and let $\mu(A) = 1$ otherwise. Show that $\mathscr{M}$ is a σ-algebra and μ is a measure on $\mathscr{M}$.

Problem 6.3. Let $X = \mathbb{R}$, and let $\mathscr{B}$ be the Borel subsets of $\mathbb{R}$. Show that every continuous real-valued function f is measurable with respect to the measurable space $(X, \mathscr{B})$.

Problem 6.4. Show that any nonnegative measurable function f is the limit of an increasing sequence of simple functions, with the obvious meaning of the limit at points where f takes the value $+\infty$.

Problem 6.5. Prove Proposition 6.2.5. **Hint:** For addition, apply Corollary 6.2.3 and Proposition 6.2.2.

Problem 6.6. State and prove a simplified form of Theorem 6.2.2 for the case that a single function g can replace all of the functions g_n.

Problem 6.7. Let $(X, \mathscr{B}, \mu)$ be a finite measure space, and let g and h be two bounded measurable functions on X. Suppose that for each integrable function f on X, $\int_X f \cdot g \, d\mu = \int_X f \cdot h \, d\mu$. Show then $g(x) = h(x)$ for μ-almost all $x \in X$.

Problem 6.8. Let $([0,1], \mathscr{B}, \mu)$ be a measure space, and let $f_n : [0,1] \mapsto [0,\infty)$ be sequence μ-integrable functions such that $\int_0^1 f_n \, d\mu = 1$ and $\int_{1/n}^1 f_n \, d\mu < 1/n$ for all $n \in \mathbb{N}$. Show that $\int_0^1 (\sup_{n\in\mathbb{N}} f_n) d\mu = +\infty$.

Problem 6.9. Let $(\Omega, \mathscr{B}, \mu)$ be a σ-finite measure space, and let $f : \Omega \mapsto \mathbb{R}$ be a measurable real-valued function. Suppose there exists a constant $c > 0$ such that for all $X \subset \Omega$ of finite measure we have $|\int_X f \, d\mu| \leq c$. Show that f is integrable on Ω.

Problem 6.10. Let $(X, \mathscr{B}, \mu)$ be a measure space, and let f be an integrable, extended-real valued function on $(X, \mathscr{B}, \mu)$. Show that for any $\varepsilon > 0$, there is a $\delta > 0$ such that if $E \in \mathscr{B}$ and $\mu(E) < \delta$, then $|\int_E f \, d\mu| \leq \int_E |f| \, d\mu < \varepsilon$.

Problem 6.11. Let $(X, \mathscr{B}, \mu)$ be a measure space, and let f be an integrable, extended-real valued function on $(X, \mathscr{B}, \mu)$. Prove the following inequality called **Chebyshev's Inequality:** Given $a > 0$ in $\mathbb{R}$, $\mu(\{|f| \geq a\}) \leq \frac{1}{a} \int |f| \, d\mu$.

Problem 6.12. **a)** Let $(X, \mathscr{B}, \mu)$ be a finite measure space. Extend Egoroff's Theorem 3.3.4 to the case of measurable real-valued functions on $(X, \mathscr{B}, \mu)$.

b) Show that $f_n \to f$ in measure if and only if every subsequence $\langle f_{n_k} \rangle$ has a further subsequence $\langle f_{n_{k_j}} \rangle$ converging μ-a.e. to f.

c) Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a continuous function, and let $\langle f_n : n \in \mathbb{N} \rangle$ be a sequence of measurable real-valued functions on X converging to f in measure. Show that $g \circ f_n \to g \circ f$ in measure.

Problem 6.13. **a)** Find a Lebesgue measurable function $f : \mathbb{R} \mapsto \mathbb{R}$ such that the Lebesgue measure $\lambda(f^{-1}[n]) > 0$ for all $n \in \mathbb{N}$.
b) Find an example of a measure space $(X, \mathscr{B}, \mu)$ and a measurable function $f : X \mapsto \mathbb{R}$ such that $0 < \mu(f^{-1}[t]) < \infty$ for all $t \in \mathbb{R}$.

Problem 6.14. **a)** Let ν be a signed measure, and let μ_1, μ_2 be positive measures on X such that $\nu = \mu_1 - \mu_2$. Show that $\mu_1 \geq \nu^+$ and $\mu_2 \geq \nu^-$, where $\nu = \nu^+ - \nu^-$ is the Jordan decomposition of ν.
b) Suppose that ν_1 and ν_2 are signed measures that either both omit the value $+\infty$ or both omit the value $-\infty$. Show that $|\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|$.

Problem 6.15. Let μ be a positive measure on $(X, \mathscr{B})$, and let f be a μ-integrable function. Show that ν defined by $\nu(E) = \int_E f \, d\mu$ for any measurable set $E \subseteq X$ is a signed measure.

Problem 6.16. Let γ be a signed measure on $(X, \mathscr{B})$.

a) Suppose f is a γ-integrable function on X. Show that $|\int f \, d\gamma| \leq \int |f| \, d|\gamma|$.
b) Extend, where possible, the definition of the integral of a function g with respect to γ to cases where g is not integrable with respect to $|\gamma|$.
c) Show that there is a simple function h that may not be integrable with respect to $|\gamma|$, such that for any $E \in \mathscr{B}$, $\int_E h \, d\gamma = |\gamma|(E)$.

Problem 6.17. Let $\langle \nu_i \rangle$ be a sequence of positive measures on $(X, \mathscr{B})$, and let μ be a positive measure on $(X, \mathscr{B})$. Suppose $\nu_i \perp \mu$ for all $i \in \mathbb{N}$. Show that $\mu \perp \sum_{i \in \mathbb{N}} \nu_i$.

Problem 6.18. **a)** Let $\mathscr{B}$ be the Borel σ-algebra on $\mathbb{R}$. Given $N \in \mathbb{N}$, give an example of a signed measure μ on $(\mathbb{R}, \mathscr{B})$ that takes every value in the interval $[-N, N]$.
b) Is there a signed measure on $(\mathbb{R}, \mathscr{B})$ that takes very real value?

Problem 6.19. Let $(X, \mathscr{A}, \mu)$ be a probability space (i.e., $\mu(X) = 1$), and fix an integrable function f on X with range contained in an open interval $J \subseteq \mathbb{R}$. Show that $y := \int_X f d\mu$ is a point in the interval J.

Chapter 7
Introduction to Metric and Normed Spaces

7.1 Metric Spaces

In this chapter, we extend the notion of distance and absolute value from the real and complex number systems to more general spaces, in particular, spaces of functions.

Definition 7.1.1. A nonempty set X supplied with a distance function d is called a **metric space**. The pair is denoted by (X,d). The distance function d is called a **metric**. It is a nonnegative, real-valued function on $X \times X$ such that for all points x, y, and z in X,

1) (**positive definite**) $d(x,y) = 0$ if and only if $x = y$,
2) (**symmetric**) $d(x,y) = d(y,x)$, and
3) (**triangle inequality**) $d(x,y) \leq d(x,z) + d(z,y)$.

Example 7.1.1. Euclidean n-space with the usual distance function is an example of a metric space. In particular, $\mathbb{R}$ is a metric space with the usual distance function $d(x,y) = |x-y|$.

Example 7.1.2. If we identify functions that are equal almost everywhere, then on a set E of finite measure in $\mathbb{R}$, convergence in measure with respect to a measure m is given by the metric $d(f,g) = \int_E |f-g| \wedge 1 \, dm$.

For the rest of this section, we work in a metric space (X,d). The notions of neighborhoods and convergence at points $x \in X$ are defined in terms of **ε-balls** of the form $B(x,\varepsilon) := \{y \in X : d(x,y) < \varepsilon\}$. These generalize the use of open intervals in $\mathbb{R}$ and open disks in $\mathbb{C}$.

Definition 7.1.2. A set O is **open** if for each $x \in O$ there is an ε-ball $B(x,\varepsilon) \subseteq O$ for some $\varepsilon > 0$. A set C is **closed** if the complement is open; that is, if for each $x \notin C$ there is an ε-ball $B(x,\varepsilon)$ with $B(x,\varepsilon) \cap C = \varnothing$.

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_7

We leave proofs of the following 4 properties of (X,d) as exercises.

Proposition 7.1.1. *For each $x \in X$ and $\varepsilon > 0$, the ε-ball $B(x,\varepsilon)$ is an open set.*

Remark 7.1.1. For emphasis one often speaks of an open ball.

Proposition 7.1.2. *If $x \neq y$ in X, then for some $\varepsilon > 0$, $B(x,\varepsilon) \cap B(y,\varepsilon) = \varnothing$.*

Proposition 7.1.3. *The set X and the empty set are both open subsets of X. The open subsets of X are stable with respect to the operations of taking finite intersections and arbitrary unions. The closed subsets of X are stable with respect to the operations of taking finite unions and arbitrary intersections.*

Definition 7.1.3. A sequence $\langle x_n : n \in \mathbb{N}\rangle$ in X **converges** to a point $x \in X$, and x is the **limit** of the sequence, if for any $\varepsilon > 0$ there is a $k \in \mathbb{N}$ such that for all $n \geq k$, $x_n \in B(x,\varepsilon)$; that is, the sequence is eventually in $B(x,\varepsilon)$. A point $x \in X$ is a **cluster point** of a sequence $\langle x_n : n \in \mathbb{N}\rangle$, if for each $\varepsilon > 0$ and each $k \in \mathbb{N}$, there is an $n \geq k$ with $x_n \in B(x,\varepsilon)$; that is, the sequence is frequently in $B(x,\varepsilon)$.

Proposition 7.1.4. *A sequence can have at most one limit, but many cluster points.*

Example 7.1.3. The sequence $n \mapsto (-1)^n$ has two cluster points in $\mathbb{R}$, while the sequence $n \mapsto n$ has none.

Proposition 7.1.5. *If x is a cluster point of a sequence $\langle x_n : n \in \mathbb{N}\rangle$ in X, then a subsequence $\left\langle x_{n_k} : k \in \mathbb{N}\right\rangle$ converges to x.*

Proof. There is an $n_1 \in \mathbb{N}$ with $x_{n_1} \in B(x,1)$. Given $k \geq 1$ and n_k, there is an $n_{k+1} > n_k$ with $x_{n_{k+1}} \in B(x, \frac{1}{k+1})$. The subsequence $\left\langle x_{n_k} : k \in \mathbb{N}\right\rangle$ converges to x.

Definition 7.1.4. Given a subset A of X, we write $\overline{A}$ for the set of all points $x \in X$ such that for each $\varepsilon > 0$, $B(x,\varepsilon) \cap A \neq \varnothing$. The set $\overline{A}$ is called the **closure** of A, and points of $\overline{A}$ are called **closure points** of A.

Proposition 7.1.6. *If A is a nonempty subset of X, then $A \subseteq \overline{A}$. Moreover, $\overline{(\overline{A})} = \overline{A}$; that is, the closure of the closure is the closure. Any cluster point of a sequence in $\overline{A}$ is in $\overline{A}$. For a metric space, it is also true that any point of $\overline{A}$ is the limit of a sequence in A.*

Proof. It is clear from the definition that $A \subseteq \overline{A}$. If $z \in \overline{(\overline{A})}$, then for each $\varepsilon > 0$ there is a point $y \in B(z,\varepsilon) \cap \overline{A}$. Since $B(z,\varepsilon)$ is open, there is a $\delta > 0$ with $B(y,\delta) \subset B(z,\varepsilon)$. Since $y \in \overline{A}$, there is an $x \in A \cap B(y,\delta) \subseteq A \cap B(z,\varepsilon)$. If follows that $\overline{(\overline{A})} = \overline{A}$. If z is a cluster point of a sequence in $\overline{A}$, then for each $\varepsilon > 0$, there is a point of the sequence in $B(z,\varepsilon)$, whence $z \in \overline{(\overline{A})} = \overline{A}$. If $w \in \overline{A}$, then for each $n \in \mathbb{N}$, there is a point $x_n \in A \cap B(w,1/n)$. The sequence $\langle x_n \rangle$ converges to w.

Remark 7.1.2. We have used the fact, true for a metric space, that for each $x \in X$ and each open set O containing x, there is an $n \in \mathbb{N}$ with $B(x,1/n) \subseteq O$.

Proposition 7.1.7. *A subset A of a metric space (X,d) is closed if and only if $A = \overline{A}$, and this is true if and only if every sequence in A converging to a point of X actually converges to a point of A.*

Proof. Exercise 7.6.

Proposition 7.1.8. *If $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$. The closure of A is a closed set; it is the intersection of all closed subsets of X containing A, and is, therefore, the smallest closed subset of X that contains A.*

Proof. Exercise 7.7.

Definition 7.1.5. A sequence $\langle x_n : n \in \mathbb{N}\rangle$ in (X,d) is a **Cauchy sequence** if for any $\varepsilon > 0$ there is a $k \in \mathbb{N}$ such that for all $n \geq k$ and $m \geq k$, $d(x_n, x_m) < \varepsilon$. A sequence $\langle x_n : n \in \mathbb{N}\rangle$ in (X,d) is **bounded** if for some $x \in X$ and $M \in \mathbb{N}$, $x_n \in B(x,M)$ for all $n \in \mathbb{N}$. A metric space (X,d) is **complete** if every Cauchy sequence in X converges to a point of X.

Example 7.1.4. By definition, a metric space is nonempty. The real and complex numbers are complete metric spaces. The continuous real-valued functions on a closed and bounded interval $[a,b]$ form a complete metric space using the distance $d(f,g) = \max_{x\in[a,b]} |f(x) - g(x)|$ (Problem 7.25). Completeness here follows from the fact that the uniform limit of continuous functions is continuous (Theorem 1.11.4).

Proposition 7.1.9. *A Cauchy sequence is a bounded sequence. A convergent sequence is a Cauchy sequence.*

Proof. Exercise 7.8.

If S is a nonempty subset of a metric space (X,d), then S is a metric space when supplied with the restriction of the metric d to pairs of points in S.

Proposition 7.1.10. *If $S \subset X$ and (S,d) is complete, then S is a closed subset of X. If $S \subset X$ is closed and (X,d) is complete, then (S,d) is complete.*

Proof. If (S,d) is complete, and $x \in X$ is the limit of a sequence in S, then that sequence is Cauchy. Since S is complete, $x \in S$. By Proposition 7.1.7, S is closed. If S is a closed subset of a complete metric space (X,d), any Cauchy sequence in S has a limit x in X. Since S is closed, $x \in S$. It follows that S is complete.

Definition 7.1.6. Given metric spaces (X,d) and (Y,ρ), let f be a function from X into Y. The function f is continuous at $x \in X$ if for any $\varepsilon > 0$ there is a $\delta > 0$ such that the image $f[B(x,\delta)] \subseteq B(f(x),\varepsilon)$. The function f is continuous on X if it is continuous at each point of X. The function f is uniformly continuous on $A \subseteq X$ if for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any pair of points x and y in A, if $d(x,y) < \delta$, then $\rho(f(x), f(y)) < \varepsilon$.

Proposition 7.1.11. *Let (X,d) and (Y,ρ) be metric spaces, and let $f : X \mapsto Y$ be uniformly continuous on X. If $\langle x_n : n \in \mathbb{N} \rangle$ is a Cauchy sequence in (X,d), then $\langle f(x_n) : n \in \mathbb{N} \rangle$ is a Cauchy sequence in (Y,ρ).*

Proof. Exercise 7.9.

Example 7.1.5. Let $X = \mathbb{R}$, and let $Y = (-1,1)$, both with the usual metric. Let $f : X \mapsto Y$ be the function defined by setting $f(x) = x/(1+|x|)$ for all $x \in \mathbb{R}$. Then f is bijective, and both f and f^{-1} are continuous. In fact, for any x and z in $\mathbb{R}$, $|f(x) - f(z)| \le |x - z|$, whence f is uniformly continuous on $\mathbb{R}$. On the other hand, $X = \mathbb{R}$ is complete, but $Y = (1,1)$ is not complete.

Definition 7.1.7. A subset A of a metric space (X,d) is **dense** in X if the closure of A equals X. A metric space is **separable** if there is a countable dense subset of the space.

Remark 7.1.3. Note that A is dense in X if and only if for every $x \in X$ and for every $\varepsilon > 0$, there is an $a \in A$ with $d(a,x) < \varepsilon$. That is, A has a nonempty intersection with every open ball.

Example 7.1.6. The rational numbers are dense in $\mathbb{R}$ supplied with the usual distance function. Therefore, $\mathbb{R}$ is a separable metric space.

Remark 7.1.4. In what follows, we write $B_c(x,r)$ for the set $\{y \in X : d(x,y) \le r\}$; the set is called the **closed ball** centered at x of radius r. One should be somewhat careful with balls in a general metric space. For example, if every point is distance 1 from every other point, as with the set of natural numbers supplied with the usual distance, then the open ball $B(x,1)$ consists just of the point x, so the closure of the open ball is not the closed ball $B_c(x,1)$. It is, however, an exercise (7.15) to show that a closed ball $B_c(x,r)$ is a closed set containing $B(x,r)$.

7.2 Baire Category

We now establish an important principle for a complete metric space that yields many unexpected results.

Theorem 7.2.1 (Baire). *Given a complete metric space (X,d), the countable intersection of dense open subsets of X is dense in X.*

Proof. Let O_n denote the nth dense open subset. Given any open ball $B(x_0,r_0)$, we must show that there is a point x that is in $B(x_0,r_0)$ and also in $\cap_n O_n$. We do this by finding a decreasing sequence of closed balls $B_c(x_n,r_n) \subseteq O_n \cap B(x_0,r_0)$ with $r_n \to 0$. We require that for every $n \in \mathbb{N}$, $B_c(x_{n+1},r_{n+1}) \subseteq B(x_n,r_n)$. Since O_1 is dense, there is a point $x_1 \in B(x_0,r_0) \cap O_1$ and an r_1 with $0 < r_1 < 1$ such that $B_c(x_1,r_1) \subseteq B(x_0,r_0) \cap O_1$. Similarly, there is a point $x_2 \in B(x_1,r_1) \cap O_2$ and an r_2 with $0 < r_2 < 1/2$ such that $B_c(x_2,r_2) \subseteq B(x_1,r_1) \cap O_2$. Continuing in this way, the x_n's form a Cauchy sequence for which the limit x must be in all of the closed balls $B_c(x_n,r_n)$. It follows that $x \in B(x_0,r_0) \cap (\cap_n O_n)$.

Definition 7.2.1. A subset of a metric space is **nowhere dense** if its closure contains no open balls; that is, the complement of the closure is dense. A set is said to be of the **Baire first category** or **meager** if it is the countable union of nowhere dense sets. A set is said to be of the **Baire second category** if it is not of the first category.

Note that the category notions are relative to the overall metric space X. We will show, for example, that the real line $\mathbb{R}$ is second category as a subset of itself. On the other hand, $\mathbb{R}$ is a closed, nowhere dense subset of the plane.

Proposition 7.2.1. *If E is a set of first category in (X,d) and $A \subseteq E$, then A is a set of the first category in X.*

Proof. Exercise 7.16(A).

Theorem 7.2.2 (Baire Category Theorem). *A set X forming a complete metric space with metric d is a set of the second Baire category as a subset of itself.*

Proof. Let E_n be a sequence of nowhere dense sets in X. The complements of the closures of the E_n's form a countable collection of dense open sets. The intersection of those complements is dense and therefore cannot be empty. It follows that the space X is not the union of the E_n's.

Proposition 7.2.2. *Let E be a subset of a complete metric space (X,d) such that the complement of E is a dense subset of X. Then any closed subset of E is nowhere dense, and any F_σ subset of E is a set of the first category in X. It follows that if E and its complement $\widetilde{E}$ are both dense in X, then at most one of them is an F_σ set.*

Proof. Exercise 7.17.

Corollary 7.2.1. *The rational numbers $\mathbb{Q}$ do not form a G_δ set in $\mathbb{R}$.*

Proof. Exercise 7.18.

Here is an important application called the **general uniform boundedness principle**.

Theorem 7.2.3. *Let $\mathscr{F}$ be a family of real-valued continuous functions on a complete metric space X. Suppose that for each $x \in X$, there is a positive constant M_x such that $\sup_{f\in\mathscr{F}} |f(x)| \leq M_x$. Then there is a nonempty open set O and a constant M such that for every $x \in O$ and every $f \in \mathscr{F}$, $|f(x)| \leq M$.*

Proof. For each $m \in \mathbb{N}$ and each $f \in \mathscr{F}$, let $E_{m,f} = \{x \in X : |f(x)| \leq m\}$. Set $E_m = \cap_{f\in\mathscr{F}} E_{m,f}$. Since each $f \in \mathscr{F}$ is continuous, each $E_{m,f}$ is closed, and so each E_m is closed. By assumption, $X = \cup_m E_m$. By the Baire Category Theorem, at least one of the E_m's contains an open set O.

7.2.1 Application to Differentiable Functions

Let C be the set of continuous functions on $[0,1]$. A metric ρ on C is given by $\rho(f,g) = \max_{x\in[0,1]} |f(x)-g(x)|$. Convergence in this metric is uniform convergence. This makes C a complete metric space (Problem 7.25). We now show that most (in terms of Baire category) of the functions of C fail to have a derivative at any point of $[0,1]$.

Suppose $f \in C$ has a right-hand derivative at some point x_f in $[0,1)$. Then f has the property that for some $n \in \mathbb{N}$, there is a δ with $0 < \delta \le 1-x_f$ such that $\left|\frac{f(x)-f(x_f)}{x-x_f}\right| \le n$, i.e., $\left|f(x)-f(x_f)\right| \le n\left(x-x_f\right)$ for all $x \in (x_f, x_f+\delta]$. Since f is bounded on $[0,1]$, we may assume, perhaps by increasing n, that $\left|f(x)-f(x_f)\right| \le n\left(x-x_f\right)$ for all $x \in [x_f, 1]$ and $x_f \le 1-\frac{1}{n}$. Let F_n denote the set of functions in C with this property for some point $x_f \in [0, 1-\frac{1}{n}]$.

For each $n \in \mathbb{N}$, F_n is a closed subset of C. That is, if $\langle f_k \rangle$ is a sequence in F_n and $f_k \to f$ uniformly, then $f \in F_n$. To see this, we set $x_k := x_{f_k}$. By taking a subsequence, we may assume that the points x_k converge to a point $\bar{x}$ with $0 \le \bar{x} \le 1-\frac{1}{n}$. Fix a point y with $\bar{x} < y \le 1$. Since $x_k \to \bar{x} < y$, we need only consider those $k \in \mathbb{N}$ with $x_k < y$. For each such k,

$$\begin{aligned}
&|f(y)-f(\bar{x})| \\
&\le |f(y)-f_k(y)| + |f_k(y)-f_k(x_k)| + |f_k(x_k)-f(x_k)| + |f(x_k)-f(\bar{x})| \\
&\le 2\cdot \max_{z\in[0,1]} |f(z)-f_k(z)| + n(y-x_k) + |f(x_k)-f(\bar{x})|\,.
\end{aligned}$$

As $k \to \infty$, $\max_{z\in[0,1]} |f(z)-f_k(z)| \to 0$ by the uniform convergence of f_k to f, $|f(x_k)-f(\bar{x})| \to 0$ since f is continuous, and $n(y-x_k) \to n(y-\bar{x})$, whence $|f(y)-f(\bar{x})| \le n(y-\bar{x})$, and thus, $f \in F_n$.

The closed set F_n is nowhere dense, since any $g \in C$ can be uniformly approximated to within an arbitrary $\varepsilon > 0$ by a polygonal path with the absolute value of the right-hand derivative everywhere $\ge 2n$. That is, the complement of F_n is dense in C. The subset of C consisting of functions that have a finite right derivative for at least one point of $[0,1)$ is the union $\cup_{n\in N} F_n$, and is therefore a set of the first category. Similarly, the functions in C that have a finite left derivative for at least one point of $(0,1]$ form a set of first category in C. The union of these two sets is also of the first category. Since a subset of a set of first category is of the first category, the functions in C with a derivative anywhere in $[0,1]$, including at the endpoints, is a set of the first category.

7.3 Normed Spaces

A norm is a mapping from a vector space, also called a **linear space**, into the non-negative real numbers. Since our application here will be to spaces of functions and equivalence classes of functions, we write f, g, etc., for elements of the linear space V. We use lower case Greek letters for elements of the scalar field, which is usually the real numbers $\mathbb{R}$, but can be the complex numbers $\mathbb{C}$. We use $|\alpha|$ to denote the absolute value of α for a real-valued scalar; it is the modulus of α if α is a complex scalar. We use 0 to denote the additive identity in both the scalar field and the linear space.

Definition 7.3.1. A **norm** is a mapping $\|\cdot\|$ from a linear space V into the nonnegative real numbers with the following properties:

1) (**positive definite**) $\|f\| = 0$ if and only if $f = 0$ in V.
2) (**homogeneous**) $\|\alpha f\| = |\alpha| \, \|f\|$ for each $f \in V$ and α in the scalar field.
3) (**triangle inequality**) $\|f+g\| \le \|f\| + \|g\|$ for all $f, g \in V$.

We call a linear space with a norm a **normed space**.

Example 7.3.1. The absolute value function on $\mathbb{R}$ and the modulus on $\mathbb{C}$ are examples of norms on normed spaces. Another example is the Euclidean norm on $\mathbb{R}^m$. That is, if $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, then $\|\mathbf{x}\| = (\sum_{i=1}^m x_i^2)^{1/2}$. This makes $\mathbb{R}^m$ into a normed space.

There is a natural **metric** d on a normed space, namely, $(f,g) \mapsto d(f,g) := \|f-g\|$. By Property 1, $d(f,g) = 0$ if and only if $f = g$. Moreover, $d(f,g) = d(g,f)$ since $\|g-f\| = \|(-1)(f-g)\|$. The triangle inequality follows from the fact that for all f, g, and h in V,

$$d(f,h) = \|f-h\| = \|(f-g)+(g-h)\| \le \|f-g\| + \|g-h\| = d(f,g) + d(g,h).$$

Remark 7.3.1. For a normed space X and all x, y in X, $\|x\| \le \|x-y\| + \|y\|$, and $\|y\| \le \|x-y\| + \|x\|$, whence $|\|x\| - \|y\|| \le \|x-y\|$. This means that $x \mapsto \|x\|$ is a continuous map from X into $\mathbb{R}^+$.

We next define some important norms on spaces of equivalence classes of measurable functions.

7.4 Classical Normed Spaces

In this section, we work with a measure space $(X, \mathscr{B}, \mu)$. We don't assume μ is a complete measure, but $f = g$ almost everywhere means that we have equality outside of a null set, that is, a set of μ-measure 0; nothing is said about what happens on the null set. Throughout this section, two functions are **equivalent** if they are equal almost everywhere. We use notation such as f for both the function f and the equivalence class it represents. Moreover, $|f|$ denotes the absolute value of f when f is extended-real valued, and the modulus of f when f is complex-valued.

Definition 7.4.1. The **essential supremum** of a nonnegative, measurable, extended-real valued function f that is bounded outside of a null set is the infimum of values α such that $f^{-1}[(\alpha, +\infty)]$ is a null set. If f is not bounded outside of a null set, then the essential supremum is $+\infty$. If f is an equivalence class of measurable functions, i.e., functions equal a.e., then $\|f\|_\infty$ denotes the essential supremum of $|f|$, where f is any representative of the equivalence class. The space $L^\infty(\mu)$ consists of those equivalence classes of measurable functions f for which $\|f\|_\infty$ is finite; such an f is said to be **essentially bounded**.

Definition 7.4.2. For $1 \le p < \infty$, the space $L^p(\mu)$ consists of those equivalence classes of measurable functions f such that $\int |f|^p \, d\mu$ is finite. For each $f \in L^p(\mu)$,

$$\|f\|_p := \left[\int |f|^p\right]^{1/p}.$$

For $1 \le p \le \infty$, $\|f\|_p$ is well-defined since if f and g are measurable functions with $f = g$ μ-a.e., then $\|f\|_p = \|g\|_p$. The space $L^p(\mu)$ (also denoted by just L^p) is a linear space with respect to the scalar field. That is, it is stable with respect to the operation of scalar multiplication, and it is stable with respect to addition since for $1 \le p < \infty$, $|f+g|^p \le 2^p(|f|^p + |g|^p)$, and the sum of two essentially bounded functions is essentially bounded.

Recall Example 6.2.1. When dealing only with counting measure on $\mathbb{N}$, we write ℓ^p instead of L^p. The space ℓ^1 is the space of absolutely summable sequences, while for real scalars, the space ℓ^2 is the space of square summable sequences. The space ℓ^∞ is the space of bounded sequences, since only the empty set has zero counting measure.

A modification of ℓ^p uses counting measure on a finite set such as an initial segment of $\mathbb{N}$. It is instructive to consider the surface S of the unit ball centered at 0 when using counting measure and real scalars on a set consisting of 2 points. Points of that space correspond to points of the plane. For $p = 1$, S is the diamond shaped curve consisting of the line segments $y = 1 - x$ for $x \in [0,1]$, $y = -1 + x$ for $x \in [0,1]$, $y = 1 + x$ for $x \in [-1,0]$, and $y = -1 - x$ for $x \in [-1,0]$. For $p = 2$, S is the circle of radius 1 about the origin. For $p = \infty$, S is the rectangle formed by the line segments $y = \pm 1$ for $x \in [-1,1]$ and $x = \pm 1$ for $y \in [-1,1]$.

Note that if $\mu(X) < +\infty$, then for $p < q$, $L^q(\mu) \subseteq L^p(\mu)$; we need only consider the integral where the integrand is greater than 1. If $\mu(X) = +\infty$, then the containment goes the other way for bounded functions; that is, consider the integral where the integrand is less than 1.

It will be important to establish relationships between L^p and L^q where $1/p + 1/q = 1$. Values p and q satisfying this equality are called **conjugate exponents**. An important example is when $p = q = 2$. The pair L^1 and L^∞ are also important in what follows.

Since $L^p(\mu)$ consists of equivalence classes, $\|f\|_p = 0$ if and only if each representative of the equivalence class f takes the value 0 μ-a.e. That is, the map $f \mapsto \|f\|_p$ is positive definite. Moreover the map is clearly homogeneous. To show, therefore, that $f \mapsto \|f\|_p$ is a norm, we need only establish the triangle inequality. For this, we need two important inequalities for the integral. We use the following special case of Jensen's Inequality.

Proposition 7.4.1. *Let x and y be nonnegative real numbers, and suppose $\alpha > 0$, $\beta > 0$, and $\alpha + \beta = 1$. Then*

$$x^\alpha y^\beta \leq \alpha \cdot x + \beta \cdot y.$$

Moreover, this inequality is an equality if and only if $x = y$.

Proof. If x or y is 0, the result is clear. Otherwise, let $x = e^s$ and $y = e^t$. By the strict convexity of the exponential function, we have

$$x^\alpha y^\beta = e^{\alpha s + \beta t} \leq \alpha \cdot e^s + \beta \cdot e^t = \alpha \cdot x + \beta \cdot y.$$

This inequality is an equality if and only if $s = t$, i.e., $x = y$.

Theorem 7.4.1 (Hölder's Inequality). *Assume either that p and q are real numbers larger than 1 with $1/p + 1/q = 1$ or that $p = 1$ and $q = \infty$. In either case, if $f \in L^p$ and $g \in L^q$, then $f \cdot g \in L^1$ and*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

The inequality is equality if the right side is 0. Otherwise, for $p = 1$, equality holds if and only if $|g(x)| = \|g\|_\infty$ for almost all x such that $f(x) \neq 0$, while for $p > 1$, equality holds if and only if there are positive constants s and t such that $s \cdot |f|^p = t \cdot |g|^q$ a.e.

Proof. For $p = 1$ and $q = \infty$, the result is Exercise 7.26. Suppose $1 < p < \infty$. We may assume that $\|f\|_p \cdot \|g\|_q \neq 0$ since otherwise the result is trivial. By Proposition 7.4.1, with $x = \left(\frac{|f|}{\|f\|_p}\right)^p$ and $y = \left(\frac{|g|}{\|g\|_q}\right)^q$, since $1/p + 1/q = 1$,

$$\frac{|f|}{\|f\|_p} \cdot \frac{|g|}{\|g\|_q} \leq \frac{1}{p} \cdot \left(\frac{|f|}{\|f\|_p}\right)^p + \frac{1}{q} \cdot \left(\frac{|g|}{\|g\|_q}\right)^q.$$

Integrating, we have

$$\frac{\|fg\|_1}{\|f\|_p \cdot \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

and the inequality follows.

By the condition for equality in Proposition 7.4.1, the inequality is actually an equality if and only if almost everywhere we have

$$\left(\frac{|f|}{\|f\|_p}\right)^p = \left(\frac{|g|}{\|g\|_q}\right)^q,$$

that is, $\|g\|_q^q \cdot |f|^p = \|f\|_p^p \cdot |g|^q$ a.e. If the latter equality holds, then clearly, there are positive constants s and t such that $s \cdot |f|^p = t \cdot |g|^q$ a.e. Conversely, if such constants exist, then

$$\|f\|_p^p = \int |f|^p = \frac{t}{s} \int |g|^q = \frac{t}{s} \cdot \|g\|_q^q,$$

so

$$\|g\|_q^q \cdot |f|^p = \frac{s}{t} \cdot \|f\|_p^p \cdot \frac{t}{s} \cdot |g|^q = \|f\|_p^p \cdot |g|^q \text{ a.e.}$$

The next result establishes the triangle inequality for $\|f\|_p$.

Theorem 7.4.2 (Minkowski's Inequality). *If for $1 \le p \le \infty$, f and g are in L^p, then so is the sum $f+g$ and*

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

Proof. For $p = 1$ and $p = \infty$, the result is an easy exercise (7.27). Assume that $1 < p < \infty$ and f and g are in L^p. As already noted,

$$|f+g|^p \le (|f|+|g|)^p \le 2^p\,(|f|^p + |g|^p),$$

so $f+g \in L^p$. We may assume that $\|f\|_p \cdot \|g\|_p \neq 0$, since otherwise the result is trivial. Now,

$$\int |f+g|^p \le \int \left(|f+g|^{p-1} \cdot |f|\right) + \int \left(|f+g|^{p-1} \cdot |g|\right).$$

Choose q so that $1/p + 1/q = 1$. Then $(p+q)/pq = 1$, so $p+q = pq$, whence $p = q\cdot(p-1)$. Now by Hölder's Inequality

$$\int \left(|f+g|^{p-1} \cdot |f|\right) \le \|f\|_p \cdot \left(\int |f+g|^p\right)^{1/q} = \|f\|_p \cdot \|f+g\|_p^{p/q}.$$

Similarly,

$$\int \left(|f+g|^{p-1} \cdot |g|\right) \le \|g\|_p \cdot \left(\int |f+g|^p\right)^{1/q} = \|g\|_p \cdot \|f+g\|_p^{p/q}.$$

It follows that

$$\|f+g\|_p^p \le \left(\|f\|_p + \|g\|_p\right) \cdot \|f+g\|_p^{p/q}.$$

Since $p - p/q = p(1-1/q) = p\,[(1/p+1/q) - 1/q] = 1$, the inequality follows.

Theorem 7.4.3. *For $1 \le p \le \infty$, the map $f \mapsto \|f\|_p$ is a norm on L^p.*

Definition 7.4.3. For L^p with $1 \le p < \infty$, norm convergence is often called **convergence in the mean of order** p. Convergence in L^∞ is often called **nearly uniform convergence**.

The last result of this section is an important application of the Fatou Lemma. We will later use the result to establish the Radon-Nikodým Derivative Theorem using Hilbert space techniques. Recall that the norm on an L^p space is used to generate a metric on the space.

Theorem 7.4.4 (Riesz–Fisher). *An L^p space, with $1 \le p \le \infty$, is a complete metric space.*

Proof. We give the proof for the case $1 \le p < \infty$; the case $p = \infty$ is an exercise (7.35(A)). We write $\|\cdot\|$ for the L^p-norm. Given a Cauchy sequence $\langle f_n : n \in \mathbb{N} \rangle$ in L^p, we replace the sequence with a subsequence such that for $m > n$, $\|f_m - f_n\| < 2^{-n}$. We will show that the subsequence converges to a limit f. As noted in Problem 7.12, it will follow that the original sequence converges to the same limit. For each $k \in \mathbb{N}$, we set

$$g_k = \sum_{n=1}^{k} |f_{n+1} - f_n|, \text{ and } \quad g = \sum_{n=1}^{\infty} |f_{n+1} - f_n|.$$

By Fatou's Lemma,

$$\begin{aligned}
\|g\|^p &= \int \left(\sum_{n=1}^{\infty} |f_{n+1} - f_n| \right)^p = \int \lim_k \left(\sum_{n=1}^{k} |f_{n+1} - f_n| \right)^p \\
&\le \liminf_k \int \left(\sum_{n=1}^{k} |f_{n+1} - f_n| \right)^p = \liminf_k \|g_k\|^p \\
&\le \liminf_k \left(\sum_{n=1}^{k} \|f_{n+1} - f_n\| \right)^p \le 1.
\end{aligned}$$

It follows that $g(x)$ is finite μ-a.e., so the series

$$f_1(x) + \sum_{n=1}^{\infty} (f_{n+1}(x) - f_n(x))$$

converges absolutely μ-a.e. Where we have convergence, denote the limit by $f(x)$; elsewhere, set $f(x) = 0$. By definition, $f_n(x) \to f(x)$ μ-a.e. Given $n \in \mathbb{N}$, we have again by Fatou's Lemma

$$\int |f - f_n|^p \le \liminf_{m \to \infty} \int |f_m - f_n|^p = \liminf_{m \to \infty} \|f_m - f_n\|^p \le (2^{-n})^p.$$

It follows that $f - f_n \in L^p$, whence $f = (f - f_n) + f_n \in L^p$. It also follows that $\lim_{n \to \infty} \|f - f_n\| = 0$.

Remark 7.4.1. A normed space, such as an L^p space, that is complete with respect to the metric generated by the norm is called a **Banach space**. Coincidently, Banach called such a space a space of type B. We will study these spaces in greater detail after first considering the special example of Hilbert spaces.

7.5 Linear Functionals

Let $(V, \|\cdot\|)$ be a normed linear space with real or complex scalars. We don't necessarily assume that V is complete with respect to the metric generated by the norm. As usual, $|\alpha|$ is the absolute value of α if $\alpha \in \mathbb{R}$, and $|\alpha|$ is the modulus of α if $\alpha \in \mathbb{C}$.

Definition 7.5.1. A linear map F from V into the scalar field is called a **linear functional**. If there is a nonnegative constant M such that for all $x \in V$, $|F(x)| \leq M \cdot \|x\|$, then F is called a **bounded linear functional** on V. The infimum of the values M that work for all x in V is called the **norm** of F; it is denoted by $\|F\|$.

Proposition 7.5.1. *If F is a bounded linear functional on V, then the norm $\|F\| = \sup_{\|x\|=1} |F(x)|$.*

Proof. $\|F\| = \sup_{x\neq 0} \frac{|F(x)|}{\|x\|} = \sup_{x\neq 0} \left| F\left(\frac{x}{\|x\|}\right)\right| = \sup_{\|x\|=1} |F(x)|$.

Theorem 7.5.1. *If a linear functional F on V is continuous at just one point, then it is bounded. On the other hand, a bounded linear functional is uniformly continuous on V. Therefore, a linear functional is continuous, and even uniformly continuous, on V if and only if it is bounded.*

Proof. Assume F is continuous at a point $x \in V$. Then for any $\varepsilon > 0$ there is a $\delta > 0$ such that the open ball $B(x,\delta) = \{y \in V : \|x-y\| < \delta\}$ about the point x maps into the interval $(F(x)-\varepsilon, F(x)+\varepsilon)$ about $F(x)$. (For complex scalars, replace an interval in $\mathbb{R}$ with the appropriate open disk in $\mathbb{C}$.) Now the translate

$$B(x,\delta) - x := \{y : y = z - x,\ z \in B(x,\delta)\}$$

is the open ball $B(\mathbf{0},\delta)$, and $F[B(\mathbf{0},\delta)]$ is contained in $(F(x)-\varepsilon, F(x)+\varepsilon)$ translated by adding $-F(x)$. This is the interval (or disk) about 0 of radius ε. It follows that F is continuous at the point $0 \in V$. Moreover, F is a bounded linear functional, since $\|x\| < \delta$ if and only if $\left\|\frac{2}{\delta}x\right\| < 2$, so

$$\|F\| = \sup_{\|x\|=1} |F(x)| \leq \sup_{\|x\|<2} |F(x)| = \frac{2}{\delta} \cdot \sup_{x\in B(\mathbf{0},\delta)} |F(x)| < \frac{2\varepsilon}{\delta}.$$

The uniform continuity follows from the fact that for all x, y in V,

$$|F(x) - F(y)| = |F(x-y)| \leq \|F\| \cdot \|x-y\|.$$

That is, F satisfies a Lipschitz condition with Lipschitz constant $\|F\|$.

Remark 7.5.1. We will show later that the bounded linear functionals form a Banach space, and $F \mapsto \|F\|$ is the norm on that space. The next result, which we will refine later, is an immediate consequence of Hölder's Inequality.

Proposition 7.5.2. *Assume either that p and q are real numbers larger than 1 with $1/p + 1/q = 1$ or that $p = 1$ and $q = \infty$. In either case, if $g \in L^q$, then the map $f \mapsto \int f \cdot g$ is a bounded linear functional F_g on L^p with $\left\|F_g\right\| \leq \|g\|_q$.*

7.6 Problems

Problem 7.1. Show that the triangle inequality holds in Example 7.1.2.

Problem 7.2. Prove Proposition 7.1.1.

Problem 7.3. Prove Proposition 7.1.2.

Problem 7.4. Prove Proposition 7.1.3.

Problem 7.5. Prove Proposition 7.1.4.

Problem 7.6. Prove Proposition 7.1.7.

Problem 7.7. Prove Proposition 7.1.8.

Problem 7.8. Prove Proposition 7.1.9.

Problem 7.9. Prove Proposition 7.1.11.

Problem 7.10. **a)** Give an example of a set X and two metrics d, ρ on X such that (X,d) is complete, but (X,ρ) is not complete.
b) Prove or give a counterexample: If X is a finite set and d is a metric on X, then (X,d) is complete.

Problem 7.11. Let (X,d) be a metric space, and let $\langle x_n \rangle$ be a Cauchy sequence on X. Show that there exists a subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ such that $\sum_{k\in\mathbb{N}} d(x_{n_k}, x_{n_{k+1}}) < \infty$.

Problem 7.12. Show that if a subsequence of a Cauchy sequence converges to a point x, then the original sequence converges to x.

Problem 7.13. Let f be a continuous mapping from a metric space (X,d) onto a metric space (Y,ρ). Show that if (X,d) is separable, then (Y,ρ) is separable.

Problem 7.14. Let (X,d) and (Y,ρ) be two metric spaces, and let f be a uniformly continuous bijection from X onto Y. Assume that $f^{-1} : Y \mapsto X$ is also continuous. If Y is complete, must X also be complete? Compare with Example 7.1.5.

Problem 7.15. Show that a closed ball $B_c(x,r)$ in a metric space is a closed set.

Problem 7.16 (A). Prove that if E is a set of first category in X and $A \subseteq E$, then A is a set of the first category in X.

Problem 7.17. Prove Proposition 7.2.2.

Problem 7.18. Prove Corollary 7.2.1.

Problem 7.19. **a)** Show that a countable union of sets of first category is also of first category.

b) To construct a general Cantor set C_n of measure $1-1/n$, one follows a process similar to the construction of a Cantor set. At each stage, however, one removes from the middle of each closed interval that remains, an open interval of length $1/n$ times a third of the length of the closed interval from which it is removed. Show that a general Cantor set C_n in $[0,1]$ is closed, has Lebesgue measure $1-\frac{1}{n}$, and is nowhere dense.

c) Does there exist a set A of the first Baire category in $[0,1]$ supplied with the usual metric such that the Lebesgue measure of A is 1?

d) Does there exist a set B of the second Baire category in $[0,1]$ supplied with the usual metric such that the Lebesgue measure of B is 0?

Problem 7.20 (A). A point x in a metric space is called **isolated** if the singleton set $\{x\}$ is open. For example, the set of natural numbers $\mathbb{N}$ with the usual distance function forms a countable, complete metric space such that all points are isolated; notice that every Cauchy sequence is eventually constant. Prove that a complete metric space without isolated points has an uncountable number of points. (Note this implies that $[0,1]$ is uncountable.)

Problem 7.21. Suppose one shows that the rational numbers in $[0,1]$ have Lebesgue measure 0 as follows: Order the rationals and center an open interval of length $1/(2^m n)$ about the mth rational so that the rationals are contained in the union, which is an open set O_n of measure at most $1/n$. Can it be that the rationals are exactly the set $\cap_{n=1}^{\infty} O_n$?

Problem 7.22. Show that the characteristic function of the rationals, i.e., $\chi_{\mathbb{Q}}$, is not the pointwise limit of a sequence of continuous function on $\mathbb{R}$. **Hint:** If there were such a sequence $\langle f_n \rangle$, what could you say about the sequence $n \mapsto f_n^{-1}[(1/2,\infty)]$?

Problem 7.23. Either give an example of the following sequence $\langle A_n : n \in \mathbb{N} \rangle$ of subsets of $\mathbb{R}$ or state why no such sequence can exist: If I is any open interval in $\mathbb{R}$, then for each set A_n, there is a point in I that is not the limit of any sequence of points in A_n, and $\mathbb{R} = \cup_n A_n$.

Problem 7.24. Recall the construction in Problem 5.12 of a function continuous at each irrational number in $[0,1]$ and discontinuous at each rational number in $[0,1]$. Also recall Problem 1.43 characterizing the set of points of continuity of a real-valued function as a G_δ set. Is there a function continuous at each rational number in $[0,1]$ and discontinuous at each irrational number in $[0,1]$?

Problem 7.25. Let C be the set of continuous functions on $[a,b]$. A metric ρ on C is given by $\rho(f,g) = \max_{x\in[a,b]} |f(x)-g(x)|$. Show that (C,ρ) is a complete metric space.

Problem 7.26. Prove Theorem 7.4.1 for the case $p=1$ and $q=\infty$.

Problem 7.27. Prove Minkowski's Inequality for the cases $p=1$ and $p=\infty$.

Problem 7.28. Let $(X,\mathscr{B},\mu)$ be a measure space, and fix p with $1 \le p \le \infty$. Show that if f and g are in $L^p(\mu)$, then $\max(f,g) \in L^p(\mu)$.

Problem 7.29. Show that $\|f\|_p = \left(\int_0^1 |f|^p \, d\lambda\right)^{1/p}$ does not define a norm on $[0,1]$ supplied with Lebesgue measure when $0 < p < 1$. **Hint:** Let $f = \chi_{[0,1/2)}$ and $g = \chi_{[1/2,1]}$.

Problem 7.30. Suppose f is a nonnegative, real-valued, Lebesgue measurable function on $[0,1]$, and the product $x \cdot (f(x))^3$ has a finite Lebesgue integral on $[0,1]$. Use the fact that $\frac{1}{3} + \frac{2}{3} = 1$ together with the fact that $x^{-1/3}\left(x^{1/3} f(x)\right) = f(x)$ for $x > 0$ to show that f has a finite integral on $[0,1]$.

Problem 7.31. For each $k \in \mathbb{N}$, fix a sequence $\left\langle a_i^{(k)} : i \in \mathbb{N} \right\rangle$ in ℓ^1 with norm $\sum_{i \in \mathbb{N}} \left| a_i^{(k)} \right| \leq 1$. For each $i \in \mathbb{N}$, assume that $\lim_{k \to \infty} a_i^{(k)}$ exists, and denote that limit by a_i. Use Fatou's Lemma to show that the sequence $\langle a_i : i \in \mathbb{N} \rangle$ is in ℓ^1 with norm $\sum_{i \in \mathbb{N}} |a_i| \leq 1$.

Problem 7.32. Fix a sequence $\langle a_i : i \in \mathbb{N} \rangle$ in ℓ^1. Show that there is a signed measure μ defined on the power set of $\mathbb{N}$ such that for every sequence $\mathbf{b} = \langle b_i : i \in \mathbb{N} \rangle$ in ℓ^∞, $\sum_{i=1}^\infty b_i \cdot a_i = \int_{\mathbb{N}} \mathbf{b} \, d\mu$.

Problem 7.33. Suppose $f_n \to f$ almost everywhere and all functions are in L^p for $1 \leq p < \infty$. Show that $\|f_n - f\| \to 0$ if and only if $\|f_n\| \to \|f\|$. **Hint:** Use the General Lebesgue Dominated Convergence Theorem, which is a consequence of Theorem 4.6.2. That is, let $g_n = 2^p |f_n|^p + 2^p |f|^p$ for each $n \in \mathbb{N}$, and note that $g_n \to g := 2^{p+1} |f|^p$ almost everywhere.

Problem 7.34. Fix a measure space $(X, \mathscr{B}, \mu)$, and fix p with $1 \leq p < \infty$. Let $\langle f_n \rangle$ be a Cauchy sequence in L^p such that for $n \geq k$ in $\mathbb{N}$, $\|f_n - f_k\|_p < 2^{-k}$ and $f_n(x)$ converges to $f(x)$ at all $x \in X$. Show that f is in L^p and $\lim_{k \to \infty} \|f - f_k\|_p = 0$. **Hint:** For each $k \in \mathbb{N}$, $f = (f - f_k) + f_k$.

Problem 7.35 (A). Fix a measure space $(X, \mathscr{B}, \mu)$.

a) Let $\langle f_n \rangle$ be a sequence in L^∞. Prove that $f_n \to f$ in L^∞ if and only if there is a set E of measure 0 such that f_n converges to f uniformly on the complement of E.
b) Prove that L^∞ is complete. Note that given the generality of the underlying measure space, this also proves that ℓ^∞ is complete.

Problem 7.36. Fix a measure space $(X, \mathscr{B}, \mu)$, and let $\langle f_n \rangle$ be a sequence in $L^2(\mu)$ such that for some $M > 0$ and all $n \in \mathbb{N}$, $\|f_n\|_2 \leq M$. Show that $f_n(x)/n \to 0$ for μ-almost all $x \in X$. **Hint:** Consider $\sum_{n \in \mathbb{N}} \int_X (f_n/n)^2 d\mu$.

Problem 7.37. Fix a measure space $(X, \mathscr{B}, \mu)$, and fix p with $1 \leq p < \infty$.

a) Prove or give a counterexample: if $f_n \to f$ in L^p, then $f_n(x) \to f(x)$ for μ-almost all $x \in X$.
b) Fix p and q with $1 \leq p < q < \infty$. Show that if $f_n \to f$ in L^p and $f_n \to g$ in L^q, then $f = g$ μ-a.e.

Problem 7.38. Let $(X,\mathscr{B},\mu)$ be a measure space, and fix p and q with $1 \le p < \infty$, $1 < q \le \infty$, and $1/p + 1/q = 1$. In particular, if $p = 1$, $q = \infty$. Let $\langle f_n \rangle$ be a sequence in $L^p(\mu)$, and let $\langle g_n \rangle$ be a sequence in $L^q(\mu)$. Suppose $f_n \to f$ in L^p and $g_n \to g$ in L^q. Show that $f_n g_n \to fg$ in measure.

Problem 7.39 (A). Given that the space c consisting of convergent real-valued sequences is a linear subspace of $\ell^\infty(\mathbb{N})$, show that it is a Banach space; that is, show that it is complete with respect to the ℓ^∞-norm.

Problem 7.40. a) Use the distance function on $L^\infty[0,1]$ (with respect to Lebesgue measure) to show that $L^\infty[0,1]$ cannot contain a countable dense subset. **Hint:** For each r, $s \in (0,1)$ with $r \ne s$, $\|\chi_{[0,r)} - \chi_{[0,s)}\|_\infty = ?$

b) Describe a countable family of step functions that are dense in $L^1[0,1]$ (with respect to Lebesgue measure).

Problem 7.41 (A). Fix a measure space $(X,\mathscr{B},\mu)$. Fix $f \in L^p(\mu)$, $1 \le p < \infty$. Given $\varepsilon > 0$, show that there is a simple function φ vanishing outside a set of finite measure such that $\|f - \varphi\|_p < \varepsilon$.

Problem 7.42. a) Show that the set E of integrable real-valued functions that take the value 1 on $(1/2,1)$ form a closed subset of $L^1(\lambda)$, where λ is Lebesgue measure on $[0,1]$. Again, we consider two functions equal a.e. to represent the same thing, so E consists of integrable functions that take the value 1 almost everywhere on $[1/2,1]$.

b) Show that the set S of **continuous** real-valued functions on $[0,1]$ that take the value 1 on $(1/2,1)$ do **not** form a closed subset of $L^1(\lambda)$, where λ is Lebesgue measure on $[0,1]$.

Problem 7.43 (A). Let $(X,\mathscr{B},\mu)$ be a measure space, and let f be a function that is in every L^p for $1 \le p \le \infty$. Show that $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$.

Problem 7.44. Let $\langle q_n : n \in \mathbb{N} \rangle$ be an enumeration of the rational numbers, and set $f_n := n \cdot \chi_{[q_n, q_n + 2^{-n}]}$ for each $n \in \mathbb{N}$. Let $f = \sum_{n\in\mathbb{N}} f_n$. Show that $f \in L^p$ with respect to Lebesgue measure on $\mathbb{R}$ for every p such that $1 \le p < \infty$, but f is not essentially bounded on any non-degenerate interval. **Hint:** For sufficiently large k, $k \le 2^{k/2p}$. Moreover, $S = \sum_{k=1}^\infty 2^{-k/2p}$ is finite, as seen by considering $a \cdot S - S$ for an appropriate value of a.

Problem 7.45. Let F_n, $n \in \mathbb{N}$, and F be bounded linear functionals on a normed space $(V, \|\cdot\|)$. Assume that $\|F_n - F\| \to 0$. Let $\langle x_n \rangle$ be a sequence in V converging to x in V. Show that $F_n(x_n) \to F(x)$.

Problem 7.46. Let $C([0,1])$ denote the space of continuous real-valued functions on $[0,1]$. For each $f \in C([0,1])$, let $\|f\| = \max_{x\in[0,1]} f(x)$ be the norm of f obtained by restricting the $L^\infty(\lambda)$ norm with respect to Lebesgue measure to $C([0,1])$. Fix $g \in C([0,1])$. Let F be the linear functional defined by setting $F(f) = \int_0^1 f(t)g(t)\,d\lambda$ for each $f \in C([0,1])$. Show that F is bounded, and $\|F\| = \|g\|_1$, i.e., the $L^1(\lambda)$ norm of g. **Hint:** Let $h(x) = 1$ on the set $\{g \geq 0\}$, and let $h(x) = -1$ on the set $\{g < 0\}$. Use Lusin's Theorem 3.3.3 to show that h is the λ-a.e. limit of a sequence $\langle f_n \rangle$ in $C([0,1])$ with $\|f_n\| = 1$ for all $n \in \mathbb{N}$.

Problem 7.47 (A). Let $C([0,1])$ denote the space of continuous real-valued functions on $[0,1]$. Note that each $g \in C([0,1])$ is a member of an L^∞-equivalence class with respect to Lebesgue measure λ on $[0,1]$. Suppose $f \in L^\infty(\lambda)$ is a point of closure of the subspace of $L^\infty(\lambda)$ formed by $C([0,1])$. Show that f is a member of that subspace.

Chapter 8
Hilbert Spaces

8.1 Basic Definitions

A Hilbert space is, among other things, a linear space with either real or complex scalars; it is stable with respect to addition and scalar multiplication. To avoid repetition, we will work with the complex case, which will include the real case. That is, for a complex number $z = x + iy$ (x and y real), the **conjugate** is $\overline{z} = x - iy$. For a real number a, the conjugate $\overline{a} = a$. Therefore, if the scalar field is just the real numbers, then the conjugation operation $a \mapsto \overline{a}$ is just the identity operation on $\mathbb{R}$. For this reason, we can treat both the real and complex scalar cases at the same time. Here is the definition of the space.

Definition 8.1.1. An **inner product space** H is a linear space with real or complex scalars for which there is an **inner product** $(x,y) : H \times H \mapsto \mathbb{C}$, such that the following holds:

i) The inner product is linear in the first variable, i.e., for all $x,y,z \in H$ and all scalars α, β,
$$(\alpha x + \beta y, z) = \alpha(x,z) + \beta(y,z).$$
ii) It is complex conjugate symmetric, i.e., for all $x,y \in H$, $(x,y) = \overline{(y,x)}$ if the scalar field is $\mathbb{C}$, and $(x,y) = (y,x)$ if the scalar field is $\mathbb{R}$.
iii) The inner product is positive definite, i.e., for each $x \in H$, (x,x) is real and nonnegative, and $(x,x) = 0$ if and only if $x = 0$.

We will show that the mapping $x \mapsto \sqrt{(x,x)}$ is a norm on an inner product space. We will work with the completion with respect to the corresponding metric.

Definition 8.1.2. A **Hilbert space** is an inner product space that is complete with respect to the metric generated by the norm $x \mapsto \|x\| := \sqrt{(x,x)}$.

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_8

A Hilbert space, for which the scalar field is $\mathbb{R}$, is called a **real Hilbert space**. Note that by Properties i and ii, $(x, \alpha y + \beta z) = \overline{\alpha}\,\overline{(y,x)} + \overline{\beta}\,\overline{(z,x)}$ for all $\alpha, \beta \in \mathbb{C}$. If the scalar field is $\mathbb{R}$, then $(x, \alpha y + \beta z) = \alpha\,(y,x) + \beta\,(z,x)$. That is, the inner product is linear in both arguments; this is also called the **bilinear** property of the inner product for a real Hilbert space.

Example 8.1.1. **a)** Euclidean spaces form real Hilbert spaces with $(x,y) = x \cdot y$. That is, if $x = (x_1, x_2, \cdots, x_m)$ and $y = (y_1, y_2, \cdots, y_m)$ with real components x_i and y_i for each i, then the usual scalar product $(x,y) = x \cdot y = \sum_{i=1}^{m} x_i y_i$ is an inner product.

b) The sequence space ℓ^2 forms a Hilbert space with either real or complex scalars. Here, $(\langle x\rangle, \langle y\rangle) = \sum_{n=1}^{\infty} x_n \cdot \overline{y_n}$. We know by Hölder's inequality that this is finite. For any element $\langle x_n : n \in \mathbb{N}\rangle \in \ell^2$, we have the norm $\|\langle x_n\rangle\|_2 = \sqrt{(\langle x_n\rangle, \langle x_n\rangle)} = \left(\sum_{n=1}^{\infty} |x_n|^2\right)^{1/2}$. If the scalar field is $\mathbb{R}$, then $|x_n|$ is the absolute value of x_n; it is the modulus of x_n if the scalar field is $\mathbb{C}$.

c) Given a measure space $(X, \mathscr{B}, \mu)$, the space L^2 formed from equivalence classes of appropriate functions is a Hilbert space with either real or complex scalars. Here, $(f,g) = \int f \cdot \overline{g}$. We know by Hölder's inequality that this is finite. Of course, for any element $f \in L^2$, $\|f\|_2 = \sqrt{(f,f)}$.

We show next that for a general inner product space, a norm can be defined by setting $\|x\| = \sqrt{(x,x)}$. In doing so, we recall that for the special case of a function in L^2, the function represents 0 if and only if it takes the value 0 almost everywhere. In general, it follows from Property iii that $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$. Moreover, if a scalar α is real, $\|\alpha x\|^2 = (\alpha x, \alpha x) = \alpha^2 (x,x) = \alpha^2 \|x\|^2$, while if α is complex, $\|\alpha x\|^2 = (\alpha x, \alpha x) = \alpha\bar{\alpha}(x,x) = |\alpha|^2 \|x\|^2$. In either case, $\|\alpha x\| = |\alpha| \cdot \|x\|$. To finish, we need the triangle inequality, and for that we generalize the L^2 case of the Hölder Inequality. We prove that generalization for both real and complex scalars.

8.2 Basic Inequalities

Theorem 8.2.1 (Cauchy–Buniakowsky–Schwarz (CBS) Inequality). *Let H be an inner product space. For all $x, y \in H$,*

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

Proof. We may assume that neither x nor y is zero. Choose α with $|\alpha| = 1$ so that $\alpha(y,x) = |(x,y)|$. In the real case, $\alpha = 1$ or $\alpha = -1$. For the complex case where $(y,x) = re^{i\theta}$, $\alpha = e^{-i\theta}$. Note that $\bar{\alpha}(x,y) = |(x,y)|$ and $\alpha\overline{\alpha} = 1$. Now, for any positive real number λ,

$$\begin{aligned} 0 &\leq (x - \lambda\alpha y, x - \lambda\alpha y) = \|x\|^2 - \lambda[\alpha(y,x) + \bar{\alpha}(x,y)] + \lambda^2 \|y\|^2 \\ &= \|x\|^2 - 2\lambda\,|(x,y)| + \lambda^2 \|y\|^2. \end{aligned}$$

Setting $\lambda = \|x\|/\|y\|$, we have

$$2|(x,y)| \leq \frac{\|y\|}{\|x\|}\|x\|^2 + \frac{\|x\|}{\|y\|}\|y\|^2 = 2(\|x\| \cdot \|y\|).$$

Corollary 8.2.1. *The inner product is a continuous function on $H \times H$.*

Proof. Exercise 8.1.

Corollary 8.2.2 (Triangle Inequality). *For all x, $y \in H$, $\|x+y\| \leq \|x\| + \|y\|$.*

Proof. Since $(x,y)+(y,x) = 2\,\mathrm{Re}(x,y) \leq 2\,|(x,y)| \leq 2\|x\| \cdot \|y\|$,

$$\|x+y\|^2 = (x+y,x+y) = \|x\|^2 + (x,y) + (y,x) + \|y\|^2 \leq (\|x\| + \|y\|)^2.$$

Corollary 8.2.3. *The mapping $x \mapsto \|x\| := \sqrt{(x,x)}$ is a norm on H.*

Note that a Hilbert space is a particular example of a Banach space; that is, a complete normed linear space. Next, we generalize the equality that holds for the diagonals of parallelograms in the plane.

Proposition 8.2.1 (Parallelogram law). *Let H be an inner product space. For all x, y in H,*

$$\|x+y\|^2 + \|x-y\|^2 = 2\,\|x\|^2 + 2\,\|y\|^2.$$

Proof. Note that

$$\begin{aligned}
\|x+y\|^2 &= (x+y,x+y) = \|x\|^2 + (y,x) + (x,y) + \|y\|^2 \\
\|x-y\|^2 &= (x-y,x-y) = \|x\|^2 - (y,x) - (x,y) + \|y\|^2.
\end{aligned}$$

Adding these equations completes the proof.

8.3 Convex Sets, Orthogonality, and Bounded Linear Functionals

In this section we work with a fixed Hilbert space $H \neq \{0\}$.

Proposition 8.3.1. *For each $y \in H$, the mapping $x \mapsto (x,y)$ is a bounded linear functional on H. (See Definition 7.5.1.) Moreover, the norm of the functional is equal to $\|y\|$.*

Proof. By the CBS inequality, for all $x \in H$, $|(x,y)| \leq \|x\| \cdot \|y\|$, and by the definition of the norm, $|(y,y)| = \|y\| \cdot \|y\|$, so the functional has norm $\|y\|$.

Remark 8.3.1. We will show in Theorem 8.3.3 below that every bounded linear functional on H has the form $x \mapsto (x,y)$ for a unique $y \in H$.

A linear subspace of H inherits the inner product. If the subspace is closed, it is complete (Proposition 7.1.10), so it is again a Hilbert space. A **translate** of a subspace Y is a set of the form $Y+x := \{y+x : y \in Y\}$ for some $x \in H$.

Definition 8.3.1. A set is **convex** if it contains every line segment joining points in the set. That is, if x and y are in the set, so is $\alpha x + \beta y$ for all nonnegative $\alpha, \beta \in \mathbb{R}$ such that $\alpha + \beta = 1$.

Example 8.3.1. Any subspace, and all of its translates, is convex (Problem 8.2).

Theorem 8.3.1. *Let E be a closed, convex subset of H. Then E contains a unique element of smallest norm. That is, there is a unique element closest to the point* 0 *in H.*

Proof. Let $a = \inf\{\|x\| : x \in E\}$. Fix x and y in E. Applying the parallelogram law (8.2.1) to $\frac{1}{2}x$ and $\frac{1}{2}y$, we have

$$\frac{1}{4}\|x-y\|^2 = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \left\|\frac{x+y}{2}\right\|^2.$$

Since E is convex, $\frac{x+y}{2} \in E$, so

$$\|x-y\|^2 \le 2\|x\|^2 + 2\|y\|^2 - 4a^2. \tag{8.3.1}$$

Equation (8.3.1) shows that there can be at most one point in E with norm a. If now x and y are replaced by elements of a sequence $\langle x_n \rangle$ in E such that $\|x_n\| \to a$, then Equation (8.3.1) shows that the sequence is Cauchy, and so the limit x, which must be in E, has norm a.

Definition 8.3.2. Elements x and y in H are called **orthogonal**, and we write $x \perp y$, if $(x,y) = 0$. We write $x^\perp$ for the set of all y orthogonal to x. Given $A \subseteq H$, $A^\perp := \bigcap_{x\in A} x^\perp$.

Proposition 8.3.2. *For each $x \in H$, the space $x^\perp$ is a closed subspace of H.*

Proof. The fact that $x^\perp$ is a linear subspace follows from the linearity of the first argument of the inner product. This space is closed since it is the inverse image of 0 with respect to the continuous functional $y \mapsto (y,x)$.

Corollary 8.3.1. *Given $A \subseteq H$, $A^\perp = \bigcap_{x\in A} x^\perp$ is a closed subspace of H.*

The next fact is an easy generalization of the Pythagorean Theorem.

Proposition 8.3.3. *If $x \perp y$ in H, then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.*

Proof.

$$\|x+y\|^2 = (x+y, x+y) = \|x\|^2 + \|y\|^2 + (x,y) + (y,x) = \|x\|^2 + \|y\|^2.$$

The following result is central to our later calculations. We will give the proof for the case of complex scalars; the real case follows since the conjugate operation is the identity map on $\mathbb{R}$.

Theorem 8.3.2. *Let M be a closed subspace of H. For each $x \in H$, let Px denote the (unique) nearest point to x in M, that is, the element $y \in M$ such that $y - x$ is the element of smallest norm in $M - x$; let Qx denote the (unique) nearest point to x in $M^{\perp}$. Then $\|x\|^2 = \|Px\|^2 + \|Qx\|^2$. The mappings $P : H \mapsto M$ and $Q : H \mapsto M^{\perp}$ are linear, and for each $x \in H$, $x = Px + Qx$. Moreover, this is the only way to decompose x into an element of M plus an element of $M^{\perp}$.*

Proof. The case $M = \{0\}$, $M^{\perp} = H$ is clear. Given $x \in H$, let z be the element of smallest norm in $M + x$. Then $z = u + x$ for some $u \in M$, so $x - z = -u \in M$. Fix an arbitrary nonzero element $y \in M$, scaled so that $\|y\| = 1$. We will show that $z \in M^{\perp}$ by showing that $\alpha := (z,y) = 0$. Since $z \in M + x$, $z - \alpha y \in M + x$. Since z has the smallest norm in $M + x$,

$$\begin{aligned}
0 &\le \|z - \alpha y\|^2 - \|z\|^2 = (z - \alpha y, z - \alpha y) - (z,z) \\
&= -\alpha(y,z) - \overline{\alpha}\,(z,y) + \alpha\overline{\alpha}\,(y,y) = -\alpha\overline{\alpha} - \overline{\alpha}\alpha + \alpha\overline{\alpha} = -\alpha\overline{\alpha} \\
&= -\,|(z,y)|^2 \le 0.
\end{aligned}$$

Therefore, $(z,y) = 0$ for all $y \in M$, i.e., $z \in M^{\perp}$.

Given any element $y \in M$, we have $x - z - y \in M$. Since $z \in M^{\perp}$,

$$\|x - y\|^2 = \|z + [(x - z) - y]\|^2 = \|z\|^2 + \|x - z - y\|^2.$$

This is a minimum when $y = x - z$, so $x - z = Px$. Given any element $w \in M^{\perp}$, we have

$$\|x - w\|^2 = \|x - z + z - w\|^2 = \|x - z\|^2 + \|z - w\|^2.$$

This is a minimum when $z = w$, so $z = Qx$.

We now have a decomposition of x into an element of M and $M^{\perp}$, namely, $x = (x - z) + z$. To show that such a decomposition must be unique, we suppose that $x = x_1 + x_2$ is another decomposition with $x_1 \in M$ and $x_2 \in M^{\perp}$. Then $Px + Qx = x = x_1 + x_2$, so

$$Px - x_1 = x_2 - Qx \in M \cap M^{\perp} = \{0\},$$

whence $x_1 = Px$ and $x_2 = Qx$. To show that P and Q are linear, we note that for any x, $y \in H$ and scalars α and β,

$$\alpha[P(x) + Q(x)] + \beta[P(y) + Q(y)] = \alpha x + \beta y = P(\alpha x + \beta y) + Q(\alpha x + \beta y),$$

so

$$P(\alpha x + \beta y) - \alpha P(x) - \beta P(y) = \alpha Q(x) + \beta Q(y) - Q(\alpha x + \beta y) \in M \cap M^{\perp} = \{0\}.$$

Corollary 8.3.2. *If M is a closed subspace of H with $M \neq H$, then there is a nonzero element $y \in H$ such that $y \perp M$.*

Theorem 8.3.3. *If L is a bounded linear functional on H, then there is a unique $y \in H$ such that $L(x) = (x,y)$ for all $x \in H$. The norm of L is $\|L\| = \|y\|$.*

Proof. The element, if it exists, must be unique since if y and z both produce the same functional, then $L(y-z) = (y-z,y) = (y-z,z)$, whence $(y-z, y-z) = 0$, so $y-z = 0$. If $L = 0$, take $y = 0$. Otherwise, let K be the kernel of L, i.e., the closed, linear subspace consisting of all elements that L maps to 0. Fix $z \in K^{\perp}$ with $\|z\| = 1$. Then $L(z) = \alpha \neq 0$. Let $y = \bar{\alpha} z$. We will show that L is represented by y. Fix $x \in H$; we must show that $L(x) - (x,y) = 0$. Now $L(L(x)z - L(z)x) = 0$. This means that $L(x)z - L(z)x \in K$. Since $z \in K^{\perp}$ and $\|z\| = 1$,

$$\begin{aligned} L(x) - (x,y) &= L(x) - \alpha(x,z) = L(x)(z,z) - L(z)(x,z) = (L(x)z,z) - (L(z)x,z) \\ &= (L(x)z - L(z)x, z) = 0. \end{aligned}$$

Corollary 8.3.3. *If L is a nonzero, bounded linear functional on H, then $K^{\perp}$ is a one-dimensional space. That is, for any $z \neq 0$ in $K^{\perp}$ and any $w \in K^{\perp}$, $w = \gamma z$ for some scalar γ.*

Proof. Exercise 8.6.

8.4 Radon-Nikodým Theorem

We now have sufficient background to establish the Radon-Nikodým Derivative Theorem. We will start with the results for just finite measures μ and ν on a measurable space $(X, \mathscr{B})$ with $\nu << \mu$. This means that $\mathscr{B}$ is a σ-algebra of subsets of a set X, and the values taken by our measures for sets in $\mathscr{B}$ are finite. It also means that ν is absolutely continuous with respect to μ, i.e., sets of μ-measure 0 have ν-measure 0. We will then extend the result to the case that μ is a σ-finite measure, that is, X is a finite or countably infinite union of measurable sets of finite μ-measure, and ν is an arbitrary measure on $(X, \mathscr{B})$ absolutely continuous with respect to μ.

We will also establish as a corollary, the Lebesgue Decomposition Theorem for pairs μ and ν of σ-finite measures on $(X, \mathscr{B})$. For this result, recall that μ and ν are called mutually singular, and we write $\nu \perp \mu$, if there are disjoint measurable sets A and B with $X = A \cup B$ and $\mu(B) = \nu(A) = 0$.

Important for this section is the fact that given a measure space $(X, \mathscr{B}, \mu)$, the function space L^2 is a Hilbert space. In particular, we use the fact that L^2 is complete with respect to the metric derived from the L^2-norm. Moreover, as shown in Theorem 8.3.3, L^2 is its own space of bounded linear functionals. We need only work here with real scalars. The proof we now give of the Radon-Nikodým Theorem is due to John von Neumann.

Proposition 8.4.1. *Let $(X, \mathscr{B}, \mu)$ be a finite measure space, and fix an integrable function $g \geq 0$ on X. The set function $E \mapsto \nu(E) := \int_E g\, d\mu$ is a finite measure on $(X, \mathscr{B})$. Moreover, if $f \geq 0$ is measurable on $(X, \mathscr{B})$, then $\int f\, d\nu = \int fg\, d\mu$.*

Proof. It follows from Corollary 4.4.3 that ν is a finite measure on $(X,\mathscr{B})$. If $f \geq 0$ is measurable on X, and φ_n is an increasing sequence of simple functions converging up to f, then for all $n \in \mathbb{N}$, $\int \varphi_n \, d\nu = \int \varphi_n g \, d\mu$, so by the Monotone Convergence Theorem applied twice,

$$\int f \, d\nu = \lim_n \int \varphi_n \, d\nu = \lim_n \int \varphi_n g \, d\mu = \int fg \, d\mu.$$

Proposition 8.4.2. *Let μ and ν be finite measures on the measurable space $(X,\mathscr{B})$, and set $\lambda := \mu + \nu$. Define $F(f) := \int f \, d\mu$ for all $f \in L^2(\lambda)$. Then F is a bounded linear functional on $L^2(\lambda)$. Fix the $g \in L^2(\lambda)$ that represents F; that is, for each $f \in L^2(\lambda)$, $F(f) = \int f \, d\mu = \int fg \, d\lambda$. Then $0 \leq g \leq 1$ λ-a.e. on X, whence we may assume that the inequality holds at all points of X.*

Proof. The map F is clearly linear on $L^2(\lambda)$. If $f \in L^2(\lambda)$ with norm $\|f\|_2$, then by Hölder's Inequality,

$$|F(f)| = \left| \int f \, d\mu \right| \leq \int |f| d\mu \leq \int |f| d\lambda = \int |f \cdot 1| \, d\lambda \leq \|f\|_2 \cdot \|1\|_2.$$

Therefore, F is bounded. Fix the $g \in L^2(\lambda)$ that represents F, i.e., $(f,g) = F(f)$ for all $f \in L^2(\lambda)$. The inequality $0 \leq g \leq 1$ λ-a.e. follows from the fact that for every $E \in \mathscr{B}$,

$$0 \leq \int \chi_E \, d\mu = F(\chi_E) = \int g \cdot \chi_E \, d\lambda = \int_E g \, d\lambda = \int_E d\mu \leq \int_E d\lambda = \lambda(E).$$

That is, if for some $n \in \mathbb{N}$ there is a set $E \in \mathscr{B}$ such that $g \leq -\frac{1}{n}$ on E, then we must have $\lambda(E) = 0$, since $0 \leq \int_E g \, d\lambda$. Therefore, $\lambda(\{g < 0\}) = 0$. If there is a set $E \in \mathscr{B}$ such that $g \geq 1 + \frac{1}{n}$ on E, then we must have $\lambda(E) = 0$, since $\int_E g \, d\lambda \leq \lambda(E)$. Therefore, $\lambda(\{g > 1\}) = 0$.

Corollary 8.4.1. *Add the additional hypothesis that $\nu << \mu$. Then $\lambda(\{g = 0\}) = 0$. Moreover, for each set $E \in \mathscr{B}$,*

$$\mu(E) = \int_E g \, d\lambda, \quad \nu(E) = \int_E (1-g) \, d\lambda,$$
$$\lambda(E) = \int_E \frac{1}{g} \, d\mu, \quad \nu(E) = \int_E (1-g) \cdot \frac{1}{g} \, d\mu.$$

Proof. Given $E \in \mathscr{B}$,

$$\mu(E) = \int \chi_E \, d\mu = \int \chi_E \cdot g \, d\lambda = \int_E g \, d\lambda.$$
$$\nu(E) = \lambda(E) - \mu(E) = \int_E 1 \, d\lambda - \int_E g \, d\lambda = \int_E (1-g) \, d\lambda.$$

If $\mu(E) = 0$, then $\nu(E) = 0$, so $\lambda(E) = 0$. That is, $\lambda << \mu$. In this case, since $\mu(\{g = 0\}) = \int_{\{g=0\}} g \, d\lambda = 0$, $\lambda(\{g = 0\}) = 0$. By Proposition 8.4.2, for any f in

$L^2(\lambda)$, $\int f\, d\mu = \int f \cdot g\, d\lambda$. Set $f = \frac{1}{g}$ where $g > 0$ and $f = 1$ where $g = 0$. For each $n \in \mathbb{N}$, set $G_n := \{g \geq 1/n\}$. If E is a measurable subset of G_n, then $f \cdot \chi_E \in L^2(\lambda)$, and by Proposition 8.4.2,

$$\lambda(E) = \int_E 1\, d\lambda = \int_E f \cdot g\, d\lambda = F(f \cdot \chi_E) = \int_E \frac{1}{g}\, d\mu.$$

Since this is true for each $n \in \mathbb{N}$, it follows from the Monotone Convergence Theorem that for every measurable set $E \subseteq \{g > 0\}$, $\lambda(E) = \int_E \frac{1}{g}\, d\mu$. Since $\mu(\{g = 0\}) = 0$, $\lambda(E) = \int_E \frac{1}{g}\, d\mu$ for any set $E \in \mathscr{B}$. Since $\nu(E) = \int_E (1-g) d\lambda$, it again follows from the Monotone Convergence Theorem that $\nu(E) = \int_E (1-g) \cdot \frac{1}{g}\, d\mu$ for any measurable set $E \subseteq \{g > 0\}$, and thus for any $E \in \mathscr{B}$.

Let g be the function defined by Proposition 8.4.2, and let $h = (1-g) \cdot \frac{1}{g}$ on the set $\{g > 0\}$ and $h = 1$ on the μ-null set $\{g = 0\}$. Then h is a representative function of the **Radon-Nikodým derivative** of ν with respect to μ, denoted by $d\nu/d\mu$. That is, for each $E \in \mathscr{B}$, $\nu(E) = \int_E \frac{d\nu}{d\mu}\, d\mu$ when μ and ν are finite measures with $\nu << \mu$. Thus, we have established for finite measures the existence part of the following result.

Theorem 8.4.1 (Radon-Nikodým). *Let μ be a σ-finite measure on a measurable space $(X, \mathscr{B})$, and let ν be a measure on $(X, \mathscr{B})$ with $\nu << \mu$. Then there is a nonnegative measurable function, denoted by $d\nu/d\mu$, such that for any measurable $E \subseteq X$, $\nu(E) = \int_E (d\nu/d\mu)\, d\mu$. The function $d\nu/d\mu$ is unique except for changes on μ-null sets.*

Proof. As noted, we have already established the existence for the case of finite measures. For that case, let h and f be two functions with the properties of the Radon-Nikodým derivative. For any $n \in \mathbb{N}$, let $E_n = \{f > h + 1/n\}$. We then have

$$\nu(E_n) = \int_{E_n} f\, d\mu \geq \int_{E_n} h\, d\mu + \frac{1}{n}\mu(E_n) = \nu(E_n) + \frac{1}{n}\mu(E_n).$$

Since $\nu(E_n)$ is finite, it follows that $\mu(E_n) = 0$, and this is true for each $n \in \mathbb{N}$, so $f \leq h$ μ-a.e. Similarly, $h \leq f$ μ-a.e., so $h = f$ μ-a.e.

Now assume that μ is a finite measure and ν is an arbitrary measure that is absolutely continuous with respect to μ on $(X, \mathscr{B})$. For each $n \in \mathbb{N}$, let A_n and B_n be the positive set and negative set, respectively, for a Hahn decomposition of X with respect to the signed measure $\nu - n\mu$. (See Definition 6.3.1.) Let $S_1 = B_1$, and for each $n > 1$, let $S_n = B_n \setminus \cup_{i=1}^{n-1} B_i$. Then the sets S_n are disjoint, and $\cup_{n=1}^{\infty} S_n = \cup_{n=1}^{\infty} B_n$. For each $n \in \mathbb{N}$, $S_n \subseteq B_n$, so for any measurable subset E of S_n, $(\nu - n\mu)(E) \leq 0$, whence $\nu(E) \leq n\mu(E) < +\infty$. Therefore, there is a Radon-Nikodým derivative h_n on S_n for ν with respect to μ. We define h on $S := \cup_n S_n$ by setting $h(x) := h_n(x)$ when $x \in S_n$. Any subset E of S is the disjoint union $\cup_n (S_n \cap E)$, so h is a Radon-Nikodým derivative on S, and any other Radon-Nikodým derivative on S equals

h μ-a.e. If E is a measurable subset of $X \setminus S$ and $\mu(E) > 0$, then for each $n \in \mathbb{N}$, $E \subseteq A_n$, so $\nu(E) - n\mu(E) \geq 0$, whence $\nu(E) = +\infty$. We complete the formation of a Radon-Nikodým derivative $d\nu/d\mu$ on X by extending h to $X \setminus S$ with the value $+\infty$. Any other Radon-Nikodým derivative must take the same value on any subset of $X \setminus S$ having positive μ-measure.

Finally, we assume that μ is a σ-finite but not finite measure, and ν is an arbitrary measure absolutely continuous with respect to μ. It then follows that X is a countably infinite union of disjoint measurable sets E_i on which μ is a finite measure. Let ν_i be the restriction of ν to the collection of measurable subsets of E_i. Clearly $\nu_i << \mu$ on E_i. Let $f_i = d\nu_i/d\mu \geq 0$ on E_i and $f_i \equiv 0$ on $X \setminus E_i$. Then $f := \sum_{i=1}^{\infty} f_i$ is measurable. Moreover, f works as the Radon-Nikodým derivative, since for each measurable subset E of X,

$$\begin{aligned}\nu(E) &= \sum_{i=1}^{\infty} \nu_i(E \cap E_i) = \sum_{i=1}^{\infty} \int_{E \cap E_i} f_i \, d\mu = \sum_{i=1}^{\infty} \int_{E \cap E_i} f \, d\mu \\ &= \lim_n \int_{E \cap \left(\cup_{i=1}^n E_i\right)} f \, d\mu = \int_E f \, d\mu.\end{aligned}$$

If g also works, then on each set E_i, $g = f_i$ μ-a.e., so $g = f$ μ-a.e. on X.

Corollary 8.4.2 (Lebesgue Decomposition Theorem). *Suppose μ and ν are σ-finite measures on a measurable space $(X, \mathscr{B})$. Then ν has a unique decomposition $\nu = \nu_0 + \nu_1$, called the **Lebesgue decomposition** of ν with respect to μ, such that $\nu_0 \perp \mu$ and $\nu_1 << \mu$.*

Proof. Let $\lambda = \nu + \mu$. It is easy to see that X is a countable union of sets of finite λ-measure (Problem 8.12). Moreover, $\mu << \lambda$. Let $f = d\mu/d\lambda$. Let $A := \{f > 0\}$, and let $B := \{f = 0\}$. For any $E \in \mathscr{B}$, set $\nu_1(E) := \nu(E \cap A)$, and set $\nu_0(E) := \nu(E \cap B)$. Then $\nu = \nu_0 + \nu_1$. Moreover, $\nu_0 \perp \mu$ since $\mu(B) = 0$ and $\nu_0(A) = 0$. If $E \in \mathscr{B}$ and $\mu(E) = 0$, then $f = 0$ λ-a.e. on E. That is, $\lambda(E \cap A) = 0$. Since $\nu << \lambda$, $\nu(E \cap A) = \nu_1(E) = 0$. Thus $\nu_1 << \mu$. To show the uniqueness of the Lebesgue decomposition, suppose $\nu = \widetilde{\nu}_0 + \widetilde{\nu}_1$ is another decomposition with $\widetilde{\nu}_0 \perp \mu$ and $\widetilde{\nu}_1 << \mu$. Let C and D be disjoint measurable sets with $X = C \cup D$ and $\mu(D) = \widetilde{\nu}_0(C) = 0$. Let S be a measurable subset of X of finite ν-measure. On S, we have $\nu = \nu_0 + \nu_1 = \widetilde{\nu}_0 + \widetilde{\nu}_1$, whence on S the signed measure $\rho := \nu_0 - \widetilde{\nu}_0 = \widetilde{\nu}_1 - \nu_1$. Since $\nu_0(A) = \widetilde{\nu}_0(C) = 0$, $S \cap A \cap C$ is a null set for ρ. Since $\mu(B) = \mu(D) = 0$, $S \cap B \cap C$, $S \cap A \cap D$, and $S \cap B \cap D$ are all null sets for ρ. Thus, any measurable subset of X of finite ν-measure is a null set for $\nu_0 - \widetilde{\nu}_0 = \widetilde{\nu}_1 - \nu_1$. It follows that the decomposition is unique.

8.5 Orthonormal Families and Fourier coefficients

In this section, $H \neq \{0\}$ is a Hilbert space with complex scalars. By an **indexed family**, we mean a set for which there is given a bijection from a set acting as index onto the family.

Definition 8.5.1. A subset $S \subseteq H$ is called **linearly independent** if every finite linear combination with total 0 has all of the coefficients equal to 0. That is, if $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$ for some scalars α_i and $x_i \in S$, then each $\alpha_i = 0$. We write $\operatorname{span}(S)$ for the set of all finite linear combinations of elements of S.

If the elements of S are linearly independent, then there is only one way to write an element of $\operatorname{span}(S)$ as a finite linear combination of the elements of S, since the difference of two combinations with the same sum is a representation of 0. The span of S is the smallest vector space containing S (Exercise 8.14).

Definition 8.5.2. An **orthogonal family** is one for which every element is orthogonal to every other element. An **orthonormal family** is an orthogonal family for which every element has norm 1.

Remark 8.5.1. An indexed family $\{u_\alpha\}$ is an orthonormal family if and only if $(u_\alpha, u_\beta) = 1$ when $\alpha = \beta$ and $(u_\alpha, u_\beta) = 0$ otherwise. While the element $0 \in H$ is orthogonal to every element of H, it cannot be an element of an orthonormal family.

Example 8.5.1. The vectors **i**, **j**, **k** in $\mathbb{R}^3$ form an orthonormal family.

Proposition 8.5.1. *An orthonormal family is linearly independent.*

Proof. If in the family, a finite linear combination $s = \alpha_1 u_1 + \cdots + \alpha_n u_n = 0$, then for each i with $1 \le i \le n$, we have $0 = (s, u_i) = \alpha_i$.

Definition 8.5.3. Given an indexed orthonormal family $\{u_\alpha\}$, the αth **Fourier coefficient** of $x \in H$ is the complex number $\hat{x}(\alpha) := (x, u_\alpha)$.

Proposition 8.5.2. *Given an indexed orthonormal family $\{u_\alpha\}$, for each index α, the map $x \mapsto \hat{x}(\alpha)$ is linear and also continuous, since*

$$|\hat{x}(\alpha) - \hat{y}(\alpha)| = |(x - y, u_\alpha)| \le \|x - y\|.$$

Moreover, $\|a u_\alpha\| = |a|$ since $\|a u_\alpha\|^2 = (a u_\alpha, a u_\alpha) = |a|^2$.

The next result is the heart of what we need for Fourier series.

Proposition 8.5.3. *Let $\{u_\alpha \in H : \alpha \in I\}$ be a* ***finite*** *orthonormal family in H (i.e., I is a finite index set), and let M_I be its span. Then M_I is closed. If $y = \sum_{\alpha \in I} a_\alpha u_\alpha \in M_I$, then for each $\alpha \in I$, $\hat{y}(\alpha) = (y, u_\alpha) = a_\alpha$, and $\|y\|^2 = (y, y) = \sum_{\alpha \in I} |\hat{y}(\alpha)|^2$. Given any $x \in H$, let $s_I(x) := \sum_{\alpha \in I} \hat{x}(\alpha) u_\alpha$. Then $(x - s_I(x)) \in M_I^\perp$, so $s_I(x)$ is the unique closest element to x in M_I, $(x - s_I(x))$ is the unique element closest to x in $M_I^\perp$, and*

$$\|x\|^2 = \|s_I(x)\|^2 + \|x - s_I(x)\|^2,$$

whence

$$\|s_I(x)\|^2 = \sum_{\alpha \in I} |\hat{x}(\alpha)|^2 \le \|x\|^2.$$

Proof. If $y = \sum a_\alpha u_\alpha \in M_I$, then for each $\alpha \in I$,

$$\hat{y}(\alpha) = (y, u_\alpha) = \sum_{\beta \in I} a_\beta (u_\beta, u_\alpha) = a_\alpha,$$

and

$$\|y\|^2 = (y,y) = \sum_{\alpha \in I} \sum_{\beta \in I} \hat{y}(\alpha)\overline{\hat{y}}(\beta)(u_\alpha, u_\beta) = \sum_{\alpha \in I} |\hat{y}(\alpha)|^2.$$

In general, the span of a finite set in a linear space is closed, but here is a simple proof for our setting. If $\langle x_n : n \in \mathbb{N} \rangle$ is a sequence in M_I with limit $y \in H$, then by the continuity of the inner product, $(x_n - y, u_\alpha) \to 0$ for each $\alpha \in I$. Let $y_I = \sum_{\alpha \in I}(y, u_\alpha)u_\alpha \in M_I$. For each $n \in \mathbb{N}$, $x_n = \sum_{\alpha \in I}(x_n, u_\alpha)u_\alpha$, so

$$\begin{aligned} \|x_n - y_I\| &= \left\| \sum_{\alpha \in I}(x_n, u_\alpha)u_\alpha - \sum_{\alpha \in I}(y, u_\alpha)u_\alpha \right\| = \left\| \sum_{\alpha \in I}(x_n - y, u_\alpha)u_\alpha \right\| \\ &\le \sum_{\alpha \in I} \|(x_n - y, u_\alpha)u_\alpha\| = \sum_{\alpha \in I} |(x_n - y, u_\alpha)| \to 0. \end{aligned}$$

Since a limit is unique in H, $x_n \to y = y_I \in M_I$. It follows that M_I is closed.

Next, we note that for any $x \in H$ with $s_I(x) = \sum_{\alpha \in I} \hat{x}(\alpha)u_\alpha$, and any $z = \sum_{\beta \in I} b_\beta u_\beta \in M_I$,

$$\begin{aligned} (x - s_I(x), z) &= \left(x, \sum_{\beta \in I} b_\beta u_\beta \right) - \left(\sum_{\alpha \in I} \hat{x}(\alpha)u_\alpha, \sum_{\beta \in I} b_\beta u_\beta \right) \\ &= \sum_{\beta \in I} \overline{b}_\beta (x, u_\beta) - \sum_{\alpha \in I} \sum_{\beta \in I} \hat{x}(\alpha)\overline{b}_\beta (u_\alpha, u_\beta) \\ &= \sum_{\beta \in I} \overline{b}_\beta \hat{x}(\beta) - \sum_{\beta \in I} \hat{x}(\beta)\overline{b}_\beta = 0. \end{aligned}$$

The rest follows from Theorem 8.3.2.

Next we use the Axiom of Choice to show that there is a maximal orthonormal family in H, i.e., one not properly contained in a larger orthonormal family. Again, we assume that $H \neq \{0\}$. Note that for any nonzero $x \in H$, the singleton set $\{x/\|x\|\}$ is an orthonormal family.

Theorem 8.5.1. *Every orthonormal family B in H is contained in a* ***maximal*** *(also known as* ***complete****) orthonormal family in H.*

Proof. We use the Hausdorff Maximal Principle, which, as shown in the appendix, is equivalent to the Axiom of Choice. We partially order by containment the orthonormal families in H that contain B. That is, a family S_1 is further along in the ordering than a family S_0 if S_1 contains S_0. Let S be the union of a maximal (with respect to containment) linearly ordered family of orthonormal sets all containing

B. Then S is orthonormal since if u and v are in S, they are in some orthonormal family contained in S. By definition, S is maximal.

We now show that for a maximal orthonormal family in a Hilbert space, the closure of the span is H. Recall that the span of a set is the collection of all finite linear combinations of members of the set. By Problem 8.14, the span of a set is a linear space, and the closure of the span is also a linear space.

Theorem 8.5.2. *An orthonormal family $\{u_\alpha : \alpha \in A\}$ is maximal in H if and only if the closure of its span is all of H.*

Proof. Let S be the span of $\{u_\alpha : \alpha \in A\}$, and let M be the closure of S. Then M is a closed linear subspace of H. By Theorem 8.3.2, $M \neq H$ if and only if $M^\perp \neq \{0\}$. If $M^\perp \neq \{0\}$, then there is an element $u_0 \neq 0$ in $M^\perp$, which by normalization we may assume has norm 1. By definition, $(u_\alpha, u_0) = 0$ for each index $\alpha \in A$, so $\{u_\alpha : \alpha \in A\}$ is not maximal since adjoining u_0 produces a larger orthonormal family. On the other hand, if $\{u_\alpha : \alpha \in A\}$ is not maximal, then there is a u_0 of norm 1 that is orthogonal to every element of $\{u_\alpha : \alpha \in A\}$, whence it is orthogonal to S. By continuity of the inner product, u_0 is then orthogonal to M, whence $M^\perp \neq \{0\}$ and $M \neq H$.

8.6 Separability

Recall that a Hilbert space is a metric space using the distance function generated by the norm. Recall that such a space is called separable if there is a countable subset with closure equal to the whole space. There is a special relationship between separable Hilbert spaces and the ℓ^2 space formed from sequences of scalars. On the other hand, much of what is true for separable Hilbert spaces works without the assumption of separability if norm convergence of sequences is replaced with the norm convergence of unordered sums. We will discuss results for both the separable case using norm convergence of sequences and the non-separable case using unordered sums. We will use the fact shown in Theorems 8.5.1 and 8.5.2 that there exists a maximal orthonormal family in a Hilbert space H, and the closure of the span of that family is all of H.

Theorem 8.6.1. *A Hilbert space H is separable if and only if every orthonormal family in H is either finite or countably infinite.*

Proof. Suppose an orthonormal family in H for which the span is all of H is finite or countably infinite. Then finite linear combinations formed from members of the family using rational real and imaginary parts for the coefficients produce a dense subset of H, whence H is separable. On the other hand, if there exists an uncountable orthonormal family $\{u_\alpha \in H : \alpha \in A\}$, then for any two distinct members u_α and u_β of the family, $\|u_\alpha - u_\beta\|^2 = (u_\alpha - u_\beta, u_\alpha - u_\beta) = 2$, so the ball of radius

$\sqrt{2}/3$ about each member of the orthonormal family contains no other member of the orthonormal family, but must contain a point of a dense set. In this case, there cannot be a countable dense subset of H. In particular, there cannot be a finite or countably infinite orthonormal family with span equal to H.

8.7 Unordered Sums and ℓ^2 Spaces

To include non-separable Hilbert spaces in our discussion, we need facts about unordered sums for general normed spaces as well as scalar fields. If we are given an infinite collection of vectors or scalars $S = \{x_\alpha : \alpha \in A\}$ indexed by an index set A, we can make the collection $\mathscr{F}$ consisting of all finite subsets of S into a partially ordered set using containment. That is, a finite subset F_1 of S is "further along in the ordering" than a finite set F_0 of S if F_1 contains F_0. Given any two members F_1 and F_2 in $\mathscr{F}$, the union $F_1 \cup F_2$ is again in $\mathscr{F}$. Therefore, the map $F \mapsto \sum_{x_\alpha \in F} x_\alpha$ is a net with directed set $\mathscr{F}$. This net converges to x_0 if for each $\varepsilon > 0$ there is an $F_\varepsilon \in \mathscr{F}$ such that for each $F \in \mathscr{F}$ with $F \supseteq F_\varepsilon$, $\|\sum_{x_\alpha \in F} x_\alpha - x_0\| < \varepsilon$. We call x_0 the **unordered sum** of the collection $\{x_\alpha : \alpha \in A\}$. If a series in the scalar field is absolutely convergent, then the sum of the terms is an unordered sum; conversely, if a scalar series has an unordered sum, then the series is absolutely convergent (Problem 8.15).

The limit of a net in a normed space is unique. To see this for our special case, assume that x_0 and x_1 are such limits. Then for any $\varepsilon > 0$, there is an $F_\varepsilon \in \mathscr{F}$ and a $G_\varepsilon \in \mathscr{F}$ such that for any $F \in \mathscr{F}$ containing F_ε and any $G \in \mathscr{F}$ containing G_ε,

$$\left\| x_0 - \sum_{x_\alpha \in F} x_\alpha \right\| < \varepsilon/2 \quad \text{and} \quad \left\| x_1 - \sum_{x_\alpha \in G} x_\alpha \right\| < \varepsilon/2,$$

whence

$$\left\| x_0 - \sum_{x_\alpha \in F_\varepsilon \cup G_\varepsilon} x_\alpha \right\| < \varepsilon/2 \quad \text{and} \quad \left\| x_1 - \sum_{x_\alpha \in F_\varepsilon \cup G_\varepsilon} x_\alpha \right\| < \varepsilon/2,$$

and so $\|x_0 - x_1\| < \varepsilon$. Since ε is arbitrary, $x_0 = x_1$.

Note that in dealing with a finite unordered sum of terms x_α in any normed space X, including $\mathbb{R}$ and $\mathbb{C}$, we must have $x_\alpha = 0$ except for at most a countable number of terms (Problem 8.16(A)).

If we think of counting measure on a set A, that is, for $x \in A$, the singleton set $\{x\}$ has measure 1, then the unordered sum of the values taken by a nonnegative, real-valued function f on A corresponds to the integral of f with respect to counting measure. Recall that the Hilbert space ℓ^2 consists of scalar sequences $\langle a_n : n \in \mathbb{N}\rangle$ for which $\sum_{n\in\mathbb{N}} |a_n|^2 < +\infty$. We will write $\ell^2(\mathbb{N})$ to indicate that the index set is $\mathbb{N}$. A larger class of Hilbert spaces is obtained using unordered sums on a set A. Here, the measure space is A with counting measure. All subsets of A are measurable, and the only null set is the empty set. This is an example of an L^2-space, but we will write $\ell^2(A)$. That is, a scalar function φ on A is in $\ell^2(A)$ if the unordered sum $\sum_{\alpha\in A}|\varphi(\alpha)|^2$ exists in $\mathbb{R}$. The following is an application of the fact that an L^2-space is a Hilbert space; it is stated for the scalar field $\mathbb{C}$.

Theorem 8.7.1. *The set $\ell^2(A)$ is a vector space over $\mathbb{C}$. If φ and ψ are any two elements of $\ell^2(A)$, then the unordered sum $(\varphi,\psi)=\sum_{\alpha\in A}\varphi(\alpha)\overline{\psi(\alpha)}$ exists in $\mathbb{C}$, and the CBS inequality holds. That is,*

$$|(\varphi,\psi)|\leq\|\varphi\|_2\cdot\|\psi\|_2.$$

It follows that $(\cdot,\cdot)$ is an inner product, $\|\cdot\|_2=\sqrt{\sum_{\alpha\in A}|\varphi(\alpha)|^2}$ is a norm, and $\ell^2(A)$ is a Hilbert space.

8.8 Fourier Series

Recall that a map T from one linear space to another is linear if for all points x and y in the domain and all scalars α and β, $T(\alpha x+\beta y)=\alpha T(x)+\beta T(y)$. We now establish the existence of a linear map between any Hilbert space H and an appropriate ℓ^2 space via the Fourier coefficient map. If H is separable, the corresponding ℓ^2 space is either finite-dimensional or $\ell^2(\mathbb{N})$. We give separate consideration to the separable and the non-separable case.

Theorem 8.8.1. *Let H be a Hilbert space. Let $\{u_\alpha:\alpha\in A\}$ be an orthonormal family in H, and let M be the closure of its span. Then M is a Hilbert subspace of H. The map $\hat{}:H\mapsto\ell^2(A)$ given by $x\mapsto\hat{x}$, where $\hat{x}(\alpha)$ is the α*th *Fourier coefficient, satisfies* ***Bessel's inequality****:*

$$\|\hat{x}\|_2^2=\sum_{\alpha\in A}|\hat{x}(\alpha)|^2\leq\|x\|^2.$$

The map $x\mapsto\hat{x}$ is linear and continuous. Its restriction to points $x\in M$ is a bijection, i.e., a one-to-one map of M onto $\ell^2(A)$. It is, in fact, an ***isometry****; that is, for each $x\in M$, $\|\hat{x}\|_2=\|x\|$. As already noted, $M=H$ if and only if $\{u_\alpha:\alpha\in A\}$ is a maximal orthonormal family in H.*

Proof. The span, and therefore, the closure of the span are stable under addition and scalar multiplication. A closed subspace of a complete space is complete, since a Cauchy sequence has a limit in the complete space, and that limit must be in the closed subspace. Therefore, the restriction of the inner product to points of M makes M a Hilbert subspace of H. By Theorem 8.5.2, $M=H$ if and only if $\{u_\alpha:\alpha\in A\}$ is a maximal orthonormal family in H.

For each $\alpha\in A$, the map $x\mapsto(x,u_\alpha)=\hat{x}(\alpha)$ is linear, so the map $x\mapsto\hat{x}$ is linear. We have seen in Proposition 8.5.3 that for any finite subset $F\subseteq A$ and any $x\in H$, $\sum_{\alpha\in F}|\hat{x}(\alpha)|^2\leq\|x\|^2$. This shows that the set of Fourier coefficients of any $x\in H$

is in $\ell^2(A)$, and that Bessel's inequality holds. It follows from Bessel's inequality that the linear map $x \mapsto \hat{x}$ is continuous since $\|\hat{x}-\hat{y}\|_2 \leq \|x-y\|$. We need to show that for each $y \in \ell^2(A)$, there is a unique $x \in M$ such that $\hat{x} = y$, and moreover, $\|x\| = \|\hat{x}\|_2$.

To show the uniqueness, suppose $\hat{x} = \hat{z}$. Let $w = x-z$. By linearity $\hat{w} = \hat{x}-\hat{z} = 0$. It follows that for each $\alpha \in A$, $(w,u_\alpha) = 0$, whence $w \perp \operatorname{span}(\{u_\alpha : \alpha \in A\})$. By continuity, $w \in M^\perp$. If x and z are both in M, then $w \in M \cap M^\perp$, whence $w = x-z = 0$ in H.

Assume that H is separable. By Theorem 8.6.1, A is either finite or countably infinite. We work with the case that $A = \mathbb{N}$, leaving the finite case as an exercise (8.19). Fix $y \in \ell^2(\mathbb{N})$. That is, $y = \{a_i : i \in \mathbb{N}\}$, and $\sum_{i=1}^\infty |a_i|^2 < +\infty$. The sequence of partial sums $\sum_{i=1}^n a_i u_i$ is a Cauchy sequence in H since for $m < n$,

$$\left\| \sum_{i=1}^n a_i u_i - \sum_{i=1}^m a_i u_i \right\|^2 = \sum_{i=m+1}^n |a_i|^2.$$

Since M is closed, that Cauchy sequence converges to a point $x_0 \in M$. Since the norm and the map $\hat{}$ are continuous,

$$\|x_0\|^2 = \lim_n \left\| \sum_{i=1}^n a_i u_i \right\|^2 = \lim_n \sum_{i=1}^n |a_i|^2 = \|y\|_2^2,$$

and $a_i = \hat{x}_0(i)$ for each $i \in \mathbb{N}$. Thus, for each $y \in \ell^2(\mathbb{N})$, there is an $x \in M$ such that $\hat{x} = y$ and $\|x\| = \|y\|_2$. It follows that the Fourier coefficient map $x \mapsto \hat{x}$ maps M onto $\ell^2(\mathbb{N})$, and the map is an isometry when restricted to M.

Now suppose H is not separable. Fix $y \in \ell^2(A)$. That is, $y = \{b_\alpha : \alpha \in A\}$, and the unordered sum $\sum_{\alpha \in A} |b_\alpha|^2$ exists in $\mathbb{R}$. Let B be the finite or countably infinite set of α's in A for which $b_\alpha \neq 0$. Let $y_B = \{b_\alpha : \alpha \in B\}$. We work with the case that B is infinite, leaving the finite case as an exercise (8.19). Let M_B be the closure of the span of $\{u_\alpha : \alpha \in B\}$. Order the set of u_α's, $\alpha \in B$, to form the enumerated orthonormal family $\{u_i : i \in \mathbb{N}\}$, and for each $i \in \mathbb{N}$, let b_i be the corresponding element of y. As is true for the separable case, $x = \lim_n \sum_{i=1}^n b_i u_i$ exists and is in $M_B \subseteq M$. Moreover, $\hat{x}(i) = y_B(i)$ for each $i \in \mathbb{N}$, and $\|x\| = \|y_B\|_2$. For each $\alpha \in A \setminus B$, $\hat{x}(\alpha) = (x, u_\alpha) = 0$. Therefore, $\hat{x} = y$. Also,

$$\|\hat{x}\|_2 = \|y_B\|_2 = \|y\|_2 = \|x\|.$$

It follows that for each $y \in \ell^2(A)$, there is an $x \in M$ such that $\hat{x} = y$, and $\|x\| = \|\hat{x}\|_2$.

Corollary 8.8.1. *The equality* $\|\hat{x}\|_2 = \sqrt{\sum_{\alpha \in A} |\hat{x}(\alpha)|^2} = \sqrt{(x,x)} = \|x\|$ *holds for every* $x \in H$ *if and only if* $\{u_\alpha : \alpha \in A\}$ *is a maximal orthonormal family in* H.

Proof. Exercise 8.23.

Corollary 8.8.2 (Parseval's Equality). *The equality* $(\widehat{x},\widehat{y}) = \sum_{\alpha\in A}\hat{x}(\alpha)\overline{\hat{y}(\alpha)} = (x,y)$ *holds for every* x *and* y *in* H *if and only if* $\{u_\alpha : \alpha \in A\}$ *is a maximal orthonormal family in* H.

Proof. If there is a u_0 of norm 1 in $M^\perp$, then $\widehat{u}_0 = 0$, so $(\widehat{u}_0,\widehat{u}_0) = 0$, while $(u_0,u_0) = 1$. If $M = H$, then the corresponding inner products in $\ell^2(A)$ and in H are equal because they can be written in terms of the norm using the following identities, called the **polarization identities**: For the real case,

$$\begin{aligned}(x,y) &= \frac{1}{4}\|x+y\|^2 - \frac{1}{4}\|x-y\|^2 \\ &= \frac{1}{4}[(x,x)+2(x,y)+(y,y)] - \frac{1}{4}[(x,x)-2(x,y)+(y,y)].\end{aligned}$$

For the complex case,

$$(x,y) = \frac{1}{4}\left[\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\right].$$

8.9 Trigonometric Series

In this section, we show that the family $\{1,\cos(nt),\sin(nt) : n\in\mathbb{N}\}$ is, after a suitable normalization, a complete orthonormal system for $L^2([0,2\pi])$. The orthogonality is obtained with the usual integrals using trigonometric identities. For each $n>0$, the normalization is given by multiplying each sine and cosine function by $1/\sqrt{\pi}$ since $\int_0^{2\pi}\cos^2(nt)\,dt = \int_0^{2\pi}\sin^2(nt)\,dt = \pi$. Since $\int_0^{2\pi}1^2\,dt = 2\pi$, the normalization for 1 is given by $1/\sqrt{2\pi}$. If we like, we can integrate from $-\pi$ to π instead of from 0 to 2π, since we are working with period 2π.

Given an orthonormal family in L^2 and $f\in L^2$, we associate f with its expansion; that is, $f \sim \sum\hat{f}(\alpha)u_\alpha$, where $\hat{f}(\alpha) = (f,u_\alpha)$. Here, we have the expansion

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty}[a_n\cos(nx)+b_n\sin(nx)].$$

Usually, one drops the normalization of the cosine and sine terms in writing the Fourier series, and one writes the constant term of the Fourier series as $a_0/2$. The normalization is put back when finding the Fourier coefficients. That is, to find the coefficient of the kth cosine term a_k, we think of f as being given by its Fourier series, and for $k\geq 1$, we evaluate the integral

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)\,dt &= \frac{1}{\pi}\int_{-\pi}^{\pi}\left(\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos(nt)+b_n\sin(nt)\right)\cdot\cos(kt)\,dt \\ &= \int_{-\pi}^{\pi} a_k\cdot\frac{\cos(kt)}{\sqrt{\pi}}\cdot\frac{\cos(kt)}{\sqrt{\pi}}\,dt = a_k.\end{aligned}$$

Similarly,

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi}\left(\frac{a_0}{2}+\sum_{n=1}^{\infty} a_n\cos(nt)+b_n\sin(nt)\right)\cdot\sin(kt)\,dt$$
$$= \int_{-\pi}^{\pi} b_k\cdot\frac{\sin(kt)}{\sqrt{\pi}}\cdot\frac{\sin(kt)}{\sqrt{\pi}}\,dt = b_k.$$

To find the value of a_0, we evaluate

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi}\left(\frac{a_0}{2}+\sum_{n=1}^{\infty} a_n\cos(nt)+b_n\sin(nt)\right)dt$$
$$= \int_{-\pi}^{\pi}\left(a_0\cdot\frac{1}{\sqrt{2\pi}}\right)\cdot\frac{1}{\sqrt{2\pi}}\,dt = a_0.$$

The maximality (also known as, completeness) of the trigonometric system for the interval $[0,2\pi]$ follows first from the fact that the space $C(\mathbb{T})$, consisting of the continuous functions taking the same value at 0 and 2π, is dense in $L^2([0,2\pi])$. We think of associating 0 and 2π so that they represent a single point. The domain is then a circle of circumference 2π (hence, the $\mathbb{T}$ for torus). One needs the fact that finite linear combinations of the trigonometric functions are dense in the space $C(\mathbb{T})$, and are, therefore, dense in $L^2([0,2\pi])$. The denseness in the space $C(\mathbb{T})$ is a consequence of the Stone-Weierstrass Theorem, which is discussed in Example 9.11.1.

8.10 Problems

Problem 8.1. Show that it follows from the CBS inequality that the inner product on an inner product space H is a continuous function on $H \times H$.

Problem 8.2. Verify Example 8.3.1.

Problem 8.3. Let $H = \mathbb{R}^2$ with the inner product be equal to the usual dot-product of vectors. Let M be the line of slope $a > 0$. That is, $M := \{(t, at) : t \in \mathbb{R}\}$.

a) Find the closest point $P(x,0)$ on the line M to the point $(x,0)$ on the real axis.
b) Show that the vector $P(x,0)-(x,0)$ is orthogonal to M.

Problem 8.4. Let H be a Hilbert space.

a) Show that if $A \subseteq B \subseteq H$, then $B^{\perp} \subseteq A^{\perp}$.
b) Show that $A^{\perp} = \left(\overline{A}\right)^{\perp}$.

Problem 8.5. Let M be a closed Hilbert subspace of a Hilbert space H. Show that $\left(M^{\perp}\right)^{\perp} = M$.

Problem 8.6. Prove Corollary 8.3.3.

Problem 8.7. Let H be a Hilbert space, and let $A \subseteq X$. If A contains an open ball $B(a,r)$ for some $a \in A$, $r > 0$, show that $A^{\perp} = \{0\}$. **Hint:** Suppose there is a point $y \in A^{\perp}$ with $\|y\| = r/2$. What can you say about $a + y$?

Problem 8.8. Show that the parallelogram law fails for the space $C([-1,1])$ consisting of the continuous real-valued functions on $[-1,1]$ and equipped with the norm $\|f\| = \max_{x \in [-1,1]} |f(x)|$.

Problem 8.9. Show that $\ell^p(\mathbb{N})$ is a Hilbert space if and only if $p = 2$. **Hint:** Check the parallelogram law.

Problem 8.10. Let λ be Lebesgue measure on the real line $\mathbb{R}$. Suppose E is a measurable subset of $\mathbb{R}$ and ν is defined on any measurable subset A by setting $\nu(A) = \lambda(A \cap E)$. Clearly, ν is absolutely continuous with respect to λ, i.e., $\nu << \lambda$. Describe the Radon-Nikodým derivative $d\nu/d\lambda$.

Problem 8.11. Let $X = \{1, 2, \ldots, n\}$, and let μ be counting measure on X. That is, $\mu(i) = 1$ for each $i \in X$. Given any measure ν on the subsets of X, calculate $\frac{d\nu}{d\mu}$.

Problem 8.12. Show that if μ and ν are σ-finite measures on a measurable space $(X, \mathscr{B})$, then X is a countable union of measurable sets on which both μ and ν are finite measures.

Problem 8.13. **a)** Let μ and ν be σ-finite measures on a measurable space $(X, \mathscr{B})$ with $\nu << \mu$. Show that if $g \in L^1(\nu)$, then $\int_X g\, d\nu = \int_X g \frac{d\nu}{d\mu}\, d\mu$.
b) Assume in addition that γ is a σ-finite measure on $(X, \mathscr{B})$ such that $\mu << \gamma$. Show that $\frac{d\nu}{d\gamma} = \frac{d\nu}{d\mu}\frac{d\mu}{d\gamma}$, γ-a.e.

Problem 8.14. **a)** Show that the span of a subset S of a Hilbert space H is a linear space. That is, S is stable with respect to addition and scalar multiplication.
b) Show that the span of S is the smallest linear subspace of H containing S.
c) Show that the closure of the span of a subset S of a Hilbert space H is a linear space.

Problem 8.15. Show the following: If a series in the scalar field is absolutely convergent, then the sum of the terms is an unordered sum. Conversely, if a scalar series has an unordered sum, then the series is absolutely convergent.

Problem 8.16 (A). Show that in dealing with a finite unordered sum of terms x_α in any normed space X, including $\mathbb{R}$ or $\mathbb{C}$, we must have $x_\alpha = 0$ except for at most a countable number of terms.

Problem 8.17. Let $\{u_1, u_2, u_3\}$ be a three-element orthonormal family in a Hilbert space H. Let M be the span of the set $\{u_1, u_2, u_3\}$. Given $h \in H$, let $a_1 = (h, u_1)$, $a_2 = (h, u_2)$, and $a_3 = (h, u_3)$.

a) Show that $h - (a_1u_1 + a_2u_2 + a_3u_3)$ is in $M^{\perp}$.
b) Describe in terms of the elements u_1, u_2, and u_3, the nearest element to h in M and the nearest element to h in $M^{\perp}$.

Problem 8.18. Show that the span of a finite orthonormal family in a real or complex inner product space is complete.

Problem 8.19 (A).

a) Prove Theorem 8.8.1 for the case that H is separable and A is a finite set.
b) Prove Theorem 8.8.1 for the case that H is not separable and B is a finite set.

Problem 8.20. Let H be a Hilbert space, and let $M \subset H$ be a closed linear subspace such that $M \neq H$ and $M \neq \{0\}$. Show that there is a maximal orthonormal basis for H consisting only of elements of M and of $M^{\perp}$.

Problem 8.21. Let H be a separable Hilbert space, and let V be a dense subspace of H. Show that V contains a maximal orthonormal family for H.

Problem 8.22. Show that a linear isometry from one normed space to another is an injection, that is, a one-to-one map.

Problem 8.23. Prove Corollary 8.8.1.

Problem 8.24. Use the fact that the family

$$\{\frac{1}{\sqrt{2\pi}}\} \cup \{\frac{1}{\sqrt{\pi}} \cos nx : n \in \mathbb{N}\} \cup \{\frac{1}{\sqrt{\pi}} \sin nx : n \in \mathbb{N}\}$$

is a maximal orthonormal family in $L^2[0,2\pi]$ with respect to Lebesgue measure to show that for any $f \in L^2[0,2\pi]$, $\lim_{n\to\infty} \int_0^{2\pi} \cos(nx) f(x) dx = 0$.

Problem 8.25. Show that for a Hilbert space with complex scalars, the inner product is given by the **polarization identity**.

$$(x,y) = \frac{1}{4}\left[\|x+y\|^2 - \|x-y\|^2 + i\,\|x+iy\|^2 - i\,\|x-iy\|^2\right].$$

Chapter 9
Topological Spaces

9.1 Neighborhoods

In this chapter, we generalize the notion of closeness given by a metric or norm. Recall that if we start with a metric space (X,d), then open balls play a central role in obtaining results. An open ball with center x and radius $r > 0$ is denoted by $B(x,r)$; it is the set $\{y \in X : d(x,y) < r\}$. In $\mathbb{R}$, $B(x,r) = (x-r, x+r)$. A set O is called open in a metric space if for each $x \in O$ there is an $r > 0$ such that the open ball $B(x,r)$ is contained in O. If $z \in B(x,r)$, then there is an open ball $B(z,\delta) \subseteq B(x,r)$. Just let $\delta = r - d(x,z)$. This property makes an open ball such as $B(x,r)$ an open set. A subset of a metric space is again a metric space when the metric is restricted to pairs of points in the subset. In a space X with a norm, such as a Hilbert space, the corresponding metric is given by $d(x,y) = \|x-y\|$. Note that for any $z \in X$, $B(0,r) + z = B(z,r)$. Also, for any $z \in X$, $\|z\| = \|z-0\| = d(z,0)$.

If x and y are points in a metric space, then the more open balls centered at x that contain y, the closer y is to x. Given two open balls centered at x, the intersection contains, and in fact equals, an open ball centered at x. We now generalize these properties so that we can deal with topologies in essentially the same way that we deal with metrics. First we note that given a point x in a metric space, the collection $\mathscr{F}_x$ of open sets containing x has the properties of what is called a filter base. That is, $\mathscr{F}_x$ is a nonempty collection of nonempty sets each containing x, and the intersection of any two members of $\mathscr{F}_x$ contains another member of $\mathscr{F}_x$. Here is what we will use.

Definition 9.1.1. Fix a nonempty set X. A **local filter base at a point** $x \in X$ is a nonempty collection $\mathscr{B}_x$ of subsets of X, each containing x, such that

$$\forall U, V \in \mathscr{B}_x,\ \exists W \in \mathscr{B}_x \text{ such that } x \in W \subseteq U \cap V.$$

Example 9.1.1. An example of a local filter base not given by a metric is one for pointwise convergence of real-valued functions on $[0,1]$. Here, each point is actually a function f, and an element of $\mathscr{B}_f$ specifies a finite set $\{r_1, \cdots, r_n\}$ in the interval

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_9

$[0,1]$ and an $\varepsilon > 0$. A function g is in the member of $\mathscr{B}_f$ given by these parameters if for $1 \le i \le n$, $|g(r_i) - f(r_i)| < \varepsilon$. To see that the necessary condition for a local filter base is met, simply take two such sets for a given f, take the union of the two sets of points in $[0,1]$ and the smaller of the two ε's. This gives a member of $\mathscr{B}_f$ contained in the two initial ones.

Definition 9.1.2. Assume we are given a local filter base $\mathscr{B}_x$, at each point x in a set X. A set $O \subseteq X$ is called **open** if for each $y \in O$ there is a $U \in \mathscr{B}_y$ with $y \in U \subseteq O$. A collection $\mathscr{T}$ of subsets of a nonempty set X is called a **topology**, and $(X, \mathscr{T})$ is called a **topological space**, if $\mathscr{T}$ contains X as well as the empty set, and $\mathscr{T}$ is stable with respect to the operations of taking arbitrary unions and finite intersections.

We will usually assume that the members of a local filter base are themselves open sets. Even without this simplifying assumption, we have the following result, used, for example, to construct a topology in [21].

Theorem 9.1.1. *Given an assignment of a local filter base $\mathscr{B}_x$ at each point x of a nonempty set X, the collection $\mathscr{T}$ of open subsets of X is a topology on X.*

Proof. Since there are no points in the empty set, the condition for the empty set to be open is vacuously satisfied. Since for each $x \in X$, $\mathscr{B}_x \neq \emptyset$, the set X itself is open. If $\{U_\alpha\}$ is a collection of open sets and the union contains x, then for some index α_0, $x \in U_{\alpha_0}$, and so for some $W \in \mathscr{B}_x$, $x \in W \subseteq U_{\alpha_0} \subseteq \bigcup_\alpha U_\alpha$. Therefore $\bigcup_\alpha U_\alpha$ is open. If $\{U_1, \ldots, U_n\}$ is a finite collection of open sets, each of which contains x, then for $1 \le i \le n$, there is a set $W_i \in \mathscr{B}_x$ with $x \in W_i \subseteq U_i$. By the properties of a local filter base, there is a set $W_0 \in \mathscr{B}_x$ with $x \in W_0 \subseteq \bigcap_{i=1}^n W_i \subseteq \bigcap_{i=1}^n U_i$. Therefore $\bigcap_{i=1}^n U_i$ is open.

Definition 9.1.3. If for each x in a set X, $x \mapsto \mathscr{B}_x$ is an assignment of a local filter base consisting of open sets, then we say that we are given an **open base** at each $x \in X$.

We have noted that for a metric space, the open balls centered at a point form an open base at the point. It is also easy to see that the base described for pointwise convergence in Example 9.1.1 consists of open sets. In much of the literature, one starts with a topology $\mathscr{T}$, calling the members "open sets"; only later is an open base at a point of the space defined. If one starts with a topology $\mathscr{T}$, the mapping $x \mapsto \{O \in \mathscr{T} : x \in O\}$ gives the largest possible open base at each x that yields $\mathscr{T}$ as the collection of open sets. On the other hand, if one is given an open base $\mathscr{B}_x$ at each point of a set X and $\mathscr{U} := \{U : \exists\, x \in X, U \in \mathscr{B}_x\}$, then the topology consists of all possible unions of sets from $\mathscr{U}$.

An open set containing a point x is often called an **open neighborhood** (or just a neighborhood) of x. In general, many choices of open bases give rise to the same topology. For the Euclidean plane, for example, open disks centered at points of the plane generate the topology of open subsets of the plane. The same topology, however, is generated by the insides of squares centered at points of the plane, and also by open balls of radius $1/n$, $n \in N$, centered at points of the plane. For this reason, the overall collection of open sets is the first thing that is usually defined when considering a topology. On the other hand, just as with open balls in a metric space, it is the generating open neighborhoods of points that one considers for the most part when actually working with a topology.

The emphasis on open neighborhoods will allow us to develop the notions of topology in parallel with the notions of metric spaces. Indeed, if we are given a **metric** space, we will usually assume that the open base at a point is the set of open balls with that point as center. Even when a metric is available, however, it is sometimes better to work with a neighborhood system not directly given by the metric. For example, the topology of uniform convergence on compact sets is of fundamental importance in complex function theory. A metric is available, but it is more natural to think of a neighborhood of a continuous function f as given by a compact subset $K \subseteq \mathbb{C}$ and an $\varepsilon > 0$; the neighborhood consists of appropriate continuous functions g such that $\max_{z\in K} |f(z) - g(z)| < \varepsilon$. (See Problem 9.1.)

Finally, we note some additional structure that can be added to a topological space to obtain results not available for generic topological spaces.

Definition 9.1.4. A subset S of a topological space $(X, \mathscr{T})$ is **dense** in X if every $O \in \mathscr{T}$ contains a point of S. That is, $\overline{S} = X$. A topological space that contains a countable dense subset is called **separable**.

Example 9.1.2. The rational numbers are dense in $\mathbb{R}$, so $\mathbb{R}$ is separable. The points with rational coordinates are dense in $\mathbb{R}^n$, so $\mathbb{R}^n$ is separable.

Definition 9.1.5. Given a topological space $(X, \mathscr{T})$, a collection $\mathscr{B} \subseteq \mathscr{T}$ is called a **base** for $\mathscr{T}$ if for each open O and each $x \in O$ there is a $U \in \mathscr{B}$ with $x \in U \subseteq O$. That is, every open set O is the union of the sets from $\mathscr{B}$ contained in O.

Remark 9.1.1. If $\mathscr{B}$ is a base for a topology $\mathscr{T}$, then $x \mapsto \{U \in \mathscr{B} : x \in U\}$ is an open base at x. On the other hand, given an open base $\mathscr{B}_x$ at each point of X, $\cup_{x\in X} \mathscr{B}_x$ is also a base for the topology. Usually, a base $\mathscr{B}$ is a smaller collection of open sets.

Example 9.1.3. For a separable metric space such as $\mathbb{R}^n$, it follows from Problem 9.3 that open balls of radius $1/n$, $n \in \mathbb{N}$, centered at points of a dense subset of X form a base for the metric topology.

Definition 9.1.6. A topological space with a countable open base $\mathscr{B}_x$ at each point is said to satisfy the **first axiom of countability**. One also says the space is "first countable." A topological space with a countable base $\mathscr{B}$ for the topology is said to satisfy the **second axiom of countability**. One also says the space is "second countable."

We will show that sequences suffice for convergence to closure points when a space satisfies the first axiom of countability. Also, a space that satisfies the second axiom of countability satisfies the first axiom, since the sets in the base containing x form a base at x.

Example 9.1.4. A metric space (X, d) satisfies the first axiom of countability since open balls of radius $1/n$, $n \in \mathbb{N}$, centered $x \in X$ form an open base at x.

Theorem 9.1.2. *A topological space $(X,\mathscr{T})$ that satisfies the second axiom of countability is separable. If $(X,\mathscr{T})$ is separable and satisfies the first axiom of countability, then it satisfies the second axiom of countability.*

Proof. Given a countable base for the topology, use the Axiom of Choice to select one point from each base set. Given a countable dense set $S \subseteq X$ and a countable open base $\mathscr{B}_x$ at each $x \in S$, the union $\cup_{x\in S}\mathscr{B}_x$ forms a countable base for the topology.

Remark 9.1.2. For the most part, we work with open bases at points of a topological space. Recall, however, that if we are given a local filter base at each point of a set X, then a set E is open in the corresponding topology if and only if for each $x \in E$, there is a set S in the local filter base at x such that $x \in S \subseteq E$. This raises the question as to whether such a set must at least contain an open subset. The answer is "not always." For example, the set $\{a,b,c\}$ with the trivial topology is generated by the local filter base that associates the set $\{a,b\}$ with a and associates the set $\{a,b,c\}$ with both b and c. Here is a partial positive answer for a Hausdorff space; that is, a space for which distinct points have disjoint open neighborhoods.

Proposition 9.1.1. *Let $(X,\mathscr{T})$ be a Hausdorff space. Let p be a point in X for which there is a countable open base at p, whence there is a sequence $\langle O_n : n \in \mathbb{N}\rangle$ of open sets with $p \in O_{n+1} \subseteq O_n$ for each n, and $\cap_n O_n = \{p\}$. Let $\mathscr{L}_p$ be a local filter base for p. Then each $S \in \mathscr{L}_p$ contains an open set.*

Proof. For each $n \in \mathbb{N}$, there is an $S_n \in \mathscr{L}_p$ with $p \in S_n \subseteq O_n$. If some $S_0 \in \mathscr{L}_p$ contains no open set, then we may assume that each S_n is contained in S_0. Now, for each $n \in \mathbb{N}$, there is a point $x_n \in O_n \setminus S_0$. Let $A = \{x_n : n \in \mathbb{N}\}$. Since $(X,\mathscr{T})$ is Hausdorff, for each point $x \neq p$ in X, there is an $m \in \mathbb{N}$ and an open set V that contains x such that $O_m \cap V = \varnothing$. Using the fact that X is Hausdorff again, there is an open neighborhood U of x contained in $X \setminus A$. Since $p \in S_0 \subseteq X \setminus A$, $X \setminus A$ is an open set. On the other hand, A is not closed since p is a point of closure of A. This contradiction establishes the result.

9.2 Metric and Topological Notions

We now establish some properties of both metric and topological spaces. Throughout this section, we work with a nonempty set X and an assignment of an open base $\mathscr{B}_x$ at each $x \in X$. When the space is a metric space, which we assume is nonempty, the open base at x consists of open balls centered at x. For this reason, our results will recapitulate results for metric spaces.

Definition 9.2.1. Given $A \subseteq X$, and $x \in X$, we say that x is a **point of closure** or a **closure point** of A if for each $U \in \mathscr{B}_x$, $U \cap A \neq \varnothing$. We write $\overline{A}$ for the set of points of closure of A. A set A is called **closed** if $A = \overline{A}$. Moreover, $\overline{A}$ is called the **closure** of A.

Proposition 9.2.1. *Fix $A, B \subseteq X$. Clearly, $A \subseteq \overline{A}$. If $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$. In general, $\overline{A \cup B} = \overline{A} \cup \overline{B}$, and $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$.*

Proof. If $C \subseteq D \subseteq X$, then a closure point of C is a closure point of D. It follows that, $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$, and $\overline{A \cup B} \supseteq \overline{A} \cup \overline{B}$. If x is in $\overline{A \cup B}$ and x is not a point of closure of A, then some neighborhood of x does not intersect A. Every neighborhood in $\mathscr{B}_x$ must then intersect B, so $x \in \overline{B}$. It follows that $\overline{A \cup B} = \overline{A} \cup \overline{B}$.

Example 9.2.1. On $\mathbb{R}$, if $A = (0,1)$ and $B = (1,2)$, then $\overline{A \cap B} = \varnothing$, but $\overline{A} \cap \overline{B} = \{1\}$.

Proposition 9.2.2. *The closure of the closure is the closure; i.e., $\overline{(\overline{A})} = \overline{A}$.*

Proof. Let x be a point of closure of $\overline{A}$, and fix an arbitrary $U \in \mathscr{B}_x$. By definition, there is a $z \in U \cap \overline{A}$. Since U is open, there is a $V \in \mathscr{B}_z$ with $z \in V \subseteq U$. Since $z \in \overline{A}$, there is a $y \in V \cap A \subseteq U \cap A$, whence $x \in \overline{A}$.

Proposition 9.2.3. *The closure of a set $A \subseteq X$ is the intersection of all closed sets containing A.*

Proof. Exercise 9.4.

Proposition 9.2.4. *The closure of a dense subset $S \subseteq X$ is all of X.*

Proof. Exercise 9.5.

Theorem 9.2.1. *A set A is closed if and only if its complement $X \setminus A$ is open.*

Proof. Exercise 9.6.

Proposition 9.2.5. *The set X and the empty set $\varnothing$ are closed. Moreover, finite unions and arbitrary intersections of closed sets are closed.*

Proof. Use De Morgan's law.

Remark 9.2.1. We have already noted in Remark 7.1.4 that a closed ball in a metric space is a closed set, and for some spaces, the closure of an open ball may not be the closed ball with the same radius.

Definition 9.2.2. The **interior** of a set $E \subseteq X$ is the set of all x for which there is a $U \in \mathscr{B}_x$ with $x \in U \subseteq E$. We write E° for the interior.

Proposition 9.2.6. *The interior of a set E is the union of all open sets contained in E.*

Proof. Exercise 9.7.

Definition 9.2.3. If A is a subset of a metric space, then A is bounded if A is contained in an open ball about some point of the metric space.

9.3 Continuous Mappings

In this section, we assume that X and Y are two sets; we let $x \mapsto \mathscr{B}_x$ be an open base at each point of X while $y \mapsto \mathscr{D}_y$ is an open base at each point of Y.

Definition 9.3.1. A function f from X into Y is continuous at $x \in X$ if for each $V \in \mathscr{D}_{f(x)}$, there is a $U \in \mathscr{B}_x$ with $f[U] \subseteq V$. We say that f is continuous or continuous on X if it is continuous at each point $x \in X$.

Theorem 9.3.1. *A function f mapping X into Y is continuous on X if and only if for each open set O contained in Y, $f^{-1}[O]$ is open in X.*

Proof. Assume that for every open $O \subseteq Y$, $f^{-1}[O]$ is open. Given $x \in X$ and $V \in \mathscr{D}_{f(x)}$, $f^{-1}[V]$ is open and contains x. Therefore, there is a $U \in \mathscr{B}_x$ with $x \in U \subseteq f^{-1}[V]$, whence $f[U] \subseteq V$. It follows that f is continuous on X. Conversely, if f is continuous on X and O is open in Y, then given $x \in f^{-1}[O]$, set $y = f(x) \in O$. Now there is a $V \in \mathscr{D}_y$ with $y \in V \subseteq O$, and there is a $U \in \mathscr{B}_x$ with $f[U] \subseteq V \subseteq O$. That is, $x \in U \subseteq f^{-1}[O]$. It follows that $f^{-1}[O]$ is open in X.

Proposition 9.3.1. *Assume now that Z is a third set with open base $\mathscr{F}_z$ at each point z of Z. Fix $x \in X$. If $f : X \mapsto Y$ is continuous at x, and $g : Y \mapsto Z$ is continuous at $f(x)$, then $g \circ f : X \mapsto Z$ is continuous at x. If this is true at each point of x, then the composition function is continuous. That is, a continuous function of a continuous function is a continuous function.*

Proof. For the second part, if O is open in Z, then $g^{-1}[O]$ is open in Y, whence $(g \circ f)^{-1}[O] = f^{-1}[g^{-1}[O]]$ is open in X.

Definition 9.3.2. A bijection f mapping X onto Y is called a **homeomorphism** between X and Y if f and f^{-1} are continuous. If such a mapping exists, we say that X and Y are **homeomorphic**.

Properties invariant under homeomorphic mappings are often called **topological properties**. If (X,ρ) and (Y,σ) are metric spaces and f is a mapping from X onto Y such that for every $x, z \in X$, $\sigma(f(x), f(z)) = \rho(x,z)$, we say that f is an **isometry** between X and Y. If such a mapping exists, we say that X and Y are **isometric** spaces. An example of a homeomorphism that is not an isometry is given by the identity mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ where the domain is supplied with the Euclidean metric ρ and the range has the distance function $\min(\rho, 1)$.

9.4 Sequences and Nets

Example 9.4.1. We want to use ordinal numbers as an example in this section. We summarize some results from the appendix on the Axiom of Choice. It follows from that axiom that any set can be well-ordered. Take an uncountable set and impose a well-ordering on it. Every nonempty subset of the set will have a first element with respect to the ordering. There is, therefore, a first element ω such that the set of its predecessors, corresponding to the elements of $\mathbb{N}$, is an infinite set. There is a first element Ω such that the set of its predecessors form an uncountable set. There is a bijection between any two well-ordered sets terminating in an element that is the first element for which the set of predecessors is an uncountable set. That terminating element is called the **first uncountable ordinal** Ω. If S is such a set, then any element smaller than Ω has an immediate successor but not necessarily an immediate predecessor. Since each $\alpha < \Omega$ has an immediate successor ν, $\gamma \leq \alpha$ if and only if $\gamma < \nu$. An open base at each $\alpha \leq \Omega$ consists of the sets $U_\beta = \{\gamma \in S : \beta < \gamma \leq \alpha\}$ for each $\beta < \alpha$ in S. The corresponding topology is called the **order topology** on S. The element Ω and its predecessors form the **ordinal numbers up to the first uncountable ordinal**.

Throughout this section, we let $(X, \mathscr{T})$ be a topological space, and for each $x \in X$, we let $\mathscr{B}_x$ be an open base such that the assignment $x \mapsto \mathscr{B}_x$ generates the topology $\mathscr{T}$. Recall that for each $x \in X$, $\mathscr{B}_x$ may consist of all open sets containing x.

Definition 9.4.1. A sequence $\langle x_n; n \in \mathbb{N}\rangle$ in X **converges** to a point $x \in X$ if it is **eventually** in each $U \in \mathscr{B}_x$. That is, for every $U \in \mathscr{B}_x$, there is an $m \in \mathbb{N}$ such that for all $n \geq m$, $x_n \in U$. The point x is a **cluster point** of x_n if the sequence x_n is **frequently** in every $U \in \mathscr{B}_x$. That is, for every $m \in \mathbb{N}$, there is an $n \geq m$ with $x_n \in U$.

A generalization of sequential convergence is given by net convergence. Recall that a **directed set** is a set $\mathscr{D}$ together with a transitive relation $\leq$ such that for any pair a, b in $\mathscr{D}$, there is a c in $\mathscr{D}$ with $a \leq c$ and $b \leq c$. A **net** in X is a mapping from a directed set into X. We write x_a for the image in X of $a \in \mathscr{D}$; we write $\langle x_a : a \in \mathscr{D}\rangle$ for the net.

Definition 9.4.2. A net $\langle x_a : a \in \mathscr{D}\rangle$ in X **converges to a point** $x \in X$ if for each $U \in \mathscr{B}_x$ there is a $c \in \mathscr{D}$ such that for all $a \geq c$ in $\mathscr{D}$, $x_a \in U$. That is, the net is eventually in every open neighborhood of x. A point $x \in X$ is a **cluster point** of the net $\langle x_a : a \in \mathscr{D}\rangle$ if for every $U \in \mathscr{B}_x$ and every $b \in \mathscr{D}$, there is a $c \geq b$ in $\mathscr{D}$ with $x_c \in U$. That is, the net is frequently in every neighborhood of x.

Example 9.4.2. We have already noted in Examples 1.9.3 and 1.9.4 the use of net convergence beyond sequential convergence in calculus. Unordered sums are also an example of net convergence. Our topological space $(X, \mathscr{T})$ provides still another example. That is, fix $x \in X$, and let the directed set $\mathscr{D} := \mathscr{B}_x$. We set $V \geq U$ in $\mathscr{D}$ if $V \subseteq U$. The property that makes $\mathscr{B}_x$ a local filter base makes $\mathscr{D}$ a directed set. Choose $x_U \in U$ for each $U \in \mathscr{B}_x$. Then the net $\langle x_U : U \in \mathscr{B}_x\rangle$ converges to x since it is eventually in every $U \in \mathscr{B}_x$. Note that such a net always converges to x, but if $(X, \mathscr{T})$ does not satisfy the first axiom of countability, there may be no sequence except one eventually identically equal to x that converges to x. An example is given by the set of ordinal numbers up to and including the first uncountable ordinal Ω described in Example 9.4.1 and Problems 9.9 and 9.20. There is no countable base at Ω for the order topology. There is no sequence of ordinals smaller than Ω that converges to Ω even though, using the order topology, Ω is a closure point of the set of its predecessors. On the other hand, the set of ordinals smaller than Ω forms a directed set, and the identity map $\alpha \mapsto \alpha$ provides a net converging to Ω.

Proposition 9.4.1. *A point x is a closure point of $E \subseteq X$ if and only if there is a net in E converging to x. If there is a countable base for the topology at x, then a sequence in E converges to x.*

Proof. If such a net exists, then every neighborhood of x contains a point of E. Conversely, if every neighborhood of x contains a point of E, then Example 9.4.2, with $x_U \in E$ for each $U \in \mathscr{B}_x$, gives a net converging to x.

Definition 9.4.3. A topological space $(X, \mathscr{T})$ is a **Hausdorff space** if distinct points x and y in X have disjoint open neighborhoods $U \in \mathscr{B}_x$ and $V \in \mathscr{B}_y$.

Remark 9.4.1. A metric space with the topology generated by the metric is a Hausdorff space. For these, and for more general spaces, the following is clear.

Proposition 9.4.2. *A net, in particular a sequence, can have at most one limit in a **Hausdorff** space.*

9.5 Subspaces

Given a subset $Y \subset X$, where (X,d) is a metric space, we can restrict the metric to $Y \times Y$. For each $x \in Y$, we then look at open balls in Y; these have the form $\{y \in Y : d(x,y) < r\}$. For spaces without a metric, we can still restrict the topology. Let $(X, \mathscr{T})$ be a topological space, and let $x \mapsto \mathscr{B}_x$ be an assignment of an open base at each $x \in X$. Fix $Y \subset X$. If $y \in Y$, then $\mathscr{B}_y^Y := \{U \cap Y : U \in \mathscr{B}_y\}$ is a local filter base at y since

$$W \subseteq U \cap V \Rightarrow W \cap Y \subseteq (U \cap Y) \cap (V \cap Y).$$

Definition 9.5.1. Sets open in Y with respect to the assignment $y \mapsto \mathscr{B}_y^Y$ are called **relatively open** in Y. The corresponding topology on Y is called the **relative topology** on Y.

Proposition 9.5.1. *For $Y \subseteq X$, a subset $W \subseteq Y$ is open in Y, i.e., relatively open, if and only if there is an open $O \subseteq X$ with $W = O \cap Y$.*

Proof. Assume W is nonempty and relatively open in Y. For each $y \in W$, choose a $U_y \in \mathscr{B}_y$ so that $y \in U_y \cap Y \subseteq W$. The set $O := \cup_{y \in W} U_y$ is the desired open subset of X. Conversely, if $W = O \cap Y$ for some open subset O of X, then for each $y \in W$, $y \in O$, so there is a $U \in \mathscr{B}_y$ with $U \subseteq O$. Since $U \cap Y \in \mathscr{B}_x^Y$, and $U \cap Y \subseteq W$. It follows that W is relatively open in Y.

Corollary 9.5.1. *For each $y \in Y$, $\mathscr{B}_y^Y$ is an open base at y with respect to the relative topology on Y.*

Corollary 9.5.2. *If $Z \subseteq Y \subseteq X$, the relative topology of Z with respect to the relative topology of Y equals the relative topology of Z with respect to the topology for X.*

Corollary 9.5.3. *A set A is relatively closed in Y if and only if there is a closed set F in X with $F \cap Y = A$. Moreover, for a general subset of Y, the relative closure in Y is the intersection of the closure in X with Y.*

Proof. Exercise 9.8.

Proposition 9.5.2. *If $\mathscr{B}$ is a base for the topology $\mathscr{T}$, and $Y \subset X$, then the restriction of the sets in $\mathscr{B}$ to Y is a base for the relative topology on Y.*

Corollary 9.5.4. *If $(X, \mathscr{T})$ is separable and satisfies the first axiom of countability, then every subspace of X is separable.*

Proof. Apply Theorem 9.1.2.

Corollary 9.5.5. *Every subspace of a separable* ***metric*** *space is separable.*

9.6 Connectedness

Connected sets in topological spaces generalize intervals in the real line.

Definition 9.6.1. A subset A of a topological space $(X, \mathscr{T})$ is **connected** if it is not the union of two disjoint, nonempty, relatively open sets. Equivalently, it is connected if it is not the union of two disjoint, nonempty, relatively closed sets. In either case, such a pair of sets is called a **disconnection** of A. If $A = X$, relatively open is just open, and relatively closed is just closed.

Theorem 9.6.1. *Suppose A is a nonempty subset of a topological space $(X, \mathscr{T})$, and f is a continuous function from A with the relative topology into a topological space $(Y, \mathscr{S})$. In particular, f may be the restriction to A of a continuous function on X. Then the inverse image of a disconnection of $f[A]$ is a disconnection of A. It follows that the continuous image of a connected set is connected.*

Proof. Suppose U and V are open sets in Y such that $f[A] \subset U \cup V$ and $U \cap V \cap f[A] = \varnothing$. Then $U \cap f[A]$ and $V \cap f[A]$ are relatively open sets in $f[A]$. It follows that $f^{-1}[U]$ and $f^{-1}[V]$ are relatively open in A, their union is A and their intersection is empty. That is, they form a disconnection of A.

Example 9.6.1. The Intermediate Value Theorem in calculus is an example of the fact that an interval is connected, and so its continuous image is connected.

Theorem 9.6.2. *If $(X, \mathscr{T})$ is a topological space, and $\{A_\gamma : \gamma \in \mathscr{I}\}$ is an indexed family of connected subsets of X such that for any α and β in $\mathscr{I}$, $A_\alpha \cap A_\beta \neq \varnothing$, then the union $\cup_{\gamma \in \mathscr{I}} A_\gamma$ is connected.*

Proof. Exercise 9.10(A).

Theorem 9.6.3. *If A is a connected subset of a topological space $(X, \mathscr{T})$, and the set $A \subset B \subseteq \overline{A}$, then B is connected. In particular, $\overline{A}$ is connected.*

Proof. Let U and V be disjoint, relatively open subsets of B such that $B = U \cup V$. If $U \cap A = \varnothing$, then since any point of $B \setminus A$ is a point of closure of A, $U = \varnothing$. A similar fact holds for V. Since A is connected, it follows from Corollary 9.5.2 that either $U = \varnothing$ or $V = \varnothing$. Therefore, B is connected.

Definition 9.6.2. Let $(X, \mathscr{T})$ be a topological space. A path in X joining $x \in X$ to $y \in X$ is a continuous function h mapping a closed interval $[a,b] \subset \mathbb{R}$ into X with $h(a) = x$ and $h(b) = y$. The space $(X, \mathscr{T})$ is **pathwise connected** if every pair of points x and y in X can be joined by a path in X.

Remark 9.6.1. One can always use a continuous transformation to change the parametrizing interval $[a,b]$ to $[0,1]$. In the literature, a pathwise connected space is also called **arcwise connected**.

Theorem 9.6.4. *Every pathwise connected topological space is connected. In particular, every interval in the real line is connected.*

Proof. First we show that every closed and bounded interval $[a,b] \subset \mathbb{R}$ is connected. Suppose U and V are disjoint open sets in $\mathbb{R}$ with $a \in U$ and $[a,b] \subseteq U \cup V$. Let $c = \sup\{x \in [a,b] : [a,x] \subseteq U\}$. Since U and V are open, $c = b$, and $c \in U$, whence $V \cap [a,b] = \varnothing$. Now suppose that $(X, \mathscr{T})$ is a nonempty, pathwise connected space, and fix $x \in X$. By Theorem 9.6.1, every path in X from x to another point in X is a connected set. Since X is the union of such paths, it follows from Theorem 9.6.2 that X is connected. In particular, any interval in $\mathbb{R}$ is connected.

Example 9.6.2. Let X be the plane $\mathbb{R}^2$ with the Euclidean topology, and let A be the graph of the function $x \mapsto \sin(1/x)$ for $0 < x \leq 1/\pi$. Since A is pathwise connected it is connected. The closure $\overline{A}$ in $\mathbb{R}^2$ adjoins the interval from -1 to 1 on the y-axis. By Proposition 9.6.3, $\overline{A}$ is connected. It is not, however, pathwise connected. See Problem 9.13.

Remark 9.6.2. Here are some further aspects of connectedness that appear in the literature. The **component** of a point is the union of all connected sets containing the point. By Theorems 9.6.2 and 9.6.3, a component is both closed and connected. A **locally connected** space is one having an open base $x \mapsto \mathscr{B}_x$ at each point consisting of connected sets.

9.7 Compact Spaces

We next consider compactness for an arbitrary topological space $(X, \mathscr{T})$. For a history of the notion, see [42].

Definition 9.7.1. An **open covering** of a set A contained in a topological space $(X,\mathscr{T})$ is a collection of open sets in X such that each point of A is in at least one of the open sets. A **finite subcover** of an open cover of A is a finite subcollection of the open cover that again forms a cover of A. A set $A \subseteq X$ is called **compact** if every open cover of A has a finite subcover. A collection of sets has the **finite intersection property** if every **finite** subcollection has a nonempty intersection.

Proposition 9.7.1. *A set $A \subset (X,\mathscr{T})$ is compact in X if and only if A is a compact topological space with respect to the relative topology.*

Proof. This follows from Proposition 9.5.1.

Fix $(X,\mathscr{T})$. A collection of sets fails to cover a set $A \subseteq X$ if there is a point of A in all of the complements. Therefore, by DeMorgan's law, we have an equivalent formulation of compactness using closed sets. For the space $(X,\mathscr{T})$ itself, relatively closed just means closed.

Theorem 9.7.1. *A set $A \subseteq X$ is compact if and only if any collection of open sets having no finite subcover of A does not itself cover A. Equivalently, $A \subseteq X$ is compact if any collection of relatively closed subsets of A with the finite intersection property has a nonempty intersection.*

Recall that a topological space is Hausdorff if any two distinct points have disjoint open neighborhoods.

Theorem 9.7.2. *A closed subset of a compact set is compact. A compact subset of a Hausdorff space is closed. A compact subset of a metric space is bounded.*

Proof. Suppose $A \subset B \subseteq X$, and A is a closed subset of B, while B is a compact subset of X. Adjoin $X \diagdown A$ to any open cover $\mathscr{O}$ of A. The augmented collection covers B, and so there is a finite subcollection that covers B. The members of that finite subcollection that are in $\mathscr{O}$ cover A. Now suppose $A \subset X$ is compact and $(X,\mathscr{T})$ is Hausdorff. If z is a point that is not in A, then for any $x \in A$, there are disjoint open sets U_x and V_z about x and z, respectively. There is a finite subcover $U_1,\ldots,U_n$ of the U's, and the corresponding finite intersection of the V's does not intersect A, so z is not a point of closure of A. It follows that A is closed. Now assume that if $A \subseteq X$ is compact and the topology on X is generated by a metric. Fix any point x_0 in A. The open balls $B(x_0,n)$, $n \in \mathbb{N}$, cover A, so A is contained in a finite number of them, whence A is bounded.

Theorem 9.7.3. *The continuous image of a compact set is compact.*

Proof. Let f be a continuous function on a compact set A. If $\{O_\alpha\}$ is an open cover of $f[A]$, then $\{f^{-1}[O_\alpha]\}$ is an open cover of A. Since A is compact, there are a finite number $O_1,\cdots,O_n$ of the O_α's such that $\{f^{-1}[O_i] : 1 \le i \le n\}$ covers A. It follows that $\{O_i : 1 \le i \le n\}$ covers $f[A]$.

Corollary 9.7.1. *A real-valued continuous function f defined on a compact set A takes a maximum and minimum value.*

Proof. The image $f[A]$ is a compact subset of $\mathbb{R}$. By Theorem 9.7.2, $f[A]$ is a closed subset of $[-n,n]$ for some $n \in \mathbb{N}$, whence $f[A]$ contains its lub and glb in $\mathbb{R}$.

Theorem 9.7.4. *A continuous function f from a compact metric space (X,d) into a metric space (Y,ρ) is uniformly continuous.*

Proof. Given $\varepsilon > 0$, cover X with open balls of half the radius that works for continuity and $\varepsilon/2$ at each $x \in X$. Take a finite subcover, and let δ be half the minimum radius for that finite subcover. If $d(x,y) < \delta$, then both x and y are in a ball $B(z,r)$ that maps into $B(f(z),\varepsilon/2)$. It follows that $\rho(f(x), f(y)) < \varepsilon$.

Proposition 9.7.2. *If $\mathscr{B}$ is a base for the topology of X, then a subset of X is compact if and only if every covering by sets from $\mathscr{B}$ has a finite subcovering.*

Proof. Replace each set from an open covering by the sets from $\mathscr{B}$ that it contains. Reduce that covering to a finite subcovering $B_1,\cdots,B_n$. Then replace each B_i with one of the original open sets that contains B_i.

Definition 9.7.2. A topological space $(X,\mathscr{T})$ has the **Bolzano-Weierstrass property** if every sequence in X has at least one cluster point in X.

Theorem 9.7.5. *If X is compact, then every net in X, and in particular every sequence in X, has a cluster point in X. That is, the space has the Bolzano-Weierstrass property. Conversely, if every net in X has a cluster point, then X is compact. On the other hand, if X satisfies the second axiom of countability, then X is compact if and only if X has the Bolzano-Weierstrass property.*

Proof. Suppose $\langle x_\alpha : \alpha \in \mathscr{D}\rangle$ is a net in X and X is compact. Then the collection of closed sets $F_\gamma = \overline{\{x_\alpha : \alpha \geq \gamma\}}$, $\gamma \in \mathscr{D}$, has the finite intersection property, so there is a point x in the intersection. For any open neighborhood U of x and any $\gamma \in \mathscr{D}$, $U \cap \{x_\alpha : \alpha \geq \gamma\} \neq \varnothing$. That is, x is a cluster point of the net.

Conversely, let $\mathscr{F}$ be a family of closed subsets of X such that $\mathscr{F}$ has the finite intersection property. Let $\mathscr{D}$ denote the collection of finite subcollections of $\mathscr{F}$. The ordering is given by $\subseteq$, with the larger set being further along in the ordering. The relation $\subseteq$ is transitive, and given two finite collections, their union contains them both. Therefore, $\langle \mathscr{D}, \subseteq\rangle$ is a directed set. For each finite subcollection of $\mathscr{F}$ in $\mathscr{D}$, let the corresponding value of the net be an element chosen in the intersection of the members of $\mathscr{F}$. (In general, we are using the Axiom of Choice here.) We are assuming that this net has a cluster point x. Therefore, for each neighborhood U of x and for each $F \in \mathscr{F}$, there is a finite subset $\{F_1,\ldots,F_n\}$ of $\mathscr{F}$ containing F, and there is a point y in the intersection $\cap_{i=1}^n F_i$ with $y \in U$. Since $y \in F$, x is a closure point of the closed set F, i.e., $x \in F$. It follows that x is in the intersection of all of the sets in $\mathscr{F}$.

Assume X has a countable base $\mathscr{B}$. To show X is compact, it is enough to show that every covering by sets from $\mathscr{B}$ has a finite subcovering. Equivalently, it is enough to consider a countable family $\{F_1, F_2, \cdots\}$ consisting of the complements of the sets in $\mathscr{B}$ with the property that for any $n \in \mathbb{N}$, $\cap_{i=1}^n F_i \neq \varnothing$. Now for each $n \in \mathbb{N}$, we pick a point $x_n \in \cap_{i=1}^n F_i$. The assumption that this sequence has a cluster point x means that x is a point of closure of each of the F_i's, whence $x \in \cap \mathscr{F}$. Therefore, to establish compactness for a space with a countable base, one need only show that the space has the Bolzano-Weierstrass property.

Example 9.7.1. In a space satisfying the first axiom of countability, such as a metric space, if a sequence $\langle x_n \rangle$ has a cluster point x, then a subsequence $\langle x_{n_i} \rangle$ converges to x. On the other hand, the space of ordinals, with the order topology, up to but not including the first uncountable ordinal Ω satisfies the first but not the second axiom of countability. (See Remark 9.4.1 and Problem 9.9.) For that space, every sequence has a cluster point with a subsequence converging to that cluster point, but the space is not compact. We will see (Example 9.9.1) that there is a compact space, namely the Stone-Čech compactification of the natural numbers, that contains a sequence with no converging subsequence even though the space must have the Bolzano-Weierstrass property.

Proposition 9.7.3. *Suppose (X,d) is a metric space with the Bolzano-Weierstrass property, and fix $n \in \mathbb{N}$. Then there is a finite subset S of X with the following properties:*

i) *For distinct points x and y in S, $d(x,y) \geq 1/n$, and*
ii) *For any $z \in X$ such that $z \notin S$, there is an $x \in S$ with $d(x,z) < 1/n$.*

Proof. Any subset of X with just one point satisfies the first property. If S satisfies the first but not the second property, then there is a point that can be added to S and the first property will still be satisfied. This addition must stop at some finite step, since a countably infinite set with the first property forms a sequence with no cluster point.

Theorem 9.7.6. *Suppose (X,d) is a metric space with the Bolzano-Weierstrass property. Then X is separable and compact.*

Proof. By Proposition 9.7.3, for each $n \in \mathbb{N}$, there is a finite set S_n such that for each $x \in X$, there is an $s \in S_n$ with $d(x,s) < 1/n$. The union $\cup_{n \in \mathbb{N}} S_n$ is a countable dense set. By Theorem 9.1.2, X has a countable base for the topology, so by Theorem 9.7.5 X is compact.

Theorem 9.7.7 (Heine-Borel). *A subset of $\mathbb{R}^m$, $m \in \mathbb{N}$, is compact if and only if it is closed and bounded.*

Proof. By Theorem 9.7.2, a compact subset of $\mathbb{R}^m$ is closed and bounded. Let A be a closed and bounded subset of $\mathbb{R}^m$. Let $\langle x_n : n \in \mathbb{N} \rangle$ be a sequence in A. Choose a subsequence $\langle x_{n_i} : i \in \mathbb{N} \rangle$ so that the first coordinates of the subsequence converge to the $\limsup$ of the first coordinates of the original sequence. Choose a subsequence of that subsequence so that the second coordinates converge. After m steps, one has a subsequence converging to a point that must be in A since A is closed. The limit of the subsequence is a cluster point of the original sequence, so A is compact.

Example 9.7.2. **Warning:** It is not true for an infinite-dimensional space, such as ℓ^2, that a closed and bounded set is compact. Let A be the basis for ℓ^2 consisting of sequences e_i that are zero except at the i^{th} place, where the value 1 is taken. Each element of A is a distance $\sqrt{2}$ from every other element. Therefore, the closed unit ball of ℓ^2, although closed and bounded, does not have the Bolzano-Weierstrass property.

Definition 9.7.3. A metric space is totally bounded if for each $\varepsilon > 0$, the space is covered by a finite number of balls of radius ε.

Theorem 9.7.8. *A metric space is compact if and only if it is both complete and totally bounded.*

Proof. Clearly, a compact metric space is totally bounded, and any Cauchy sequence has a cluster point to which the whole sequence must converge. For the converse, we assume that (X,d) is complete and totally bounded. For each $n \in \mathbb{N}$, take a finite covering C_n of X by balls of radius $1/n$. Let S_n be the finite set of centers of those balls. The union $\cup_{n\in\mathbb{N}} S_n$ is a countable dense set, so by Theorem 9.1.2, (X,d) has a countable base for the topology. Let $\langle x_n : n \in \mathbb{N}\rangle$ be a sequence in X. A subsequence is in some ball $B(y_1,1) \in C_1$. A subsequence of that sequence is in some ball $B(y_2,1/2) \in C_2$, etc. The diagonal sequence, formed by mapping each $n \in \mathbb{N}$ to the n^{th} point in the n^{th} subsequence, is Cauchy, and therefore has a limit. That limit is a cluster point of the original sequence.

Example 9.7.3. The map $\tan x$ is a homeomorphism from $(-\frac{\pi}{2},\frac{\pi}{2})$ onto $\mathbb{R}$. Note that $(-\frac{\pi}{2},\frac{\pi}{2})$ is totally bounded, but $\mathbb{R}$ is not. This shows that total boundedness may not be preserved by a homeomorphic map. It is, however, preserved by any uniformly continuous map. See Problem 9.14.

9.8 Properties of Topologies

Two topologies that always exist on any nonempty set X are given by $\mathscr{T} = \{X, \varnothing\}$ and by $\mathscr{T} = \mathscr{P}(X)$, i.e., the set of all subsets of X. The first is called the trivial topology; the second is called the discrete topology, since every point forms an open set. The discrete topology can be generated by the following metric: Set $d(x,y) = 1$ for all x, $y \in X$.

It is common to have two topologies, $\mathscr{T}_1$ and $\mathscr{T}_2$, on X where every set open for the first is open for the second; that is, $\mathscr{T}_1 \subseteq \mathscr{T}_2$. We say $\mathscr{T}_1$ is weaker or coarser than $\mathscr{T}_2$ and $\mathscr{T}_2$ is stronger or finer than $\mathscr{T}_1$. Of course, the trivial topology is the weakest possible topology, and the discrete topology is the strongest possible topology on a set. A topology $\mathscr{T}_2$ is stronger than $\mathscr{T}_1$ on a set X if for each point $x \in X$, any $\mathscr{T}_1$-open neighborhood of x contains a $\mathscr{T}_2$-open neighborhood of x. Of course, if each topology is stronger than the other, then they are equal. Recall that in a Hausdorff space, distinct points have disjoint open neighborhoods.

Theorem 9.8.1. *Suppose $(X,\mathscr{T})$ is a compact Hausdorff space. Then any strictly weaker topology is not a Hausdorff topology, and with any strictly stronger topology, the space is not compact.*

Proof. Assume that $\mathscr{T}_0$ is strictly weaker than $\mathscr{T}$, and let f be the identity map $x \mapsto x$ on X mapping $(X,\mathscr{T})$ onto $(X,\mathscr{T}_0)$. The map f is continuous. Fix O in $\mathscr{T}$ such that $O \notin \mathscr{T}_0$, and let $C = X \diagdown O$. By Theorems 9.7.2 and 9.7.3, C is compact with respect to $\mathscr{T}$, so C is compact with respect to $\mathscr{T}_0$. On the other hand, C is not closed with respect to $\mathscr{T}_0$, so $(X,\mathscr{T}_0)$ is not Hausdorff. The rest is Problem 9.16.

Any collection of subsets of X generates a topology on X. The generated topology is the intersection of all topologies (including the discrete topology) containing the given collection. It is the weakest topology containing the collection. (See Problem 9.21.)

Definition 9.8.1. We say that disjoint sets A and B in a topological space can be separated by disjoint open sets if there are open sets U, V with $U \cap V = \varnothing$ and $A \subseteq U, B \subseteq V$.

Definition 9.8.2. Let (X,T) be a topological space. The following is a list of separation properties, given in increasing order of strength, with all properties assuming that singleton sets in X are closed. That is, if $x \in X$, then $X \diagdown \{x\}$ is open.

Hausdorff) If x and y are distinct points in X, then there are disjoint open neighborhoods of x and y, respectively.
Regular) If F is a closed subset of X and $x \in X \diagdown F$, then there are disjoint open sets containing F and x, respectively.
Completely Regular) The family $\mathscr{F}$ of bounded, real-valued, continuous functions on X **separates points and closed sets**. That is, if F is a closed subset of X and $x \in X \diagdown F$, then there is an $f \in \mathscr{F}$ such that $f(x) \notin \overline{f[F]}$.
Normal) Disjoint closed subsets of X can be separated by disjoint open sets.

Remark 9.8.1. Given that singleton sets are closed, it is easy to see that regular spaces are Hausdorff, and completely regular spaces are regular. The fact that normal spaces are completely regular follows from Urysohn's Lemma, which is presented next. We also note that a convenient formulation of regularity is that for any point $x \in X$ and any open set U with $x \in U$, there is an open set V with $x \in V \subseteq \overline{V} \subseteq U$. See Problem 9.22.

Theorem 9.8.2 (Urysohn's Lemma). *Let A and B be disjoint closed subsets of a normal space X. There is a continuous function $f : X \mapsto [0,1]$ such that $f \equiv 0$ on A and $f \equiv 1$ on B.*

Proof. If a space is normal, then for any closed set F and any open set W with $F \subseteq W$, there is an open set O with $F \subseteq O \subseteq \overline{O} \subseteq W$. Recall that dyadic rational numbers in $[0,1]$ have the form $k/2^n$ for $k \in \mathbb{N}$, $0 \leq k \leq 2^n$. Let $O_1 = X - B$, and fix an open set O_0 with $A \subseteq O_0 \subseteq \overline{O}_0 \subseteq O_1 = X - B$. Choose $O_{1/2}$ with $\overline{O}_0 \subseteq O_{1/2} \subseteq \overline{O}_{1/2} \subseteq O_1$. Between $\overline{O}_0$ and $O_{1/2}$, and between $\overline{O}_{1/2}$ and O_1, we can fit open sets $O_{1/4}$ and $O_{3/4}$

with similar properties. Continuing in this way, there is a countable family of open sets $\{O_r\}$ indexed by the dyadic rationals in $[0,1]$ such that the following holds: If $r < s$, then $A \subseteq O_r \subseteq \overline{O}_r \subseteq O_s \subseteq X - B$, etc. Now as shown in Problem 9.23, the desired function f is obtained by setting $f(x) = \inf\{r : x \in O_r\}$.

Next we sketch an application of Urysohn's Lemma generalizing the useful property that a continuous function on a closed subset of $\mathbb{R}$ has a continuous extension to all of $\mathbb{R}$.

Theorem 9.8.3 (Tietze's Extension Theorem). *Any continuous real-valued function f defined on a closed set S in a normal space X has a continuous extension to all of X; the extension can keep the same bounds as the original function.*

Proof. By composing f with a homeomorphism, we may assume that we have $\sup_{x\in S} f(x) = 1$ and $\inf_{x\in S} f(x) = -1$. The final step of the proof is to compose the constructed function with the inverse of the homeomorphism. Let $A = \{x \in S : f(x) \leq -1/3\}$, and let $B = \{x \in S : f(x) \geq 1/3\}$. Both sets are closed. Let h_1 be a continuous function on X such that $h_1 : X \mapsto [-1/3, 1/3]$, $h_1 \equiv -1/3$ on A and $h_1 \equiv 1/3$ on B. Since $|f(x)| \leq 1/3$ on $S \setminus (A \cup B)$, $-2/3 \leq f - h_1 \leq 2/3$ on S. Let $g = \frac{3}{2} \cdot (f - h_1)$. Applying the previous step, there is a continuous function h_2 on X such that $\frac{3}{2}h_2 : X \mapsto [-1/3, 1/3]$, and $-2/3 \leq g - \frac{3}{2}h_2 \leq 2/3$ on S. It follows that $h_2 : X \mapsto [-\frac{2}{3^2}, \frac{2}{3^2}]$, and $-(2/3)^2 \leq f - h_1 - h_2 \leq (2/3)^2$ on S. It now follows that for each $n \geq 2$, there is a collection of continuous functions $\{h_1, h_2, \cdots, h_n\}$ on X such that $\sum_{i=1}^n |h_i| \leq \sum_{i=1}^n \frac{2^{i-1}}{3^i}$, whence the sum $h := \sum_{i=1}^\infty h_i$ is continuous on X. Moreover, $|f - \sum_{i=1}^n h_i| \leq (2/3)^n$, whence $h = f$ on S.

Definition 9.8.3. A topological space is **Lindelöf** if any collection of open sets has a countable subcollection with the same union.

Proposition 9.8.1. *A topological space with a countable base for the topology is Lindelöf.*

Proof. For each point in the union of the original collection of open sets, choose a base set containing the point and contained in one of the open sets of the collection. This collection of base sets is countable. Replace each one of these with one of the original open sets containing it.

Theorem 9.8.4. *A regular, Lindelöf space is normal. In particular, a regular, second countable space is normal.*

Proof. Given disjoint closed sets A and B in the space X, we may enclose each point $x \in A$ in an open set O such that $x \in O \subseteq \overline{O} \subseteq X \setminus B$. We may do the same thing for each point in B. By the Lindelöf property, A is contained in the union of a countable family of open sets U_i such that $\overline{U}_i \cap B = \varnothing$ for each i. Similarly, B is contained in the union of a countable family of open sets V_i such that $\overline{V}_i \cap A = \varnothing$ for each i. For each $n \in \mathbb{N}$, let $O_n = U_n \setminus \cup_{i \leq n} \overline{V}_i$ and let $W_n = V_n \setminus \cup_{i \leq n} \overline{U}_n$. Then $O = \cup_{n\in\mathbb{N}} O_n$ and $W = \cup_{n\in\mathbb{N}} W_n$ are open sets with $A \subseteq O$ and $B \subseteq W$. If $x \in O$, then there is a first n with $x \in O_n$. It follows that $x \notin V_i$ for $1 \leq i \leq n$. It also follows that $x \notin W_j$ for $j > n$. Therefore, $x \notin W$. Thus $O \cap W = \varnothing$.

Proposition 9.8.2. *A compact Hausdorff space is normal.*

Proof. Given disjoint closed and therefore compact sets A and B, for each $x \in A$ and $y \in B$, there are disjoint open sets U_x and U_y containing x and y, respectively. Given $x \in A$, take a finite subcover of the U_y's, $y \in B$; let W_x be the finite intersection of the corresponding U_x's. This yields disjoint open sets W_x and V_x with $x \in W_x$ and $B \subseteq V_x$. Cover A with the open sets W_x, $x \in A$, and take a finite subcover. For each W_x in that finite subcover, there is an open V_x with $B \subseteq V_x$ and $W_x \cap V_x = \varnothing$. Let W be the union of the finite subcover of the W_x's, and let V be the finite intersection of the corresponding V_x's. The sets W and V are disjoint open sets containing A and B, respectively.

9.9 Product Spaces, Metrization, and Compactification

In $\mathbb{R}^n$, each point $(x_1, x_2, \cdots, x_n)$ is a function from the index set $\{1, 2, \cdots, n\}$ into the real numbers. It is this point of view that we extend to an index set that is not necessarily finite, and with not necessarily real value at each index.

Definition 9.9.1. Let $\mathscr{I}$ be an index set, and for each $\alpha \in \mathscr{I}$, let X_α be a topological space. The **product** $\Pi_{\alpha \in \mathscr{I}} X_\alpha$ is the set of all functions f on $\mathscr{I}$ with $f(\alpha) \in X_\alpha$ for each $\alpha \in \mathscr{I}$. Each space X_α is called a **factor** of the product. The **product topology**, also known as the topology of pointwise convergence, is generated by the following open base $f \mapsto \mathscr{B}_f$ at each element of this space of functions: An element $U \in \mathscr{B}_f$ is determined by a finite set F of indices α and an open set V_α in X_a containing $f(\alpha)$ for each $\alpha \in F$. The set U consists of all elements $g \in \Pi_{\alpha \in \mathscr{I}} X_\alpha$ such that for each $\alpha \in F$, $g(\alpha) \in V_\alpha$; for an index $\gamma \notin F$, $g(\gamma)$ can be any value in X_γ. We say that U is **restricted** at the indices of F. The set $\Pi_{\alpha \in \mathscr{I}} X_\alpha$ with the product topology is called a **product space**. The mapping that takes each $f \in \Pi_{\alpha \in \mathscr{I}} X_\alpha$ to the restriction on a subset (possibly a singleton set) $\mathscr{J}$ of $\mathscr{I}$ is called the **projection** on $\Pi_{\alpha \in \mathscr{J}} X_\alpha$.

A helpful way to picture a product space is to think of the index set as points on the x-axis in the plane. For each index α, the topological space X_α is represented by a vertical line above α. An element f of the product is represented by a path passing through each of the vertical lines. An element U of the open base $\mathscr{B}_f$ at f is given by a finite set F of indices α and an open interval V_α containing $f(\alpha)$ in the vertical line above α. An element g of the product is in U if it takes a value in V_α at each α in F. A special case of such an element in $\mathscr{B}_f$ is given by specifying a single α to form by itself the finite subset of the index set. The corresponding U_α in $\mathscr{B}_f$ must be open, and finite intersections of these form a typical element of $\mathscr{B}_f$. Since any topology containing the product topology must contain such sets U_α, we have the following result.

Proposition 9.9.1. *The product topology is the weakest topology for which each projection $f \mapsto f(\alpha)$ from the product space to X_α, $\alpha \in \mathscr{I}$, is a continuous map.*

Proposition 9.9.2. *Each factor X_α of a product space is Hausdorff if and only if the product space $\Pi_{\alpha\in\mathscr{I}} X_\alpha$ is Hausdorff, in which case, each subspace of the product is Hausdorff.*

Proof. Exercise 9.25.

We can indicate with appropriate notation a subset of a product space where the elements are restricted at certain indices. For example, suppose β and γ are indices and $S_\beta \subset X_\beta$ while $S_\gamma \subset X\gamma$. Let T be the subset of the product space consisting of those elements for which the value at β is restricted to S_β and the value at γ is restricted to S_γ, and there are no other restrictions. Let $\mathscr{J} = \mathscr{I} \setminus \{\beta, \gamma\}$. Then $T = S_\beta \times S_\gamma \times \Pi_{\alpha\in\mathscr{J}} X_\alpha$.

An important special type of a product space is the one where each factor X_α is a compact interval $I_\alpha \subset \mathbb{R}$. If the index set is countably infinite, we will assume it is $\mathbb{N}$, and write $\Pi_{n\in\mathbb{N}} I_n$ for the product.

Definition 9.9.2. A topological space $(X, \mathscr{T})$ is **metrizable** if there is a metric on X such that the metric topology equals $\mathscr{T}$. We also say that $\mathscr{T}$ is metrizable.

Theorem 9.9.1. *A finite or countably infinite product of compact intervals with the product topology is compact. Moreover, the product topology is metrizable.*

Proof. The proof for a finite product is Problem 9.27. Let P denote the product $\Pi_{n\in\mathbb{N}} I_n$. An element of a countable base for P is formed by a finite set $F = \{n_1, \cdots, n_k\}$ from $\mathbb{N}$, and an open interval U_{n_i} with rational endpoints contained in I_{n_i} for $i = 1, \cdots, k$. It consists of all elements of P that take values in U_{n_i} at the index n_i. That is the product

$$U_{n_1} \times \cdots \times U_{n_k} \times \Pi_{n\in\mathbb{N}\setminus F} I_n.$$

Since P has a countable base for the topology, sequences suffice to establish compactness. Given a sequence $\langle f_m : m \in \mathbb{N}\rangle$, choose a subsequence $\langle f_m^1\rangle$ that converges in I_1. Given $\langle f_m^k\rangle$ chose a subsequence $\langle f_m^{k+1}\rangle$ that converges in I_{k+1}. The diagonal sequence $\langle f_m^m : m \in \mathbb{N}\rangle$ converges in P, and the limit is a cluster point of the original sequence.

There is a metric d_n on the n^{th} factor of P, namely, $d_n(x,y) = |x-y| \wedge 1$; that is, we take the smaller of the usual distance and 1. The sum $d := \sum_{n=1}^{\infty} 2^{-n} \cdot d_n$ is a metric on P. The fact that d is symmetry and has the transitive property follows from those properties for each d_n. Moreover, $f = g$ in P if and only if $f(n) = g(n)$ for every index n, so $d(f,g) = 0$ if and only if $f = g$. To show that d generates the product topology on P, fix $f \in P$. Fix a basic product neighborhood of f consisting of elements $g \in P$ for which the values at indices in $F = \{n_1, \cdots, n_k\}$ are restricted to small open intervals containing the values taken by f. There is an $\varepsilon > 0$ such that the elements $g \in P$ for which $d(f,g) < \varepsilon$ all meet those finite number of restrictions. Conversely, given an open metric ball $B(f,r)$ centered at f, there is a $k \in \mathbb{N}$ such that $\sum_{n=k+1}^{\infty} \frac{1}{2^n} < r/2$. Fix a positive $\delta < \frac{r}{2k}$. The product open neighborhood of f consisting of elements $g \in P$ such that $|f(n) - g(n)| < \delta$ for $1 \le n \le k$ is contained in $B(f,r)$. It follows that the metric topology equals the product topology on P.

Remark 9.9.1. With a similar proof, the same result is valid for a countable product of second countable, compact metric spaces. We next show, however, that any product of compact spaces is compact. The proof is taken from the author's 1965 article [26]. It uses the Axiom of Choice, which, by a result of Kelley [24], is unavoidable for the general case. Also see the proof of Theorem C.12.2.

Lemma 9.9.1. *Let X and Y be topological spaces, and assume that X is compact. Let $\mathscr{O}$ be a collection of sets of the form $U \times V$, where U is open in X and V is open in Y. Assume that there is no finite subset of $\mathscr{O}$ for which the union is all of $X \times Y$. Then there is a nonempty closed set $C \subseteq X$ such that for each $x \in C$ and each finite set $\{U_i \times V_i : 1 \le i \le n\} \subseteq \mathscr{O}$ with $x \in \cap_i^n U_i$, $Y \setminus \cup_{i=1}^n V_i \ne \varnothing$.*

Proof. Let $\mathscr{F}$ be the collection of all finite sets $\{U_i \times V_i : 1 \le i \le n\} \subseteq \mathscr{O}$ such that $Y = \cup_{i=1}^n V_i$. For each such set F in $\mathscr{F}$, let $W_F := \cap_{i=1}^n U_i$. If (x,y) is in $X \times Y$ and $x \in W_F$, then $y \in \cup_{i=1}^n V_i$, so (x,y) is in one of the members of F. Therefore, the family of open set W_F, $F \in \mathscr{F}$, does not cover X, for if it did, then a finite subcollection $\{W_{F_j} : 1 \le j \le k\}$ would cover X, and the corresponding collection $\cup_{j=1}^k F_j$ would cover $X \times Y$. Let $C = X \setminus \cup_{F \in \mathscr{F}} W_F$. Now if $x \in C$, and $F = \{U_i \times V_i : 1 \le i \le n\}$ is a finite subset of $\mathscr{O}$ with $x \in U_i$ for each i, then $Y \ne \cup_{i=1}^n V_i$.

Theorem 9.9.2 (Tychonoff Product Theorem). *The product of compact spaces is compact.*

Proof. We will call a collection of sets from a base for the topology "admissible" if no finite subcollection covers the space. Let $\mathscr{O}$ be an admissible collection of sets from the base for the product topology. By Proposition 9.7.2, we need to find an element f in the product $\Pi_{\alpha \in \mathscr{I}} X_\alpha$ that is in none of the sets in $\mathscr{O}$. Using the Axiom of Choice, we assume that the index set $\mathscr{I}$ is well-ordered; that is, any nonempty set has a first member in the ordering. If the index set is supplied with an enumeration, such as for $\mathbb{N}$, a well-ordering is available without the Axiom of Choice. By moving the first element of $\mathscr{I}$ to the last position, we may also assume that $\mathscr{I}$ has a terminating element τ. We also assume that there is a function T that chooses an element from each closed subset of each factor. For sets such as intervals $[a,b]$, a function T is available without the Axiom of Choice.

Let X_{α_1} denote the first factor in the product, and let $Y_{\alpha_1} := \Pi_{\alpha > \alpha_1} X_\alpha$. By the lemma, there is a closed subset C_{α_1} of X_{α_1} and a point $f(\alpha_1)$ chosen by T from C_{α_1} such that if $\mathscr{O}_{\alpha_1}$ is the subset of $\mathscr{O}$ consisting of sets for which the projection on X_{α_1} contains $f(\alpha_1)$, then the projection of the sets in $\mathscr{O}_{\alpha_1}$ on Y_{α_1} is an admissible collection on Y_{α_1}.

Now let $\mathscr{K}$ be the set of indices $\alpha \in \mathscr{I}$ such that a value $f(\gamma)$ has been chosen for all indices $\gamma \le \alpha$. Let $\mathscr{O}_\alpha$ consist of those members O of $\mathscr{O}$ such that for each $\gamma \le \alpha$, the projection of O on X_γ contains $f(\gamma)$. We assume the further property for $\mathscr{K}$ that for each $\alpha \in \mathscr{K}$, the projection of the sets in $\mathscr{O}_\alpha$ on $\Pi_{\gamma>\alpha}X_\gamma$ is an admissible collection on $\Pi_{\gamma>\alpha}X_\gamma$. We have shown that $\alpha_1 \in \mathscr{K}$. Suppose all indices $\alpha < \beta$ are in $\mathscr{K}$. Let $\mathscr{O}_{\beta-} := \cap_{\alpha<\beta}\mathscr{O}_\alpha$. If β has an immediate predecessor ξ, then since $\xi \in \mathscr{K}$, the projection of the sets in $\mathscr{O}_{\beta-}$ is an admissible collection $\Pi_{\gamma\geq\beta}X_\gamma$. Suppose β has no immediate predecessor. (An example would be the first index bigger than those indices that have a finite number of predecessors.) Let F be a finite subset of $\mathscr{O}_{\beta-}$. Each $O \in F$ is restricted at only a finite number of indices, and since F is finite, there is a $\xi < \beta$ such that no O in F is restricted at indices γ with $\xi < \gamma < \beta$. Since $\xi \in \mathscr{K}$, the projection of the sets in F on $\Pi_{\alpha>\xi}X_\alpha$ does not cover $\Pi_{\alpha>\xi}X_\alpha$, so the projection of the sets in F on $\Pi_{\alpha\geq\beta}X_\alpha$ does not cover $\Pi_{\alpha\geq\beta}X_\alpha$. In either case, it follows that the projection of the sets in $\mathscr{O}_{\beta-}$ on $\Pi_{\gamma\geq\beta}X_\gamma$ is an admissible collection on $\Pi_{\gamma\geq\beta}X_\gamma$. Applying the lemma to the projection of the sets in $\mathscr{O}_{\beta-}$ on $\Pi_{\gamma\geq\beta}X_\gamma$ where $X = X_\beta$ and $Y = \Pi_{\alpha>\beta}X_\alpha$, we may choose $f(\beta)$ so that $\beta \in \mathscr{K}$.

It now follows, since $\mathscr{I}$ is well-ordered, that all indices smaller than the terminal index τ are in $\mathscr{K}$. Therefore, the projection of the sets in $\mathscr{O}_{\tau-} := \cap_{\alpha<\tau}\mathscr{O}_\alpha$ on X_τ is a collection of open subsets of X_τ with no finite subcover. Since X_τ is compact, there is a value $f(\tau)$ in the complement of all of those projections. This completes the choice of f on $\mathscr{I}$ so that f is in no $O \in \mathscr{O}$.

The following construction of an imbedding in a product space is the foundation of metrization and classical compactifications.

Theorem 9.9.3. *Let $(X, \mathscr{T})$ be a completely regular space, and let $\mathscr{F}$ be family of continuous functions on X such that each $f \in \mathscr{F}$ takes its values in a compact interval $I_f \subset \mathbb{R}$. We assume that the family $\mathscr{F}$ separates points and closed sets. Let $P := \Pi_{f\in\mathscr{F}}I_f$. For each $x \in X$, let $\phi(x)$ be the element of P such that for each index $f \in \mathscr{F}$, $\phi(x)(f) = f(x)$. Then ϕ is a homeomorphism of X onto $\phi[X] \subseteq P$.*

Proof. If $x \neq y$ in X, then some $f \in \mathscr{F}$ takes a value $f(x) \neq f(y)$, so ϕ is an injective map. Given $x \in X$, let S be a basic open neighborhood of $\phi(x)$. The set S is determined by a finite number of functions $f_1, \cdots, f_n$ in $\mathscr{F}$ and open sets V_i containing $f_i(x)$ in I_{f_i} for $i = 1, \cdots, n$. By continuity, there is an open neighborhood W of x such that for each $i = 1, \cdots n$, $f_i[W] \subseteq V_i$. This shows that ϕ is continuous on X. Now fix an open neighborhood W of x in X and a function g in $\mathscr{F}$ such that $g(x) \in V := I_g \setminus \overline{g[X \setminus W]}$. Note that V is an open subset of I_g. If $y \notin W$, then $g(y) \in \overline{g[X \setminus W]}$, whence $g(y) \notin V$. Therefore, if $g(z) \in V$, then $z \in W$. Let U be the open subset of P consisting of those elements of P taking values in V at f_0 and unrestricted at other indices in $\mathscr{F}$. Now,

$$\phi^{-1}[U] = \phi^{-1}\left[V \times \Pi_{f\neq g}I_f\right] = g^{-1}[V] \subseteq W.$$

This now shows that, ϕ is a homeomorphism of X onto $\phi[X] \subseteq P$.

Theorem 9.9.4 (Urysohn Metrization). *A regular, second countable topological space $(X, \mathscr{T})$ is metrizable.*

Proof. By Theorem 9.8.4, $(X,\mathscr{T})$ is normal. Let $\mathscr{B}$ be a countable base of open sets for $\mathscr{T}$. Fix a closed set $C \neq \varnothing$ and a point $x \notin C$. There are open sets U and V in $\mathscr{B}$ such that

$$x \in V \subseteq \overline{V} \subseteq U \subseteq \overline{U} \subset X \diagdown C.$$

By Urysohn's Lemma 9.8.2, there is a continuous function $f : X \mapsto [0,1]$, such that $f(x) \equiv 0$ on $\overline{V}$ and $f(x) \equiv 1$ on $X \diagdown U$. Since $\mathscr{B}$ is countable, it follows from Theorem 9.9.3 that there is a homeomorphism from $(X,\mathscr{T})$ onto a subspace of a countable product of compact intervals. By Theorem 9.9.1, $(X,\mathscr{T})$ is metrizable.

We now turn to imbedding spaces as dense subsets of compact spaces. By Theorem 9.9.3, we may think of the original space as a subspace of an appropriate product space. The set of elements that are adjoined to the original space to form the compact space is called the **remainder**.

Definition 9.9.3. Let $(X,\mathscr{T})$ be a topological space, and let Q be a family of continuous, bounded, real-valued functions on X. Let Z be a compact space containing X as a dense subspace such that each $f \in Q$ has a continuous extension $\widetilde{f}$ to Z, and the set of extensions **separates the points** of $Z \diagdown X$. That is, if ξ and η are in $Z \diagdown X$, then there is an $f \in Q$ such that $\widetilde{f}(\xi) \neq \widetilde{f}(\eta)$. Then Z is called a ***Q*-compactification** of X.

Proposition 9.9.3. *Suppose $(X,\mathscr{T})$ is a dense subspace of a compact Hausdorff space $(Z,\mathscr{S})$; that is, $\mathscr{T}$ is the relative S-topology on X. Then for some family Q, $(Z,\mathscr{S})$ is a Q-compactification of X. Moreover, we may assume that Q contains sufficient functions to separate points and closed subsets of X.*

Proof. For each pair of distinct points ξ and η in $Z \diagdown X$, there is a continuous function f on Z taking values in $[0,1]$ with $f(\xi) = 0$ and $f(\eta) = 1$. Let Q contain the restriction to X of one such function for each pair of points in $Z \diagdown X$. Recall that a $\mathscr{T}$-closed subset of X is $C \cap X$, where C is $\mathscr{S}$-closed. If $x \in X \diagdown C$, then there is a continuous function g on Z taking values in $[0,1]$ with $f(x) = 0$ and $f(z) \equiv 1$ on C. We may assume that the restriction of g to X is also in Q.

Theorem 9.9.5. *Let $(X,\mathscr{T})$ be a completely regular space, and let Q be a family of continuous functions on X such that each $f \in Q$ takes its values in a compact interval $I_f \subset \mathbb{R}$. Suppose Q separates points and closed subsets of X. Let $P := \Pi_{f \in Q} I_f$, and let ϕ be the homeomorphism of $(X,\mathscr{T})$ onto the subspace $Y := \phi[X] \subseteq P$ given by Theorem 9.9.3. Then the closure $\overline{Y}$ is a compact Hausdorff space containing Y as a dense subspace. For each $f \in Q$, let $\widetilde{f}$ be the function on Y given by $\widetilde{f}(\phi(x)) = \phi(x)(f) = f(x)$. Then each $\widetilde{f}$ has a continuous extension $\widetilde{f}$ to the remainder, and that extension separates points of the remainder. Thus, if we associate X with its homeomorphic image Y, then $\overline{Y}$ is a Q-compactification.*

Proof. Since a closed subset of a compact space is compact, $\overline{Y}$ is compact. By Proposition 9.9.2, $\overline{Y}$ is Hausdorff, and by definition, Y is dense in $\overline{Y}$. Fix $f \in Q$, and let p_f be the projection of the product space P onto the space I_f. Since p_f is continuous, the restriction to $\overline{Y}$ is continuous, but that restriction on Y is $\widetilde{f}$. If $\xi \neq \eta$ in P, then for some $f \in Q$, $\xi(f) \neq \eta(f)$, whence if ξ and η are points in the remainder, then $\widetilde{f}(\xi) \neq \widetilde{f}(\eta)$.

Definition 9.9.4. A space is locally compact if there is an open base $x \mapsto \mathscr{B}_x$ at each point consisting of sets with compact closure.

In the following result, we do not distinguish between a topological space X and its homeomorphic image in an appropriate product space.

Theorem 9.9.6. *Let $(X, \mathscr{T})$ be dense subspace of a compact Hausdorff space $(Z, \mathscr{S})$. That is, $\mathscr{T}$ is the relative S-topology on X. Then X is $\mathscr{S}$-open in Z if and only if $(X, \mathscr{T})$ is locally compact.*

Proof. Exercise 9.30(A).

Definition 9.9.5. A continuous function with **compact support** on a topological space $(X, \mathscr{T})$ is a continuous function that is identically equal to 0 outside of some compact subset of X.

Proposition 9.9.4. *Let $(X, \mathscr{T})$ be a locally compact Hausdorff space. Let $\mathscr{F}$ be the set of continuous functions with compact support mapping X into $[0,1]$. Then $\mathscr{F}$ separates points and closed subsets of X, whence $(X, \mathscr{T})$ is completely regular.*

Proof. Let $C \neq \varnothing$ be a closed subset of X, and fix $x \notin C$. Let U be an open neighborhood of x with compact closure $\overline{U}$ such that $U \cap C = \varnothing$. If, as for example, in a discrete space, $\overline{U} = U$, let $h \equiv 1$ on U and $h \equiv 0$ on $X \diagdown \overline{U}$. Otherwise, by Proposition 9.8.2, $\overline{U}$ with the relative topology is normal, so there is a continuous function g on $\overline{U}$ taking its values in $[0,1]$ such that $g(x) = 1$ and $g \equiv 0$ on $\overline{U} \diagdown U$. Let $h = g$ on $\overline{U}$ and $h \equiv 0$ on $X \diagdown \overline{U}$. By Problem 9.31, h is continuous on X.

Finally we consider what compactifications are produced by various choices of a family Q. If $(X, \mathscr{T})$ is a locally compact Hausdorff space, we will follow Constantinescu and Cornea [12], and assume that every continuous function with compact support mapping X into $[0,1]$ is in the family Q. These functions will have an extension that takes the value 0 at the remainder points of the compactification. With these functions, and possibly more, it follows from Theorem 9.9.6 that the space X will be an open subset of the Q-compactification. For more general completely regular spaces, we will assume that Q has members sufficient to separate points and closed subsets of X.

Definition 9.9.6. The **Alexandroff one-point compactification** of a locally compact space adjoins a single point for which the open neighborhoods consist of complements of compact sets. It is the Q-compactification, where Q is the family of continuous, real-valued functions with compact support.

Definition 9.9.7. The **Stone-Čech compactification** βX of a completely regular space $(X, \mathscr{T})$ is the Q-compactification where Q consists of all bounded, continuous, real-valued function on X.

Remark 9.9.2. As noted, if $(X,\mathscr{T})$ is a dense subspace of a compact Hausdorff space Y, then Y is a Q-compactification of X for some family Q. An appropriate family Q contains sufficient functions to separate points and closed subsets of X. Also as noted, a Q-compactification of $(X,\mathscr{T})$ can be constructed by embedding $(X,\mathscr{T})$ as a homeomorphic image in a product of compact intervals with Q acting as the index set. The bigger the family Q, the bigger the corresponding product. For a family Q_1 that contains a family Q_0, the projection from the larger product to the product for the smaller index set yields a continuous map from the Q_1-compactification onto the Q_0-compactification that leaves the image of X essentially fixed. In this sense, the Stone-Čech compactification of a completely regular space $(X,\mathscr{T})$ is the largest compactification of X.

Example 9.9.1. The Stone-Čech compactification $\beta\mathbb{N}$ of the natural numbers is an example of a compact space containing a sequence with a cluster point but no convergent subsequence. The set of natural numbers itself forms an example of such a sequence. If $\langle n_i : i\in\mathbb{N}\rangle$ were a strictly increasing subsequence of $\mathbb{N}$ converging to a limit point x, then the restriction of every bounded function f on $\mathbb{N}$ to $\langle n_i : i\in\mathbb{N}\rangle$ would have to have a limit at x, since f has a continuous extension to $\beta\mathbb{N}$. On the other hand, if $f(n_i)=(-1)^i$, no such limit can exist.

Remark 9.9.3. For a different construction of the remainder set and an extension of Q-compactifications, see [27]. For an alternative to the embedding construction, see [21].

9.10 Ascoli-Arzelá Theorem

In this section, we work with a family $\mathscr{F}$ consisting of functions from a Hausdorff space $(X,\mathscr{T})$ into a metric space (Y,ρ).

Definition 9.10.1. The family $\mathscr{F}$ is **equicontinuous** at $x\in X$ if for any $\varepsilon>0$, there is an open neighborhood U_ε of x that works in terms of continuity at x for every $f\in\mathscr{F}$, that is,

$$\forall y\in U_\varepsilon,\ \ \forall f\in\mathscr{F},\ \rho(f(y),f(x))<\varepsilon.$$

The family $\mathscr{F}$ is **equicontinuous on** X if it is equicontinuous at each $x\in X$.

Note that if $\mathscr{F}$ is equicontinuous on X, then each $f\in\mathscr{F}$ is continuous on X, and any subfamily of $\mathscr{F}$ is equicontinuous on X. We assume that for each $x\in X$, the set $\{f(x): f\in\mathscr{F}\}$ is contained in a closed subset $Y_x\subseteq Y$, so $\mathscr{F}$ is a subset of the product space $\Pi=\Pi_{x\in X}Y_x$. Let $\overline{\mathscr{F}}$ denote the closure, also called the **pointwise closure**, of $\mathscr{F}$ in Π. That is, g is in the closure of $\mathscr{F}$ if for any $\varepsilon>0$ and any finite number of points $x_1,\cdots,x_n$ in X, there is an $f\in\mathscr{F}$ with $\rho(f(x_i),g(x_i))<\varepsilon$ for $1\le i\le n$.

Proposition 9.10.1. *If $\mathscr{F}$ is equicontinuous on X, then the pointwise closure $\overline{\mathscr{F}}$ is also equicontinuous on X.*

Proof. Exercise 9.32(A).

Definition 9.10.2. The **topology of uniform convergence on compact sets,** also called the **ucc topology**, on a family $\mathscr{F}$ is generated by a local filter base $\mathscr{B}_g$ at each $g \in \mathscr{F}$ that is determined as follows: Fix a compact subset K of the domain X and an $\varepsilon > 0$. The corresponding element $U \in \mathscr{B}_g$ consists of all $h \in \mathscr{F}$ such that $\rho(g(x), h(x)) < \varepsilon$ for all $x \in K$.

By Problem 9.33 the ucc-local filter base is an open base for each $g \in \mathscr{F}$. The ucc topology is like the topology of pointwise convergence. For the ucc topology, the collection of compact sets that are finite point sets in the domain is replaced with the collection of all compact subsets of the domain. It follows that there are more ucc neighborhoods of a point than product neighborhoods. That is, the product topology is weaker than the ucc topology.

Ordinarily, weaker topologies produce larger closures since there are fewer conditions for a point to be a point of closure. For an equicontinuous family, however, we now show that the pointwise closure of a set is not just a superset of the ucc closure, it equals the ucc closure. Recall that we are assuming our family $\mathscr{F}$ is a subset of the product $\Pi = \Pi_{x \in X} Y_x$.

Proposition 9.10.2. *If $\mathscr{F}$ is equicontinuous on X, then the product topology for $\Pi = \Pi_{x \in X} Y_x$ when restricted to $\mathscr{F}$ equals the topology of uniform convergence on compact subsets of X.*

Proof. Fix $g \in \mathscr{F}$. Since finite subsets of X are compact, any basic product neighborhood of g restricted to $\mathscr{F}$ is itself a basic ucc neighborhood restricted to $\mathscr{F}$. Fix $\varepsilon > 0$ and a compact subset K of X; let O be the corresponding ε, K-ucc neighborhood of g. That is, $h \in O$ if $\sup_{x \in K} \rho(h(x), g(x)) < \varepsilon$. For each $x \in K$, let W_x be an open neighborhood of x in X such that for every $y \in W_x$ and every $h \in \mathscr{F}$, $\rho(h(y), h(x)) < \varepsilon/3$. Find a finite subcover $W_{x_1}, \cdots, W_{x_n}$ of K. Let V be the product neighborhood of g consisting of all functions h for which $\rho(h(x_i), g(x_i)) < \varepsilon/3$, $1 \le i \le n$. We now show that $V \cap \mathscr{F} \subseteq O \cap \mathscr{F}$. Fix $h \in V \cap \mathscr{F}$ and any $y \in K$. Since $y \in W_i$ for some i,

$$\rho(h(y), g(y)) \le \rho(h(y), h(x_i)) + \rho(h(x_i), g(x_i)) + \rho(g(x_i), g(y)) < \varepsilon.$$

That is, $V \cap \mathscr{F}$ is contained in the ε, K-ucc neighborhood O of g. It follows that the product topology and the ucc topology are the same when restricted to $\mathscr{F}$.

Corollary 9.10.1. *Assume $\mathscr{F}$ is equicontinuous on X, and let $\overline{\mathscr{F}}$ be the product topology closure of $\mathscr{F}$. Fix $A \subseteq \mathscr{F}$. The product topology closure of A in $\overline{\mathscr{F}}$ equals the closure with respect to the topology of uniform convergence on compact sets of X.*

Proof. The result follows from the fact that the product topology and the ucc topology are the same when restricted to $\overline{\mathscr{F}}$ since $\overline{\mathscr{F}}$ is equicontinuous.

Theorem 9.10.1 (Ascoli-Arzelá). *Let $\mathscr{F}$ be an equicontinuous family of functions from a separable Hausdorff space X into a metric space Y. Assume that for each $x \in X$, the values $\{f(x) : f \in \mathscr{F}\}$ are contained in a compact subset $Y_x \subseteq Y$. Then any sequence $\langle f_n : n \in \mathbb{N} \rangle$ in $\mathscr{F}$ has a subsequence that converges uniformly on compact subsets of X to a continuous function g.*

Proof. Let $D = \{x_1, x_2, \cdots\}$ be a countable dense subset of X. Fix $\langle g_n : n \in \mathbb{N} \rangle$ in $\mathscr{F}$. Choose a subsequence $\langle g_m^1 \rangle$ that converges at x_1. Given $\langle g_m^k \rangle$, choose a subsequence $\langle g_m^{k+1} \rangle$ that converges at x_{k+1}. The diagonal sequence $\langle g_m^m : m \in \mathbb{N} \rangle$ converges at all points of D. Let $\langle f_n : n \in \mathbb{N} \rangle$ be that subsequence. We next show that for every $x \in X$, $\langle f_n(x) \rangle$ is a Cauchy sequence, and therefore a convergent sequence. Fix $x \in X$ and $\varepsilon > 0$. Choose a neighborhood W of x such that for all $z \in W$ and all $h \in \mathscr{F}$, $\rho(h(z), h(x)) < \varepsilon/3$. Fix an $x_i \in D \cap W$. There is an $n_0 \in \mathbb{N}$ such that for all $n, m \geq n_0$, $\rho(f_n(x_i), f_m(x_i)) < \varepsilon/3$, whence

$$\rho(f_n(x), f_m(x)) \leq \rho(f_n(x), f_n(x_i)) + \rho(f_n(x_i), f_m(x_i)) + \rho(f_m(x_i), f_m(x)) < \varepsilon.$$

It follows that f_n converges pointwise (and therefore uniformly on compact sets) to a function g in the product topology closure $\overline{\mathscr{F}} \subseteq \Pi_{x \in X} Y_x$. Since $\overline{\mathscr{F}}$ is equicontinuous, g is continuous.

Example 9.10.1. In the context of complex function theory, let $\mathscr{F}$ be the family of holomorphic (i.e., analytic) mappings of the open unit disk $\Delta := \{z \in \mathbb{C} : |z| < 1\}$ into a closed disk of radius M centered at 0. Given any point $z_0 \in \Delta$, there is a circle C_r of radius r with center z_0 contained in Δ. By a formula due to Cauchy, if $f \in \mathscr{F}$, then for any point z_1 with $|z_1 - z_0| < r/2$,

$$f(z_1) = \frac{1}{2\pi i} \oint_{C_r} \frac{f(w)}{w - z_1} dw,$$

where the integral path is in the counterclockwise direction. For any $w \in C_r$, $|w - z_1| \geq \frac{r}{2}$, whence

$$\begin{aligned}
|f(z_1) - f(z_0)| &\leq \left| \frac{1}{2\pi i} \oint_{C_r} \frac{f(w)}{w - z_1} dw - \frac{1}{2\pi i} \oint_{C_r} \frac{f(w)}{w - z_0} dw \right| \\
&\leq \frac{2\pi r}{2\pi} \cdot \max_{w \in C_r} \left| \frac{1}{w - z_1} - \frac{1}{w - z_0} \right| \cdot M \\
&\leq r \cdot \max_{w \in C_r} \left| \frac{z_1 - z_0}{(w - z_1)(w - z_0)} \right| \cdot M \\
&\leq r \cdot \frac{2}{r} \cdot \frac{1}{r} \cdot |z_1 - z_0| \cdot M = \frac{2}{r} \cdot |z_1 - z_0| \cdot M.
\end{aligned}$$

It follows that $\mathscr{F}$ is equicontinuous at z_0. By the Ascoli-Arzelà theorem, any sequence $\langle f_n : n \in \mathbb{N} \rangle$ in $\mathscr{F}$ has a subsequence converging uniformly on compact subsets of Δ to a function f. By a theorem due to Cauchy, the integral of each f_n

around a triangular path contained in Δ is 0, so the same is true of the uniform limit f. By a theorem of Morera, $f \in \mathscr{F}$. That is, a uniformly bounded family of analytic functions on Δ is stable with respect to the operation of taking ucc limits of sequences.

9.11 Stone-Weierstrass Theorem

We want to establish a general result that has as a corollary the following theorem of Weierstrass: Every continuous function on a closed and bounded set X in $\mathbb{R}^n$ can be uniformly approximated on X by polynomials in the variables corresponding to the coordinates. (An example of such a polynomial defined on $\mathbb{R}^2$ is $x^2 + xy + y^5$.)

For the general case, we will work with a compact Hausdorff space $(X, \mathscr{T})$. We let $C(X)$ denote the space of all continuous real-valued functions on X. Now $(X, \mathscr{T})$ is a normal space, so by Urysohn's Lemma 9.8.2, for any pair of distinct points x and y in X, there is an $f \in C(X)$ that takes different values at x and y. We say the family $C(X)$ **separates the points** of X. The space $C(X)$ is not only a vector space with scalars in $\mathbb{R}$, it is an **algebra**. That is, given $f, g \in C(X)$, the pointwise product $f \cdot g$ is also in $C(X)$. The space $C(X)$ is a normed linear space with the norm given by $\|f\| = \max_{x \in X} |f(x)|$. When we speak of the closure of a set in $C(X)$ we will be speaking of the closure with respect to this norm and call it the **uniform closure**. We will show that the uniform closure of any subalgebra that contains the constant functions and separates points of X is all of $C(X)$.

The space $C(X)$ is also a real **vector lattice**. That is, it is a vector space with scalars in $\mathbb{R}$, and if f and g are in $C(X)$, so are $f \wedge g$ (the pointwise minimum) and $f \vee g$ (the pointwise maximum.) Note that $|f| = (f \vee 0) + (-f \vee 0)$. Conversely, given the absolute value function, we have

$$f \vee g = (1/2)[f + g + |f - g|], \qquad f \wedge g = (1/2)[f + g - |f - g|].$$

We show first that the uniform closure of any vector sublattice of $C(X)$ is all of $C(X)$ if it separates points and contains the constant functions.

9.11.1 Vector Lattices

Fix a nonempty vector lattice L of continuous real-valued functions on X. We assume that L separates points and contains the constants.

Lemma 9.11.1. *Given a, $b \in \mathbb{R}$ and x, $y \in X$ with $x \neq y$, there is an $f \in L$ with $f(x) = a$ and $f(y) = b$.*

Proof. Fix $g \in L$ with $g(x) \neq g(y)$. Set

$$f = \frac{a-b}{g(x)-g(y)} \cdot g + \frac{b \cdot g(x) - a \cdot g(y)}{g(x)-g(y)}.$$

Then

$$\begin{aligned} f(x) &= \frac{a-b}{g(x)-g(y)} \cdot g(x) + \frac{b \cdot g(x) - a \cdot g(y)}{g(x)-g(y)} \\ &= \frac{ag(x) - bg(x) + bg(x) - ag(y)}{g(x)-g(y)} = a. \\ f(y) &= \frac{a-b}{g(x)-g(y)} \cdot g(y) + \frac{b \cdot g(x) - a \cdot g(y)}{g(x)-g(y)} \\ &= \frac{ag(y) - bg(y) + bg(x) - ag(y)}{g(x)-g(y)} = b. \end{aligned}$$

Lemma 9.11.2. *Fix $a, b \in \mathbb{R}$ with $a \leq b$. Given a nonempty closed set $F \subset X$ and a point $p \notin F$, there is an $f \in L$ with $a \leq f \leq b$ such that $f(p) = a$ and $f \equiv b$ on F.*

Proof. For each $x \in F$, find an $f_x \in L$ with $f_x(p) = a$ and $f_x(x) = b + 1$. Let O_x be an open neighborhood of x on which $f_x > b$. Take a finite subcover $\{O_{x_i} : i = 1 \cdots k\}$ of this collection of open sets. Let $f = f_{x_1} \vee \ldots \vee f_{x_k}$. Then $f \in L$, $f(p) = a$, and $f|F > b$. Replace f with $a \vee f \wedge b$.

Theorem 9.11.1 (Lattice form of Stone-Weierstrass). *Given any continuous, real-valued function h on X and any $\varepsilon > 0$, there is a function $g \in L$ with $h \leq g \leq h + \varepsilon$ uniformly on X. Therefore, $C(X)$ is the uniform closure of L.*

Proof. Let $M = \max_X h$. Let $L' = \{f \in L : h \leq f\}$. Since the constant function M is in L', L' is a nonempty lattice. Fix $x \in X$ and $\varepsilon > 0$. Choose an open neighborhood O of x so that $h(x) < h(x) + \varepsilon$ on O. There is a function $f \in L$ with $h + \varepsilon \leq f \leq M + 1$ on X, $f(x) = h(x) + \varepsilon$, and $f(y) \equiv M + 1$ on $X \setminus O$. It follows that $f \in L'$. Since ε is arbitrary, $h(x) = \inf_{f \in L'} f(x)$. Moreover, this is true for every $x \in X$.

Now for each $x \in X$, find an open neighborhood O_x of x and a function $g_x \in L'$ such that $g_x(x) < h(x) + \varepsilon/3$, $|g_x(x) - g_x(y)| < \varepsilon/3$, and $|h(x) - h(y)| < \varepsilon/3$ for every $y \in O_x$. Fix a finite subcovering of the O_x's, and note the corresponding points $x_1, \cdots, x_n$. Let $g = g_{x_1} \wedge g_{x_2} \wedge \cdots \wedge g_{x_n}$. Then $g \in L'$. For each $y \in X$, $y \in O_{x_i}$ for some i, so

$$h(y) \leq g(y) \leq g_{x_i}(y) < g_{x_i}(x_i) + \varepsilon/3 < h(x_i) + 2\varepsilon/3 < h(y) + \varepsilon.$$

9.11.2 Algebras

We now show that if $\mathscr{A}$ is a subalgebra of $C(X)$, and $\mathscr{A}$ separates points and contains the constant functions, then the uniform closure $\overline{\mathscr{A}}$ is a uniformly closed sublattice of $C(X)$, and hence equals $C(X)$.

Proposition 9.11.1. *If $\mathscr{A}$ is a subalgebra of $C(X)$, then so is the uniform closure $\overline{\mathscr{A}}$.*

Proof. Exercise 9.34(A).

Lemma 9.11.3. *Given $\varepsilon > 0$, there is a polynomial P in one variable on the real line such that the constant term $P(0) = 0$ and for all $s \in [-1,1]$, $|P(s) - |s|| < \varepsilon$.*

Proof. Let $\delta = \varepsilon/4$. If z is a point in the open disk $|z| < 1+\delta^2$ in the complex plane, then $1+\delta^2 - z$ is a point in the open disk of radius $1+\delta^2$ centered at $1+\delta^2$. We use the branch of the complex square root defined on the complement of the negative real axis that maps 1 to 1. Now $(1+\delta^2-z)^{1/2}$ has a complex power series about 0 that converges on closed disks centered at 0 of radius R for $1 < R < 1+\delta^2$. Therefore, the real-valued series for $(1+\delta^2-t)^{1/2}$ about $t = 0$ converges uniformly on $[-1,1]$ and in particular on $[0,1]$. Choose a partial expansion $Q_N(t) = \Sigma_{n=0}^{N} c_n t^n$ so that for all $t \in [0,1]$, $|(1+\delta^2-t)^{1/2} - Q_N(t)| < \delta$. Replace t with $1-s^2$, $s \in [-1,1]$, and set $P_c(s) := Q_N(1-s^2)$. This is a polynomial in s on the interval $[-1,1]$, and for all $s \in [-1,1]$, $\left|P_c(s) - (s^2+\delta^2)^{1/2}\right| < \delta$.

Using the same branch of the square root, for all $s \in [-1,1]$ we have the inequality $|s|\,(s^2+\delta^2)^{1/2} > s^2$, so

$$[(s^2+\delta^2)^{1/2} - |s|]^2 = 2s^2 + \delta^2 - 2|s|(s^2+\delta^2)^{1/2} < \delta^2,$$

whence $\left|(s^2+\delta^2)^{1/2} - |s|\right| < \delta$. It now follows that for all $s \in [-1,1]$,

$$|P_c(s) - |s|| \le |P_c(s) - (s^2+\delta^2)^{1/2}| + |(s^2+\delta^2)^{1/2} - |s|| < 2\delta.$$

Moreover, the constant term is $P_c(0)$, and $|P_c(0)| < 2\delta$. The desired polynomial is $P := P_c - P_c(0)$.

Proposition 9.11.2. *A uniformly closed subalgebra $\mathscr{A}$ of $C(X)$ is a lattice.*

Proof. Given $f \in \mathscr{A}$ and $\varepsilon > 0$, let P be the polynomial of the Lemma corresponding to ε. Now $f/\|f\| \in \mathscr{A}$, and so $P(f/\|f\|) \in \mathscr{A}$. Moreover,

$$\left| P\left(\frac{f}{\|f\|}\right) - \left(\frac{|f|}{\|f\|}\right) \right| < \varepsilon.$$

Since ε is arbitrary and $\mathscr{A}$ is uniformly closed, $|f|/\|f\| \in \mathscr{A}$, whence $|f| \in \mathscr{A}$. Since $f \vee g = (1/2)[f+g+|f-g|]$, and $f \wedge g = (1/2)[f+g-|f-g|]$, $\mathscr{A}$ is a lattice.

Theorem 9.11.2 (Stone-Weierstrass). *Let $\mathscr{A}$ be an algebra of continuous real-valued functions on X that separates points of X and contains the constant functions. Then $\mathscr{A}$ is dense in $C(X)$. That is, the uniform closure $\overline{\mathscr{A}}$ equals $C(X)$.*

Proof. The uniform closure of $\mathscr{A}$ is an algebra $\overline{\mathscr{A}}$ that is also a lattice; it separates points and contains the constant functions. By Theorem 9.11.1, $\overline{\mathscr{A}} = C(X)$.

Corollary 9.11.1. *Every continuous function on a closed and bounded set X in $\mathbb{R}^n$ can be uniformly approximated on X by a polynomial in the variables corresponding to the coordinates.*

Example 9.11.1. Consider the family of continuous real-valued functions of period 2π on $\mathbb{R}$. We may think of these as functions on a circle $\mathbb{T}$ of circumference 2π, so the domain is compact. The functions $\{\sin\, x, \cos\, x\}$ separate points of $\mathbb{T}$. It follows from Theorem 9.11.2 that the uniform closure of the smallest algebra $\mathscr{A}$ containing the functions $1,\;\sin\, x$, and $\cos\, x$ is all of $C(\mathbb{T})$. Now

$$e^{imx} = \cos(mx) + i\,\sin(mx) = [e^{ix}]^m = [\cos\, x + i\,\sin\, x]^m,$$

and both the real and imaginary parts of $[\cos\, x + i\,\sin\, x]^m$ are in $\mathscr{A}$. Therefore, it follows that finite sums of the form

$$a_0 + \Sigma_{n=1}^{N}(a_n \cos\, nx + b_n \sin\, nx)$$

are also in $\mathscr{A}$. We now only need to show that the collection $\mathscr{B}$ consisting of all such sums forms an algebra. Since $\mathscr{B} \subseteq \mathscr{A}$, we will then know that $\mathscr{B} = \mathscr{A}$. To show that $\mathscr{B}$ is an algebra, we note that we can add two sums in $\mathscr{B}$ and we again have an element of $\mathscr{B}$. We can multiply such a sum by a constant and again have an element of $\mathscr{B}$. The product of two sums in $\mathscr{B}$ is a sum with terms of the form

$$\alpha(\cos\, nx)(\cos\, mx) \qquad \beta(\cos\, nx)(\sin\, mx) \qquad \gamma(\sin\, nx)(\sin\, mx).$$

Since

$$(\cos\, nx)(\cos\, mx) = (1/2)[\cos(n+m)x + \cos(n-m)x]$$
$$(\cos\, nx)(\sin\, mx) = (1/2)[\sin(n+m)x - \sin(n-m)x]$$
$$(\sin\, nx)(\sin\, mx) = (1/2)[\cos(n-m)x - \cos(n+m)x],$$

and $\cos(-nx) = \cos(nx)$, while $\sin(-nx) = -\sin(nx)$, all of these terms are in $\mathscr{B}$. Therefore, $\mathscr{B}$ is an algebra in $C(\mathbb{T})$ that contains the constants and separates points, so the uniform closure of $\mathscr{B}$ is $C(\mathbb{T})$.

Example 9.11.2. Let $X \times Y$ be a product of compact Hausdorff spaces. By Theorem 9.9.2, such a product is compact. Let $\mathscr{F}$ consist of functions of the form $g(x) \cdot h(y)$, where g is a continuous function (perhaps constant) on X and h is a continuous function (perhaps constant) on Y. The functions in $\mathscr{F}$ separate the points of $X \times Y$, so the polynomials in finite numbers of elements of $\mathscr{F}$ are uniformly dense in $C(X \times Y)$. A typical such polynomial has the form $\sum_{i=1}^{n} g_i(x)h_i(y)$.

9.12 Problems

Problem 9.1. a) Let V be a vector space of continuous functions on the complex plane $\mathbb{C}$; the scalar field is either $\mathbb{R}$ or $\mathbb{C}$. For each $f \in V$, let $\mathscr{B}_f$ be given by a nonempty compact set $K \subset \mathbb{C}$ and an $\varepsilon > 0$. The collection

$$\mathscr{B}_f := \left\{ g \in V : \max_{z \in K} |g(z) - f(z)| < \varepsilon \right\}.$$

Show that $\mathscr{B}_f$ is an open base at f. (The assignment $f \mapsto \mathscr{B}_f$ generates the topology of uniform convergence on compact sets on V.)

b) What is the relationship between the open base $f \mapsto \mathscr{B}_f$ described in Part a, and the base described in Example 9.1.1?

Problem 9.2. a) Fix $n \in \mathbb{N}$. A closed dyadic cube in $\mathbb{R}^n$ is a set of the form

$$\left[\frac{a_1}{2^m}, \frac{a_1+1}{2^m}\right] \times \cdots \times \left[\frac{a_n}{2^m}, \frac{a_n+1}{2^m}\right]$$

for integers $m, a_1, \ldots, a_n$. Given a point $x \in \mathbb{R}^n$, show that the collection of closed dyadic cubes in $\mathbb{R}^n$ containing x forms a local filter base at x.

b) Prove that

$$\mathscr{T}_1 := \{\bigcup_{i=1}^{\infty} Q_i : Q_i = \varnothing \text{ or } Q_i \text{ is a closed dyadic cube in } \mathbb{R}^n \text{ for each } i\}.$$

forms a topology on $\mathbb{R}^n$.

c) Is it true that

$$\mathscr{T}_2 := \{\text{finite (possibly empty) unions of closed dyadic cubes}\}$$

forms a topology on $\mathbb{R}^n$?

Problem 9.3. Show that for a separable metric space (X,d), the open balls of radius $1/n$, $n \in \mathbb{N}$, centered at points of a dense subset of X form a base for the topology.

Problem 9.4. Prove Proposition 9.2.3.

Problem 9.5. Prove Proposition 9.2.4.

Problem 9.6. Prove Theorem 9.2.1.

Problem 9.7. Prove Proposition 9.2.6.

Problem 9.8. Prove Corollary 9.5.3.

Problem 9.9 (A). Let Y be the set of ordinal numbers strictly smaller than the first uncountable ordinal Ω. Let S be a countable subset of Y. For example, S may be the range of a sequence in Y. Show that S has an upper bound that is in the set Y. It then follows, since Y is well-ordered, that there is a least upper bound of S in Y. That upper bound is a point of closure of S.

Problem 9.10 (A). Prove Theorem 9.6.2.

Problem 9.11. Suppose $(X, \mathscr{T})$ is a topological space, and $\{A_\gamma : \gamma \in \mathscr{I}\}$ is an indexed family of connected subsets of X such that the index set $\mathscr{I}$ is well-ordered, and for each $\beta \in \mathscr{I}$ and some $\alpha < \beta$ in $\mathscr{I}$, $A_\alpha \cap A_\beta \neq \varnothing$. Show that the union $\cup_{\gamma \in I} A_\gamma$ is connected.

Problem 9.12. A **locally pathwise connected** space is one, such as $\mathbb{R}^n$, having an open base $x \mapsto \mathscr{B}_x$ at each point consisting of pathwise connected sets. Show that if O is an open, connected, and locally pathwise connected subset of a topological space, then O is pathwise connected.

Problem 9.13 (A). Show that the set $\overline{A}$ in Example 9.6.2 is not pathwise connected.

Problem 9.14. Show that total boundedness is preserved by any uniformly continuous map from one metric space to another.

Problem 9.15. **a)** What is the collection of all continuous real-valued functions on a nonempty set X supplied with the trivial topology?
b) What is the collection of all continuous real-valued functions on a nonempty set X supplied with the discrete topology?

Problem 9.16. Show that if $(X, \mathscr{T})$ is a compact Hausdorff space, then with any strictly stronger topology, the space is not compact.

Problem 9.17. Let (X, T) be a Hausdorff space, and let p be a point in X for which there is a sequence $\langle O_n : n \in \mathbb{N} \rangle$ of open sets with $p \in O_{n+1} \subseteq O_n$ for each n, and $\cap_n O_n = \{p\}$. Let $\mathscr{L}_p$ be a local filter base for p. That is, if U is open and $p \in U$, then for some $S \in \mathscr{L}_p$, $p \in S \subseteq U$. Therefore, for each $n \in \mathbb{N}$, there is an $S_n \in \mathscr{L}_p$ with $p \in S_n \subseteq O_n$. Show that each S_n contains an open set. **Hint:** Suppose S_1 contains no open set. We may assume $S_n \subseteq S_1$ for each n. Therefore, for each $n \in \mathbb{N}$, there is a point $x_n \in O_n \diagdown S_1$. Let $A = \{x_n : n \in \mathbb{N}\}$. Show that $X \diagdown A$ is open. Show that A is not closed.

Problem 9.18. Give a proof that a compact subset of a metric space (X, d) is closed by showing that if $x \in \overline{A} \diagdown A$, then there is an open cover of A with no finite subcover.

Problem 9.19. Show that any continuous, one-to-one, function f with compact domain K in a Hausdorff space $(X, \mathscr{T})$ and range $f[K]$ in a Hausdorff space $(Y, \mathscr{S})$ is a homeomorphism. That is, the inverse function $f^{-1} : f[K] \mapsto K$ is continuous.

Problem 9.20. Show that the set of ordinal numbers less than or equal to the first uncountable ordinal Ω supplied with the order topology is compact. See Example 9.4.1 and Problem 9.9.

Problem 9.21. Given a collection $\mathscr{S}$ of subsets of a nonempty set X, show that the intersection of all topologies (including the discrete topology) containing $\mathscr{S}$ is a topology on X.

Problem 9.22. Show that a topological space for which every singleton set is closed is a regular space if and only if for any point $x \in X$ and any open set U with $x \in U$, there is an open set V with $x \in V \subseteq \overline{V} \subseteq U$.

Problem 9.23. Show that the function f constructed in the proof of Theorem 9.8.2 is continuous with $f : X \mapsto [0,1]$, and f restricted to $A \equiv 0$ and f restricted to $B \equiv 1$.

Problem 9.24. Show that any topology on a countable set X has the Lindelöf property (Definition 9.8.3.)

Problem 9.25. Prove Proposition 9.9.2.

Problem 9.26. Show that a net $\langle h_\gamma : \gamma \in D \rangle$ converges in a product space $\Pi_{\alpha \in \mathscr{I}} X_\alpha$ if and only if for each index α, the net $\langle h_\gamma(\alpha) : \gamma \in D \rangle$ converges in X_α.

Problem 9.27. Prove Theorem 9.9.1 for a finite product of closed and bounded intervals.

Problem 9.28. Let D be a countable set, and for each $x \in D$, let Y_x be a compact metric space. Show that any sequence f_n in the product space $\Pi_{x \in D} Y_x$ has a subsequence that converges in the product topology.

Problem 9.29. Let $(X, \mathscr{T})$ be dense subspace of a compact Hausdorff space $(Y, \mathscr{S})$. That is, $\mathscr{T}$ is the relative S-topology on X. Show that any $\mathscr{T}$-compact subset of X is $\mathscr{S}$-compact and $\mathscr{S}$-closed in Y.

Problem 9.30 (A). Prove Theorem 9.9.6.

Problem 9.31. Show that the function h defined in the proof of Proposition 9.9.4 is continuous on X.

Problem 9.32 (A). Prove Proposition 9.10.1.

Problem 9.33. Show that the ucc-local filter base is an open base for each member g in a family of functions from a Hausdorff space to a metric space.

Problem 9.34 (A). Prove Proposition 9.11.1.

Problem 9.35. Show that polynomials form a dense subset of the continuous real-valued functions on $[0,1]$ with the relative topology generated by the L^∞-norm.

Problem 9.36 (A). Let $\mathscr{A}$ be an algebra of continuous real-valued functions on a compact Hausdorff space X, and assume that $\mathscr{A}$ separates points of X. Do **not** assume that $\mathscr{A}$ contains constant functions. Show that the uniform closure $\overline{\mathscr{A}}$ of $\mathscr{A}$ is either $C(X)$ or $\{f \in C(X) : f(p) = 0\}$ for a unique point $p \in X$.

Problem 9.37. Show that convergence Lebesgue almost everywhere on the unit interval $[0,1]$ is not defined in terms of a topology.

Problem 9.38. Let X be the Euclidean plane with the topology generated at each point except the origin by the usual metric balls centered at the point. Let the local filter base $\mathscr{L}_0$ at the origin consists of sets S_n, where for each $n \in \mathbb{N}$, S_n is the union of the open interval $(-1/n, 1/n)$ on the x-axis and the open interval $(-1/n, 1/n)$ on the y-axis. How does this example relate to Proposition 9.9.1? Explain.

Chapter 10
Measure Construction

10.1 Measures from Outer Measures

Like measures on the real line, one can construct a general measure from an outer measure. We do not, however, construct an outer measure from an integrator; instead, we assume that we start with an outer measure that has the necessary properties.

Recall that an algebra in a set X is a collection of subsets of X; the collection contains X itself and is stable with respect to the operations of taking complements and finite unions. It is therefore stable with respect to the operation of taking finite intersections. A σ-algebra in X is an algebra that is stable with respect to the operation of taking countable unions, and therefore, countable intersections. As before, $\widetilde{E}$ and $\complement E$ both denote the complement of a set E. A measure is complete if subsets of sets of measure 0 are measurable. Also recall that the power set of X is the collection of all subsets of X.

Definition 10.1.1. An **outer measure** μ^* on a set X is a nonnegative, extended-real valued set function defined on the power set of X such that

i) $\mu^*(\varnothing) = 0$,
ii) μ^* is monotone increasing, i.e., the bigger the set, the bigger the value of μ^*, and
iii) μ^* is countably subadditive, i.e., $\mu^*(\cup_{i\in\mathbb{N}} A_i) \leq \Sigma_{i\in\mathbb{N}}\, \mu^*(A_i)$.

Definition 10.1.2 (Carathéodory). A set E is called **measurable** with respect to an outer measure μ^* on a set X if for each $A \subseteq X$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap \widetilde{E}).$$

Remark 10.1.1. As for the real line, a set $E \subseteq X$ is measurable if for those sets $A \subseteq X$ with $\mu^*(A) < +\infty$, $\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap \widetilde{E})$.

Theorem 10.1.1. *The class $\mathscr{B}$ of μ^*-measurable sets is a σ-algebra, and μ^* restricted to $\mathscr{B}$ is a complete measure.*

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_10

Proof. Clearly, $\varnothing \in \mathscr{B}$, and if $E \in \mathscr{B}$, so is $\widetilde{E}$. Fix E_1 and E_2 in $\mathscr{B}$ and $A \subseteq X$ with $\mu^*(A) < +\infty$. Then by subadditivity,

$$\begin{aligned}\mu^*(A) &= \mu^*(A \cap E_2) + \mu^*(A \cap \widetilde{E}_2)\\ &= \mu^*(A \cap E_2) + \mu^*(A \cap \widetilde{E}_2 \cap E_1) + \mu^*(A \cap \widetilde{E}_2 \cap \widetilde{E}_1)\\ &\geq \mu^*(A \cap [E_1 \cup E_2]) + \mu^*(A \cap \widetilde{E}_2 \cap \widetilde{E}_1)\\ &= \mu^*(A \cap [E_1 \cup E_2]) + \mu^*(A \cap \complement[E_1 \cup E_2]).\end{aligned}$$

It follows that $\mathscr{B}$ is an algebra. Moreover, if $E_1 \cap E_2 = \varnothing$, then $[E_1 \cup E_2] \cap E_1 = E_1$ and $[E_1 \cup E_2] \cap \widetilde{E}_1 = E_2$, so

$$\begin{aligned}\mu^*(E_1 \cup E_2) &= \mu^*([E_1 \cup E_2] \cap E_1) + \mu^*\left([E_1 \cup E_2] \cap \widetilde{E}_1\right)\\ &= \mu^*(E_1) + \mu^*(E_2).\end{aligned}$$

This shows that, μ^* restricted to $\mathscr{B}$ is finitely additive.

Now let $E = \cup_{i=1}^{\infty} E_i$ be a countable union of sets in $\mathscr{B}$. Since $\mathscr{B}$ is an algebra, we may assume that the sets E_i are pairwise disjoint. We now show that $E \in \mathscr{B}$ and $\mu^*(E) = \Sigma_{i=1}^{\infty} \mu^*(E_i)$. For each $n \in \mathbb{N}$, let $G_n = \cup_{i=1}^{n} E_i$, and let A be a subset of X. Then

$$\begin{aligned}\mu^*(A) &= \mu^*(A \cap G_n) + \mu^*(A \cap \widetilde{G}_n) \geq \mu^*(A \cap G_n) + \mu^*(A \cap \widetilde{E})\\ &= \mu^*(A \cap E_1) + \mu^*(A \cap (G_n \setminus E_1)) + \mu^*(A \cap \widetilde{E}) = \cdots\\ &= \Sigma_{i=1}^{n} \mu^*(A \cap E_i) + \mu^*(A \cap \widetilde{E}).\end{aligned}$$

Therefore, by subadditivity,

$$\mu^*(A) \geq \Sigma_{i=1}^{\infty} \mu^*(A \cap E_i) + \mu^*(A \cap \widetilde{E}) \geq \mu^*(A \cap E) + \mu^*(A \cap \widetilde{E}),$$

It follows that E is measurable, whence $\mathscr{B}$ is a σ-algebra. Replacing A with E and using subadditivity, we have $\mu^*(E) = \Sigma_{i=1}^{\infty} \mu^*(E_i)$. That is, μ^* restricted to $\mathscr{B}$ is σ-additive. If $S \subseteq X$ and $\mu^*(S) = 0$, then it is immediate that $S \in \mathscr{B}$, so μ^* is a complete measure when restricted to $\mathscr{B}$.

10.2 The Carathéodory Extension Theorem

In this section, we show that measures on algebras of sets (Definition 10.2.1) can be extended to measures on σ-algebras. In general, topologies are not part of the structure of an algebra of sets. Recall that the intersection of all σ-algebras containing a collection $\mathscr{C}$ of subsets of X is the smallest σ-algebra in X containing $\mathscr{C}$; it is denoted by $\sigma(\mathscr{C})$.

Definition 10.2.1. A nonnegative, extended-real valued set function μ on an algebra $\mathscr{A}$ is called **a measure on** $\mathscr{A}$ if $\mu(\varnothing) = 0$ and for any pairwise disjoint sequence $\langle A_n : n \in \mathbb{N}\rangle$ in $\mathscr{A}$ for which the union is also in $\mathscr{A}$, we have $\mu(\cup_{n\in\mathbb{N}} A_n) = \Sigma_{n\in\mathbb{N}}\mu(A_n)$. We say μ is **σ-additive** on $\mathscr{A}$.

Remark 10.2.1. Clearly, σ-additivity is necessary for μ to be extendable to a σ-additive set function on a larger collection. Using an outer measure, we now show that it is also a sufficient condition. The proof is presented in several steps.

Theorem 10.2.1 (Carathéodory Extension Theorem). *If μ is a measure on an algebra $\mathscr{A}$ of subsets of a set X, then μ has a σ-additive extension to the completion of the smallest σ-algebra $\sigma(\mathscr{A})$ containing $\mathscr{A}$. If X is a finite or countably infinite union of sets from $\mathscr{A}$ having finite μ-measure, then the extension of μ from $\mathscr{A}$ to $\sigma(\mathscr{A})$ is unique.*

Definition 10.2.2. Given a measure μ on $\mathscr{A}$, for each set $E \subseteq X$, let $\mathscr{C}(E)$ denote the family of all sequences in $\mathscr{A}$ that cover E. That is, a sequence $\langle A_n : n \in \mathbb{N}\rangle$ is a member of $\mathscr{C}(E)$ if and only if for each $n \in \mathbb{N}$, $A_n \in \mathscr{A}$, and $\cup_{n\in\mathbb{N}} A_n \supseteq E$. We set $\mu^*(E) = \inf_{\langle A_n\rangle \in \mathscr{C}(E)} \left(\Sigma_{n=1}^{\infty}\mu(A_n)\right)$.

Proposition 10.2.1. *The set function μ^* is an outer measure. If $E \in \mathscr{A}$, then $\mu E = \mu^* E$.*

Proof. First we fix $E \in \mathscr{A}$ and show that $\mu(E) = \mu^*(E)$. The sequence $\langle A_n\rangle$ where $A_1 = E$ and $A_n = \varnothing$ for $n > 1$ covers E, so $\mu^*(E) \le \mu(E)$. On the other hand, if $\langle A_n\rangle$ is a sequence from $\mathscr{A}$ that covers E, then we may replace A_1 with $B_1 = E \cap A_1$, and for $n > 1$, replace A_n with $B_n = (E \cap A_n) \setminus \bigcup_{i=1}^{n-1} B_i$, which is in $\mathscr{A}$. Moreover, $\mu(B_n) \le \mu(A_n)$ and $E = \bigcup_{n=1}^{\infty} B_n$, so by the σ-additivity of μ on $\mathscr{A}$,

$$\mu(E) = \sum_{n=1}^{\infty} \mu(B_n) \le \sum_{n=1}^{\infty} \mu(A_n).$$

It follows that $\mu^*(E) = \mu(E)$.

Clearly, μ^* is a nonnegative, monotone increasing set function defined on all subsets of X. Since $\varnothing \in \mathscr{A}$, $\mu^*(\varnothing) = \mu(\varnothing) = 0$. We need to show that μ^* is countably subadditive; that is, if $E \subseteq \bigcup_{n=1}^{\infty} E_n$, then $\mu^*(E) \le \sum_{n=1}^{\infty} \mu^*(E_n)$. This is clear if any of the terms in the sum is infinite. Assuming each is finite, we fix $\varepsilon > 0$, and for each $n \in \mathbb{N}$, we choose a sequence $\langle A_i^n : i \in \mathbb{N}\rangle$ in $\mathscr{A}$ that covers E_n such that $\sum_{i=1}^{\infty} \mu(A_i^n) \le \mu^*(E_n) + \frac{\varepsilon}{2^n}$. Since the countable collection $\{A_i^n : i, n \in \mathbb{N}\}$ covers E, we have

$$\mu^*(E) \le \sum_{n=1}^{\infty}\sum_{i=1}^{\infty} \mu(A_i^n) \le \left(\sum_{n=1}^{\infty} \mu^*(E_n)\right) + \varepsilon.$$

Since ε is an arbitrary positive number, subadditivity of μ^* is established.

Proposition 10.2.2. *Each $E \in \mathscr{A}$ is μ^*-measurable.*

Proof. Fix $E \in \mathscr{A}$. Fix a set $S \subseteq X$ of finite outer measure, and fix an $\varepsilon > 0$. Choose a sequence $\langle F_n : n \in \mathbb{N} \rangle$ from $\mathscr{A}$ that covers S with total measure less than $\mu^*(S) + \varepsilon$. Since E is in $\mathscr{A}$, for each $n \in \mathbb{N}$, $\mu(F_n) = \mu(F_n \cap E) + \mu(F_n \cap \widetilde{E})$. Therefore,

$$\begin{aligned}\mu^*(S) + \varepsilon > \sum_{n=1}^{\infty} \mu(F_n) &= \sum_{n=1}^{\infty} \mu(F_n \cap E) + \sum_{n=1}^{\infty} \mu(F_n \cap \widetilde{E}) \\ &\geq \mu^*(S \cap E) + \mu^*(S \cap \widetilde{E}).\end{aligned}$$

Since ε is an arbitrary positive number, it follows that E is μ^*-measurable.

We write μ for the restriction of μ^* to the σ-algebra $\mathscr{B}$ consisting of μ^*-measurable sets. This gives the desired extension of μ. We call a set $E \in \mathscr{B}$ with $\mu(E) = 0$ a **null set**. We will write $\mathscr{A}_\sigma$ for the family of countable unions of sets in $\mathscr{A}$, and $\mathscr{A}_{\sigma\delta}$ for countable intersections of sets in the family $\mathscr{A}_\sigma$. We have now established the following result: The class $\mathscr{B}$ of μ^*-measurable sets is a σ-algebra containing $\mathscr{A}$, and μ^* restricted to $\mathscr{B}$ is a complete measure.

Proposition 10.2.3. *The outer measure μ^* derived from a measure μ on an algebra $\mathscr{A}$ has the property that for any set $E \subseteq X$ and any $\varepsilon > 0$, there is an $\mathscr{A}_\sigma$-set $A \supseteq E$ such that $\mu(A) \leq \mu^*(E) + \varepsilon$, and there is an $\mathscr{A}_{\sigma\delta}$-set $B \supseteq E$ with $\mu(B) = \mu^*(E)$.*

Proof. If $\mu^*(E) = +\infty$, let $A = B = X$. Otherwise, let A be the union of sets from $\mathscr{A}$ that cover E with total measure less than $\mu^*(E) + \varepsilon$, and then let B be the intersection of such $\mathscr{A}_\sigma$-sets for each value $\varepsilon = 1/n$.

Corollary 10.2.1. *If $E \in \mathscr{B}$ has finite measure, then E equals an $\mathscr{A}_{\sigma\delta}$-set from which a null set in $\mathscr{B}$ has been removed.*

Corollary 10.2.2. *The extension of a measure μ on an algebra $\mathscr{A}$ to $\sigma(\mathscr{A})$ is unique if X is the finite or countably infinite union of sets in $\mathscr{A}$ of finite measure.*

Proof. Assume first that $\mu(X) < +\infty$. Any $\mathscr{A}_\sigma$-set that is not in $\mathscr{A}$ is the union of a pairwise disjoint countably infinite sequence of sets in $\mathscr{A}$, so the extension of μ from $\mathscr{A}$ to $\mathscr{A}_\sigma$ is unique. Let $\widetilde{\mu}$ be any other measure extending μ from $\mathscr{A}$ to $\sigma(\mathscr{A})$. We must have $\widetilde{\mu} = \mu$ on $\mathscr{A}_\sigma$. Fix A in $\sigma(\mathscr{A})$ and $\varepsilon > 0$. There is a $B \in \mathscr{A}_\sigma$ with $B \supseteq A$ and $\mu(B) \leq \mu(A) + \varepsilon$, whence $\widetilde{\mu}(A) \leq \widetilde{\mu}(B) = \mu(B) \leq \mu(A) + \varepsilon$. Since this is true for any $\varepsilon > 0$, $\widetilde{\mu}(A) \leq \mu(A)$. Since A is arbitrary in $\sigma(\mathscr{A})$, we also have $\widetilde{\mu}(B \setminus A) \leq \mu(B \setminus A) < \varepsilon$. On the other hand,

$$\mu(A) \leq \mu(B) = \widetilde{\mu}(B) = \widetilde{\mu}(A) + \widetilde{\mu}(B \setminus A) < \widetilde{\mu}(A) + \varepsilon.$$

Since ε is an arbitrary positive number, $\mu(A) \leq \widetilde{\mu}(A)$, whence $\mu(A) = \widetilde{\mu}(A)$. That is, $\widetilde{\mu} = \mu$ on $\sigma(\mathscr{A})$.

Now assume that $X = \cup_{i=1}^{\infty} X_i$, where each X_i is in $\mathscr{A}$ and has finite μ-measure. We may also assume the sets X_i are pairwise disjoint. In this case, if $\widetilde{\mu}$ is another measure extending μ to $\sigma(\mathscr{A})$, then we have shown that for each $i \in \mathbb{N}$ and $B \in \sigma(\mathscr{A})$, $\widetilde{\mu}(B \cap X_i) = \mu(B \cap X_i)$. It follows that $\widetilde{\mu} = \mu$ on $\sigma(\mathscr{A})$.

Example 10.2.1. Let X consist of the rational numbers in $(0,1]$, and let $\mathscr{A}$ be the algebra formed by disjoint unions of intervals $(a,b] \cap X$. Let $\mu(\varnothing) = 0$, and for all other sets $E \in \mathscr{A}$, let $\mu(E) = +\infty$. The extension of μ to $\sigma(\mathscr{A})$ is not unique (Problem 10.1).

10.3 Lebesgue Measure on Euclidean Space

Lebesgue measure on Euclidean space generalizes area for the plane and volume for 3-space. Just as for the line, it can be generated by an outer measure. First we need some basic properties of an n-dimensional Euclidean space $\mathbb{R}^n$. We assume that a coordinate system has been fixed. We use $|x-y|$ to denote the Euclidean distance from a point x to a point y in $\mathbb{R}^n$. This is a metric on $\mathbb{R}^n$, and so metric notions apply. The topology on $\mathbb{R}^n$ is the metric topology with a local open base at each point consisting of open balls centered at the point.

A point $x \in \mathbb{R}^n$ is called **rational** if its coordinates are rational. Every open set $O \subseteq \mathbb{R}^n$ equals the union of the open balls contained in O having rational centers and rational radii. Therefore, the topology satisfies the second axiom of countability, and by Proposition 9.8.1, any collection of open sets has a finite or countably infinite subcollection with the same union. Therefore, to show that a subset is compact, it is enough to show that every countably infinite covering by open sets has a finite subcovering. By Theorem 9.7.5, a set is compact if and only if it has the Bolzano-Weierstrass property. By the Heine-Borel Theorem 9.7.7, a set is compact if and only if it is closed and bounded.

If X is the x-y plane, then the product of finite closed intervals, one in each of the coordinate axes, is a finite rectangle with its boundary. The product of finite open intervals in the coordinate axes is a finite rectangle without its boundary. In 3-space the analogous products form rectangular parallelepipeds with and without a boundary. In general, we have the following property for such a product; the proof is Exercise 10.2.

Proposition 10.3.1. *If P is the product of finite open intervals, one in each of the coordinate axes of $\mathbb{R}^n$, then P is an open set. If P is the product of finite closed intervals in the coordinate axes of $\mathbb{R}^n$, then P is a closed and bounded, and therefore compact, set.*

We next define Lebesgue outer measure λ_n^* on n-dimensional Euclidean space $\mathbb{R}^n$ and the corresponding Lebesgue measure. Given a product P of finite open intervals or finite closed intervals in the coordinate axes, we let $V(P)$ denote the product of the length of those intervals. If the intervals are open, we will call P a **general open rectangle**, and $V(P)$ the **volume** of P. For $n = 2$, $V(P)$ is the area of a rectangle without boundary. For $n = 3$, $V(P)$ is the volume of a rectangular parallelepiped without boundary.

Given a set $A \subseteq X$, we let $\mathscr{C}(A)$ denote the family of all collections of such general open rectangles that cover A. That is, $\mathscr{I}$ is a member of $\mathscr{C}(A)$ if and only if $\mathscr{I}$ is a set of general open rectangles in X, and the union of the elements of $\mathscr{I}$ contains the set A. By $\sum_{P\in\mathscr{I}} V(P)$ we mean the unordered sum of the volumes of the sets in $\mathscr{I}$. Recall that this is the supremum of the sums obtained by adding the volumes of finite subsets of $\mathscr{I}$. If $\mathscr{I}$ is an uncountable collection, then by the Lindelöf property, a finite or countably infinite subfamily of $\mathscr{I}$ also covers A and has a sum of volumes that is no greater than the sum for the whole family. Therefore, in applying the following definition, we usually consider just finite and countably infinite families of general open rectangles that cover A. Every enumeration of a countably infinite family of general open rectangles will produce the same sum of volumes, which is the usual limit of partial sums.

Definition 10.3.1 (Lebesgue outer measure). For each subset $A \subseteq \mathbb{R}^n$, the Lebesgue outer measure, $\lambda_n^*(A)$, is obtained as follows:

$$\lambda_n^*(A) = \inf_{\mathscr{I}\in\mathscr{C}(A)} \left(\sum_{P\in\mathscr{I}} V(P) \right).$$

Proposition 10.3.2. *For a general open rectangle P, the outer measure $\lambda_n^*(P) = V(P)$. For each $A \subseteq \mathbb{R}^n$, $\lambda_n^*(A) \geq 0$, $\lambda_n^*(\varnothing) = 0$, $\lambda_n^*(\mathbb{R}^n) = +\infty$, and if $A \subseteq B \subseteq \mathbb{R}^n$, then $\lambda_n^*(A) \leq \lambda_n^*(B)$. Moreover, λ_n^* is **countably subadditive**, that is, for any sequence $\langle E_i : i \in \mathbb{N}\rangle$ in $\mathbb{R}^n$, $\lambda_n^*(\cup_{i\in\mathbb{N}} E_i) \leq \Sigma_{i\in\mathbb{N}}\, \lambda_n^*(E_i)$. Therefore, λ_n^* is indeed an outer measure.*

Proof. Since P covers itself, $\lambda_n^*(P) \leq V(P)$. Let $\overline{R}$ be any closed general rectangle with $\overline{R} \subset P$. We show next that $V(\overline{R}) \leq \lambda_n^*(\overline{R})$. It will then follow that $V(P) \leq \lambda_n^*(P)$, whence $V(P) = \lambda_n^*(P)$. Since $\overline{R}$ is compact, we need only show that if $\{P_i : i = 1, \cdots, m\}$ is a finite covering of $\overline{R}$ by general open rectangles, then $V(\overline{R}) \leq \sum_{i=1}^m V(P_i)$. That proof is Exercise 10.4(A). The rest is clear except for countable subadditivity, that is, if $E \subseteq \bigcup_{i=1}^\infty E_i$, then $\lambda_n^*(E) \leq \sum_{i=1}^\infty \lambda_n^*(E_i)$. This is clear if any of the terms in the sum is infinite. Assuming each is finite, we fix $\varepsilon > 0$, and for each $i \in \mathbb{N}$, we choose a covering of E_i by general open rectangles such that the sum of their volumes is less than $\lambda_n^*(E_i) + \frac{\varepsilon}{2^i}$. The total volume of all of the covering general open rectangles, which together cover E, is less than $(\sum_{i=1}^\infty \lambda_n^*(E_i)) + \varepsilon$. Since ε is arbitrary, subadditivity is established.

Proposition 10.3.3. *Given $A \subseteq \mathbb{R}^n$ and $\varepsilon > 0$, there is an open set O with $A \subseteq O$ and $\lambda_n^*(O) \leq \lambda_n^*(A) + \varepsilon$. Moreover, there is a countable intersection of open sets $S \supseteq A$ with $\lambda_n^*(S) = \lambda_n^*(A)$.*

Proof. Exercise 10.5.

Definition 10.3.2. To obtain Lebesgue measure on $\mathbb{R}^n$, restrict λ_n^* to the σ-algebra of λ_n^*-measurable sets. The restriction is a complete, countable additive measure extending the volume of general rectangles. The family of λ_n^*-measurable sets is called the **Lebesgue measurable sets**, and the restriction of λ_n^* is called **Lebesgue measure on** $\mathbb{R}^n$.

10.4 Product Measures

We have extended Lebesgue measure to $\mathbb{R}^n$, but we have not yet discussed integration, in particular, iterated integrals. For that discussion, given in a general context, we consider just two complete measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$. The extension to 3 or more spaces will be clear.

Definition 10.4.1. A **measurable rectangle** in $X \times Y$ is the product $A \times B$ of a measurable set A in X with a measurable set B in Y. The collection of measurable rectangles is denoted by $\mathscr{R}$. For each $A \times B \in \mathscr{R}$, we set $\gamma(A \times B) = \mu(A) \cdot \nu(B)$. We let $\mathscr{Q}$ be the collection of finite unions of disjoint measurable rectangles, and we extend γ to $\mathscr{Q}$ by summing its values on the rectangles forming any member of $\mathscr{Q}$.

Proposition 10.4.1. *The family $\mathscr{Q}$ is an algebra of subsets of $X \times Y$, and γ is well-defined on $\mathscr{Q}$. In fact, γ is a measure on $\mathscr{Q}$.*

Proof. The proof that $\mathscr{Q}$ is an algebra of subsets of $X \times Y$ is Exercise 10.10. To show that γ is well-defined and even a measure on $\mathscr{Q}$, we fix a finite or countably infinite disjoint collection $\{(A_i \times B_i)\}$ of measurable rectangles with union a measurable rectangle $A \times B$, and we show that $\gamma(A \times B) = \Sigma\, \gamma(A_i \times B_i)$. Fix $x \in A$. For all $y \in B$, there is a unique i with $(x, y) \in A_i \times B_i$. Therefore, B is the disjoint union of sets B_i such that x is in the corresponding A_i. For those sets B_i, we have

$$\Sigma_i \nu(B_i) = \Sigma_i (\nu(B_i) \cdot \chi_{A_i}(x)) = \nu(B).$$

Therefore, for $x \in A$, $\Sigma_i (\nu(B_i) \cdot \chi_{A_i}(x)) = \nu(B) \cdot \chi_A(x)$, and if $x \notin A$, the equation still holds. Now by the Monotone Convergence Theorem,

$$\begin{aligned} \Sigma \gamma(A_i \times B_i) &= \Sigma_i \int (\nu(B_i) \cdot \chi_{A_i})\, d\mu = \int \Sigma_i (\nu(B_i) \cdot \chi_{A_i}(x))\, d\mu \\ &= \int (\nu(B) \cdot \chi_A)\, d\mu = \nu(B) \cdot \mu(A) = \gamma(A \times B). \end{aligned}$$

Since γ is a measure on the algebra $\mathscr{Q}$, by Theorem 10.2.1 there is an extension to a σ-additive measure on the completion of the σ-algebra $\sigma(\mathscr{Q})$.

Definition 10.4.2. We denote the collection of measurable sets in the completion of $\sigma(\mathscr{Q})$ by $\mathscr{S}$, and we write $\mu \times \nu$ for the extension of γ to $\mathscr{S}$. The measure $\mu \times \nu$ is called a **product measure**.

Proposition 10.4.2. *If μ and ν are both finite (or σ-finite), then $\mu \times \nu$ is finite (or σ-finite).*

Proof. Exercise 10.11.

Example 10.4.1. As an application of the construction of product measures, one can construct two-dimensional Lebesgue measure and then higher-dimensional Lebesgue measure by starting with Lebesgue measure on the real line.

We will need cross sections to describe the structure of product measurable sets. We use the following notation: $E_x := \{y \in Y : (x,y) \in E\}$, $E^y := \{x \in X : (x,y) \in E\}$. Recall that $\widetilde{E}$ and $\complement E$ both denote the complement of a set E.

Proposition 10.4.3. *Given a set $E \subseteq X \times Y$, the value $\chi_{E_x}(y) = \chi_E(x,y)$, $\left(\complement E\right)_x = \complement\,(E_x)$, and for any collection of sets $\{E_\alpha\}$, $(\cup E_\alpha)_x = \cup_\alpha (E_\alpha)_x$.*

Proof. To see that $\left(\complement E\right)_x = \complement\,(E_x)$, note the following:

$$y \in \left(\complement E\right)_x \Leftrightarrow (x,y) \in \complement E \Leftrightarrow (x,y) \notin E \Leftrightarrow y \notin E_x \Leftrightarrow y \in \complement\,(E_x)\,.$$

The rest is clear.

Lemma 10.4.1. *If $E \in \mathscr{R}_{\sigma\delta}$, then for every $x \in X$, $E_x \in \mathscr{B}$ and for every $y \in Y$, $E^y \in \mathscr{A}$.*

Proof. The result is trivial if $E \in \mathscr{R}$. If $E = \cup_i E_i$ where each $E_i \in \mathscr{R}$, then for every $x \in X$ and every $y \in Y$,

$$\chi_{E_x}(y) = \chi_E(x,y) = \sup_i \chi_{E_i}(x,y) = \sup_i \chi_{(E_i)_x}(y).$$

It follows that χ_{E_x} is a measurable function of y, whence $E_x \in \mathscr{B}$. If now we assume that $E = \cap_i E_i$ where each $E_i \in \mathscr{R}_\sigma$, then for every $x \in X$ and every $y \in Y$,

$$\chi_{E_x}(y) = \chi_E(x,y) = \inf_i \chi_{E_i}(x,y) = \inf_i \chi_{(E_i)_x}(y).$$

It follows that χ_{E_x} is a measurable function of y, whence $E_x \in \mathscr{B}$. The proof that for every $y \in Y$, $E^y \in \mathscr{A}$ is similar.

If we are dealing with the product of finite or σ-finite measure spaces, then the product is also at least σ-finite. It then follows from Proposition 10.2.3 and its corollaries that every measurable set having finite measure is an $\mathscr{R}_{\sigma\delta}$ set minus a null set. Moreover, every measurable set is a countable union of $\mathscr{R}_{\sigma\delta}$ sets of finite measure where a null set has been removed from each. The next two lemmas establish properties of $\mathscr{R}_{\sigma\delta}$ sets of finite measure and null sets.

Lemma 10.4.2. *Let E be a set in $\mathscr{R}_{\sigma\delta}$ with $(\mu \times \nu)(E) < +\infty$. Then $x \to \nu(E_x)$ is a measurable function of x, and $\int \nu(E_x)\,d\mu = (\mu \times \nu)(E)$. A similar statement holds when the roles of x and y are reversed.*

Proof. The result is trivial if $E \in \mathscr{R}$. Any finite union of measurable rectangles can be written as a finite disjoint union. Therefore, given $E \in \mathscr{R}_\sigma$, we may assume it is a pairwise disjoint union of measurable rectangles E_i. Hence, the measurability of $x \to \nu(E_x)$ is clear. Moreover, for each $x \in X$, $\nu(E_x) = \Sigma_i \nu((E_i)_x)$, and so

$$\int \nu(E_x)\, d\mu = \Sigma_i \int \nu((E_i)_x)\, d\mu = \Sigma_i (\mu \times \nu)(E_i) = (\mu \times \nu)(E).$$

Now let E be a set in $\mathscr{R}_{\sigma\delta}$ for which the product measure is finite. Note that by the construction of the product measure, there is a sequence of measurable rectangles A_i such that $E \subseteq \cup_i A_i$ and $(\mu \times \nu)(E) \leq \Sigma_i (\mu \times \nu)(A_i) \leq (\mu \times \nu)(E) + 1$. Let $E_1 = \cup A_i \in \mathscr{R}_\sigma$. We now find $\mathscr{R}_\sigma$ sets E_n so that the sequence E_n is decreasing and $E = \cap E_n$; we use the following method: Given an $\mathscr{R}_\sigma$ set $F = \cup_i A_i$ and a second $\mathscr{R}_\sigma$ set $G = \cup_j B_j$, each $A_i \cap B_j \in \mathscr{R}$, and so we can form the $\mathscr{R}_\sigma$ set $\cup_i \cup_j (A_i \cap B_j) \subseteq F \cap G$. Starting with E_1 and continuing in this way, we may assume that the sequence E_n from $\mathscr{R}_\sigma$ is decreasing, and since $E \in \mathscr{R}_{\sigma\delta}$, we may assume that $E = \cap_n E_n$. Now for each $x \in X$, $E_x = \cap_n (E_n)_x$. Therefore, $\nu(E_n)_x$ is a decreasing sequence of integrable functions on X, and $\nu(E_x) = \lim_n \nu(E_n)_x$ for each $x \in X$. It follows that $\nu(E_x)$ is a measurable function of x, and by the Lebesgue dominated convergence theorem and properties of measures,

$$\int \nu(E_x)\, d\mu = \lim_n (\mu \times \nu)(E_n) = (\mu \times \nu)(E).$$

Lemma 10.4.3. *Let E be a set for which $(\mu \times \nu)(E) = 0$. Then for μ-almost all $x \in X$, $E_x \in \mathscr{B}$ and $\nu(E_x) = 0$. A similar statement is true for ν-almost all $y \in Y$.*

Proof. There is an $\mathscr{R}_{\sigma\delta}$ set $F \supseteq E$ such that $0 = (\mu \times \nu)(F) = \int \nu(F_x)\, d\mu$. Since $E_x \subseteq F_x$ and for μ-almost all $x \in X$, $\nu(F_x) = 0$, it follows from the completeness of ν that for μ-almost all $x \in X$, $E_x \in \mathscr{B}$ and $\nu(E_x) \leq \nu(F_x) = 0$.

Proposition 10.4.4. *Let E be a set of finite product measure. Then for μ-almost all $x \in X$, $E_x \in \mathscr{B}$ and $x \to \nu(E_x)$ defines a measurable function of x on X when the value is set equal to 0 where $\nu(E_x)$ is not defined. Moreover, $\int \nu(E_x)\, d\mu = (\mu \times \nu)(E)$. A similar statement holds for ν-almost all $y \in Y$.*

Proof. There is an $\mathscr{R}_{\sigma\delta}$ set $F \supseteq E$ with $(\mu \times \nu)(F \setminus E) = 0$. The result is true for F and for $F \setminus E$, so it is true for E. That is, for μ-almost all $x \in X$, E_x is a measurable subset of Y and $\nu(E_x) = \nu(F_x)$. The rest is clear.

Theorem 10.4.1 (Fubini). *Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be two complete measure spaces, and let f be an integrable function on $X \times Y$. Then*

a) *for μ-almost all $x \in X$, the function $y \mapsto f(x,y)$ is an integrable function on Y;*
b) *for ν-almost all $y \in Y$, the function $x \mapsto f(x,y)$ is an integrable function on X;*
c) *the function $x \mapsto \int_Y f(x,y)\nu(dy)$ is an integrable function on X;*
d) *the function $y \mapsto \int_X f(x,y)\mu(dx)$ is an integrable function on Y; and*

$$\int_X \left(\int_Y f d\nu \right) d\mu = \int_{X \times Y} f d(\mu \times \nu) = \int_Y \left(\int_X f d\mu \right) d\nu.$$

Proof. The collection of functions on $X \times Y$ for which the theorem is true is a vector subspace of the space of integrable functions on $X \times Y$. That subspace contains the space of simple functions that vanish off of sets of finite measure. We need, therefore, only prove the result for a nonnegative function f. By symmetry, we need only establish Properties **a** and **c** and the first equality between the iterated integral and $\int_{X\times Y} fd(\mu\times\nu)$. Given a nonnegative integrable function f on $X\times Y$, there is an increasing sequence of simple, nonnegative functions φ_n with limit f. Since f is integrable, each φ_n is integrable, and so must vanish off of a set of finite measure. For μ-almost all $x\in X$, each function $\varphi_n(x,\cdot)$ is measurable on Y. For such an x, $f(x,\cdot)$ is measurable on Y. Now, consider the function on X taking the value $\int_Y f(x,y)\nu(dy)$ when the integral is defined, and the value 0 when the integral is not defined. By the Monotone Convergence Theorem, $\int_Y f(x,y)\nu(dy) = \lim_n \int_Y \varphi_n(x,y)\nu(dy)$ for μ-almost all x, so the function $x\mapsto \int_Y f(x,y)\nu(dy)$ is measurable. Again by the Monotone Convergence Theorem,

$$\begin{aligned}\int_X\Big(\int_Y f\,d\nu\Big)d\mu &= \lim_n \int_X\Big(\int_Y \varphi_n d\nu\Big)d\mu \\ &= \lim_n \int_{X\times Y}\varphi_n d(\mu\times\nu) = \int_{X\times Y} f\,d(\mu\times\nu) < +\infty.\end{aligned}$$

It follows that $x\mapsto \int_Y f(x,y)\nu(dy)$ is a μ-integrable function on X, so Property **c** holds. Therefore, for μ-almost all $x\in X$, $\int_Y f(x,y)\nu(dy)$ is finite, so Property **a** holds. We also have established the first equality between the iterated integral and $\int_{X\times Y} fd(\mu\times\nu)$.

The finiteness condition that we needed to go from characteristic functions of sets of finite product measure to the general case is obtained in the Fubini theorem via the assumption that f is integrable. Here is another way to go from characteristic functions to the general case.

Theorem 10.4.2 (Tonelli). *Let $(X,\mathscr{A},\mu)$ and $(Y,\mathscr{B},\nu)$ be two complete σ-finite measure spaces, and let f be a nonnegative measurable function on $X\times Y$. Then*

a) *for μ-almost all $x\in X$, the function $y\mapsto f(x,y)$ is a measurable function on Y;*
b) *for ν-almost all $y\in Y$, the function $x\mapsto f(x,y)$ is a measurable function on X;*
c) *the function $x\mapsto \int_Y f(x,y)\nu(dy)$ is a measurable function on X;*
d) *the function $y\mapsto \int_X f(x,y)\mu(dx)$ is a measurable function on Y; and*

$$\int_X\left(\int_Y fd\nu\right)d\mu = \int_{X\times Y} fd(\mu\times\nu) = \int_Y\left(\int_X fd\mu\right)d\nu.$$

Proof. With the assumption of σ-finiteness, we can take an increasing sequence of simple functions φ_n converging up to a measurable, nonnegative f on $X\times Y$ so that each φ_n vanishes off of a set of finite measure. The proof is the same as for the Fubini theorem, except the result is now stated for nonnegative functions, so the space for which the result holds is stable under addition and multiplication only by nonnegative scalars.

10.5 Other Integrals

There are other approaches to integration that may interest the reader. For example, recall that a real vector lattice on a set X is a vector space of functions $\mathscr{L}$ with scalars in $\mathbb{R}$ such that if f and g are in $\mathscr{L}$ so are $f \wedge g$ (the pointwise minimum) and $f \vee g$ (the pointwise maximum.) Note that $|f| = (f \vee 0) + (-f \vee 0)$. Conversely, given the absolute value function, we have

$$f \vee g = (1/2)[f+g+|f-g|], \qquad f \wedge g = (1/2)[f+g-|f-g|].$$

Definition 10.5.1. A **positive linear functional** I on $\mathscr{L}$ is a linear map of $\mathscr{L}$ into $\mathbb{R}$ taking nonnegative functions to nonnegative values. In particular, it is increasing in the sense that if $f \leq g$, then $I(f)$ is less than or equal to $I(g)$. Let I be such a positive linear functional with the additional property that if $\langle \varphi_n : n \in \mathbb{N} \rangle$ is a decreasing sequence in $\mathscr{L}$ with limit 0 at each point of X, then $I(\varphi_n) \searrow 0$. The corresponding **Daniell integral** is the extension of I to a larger vector lattice.

Example 10.5.1. Let $\mathscr{L}$ be the family of continuous functions with compact support on $\mathbb{R}$, and let I be the Riemann integral on $\mathscr{L}$. The corresponding Daniell integral is the Lebesgue integral on $\mathbb{R}$.

Example 10.5.2. Let $\mathscr{A}$ be an algebra of subsets of a set X, and let μ be a finite, σ-additive measure on $(X, \mathscr{A})$. Let $\mathscr{L}$ be the family of simple, measurable functions on X, and let I be the corresponding integral with respect to μ. The Daniell integral is the integral with respect to the σ-additive extension of μ to the σ-algebra generated by $\mathscr{A}$. See Theorem 10.2.1.

Another example, in [32], gives an alternate approach to the integral for the measure theory described in the Appendix on Infinitesimal Analysis and Measure Theory. For the general construction of the Daniell integral, see [45].

There is a large body of literature on the Henstock–Kurzweil integral. That integral uses sums like those for the Riemann integral, but the size of an interval in which the integrand is evaluated at a point t depends on the value of the integrand at t. The corresponding integral is a version of the Lebesgue integral. For the connection with “points of approximate continuity”, see [33].

10.6 Problems

Problem 10.1. Let X consist of the rational numbers in $(0,1]$, and let $\mathscr{A}$ be the algebra formed by disjoint unions of intervals $(a,b] \cap X$. Let $\mu(\varnothing) = 0$, and for all other sets $E \in \mathscr{A}$, let $\mu(E) = +\infty$. Show that the extension of μ to $\sigma(\mathscr{A})$ is not unique. **Hint:** For each $x \in X$, $\{x\} \in \sigma(\mathscr{A})$.

Problem 10.2. Prove Proposition 10.3.1.

Problem 10.3. There are topological spaces so disconnected that there is a base for the topology consisting of **clopen** sets. Clopen sets are sets that are simultaneously both open and closed. Let X be such a space.

a) Show that the collection $\mathscr{A}$ of clopen sets in X is an algebra in X.
b) Show that if X with the topology generated by clopen sets is compact, then any finitely additive measure on the algebra $\mathscr{A}$ of clopen sets can be extended to a countably additive measure on the completion of the smallest σ-algebra containing $\mathscr{A}$.

Problem 10.4 (A). In $\mathbb{R}^n$, let $\overline{R}$ be a general closed rectangle, and let the collection $\{P_i : i = 1, \cdots, m\}$ be a finite covering of $\overline{R}$ by general open rectangles. Show that $V(\overline{R}) \leq \sum_{i=1}^{m} V(P_i)$.

Problem 10.5. Prove Proposition 10.3.3.

Problem 10.6. Fix nonempty sets A and $B \subseteq \mathbb{R}^n$ such that

$$d(A,B) := \inf\{|x-y| : x \in A, y \in B\} = a > 0.$$

Show that Lebesgue outer measure $\lambda^*(A \cup B) = \lambda^*(A) + \lambda^*(B)$.

Problem 10.7. Show that Lebesgue outer measure on $\mathbb{R}^n$ is translation invariant.

Problem 10.8. The family of **Borel sets** in $\mathbb{R}^n$ is the smallest σ-algebra containing the open subsets of $\mathbb{R}^n$. Extend Proposition 2.4.2 and Corollaries 2.5.1 and 2.5.2 of Theorem 2.5.1, as they apply to Lebesgue measure on $\mathbb{R}$, to Lebesgue measure on $\mathbb{R}^n$.

Problem 10.9. A measure defined on the family of Borel sets is called a **Borel measure**. A Borel measure μ on $\mathbb{R}^n$ is called a **doubling measure** if there exists a constant C such that for every open ball $B(x,r)$, $\mu(B(x,2r)) \leq C\mu(B(x,r))$. Let μ and η be two doubling Borel measures on $\mathbb{R}$. Prove that the Borel measure $\mu \times \eta$ is a doubling measure on $\mathbb{R}^2$. **Hint:**

$$[x-r/2, x+r/2] \times [y-r/2, y+r/2] \subseteq B((x,y),r)$$
$$B((x,y),2r) \subseteq [x-2r, x+2r] \times [y-2r, y+2r].$$

Problem 10.10. Fix two measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$. Let $\mathscr{Q}$ be the collection of finite unions of disjoint measurable rectangles. Show that $\mathscr{Q}$ is an algebra of subsets of $X \times Y$.

Problem 10.11. Prove Proposition 10.4.2.

Chapter 11
Banach Spaces

11.1 Banach Spaces

In this chapter, we continue developing properties of normed linear spaces, and in particular, **Banach spaces**, i.e., spaces that are complete with respect to the metric generated by the norm. We have already considered the important examples of Hilbert spaces and L^p spaces. Further aspects of the latter will be developed in this chapter. The field of scalars for our work is either the set of real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$. These fields are examples of normed linear spaces; the absolute value function is the norm for $\mathbb{R}$, and the modulus is the norm for $\mathbb{C}$.

A linear subspace of a normed space is often called a **linear manifold**. Linear independence and dimension are the usual notions defined for all vector spaces. (See, for example, Definition 8.5.1.) A finite-dimensional linear manifold in a normed space is always closed with respect to the topology generated by the norm. (See Problem 11.2(A).)

Definition 11.1.1. Equivalent norms on a normed space are norms that generate the same topology.

Remark 11.1.1. Two norms are equivalent if each norm multiplied by an appropriate constant dominates the other. To see this, we note that open neighborhoods of a point are translates of open neighborhoods of 0. Therefore, we need only have an open ball about 0 in terms of each one of the norms contain an open ball about 0 in terms of the other norm. Recall that for a normed space X and all x, y in X, $|\|x\| - \|y\|| \leq \|x - y\|$, whence $x \mapsto \|x\|$ is a uniformly continuous map from X into $\mathbb{R}^+$.

Definition 11.1.2. A map from one vector space into another preserving addition and scalar multiplication is called a **linear map** or **linear transformation**.

Recall that linear functionals on a normed linear space are linear maps into the scalar field. Properties of such functionals, such as Theorem 7.5.1, hold for more general maps.

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2_11

Definition 11.1.3. A linear transformation T from one normed linear space X to another is **bounded** if the following quantity is finite:

$$\|T\| := \sup_{x\in X, x\neq 0} \frac{\|T(x)\|}{\|x\|} = \sup_{\|x\|=1} \|T(x)\| = \sup_{\|x\|\leq 1} \|T(x)\|.$$

The nonnegative real number $\|T\|$ is called the **norm** of the linear transformation.

Theorem 11.1.1. *If a linear transformation T is continuous at any point of a normed vector space X (not necessarily complete), then it is bounded. If T is bounded, then it is uniformly continuous on all of X.*

Proof. If there is an open ball $B(x,\delta)$ about x that maps into a ball $B(T(x),\varepsilon)$ about $T(x)$, then the translate

$$B(x,\delta) - x = \{y : y = z - x,\ z \in B(x,\delta)\}$$

is the open ball $B(\mathbf{0},\delta)$, and

$$T[B(\mathbf{0},\delta)] \subseteq B(T(x),\varepsilon) - T(x) = B(\mathbf{0},\varepsilon).$$

It follows that if T is continuous at x, then it is continuous at $\mathbf{0}$. Moreover, T is bounded since $\|x\| < \delta$ if and only if $\left\|\frac{2}{\delta}x\right\| < 2$, so

$$\sup_{\|x\|\leq 1} \|T(x)\| \leq \sup_{\|x\|<2} \|T(x)\| \leq \frac{2}{\delta} \cdot \sup_{x\in B(0,\delta)} \|T(x)\| \leq \frac{2\varepsilon}{\delta}.$$

The uniform continuity follows from the Lipschitz inequality that for all x and y in X, $\|T(x) - T(y)\| \leq \|T\| \cdot \|x - y\|$.

Definition 11.1.4. The space of bounded linear functionals on X is called the **dual space** or **conjugate** space of X. It is denoted by X^*.

Theorem 11.1.2. *The space $\mathscr{B}$ of bounded linear maps from a normed space X into a Banach space Y forms a Banach space. The norm of each bounded linear transformation T is the nonnegative real number $\|T\|$.*

Proof. The set $\mathscr{B}$ forms a vector space where addition and scalar multiplication are the usual pointwise operations for functions from one vector space into another. To show that $\|\cdot\|$ is a norm on $\mathscr{B}$, we note that $\|T\| = 0$ if and only if $T = 0$. For any scalar α,

$$\|\alpha T\| = \sup_{\|x\|=1} \|\alpha T(x)\| = |\alpha| \sup_{\|x\|=1} \|T(x)\| = |\alpha| \cdot \|T\|.$$

Given A, $B \in \mathscr{B}$, we have

$$\begin{aligned} \|A+B\| &= \sup_{\|x\|=1} \|A(x) + B(x)\| \leq \sup_{\|x\|=1} (\|A(x)\| + \|B(x)\|) \\ &\leq \sup_{\|x\|=1} \|A(x)\| + \sup_{\|x\|=1} \|B(x)\| = \|A\| + \|B\|. \end{aligned}$$

It only remains to show that $\mathscr{B}$ is complete. Let $\langle A_n : n \in \mathbb{N} \rangle$ be a Cauchy sequence in $\mathscr{B}$. Then for all $x \in X$, $\langle A_n(x) \rangle$ is a Cauchy sequence in Y since for all n, m in $\mathbb{N}$, $\|A_n(x) - A_m(x)\| \leq \|A_n - A_m\| \, \|x\|$. Let $A(x)$ denote the limit. Since A is the pointwise limit of linear maps, it too is linear. For example, the sequence $\langle A_n(x) + A_n(y) \rangle$ has limit $A(x) + A(y)$ and also $A(x+y)$, and so these two limits are equal. To show that A is a bounded linear map, we find an $N \in \mathbb{N}$ so that for $n \geq N$, we have

$$\big| \|A_n\| - \|A_N\| \big| \leq \|A_n - A_N\| \leq 1.$$

It follows that for all $x \in X$ with $\|x\| \leq 1$, and for all $n \geq N$, $\|A_n(x)\| \leq \|A_N\| + 1$. Since $\|\cdot\|$ is continuous on Y, $\|A(x)\| \leq \|A_N\| + 1$ for all x with $\|x\| \leq 1$, whence A is bounded. To show that $A_n \to A$, with respect to the norm $\|\cdot\|$ on $\mathscr{B}$, we fix $\varepsilon > 0$, and choose $N \in \mathbb{N}$ so that for n, $m \geq N$, $\|A_n - A_m\| < \varepsilon$. For any x with $\|x\| \leq 1$, and for any $n \geq N$,

$$\|A_n(x) - A(x)\| = \lim_m \|A_n(x) - A_m(x)\| \leq \lim_m \|A_n - A_m\| \leq \varepsilon,$$

whence $\|A_n - A\| \leq \varepsilon$.

Corollary 11.1.1. *The dual space X^* of a normed space X is a Banach space. The norm of each bounded linear functional F is $\|F\| := \sup_{\|x\| \leq 1} |F(x)|$.*

Theorem 11.1.3. *A linear functional f on a normed space X is bounded if and only if its kernel $K := \{x \in X : f(x) = 0\}$ is closed in X.*

Proof. Exercise 11.3(A).

11.2 Return to Classical Normed Spaces

We work with a measure space $(X, \mathscr{B}, \mu)$. We don't assume μ is complete; $f = g$ almost everywhere means that we have equality outside of a μ-null set. Two functions are equivalent if they are equal almost everywhere. We use notation such as f for both the function f and the equivalence class it represents. If the scalar field is $\mathbb{R}$, then $|f|$ denotes the absolute value of f. If the scalar field is $\mathbb{C}$, then $|f|$ denotes the modulus of f. For $1 \leq p < \infty$, the space $L^p(\mu)$ consists of those equivalence classes of measurable functions f such that $\int |f|^p \, d\mu$ is finite. For each $f \in L^p(\mu)$, $\|f\|_p := [\int (|f|)^p]^{1/p}$. For $p = \infty$, $\|f\|_\infty$ denotes the essential supremum of $|f|$. The space $L^\infty(\mu)$ consists of those equivalence classes of measurable functions f for which $\|f\|_\infty$ is finite.

For $1 \leq p \leq \infty$, $L^p(\mu)$ (also denoted by just L^p) is a linear space with respect to the scalar field. Real values p and q greater than 1 satisfying the equality $1/p + 1/q = 1$ are called **conjugate exponents**. The values $p = 1$ and $q = \infty$ are also paired in the theory. For $1 \leq p \leq \infty$, the map $f \mapsto \|f\|_p$ is a norm on L^p, and L^p is complete with respect to the metric generated by its norm; that is, it is a

Banach space. When the measure space is $\mathbb{N}$ with counting measure, we write ℓ^p instead of L^p. For example, ℓ^∞ is the space of bounded sequences. The reader can show in Problem 7.43(A) that if f is a function in every L^p for $1 \le p \le \infty$, then $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$.

We recall now the inequalities of Hölder and Minkowski, and add a condition for equality in Minkowski's Inequality.

Lemma 11.2.1. *For real-valued and complex-valued functions f and g, $|f+g| = |f|+|g|$ on a set E if and only if where the product is not 0, the ratio is positive.*

Proof. Exercise 11.4.

Theorem 11.2.1 (Hölder's Inequality). *Assume either that p and q are real numbers larger than 1 with $1/p + 1/q = 1$ or that $p = 1$ and $q = \infty$. In either case, if $f \in L^p$ and $g \in L^q$, then $f \cdot g \in L^1$ and $\|fg\|_1 \le \|f\|_p \cdot \|g\|_q$. The inequality is equality if the right side is 0. Otherwise, for $p = 1$, equality holds if and only if $|g(x)| = \|g\|_\infty$ for almost all x such that $f(x) \ne 0$, while for $p > 1$, equality holds if and only if there are positive constants s and t such that $s \cdot |f|^p = t \cdot |g|^q$ a.e.*

Proof. Presented in Theorem 7.4.1.

Theorem 11.2.2 (Minkowski's Inequality). *If f and g are in L^p, then so is $f+g$ and $\|f+g\|_p \le \|f\|_p + \|g\|_p$. Assume that neither $\|f\|_p$ nor $\|g\|_p$ is 0. Then we have equality for $p = 1$ if and only if $|f(x)+g(x)| = |f(x)| + |g(x)|$ for almost all x. We have equality for $1 < p < \infty$ if and only if $g = \gamma \cdot f$ a.e. for some $\gamma > 0$. We have equality for $p = \infty$ if and only if for each $\varepsilon > 0$ there is a set E of strictly positive measure such that the following holds almost everywhere on E: $|f(x)+g(x)| \ge |f(x)| + |g(x)| - \varepsilon$, $|f(x)| \ge \|f\|_\infty - \varepsilon$, and $|g(x)| \ge \|g\|_\infty - \varepsilon$.*

Proof. The inequality is established in Theorem 7.4.2; we repeat the proof here for the case $1 < p < \infty$. We assume that neither $\|f\|_p$ nor $\|g\|_p$ is 0. Equality for the case $p = 1$ is Exercise 11.5. Suppose $1 < p < \infty$ and $g = \gamma \cdot f$ a.e. for some $\gamma > 0$. Then we have equality since

$$\left(\int |f+g|^p\right)^{1/p} = \left(\int |(1+\gamma)f|^p\right)^{1/p} = (1+\gamma)\left(\int |f|^p\right)^{1/p} = \|f\|_p + \|g\|_p .$$

On the other hand,

$$\int |f+g|^p \le \int \left(|f+g|^{p-1} \cdot |f|\right) + \int \left(|f+g|^{p-1} \cdot |g|\right).$$

Equality holds here if and only if $|f+g| = |f| + |g|$ a.e. We employ Hölder's Inequality with $\frac{1}{p} + \frac{1}{q} = 1$ so $pq - q = p$.

$$\int \left(|f+g|^{p-1} \cdot |f|\right) \le \|f\|_p \cdot \left(\int |f+g|^p\right)^{1/q} = \|f\|_p \cdot \|f+g\|_p^{p/q},$$

with equality implying the existence of constants $s > 0$ and $t > 0$ such that

$$s \cdot |f|^p = t \cdot |f+g|^p \text{ a.e.}$$

Similarly,

$$\int \left(|f+g|^{p-1} \cdot |g|\right) \leq \|g\|_p \cdot \left(\int |f+g|^p\right)^{1/q} = \|g\|_p \cdot \|f+g\|_p^{p/q},$$

with equality implying the existence of constants $u > 0$ and $v > 0$ such that

$$u \cdot |g|^p = v \cdot |f+g|^p \text{ a.e.}$$

Now we have

$$\int |f+g|^p \leq \|f\|_p \cdot \|f+g\|_p^{p/q} + \|g\|_p \cdot \|f+g\|_p^{p/q}.$$

Since $\frac{1}{p} + \frac{1}{q} = 1$, $p = 1 + \frac{p}{q}$, so $p - \frac{p}{q} = 1$. Therefore,

$$\|f+g\|_p = \|f+g\|_p^{p-p/q} \leq \|f\|_p + \|g\|_p.$$

This again establishes the inequality for the case $1 < p < \infty$. If we have equality, then $|f+g| = |f| + |g|$ a.e., and

$$s \cdot |f|^p = t \cdot |f+g|^p = v \cdot \frac{t}{v} \cdot |f+g|^p = u \cdot \frac{t}{v} \cdot |g|^p \text{ a.e.},$$

so for some $\gamma > 0$, $|g| = \gamma \cdot |f|$ a.e. Therefore, outside of a null set, $f = 0$ if and only if $g = 0$, and by Lemma 11.2.1, where neither is 0, the ratio is positive. It follows that almost everywhere, $g = \gamma \cdot f$.

Finally, suppose $p = \infty$. Let f and g be representative functions for their equivalence classes. For each $\varepsilon > 0$, let $F_\varepsilon := \{|f(x)| \geq \|f\|_\infty - \varepsilon\}$, and let $G_\varepsilon := \{|g(x)| \geq \|g\|_\infty - \varepsilon\}$. If for each $\varepsilon > 0$ there is a set $E_\varepsilon \subseteq F_\varepsilon \cap G_\varepsilon$ with strictly positive measure such that $|f+g| \geq |f| + |g| - \varepsilon$ on E_ε, then

$$\begin{aligned}\|f+g\|_\infty \geq \|(f+g) \cdot \chi_{E_\varepsilon}\|_\infty &\geq \inf_{x \in E_\varepsilon} |f(x) + g(x)| \\ &\geq \inf_{x \in E_\varepsilon} (|f(x)| + |g(x)|) - \varepsilon \geq \|f\|_\infty + \|g\|_\infty - 3\varepsilon.\end{aligned}$$

Since ε is arbitrary, $\|f+g\|_\infty \geq \|f\|_\infty + \|g\|_\infty$. We must then have equality, and so the condition of the theorem is sufficient for equality.

Suppose the condition fails for some $\varepsilon > 0$. For $x \in X \setminus (F_\varepsilon \cap G_\varepsilon)$, $|f(x)| < \|f\|_\infty - \varepsilon$ or $|g(x)| < \|g\|_\infty - \varepsilon$ or both, in which case

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty - \varepsilon \text{ a.e.}$$

Let A be the set of all $x \in F_\varepsilon \cap G_\varepsilon$ such that $|f(x)+g(x)| \geq |f(x)|+|g(x)|-\varepsilon$. Since the condition fails for ε, A has measure 0. For $x \in (F_\varepsilon \cap G_\varepsilon) \setminus A$,

$$|f(x)+g(x)| < |f(x)|+|g(x)|-\varepsilon \leq \|f\|_\infty + \|g\|_\infty - \varepsilon \text{ a.e.}$$

It follows that the condition of the theorem is necessary for equality.

Remark 11.2.1. If f and g are real-valued functions, the condition that for every $\varepsilon > 0$, $|f(x)+g(x)| \geq |f(x)|+|g(x)|-\varepsilon$ on E is the condition that $f(x)$ and $g(x)$ have the same sign at all points of E.

11.3 Dual Space of L^p

For the rest of this chapter, we assume that the scalar field is just $\mathbb{R}$. As a corollary of Theorem 11.1.2, we have seen that the dual space X^* of a normed space X is a Banach space, with norm of each bounded linear functional F equal to $\sup_{\|x\|\leq 1} \|F(x)\|$. It is convenient to have a more concrete representation for the dual space of an L^p space in terms of a space that is isometrically isomorphic to the dual space. Two normed linear spaces are called isometrically isomorphic when there is a linear bijection between them such that the norm of a point equals the norm of its image.

Recall that as an immediate consequence of Hölder's Inequality, we noted that for conjugate pairs p and q, and also for $p = 1$ and $q = \infty$, when $g \in L^q$, the map $f \mapsto \int f \cdot g$ is a bounded linear functional F_g on L^p with $\left\|F_g\right\| \leq \|g\|_q$. As promised, we now refine that result. If $p = 1$, the measure should be σ-finite, as we show with the following example.

Example 11.3.1. Suppose $X = \{3,5\}$, with $\mu(\{3\}) = 1$ and $\mu(\{5\}) = +\infty$. Then μ is not σ-finite. Any L^1 function must take the value 0 at 5. If $g(3) = 1$ and $g(5) = 2$, then $g \in L^\infty$, and g has L^∞-norm 2. The functional F_g, however, has norm 1. Moreover, we can change the value of g at 5 and still represent the same functional on L^1.

Proposition 11.3.1. *Assume that μ is a σ-finite measure on $(X, \mathscr{B})$, and let $p = 1$. If $g \in L^\infty(\mu)$, then the mapping $f \mapsto \int f \cdot g \, d\mu$ defines a bounded linear functional F_g on $L^1(\mu)$ with $\left\|F_g\right\| = \|g\|_\infty$.*

Proof. By Hölder's Inequality, $\left|F_g(f)\right| \leq \int |fg| \leq \|f\|_1 \cdot \|g\|_\infty$, so F_g is bounded with $\left\|F_g\right\| \leq \|g\|_\infty$. To show the reverse inequality, we may assume that $\|g\|_\infty \neq 0$. Let g be a representative of its equivalence class. Fix ε with $0 < \varepsilon < \|g\|_\infty$. Since the space X can be decomposed into a countable number of measurable sets of finite measure, there is a measurable set A of finite, strictly positive measure such that $|g(x)| \geq \|g\|_\infty - \varepsilon$ for all $x \in A$. Set $f(x) := \frac{1}{\mu(A)} \cdot (g(x)/|g(x)|) \cdot \chi_A$. Then $\|f\|_1 = 1$ and

$$|F_g(f)| = \int f \cdot g = \frac{1}{\mu(A)} \cdot \int_A |g| \geq \|g\|_\infty - \varepsilon.$$

Since ε is arbitrary in its range, $\|F_g\| = \|g\|_\infty$.

Proposition 11.3.2. *Assume that μ is any measure on $(X, \mathscr{B})$, and fix p and q in $\mathbb{R}^+$ with $1/p + 1/q = 1$. If $g \in L^q(\mu)$, then the mapping $f \mapsto \int f \cdot g \, d\mu$ defines a bounded linear functional F_g on $L^p(\mu)$ with $\|F_g\| = \|g\|_q$.*

Proof. By Hölder's Inequality, $|F_g(f)| \leq \int |fg| \leq \|f\|_p \cdot \|g\|_q$, so F_g is bounded with $\|F_g\| \leq \|g\|_q$. Assume that $\|g\|_q \neq 0$, and let g be a representative of its equivalence class. Set $f(x) := |g(x)|^{q/p} \cdot (g(x)/|g(x)|)$ for all $x \in \{g \neq 0\}$, and set $f(x) = 0$ otherwise. Since $q/p + 1 = q(1/p + 1/q) = q$,

$$|f|^p = |g|^q = |g|^{q/p+1} = |g|^{q/p} \cdot |g| = f \cdot g.$$

Since $g \in L^q$, it follows that, $f \in L^p$ and $\|f\|_p = (\int |g|^q)^{1/p} = \|g\|_q^{q/p}$. Moreover,

$$|F_g(f)| = \int fg = \int |g|^{q/p+1} = \|g\|_q^q = \|g\|_q^{(q/p+1)} = \|f\|_p \cdot \|g\|_q.$$

Theorem 11.3.1. *Let $(X, \mathscr{B}, \mu)$ be a measure space, and fix p with $1 \leq p < \infty$. For $p = 1$, assume that μ is σ-finite, and set $q = \infty$. For $p > 1$, fix q with $1/p + 1/q = 1$. If F is a bounded linear functional on L^p, then there is a unique $g \in L^q$ such that $F = F_g$. Therefore, the map $g \mapsto F_g$ is an isometric isomorphism of L^q onto the dual space, $(L^p)^*$, of L^p.*

Proof. To establish uniqueness, suppose $F_g = F_h$ on L^p. Let g and h be representatives of their equivalence classes. If $p = 1$, let $Z = X$. If $p > 1$, set $Z = \{|g| > 0\} \cup \{|h| > 0\}$. In either case, $Z = \cup_j Z_j$, where j ranges over a finite or countably infinite set, and each Z_j has finite μ-measure. Given j, and given $n \in \mathbb{N}$, set $E := \{g \geq h + 1/n\} \cap Z_j$. Then $\chi_E \in L^p$, and

$$F_g(\chi_E) = \int_E g d\mu \geq \int_E h d\mu + \frac{1}{n}\mu(E) = F_g(\chi_E) + \frac{1}{n}\mu(E).$$

Since $F_g(\chi_E)$ is finite, $\mu(E) = 0$, and this is true for each j and each $n \in \mathbb{N}$. It follows that $g \leq h$ μ-a.e. on X. Similarly, $h \leq g$ μ-a.e. on X, so $h = g$ μ-a.e. on X.

Now we assume that μ is a finite measure on $(X, \mathscr{B})$. Fix a bounded linear functional F on $L^p(\mu)$. Set $\nu(E) = F(\chi_E)$ for each measurable set E. Since F is linear, $\nu(\varnothing) = F(0) = 0$, and for disjoint measurable sets A and B,

$$\nu(A \cup B) = F(\chi_A + \chi_B) = F(\chi_A) + F(\chi_B) = \nu(A) + \nu(B).$$

Suppose $\langle E_i \rangle$ is a pairwise disjoint sequence of measurable sets with union A, and A_k is the union of the first k of the sets E_i. Then as $k \to \infty$,

$$\left\|\chi_A - \chi_{A_k}\right\|_p = \left(\int_{A\setminus A_k} 1^p d\mu\right)^{1/p} = (\mu(A\setminus A_k))^{1/p} \to 0.$$

Since F is continuous, it follows that as $k\to\infty$,

$$\sum_{i=1}^{k} \nu(E_i) = \nu(A_k) = F(\chi_{A_k}) \to F(\chi_A) = \nu(A) = \nu\left(\cup_{i=1}^{\infty} E_i\right).$$

Therefore, ν is a signed measure on X. By assumption, $\mu(X)$ is finite, so χ_X is integrable, whence ν is a finite, signed measure on X. Moreover, for any measurable set A, in particular for a positive set or a negative set,

$$|\nu(A)| = |F(\chi_A)| \le \|\chi_A\|_p \|F\| = \left(\int_A 1^p d\mu\right)^{1/p} \|F\| = (\mu(A))^{1/p}\|F\|.$$

Therefore, if $\mu(A)=0$, then A is a null set for ν.

Let $\nu = \nu^+ - \nu^-$ be the Jordan decomposition for ν. Recall that $\nu^+ \perp \nu^-$ and $|\nu| := \nu^+ + \nu^-$. We have seen that $|\nu| << \mu$. Let g be the Radon-Nikodým derivative of ν with respect to μ, so $g = \frac{d\nu^+}{d\mu} - \frac{d\nu^-}{d\mu}$. It now follows that g is integrable, and for each simple function φ, $F(\varphi) = \int \varphi \cdot g \, d\mu$. Let g be a real-valued representative of its a.e. equivalence class.

For the case $p=1$, fix $\varepsilon>0$. Set $A := \{g > \|F\| + \varepsilon\}$, $B := \{g < -\|F\| - \varepsilon\}$. Then

$$\begin{aligned}(\|F\|+\varepsilon)\cdot\mu(A) &\le \int_A g\,d\mu = |F(\chi_A)| \le \|F\|\cdot\|\chi_A\|_1 = \|F\|\cdot\mu(A),\\ (-\|F\|-\varepsilon)\cdot\mu(B) &\ge \int_B g\,d\mu = F(\chi_B)\\ &\ge -|F(\chi_B)| \ge -\|F\|\cdot\|\chi_B\|_1 = -\|F\|\cdot\mu(B).\end{aligned}$$

It follows that $\mu(A)=0$ and $\mu(B)=0$. Since $\varepsilon>0$ is arbitrary, $|g|\le\|F\|$ μ-a.e., whence $g\in L^\infty$.

For the case $p>1$, set $E_n := \{0<|g|\le n\}$, and let $h = \chi_{E_n}\cdot|g|^q/g$, so $|h| = \chi_{E_n}\cdot|g|^{q-1}$. Now $1/p+1/q=1$, so $q+p=pq$, and $q=pq-p$. Therefore, $|h|^p = |g|^q\cdot\chi_{E_n}$. Moreover, h is bounded, and since $h\cdot g = \chi_{E_n}\cdot|g|^q$,

$$\int_{E_n} |g|^q\,d\mu = \int_X h\cdot g\,d\mu = F(h) \le \|F\|\cdot\|h\|_p = \|F\|\cdot\left(\int_{E_n}|g|^q\,d\mu\right)^{1/p},$$

whence

$$\left(\int_{E_n}|g|^q\,d\mu\right)^{1-1/p} = \left(\int_{E_n}|g|^q\,d\mu\right)^{1/q} \le \|F\|.$$

Using the Monotone Convergence Theorem, if follows that $g\in L^q$.

Since $g\in L^q$, by Propositions 11.3.1 and 11.3.2, the mapping $f\mapsto F_g(f) = \int_X f\cdot g\,d\mu$ is a continuous linear functional on L^p. By Problem 7.41(A), simple functions form a dense subset of L^p. Given $f \in L_p$, let $\langle \varphi_n \rangle$ be a sequence of simple functions with $\|\varphi_n - f\|_p \to 0$. By the continuity of the functionals F_g and F,

$$\int_X f \cdot g \, d\mu = F_g(f) = \lim_{n\to\infty} F_g(\varphi_n) = \lim_n \int_X \varphi_n \cdot g \, d\mu = \lim_n F(\varphi_n) = F(f).$$

This proves the result for a finite measure space.

Now assume that μ is σ-finite. Let X_n be an increasing sequence of sets of finite measure with union X. Let χ_n denote the characteristic function of X_n. For each n, fix g_n vanishing off of X_n such that if $f \in L^p$ and f vanishes off of X_n, then $F(f) = \int_X f \cdot g_n d\mu$. The functions g_n are unique up to sets of measure 0, and so we may assume that for $m < n$, $g_n \cdot \chi_m = g_m$. This determines a function g on all of X with $\int_X |g|^q \leq \|F\|^q$ since this is true for g_n on X_n for each n. If $f \in L^p$, then $\|f - f \cdot \chi_n\|_p^p = \int_{X \setminus X_n} |f|^p \, d\mu \to 0$ by the Lebesgue Dominated Convergence Theorem. Therefore, $f \cdot \chi_n \to f$ in L^p. By Hölder's Inequality, $|fg|$ is integrable, and of course, it dominates $|f \cdot g \cdot \chi_n|$. Since F is continuous on L^p,

$$F(f) = \lim_n F(f \cdot \chi_n) = \lim_n \int f \cdot g \cdot \chi_n d\mu = \int_X f \cdot g d\mu.$$

Now assume that μ is not σ-finite and fix p with $1 < p < \infty$. For each set E of σ-finite measure, we can find a function g_E that vanishes off of E and represents F with respect to all L^p functions that vanish off of E. We must have $\int_X |g_E|^q \leq \|F\|^q$. Let M be the supremum of the values $\int_X |g_E|^q$. There is a set H of σ-finite measure such that $g := g_H$ gives the value M. We may assume that $g = g_E$ when $E \subset H$. Moreover, by the definition of M, for any set A of σ-finite measure, $g_A = 0$ μ-a.e. on $A \setminus H$. If $f \in L^p$, then f vanishes off of a set A of σ-finite measure, so

$$F(f) = \int_X f \cdot g_A \, d\mu = \int_{H \cap A} f \cdot g_A d\mu = \int_{H \cap A} f \cdot g d\mu = \int_X f \cdot g d\mu.$$

This completes the proof.

11.4 Hahn-Banach Theorem

We next take up an important condition that enables a linear functional to be appropriately extended beyond its original domain.

Definition 11.4.1. A real-valued function p on a vector space X is subadditive if for every x and y in X, $p(x+y) \leq p(x) + p(y)$. It is positive homogeneous if for every $\alpha \geq 0$ and every x in X, $p(\alpha x) = \alpha p(x)$.

Example 11.4.1. A norm on a linear space X is both subadditive and positive homogeneous.

Theorem 11.4.1 (Hahn-Banach). *Suppose X is a vector space, and let p be a real-valued, subadditive, positive homogeneous function defined on X. Let S be a vector* *subspace of X, and let f be a real-valued linear functional defined on S with $f \leq p$ on S. Then there is a linear functional G defined on X such that $G \leq p$ on X and the restriction of G to S is equal to f.*

Proof. If $S = X$, we are done. Let $y \in X$ be a point not in S, and set Y equal to the vector space generated by S and y. Elements of Y have the form $s + \alpha y$ for $s \in S$ and $\alpha \in \mathbb{R}$. We now show that the result holds for $X = Y$. Given s_1 and s_2 in S, since

$$f(s_1) + f(s_2) = f(s_1 + s_2) \leq p(s_1 + s_2) \leq p(s_1 - y) + p(s_2 + y),$$

so

$$-p(s_1 - y) + f(s_1) \leq p(s_2 + y) - f(s_2).$$

Let I_y be the closed, nonempty interval given by

$$I_y := \left[\sup_{s \in S}(-p(s-y) + f(s)), \inf_{s \in S}(p(s+y) - f(s))\right].$$

Fix a value $h(y) \in I_y$. Given $s + \alpha y$ in Y, set

$$h(s + \alpha y) := f(s) + \alpha h(y).$$

If $\alpha = 0$, $h(s + \alpha y) = f(s) \leq p(s) = p(s + \alpha y)$. If $\alpha > 0$, then by the choice of $h(y)$,

$$\begin{aligned} h(s + \alpha y) &= f(s) + \alpha h(y) = \alpha[f(s/\alpha) + h(y)] \\ &\leq \alpha[f(s/\alpha) + (p(s/\alpha + y) - f(s/\alpha))] = p(s + \alpha y). \end{aligned}$$

If $\alpha = -\beta < 0$, then by the choice of $h(y)$,

$$\begin{aligned} h(s + \alpha y) &= f(s) + \alpha h(y) = \beta[f(s/\beta) - h(y)] \\ &\leq \beta[f(s/\beta) + (p(s/\beta - y) - f(s/\beta))] = p(s - \beta y) = p(s + \alpha y). \end{aligned}$$

This means we can find a linear extension of f to Y so that f is dominated by p.

There is a partial ordering on all linear extensions of f that are dominated by p. For such extensions g and h, the ordering sets $g \preceq h$ if the domain of g is contained in the domain of h and $h = g$ on the domain of g. By the Hausdorff Maximal Principle (see the appendix on the Axiom of Choice), there is a maximal family $\mathscr{F}$ of these extensions of f such that the restriction of the ordering $\preceq$ to $\mathscr{F}$ is a linear ordering. Let Z be the union of the domains of the extensions that are in $\mathscr{F}$. Given each $z \in Z$, set $G(z) = g(z)$ for any $g \in \mathscr{F}$ such that $g(z)$ is defined. By definition of the ordering $\preceq$ on $\mathscr{F}$, the function G is well-defined. Given z, $w \in Z$ and α and β in $\mathbb{R}$, there is a $g \in \mathscr{F}$ such that $g(z)$ is defined, and there is an $h \in \mathscr{F}$ such that $h(w)$ is defined. Suppose $h \preceq g$. Then $g(w)$ is defined, $\alpha z + \beta w \in Z$, and $g(\alpha z + \beta w) = \alpha g(z) + \beta g(w)$. A similar conclusion holds for h if $g \preceq h$. It follows that Z is a linear space and G is linear on Z; moreover, $G \leq p$ on Z. If there were a point $w \in X \setminus Z$, we could extend G to the space generated by Z and w as before, and then add the extension to $\mathscr{F}$. Since $\mathscr{F}$ is maximal on which the ordering $\preceq$ is a linear ordering, no such w can exist, that is, $Z = X$.

We follow with some applications of the theorem for a normed linear space X.

Lemma 11.4.1. *If F is a linear functional on a normed space X, then $F(x) \leq \|x\|$ for all $x \in X$ if and only if $\|F\| \leq 1$.*

Proof. Exercise 11.8(A).

Proposition 11.4.1. *Given a normed linear space $X \neq \{\mathbf{0}\}$, for each $w \neq 0$ in X, there exists an F in X^* with $\|F\| = 1$ such that $F(w) = \|w\|$.*

Proof. Fix $w \neq 0$ in X. Let S be the subspace generated by w. For each $\alpha \in \mathbb{R}$, set $f(\alpha w) := \alpha\|w\|$. This is a linear map on S, and since $\alpha\|w\| \leq \|\alpha w\|$, the hypotheses of the Hahn-Banach theorem are satisfied with p equal to the norm on X. Let F be the extension of f from S to X such that for each $x \in X$, $F(x) \leq \|x\|$. By Lemma 11.4.1, $\|F\| \leq 1$, and since $F(w) = f(w) = \|w\|$, we have $|F(w)| / \|w\| = 1$, so $\|F\| = 1$.

Proposition 11.4.2. *Suppose X is a normed linear space and S is a subspace such that the closure $\overline{S} \neq X$. Fix $y \in X \setminus \overline{S}$, and let $\delta = \inf_{s \in S} \|y - s\|$. Then there is an $F \in X^*$ such that $F(s) = 0$ for all $s \in S$, $\|F\| \leq 1$, and $F(y) = \delta$.*

Proof. On the subspace spanned by S and y, let $f(\alpha y + s) = \alpha\delta$ for all $\alpha \in \mathbb{R}$ and $s \in S$. Then f is linear, $f(s) = 0$ for all $s \in S$, and $f(y) = \delta$. For $\alpha \neq 0$ and $s \in S$,

$$f(\alpha y + s) = \alpha\delta \leq |\alpha| \cdot \|y - (-s/\alpha)\| = \|\alpha y + s\|,$$

so in general, $f \leq \|\cdot\|$ on the subspace spanned by S and y. Extend f to F on all of X with $F \leq \|\cdot\|$. By Lemma 11.4.1, $\|F\| \leq 1$.

Proposition 11.4.3. *For Lebesgue measure λ on $\mathbb{R}$, the dual space of $L^\infty[0,1]$ is not $L^1[0,1]$.*

Proof. Let $C[0,1]$ denote the space of real-valued continuous functions on $[0,1]$ supplied with the norm inherited as a subspace of $L^\infty[0,1]$. Let $g(f)$ be the linear functional given by $g(f) = f(1)$ for each $f \in C[0,1]$. With p equal to the norm on $L^\infty[0,1]$, the hypotheses of the Hahn-Banach theorem are satisfied. Extend g to a bounded linear functional G defined on all of $L^\infty[0,1]$. If $h \in L^1[0,1]$, then for each $n \in \mathbb{N}$, setting $f_n : x \mapsto x^n \in C[0,1]$, $h \cdot f_n \in L^1[0,1]$, $G(f_n) = 1$, but $\lim_n f_n(x) = 0$ a.e. It follows from the Lebesgue Dominated Convergence Theorem that

$$1 = \lim_n G(f_n) \neq \lim_n \int h \cdot f_n = 0.$$

Remark 11.4.1. As we shall show when we consider the dual space of $C[0,1]$, the functional g on $C[0,1]$ is represented by a measure, in this case unit mass at 1. With additional tools, one can show that the functional G on L^∞ is represented by a finitely additive measure.

11.5 Imbedding Into a Second Dual

Definition 11.5.1. Let X be a Banach space, and let X^* be the dual space of X. The dual space X^{**} of X^* is called the **second dual** of X.

Theorem 11.5.1. *There is a natural isometric isomorphism φ mapping a Banach space X into X^{**}. It is defined for each $x \in X$ and $f \in X^*$ by setting $\varphi(x)(f) = f(x)$. It follows that $\varphi[X]$ is a closed linear subspace of X^{**}.*

Proof. We may assume X contains nonzero elements. Fix any x and y in X, f and g in X^*, and α and β in $\mathbb{R}$. Then

$$\begin{aligned}
\varphi(x)(\alpha f+\beta g) &= (\alpha f+\beta g)(x) = \alpha f(x)+\beta g(x) = \alpha\varphi(x)(f)+\beta\varphi(x)(g).\\
\varphi(\alpha x+\beta y)(f) &= f(\alpha x+\beta y) = \alpha f(x)+\beta f(y)\\
&= \alpha\varphi(x)(f)+\beta\varphi(y)(f) = (\alpha\varphi(x)+\beta\varphi(y))(f).
\end{aligned}$$

Therefore, the mapping $x \mapsto \varphi(x)$ is linear on X, and for each $x \in X$, the mapping $\varphi(x)$ is linear on X^*. As a functional on X^*,

$$\|\varphi(x)\| = \sup_{f\in X^*, \|f\|=1} |\varphi(x)(f)| = \sup_{f\in X^*, \|f\|=1} |f(x)| \le \|x\|.$$

To show that the inequality is actually equality, we note that by Proposition 11.4.1, for each $x \neq 0$ in X, there is an $f \in X^*$ with $\|f\| = 1$ and $f(x) = \|x\|$. Therefore, for each $x \in X$, including $x = 0$, $\|\varphi(x)\| = \|x\|$; that is, φ is a linear isometry into X^{**}. To show that $\varphi[X]$ is closed, let $\langle\varphi(x_n)\rangle$ be a sequence in $\varphi[X]$ converging to an element $y \in X^{**}$. The corresponding sequence $\langle x_n\rangle$ is a Cauchy sequence in the Banach space X converging to some x. Since φ is an isometry, it is continuous, so $\varphi(x_n)$ has two limits; they are $\varphi(x)$ and y. Therefore so $y = \varphi(x)$. It follows that $\varphi[X]$ is closed in X^{**}.

Definition 11.5.2. A Banach space X is **reflexive** if the natural isomorphism φ mapping X into X^{**} is a surjection.

Example 11.5.1. Examples of reflexive spaces are the L^p spaces for $1 < p < \infty$. Since X^{**} is complete, a reflexive space X has to be a Banach space. For Lebesgue measure on $[0,1]$, it follows from Proposition 11.4.3 that $L^1[0,1]$ is not reflexive.

Proposition 11.5.1. *A Banach space X is reflexive if and only if its dual space X^* is reflexive.*

Proof. Exercise 11.9(A).

11.6 Properties of Banach Spaces

Theorem 11.6.1 (Uniform Boundedness Principle). *Let X be a Banach Space, and let $\mathscr{F}$ be a family of bounded linear mappings from X to a normed space Y. Suppose that for every $x \in X$, there is a constant M_x such that $\|Tx\| \le M_x$ for every $T \in \mathscr{F}$. Then there is a constant M with $\|T\| \le M$ for all $T \in \mathscr{F}$.*

Proof. We may assume $X \neq \{0\}$. For every $T \in \mathscr{F}$, the map $x \mapsto \|Tx\|$ is a continuous, real-valued function on X. By the general uniform boundedness principle 7.2.3, which is a consequence of the Baire Category Theorem, there is an open Ball $B(y,r) \subset X$ and a constant K such that $\|Tx\| \leq K$, for all $T \in \mathscr{F}$ and $x \in B(y,r)$. For every point z in the open ball $B(0,r)$ and each $T \in \mathscr{F}$,

$$\|Tz\| = \|T(z+y) - T(y)\| \leq \|T(z+y)\| + \|T(y)\| \leq K + M_y.$$

For points x with $\|x\| = 1$, $(r/2)x \in B(0,r)$, so

$$\|Tx\| = (2/r)\|T((r/2)x)\| \leq (2/r)[K + M_y].$$

Corollary 11.6.1. *Let X be a Banach Space, and let T be the pointwise limit of a sequence $\langle T_n : n \in \mathbb{N} \rangle$ of bounded linear maps from X into a normed linear space Y. Then T is a bounded linear map from X into Y.*

Proof. We have already seen in Theorem 11.1.2 that the pointwise limit of linear maps is linear. To show that the pointwise limit is bounded, we can use the Uniform Boundedness Principle as follows: For each $x \in X$, $T_n(x)$ converges, so $\|T_n(x)\|$ is bounded. Therefore, there is a constant M such that for all n, $\|T_n\| \leq M$, whence $\|T_n(x)\| \leq M$ for all x with $\|x\| = 1$. Since the norm is a continuous function on Y, $\|T(x)\| \leq M$ for all x with $\|x\| = 1$, and so $\|T\| \leq M$.

Definition 11.6.1. A mapping from a topological space X onto a topological space Y is called **open** if the direct image of an open subset of X is open in Y.

Remark 11.6.1. We have seen that a mapping f from X onto Y is continuous if the inverse image of each open set in Y is open in X. A continuous, open bijection is a homeomorphism. For the following results, it is important that the mappings are surjections.

Lemma 11.6.1. *Assume that X and Y are Banach spaces, and let T be a continuous linear map from X onto Y. Then for any $\delta > 0$, the image $T[B(0,\delta)]$ contains an open ball about 0 in Y.*

Proof. For each $n \in \mathbb{N} \cup \{0\}$, let $S_n = B(0, 1/2^n)$ in X. By scaling, $X = \cup_{k \in \mathbb{N}} kS_1$, and since T is a surjection, $Y = \cup_{k \in \mathbb{N}} kT[S_1]$. By the Baire Category Theorem, since Y is complete, $\overline{kT[S_1]}$ contains an open ball for some $k \in \mathbb{N}$. Using scaling, this means that $\overline{T[S_1]}$ contains an open ball $B_Y(p,\eta)$ in Y. In particular, the center p is in $\overline{T[S_1]}$. Therefore, $\overline{T[S_1]} - p$ contains the open ball $B_Y(0,\eta)$ in Y. Given any $y \in B_Y(0,\eta)$, the point y is the limit of a sequence $T(x_n) - T(z_n) = T(x_n - z_n)$ where $\langle x_n \rangle$ and $\langle z_n \rangle$ are sequences in S_1 with $z_n \to p$. But for each n, $\|x_n - z_n\| \leq \|x_n\| + \|z_n\| < 1$. Since $y \in \overline{T[S_0]}$, we have shown that $\overline{T[S_0]} \supseteq B_Y(0,\eta)$.

We now show, without taking a closure, that $T[S_0] \supseteq B_Y(0,\eta/2)$. It will follow by scaling that for any $\delta > 0$ and any n with $1/2^n < \delta$, $T[S_n] \supseteq B_Y(0,\eta/2^{n+1})$. Fix $y \in B_Y(0,\eta/2)$. Since $\overline{T[S_1]} \supseteq B_Y(0,\eta/2)$, there is a point $x_1 \in S_1$ with $\|y - T(x_1)\| < \eta/4$. Since $\overline{T[S_2]} \supseteq B_Y(0,\eta/4)$, there is a point $x_2 \in S_2$ with $\|(y - T(x_1)) - T(x_2)\| < \eta/8$. In this way, we choose a sequence $\langle x_n \rangle$ in X with $\|x_n\| < 1/2^n$ such that

$$\left\| y - \sum_{k=1}^{n} T(x_k) \right\| < \frac{\eta}{2^{n+1}}.$$

Now the series formed by the x_n's is absolutely convergent, and the sum x is in S_0. Since T is continuous,

$$T(x) = T\left(\sum_{k=1}^{\infty} x_k\right) = \sum_{k=1}^{\infty} T(x_k) = y.$$

This shows that $T[S_0] \supseteq B_Y(0, \eta/2)$.

Theorem 11.6.2 (Open Mapping). *A continuous linear mapping T from one Banach space X onto another Banach space Y maps open sets onto open sets. Thus, if the mapping is a bijection, so the inverse exists, that inverse is continuous.*

Proof. Let O be open in X, and fix $y \in T[O]$ and $x \in O$ with $T(x) = y$. Since O is open, there is an open ball centered at x, $B(x, \delta) \subseteq O$. Now $B(x, \delta) - x = B(0, \delta)$, and by Lemma 11.6.1, $T[B(0, \delta)] \supseteq B_Y(0, \gamma)$ in Y for some $\gamma > 0$. Therefore,

$$T[O] \supseteq T[B(x, \delta)] = T[B(0, \delta) + x] \supseteq B(0, \gamma) + y = B(y, \gamma).$$

It follows that $T[O]$ is open in Y.

Example 11.6.1. The projection map from the plane into the plane that sends each point (x, y) to the point $(x, 0)$ is bounded and linear. It is not, however, an open map from $\mathbb{R}^2$ to $\mathbb{R}^2$.

Proposition 11.6.1. *Let X be a linear space with two norms $\|\cdot\|$ and $|||\cdot|||$, for both of which X is complete. Assume that for some constant C, $\|\cdot\| \leq C|||\cdot|||$. Then the same is true with the roles of the two norms reversed, so they are equivalent.*

Proof. The identity map from $(X, |||\cdot|||)$ to $(X, \|\cdot\|)$ is a continuous surjection, so the inverse, which is also the identity map, is continuous, and therefore bounded.

Here is a formalization of the idea that a function is continuous if the graph has no breaks.

Theorem 11.6.3 (Closed Graph). *Let T be a linear mapping from a Banach space X into a Banach space Y, and assume that the graph of T is a closed set. That is, if $x_n \to x$ in X and $T(x_n) \to y$ in Y, then $y = T(x)$. It then follows that T is bounded.*

Proof. For each $x \in X$, set $|||x||| = \|x\| + \|T(x)\|$. By Problem 11.11, this is a norm on X for which X is complete. Of course, $\|\cdot\| \leq 1 \cdot |||\cdot|||$ on X, so by Proposition 11.6.1, there is a constant C such that $|||\cdot||| \leq C\|\cdot\|$. Therefore, if $\|x\| = 1$, then $\|T(x)\| \leq |||x||| \leq C$, so T is bounded.

Example 11.6.2. The following is an application of the Closed Graph Theorem: Let $(\Omega, \mathscr{A}, P)$ be a probability space. Let X be a Banach space and $Y := X^*$. Let φ be the natural isometric isomorphism of X into X^{**}. Given $x \in X$ and $y \in Y$, let $\varphi(x)(y)$ be denoted by $\langle x, y \rangle$. A function f from $(\Omega, \mathscr{A}, P)$ to Y is said to be Gelfand P-integrable if for each $x \in X$, the real-valued function $\langle x, f(\cdot) \rangle$ is integrable on $(\Omega, \mathscr{A}, P)$. It follows from the Closed Graph Theorem that there is a unique element $y \in Y$ such that $\langle x, y \rangle = \int_\Omega \langle x, f(\omega) \rangle \; dP$ for all $x \in X$; that element y is called the **Gelfand integral**. It is a special case of the Gelfand-Pettis integral.

11.7 Weak and Weak* Topologies

A weak topology on a topological space is formed from a family of continuous functions on the space. It is the weakest topology making the functions in the family continuous.

Definition 11.7.1. The **strong topology** on a normed space X is the metric topology generated by the norm. The **weak topology** on X is the weakest topology making the functions in the dual space X^* continuous on X. The **weak* topology**, also called the **vague topology**, is the weakest topology on X^* for which the functions $\varphi(x)$, $x \in X$, are continuous on X^*. Here φ is the natural isometric isomorphism of X into X^{**}.

A typical weak neighborhood of a point $x \in X$ is given by a finite number of functionals $f_1, \cdots, f_n$ in X^* and an $\varepsilon > 0$. It is the set of all points $z \in X$ such that $|f_i(x) - f_i(z)| < \varepsilon$ for $1 \le i \le n$. A typical weak* neighborhood of a functional $f \in X^*$ is given by a finite number of points $x_1, \cdots, x_n$ in X and an $\varepsilon > 0$. It is the set of all functionals $g \in X^*$ such that $|\varphi(x_i)(g) - \varphi(x_i)(f)| = |g(x_i) - f(x_i)| < \varepsilon$ for $1 \le i \le n$. Recall that weakening a topology increases the corresponding closure of sets.

Proposition 11.7.1. *A linear subspace S of a normed space X is weakly closed if and only if it is strongly closed.*

Proof. If S is already weakly closed, it is strongly closed since a strong closure point is a weak closure point. That is, if $\|x_n - x\| \to 0$ in X, then for any $f \in X^*$, $\|f(x_n) - f(x)\| \to 0$. Assume S is strongly closed and $y \notin S$. Then the distance between y and S is $\delta > 0$. By Proposition 11.4.2, which is an application of the Hahn-Banach theorem, there is a continuous linear functional f taking the value 0 on S such that f takes the value δ at y. Since f helps determine the weak neighborhoods in X, y is not in the weak closure of S.

Example 11.7.1. For an example of a weak* topology, let X be the space $C(K)$ consisting of continuous real-valued functions on a compact set K. We will see later in this chapter that X^* is isometrically isomorphic to the space of signed "regular" Borel measures on K. A weak* neighborhood of such a measure μ is given by an

$\varepsilon > 0$ and a finite set $\{g_1, \cdots, g_n\} \subseteq C(K)$. A signed measure ν is in this neighborhood if for $1 \le i \le n$, $|\int_K g_i d\mu - \int_K g_i d\nu| < \varepsilon$. Convergence of measures in this topology is often called **weak convergence of measures**; from the functional analysis point of view, however, it should be called weak* convergence.

One reason the weak* topology is important is that a sequence of probability measures always has a weak* cluster point. Here is the general result stated for the closed unit ball in X^*. By Problem 11.12, scaling and translation are continuous operations in the weak* topology, so the following result is actually valid for any closed ball in X^*.

Theorem 11.7.1 (Alaoglu). *The closed unit ball $S^* := \{f \in X^* : \|f\| \le 1\}$ in the dual X^* of a normed linear space X is a compact subset of X^* when supplied with the weak* topology.*

Proof. For each $x \in X$, let I_x be the closed interval $[-\|x\|, \|x\|]$. If $f \in S^*$, then for each $x \in X$, $f(x) \in I_x$. Recall that the product $\Pi_{x\in X} I_x$ is the set of functions on the set X, treated as an index set, with each function taking a value at x in I_x. Since the functions need not be linear functionals, the containment $S^* \subset \Pi_{x\in X} I_x$ is proper. By the Tychonoff Product Theorem 9.9.2, the product space is compact in the product topology. Moreover, the weak* topology on S^* is the restriction to S^* of the product topology. That is, a typical weak* neighborhood of a functional $f \in S^*$ is given by a finite number of points $x_1, \cdots, x_n$ in X and an $\varepsilon > 0$. It is the set of all functionals $g \in S^*$ such that $|g(x_i) - f(x_i)| < \varepsilon$ for $1 \le i \le n$. If S^* is closed in the product space, then it is compact. Therefore, we only have to show that if $h \in \overline{S^*}$, then $h \in S^*$. Given $\alpha \in \mathbb{R}$ and x and y in X, there is for every $\varepsilon > 0$ a function $f \in S^*$ such that $|h(z) - f(z)| < \varepsilon$ for $z \in \{x, y, x+y, \alpha x\} \subset X$. It follows that

$$\begin{aligned} |h(x)+h(y)-h(x+y)| &= |h(x)+h(y)-h(x+y)-(f(x)+f(y)-f(x+y))| \\ &\le |h(x)-f(x)|+|h(y)-f(y)|+|h(x+y)-f(x+y)| \\ &< 3\varepsilon \end{aligned}$$

$$\begin{aligned} |h(\alpha x)-\alpha h(x)| &= |h(\alpha x)-\alpha h(x)-(f(\alpha x)-\alpha f(x))| \\ &\le |h(\alpha x)-f(\alpha x)|+|\alpha h(x)-\alpha f(x)| \\ &< \varepsilon+|\alpha|\varepsilon. \end{aligned}$$

Since ε is arbitrary in $\mathbb{R}^+$, h is a linear mapping. Moreover, if $\|x\| = 1$, then $|h(x)| \le |h(x) - f(x)| + |f(x)| \le \varepsilon + 1$. Since ε is arbitrary, $h \in S^*$.

11.8 Functionals on Continuous Functions

An important space for many aspects of analysis is the space of continuous real-valued functions with compact support; that is, each function vanishes off of a compact set. A linear functional on such a space is positive if it yields a nonnegative value for nonnegative functions. Viewed as a space of measures, these functionals play a central role in many applications. Throughout this section, X is a compact, or at least locally compact, Hausdorff space. Recall that a space is locally compact if every point $x \in X$ is contained in an open set U with compact closure in X. It follows that if V is an open neighborhood of x and $V \subseteq U$, then $\overline{V}$ is also compact. Also recall that the collection of Borel sets in a topological space is the smallest σ-algebra containing the open sets.

Definition 11.8.1. The space of continuous real-valued functions with compact support on X is denoted by C_c. The notation $K \prec f$ means that $f \in C_c$, K is a compact subset of X, $0 \leq f \leq 1$ on X, and $f(x) = 1$ at every $x \in K$. The notation $f \prec V$ means that $f \in C_c$, V is an open set containing the compact support of f, and $0 \leq f \leq 1$ on X.

Remark 11.8.1. The combined notation $K \prec f \prec V$ means that $\chi_K \leq f \leq \chi_V$ and f has compact support contained in V. We establish next a version of Urysohn's Lemma 9.8.2 for a locally compact Hausdorff space. That result says that if K is compact and $V \supseteq K$ is open, then there is an $f \in C_c$ with $K \prec f \prec V$. First, we need to extend a "regularity" property to compact sets that are not just points. In doing so, we keep in mind the fact that X may be a disconnected space such as $\mathbb{N}$.

Proposition 11.8.1. *Let K be compact and U an open subset of X with $K \subseteq U$. There is an open set V with compact closure $\overline{V}$ such that $K \subseteq V \subseteq \overline{V} \subseteq U$.*

Proof. For each point $x \in K$, there is an open set W_x such that $\overline{W}_x \subseteq U$, and $\overline{W}_x$ is compact. To see this, fix $x \in K \subseteq U$. There is an open set O such that $x \in O$ and $\overline{O}$ is compact. Replace O with $O \cap U$; the closure is still compact. If $\overline{O} \subseteq U$, we are done; otherwise, note that, $\overline{O} \setminus O$ is compact. For each $y \in \overline{O} \setminus O$, there is a pair of disjoint open sets S_y and T_y with $y \in S_y$ and $x \in T_y$. Cover $\overline{O} \setminus O$ with a finite number of the sets S_y. The desired set W_x is the intersection of the corresponding sets T_y, since $\overline{W}_x \subseteq U$. Now the sets W_x cover K; take a finite subcover. The union of the open sets for this subcover is the desired set V.

Proposition 11.8.2 (Urysohn). *Let X be a locally compact Hausdorff space. Let K be compact and U an open subset of X with $K \subseteq U$. There is a continuous function f with $K \prec f \prec U$.*

Proof. By Proposition 11.8.1, there is an open set V with compact closure $\overline{V}$ such that $K \subseteq V \subseteq \overline{V} \subseteq U$. By Proposition 9.8.2, $\overline{V}$ is a normal subspace of X. It follows from Urysohn's Lemma 9.8.2 that there is a continuous function f defined on $\overline{V}$ taking values in $[0,1]$ such that $f(x) = 1$ for all $x \in K$, and $f(x) = 0$ for all $x \in \overline{V} \setminus V$. Set $f(x) = 0$ for all $x \in X \setminus \overline{V}$. If $y \in \overline{V} \setminus V$, then for any $\varepsilon > 0$, there is an open neighborhood W of y in X such that $0 \leq f(x) < \varepsilon$ for all $x \in W \cap \overline{V}$, and therefore, for all $x \in W$. It follows that f is continuous on X.

Engraved on money of the United States is the Latin "E Pluribus Unum", meaning "from many one." The following result is the reverse.

Theorem 11.8.1 (Partition of Unity). *Let $V_1,\cdots,V_n$ be a finite open covering of a compact set $K \subseteq X$. For each i, $1 \le i \le n$, there is an $h_i \prec V_i$ such that for all $x \in K$, $h = \Sigma_i h_i(x) = 1$, and $0 \le h \le 1$.*

Proof. For each $x \in K$, choose an i with $x \in V_i$ and an open neighborhood W_x of x with compact closure contained in V_i. Choose a finite subcovering $W_{x_1},\cdots,W_{x_m}$ of K. For each i, $1 \le i \le n$, let H_i be the union of the closures of those W_{x_j}'s that have compact closures contained in V_i, and fix a $g_i \in C_c$ with $H_i \prec g_i \prec V_i$. Let

$$h_1 = g_1, h_2 = (1-g_1)g_2, \cdots, h_n = (1-g_1)(1-g_2)\cdots(1-g_{n-1})g_n.$$

For each i, $h_i \le g_i$, $h_i \in C_c$, and $h_i \prec V_i$. It follows by induction that for each $k \le n$,

$$h_1 + h_2 + \cdots + h_k = 1-(1-g_1)(1-g_2)\cdots(1-g_k).$$

That is, $h_1 = 1-(1-g_1)$. If the formula holds for $k < n$, then

$$\begin{aligned} h_1 + h_2 + \cdots + h_k + h_{k+1} &= [1-(1-g_1)(1-g_2)\cdots(1-g_k)] \\ &\quad + [(1-g_1)(1-g_2)\cdots(1-g_k)g_{k+1}] \\ &= 1 + [(1-g_1)(1-g_2)\cdots(1-g_k)](g_{k+1}-1) \\ &= 1-(1-g_1)(1-g_2)\cdots(1-g_k)(1-g_{k+1}). \end{aligned}$$

Letting $k = n$, we see that for all $x \in K$, at least one of the values $g_i(x)$ is 1, so $h_1 + h_2 + \cdots + h_n = 1$ on K. Moreover, the sum for $k = n$ is positive and nowhere greater than 1.

Definition 11.8.2. A **positive linear functional** on C_c is a linear map of C_c into $\mathbb{R}$ taking nonnegative functions to nonnegative values. In particular, it is increasing in the sense that if $f \le g$, then the value at f is less than or equal to the value at g.

Definition 11.8.3. A **Radon measure** on X is a complete measure μ defined for sets in a σ-algebra $\mathscr{B}$ containing the Borel sets such that the following conditions hold:

a) for every compact set $K \subseteq X$, $\mu(K) < +\infty$;
b) (outer regularity) for every $E \in \mathscr{B}$, $\mu(E) = \inf\{\mu(V) : E \subseteq V, V \text{ open}\}$;
c) (restricted inner regularity) for every open set O, $\mu(O) = \sup\{\mu(K) : K \subseteq O, K \text{ compact}\}$, and
for every set $E \in \mathscr{B}$ with $\mu(E) < +\infty$, $\mu(E) = \sup\{\mu(K) : K \subseteq E, K \text{ compact}\}$.

Theorem 11.8.2 (Riesz Representation Theorem for C_c). *Let X be a locally compact (perhaps compact) Hausdorff space, and let C_c denote the space of continuous real-valued functions on X with compact support. Let Λ be a positive linear functional on C_c. There is a σ-algebra $\mathscr{B}$ containing the Borel sets and a unique Radon measure μ on $(X,\mathscr{B})$ such that μ represents Λ in the sense that for all $f \in C_c$, $\Lambda(f) = \int f\, d\mu$.*

We present the proof in several parts, much of which is derived from [46].

Proposition 11.8.3. *A Radon measure that represents Λ is unique.*

Proof. Let μ and ν be two measures that satisfy the properties of the theorem. Fix a compact set K, an $\varepsilon > 0$, and an open set $V \supseteq K$ with $\nu(V) < \nu(K) + \varepsilon$. Fix f with $K \prec f \prec V$. Then

$$\mu(K) = \int \chi_K \, d\mu \leq \int f \, d\mu = \Lambda(f) = \int f \, d\nu \leq \int \chi_V \, d\nu = \nu(V) < \nu(K) + \varepsilon.$$

Since ε is arbitrary $\mu(K) \leq \nu(K)$, and by symmetry, $\nu(K) \leq \mu(K)$, whence we have equality. It follows from Properties b and c that for every open set O, $\mu(O) = \nu(O)$, and so for every set $E \in \mathscr{B}$, $\mu(E) = \nu(E)$.

Definition 11.8.4. For each open set V, set $\mu(V) := \sup\{\Lambda f : f \prec V\}$. For each subset A of X, set $\mu^*(A) := \inf\{\mu(V) : A \subseteq V, \ V \text{ open}\}$.

Proposition 11.8.4. *The set function μ^* is an outer measure on X. For every open set $O \subseteq X$, $\mu^*(O) = \mu(O)$.*

Proof. If O is an open set, then O is the smallest open set containing O, so $\mu^*(O) = \mu(O)$. Since the empty set is both compact and open, the function identically equal to 0 has compact support. Since $\Lambda(0) = 0$, it follows that $\mu^*(\varnothing) = 0$. Clearly, μ^* is monotone increasing, that is, if $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$. It is only left to show that μ^* is countably subadditive.

First we show that μ is countably subadditive on any finite collection of open sets, and for that we need only consider two open sets V_1 and V_2. Fix $g \prec V_1 \cup V_2$. Employ Theorem 11.8.1 to find $h_1 \prec V_1$ and $h_2 \prec V_2$ so that $h_1 + h_2 = 1$ on the support of g. Now $g = gh_1 + gh_2$, and $gh_1 \prec V_1$, while $gh_2 \prec V_2$. It follows that $\Lambda(g) = \Lambda(gh_1) + \Lambda(gh_2) \leq \mu(V_1) + \mu(V_2)$, whence $\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2)$.

Let $\langle E_i \rangle$ be a finite or countably infinite sequence of subsets of X. We must show that $\mu(\cup_i E_i) \leq \Sigma_i \mu(E_i)$. We may assume that $\mu(E_i) < +\infty$ for each i. Fix $\varepsilon > 0$ and open sets $V_i \supseteq E_i$ such that $\mu(V_i) < \mu^*(E_i) + \varepsilon/2^i$. Let $V = \cup V_i$; then, $\cup_i E_i \subseteq V$. Fix $f \prec V$. Since f has compact support, there is an $n \in \mathbb{N}$ with $f \prec V_1 \cup \cdots \cup V_n$. Therefore,

$$\Lambda(f) \leq \mu(V_1 \cup \cdots \cup V_n) \leq \Sigma_{i=1}^{n} \mu(V_i) \leq \Sigma_{i=1}^{\infty} \mu^*(E_i) + \varepsilon.$$

Since f is arbitrary with $f \prec V$,

$$\mu^*(\cup_i E_i) \leq \mu(V) \leq \Sigma_{i=1}^{\infty} \mu(E_i) + \varepsilon.$$

Since ε is arbitrary, we are done.

Proposition 11.8.5. *For each compact set K,*

$$\mu^*(K) = \inf\{\Lambda(f) : K \prec f\} < +\infty.$$

For each open set V,

$$\mu(V) = \sup\{\mu^*(K) : K \subseteq V, K \textit{ compact}\}.$$

Proof. Let K be compact; we may assume $K \neq \varnothing$. Fix $f \in C_c$ with $K \prec f$, and fix α with $0 < \alpha < 1$. Let $V_\alpha = \{f > \alpha\}$. Then $K \subseteq V_\alpha$. Moreover, if $g \prec V_\alpha$, then $g \leq \frac{1}{\alpha} f$. Therefore,

$$\mu^*(K) \leq \mu(V_\alpha) = \sup\{\Lambda(g) : g \prec V_\alpha\} \leq \frac{1}{\alpha}\Lambda(f) < +\infty.$$

Letting $\alpha \to 1$, we see that $\mu^*(K) \leq \Lambda(f)$. Now suppose V is any open set with $K \subseteq V$. Given $f \in C_c$ with $K \prec f \prec V$, we now have $\mu^*(K) \leq \Lambda(f) \leq \mu(V)$, whence $\mu^*(K) = \inf\{\Lambda(f) : K \prec f\}$.

Finally, let V be any open subset of X. Fix $f \prec V$, and let K be the support of f. Let W be an arbitrary open set with $K \subseteq W$. We may assume that $W \subseteq V$. Now $f \prec W$, so $\Lambda f \leq \mu(W)$. It follows that $\Lambda(f) \leq \mu^*(K) \leq \mu(V)$. Since f is arbitrary with $f \prec V$,

$$\mu(V) = \sup\{\Lambda(f) : f \prec V\} = \sup\{\mu^*(K) : K \subseteq V, K \text{ compact}\}.$$

Definition 11.8.5. We set $\mathscr{A}$ equal to the collection of all $E \subseteq X$ for which $\mu^*(E) < +\infty$ and $\mu^*(E) = \sup\{\mu^*(K) : K \subseteq E,\ K \text{ compact}\}$. We write μ for μ^* on $\mathscr{A}$.

Remark 11.8.2. The collection $\mathscr{A}$ contains every open set V with $\mu(V) < +\infty$. Since a compact set K is its largest compact subset, the collection $\mathscr{A}$ contains every compact set K. The set function μ is both outer regular and inner regular on $\mathscr{A}$.

Proposition 11.8.6. *If $\langle E_i \rangle$ is a finite or countably infinite disjoint sequence of sets in $\mathscr{A}$, and $E = \cup_i E_i$, then $\mu(E) = \Sigma_i \mu(E_i)$, and $E \in \mathscr{A}$ if $\mu(E) < +\infty$.*

Proof. First we prove the result for a finite family of pairwise disjoint compact sets. We need only do so for two disjoint compact sets K_1 and K_2. Fix f_1 and f_2 with $K_1 \prec f_1 \prec X \setminus K_2$, $K_2 \prec f_2 \prec X \setminus K_1$, and $f_2 \leq 1 - f_1$. Fix $\varepsilon > 0$ and g with $K_1 \cup K_2 \prec g$ and $\Lambda(g) \leq \mu(K_1 \cup K_2) + \varepsilon$. Now,

$$\mu(K_1) + \mu(K_2) \leq \Lambda(f_1 \cdot g) + \Lambda(f_2 \cdot g) = \Lambda((f_1 + f_2) \cdot g) \leq \Lambda(g) \leq \mu(K_1 \cup K_2) + \varepsilon.$$

Since ε is arbitrary and μ is subadditive on $\mathscr{A}$, $\mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2)$.

For the general result, we note that we may form an infinite sequence from a finite one by filling in with empty sets. If $\mu(E) = +\infty$, we are done. Assume $\mu(E) < +\infty$, and fix $\varepsilon > 0$ and compact sets $H_i \subseteq E_i$ with $\mu(H_i) > \mu(E_i) - \varepsilon/2^i$. For each $n \in \mathbb{N}$, let $K_n = \cup_{i=1}^n H_i$. Then

$$\mu(E) \geq \mu(K_n) = \Sigma_{i=1}^n \mu(H_i) \geq \Sigma_{i=1}^n \mu(E_i) - \varepsilon.$$

Since n and ε are arbitrary and μ is subadditive,

$$\mu(E) = \Sigma_{i=1}^\infty \mu(E_i) = \lim_{n\to\infty} \Sigma_{i=1}^n \mu(E_i) = \lim_{n\to\infty} \mu(K_n).$$

It follows that $E \in \mathscr{A}$.

Proposition 11.8.7. *A set E is in $\mathscr{A}$ if and only if for each $\varepsilon > 0$, there is a compact set K and an open set V with $K \subseteq E \subseteq V$ and $\mu(V \setminus K) < \varepsilon$.*

Proof. Fix $E \in \mathscr{A}$. For any $\varepsilon > 0$, there is a compact set K and an open set V with $K \subseteq E \subseteq V$ such that $\mu(V) - \varepsilon/2 < \mu(E) < \mu(K) + \varepsilon/2$. Since $V \setminus K$ is an open set of finite measure, $V \setminus K \in \mathscr{A}$, and so by Proposition 11.8.6, $\mu(K) + \mu(V \setminus K) = \mu(V) < \mu(K) + \varepsilon$, whence $\mu(V \setminus K) < \varepsilon$.

Now fix $E \subseteq X$ such that for any $\varepsilon > 0$ there is a compact set K and an open set V with $K \subseteq E \subseteq V$ and $\mu(V \setminus K) < \varepsilon$. Again, $V \setminus K \in \mathscr{A}$ and $\mu(V \setminus K) + \mu(K) = \mu(V)$. Therefore,

$$\mu(K) \leq \mu^*(E) \leq \mu(V) < \mu(K) + \varepsilon.$$

Since ε is arbitrary, $E \in \mathscr{A}$.

Proposition 11.8.8. *If $A \in \mathscr{A}$ and $B \in \mathscr{A}$, then $A \setminus B$, $A \cup B$, and $A \cap B$ are in $\mathscr{A}$.*

Proof. Given $\varepsilon > 0$, fix compact sets H, K and open sets U, V such that

$$H \subseteq A \subseteq U, \quad K \subseteq B \subseteq V, \quad \mu(U \setminus H) < \varepsilon/2, \quad \text{and} \quad \mu(V \setminus K) < \varepsilon/2.$$

Then

$$H \setminus V \subseteq A \setminus B \subseteq U \setminus K \subseteq (U \setminus H) \cup (H \setminus V) \cup (V \setminus K).$$

Since $H \setminus V$ is a compact subset of $A \setminus B$ with

$$\mu(H \setminus V) \leq \mu(A \setminus B) \leq \mu(H \setminus V) + \varepsilon,$$

$A \setminus B \in \mathscr{A}$. By Proposition 11.8.6, $A \cup B = (A \setminus B) \cup B \in \mathscr{A}$. Since $A \cap B = A \setminus (A \setminus B)$, $A \cap B \in \mathscr{A}$.

Definition 11.8.6. We set $\mathscr{B}$ equal to the collection of all sets $E \subseteq X$ such that for any compact set $K \subseteq X$, $E \cap K \in \mathscr{A}$.

Proposition 11.8.9. *The collection $\mathscr{A} = \{E \in \mathscr{B} : \mu(E) < +\infty\}$.*

Proof. If $E \in \mathscr{A}$, then by Proposition 11.8.8, for any compact set $K \subseteq X$, $E \cap K \in \mathscr{A}$, whence $E \in \mathscr{B}$. Now fix $E \in \mathscr{B}$ with $\mu(E) < +\infty$. Fix $\varepsilon > 0$ and an open set $V \supseteq E$ with $\mu(V) < +\infty$. Fix a compact set $K \subseteq V$ with $\mu(V \setminus K) < \varepsilon/2$. Since $E \cap K \in \mathscr{A}$, there is a compact set $H \subseteq E \cap K \subseteq E$ with $\mu(E \cap K) < \mu(H) + \varepsilon/2$. Now, $E \subseteq (E \cap K) \cup (V \setminus K)$, so $\mu(E) \leq \mu(H) + \varepsilon$. Therefore, $E \in \mathscr{A}$.

Proposition 11.8.10. *The collection $\mathscr{B}$ is a σ-algebra containing the Borel sets.*

Proof. Fix an arbitrary compact set K in X. Given $E \in \mathscr{B}$, by definition $E \cap K \in \mathscr{A}$, so $(X \setminus E) \cap K = K \setminus (E \cap K) \in \mathscr{A}$. It follows that $X \setminus E \in \mathscr{B}$. Given a sequence $\langle E_n \rangle$ in $\mathscr{B}$ with $E = \cup_{i=1}^{\infty} E_i$, define $B_n \in \mathscr{A}$ using induction by setting $B_1 = E_1 \cap K$, and for every $n > 1$, set $B_n = (E_n \cap K) \setminus \cup_{i=1}^{n-1} B_i$. Then $E \cap K = \cup_{n=1}^{\infty} B_n \in \mathscr{A}$, so $E \in \mathscr{B}$. Therefore, $\mathscr{B}$ is a σ-algebra. If C is closed in X, then $C \cap K$ is compact, so $C \cap K \in \mathscr{A}$, whence $C \in \mathscr{B}$. It follows that the σ-algebra $\mathscr{B}$ contains the Borel sets.

Corollary 11.8.1. *For every set $E \in \mathscr{B}$ with $\mu(E) < +\infty$, $\mu(E) = \sup\{\mu(K) : K \subseteq E,\ K \text{ compact}\}$. Such an E equals a Borel set, in fact, a countable union of compact sets, to which a set of μ-measure 0 is adjoined.*

Proposition 11.8.11. *The set function μ is a complete, outer-regular measure on $\mathscr{B}$.*

Proof. Let $\langle E_i \rangle$ be a countable disjoint sequence of sets in $\mathscr{B}$, and let $E = \cup_i E_i$. If $\mu(E) = +\infty$, countable additivity follows from the countable subadditivity of μ. If $\mu(E) < +\infty$, then since E and each E_i belong to $\mathscr{A}$, countable additivity follows from Proposition 11.8.6. By definition, μ is outer regular in the sense that the value for any set is the infimum of the values of open supersets. Also, if $E \subset X$ and $\mu(E) = 0$, then $E \in \mathscr{A}$ and $E \in \mathscr{B}$. Therefore, the restriction of μ to $\mathscr{B}$ is a complete measure.

Here is the final step.

Proposition 11.8.12. *For every $f \in C_c$, $\Lambda(f) = \int f\, d\mu$.*

Proof. Let f be arbitrary in C_c. By showing that $\Lambda(f) \leq \int f\, d\mu$, it will follow that $\Lambda(-f) \leq -\int f\, d\mu$, whence $\Lambda(f) = \int f\, d\mu$. Let K be the support of f, and let $[a,b]$ be a finite interval containing the range of f. Fix $\varepsilon > 0$ and a finite set of points

$$y_0 < a < y_1 < \cdots < y_n = b$$

with $y_i - y_{i-1} < \varepsilon$ for each i. Let $E_i = \{x \in K : y_{i-1} < f(x) \leq y_i\}$ for $1 \leq i \leq n$. This gives a finite, pairwise disjoint, Borel measurable partition of K. For each i, choose an open set $V_i \supseteq E_i$ such that $\mu(V_i) < \mu(E_i) + \varepsilon/n$ and $f(x) < y_i + \varepsilon$ for all $x \in V_i$. Using Theorem 11.8.1, we fix $h_i \prec V_i$ for each i so that $\Sigma_i h_i(x) = 1$ for all $x \in K$ and $0 \leq \Sigma_i h_i \leq 1$. Now $f = \Sigma_i h_i f$, and $\mu(K) \leq \Lambda(\Sigma_i h_i) = \Sigma_i \Lambda(h_i)$. Moreover, $\Sigma_{i=1}^n \mu(E_i) = \mu(K)$, $h_i f \leq (y_i + \varepsilon) h_i$ for each i, and $y_i - \varepsilon < f(x)$ for all $x \in E_i$. Therefore,

$$\begin{aligned}
\Lambda(f) &= \Sigma_{i=1}^n \Lambda(h_i f) \\
&\leq \Sigma_{i=1}^n (y_i + \varepsilon)\Lambda(h_i) = \Sigma_{i=1}^n (|a| + y_i + \varepsilon)\Lambda(h_i) - |a|\Sigma_{i=1}^n \Lambda(h_i) \\
&\leq \Sigma_{i=1}^n (|a| + y_i + \varepsilon)[\mu(V_i)] - |a|\Lambda(\Sigma_{i=1}^n h_i) \\
&\leq \Sigma_{i=1}^n (|a| + y_i + \varepsilon)[\mu(E_i) + \frac{\varepsilon}{n}] - |a|\mu(K) \\
&= \Sigma_{i=1}^n (y_i + \varepsilon)\mu(E_i) + \frac{\varepsilon}{n}\Sigma_{i=1}^n (|a| + y_i + \varepsilon) \\
&= \Sigma_{i=1}^n (y_i - \varepsilon)\mu(E_i) + 2\varepsilon\mu(K) + \frac{\varepsilon}{n}\Sigma_{i=1}^n (|a| + y_i + \varepsilon) \\
&\leq \int f d\mu + \varepsilon[2\mu(K) + |a| + |b| + \varepsilon].
\end{aligned}$$

Since ε is arbitrary, the result follows.

Example 11.8.1. We define a distance ρ on the plane $\mathbb{R}^2$ as follows: For points (x_1,y_1) and (x_2,y_2) with $x_1 = x_2$, the distance is $|y_2 - y_1|$. For points (x_1,y_1) and (x_2,y_2) with $x_1 \neq x_2$, the distance is $|y_2 - y_1| + 1$. It is left to the reader (Problem 11.14) to show that ρ is indeed a metric on $\mathbb{R}^2$, and the space $\left(\mathbb{R}^2,\rho\right)$ is locally compact. Given $f \in C_c(X)$, there is a finite set A_f of x's such that for all $x \notin A_f$, $f(x,y) = 0$ for all values of y. This follows from the fact that a covering of the support of f by open balls of radius $1/2$ is a covering by vertical intervals. A finite subcover can contain only a finite number of such intervals. For each $f \in C_c(X)$, set

$$\Lambda f = \sum_{x\in A_f} \int_{-\infty}^{+\infty} f(x,y)\,dy.$$

Let μ be the measure associated with Λ given by the Riesz Representation Theorem 11.8.2. On any vertical line, the measure μ is generated by the integrator that is increasing length, and is therefore Lebesgue measure. Since any open subset containing the interval $[0,1]$ on the real line will contain an uncountable number of open vertical intervals having positive length, the measure of $[0,1]$ will be infinite. Any subset of $[0,1]$ that is compact in the topology generated by the metric ρ will contain only a finite number of points. The μ-measure of any point, and therefore of any finite set, is 0.

Remark 11.8.3. At this point, since we want a Borel measure to represent Λ, we have assumed outer regularity. Therefore, $[0,1]$ has infinite measure. If, instead, we assume inner regularity for all measurable sets, then the interval $[0,1]$ has measure 0. The choice does not change the measure on the **Baire sets**, i.e., the smallest σ-algebra making the functions in $C_c(X)$ measurable. The Baire sets form the smallest σ-algebra containing the compact G_δ sets. We have chosen to work with Radon measures rather than restrict our measures to the Baire sets as is done in some of the literature. For more details, see [45].

Definition 11.8.7. A Borel measure is **regular** if it is both inner and outer regular. That is, the measure of any measurable set is the supremum of the measures of compact sets it contains and the infimum of the measures of open sets containing it. A set is σ-compact if it is a countable union of compact sets.

Recall that a measure on a space X is σ-finite if X is a countable union of sets of finite measure. Example 11.8.1 raises the question as to when Borel measures are automatically regular. We give now a sufficient condition for σ-finite measures.

Proposition 11.8.13. *Suppose X is a second countable, locally compact Hausdorff space. Then every open set is σ-compact.*

Proof. Let $\mathscr{F}$ be a countable base for the topology on X. Let U be an open subset of X. By Proposition 11.8.1, every $x \in U$ is contained in an open set V_x with $x \in V_x \subset \overline{V}_x \subset U$ such that $\overline{V}_x$ is compact. We may assume, by taking a subset if necessary, that $V_x \in \mathscr{F}$. Therefore, the union of such compact sets $\overline{V}_x \subset U$ with $V_x \in \mathscr{F}$ is all of U, since no $x \in U$ can be outside the union. Moreover, since $\mathscr{F}$ is countable there can be no more than a countable number of such sets $\overline{V}_x$.

Theorem 11.8.3. *Suppose X is a second countable, locally compact Hausdorff space. Then every Borel measure that takes finite values on compact sets is σ-finite and regular.*

Proof. Let ν be such a measure. Let Λ be the functional on C_c it generates by integration, and let μ be the Radon measure representing Λ in the sense of Theorem 11.8.2. By Proposition 11.8.13, X is σ-compact, and therefore μ is σ-finite. We know that μ is outer regular. It follows from Problem 11.15 that μ is inner regular. We need to show $\mu = \nu$. Let U be a nonempty open subset of X. For each $f \in C_c$ with $U \succ f$, $\nu(U) \geq \int f d\nu = \Lambda(f)$, so $\nu(U) \geq \mu(U)$. Let K_n be an increasing sequence of compact sets with union equal to U, and let $\langle f_n \rangle$ be a sequence in C_c with $K_n \prec f_n \prec U$ for each n. Then

$$\nu(U) \geq \mu(U) \geq \sup \int f_n d\mu = \sup \Lambda(f_n) = \sup \int f_n d\nu \geq \sup \nu(K_n) = \nu(U).$$

Thus, ν and μ agree on open sets. Now let K be a compact set, and let U be an open set containing K with $\mu(U) < +\infty$. Since

$$\mu(U \setminus K) + \mu(K) = \mu(U) = \nu(U) = \nu(U \setminus K) + \nu(K),$$

$\mu(K) = \nu(K)$. Therefore, ν and μ agree on compact sets. Since X is σ-compact, we need only show that μ and ν agree on Borel subsets of compact sets. But these sets can be approximated in terms of μ from the inside by compact sets and from the outside by open sets. Therefore, $\mu = \nu$.

Definition 11.8.8. For a compact Hausdorff space X, we let $C(X)$ denote the space of continuous real-valued functions on X. There is a norm on $C(X)$, called the sup-norm. It is given by $f \mapsto \|f\| = \sup_{x\in X} |f(x)|$.

Proposition 11.8.14. *If X is a compact Hausdorff space, then a positive linear functional F on $C(X)$ is bounded, and therefore continuous.*

Proof. If for $f \in C(X)$, the sup-norm $\|f\| \leq 1$, then

$$|F(f)| = \left|F(f^+) - F(f^-)\right| \leq F(f^+) + F(f^-) = F(|f|) \leq F(1).$$

For a locally compact space that is not compact, positivity of a linear functional does not imply continuity with respect to bounded, continuous real-valued functions supplied with the sup-norm topology.

Example 11.8.2. Let μ be the measure on the natural numbers $\mathbb{N}$ such that $\mu(\{n\}) = n$ for each $n \in \mathbb{N}$. The integral with respect to μ is a positive linear functional on $C_c(\mathbb{N})$, i.e., the sequences that vanish after a finite number of entries. The sequence $\left\langle \frac{1}{n}\chi_{\{n\}} : n \in \mathbb{N} \right\rangle$ in $C_c(\mathbb{N})$ tends to 0 in the sup-norm, but the integral $\int \frac{1}{n}\chi_{\{n\}} d\mu = 1$ for each n.

Example 11.8.3. Let Λ be given by the Riemann integral on $C_c(\mathbb{R})$. Then Λ is a positive linear functional, and the representing measure is Lebesgue measure. For each $n \in \mathbb{N}$, fix f_n with $[-n,n] \prec f_n \prec (-n-1,n+1)$. The sup-norms of the sequence $\langle \frac{1}{n} f_n : n \in \mathbb{N} \rangle$ have limit 0, but for each n, $\Lambda(\frac{1}{n} f_n) \geq 2$.

For the rest of this section, we assume that X is a compact Hausdorff space. We have seen that a positive linear functional on $C(X)$ is bounded. We now show that a bounded linear functional is the difference of two positive linear functionals. Recall that for a compact space, a measure μ satisfying the conditions of Theorem 11.8.2 is finite and regular. Again we work with Radon measures, i.e., finite, complete, regular Borel measures, rather than restricting our measures to the Baire sets.

Example 11.8.4. Let A be an uncountable set with the discrete topology, i.e., each point of A forms an open set. Let X be the one-point compactification. That is, $X = A \cup \{p\}$ where open neighborhoods of p are the complements of finite subsets of X. Each continuous function f on X is constant off of some countable set. Mapping f to that constant value is a positive linear functional Λ. Unit mass at p is the Radon measure that represents Λ, but $\{p\}$ is not a Baire set. That is, it is not in the smallest σ-algebra containing the compact G_δ-sets.

Theorem 11.8.4. *Let X be a compact Hausdorff space. Let F be a bounded linear functional on $C(X)$. There are two unique, positive linear functionals F^+ and F^- with representing Radon measures μ^+ and μ^-, respectively, such that $F = F^+ - F^-$. That is, for $\mu := \mu^+ - \mu^-$ and for each $f \in C(X)$, $F(f) = \int f d\mu$. Moreover, (μ^+, μ^-) is the Jordan decomposition of μ. The σ-algebra of measurable sets is the completion of the Borel sets with respect to the total variation $|\mu| = \mu^+ + \mu^-$. The signed measure μ is the only regular (therefore finite) signed measure that represents F on that σ-algebra, and $\|F\| = |\mu|(X)$.*

Proof. For each nonnegative $f \in C(X)$, set $P(f) := \{\varphi \in C(X), 0 \leq \varphi \leq f\}$. Clearly, $f \in P(f)$, but F may not be a positive functional. Therefore,

$$F^+(f) := \sup_{\varphi \in P(f)} F(\varphi) \geq F(f).$$

On the other hand, for each $\varphi \in P(f)$, $|F(\varphi)| \leq \|f\| \cdot \|F\|$, so $0 \leq F^+(f) < +\infty$. Also, F^+ is positive and homogeneous; that is, for each constant $c \geq 0$ and each continuous f, $F^+(cf) = cF^+(f)$. Let f and g be two nonnegative continuous functions. If $\varphi \in P(f)$ and $\psi \in P(g)$, then $0 \leq \varphi + \psi \leq f + g$, and so $F(\varphi) + F(\psi) \leq F^+(f+g)$. It follows that $F^+(f) + F^+(g) \leq F^+(f+g)$. To reverse the inequality, we note that if $\psi \in P(f+g)$, then $\psi \leq (\psi \wedge f) + g$, and so $0 \leq \psi - (\psi \wedge f) \leq g$. Therefore,

$$F(\psi) = F(\psi \wedge f) + F(\psi - (\psi \wedge f)) \leq F^+(f) + F^+(g).$$

It follows that $F^+(f+g) \leq F^+(f) + F^+(g)$, and so we have equality. That is, F^+ preserves addition for nonnegative continuous functions.

Now if f is an arbitrary continuous function on X and $f \geq -a-b$ for positive constants a and b, then

$$F^+(f+a+b) = F^+(f+a) + F^+(b) = F^+(f+b) + F^+(a)$$
$$F^+(f+a) - F^+(a) = F^+(f+b) - F^+(b).$$

We can therefore set $F^+(f) = F^+(f+a) - F^+(a)$, and this is independent of the choice of a with $f \geq -a$. Note that this extension does not change the definition of F^+ for nonnegative continuous functions. Now for any continuous f and g on X, if $f \geq -a$ and $g \geq -b$ for nonnegative constants a and b, then

$$\begin{aligned} F^+(f+g) &= F^+(f+g+a+b) - F^+(a+b) \\ &= F^+(f+a) - F^+(a) + F^+(g+b) - F^+(b) = F^+(f) + F^+(g). \end{aligned}$$

That is, F^+ preserves addition for all continuous functions on X. Moreover, $F^+(0) = F^+(0) + F^+(0)$, so $F^+(0) = 0$. Also, for any continuous f, $F^+(f) + F^+(-f) = 0$, whence $F^+(-f) = -F^+(f)$. If c and a are positive and $f \geq -a$, then $cf \geq -ca$, so

$$F^+(cf) := F^+(cf + ca) - F^+(ca) = cF^+(f+a) - cF^+(a) = cF^+(f).$$

Moreover, $F^+((-c)f) = -F^+(cf) = -cF^+(f)$. It follows that F^+ is a positive linear functional on $C(X)$.

By Theorem 11.8.2, F^+ is represented by a unique Radon measure μ^+. Set $F^- := F^+ - F$. For a continuous $f \geq 0$, $F^+(f) \geq F(f)$, so F^- is also a positive linear functional on $C(X)$, and it is represented by a unique Radon measure μ^-. Clearly, $F = F^+ - F^-$. Let $\mu := \mu^+ - \mu^-$. Then for each $f \in C(X)$, $F(f) = \int f d\mu$.

Let ν be another finite, signed, regular measure defined at least on the Borel sets such that for each $f \in C(X)$, $F(f) = \int f d\nu$. Let (ν^+, ν^-) be the Jordan decomposition of ν. We want to show that $\nu^+ = \mu^+$ and $\nu^- = \mu^-$. It will follow that if $F = G^+ - G^-$ is another decomposition of F with corresponding measures ν^+ and ν^- forming the Jordan decomposition of ν, then $F^+ = G^+$ and $F^- = G^-$.

We show first that for any continuous $f \geq 0$, $F^+(f) = \int f d\nu^+$. We may assume that $\|f\| = 1$. Now since ν^+ is a nonnegative measure,

$$F^+(f) = \sup_{\varphi \in P(f)} F(\varphi) = \sup_{\varphi \in P(f)} \int \varphi d\nu \leq \sup_{\varphi \in P(f)} \int \varphi d\nu^+ = \int f d\nu^+.$$

To show that the reverse inequality holds for the same f, we choose a Hahn decomposition. Recall that this is a pair of disjoint measurable sets A and B so that $X = A \cup B$, and $\nu^+(B) = \nu^-(A) = 0$. Given $\varepsilon > 0$, we choose compact sets K and H contained in A and B, respectively, so that $\nu^+(K) > \nu^+(A) - \varepsilon/2$ and $\nu^-(H) > \nu^-(B) - \varepsilon/2$. Choose ψ with $K \prec \psi \prec X \setminus H$. Then

$$
\begin{aligned}
\int_X f\,d\nu^+ &= \int_A f d\nu \le \int_K f d\nu + \varepsilon/2 = \int_K f\cdot\psi d\nu + \varepsilon/2 \\
&\le \int_A f\cdot\psi d\nu + \varepsilon/2 = \int_A f\cdot\psi d\nu + \varepsilon/2 + \int_{B\setminus H} f\cdot\psi d\nu - \int_{B\setminus H} f\cdot\psi d\nu \\
&\le \int_X f\cdot\psi d\nu + \varepsilon/2 + \nu^-(B\setminus H) \le \int_X f\cdot\psi d\nu + \varepsilon \le F^+(f) + \varepsilon.
\end{aligned}
$$

It follows that ν^+ represents F^+ on the nonnegative continuous functions, and therefore on all of the continuous functions. Since the measure representing F^+ is unique, $\nu^+ = \mu^+$, and it follows that $\nu^- = \mu^-$. This shows the uniqueness of μ, and the fact that (μ^+, μ^-) is the Jordan decomposition of μ.

It is clear that $\|F\| \le \|F^+\| + \|F^-\| = \int 1 d\mu^+ + \int 1 d\mu^- = |\mu|(X)$. To establish the inequality in the other direction, we fix $\varphi \in P(1)$. Then $|2\varphi - 1| \le 1$. Therefore, $\|F\| \ge F(2\varphi - 1) = 2F(\varphi) - F(1)$. Taking the supremum over all such φ, we have

$$\|F\| \ge 2F^+(1) - F(1) = F^+(1) + F^-(1) = |\mu|(X).$$

Corollary 11.8.2 (Riesz Representation for $C(X)$). *The dual space of $C(X)$ can be represented by the space of finite, signed, regular Borel measures on X with the norm defined by the total variation evaluated at X.*

As indicated in Example 11.7.1, the following corollary is a consequence of the Alaoglu Theorem 11.7.1.

Corollary 11.8.3. *Let X be a compact Hausdorff space. The space of regular, Borel probability measures μ on X, that is, with $\mu(X) = 1$, is compact in the weak* topology, i.e., the topology generated by $C(X)$.*

Example 11.8.5. **Harmonic functions** on the unit disk in the complex plane are real-valued, continuous functions such that the Laplacian $\Delta h = \frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} = 0$. Let D_r denote the open disk $\{z \in \mathbb{C} : |z| < r\}$, and let $D = D_1$. Let C_r be the circle $\{z \in \mathbb{C} : |z| = r\}$, and let $C = C_1$. Let $\mathbf{P}(z,x)$ be the Poisson Kernel $(|z|^2 - |x|^2)/|z-x|^2$, and let x_0 denote the origin. Let the space $\mathscr{H}^1$ consist of all positive harmonic functions on D taking the value 1 at x_0. It is well-known that every continuous function on C has a harmonic extension on D, but that not every harmonic function on D is obtained in this way. On the other hand, for each $h \in \mathscr{H}^1$, there is a probability measure ν_h on C such that for each $x \in D$,

$$h(x) = \int_C \mathbf{P}(z,x)\,\nu_h(dz).$$

Each $h \in \mathscr{H}^1$ when restricted to a circle C_r, $0 < r < 1$, produces a continuous function there. Let λ_r be uniform probability measure on C_r. It is essentially normalized Lebesgue measure with total mass 1 defined on C_r. For any $x \in D_r$, $h(x) = \int_{C_r} \mathbf{P}(z,x)h(z)\lambda_r(dz)$. Given $h \in \mathscr{H}^1$, the measures $h\cdot\lambda_r$, $0 < r < 1$, are probability measures on the closed unit disk. The space of such measures is compact in the weak* topology, and these measures have a weak* limit ν_h (also called a weak limit) as the radius r tends to 1. If $x \in D_s$, then for $|z| > s$, the Poisson Kernel $\mathbf{P}(z,x)$ is a continuous function of z out to and including at points of C. Therefore ν_h, which is the weak* limit of $h \cdot \lambda_r$ continues to integrate against the Poisson Kernel to produce the value of h at x.

Remark 11.8.4. Using the space of continuous functions with compact support as "test functions", we have obtained a great deal of information about spaces of measures as functionals. By requiring more structure for the test functions, that is, by reducing the space of test functions, we enlarge the space of functionals. Functionals in an appropriate setting on the space of infinitely differentiable functions with compact support form the space of distributions, also called generalized functions. A test function vanishes off of a compact set. Therefore, integration by parts is used in one dimension to define the action of the "derivative" of a distribution with a test function. That action is the negative of the action of the distribution itself against the derivative of the test function. Since derivation has a meaning in higher dimensions as well, distributions play an important role in the theory of ordinary and partial differential equations.

11.9 Problems

Problem 11.1. Show that the following norms on $\mathbb{R}^2$ are equivalent.

a) The ℓ^1-norm: $\|(x,y)\| = |x| + |y|$.
b) The ℓ^2-norm: $\|(x,y)\| = \sqrt{x^2+y^2}$.
c) The ℓ^∞-norm: $\|(x,y)\| = \max\{|x|,|y|\}$.

Problem 11.2 (A). . Show that a finite-dimensional linear subspace of a normed linear space X must be a closed subspace of X with respect to the topology generated by the norm.

Problem 11.3 (A). . Show that a linear functional f on a normed space X is bounded if and only if its kernel $K := \{x \in X : f(x) = 0\}$ is closed in X.

Problem 11.4. Prove Lemma 11.2.1.

Problem 11.5. Show that for a measure space $(X, \mathscr{B}, \mu)$ and L^1 functions f and g of strictly positive norm, $\|f+g\|_1 = \|f\|_1 + \|g\|_1$ if and only if $|f(x)+g(x)| = |f(x)| + |g(x)|$ for almost all x.

Problem 11.6. The space c consisting of convergent real-valued sequences is a subspace of $\ell^\infty(\mathbb{N})$.

a) Show that c is a Banach space. **Hint:** A closed subset of a complete space is complete.
b) Show that the space c_0 consisting of sequences in c that converge to 0 is a Banach space.

Problem 11.7. Using Lebesgue measure λ, define $G : L^1(\mathbb{R}) \mapsto \mathbb{R}$ by

$$G(f) = \sum_{n=1}^{\infty} \frac{1}{n} \int_{[n-1,n]} f \, d\lambda .$$

Show that G defines a bounded linear functional on $L^1(\mathbb{R})$ and find the norm of G. Then find $g \in L^\infty(\mathbb{R})$ such that $G = F_g$ as in Theorem 11.3.1.

Problem 11.8 (A). Prove Lemma 11.4.1.

Problem 11.9 (A). Show that a Banach space X is reflexive if and only if its dual space X^* is reflexive.

Problem 11.10. Let S be a linear subspace of the space $C[0,1]$ of continuous real-valued functions on $[0,1]$. Assume that S is closed with respect to the L^2-norm using Lebesgue measure. Show that for some constant K and all $f \in S$, $\|f\|_\infty \le K \|f\|_2$.

Problem 11.11. Let T be a linear mapping from a Banach space X into a Banach space Y, and assume the graph of T is a closed set. For each $x \in X$, set $|||x||| = \|x\| + \|T(x)\|$. Show that this is a norm on X for which X is complete.

Problem 11.12. Let X be a normed linear space. Show that for any $\alpha \in \mathbb{R}$ and any g in the dual space X^*, the operations $f \mapsto \alpha f$ and $f \mapsto f + g$ are continuous on X^* when X^* is supplied with the weak* topology.

Problem 11.13 (A). Let X be a Banach space, and let A be a weakly compact subset of X. That is, A is a subset of X compact in the topology generated by X^*. Use the canonical injection φ of X into the second dual X^{**} to show that there is an $M \ge 0$ in $\mathbb{R}$ such that $\|x\| \le M$ for all $x \in A$.

Problem 11.14. Show that the function ρ in Example 11.8.1 is a metric on $\mathbb{R}^2$, and $(\mathbb{R}^2, \rho)$ is locally compact.

Problem 11.15. Show that a Radon measure on a σ-compact, locally compact Hausdorff space is regular.

Problem 11.16 (A). Let X be a compact Hausdorff space, and let $\mathscr{F} = \{f_\alpha : \alpha \in \mathscr{I}\}$ be a family of continuous real-valued functions on X. Suppose there is a corresponding family $\{c_\alpha : \alpha \in \mathscr{I}\}$ of constants such that for each finite set $\{\alpha_1, \cdots, \alpha_n\}$ of indices, there is a signed Radon measure ν with total variation $|\nu|(X) \le 1$, such that $\int f_{\alpha_i} \, d\nu = c_{\alpha_i}$ for $1 \le i \le n$. Show that there is a measure ν that works this way for all of the indices at the same time.

Appendix A
Appendix on the Axiom of Choice

We repeat the following definition from Chapter 1.

Definition A.1. A **partial ordering** on a nonempty set E is a relation $\leq$ on E such that

1) $\forall\, a \in E$, $a \leq a$, (**reflexive property**) and
2) $\forall\, a, b, c \in E$, $a \leq b$ and $b \leq c \Rightarrow a \leq c$. (**transitive property**)
We sometimes write $b \geq a$ for $a \leq b$.
We call a partial ordering **antisymmetric** if
3) $\forall\, a, b \in E$, $a \leq b$ and $b \leq a \Rightarrow a = b$.
For an antisymmetric ordering we write $a < b$ for $a \leq b$ but $a \neq b$.
A partial ordering $\leq$ is called a **total ordering** if
4) $\forall\, a, b \in E$, either $a \leq b$ or $b \leq a$.
A total ordering $\leq$ is called a **linear ordering** if it is antisymmetric.

Recall that for a set E with a partial ordering $\leq$, a **maximal element** of E is an element z such that for each $y \in E$, y need not be related to z, but if $y \geq z$, then we also have $y \leq z$. For an antisymmetric ordering, this means that $y = z$.

As noted in Chapter 1, the **Axiom of Choice** states that if $\{S_\alpha : \alpha \in \mathscr{I}\}$ is a nonempty collection of nonempty sets, then there is a function $T : \mathscr{I} \to \bigcup_{\alpha \in \mathscr{I}} S_\alpha$ such that for every $\alpha \in \mathscr{I}$, $T(\alpha) \in S_\alpha$. If the sets are disjoint, we may think of a parliament. Recall the example given by Bertrand Russell in terms of pairs of shoes and pairs of socks: Given a finite or even countably infinite set of pairs of shoes, one can always pick one shoe from each pair, e.g., the left shoe. Given a finite collection of pairs of socks, one can pick one sock from each pair, but what happens with an infinite collection of pairs of socks? The Axiom of Choice says there is a set consisting of exactly one sock from each pair.

We start with a result which, as we shall show, is equivalent to the Axiom of Choice. It is a modification of results and proofs in [17]. We will call a nonempty family $\mathscr{C}$ of subsets of a set S a **chain** if for any A and $B \in \mathscr{C}$ either $A \subseteq B$ or $B \subseteq A$. We will let $\cup\mathscr{C}$ denote the union of the sets in a chain $\mathscr{C}$.

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2

Theorem A.1 (Hausdorff Maximal Principle for Linear Orderings). *Let $\leq$ be a partial ordering on a nonempty set E. Let F_0 be a nonempty subset of E that is linearly ordered with respect to $\leq$. There is a maximal (with respect to $\subseteq$) subset $F \subseteq E$ such that $F_0 \subseteq F$ and $\leq$ is a linear ordering on F.*

Proof. We let $\mathscr{F}$ be the family of all $F \subseteq E$ such that the restriction of the ordering $\leq$ to F is a linear ordering and $F \supseteq F_0$. Let $T(\varnothing) = F_0$, and for each $F \in \mathscr{F}$, let $T(F)$ be an element of $\mathscr{F}$ that strictly contains F if there is one, and otherwise let $T(F) = F$. Here we have used the Axiom of Choice to define T. We want to show that there is an $F \in \mathscr{F}$ with $T(F) = F$. By the definition of T, this F is a maximal linear ordered subset of E containing F_0.

Now, the union of the sets in a chain $\mathscr{C}$ consisting of members of $\mathscr{F}$ is again in $\mathscr{F}$. To see this, let x and y be members of this union. We have $x \in A \in \mathscr{C}$, and $y \in B \in \mathscr{C}$, and either $A \subseteq B$ or $B \subseteq A$. Suppose the first. Then both x and y are in B, so since $\leq$ is a linear ordering on B, either $x \leq y$ or $y \leq x$; if both then $x = y$. The same is true if $B \subseteq A$.

Let $\mathscr{W}$ be the collection of all chains $\mathscr{C}$ consisting of members of $\mathscr{F}$ such that for each $F \in \mathscr{C}$ with $F \supsetneq F_0$, $\cup\{A \in \mathscr{C} : A \subsetneq F\} \in \mathscr{C}$, and for each $F \in \mathscr{C}$ such that $F \neq \cup\{A \in \mathscr{C} : A \subsetneq F\}$, $F = T(\cup\{A \in \mathscr{C} : A \subsetneq F\})$. This means that T uses the Axiom of Choice to supply the next step in the containment after the union. Note that the singleton $\{F_0\}$ is in $\mathscr{W}$.

Given $\mathscr{C} \in \mathscr{W}$, we say that a chain $\mathscr{D}$ is an initial chain of $\mathscr{C}$ if $\mathscr{D} \in \mathscr{W}$, each $F \in \mathscr{D}$ is a member of the chain $\mathscr{C}$, and if $A \in \mathscr{C}$ and $A \subset F$, then $A \in \mathscr{D}$. The union $\mathscr{G}$ of a nonempty family of initial chains of $\mathscr{C}$ is an initial chain of $\mathscr{C}$. To see this, note that F_0 is in $\mathscr{G}$, and each member of $\mathscr{G}$ is a member of $\mathscr{C}$. Moreover, if A and B are members of $\mathscr{G}$, then as elements of the chain $\mathscr{C}$, either $A \subseteq B$ or $B \subseteq A$. Also, each $A \in \mathscr{G}$ is an element of an initial chain $\mathscr{D} \in \mathscr{W}$, so if $B \in \mathscr{C}$ and $B \subset A$, then $B \in \mathscr{D}$. It follows that $\mathscr{G} \in \mathscr{W}$.

If $\mathscr{C}$ and $\mathscr{D}$ are chains in $\mathscr{W}$, then either $\mathscr{C}$ is an initial chain of $\mathscr{D}$ or $\mathscr{D}$ is an initial chain of $\mathscr{C}$. To see this, let $\mathscr{G}$ be the union of all chains in $\mathscr{W}$ that are both initial chains of $\mathscr{C}$ and initial chains of $\mathscr{D}$. The singleton $\{F_0\}$ is such a chain. As before, $\mathscr{G} \in \mathscr{W}$, and $\mathscr{G}$ is the largest initial chain of both $\mathscr{C}$ and $\mathscr{D}$. Suppose $\mathscr{C} \neq \mathscr{G}$. Let $F = \cup\mathscr{G}$. If $F \notin \mathscr{G}$, let $\mathscr{G}' = \mathscr{G} \cup \{F\}$. If $F \in \mathscr{G}$, let $\mathscr{G}' = \mathscr{G} \cup \{T(F)\}$. Then $\mathscr{G}' \in \mathscr{W}$, and $\mathscr{G}'$ is strictly larger than $\mathscr{G}$ initial chain of $\mathscr{C}$. Similarly, if $\mathscr{D} \neq \mathscr{G}$, then $\mathscr{G}'$ is strictly larger than $\mathscr{G}$ initial chain of $\mathscr{D}$. Since we cannot have both, either $\mathscr{G} = \mathscr{C}$ or $\mathscr{G} = \mathscr{D}$, or both.

Now let $\mathscr{G}$ be the union of all chains in $\mathscr{W}$. Then $\mathscr{G} \in \mathscr{W}$. To see this, note that the singleton $\{F_0\}$ is in $\mathscr{G}$. If A and B are in $\mathscr{G}$, then $A \in \mathscr{C} \in \mathscr{W}$, and $B \in \mathscr{D} \in \mathscr{W}$. If $\mathscr{C}$ is an initial chain of $\mathscr{D}$, then both A and B are in $\mathscr{D}$, so either $A \subseteq B$ or $B \subseteq A$. Similarly, if $\mathscr{D}$ is an initial chain of $\mathscr{C}$, then both A and B are in $\mathscr{C}$, so either $A \subseteq B$ or $B \subseteq A$. Therefore, $\mathscr{G}$ is a chain. If $F \in \mathscr{G}$, then $F \in \mathscr{C}$ for some $\mathscr{C} \in \mathscr{W}$. Therefore, if $F \neq F_0$, $\cup\{A \in \mathscr{C} : A \subsetneq F\} \in \mathscr{C} \subseteq \mathscr{G}$. If $F \neq \cup\{A \in \mathscr{C} : A \subsetneq F\}$, then $F = T(\cup\{A \in \mathscr{C} : A \subsetneq F\})$. Since $\mathscr{G} \in \mathscr{W}$ and $\mathscr{G}$ is the union of all chains in $\mathscr{W}$, $\cup\mathscr{G} \in \mathscr{G}$. Since $\mathscr{G} \cup \{T(\cup\mathscr{G})\}$ must equal $\mathscr{G}$, $\cup\mathscr{G}$ is a maximal linearly ordered subset of E containing F_0.

Remark A.1. A similar proof, left to the reader, establishes the **Hausdorff Maximal Principle for total orderings**.

Theorem A.2. *The following principles are equivalent:*

1) **Axiom of Choice.**
2) **Hausdorff Maximal Principle for Linear Orderings.**
3) **Zorn's Lemma:** Let $\leq$ be a partial ordering on a nonempty set E. If every linearly ordered subset of E has an upper bound in E, then there is a maximal element with respect to $\leq$ contained in E.
4) **Well-Ordering Principle:** Every set can be ordered with a well-ordering.

Proof. **1) $\Rightarrow$ 2):** This is Theorem A.1.

2) $\Rightarrow$ 3): By the assumption, we can find a maximum (under containment) linearly ordered subset F of E. Let z be an upper bound of F. We will show that z is a maximal element of E. Consider any y such that $z \leq y$. We must show that we also have $y \leq z$; they do not have to be equal. Now since z is an upper bound of F, for each $x \in F$, $x \leq z \leq y$. So $F \cup \{y\}$ is totally ordered with respect to $\leq$. Since F is a maximal linearly ordered set in E, it must also be true that for some $x \in F$, $y \leq x$, whence, $y \leq z$.

3) $\Rightarrow$ 4): Given a nonempty set E, consider subsets F of E with a well-orderings $\leq_F$. Let $\mathscr{F}$ be the set of all such pairs $(F, \leq_F)$. (A given set may appear more than once but with different orderings. In this sense, the subscript on the ordering is a bit misleading.) We can put an antisymmetric, partial ordering $\preceq$ on $\mathscr{F}$ by setting $(F, \leq_F) \preceq (G, \leq_G)$ when

i) $F \subseteq G$,
ii) $\leq_G = \leq_F$ on $F \times F$, [i.e., if $x, z \in F$, then $x \leq_F z$, if and only if $x \leq_G z$], and
iii) for every pair (x, y) with $x \in F$ and $y \in G \setminus F$, $x \leq_G y$.

We wish to show that $\preceq$ is, in fact, an antisymmetric, partial ordering on $\mathscr{F}$. Clearly, for each element $(F, \leq_F)$ in $\mathscr{F}$, $(F, \leq_F) \preceq (F, \leq_F)$ [(iii) is vacuously satisfied]. Moreover, $(F, \leq_F) \preceq (G, \leq_G)$ and $(G, \leq_G) \preceq (H, \leq_H) \Rightarrow (F, \leq_F) \preceq (H, \leq_H)$. To see this, we note that Properties i and ii are clear, so we only check (iii). If $y \in H \setminus F$ and $x \in F$, then for the case $y \in H \setminus G$, we have $x \leq_H y$ since $x \in G$. If $y \in G \setminus F$, then we have $x \leq_G y$, so $x \leq_H y$ by (ii). To see that $\preceq$ is antisymmetric, we assume that we have both $(F, \leq_F) \preceq (G, \leq_G)$ and $(G, \leq_G) \preceq (F, \leq_F)$. By (i), $F = G$. By (ii), $\leq_G = \leq_F$.

Fix a subset $\mathscr{G} = \{(F_\alpha, \leq_F)\}$ of $\mathscr{F}$, linearly ordered with respect to $\preceq$. We set $G = \cup_\alpha F_\alpha$, and we define an ordering on G by setting $x \leq_G y$ if $x \leq_{F_\alpha} y$ for some α. By (ii), this ordering is well-defined. Moreover, it is a linear ordering. To see this, we note first that for each $x \in G$, there is an α with $x \in F_\alpha$ so $x \leq_{F_\alpha} x$, whence $x \leq_G x$. If $x \leq_G y$ and $y \leq_G z$, then for some α, x, y, and z all belong to F_α. By (ii), it follows that $x \leq_{F_\alpha} y$ and $y \leq_{F_\alpha} z$, so $x \leq_{F_\alpha} z$, whence $x \leq_G z$. If x, y are in G, then again for some α, x and y belong to F_α, so either $x \leq_{F_\alpha} y$ or $y \leq_{F_\alpha} x$, whence either $x \leq_G y$ or $y \leq_G x$; if both, $x = y$.

To show that $\leq_G$ is a well-ordering of G, we consider a subset $A \subseteq G$ with an element $x \in A$. Then $x \in F_\alpha$ for some α, and by (iii), every element of $A \cap F_\alpha$ is smaller than every element of $A \setminus F_\alpha$ with respect to $\leq_G$. Let a be the first element of $A \cap F_\alpha$. Since each element of $A \setminus F_\alpha$ is larger than a, a is the first element of A.

It now follows that $(G, \leq_G)$ is an upper bound of $\mathscr{G}$. By Zorn's Lemma, there is a maximum element $(H, \leq_H)$ in $\mathscr{F}$. We need only show that $H = E$. If not, then there is a $z \in E \setminus H$. We then consider the set $K = H \cup \{z\}$ and set $x \leq_K y$ if either $x, y \in H$ and $x \leq_H y$ or if $x \in H$ and $y = z$. Now $(H, \leq_H) \preceq (K, \leq_K)$, so by the maximality of $(H, \leq_H)$, $(H, \leq_H) = (K, \leq_K)$. This is a contradiction since $H \neq K$.

4) ⇒ 1): Given a nonempty family of nonempty sets A_α, well-order $\cup_\alpha A_\alpha$. For each α, let $f(\alpha)$ be the first element in A_α.

Remark A.2. Suppose $\leq$ is a partial ordering on a nonempty set E. To apply Zorn's Lemma, it is only necessary to find an upper bound for each linearly ordered subset. By looking at the set of points (x, y) in the plane with $0 \leq y \leq 1$ and $(x_1, y_1) \leq (x_2, y_2)$ if and only if $x_1 = x_2$ and $y_1 \leq y_2$, it is clear that one cannot hope to get a biggest element. If the ordering $\leq$ is an antisymmetric partial ordering, then the Hausdorff Maximal Principle for Total Orderings implies Zorn's Lemma for $\leq$. Recall that the ordering in the proof **3) ⇒ 4)** is antisymmetric.

Example A.1. A **filter** $\mathscr{F}$ of subsets of a set X is a nonempty collection of subsets such that $\varnothing \notin \mathscr{F}$, if A and B are in $\mathscr{F}$, then $A \cap B \in \mathscr{F}$, and if $A \in \mathscr{F}$ and $A \subseteq B \subseteq X$, then $B \in \mathscr{F}$. An example is the **Fréchet filter** in $\mathbb{N}$ consisting of complements of finite subsets of $\mathbb{N}$. A filter base $\mathscr{B}$ of subsets of X is a collection of subsets such that $\varnothing \notin \mathscr{B}$, and if A and B are in $\mathscr{B}$, then for some $C \in \mathscr{B}$, $C \subseteq A \cap B$. The filter generated by a filter base is the set of all supersets of sets in the filterbase. An **ultrafilter** in a set X is a filter of subsets of X such that no strictly larger collection of subsets of X forms a filter. Given an ultrafilter $\mathscr{U}$ of subsets of X, each nonempty $A \subset X$ is either in $\mathscr{U}$ or $X \diagdown A$ is in $\mathscr{U}$, but not both. To see this, note that if $B \in \mathscr{U}$ and $A \cap B = \varnothing$, then $B \subseteq X \diagdown A$, so $X \diagdown A \in \mathscr{U}$. If $A \cap B \neq \varnothing$ for each $B \in \mathscr{U}$, then the collection $\{A \cap B : B \in \mathscr{U}\}$ forms a filter base, and A is a superset of each element $A \cap B$. Since the generated filter must equal $\mathscr{U}$, $A \in \mathscr{U}$.

Theorem A.3. *Any filter $\mathscr{F}_0$ of subsets of X is contained in an ultrafilter $\mathscr{U}$ of subsets of X.*

Proof. Order the filters $\mathscr{F}$ of subsets of X by containment. Any containment chain of filters containing $\mathscr{F}_0$ has an upper bound, namely the filter formed by the union. By Zorn's Lemma, there is a maximal filter, i.e., an ultrafilter containing $\mathscr{F}_0$.

Remark A.3. Given an ultrafilter $\mathscr{U}$ of subsets of X, the set function that assigns the value 1 to each $A \in \mathscr{U}$ and the value 0 to $X \diagdown A$ is a finitely additive measure on the family of subsets of X. We may say a property is true almost everywhere on X if it is true on an element of $\mathscr{U}$. Using an ultrafilter such as one containing the Fréchet filter in $\mathbb{N}$, and calling two real-valued sequences equivalent if they are equal almost everywhere, one forms an ordered field containing the real numbers and also containing infinitesimal elements. See Appendix on Nonstandard Analysis.

Appendix B
Appendix on Limit and Covering Theorems

Many theorems in analysis and probability theory produce a Radon-Nikodým derivative as a limit. In this appendix we show that the desired result is established once one shows that the limit is 0 where the input is 0. A major consequence is a simple proof of the Lebesgue Differentiation Theorem for measures. That result needs a covering theorem, and that is the other focus of this appendix.

Covering theorems are used to show that a set E where a desired property fails has measure zero. One begins by covering each point $\mathbf{x}$ in E with a special kind of Borel set. For many covering theorems, the covering sets are closed balls $B(\mathbf{x}, r)$ with center $\mathbf{x}$ and radius $r > 0$. Given a Borel measure μ, the covering theorem produces a constant C that depends on the space and perhaps on the measure. It also produces a disjoint collection $\{A_i\}$ of covering sets so that the outer measure $\mu^*(E) \leq C \cdot \Sigma_i \mu(A_i)$. The result is used in working with maximal functions, and measure derivatives.

Theorem 5.2.1 is the optimal covering theorem for the real line. The best known covering theorem for both the real line and higher dimensions is that of Vitali. (See, for example, [9].) That result is applicable for balls and measures, like Lebesgue measure, that increase at a known rate as the radius of the ball is increased. For finite-dimensional normed vector spaces, however, there are better, more general results. These are presented in this appendix.

B.1 A General Limit Theorem

We will give the results of this section in terms of filters, filter bases, and the corresponding limits. The results, due to Jürgen Bliedtner and the author, are taken from [9]. Recall the following definitions.

Definition B.1.1. A collection $\mathscr{F}$ of subsets of a set X is a **filter base** if it does not contain the empty set and the intersection of any finite subcollection of $\mathscr{F}$ contains

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2

a set in $\mathscr{F}$. A metric space valued function f converges to a limit L along a filter base $\mathscr{F}$ if for any $\varepsilon > 0$, there is a set $A \in \mathscr{F}$ such that for all $x \in A$, the metric $\rho(f(x), L) < \varepsilon$.

In what follows, $(X, \mathscr{B})$ is a measurable space and M is the set of all nonnegative finite measures on $(X, \mathscr{B})$. We use the notation μ_E for the restriction of a measure μ to a measurable set E; i.e., $\mu_E(A) = \mu(A \cap E)$. We let $\mathbb{R}^+$ denote the nonnegative real numbers, $\mathbb{N}$ the natural numbers, and $\complement E$ the complement of a set E in X. If μ and η are measures, we write $\mu \leq \eta$ if $\mu(A) \leq \eta(A)$ for each $A \in \mathscr{B}$.

We work with a nonzero measure $\sigma \in M$, called a reference measure. The theorem here is stated for all of M, but it is also true for subclasses such as all multiples of σ by an $L^p(\sigma)$ density for some fixed p, $1 \leq p \leq \infty$. Our theorem is given in terms of a class $\mathscr{F}$ of linear functionals mapping M into $\mathbb{R}^+$ with $F(\sigma) > 0$ for each $F \in \mathscr{F}$. Moreover, each $F \in \mathscr{F}$ has the property that if $\mu \leq \eta$, then $F(\mu) \leq F(\eta)$. We associate with every $x \in X$, a filter base $\mathscr{F}(x)$ on $\mathscr{F}$.

Example B.1.1. In a bounded open set S in $\mathbb{R}^n$, we can let $\mathscr{F}$ be the set of functionals determined by balls, with σ a regular Borel measure giving nonzero weight to balls in S. That is, each $F \in \mathscr{F}$ is given by a ball B in the sense that for each Borel measure μ, $F(\mu) = \mu(B)$. The filter $\mathscr{F}(x)$ can consist of sets of balls, with a typical element of $\mathscr{F}(x)$ being all balls of radius at most R with center x. The limit theorem we are about to establish will then, together with a covering theorem, yield the **Lebesgue Differentiation Theorem** discussed below. A simple version, Theorem 5.4.1, was presented in Chapter 5.

Example B.1.2. Recall Example 11.8.5. Harmonic functions on the unit disk D in the complex plane are real-valued, continuous functions such that the Laplacian $\Delta h = \frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} = 0$. Let $\mathbf{P}(z,x)$ be the Poisson Kernel $(|z|^2 - |x|^2)/|z-x|^2$. For each finite measure μ on the boundary of the unit disk D, we can extend μ inside the disk with a harmonic function h_μ using the Poisson Kernel. For each $y \in D$, we let $F_y(\mu) = h_\mu(y)$. Now we have various boundary limit theorems at points x of the boundary that go by the names "Fatou", "Ratio Fatou", "non-tangential", and "fine". For example, when σ is a reference measure on the boundary, $\lim_{r \to 1+} h_\mu(rx)/h_\sigma(rx) = d\mu/d\sigma(x)$ for σ-a.e. x on the boundary. There are fatter sets of points that can be used to replace segments of the line from 0 to x. These give better limit theorems.

In what follows, we write $\lim\limits_{F, \mathscr{F}(x)} F(\mu)/F(\sigma) = a$ if for each $\varepsilon > 0$ there is a set $\mathscr{S} \in \mathscr{F}(x)$ such that for all $F \in \mathscr{S}$, $|F(\mu)/F(\sigma) - a| < \varepsilon$. Also, we write $d\mu/d\sigma$ for the Radon-Nikodým derivative of the absolutely continuous part of μ with respect to σ.

Theorem B.1.1. *The following are equivalent:*

(1) For all $\mu \in M$, $\lim\limits_{F, \mathscr{F}(x)} \frac{F(\mu)}{F(\sigma)} = \frac{d\mu}{d\sigma}(x)$ for σ-a.e. $x \in X$.

(2) Given $E \in \mathscr{B}$ with $\sigma(E) > 0$, and $\nu \in M$ with $\nu(E) = 0$, we have

$$\lim_{F,\mathscr{F}(x)} \frac{F(\nu)}{F(\sigma)} = 0 \text{ for } \sigma\text{-a.e. } x \in E.$$

(3) Given $E \in \mathscr{B}$ with $\sigma(E) > 0$, and $\nu \in M$ with $\nu(E) = 0$, for σ-a.e. $x \in E$ there is a set $\mathscr{S} \in \mathscr{F}(x)$ with $F(\nu)/F(\sigma) \le 1$ for all $F \in \mathscr{S}$, whence

$$\limsup_{F,\mathscr{F}(x)} \frac{F(\nu)}{F(\sigma)} \le 1 \text{ for } \sigma\text{-a.e. } x \in E.$$

Proof. $(1 \Rightarrow 3)$ Assume 1 holds, and fix $E \in \mathscr{B}$ with $\sigma(E) > 0$ and $\nu \in M$ with $\nu(E) = 0$. Then $\lim_{F,\mathscr{F}(x)} \frac{F(\nu)}{F(\sigma)} = \frac{d\nu}{d\sigma}(x) = 0$ for σ-a.e. $x \in E$. For each such point x, there is a set $\mathscr{S} \in \mathscr{F}(x)$ with $F(\nu)/F(\sigma) \le 1$ for all $F \in \mathscr{S}$.

$(3 \Rightarrow 2)$ Fix $E \in \mathscr{B}$ with $\sigma(E) > 0$, $\nu \in M$ with $\nu(E) = 0$, and $k \in \mathbb{N}$. Since $k\nu(E) = 0$, it follows from 3 that for σ-a.e. $x \in E$ there is a set $\mathscr{S} \in \mathscr{F}(x)$ such that for all $F \in \mathscr{S}$, $F(k\nu)/F(\sigma) \le 1$, whence $F(\nu)/F(\sigma) \le 1/k$. Since this is true for each $k \in \mathbb{N}$,

$$\lim_{F,\mathscr{F}(x)} \frac{F(\nu)}{F(\sigma)} = 0 \text{ for } \sigma\text{-a.e. } x \in E.$$

$(2 \Rightarrow 1)$ By assumption, 1 holds for any measure that is singular with respect to σ. Therefore, given a finite, nonnegative, integrable function h on X, we must show that for some measurable set A with $\sigma(A) = 0$ and for all $x \in X \setminus A$, $\lim_{F,\mathscr{F}(x)} \frac{F(h\sigma)}{F(\sigma)} = h(x)$. Choose an $n \in \mathbb{N}$, and partition $\mathbb{R}^+$ into intervals of length $1/4n$. Let E be the inverse image with respect to h of one of the intervals $[r, r + 1/4n]$. If $\sigma(E) = 0$, adjoin E to A. Assume $\sigma(E) > 0$. Now for any $x \in E$ and any $F \in \mathscr{F}$, we have

$$\begin{aligned}
&|F(h\sigma) - h(x) \cdot F(\sigma)| \\
&\le |F(h\sigma) - r \cdot F(\sigma)| + |r \cdot F(\sigma) - h(x) \cdot F(\sigma)| \\
&\le |F(h\sigma_{\complement E}) - r \cdot F(\sigma_{\complement E})| + |F(h\sigma_E) - r \cdot F(\sigma_E)| + (h(x) - r)F(\sigma) \\
&\le F(h\sigma_{\complement E}) + r \cdot F(\sigma_{\complement E}) + \frac{1}{4n} \cdot F(\sigma_E) + \frac{1}{4n} \cdot F(\sigma).
\end{aligned}$$

Dividing by $F(\sigma)$, we have

$$\left| \frac{F(h\sigma)}{F(\sigma)} - h(x) \right| \le \frac{F(h\sigma_{\complement E})}{F(\sigma)} + r \cdot \frac{F(\sigma_{\complement E})}{F(\sigma)} + \frac{1}{2n}.$$

By assumption, for σ-almost all $x \in E$,

$$\lim_{F,\mathscr{F}(x)} \frac{F(h\sigma_{\complement E})}{F(\sigma)} = 0, \text{ and } \lim_{F,\mathscr{F}(x)} r \cdot \frac{F(\sigma_{\complement E})}{F(\sigma)} = 0.$$

Therefore, for σ-a.e. $x \in E$, there is a set $\mathscr{S}_n(x) \in \mathscr{F}(x)$ such that $|F(h\sigma)/F(\sigma) - h(x)| < 1/n$ for all $F \in \mathscr{S}_n(x)$. We obtain the desired result by putting all of the sets of measure 0 together for the sets E corresponding to the partition, and repeating the operation for each $n \in \mathbb{N}$.

Example B.1.3. For measure differentiation on $\mathbb{R}^n$ with respect to closed balls $B(\mathbf{x}, r)$ and a finite reference measure σ, one wants to show that for all finite Borel measures μ,

$$\lim_{r \to 0+} \frac{\mu(B(\mathbf{x},r))}{\sigma(B(\mathbf{x},r))} = \frac{d\mu}{d\sigma}(\mathbf{x}) \quad \sigma\text{-a.e}$$

where $d\mu/d\sigma$ is the Radon-Nikodým derivative of the absolutely continuous part of μ with respect to σ. The above limit Theorem B.1.1 is employed in [9] (and with an application in [10]), to show that this result is established when the conditions of the theorem hold. For that proof here, fix a Borel set E in the support of σ and a finite Borel measure ν with $\nu(E) = 0$. Let F be the set of points $\mathbf{x} \in E$ such that $\nu(B(\mathbf{x},r)) \geq \sigma(B(\mathbf{x},r))$ for a sequence of values r with limit 0. Here is the proof using a covering theorem that the outer measure $\sigma^*(F) = 0$:

Fix $\varepsilon > 0$ and a nonempty compact set $K \subseteq \mathbb{R}^n \setminus E$ with $\nu(\mathbb{R}^n \setminus K) < \varepsilon/C$, where C is the covering theorem constant. By assumption, for each $\mathbf{x} \in F$, there is an $r_{\mathbf{x}} > 0$ such that $B(\mathbf{x}, r_{\mathbf{x}}) \subseteq \mathbb{R}^n \setminus K$, and $\sigma(B(\mathbf{x}, r_{\mathbf{x}})) \leq \nu(B(\mathbf{x}, r_{\mathbf{x}}))$. For the disjoint subcollection $\{B_n\}$ of balls given by the covering theorem,

$$\begin{aligned}\sigma^*(F) &\leq C \cdot \Sigma_n \sigma(B_n) \leq C \cdot \Sigma_n \nu(B_n)\\ &\leq C \cdot \nu(\mathbb{R}^n \setminus K) < \varepsilon.\end{aligned}$$

B.2 Besicovitch and Morse Covering Theorems

Let $(X, \|\cdot\|)$ be a normed vector space of dimension $d < \infty$ over the real numbers $\mathbb{R}$. The covering theorems of Besicovitch [8] and of Morse [38] hold for arbitrary Borel measures on X. These theorems have two parts. The first part uses transfinite induction, and the second part uses geometric reasoning to find an upper bound κ for the number of covering sets that can form what we shall now define as a τ-satellite configuration.

Definition B.2.1. Fix $\tau > 1$. Let $\{S_i : 1 \leq i \leq n\}$ be an ordered collection of subsets of X with each S_i having finite diameter $\Delta(S_i)$ and containing a point $\mathbf{a}_i$ in its interior, $\text{int}(S_i)$. We say that the ordered collection of sets S_i is in τ-satellite configuration with respect to the ordered set of points $\mathbf{a}_i$ if: **(i)** For all $i \leq n$, $S_i \cap S_n \neq \varnothing$ and **(ii)** For all pairs $i < j \leq n$, $\mathbf{a}_j \notin \text{int}(S_i)$ and $\Delta(S_j) < \tau \cdot \Delta(S_i)$.

Remark B.2.1. We will use the geometry of X to show that for classes of sets, even beyond convex sets, and for $1 < \tau \leq 2$, there is an upper bound κ to the value of n. This upper bound will also hold for manifolds and a τ-satellite configuration isometric to one in X. Therefore the following theorem can be extended to metric spaces somewhat more general than a normed linear space X.

Theorem B.2.1. *Let A be an arbitrary nonempty subset of X. With each point $\mathbf{a} \in A$, associate a set $S(\mathbf{a})$ containing $\mathbf{a}$ in its interior so that the diameters of the sets $S(\mathbf{a})$ have a finite upper bound. Fix τ with $1 < \tau \leq 2$, and suppose that there can be no more than $\kappa < \infty$ sets from the collection $\{S(\mathbf{a}) : \mathbf{a} \in A\}$ forming a τ-satellite configuration in X. Then for some $m \leq \kappa$, there are pairwise disjoint subsets $A_1, \ldots, A_m$ of A such that $A \subseteq \cup_{j=1}^{m} \cup_{\mathbf{a} \in A_j} \operatorname{int}(S(\mathbf{a}))$ and for each j, $1 \leq j \leq m$, the elements of the collection $\{S(\mathbf{a}) : \mathbf{a} \in A_j\}$ are pairwise disjoint.*

Proof. We give both an informal and a formal description of the steps of the proof.

[Informal: Since one can't necessarily choose an $S(\mathbf{a})$ with maximal diameter, we choose one with essentially the maximal diameter. That is, τ times the diameter is bigger than all the other diameters. After that, at any stage, choose a point not in the interior of any set chosen before with the corresponding set having essentially the largest diameter among the competing sets. Stop when A is covered.**]**

[Formal: Put a well-ordering on A. By moving the first element of A to the last position, we may assume that A has a terminating element in the ordering. Let $\mathscr{I} = A$ with its well-ordering. We will use $\mathscr{I}$ as an index set. Each nonempty subset B of A is well-ordered as a subset of $\mathscr{I}$. Let $T(B)$ be the first point $\mathbf{b} \in B$ with $\tau \cdot \Delta(S(\mathbf{b})) > \sup_{\mathbf{a} \in B} \Delta(S(\mathbf{a}))$. Form a one-to-one correspondence between an initial segment of $\mathscr{I}$ and a subcollection of A as follows: Set $B_1 = A$ and $\mathbf{a}_1 = T(B_1)$. Having chosen $\mathbf{a}_\alpha$ for $\alpha < \beta$ in $\mathscr{I}$, let $B_\beta = A \setminus \cup_{\alpha<\beta} \operatorname{int}(S(\mathbf{a}_\alpha))$. If $B_\beta \neq \varnothing$, set $\mathbf{a}_\beta = T(B_\beta)$. There exists a first $\gamma \in \mathscr{I}$ for which $B_\gamma = \varnothing$.**]**

[Informal and Formal: This gives a well-ordering $\prec$ on a subset A_c of A such that $A \subseteq \cup_{\mathbf{a} \in A_c} \operatorname{int}(S(\mathbf{a}))$ and for $\mathbf{a} \prec \mathbf{b}$ in A_c, we have $\mathbf{b} \notin \operatorname{int}(S(\mathbf{a}))$ and $\Delta(S(\mathbf{b})) < \tau \cdot \Delta(S(\mathbf{a}))$.**]**

[Informal: Choosing the first element $\mathbf{a}(1)$ of A_c, and then choose the next element in the ordering $\prec$ for which the set $S(\mathbf{a})$ does not intersect $S(\mathbf{a}(1))$, etc. In this way, we find a maximal subset A_1 of A_c for which the corresponding sets $S(\mathbf{a})$ are pairwise disjoint. Start again with $A_c \setminus A_1$, etc. This gives a sequence of sets $A_1, A_2 \cdots$.**]**

[Formal: Given any nonempty subset B of A_c, form a one-to-one correspondence between an initial segment of $\mathscr{I}$ and a subset $V(B)$ of B as follows. Set $B_1 = B$, and let $\mathbf{a}(1)$ be the first element (with respect to $\prec$) of B_1. Having chosen $\mathbf{a}(\alpha)$ for $\alpha < \beta$, let

$$B_\beta = \{\mathbf{b} \in B : \forall \alpha < \beta \text{ in } \mathscr{I},\ S(\mathbf{b}) \cap S(\mathbf{a}(\alpha)) = \varnothing\}.$$

If $B_\beta \neq \varnothing$, let $\mathbf{a}(\beta)$ equal the first element (with respect to $\prec$) of B_β. There exists a first $\eta \in \mathscr{I}$ for which $B_\eta = \varnothing$. Let $V(B) = \{\mathbf{a}(\alpha) : \alpha < \eta\}$.

Now for $i \geq 1$ in $\mathbb{N}$, form sets $A_i \subseteq A_c$ as follows. Set $A_1 = V(A_c)$. Having chosen A_i for $i < n$ in $\mathbb{N}$, let $B_n = A_c \setminus \cup_{i=1}^{n-1} A_i$. Stop if $B_n = \varnothing$. Otherwise, set $A_n = V(B_n)$.**]**

[Informal and Formal: Note that for any $\mathbf{b} \in A_c$ with $\mathbf{b} \notin \cup_{i=1}^{n-1} A_i$ and each i between 1 and $n-1$, there is a first (with respect to $\prec$) $\mathbf{a}_i \in A_i$ with $S(\mathbf{a}_i) \cap S(\mathbf{b}) \neq \varnothing$;

clearly, $\mathbf{a}_i \prec \mathbf{b}$ in A_c. It now follows that the set $\{S(\mathbf{a}_1), \ldots, S(\mathbf{a}_{n-1}), S(\mathbf{b})\}$ is in τ-satellite configuration with respect to the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}, \mathbf{b}\}$ when each set is given the ordering inherited from $\mathscr{I}$. Therefore, n cannot be larger than κ, i.e., for some $m \leq \kappa$, $A \subseteq \cup_{j=1}^{m} \cup_{\mathbf{a}\in A_j} \operatorname{int}(S(\mathbf{a}))$.]

Corollary B.2.1. *For any finite Borel measure μ on X, there is a j with $1 \leq j \leq m$ such that*

$$\mu^*(A) \leq \kappa \cdot \sum_{\mathbf{a}\in A_j} \mu(\operatorname{int}(S(\mathbf{a}))).$$

Proof. Take the first $j \leq m$ that maximizes the sum $\Sigma_{\mathbf{a}\in A_j} \mu(\operatorname{int}(S(\mathbf{a})))$.

Remark B.2.2. If we replace κ with 2κ, we may even take a finite subcollection of the collection $\{S(\mathbf{a}) : a \in A_j\}$.

For the **Besicovitch Covering Theorem**, each set $S(\mathbf{a})$ is a closed ball $B(\mathbf{a}, r)$. Note that the open balls cover A and the closed balls form the disjoint families. For the **Morse Covering Theorem**, the sets $S(\mathbf{a})$ are more general than balls. There is a constant $\lambda \geq 1$, so that for each $\mathbf{a} \in A$,

$$\mathbf{a} \in B(\mathbf{a}, r) \subseteq S(\mathbf{a}) \subseteq B(\mathbf{a}, \lambda \cdot r).$$

Moreover, $S(\mathbf{a})$ must be "starlike" with respect to each point in $B(\mathbf{a}, r)$. That is, for each point $\mathbf{x}$ in $S(\mathbf{a})$ and each point $\mathbf{y}$ in $B(\mathbf{a}, r)$, the line segment joining these two points is also in $S(\mathbf{a})$. Again, the interiors of the sets given by the Morse theorem cover A.

B.3 The Best Constant and Proof of Besicovitch's Theorem

Now, for $1 < \tau \leq 2$, we let $\kappa(\tau)$ denote the maximum number of closed balls that can form a τ-satellite configuration in the space X. Z. Füredi and the author showed in [19] that for values of τ close to 1, $\kappa(\tau)$ is a packing constant K; it is the best constant for the Besicovitch theorem in terms of all known proofs. An upper bound for K is 5^d where d is the dimension of X.

Fix a τ-satellite configuration. Recall that later centers are not in the interior of earlier balls. Take the last ball in the collection out of the ordering, translate, and scale, so that it becomes the ball $B(\mathbf{0}, 1)$. We then have for the remaining indices $1 \leq i \leq n$ and corresponding balls $B(\mathbf{c}_i, r_i)$:

1) $B(\mathbf{c}_i, r_i) \cap B(\mathbf{0}, 1) \neq \varnothing$,
2) for $1 \leq i < j \leq n$, $\|\mathbf{c}_i - \mathbf{c}_j\| \geq r_i > r_j/\tau$, and
3) $\|\mathbf{c}_i\| = \|\mathbf{c}_i - \mathbf{0}\| \geq r_i > 1/\tau$.

In what follows, we will also consider this configuration for $\tau = 1$, but then with the strict inequalities $>$ replaced by $\geq$. To begin, we let K be the packing constant equal to the maximum number of points that can be put into the closed ball $B(\mathbf{0}, 2)$

when one of the points is at $\mathbf{0}$ and the distance between distinct points is at least 1. Given such a packing of points in $B(\mathbf{0},2)$, by centering a ball of radius 1 about each of the points, one obtains a set of balls satisfying Conditions 1–3 for $\tau = 1$, and so $\kappa(1) \geq K$. It is easy to see that $\kappa(\tau)$ is an increasing function of τ. We will show that for τ sufficiently close to 1,

$$\kappa(\tau) = K, \quad \text{whence,} \quad K = \kappa(1) = \kappa(\tau).$$

First, a diagonalization argument shows that for some $\delta > 0$, K is also the maximum when the distance 1 is replaced by $1-\delta$. That is, if an additional point can be placed in $B(\mathbf{0},2)$ for $\tau = 1 - \frac{1}{n}$ for large $n \in \mathbb{N}$, then in the limit, an additional point can be placed in $B(\mathbf{0},2)$ for $\tau = 1$. Fix τ with $1 \leq \tau < 1 + \delta/4$. Since $\delta < 1$, $\tau < \frac{5}{4}$. Fix a set of balls satisfying Conditions 1–3. Set $\mathbf{b}_0 = \mathbf{0}$. Given $1 \leq i \leq n$, if $\|\mathbf{c}_i\| \leq 2$, we set $\mathbf{b}_i = \mathbf{c}_i$; if $\|\mathbf{c}_i\| > 2$, we replace $\mathbf{c}_i$ with $\mathbf{b}_i = (2/\|\mathbf{c}_i\|)\mathbf{c}_i$. That is, in the latter case we project $\mathbf{c}_i$ onto the surface of the ball $B(\mathbf{0},2)$. We now need only show that the distance between pairs of distinct points $\mathbf{b}_i$ and $\mathbf{b}_j$ is at least $1-\delta$. This proof consists of three cases.

I) $\|\mathbf{c}_i\| \leq 2$ and $\|\mathbf{c}_j\| \leq 2$. By Conditions 2 and 3, $\|\mathbf{b}_i - \mathbf{b}_j\| = \|\mathbf{c}_i - \mathbf{c}_j\| \geq 1/\tau$, and $1 - 1/\tau \leq \tau - 1 < \delta$, so $1/\tau > 1 - \delta$.

II) $\|\mathbf{c}_i\| \leq 2$ and $\|\mathbf{c}_j\| > 2$. Since $B(\mathbf{c}_j, r_j) \cap B(\mathbf{0},1) \neq \varnothing$, $B(\mathbf{b}_j,1) \subseteq B(\mathbf{c}_j, r_j)$. If $j < i$, then $\|\mathbf{c}_i - \mathbf{c}_j\| \geq r_j$, so $\|\mathbf{b}_i - \mathbf{b}_j\| \geq 1$, and we are done. The alternative is that $i < j$, so $\|\mathbf{c}_i - \mathbf{c}_j\| \geq r_i \geq r_j/\tau$. By Condition 3, $r_i \leq \|\mathbf{c}_i\| = \|\mathbf{b}_i\| \leq 2$, so $r_j \leq \tau \cdot r_i \leq 2 \cdot \tau$. Since $\tau < 5/4$,

$$r_j - r_j/\tau \leq 2\tau(1 - 1/\tau) \leq 2\tau(\tau - 1) < 2\tau \cdot \frac{\delta}{4} < \delta,$$

so $r_j - \delta < r_j/\tau \leq \|\mathbf{c}_i - \mathbf{c}_j\|$. By Condition 1,

$$\|\mathbf{b}_i - \mathbf{b}_j\| = \|\mathbf{c}_i - \mathbf{b}_j\| \geq \|\mathbf{c}_i - \mathbf{c}_j\| - \|\mathbf{c}_j - \mathbf{b}_j\| \geq (r_j - \delta) - (r_j - 1) = 1 - \delta.$$

III) $\|\mathbf{c}_i\| \geq \|\mathbf{c}_j\| > 2$. Set $s = \|\mathbf{c}_j\|$ and $\mathbf{x} = (s/\|\mathbf{c}_i\|) \cdot \mathbf{c}_i$. Now,

$$\|\mathbf{c}_i - \mathbf{c}_j\| \leq \|\mathbf{c}_i - \mathbf{x}\| + \|\mathbf{x} - \mathbf{c}_j\| = \|\mathbf{c}_i\| - \|\mathbf{c}_j\| + \|\mathbf{x} - \mathbf{c}_j\|,$$

so we have the "Bow and Arrow" Inequality

$$\|\mathbf{x} - \mathbf{c}_j\| \geq \|\mathbf{c}_j\| + \|\mathbf{c}_i - \mathbf{c}_j\| - \|\mathbf{c}_i\|.$$

By Condition 1, $\|\mathbf{c}_i\| \leq r_i + 1$, so

$$\|\mathbf{b}_i - \mathbf{b}_j\| = \|\frac{2}{s} \cdot \mathbf{x} - \frac{2}{s} \cdot \mathbf{c}_j\| = \frac{2}{s}\|\mathbf{x} - \mathbf{c}_j\| \geq \frac{2}{s}\left(s - 1 + \|\mathbf{c}_i - \mathbf{c}_j\| - r_i\right).$$

If $\|\mathbf{c}_i - \mathbf{c}_j\| - r_i \geq 0$, then since $2/s < 1$,

$$\|\mathbf{b}_i - \mathbf{b}_j\| \geq \frac{2}{s}(s-1) > 2 - 1 = 1$$

If $\|\mathbf{c}_i - \mathbf{c}_j\| < r_i$, then $j < i$. By Condition 3, $s \geq r_j$. By Condition 2, $\|\mathbf{c}_i - \mathbf{c}_j\| \geq r_j > r_i/\tau$, so

$$r_i - \|\mathbf{c}_i - \mathbf{c}_j\| \leq r_i - r_j < r_j(\tau - 1) \leq s(\tau - 1).$$

Since $1 \leq \tau < 1 + \delta/4$ and $\frac{2}{s} < 1$,

$$\begin{aligned}\|\mathbf{b}_i - \mathbf{b}_j\| &\geq \frac{2}{s}\left[s - 1 - (r_i - \|\mathbf{c}_i - \mathbf{c}_j\|)\right] \geq \frac{2}{s}\left[s - 1 - \tau s + s\right] = 4 - \frac{2}{s} - 2\tau \\ &> 3 - 2\tau > \ 3 - 2\cdot(1 + \delta/4) = 1 - \delta/2.\end{aligned}$$

Independent of the work with Füredi in [19], John Sullivan [50] obtained a similar bound for the case of Euclidean spaces. Reifenberg [43] and Bateman-Erdős [6] showed that for disks in the plane with the Euclidean norm, $\kappa(1) = K = 19$. In general, for any norm, $K \leq 5^d$ where d is the dimension of X. For the ℓ^∞ norm on $\mathbb{R}^d$, $K = 5^d$.

B.4 Proof of Morse Covering Theorem

Recall that there is a constant $\lambda \geq 1$ used to define the sets in the Morse Covering Theorem. One associates a set S with each $\mathbf{a} \in A$, so that for some $r > 0$, $\mathbf{a} \in B(\mathbf{a}, r) \subseteq S \subseteq B(\mathbf{a}, \lambda \cdot r)$, and S is starlike with respect to each point in $B(\mathbf{a}, r)$. If $\lambda = 1$, this is the Besicovitch result. The "starlike" condition means that for each $\mathbf{y} \in B(\mathbf{a}, r)$ and each $\mathbf{x} \in S$, the line segment $\alpha\mathbf{y} + (1-\alpha)\mathbf{x}$, $0 \leq \alpha \leq 1$, is contained in S. The proof we now give from [34] of the Morse theorem modifies arguments in [38] and [9].

Proposition B.4.1. *If $\|\mathbf{y} - \mathbf{a}\| < r$, i.e., if $\mathbf{y}$ is in the interior of $B(\mathbf{a}, r)$, and $\mathbf{x}$ is in the closure, $\mathrm{cl}(S)$, of S, then every point of the form $\alpha\mathbf{y} + (1-\alpha)\mathbf{x}$, $0 < \alpha \leq 1$, is in the interior of S.*

Proof. Fix $\rho > 0$ so that $B(\mathbf{y}, \rho) \subset B(\mathbf{a}, r)$, and fix α with $0 < \alpha < 1$. Assume first that $\mathbf{x} \in S$, and translate so that $\mathbf{x} = \mathbf{0}$. Then $\alpha y \in S$. Moreover, the ball $B(\alpha\mathbf{y}, \alpha\rho) \subseteq S$ since

$$\begin{aligned}\|\alpha\mathbf{y} - \mathbf{z}\| \leq \alpha\rho \Rightarrow \left\|\mathbf{y} - \tfrac{1}{\alpha}\mathbf{z}\right\| \leq \rho \Rightarrow \tfrac{1}{\alpha}\mathbf{z} \in B(\mathbf{a}, r) \\ \Rightarrow \mathbf{z} = \alpha\left(\tfrac{1}{\alpha}\mathbf{z}\right) + (1-\alpha)\mathbf{0} \in S.\end{aligned}$$

Now for the case that $\mathbf{x} \in \mathrm{cl}(S)$, choose a point $\mathbf{w} \in S$ so that $\frac{1-\alpha}{\alpha}\|\mathbf{x} - \mathbf{w}\| < \rho$. The result follows from the previous case since

$$\alpha\mathbf{y} + (1-\alpha)\mathbf{x} = \alpha\left(\mathbf{y} + \tfrac{1-\alpha}{\alpha}(\mathbf{x} - \mathbf{w})\right) + (1-\alpha)\mathbf{w}.$$

For each $\gamma \geq 1$, we will let $N(\gamma)$ be an upper bound for the number of points that can be packed into the closed ball $B(\mathbf{0},1)$ when the distance between distinct points is at least $1/\gamma$ and one point is at $\mathbf{0}$. We will write $N_S(\gamma)$ for the similar constant when all points are on the surface of $B(\mathbf{0},1)$. Given nonzero points $\mathbf{b}$ and $\mathbf{c}$ in X, we set $V(\mathbf{b},\mathbf{c}) := \left\| \frac{\mathbf{b}}{\|\mathbf{b}\|} - \frac{\mathbf{c}}{\|\mathbf{c}\|} \right\|$.

We now fix τ with $1 < \tau \leq 2$, and we fix $\lambda > 1$. We fix $\{S_i : 1 \leq i \leq n\}$ in τ-satellite configuration with respect to an ordered set of points $\{\mathbf{a}_i : 1 \leq i \leq n\}$. For each i, we fix a positive r_i so that $B(\mathbf{a}_i, r_i) \subseteq S_i \subseteq B(\mathbf{a}_i, \lambda r_i)$ and S_i is starlike with respect to every $\mathbf{y} \in B(\mathbf{a}_i, r_i)$. We also translate so that $\mathbf{a}_n = \mathbf{0}$, and we set $r = r_n$ and $S = S_n$. We must find an upper bound for n.

Proposition B.4.2. *Suppose there are two constants $C_0 \geq 1$ and $C_1 \geq 1$ such that if $\mathbf{a}_i$ and $\mathbf{a}_j$ are centers with the properties that $C_0 r < \|\mathbf{a}_i\| \leq \|\mathbf{a}_j\|$ and $V(\mathbf{a}_i, \mathbf{a}_j) \leq 1/C_1$, then $\mathbf{a}_i$ must be in the interior of S_j. It then follows that*

$$n \leq N(2\lambda C_0) + N(8\lambda^2)\, N_S(C_1).$$

Proof. We give both an informal and a formal description of the steps of the proof.

[Informal and Formal: For $1 \leq i < j \leq n$, we have

$$\|\mathbf{a}_i - \mathbf{a}_j\| \geq r_i \geq \Delta(S_i)/(2\lambda) \geq \Delta(S)/(4\lambda) \geq r/(2\lambda).$$

Scaling by $1/(C_0 r)$, one sees that there can be at most $N(2\lambda C_0)$ indices i for which $\|\mathbf{a}_i\| \leq C_0 r$. We only have to show, therefore, that there are at most $N(8\lambda^2)\, N_S(C_1)$ indices in the set $J := \{j < n : C_0 r < \|\mathbf{a}_j\|\}$.**]**

[Formal: Suppose $i \neq j$ are members of J with $\mathbf{a}_i \in \operatorname{int}(S_j)$. Then $i < j$ and

$$\mathbf{a}_j \in B(\mathbf{a}_i, \Delta(S_j)) \subseteq B(\mathbf{a}_i, 2\Delta(S_i)) \subseteq B(\mathbf{a}_i, 4\lambda r_i).$$

Moreover, $\|\mathbf{a}_j - \mathbf{a}_i\| \geq r_i \geq r_i/(2\lambda)$. If also $j < k$ in J, and $\mathbf{a}_i \in \operatorname{int}(S_k)$, then $\mathbf{a}_k \in B(\mathbf{a}_i, 4\lambda r_i)$ and

$$\|\mathbf{a}_k - \mathbf{a}_j\| \geq r_j \geq \Delta(S_j)/(2\lambda) \geq \|\mathbf{a}_j - \mathbf{a}_i\|/(2\lambda) \geq r_i/(2\lambda).$$

Scaling by $1/(4\lambda r_i)$, it follows that for each $i \in J$, the cardinality $\operatorname{Card}\{j \in J : \mathbf{a}_i \in \operatorname{int}(S_j)\} \leq N(8\lambda^2)$.

Now construct $J' \subseteq J$ by induction as follows. Set $J_1 = J$. At the k^{th} step for $k \geq 1$, if J_k is empty, stop. Otherwise, choose the first $i_k \in J_k$ so that for all $j \in J_k$, $\|\mathbf{a}_{i_k}\| \leq \|\mathbf{a}_j\|$. Put i_k in J'. Form the set J_{k+1} by discarding from J_k the index i_k and all other indices j such that $\mathbf{a}_{i_k} \in \operatorname{int}(S_j)$.**]**

[Informal: Choose the first element i of J with minimal norm. Move that element from J to J', and throw away all other elements j of J for which the chosen center a_i is in the interior of S_j. You have thrown away at most $N(8\lambda^2) - 1$ elements. Now repeat with what is left of J until J is exhausted.**]**

[Informal and Formal: Now, if $i \neq j$ in J', $V(\mathbf{a}_i, \mathbf{a}_j) > 1/C_1$. Therefore, $\operatorname{Card}(J') \leq N_S(C_1)$, and so $\operatorname{Card}(J) \leq N(8\lambda^2)\, N_S(C_1)$.**]**

Theorem B.4.1. *The constants* $C_0 = 32\lambda^2$ *and* $C_1 = 16\lambda$ *work, whence*

$$n \leq N(64\lambda^3) + N(8\lambda^2)N_S(16\lambda).$$

Proof. Suppose i and j are indices such that $32\lambda^2 r < \|\mathbf{a}_i\| \leq \|\mathbf{a}_j\|$ and $V(\mathbf{a}_i, \mathbf{a}_j) \leq 1/(16\lambda)$. By Proposition B.4.2, we only have to show that $\mathbf{a}_i$ must be in the interior of S_j. To simplify notation, let $\mathbf{b} = \mathbf{a}_i$ and $\mathbf{c} = \mathbf{a}_j$. Fix $\mathbf{x} \in S \cap S_j$. Since $\|\mathbf{x}\| \leq \lambda r < 32\lambda^2 r < \|\mathbf{b}\|$, $\mathbf{x} \neq \mathbf{b}$. Let $s = \|\mathbf{c}\|/\|\mathbf{b}\|$ and $t = 1/s$. Set $\mathbf{y} = (1-s)\mathbf{x} + s\mathbf{b}$. Then $\mathbf{b} = (1-t)\mathbf{x} + t\mathbf{y}$. To show that $\mathbf{b} \in \operatorname{int}(S_j)$, we only have to show that $\|\mathbf{y} - \mathbf{c}\| < r_j$. Now $16\lambda\Delta(S) \leq 32\lambda^2 r < \|\mathbf{b}\|$, whence $\|\mathbf{x}\| \leq \Delta(S) \leq \min(\|\mathbf{b}\|/(16\lambda), 2\Delta(S_j))$. Therefore, since $|1-s| = s-1 < s$,

$$\begin{aligned}
\|\mathbf{y} - \mathbf{c}\| &= \left\|(1-s)\mathbf{x} + \|\mathbf{c}\| \left(\tfrac{\mathbf{b}}{\|\mathbf{b}\|} - \tfrac{\mathbf{c}}{\|\mathbf{c}\|}\right)\right\| \\
&< s\|\mathbf{x}\| + \|\mathbf{c}\|/(16\lambda) \\
&\leq s\|\mathbf{b}\|/(16\lambda) + \|\mathbf{c}\|/(16\lambda) = \|\mathbf{c}\|/(8\lambda) \\
&\leq (\|\mathbf{c} - \mathbf{x}\| + \|\mathbf{x}\|)/(8\lambda) \\
&< \Delta(S_j)/(2\lambda) \leq r_j.
\end{aligned}$$

Appendix C
Appendix on Infinitesimal Analysis and Measure Theory

C.1 Basic Nonstandard Analysis

In this section, we present a brief introduction to nonstandard analysis, including the extension of the real numbers with infinitely large and infinitely small numbers. We will conclude with the application to measure theory. A more extensive introduction as well as a large body of applications can be found in [35] as well as [15], [49], and [1].

One thinks of a standard mathematical model as a world that exists in some sense. For example, we think of the real numbers as having an existence independent of what we may know about them. Theorems in an appropriate formal language form correct statements about such a model. It is important to recognize the distinction between the names of objects in a standard model along with statements using such names in a formal language, and the objects themselves. For example, the number five has many names such as 5 in base ten, 101 in binary, and *V* in Roman numerals. The reason to emphasize this distinction is that for each standard mathematical model there are other mathematical objects, called nonstandard models, for which all the names and theorems for the standard model have a meaning and are correct for each nonstandard model. Informally, if we fix a nonstandard model, what we have are two worlds, the standard and the nonstandard, and the theorems about the first are also correct statements about the second. The foundation for the application of this fact to analysis, called nonstandard analysis, is due to Abraham Robinson [44]. His nonstandard models for the real number system contain infinitely large and infinitely small positive numbers together with all of the numbers in the original real number system.

One way to explain Robinson's result is to invoke a theorem of Kurt Gödel. Take a name not used for anything in the standard number system – for example, George. To the theorems about the standard real number system add new statements: "George is bigger than 1", "George is bigger than 2", etc. Add one such statement for each natural number. The standard number system is not a model for the collection of theorems augmented by these statements about George. There is no number

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2

simultaneously bigger than $1,2,3$, etc. The standard number system is, however, a model for any finite subset of the augmented collection of statements. To see this, fix a finite subset of the augmented collection. Find the biggest number named in these statements, and let George be the name of a number that is even bigger. Since every finite subset of our augmented collection of statements has a model, it follows from a result of Gödel that the entire augmented collection of statements has a model. That is, there is a number system for which all the theorems about the real numbers hold, but there is a number in that system, call it George, that is bigger than $1,2,3$, etc. George's reciprocal, 1 divided by George, is then a positive infinitesimal number.

Another approach to understanding Robinson's result is to construct a simple number system ${}^*\mathbb{R}$ with infinitesimals using sequences of real numbers and a free ultrafilter $\mathscr{U}$ on the natural numbers. We consider this next.

C.2 A Simple Extension of the Real Numbers

Recall the definition of a filter and ultrafilter in Example A.1.

Definition C.2.1. A **free ultrafilter** on $\mathbb{N}$ is a collection $\mathscr{U}$ consisting of subsets of $\mathbb{N}$ such that

1) $\varnothing \notin \mathscr{U}$,
2) $A \in \mathscr{U}$ & $B \in \mathscr{U} => A \cap B \in \mathscr{U}$,
3) $A \in \mathscr{U}$ & $A \subseteq B => B \in \mathscr{U}$,
4) $A \subset \mathbb{N}$ & $A \notin \mathscr{U} => \mathbb{N} \setminus A \in \mathscr{U}$, and
5) S a finite subset *of* $\mathbb{N} => \mathbb{N} \setminus S \in \mathscr{U}$.

Remark C.2.1. The ultrafilter $\mathscr{U}$ is generated by the Fréchet filter of Example A.1. The existence of such ultrafilters is established in Theorem A.3. The word "free" refers to Property 5. If $\mathscr{U}$ is a free ultrafilter on $\mathbb{N}$, and A_1, $A_2, \cdots$, A_m are disjoint subsets of $\mathbb{N}$ with union equal to $\mathbb{N}$, then one and only one of the sets A_i is in $\mathscr{U}$. Without Property 5, one could fix an $m \in \mathbb{N}$, and let $\mathscr{U}$ consist of all subsets of $\mathbb{N}$ containing m. A free ultrafilter however corresponds to a finitely additive measure on the power set of $\mathbb{N}$ taking either the value 0 or 1 on each set, and taking the value 1 on complements of finite sets.

For the rest of this section, we will work with a fixed free ultrafilter $\mathscr{U}$ on $\mathbb{N}$. We will say that a property holds almost everywhere, abbreviated a.e., if for some $U \in \mathscr{U}$, the property holds for all elements $n \in U$. For example, the sequence $x_n = 1/n$ is less than $1/3$ a.e. since by (5), the complement of the set $\{1,2,3\}$ must be in $\mathscr{U}$.

Notice that by (4), a property either holds a.e. or its negation holds a.e. It is this consequence of using an ultrafilter that will allow us to transfer properties valid for $\mathbb{R}$ to a number system built from sequences. That is, whatever we say about a sequence will either be true a.e. or false a.e. In what follows, we will use the notation $\langle r_i \rangle$ to denote the sequence mapping $i \in \mathbb{N}$ to r_i.

Definition C.2.2. A sequence $\langle r_i \rangle$ is equivalent to a sequence $\langle s_i \rangle$, and we write $\langle r_i \rangle \equiv \langle s_i \rangle$, when $r_i = s_i$ a.e. That is, when $\{i \in \mathbb{N} : r_i = s_i\}$ is in $\mathscr{U}$.

Definition C.2.3. We write $[< r_i >]$ for the equivalence class containing the sequence $\langle r_i \rangle$, and we say that $\langle r_i \rangle$ represents $[< r_i >]$. We use $^*\mathbb{R}$ to denote the collection of equivalence classes in the set of real-valued sequences. The set $^*\mathbb{R}$ is called the set of **nonstandard real** numbers or **hyperreal** numbers. Its formation using $\mathscr{U}$ is called an **ultrapower construction**.

Definition C.2.4. Given real-valued sequences $< r_i >$ and $< s_i >$, we set

$$[< r_i >] + [< s_i >] = [< r_i + s_i >],$$
$$[< r_i >] \cdot [< s_i >] = [< r_i \cdot s_i >],$$

$$|[< r_i >]| = [< |r_i| >]$$

$$[< r_i >] < [< s_i >] \text{ if } r_i < s_i \text{ a.e.}$$

Proposition C.2.1. *The operations + and $\cdot$, the mapping given by $[< r_i >] \to |[< r_i >]|$, and the ordering $<$ are independent of the choice of representing sequences.*

Proof. Left to the reader.

Note that the set of real numbers $\mathbb{R}$ (these are also called the **standard** numbers) is imbedded in the set of nonstandard real numbers $^*\mathbb{R}$. The imbedding is accomplished with the map that takes each $c \in \mathbb{R}$ first to the constant sequence $< c >$ and then to the element $[< c >] \in {}^*\mathbb{R}$. For example, 5 is mapped to the equivalence class $[< 5 >]$ containing the constant sequence $< 5 >$. We write $^*c$ for $[< c >]$, but we will later drop the star. Similarly, each n-tuple $\langle c_1, c_2, \cdots, c_n \rangle$ of real numbers is mapped to the n-tuple $\langle {}^*c_1, {}^*c_2, \cdots, {}^*c_n \rangle$ of nonstandard real numbers. The operations $+$ and $\cdot$ on $^*\mathbb{R}$ are extensions of the operations $+$ and $\cdot$ on $\mathbb{R}$. That is, one gets equivalent results working with either a pair of real numbers or the corresponding imbedded pair in $^*\mathbb{R}$. In a similar sense, the mapping $[< r_i >] \to |[< r_i >]|$ extends the absolute value function from $\mathbb{R}$ to $^*\mathbb{R}$, and the ordering $<$ extends the ordering $<$ from $\mathbb{R}$ to $^*\mathbb{R}$. The set $^*\mathbb{R}$ forms an ordered field extension of $\mathbb{R}$ with the equivalence classes containing the constant sequences 0 and 1 acting as the additive identity and the multiplicative identity, respectively.

The purpose of extending $\mathbb{R}$ to $^*\mathbb{R}$ is to obtain a number system with infinitesimal numbers such as $[< 1/i >]$. The reciprocals of such numbers will be infinitely large in absolute value. We measure this quality of magnitude in terms of standard natural numbers, i.e., natural numbers in the imbedded set $\mathbb{R}$.

Definition C.2.5. For any $r \in {}^*\mathbb{R}$,

1) *r is **infinite** or **unlimited** (positive or negative) if $|r| > n$ for every standard $n \in \mathbb{N}$*;
2) *r is **finite** or **limited** if $|r| < n$ for some standard $n \in \mathbb{N}$*; and
3) *r is **infinitesimal** if $|r| < 1/n$ for every standard $n \in \mathbb{N}$.*

Note that 0 is the only standard infinitesimal. The equivalence class $[< 1/i >]$ is infinitesimal and $[< i >]$ is a positive, unlimited number in ${}^*\mathbb{R}$.

Example C.2.1. To show that there are infinitely many elements in ${}^*\mathbb{R}$ greater than $\omega := [< 1, 2, \cdots, n, \cdots >]$, we note that for each $m \in \mathbb{N}$, $m\omega = [< m, 2m, \cdots, mn, \cdots >] > \omega$. It follows that there are infinitely many elements in ${}^*\mathbb{R}$ greater than 0 but strictly smaller than $\gamma := [< 1, 1/2, \cdots, 1/n, \cdots >]$.

Now we can define the extensions to ${}^*\mathbb{R}$ of relations, such as functions, that are defined on $\mathbb{R}$. The extensions are called $*$-transforms. A unary relation is a set.

Definition C.2.6. The $*$-transform of an n-ary relation P on $\mathbb{R}$ is the n-ary relation ${}^*P$ on ${}^*\mathbb{R}$, where $\langle [< r_i^1 >], \cdots, [< r_i^n >] \rangle \in {}^*P$ if and only if $\{ i \in \mathbb{N} : \langle r_i^1, \cdots, r_i^n \rangle \in P \}$ is a set in the ultrafilter $\mathscr{U}$, that is, for almost every $i \in \mathbb{N}$, one has $\langle r_i^1, \cdots, r_i^n \rangle \in P$.

Proposition C.2.2. *If P is an n-ary relation, then ${}^*P$ extends P. That is, if $\langle a_1, \cdots, a_n \rangle \in P$, then $\langle {}^*a_1, \cdots, {}^*a_n \rangle \in {}^*P$.*

Example C.2.2. The extension of the unit interval ${}^*[0,1]$ contains all nonstandard reals between 0 and 1.

Remark C.2.2. Note that when we say ${}^*1/2 \in {}^*[0,1]$ we mean the i^{th} entry of the constant sequence $\langle 1/2_i \rangle$ is in the i^{th} entry of the sequence $\langle [0,1]_i \rangle$ for all i in some element U of $\mathscr{U}$. This is not yet the true $\in$-relation, but in the construction, we then replace ${}^*[0,1]$ with the set of all elements in ${}^*\mathbb{R}$ that satisfy this relation. Working up the set-theoretic hierarchy, we then have the true $\in$-relation. Not every set you see can be obtained in this way. For example, if you try to get the set of standard natural numbers in this way, you are forced to also have some unlimited natural numbers in the set.

Proposition C.2.3. *If f is a function of n variables, then ${}^*f$ extends f, and ${}^*f$ is again a function. Moreover, if D is the domain of f, then ${}^*D$ is the domain of ${}^*f$.*

Remark C.2.3. The construction of ${}^*\mathbb{R}$ in terms of an ultrafilter $\mathscr{U}$ takes in account not just the finite limit of a convergent sequence, but how that limit is approached in terms of $\mathscr{U}$. To get more properties, one works with a bigger index set than $\mathbb{N}$.

C.3 The Transfer Principle

It is possible to work with ${}^*\mathbb{R}$ using the above construction with a larger index set to establish all needed properties. A shortcut, however, makes it a much more useful number system. A parallel can be found in the construction of the real numbers themselves. Recall that one can form the set of real numbers as the set of equivalence classes of Cauchy sequences of rational numbers. Essentially, $\mathbb{R}$ consists of all limits of Cauchy sequences of rational numbers. While it would be possible to work with the real numbers just in terms of this construction, such an approach would severely limit their utility. In practice, one works with the real number system using its properties. As is true for the real numbers, it is best here to put aside the construction of $^*\mathbb{R}$ as a collection of equivalence classes of real-valued sequences, and work instead with the corresponding properties. The basic fact needed to establish these properties is the **Transfer Principle**.

The Transfer Principle asserts, essentially, that any property that is correct when formally stated for $\mathbb{R}$, or $\mathbb{R}$ together with another structure such as a Banach space, is also valid for $^*\mathbb{R}$ and the extension of the other structure, when the names in the statement are replaced with the names of the corresponding transferred objects. The *-transform of a formal sentence is the same sentence with the names replaced with the names of the corresponding *-transformed objects. Here is that principle.

Theorem C.3.1 (Transfer Principle). *If a sentence in a formal language is a true statement about a standard structure, then its *-transform when properly interpreted is a true statement about the nonstandard extension of that structure.*

Corollary C.3.1 (Downward Transfer Principle). *If the *-transform of a sentence is a true statement about the nonstandard extension of a structure, then the original sentence is a true statement about the standard structure.*

Proof. If Φ is the sentence about the standard structure, and $\neg\Phi$ is a true statement about that structure, then the *-transform of $\neg\Phi$ would have to be a true statement about the nonstandard extension of that structure.

What is meant by saying "when properly interpreted"? Briefly, when we say "all" subsets of a given set, we can't formally specify what we mean. Even for the set of natural numbers, the idea of all subsets cannot be formalized. Ordinary language, for example, can only describe at most countably many subsets of the natural numbers. This inability to formalize the notion of "all subsets" means that when interpreting theorems in the nonstandard structure, we can cheat. We don't interpret the word "all" to really mean "all". We work instead with what are called **internal** sets, and interpret "all sets" to mean all internal sets.

If A is a set in the standard model, then $^*A$, called the nonstandard extension of A, is the set in the nonstandard model with the same name and formal properties as A. Nonstandard extensions of standard sets and sets that are elements of nonstandard extensions of standard sets are internal sets. Any object that can be described using only the names of known internal objects is also internal. An object that is not internal is called **external**.

Important for applications is the fact that the set of natural numbers $\mathbb{N} = \{1,2,3,\cdots\}$ has been extended along with the real numbers. The theorem that says that every positive real number is within distance one of a natural number is still valid for the extended number system. Therefore, we now have infinitely large natural numbers. We use $^*\mathbb{N}$ to denote the extended system of natural numbers, and $^*\mathbb{N}_\infty$ to denote the new members of that system, all of which are larger than any ordinary natural number. The set ${}^*\mathbb{N}_\infty$ is clearly external, since every nonempty, internal subset of ${}^*\mathbb{N}$ must have a first element. If ${}^*\mathbb{N}_\infty$ were internal, then subtracting 1 from a first element of ${}^*\mathbb{N}_\infty$ would yield the last ordinary natural number. Since ${}^*\mathbb{N}_\infty$ is external, it follows that the set of ordinary natural numbers is also external since ${}^*\mathbb{N}_\infty$ can be described as the complement of that set in ${}^*\mathbb{N}$. We have seen with the ultrapower construction that we cannot form a set in the nonstandard model consisting just of standard natural numbers. An important principle is the following consequence of the fact that the break between $\mathbb{N}$ and ${}^*\mathbb{N}_\infty$ is external.

Proposition C.3.1 (Spill-over Principle). *If $\langle A_n : n \in {}^*\mathbb{N}\rangle$ is an internal sequence (i.e., an internal function with domain ${}^*\mathbb{N}$) and an internal property P (i.e., stated in terms of internal objects) holds for all standard $n \in \mathbb{N}$, then that property holds for all $n \in {}^*\mathbb{N}$ up to some $\omega \in {}^*\mathbb{N}_\infty$. If an internal property holds for all $\omega \in {}^*\mathbb{N}_\infty$ smaller than some $\omega_0 \in {}^*\mathbb{N}_\infty$, then that property holds for all $n \in \mathbb{N}$ greater than some $n_0 \in \mathbb{N}$.*

C.4 Using the Transfer Principle

We now use the transfer principle to develop a fuller picture of ${}^*\mathbb{R}$. In applying the transfer principle, we shall refer to an object before taking its *-transform as **standard**, and we shall refer to its *-transform as its **nonstandard extension**. From now on, we will write a for both a standard element $a \in \mathbb{R}$ and the corresponding nonstandard extension in ${}^*\mathbb{R}$; we think of the set $\mathbb{R}$ as being imbedded in ${}^*\mathbb{R}$. We will also write equality and set membership as well as the arithmetic operations and the ordering relations using the same notation for the originals and their *-transforms.

Along with the Transfer Principle, a consequence of the construction of ${}^*\mathbb{R}$ is the fact that ${}^*\mathbb{R}$ contains nonzero infinitesimal elements as well as their multiplicative inverses, which are unlimited elements of ${}^*\mathbb{R}$. Listing the properties of $\mathbb{R}$ that make it an ordered field, one notes that they are simply stated in terms of addition $+$, multiplication $\cdot$, and the strictly positive elements $\mathbb{R}^+$ of $\mathbb{R}$. Moreover, $-x$ is notation for $(-1)\cdot x$.

Theorem C.4.1. *The number system $({}^*\mathbb{R}, +, \cdot, <)$ is an ordered field extension of the ordered field $(\mathbb{R}, +, \cdot, <)$.*

Proof. By the Transfer Principle, each property of $(\mathbb{R}, +, \cdot, <)$ holds for ${}^*\mathbb{R}$ when stated in terms of the nonstandard extensions of $\mathbb{R}$, $\mathbb{R}^+$ and the operations $+$ and $\cdot$.

Proposition C.4.1. *If f is a function of n variables on $\mathbb{R}$, then ${}^*f$ is a function of n variables on ${}^*\mathbb{R}$; it is an extension of f with ${}^*(\operatorname{dom} f) = \operatorname{dom}({}^*f)$ and ${}^*(\operatorname{range} f) = \operatorname{range}({}^*f)$.*

Proposition C.4.2. *For subsets of $\mathbb{R}$ and sets in $\mathbb{R}^n$, i.e., n-ary relations, we have the following, where $\complement A$ denotes the complement of a set A as appropriate in $\mathbb{R}$, $\mathbb{R}^n$ or their nonstandard extensions.*

i) ${}^*\varnothing = \varnothing$.
ii) ${}^*(A \cup B) = {}^*A \cup {}^*B$, ${}^*(A \cap B) = {}^*A \cap {}^*B$, $\complement({}^*A) = {}^*(\complement A)$.
iii) *For A_i, $i \in I$, a family of subsets of $\mathbb{R}^n$, $n = 1, 2, \cdots$*

$$\cup_{i\in I}{}^*A_i \subseteq {}^*(\cup_{i\in I}A_i), \text{ and } \cap_{i\in I}{}^*A_i \supseteq {}^*(\cap_{i\in I}A_i).$$

Proof. **i)** Since $\chi_\varnothing$ is identically 0, so is ${}^*\chi_\varnothing$.
ii) This can be proved using characteristic functions since

$$\chi_{(A\cap B)} = \chi_A \cdot \chi_B, \ \chi_{(A\cup B)} = \chi_A + \chi_B - \chi_A \cdot \chi_B, \ \chi_{\complement A} = 1 - \chi_A.$$

iii) For $n = 1$ and each $j \in I$, we transfer the sentences

$$(\forall x \in \mathbb{R})[x \in A_j \Rightarrow x \in \cup_{i\in I}A_i]$$

$$(\forall x \in \mathbb{R})[x \in \cap_{i\in I}A_i \Rightarrow x \in A_j].$$

The proof for $n > 1$ is similar.

Recall that for a given $\rho \in {}^*\mathbb{R}$, we say that ρ is **unlimited** or infinite if $|\rho| > n$ for all standard $n \in \mathbb{N}$, ρ is **limited** or finite if $|\rho| < n$ for some standard $n \in \mathbb{N}$, and ρ is **infinitesimal** if $|\rho| < 1/n$ for all standard $n \in \mathbb{N}$. The number 0 is the only real infinitesimal. It follows easily from the Transfer Principle that a number $\rho \in {}^*\mathbb{R}$ is positive and unlimited if and only if $1/\rho$ is strictly positive and infinitesimal, while $\rho \in {}^*\mathbb{R}$ is negative and unlimited if and only if $1/\rho$ is strictly negative and infinitesimal.

Definition C.4.1. The set of infinitesimal elements in ${}^*\mathbb{R}$ is called the **monad** of 0; it is denoted by $m(0)$.

Example C.4.1. The set $\bigcup_{\substack{n=1\\ n\in\mathbb{N}}}^{\infty} {}^*[-n,n]$ is the set of limited numbers in ${}^*\mathbb{R}$, while ${}^*\left(\bigcup_{\substack{n=1\\ n\in\mathbb{N}}}^{\infty} [-n,n]\right)$ is all of ${}^*\mathbb{R}$. The set $\bigcap_{\substack{n=1\\ n\in\mathbb{N}}}^{\infty} {}^*\left(\frac{-1}{n}, \frac{1}{n}\right) = m(0)$, but ${}^*\left(\bigcap_{\substack{n=1\\ n\in\mathbb{N}}}^{\infty} \left(\frac{-1}{n}, \frac{1}{n}\right)\right)$ is just the singleton set $\{0\}$.

Proposition C.4.3. *Let B be a subset of $\mathbb{R}^n$. Then ${}^*B \cap \mathbb{R}^n = B$.*

Proof. We give the proof for $n = 1$. If r is a real number not in B, then the sentence $r \notin B$ holds for B and thus for the extension of B.

C.5 Properties of ${}^*\mathbb{R}$

Theorem C.5.1. *The following properties hold for ${}^*\mathbb{R}$:*

i) *Finite sums, differences, and products of limited numbers are limited.*
ii) *Finite sums, differences, and products of infinitesimal numbers are infinitesimal.*

iii) *The infinitesimal numbers form an ideal in the ring of limited numbers; i.e., the product of a limited and an infinitesimal number is infinitesimal.*
iv) *The limited numbers form vector spaces over $\mathbb{R}$, and the infinitesimal numbers form vector spaces over $\mathbb{R}$.*

Proof. If $|\rho| < n$, and $|\tau| < m$ for n, $m \in \mathbb{N}$, then $n+m+n\cdot m$ bounds the sum, difference, and product of ρ and τ. Fix ρ and τ infinitesimal, and fix α limited in ${}^*\mathbb{R}$. Given any $n \in \mathbb{N}$, $|\rho| < 1/(2n)$ and $|\tau| < 1/(2n)$, so $|\rho+\tau| < 1/n$. There is an $m \in \mathbb{N}$ such that $|\alpha| < m$. Since $|\rho| < 1/(m\cdot n)$, $|\alpha\cdot\rho| < m/(m\cdot n) = 1/n$. The rest is left to the reader.

Definition C.5.1. We say that x and y are **infinitely close** if $x-y$ is infinitesimal. Here we write $x \simeq y$. If $x-y$ is limited, we write $x \sim y$. (Both $\simeq$ and $\sim$ are equivalence relations.) The equivalence class for $\simeq$ containing x is called the **monad** of x and written $m(x)$. That is, $m(x) = \{y \in {}^*\mathbb{R} : y \simeq x\}$. The equivalence class for $\sim$ containing x is called the **galaxy** of x and written $G(x)$.

Remark C.5.1. The monad of 0, $m(0)$, is the set of infinitesimals. Moreover, for all $x \in {}^*\mathbb{R}$, $m(x) = x+m(0)$. The galaxy of 0, $G(0)$ is the set of limited elements of ${}^*\mathbb{R}$. It is also denoted by $\mathrm{Fin}\,({}^*\mathbb{R})$. For each $x \in {}^*\mathbb{R}$, $G(x) = x+G(0)$.

We usually "center" monads at standard real numbers, and speak of the monad of r for $r \in \mathbb{R}$. The next result uses the property that any nonempty subset of $\mathbb{R}$ that has an upper bound has a least upper bound. In fact, the next result can be shown to be equivalent to this property.

Theorem C.5.2. *Every limited $\rho \in {}^*\mathbb{R}$ is in the monad of a unique $r \in \mathbb{R}$.*

Proof. Fix a limited $\rho \in {}^*\mathbb{R}$, and set $A := \{s \in \mathbb{R} : s \leq \rho\}$. Since A has an ordinary integer as an upper bound, we may let r be its least real upper bound. Now $\rho \simeq r$, for if not, then for some $n \in \mathbb{N}$, $1/n \leq |r-\rho|$. In this case, either $r < \rho$, and then $r+1/n$ is still in A, so r is not an upper bound of A or on the other hand $r > \rho$, and then $r-1/n$ is an upper bound of A, and so r is not the least upper bound of A. It follows that, $r \simeq \rho$. If we also have $s \in \mathbb{R}$ and $s \simeq \rho$, then $r-s \simeq 0$. Since 0 is the only infinitesimal in $\mathbb{R}$, $s = r$.

Definition C.5.2. If ρ is finite, then the unique real number r with $\rho \simeq r$ is called the **standard part** of ρ. We write $r = \mathrm{st}(\rho)$ or $r = {}^\circ\rho$. The mapping $\mathrm{st} : G(0) \to \mathbb{R}$ is called the **standard part map**.

Remark C.5.2. The reader should note that the standard part map is quite important in applications.

Theorem C.5.3. *The standard part of a sum, difference, product, or quotient of two limited numbers is, respectively, the sum, difference, product, or quotient of the standard parts of those numbers, with the exception that a denominator must not be infinitesimal. If $\rho \leq \tau$, then ${}^\circ\rho \leq {}^\circ\tau$.*

Proof. Fix limited numbers $r+\varepsilon$ and $s+\delta$, where r and s are real numbers, and ε and δ are infinitesimal (possibly 0). Then

$$(r+\varepsilon)\pm(s+\delta)=(r\pm s)+(\varepsilon\pm\delta)\simeq r\pm s,$$

$$(r+\varepsilon)\cdot(s+\delta)=(r\cdot s)+(r\cdot\delta)+\varepsilon\cdot(s+\delta)\simeq r\cdot s.$$

To establish the rule for quotients, we assume that $r>0$, and we note that for some $n,m\in\mathbb{N}$, $\frac{1}{n}<r^2-\frac{1}{m}$, and of course $\frac{1}{m}>|r\cdot\varepsilon|\simeq 0$, so $\frac{1}{n}<r^2+\varepsilon\cdot r$, whence $n>\frac{1}{r^2+\varepsilon\cdot r}>0$. This shows that $\frac{1}{r^2+\varepsilon\cdot r}$ is limited. Now we have

$$\frac{1}{r}-\frac{1}{r+\varepsilon}=\frac{\varepsilon}{r^2+r\cdot\varepsilon}\simeq 0.$$

The proof for $r<0$ is similar, and the rest follows from the product rule. If $(r+\varepsilon)\leq(s+\delta)$, then $r\leq s+(\delta-\varepsilon)<s+1/n$ for any $n\in\mathbb{N}$. It follows that $r\leq s$.

C.6 The Nonstandard Natural Numbers and Hyperfinite Sets

Recall that we write ${}^*\mathbb{N}_\infty$ to denote the set of unlimited elements of ${}^*\mathbb{N}$. The only limited elements of ${}^*\mathbb{N}$ are the standard natural numbers, so ${}^*\mathbb{N}_\infty={}^*\mathbb{N}\setminus\mathbb{N}$. If A is an infinite subset of $\mathbb{N}$, then ${}^*A$ contains arbitrarily large unlimited elements. In particular, ${}^*A\cap{}^*\mathbb{N}_\infty$ is not empty.

Definition C.6.1. The set ${}^*\mathbb{N}$ is called the set of **nonstandard natural numbers**. The extension of the integers ${}^*\mathbb{Z}$ is called the set of **nonstandard integers**. (It is formed from ${}^*\mathbb{N}$ in the same way that $\mathbb{Z}$ is formed from $\mathbb{N}$.)

Remark C.6.1. We note the following:

1) If $A=\{a_1,\cdots,a_n\}$ is a finite set in $\mathbb{R}$, then ${}^*A=A$. This follows because one can list the elements of A.
2) Not every subset of ${}^*\mathbb{R}$ is the extension of a standard one. For example, $\mathbb{N}$ cannot be the extension of a finite subset A of $\mathbb{N}$, since if A is finite, then ${}^*A=A$. On the other hand, if A is an infinite subset of $\mathbb{N}$, then ${}^*A$ contains unlimited elements.

Definition C.6.2. A hyperfinite set is a set in internal one-to-one correspondence with an initial segment of ${}^*\mathbb{N}$.

Remark C.6.2. Hyperfinite sets are important for many applications of nonstandard analysis. A feature that has powerful applications is the fact that if A is an infinite set in the standard structure, then there is a hyperfinite set $S\subset{}^*A$ such that for each $a\in A$, ${}^*a\in S$.

C.7 Sequences

A sequence is a function from $\mathbb{N}$ into $\mathbb{R}$. Therefore, every sequence has an extension that maps ${}^*\mathbb{N}$ into ${}^*\mathbb{R}$. We write $\langle s_n \rangle$ for the original sequence and $\langle {}^*s_n \rangle$ for its extension. Note that for all $n \in \mathbb{N}$, ${}^*s_n = s_n$. We write $s_n \to L$ when a sequence $\langle s_n \rangle$ has limit L. The results in this section are due to Robinson [44].

Theorem C.7.1. *A sequence $\langle s_n \rangle$ has limit L if and only if for every $\eta \in {}^*\mathbb{N}_\infty$, ${}^*s_\eta \simeq L$; that is, $L = \mathrm{st}({}^*s_\eta)$ for every $\eta \in {}^*\mathbb{N}_\infty$.*

Proof. Assume $s_n \to L$. Given an $\varepsilon > 0$ in $\mathbb{R}$, there is a $k \in \mathbb{N}$ for which the sentence $(\forall n \in \mathbb{N})[n \geq k \to |s_n - L| < \varepsilon]$ holds for $\mathbb{R}$. It follows by transfer that $\forall \eta \in {}^*\mathbb{N}_\infty$, $|{}^*s_\eta - L| < \varepsilon$. Since ε is arbitrary in $\mathbb{R}^+$, $|{}^*s_\eta - L| \simeq 0 \; \forall \eta \in {}^*\mathbb{N}_\infty$. Now assume that for all $\eta \in {}^*\mathbb{N}_\infty$, ${}^*s_\eta \simeq L$. Given $\varepsilon > 0$, let $A = \{n \in \mathbb{N} : |s_n - L| \geq \varepsilon\}$. Then A is a finite subset of $\mathbb{N}$ since there are no unlimited elements in ${}^*A$; that is, there is a finite maximum in A.

Example C.7.1. The sequence $\langle 1/n : n \in \mathbb{N} \rangle$ becomes $\langle 1/n : n \in {}^*\mathbb{N} \rangle$. For each unlimited η, $1/\eta \simeq 0$, so $1/n \to 0$.

Theorem C.7.2. *Assume $s_n \to L$ and $t_n \to M$. Then*

i) *$s_n + t_n \to L + M$,*
ii) *$s_n \cdot t_n \to L \cdot M$, and*
iii) *$s_n / t_n \to L/M$ provided $M \neq 0$.*

Proof. Left to the reader.

Proposition C.7.1. *A sequence $\langle s_n \rangle$ is Cauchy if and only if for every η and γ in ${}^*\mathbb{N}_\infty$, ${}^*s_\eta \simeq {}^*s_\gamma$.*

Proof. Left to the reader.

Theorem C.7.3. *A real number L is a cluster point of a sequence $\langle s_n \rangle$ if and only if for some $\eta \in {}^*\mathbb{N}_\infty$, ${}^*s_\eta \simeq L$.*

Proof. Assume that L is a cluster point of $\langle s_n \rangle$. There is a function $\psi : \mathbb{R}^+ \times \mathbb{N} \to \mathbb{N}$ such that

$$(\forall \varepsilon \in \mathbb{R}^+)(\forall k \in \mathbb{N})[\psi(\varepsilon, k) \geq k \wedge \left|s_{\psi(\varepsilon,k)} - L\right| < \varepsilon].$$

By transfer, if ε is a positive infinitesimal and $\eta \in {}^*\mathbb{N}_\infty$, then $\lambda = {}^*\psi(\varepsilon, \eta) \in {}^*\mathbb{N}_\infty$ and ${}^*s_\lambda \simeq L$. Conversely, if for some $\eta \in {}^*\mathbb{N}_\infty$, ${}^*s_\eta \simeq L$, then given $\varepsilon > 0$ and $k \in \mathbb{N}$, the sentence "$\exists n > k$ with $|s_n - L| < \varepsilon$" is true in the nonstandard model, so it is true in the standard model.

Example C.7.2. If $s_n = (-1)^n(1 - 1/n)$, then for any unlimited η, ${}^*s_\eta \simeq 1$ if η is even and ${}^*s_\eta \simeq -1$ if η is odd.

Theorem C.7.4 (Bolzano-Weierstrass). *Every bounded sequence has a limit point.*

Proof. If $\langle s_n \rangle$ is bounded and $\eta \in {}^*\mathbb{N}_\infty$, then ${}^*s_\eta$ is limited. Let $L = \mathrm{st}({}^*s_\eta)$. Now, there is an unlimited element of ${}^*\mathbb{N}$, namely η, such that ${}^*s_\eta \simeq L$.

C.8 Open Sets in the Reals

Recall that the monad $m(a)$ of a real number $a \in \mathbb{R}$ consists of the points infinitely close to a in ${}^*\mathbb{R}$. Many of the definitions and results of this section have obvious generalizations to $\mathbb{R}^n$.

Theorem C.8.1. *Let A be a subset of $\mathbb{R}$.*

i) *A is open if and only if for every $a \in A$, $m(a) \subset {}^*A$. (That is, all points an infinitesimal distance away from a are still in ${}^*A$.)*
ii) *A is closed if and only if for every $a \in \complement A$, $m(a) \cap {}^*A = \varnothing$.*

Proof. Clearly, (ii) follows from (i). If A is open and $a \in A$, then for some $\delta > 0$, $(\forall x \in \mathbb{R})[|x-a| < \delta \Rightarrow x \in A]$ holds for $\mathbb{R}$. If $x \in {}^*\mathbb{R}$, and $x \in m(a)$, then $|x-a| < \delta$, so by transfer, $x \in {}^*A$. If A is not open, then there is an $a \in A$ and a sequence $\langle s_n \rangle$ such that

$$(\forall n \in \mathbb{N})[s_n \in \complement A \wedge |s_n - a| < 1/n].$$

By transfer, for $\eta \in {}^*\mathbb{N}_\infty$, ${}^*s_\eta \simeq a$ and ${}^*s_\eta \in {}^*\left(\complement A\right) = \complement\,({}^*A)$, whence $m(a)$ is not contained in ${}^*A$.

Theorem C.8.2. *A point c is an accumulation point of $A \subseteq \mathbb{R}$ if and only if there is an $x \in {}^*A$ with $x \neq c$ but $x \simeq c$.*

Proof. Left to the reader.

Theorem C.8.3. *The closure $\overline{A}$ of $A \subseteq \mathbb{R}$ is the set $\{x \in \mathbb{R} : m(x) \cap {}^*A \neq \varnothing\}$.*

Proof. Left to the reader.

C.9 Limits and Continuity

Theorem C.9.1. *Suppose a is an accumulation point of A and $f : A \mapsto \mathbb{R}$. Then* $\lim_{x\to a} f(x)$ *exists and equals $L \in \mathbb{R}$ if and only if for every $x \in {}^*A$ with $x \simeq a$ but $x \neq a$, ${}^*f(x) \simeq L$.*

Proof. Assume that $\lim_{x\to a} f(x) = L$. Fix $\varepsilon > 0$ in $\mathbb{R}$. Then there is a $\delta > 0$ such that

$$(\forall x \in A)[0 < |x-a| < \delta \Rightarrow |f(x) - L| < \varepsilon]$$

holds for $\mathbb{R}$. By transfer, if $x \in (m(a) \cap {}^*A) \setminus \{a\}$, then $|f(x) - L| < \varepsilon$, and this is true for any $\varepsilon > 0$ in $\mathbb{R}$, whence $f(x) \simeq L$.

Conversely, if for every $x \in {}^*A$ with $x \simeq a$ but $x \neq a$, ${}^*f(x) \simeq L$, then given $\varepsilon > 0$ in $\mathbb{R}$, the sentence "$\exists \delta > 0$ such that if $x \in A$ and $0 < |x-a| < \delta$, then $|f(x) - L| < \varepsilon$" holds for the nonstandard model, just let δ be a positive infinitesimal, so the sentence holds for the standard model.

Theorem C.9.2. *If* $\lim_{x\to a} f(x) = L$ *and* $\lim_{x\to a} g(x) = M$*, then*

i) $\lim_{x\to a}\,(f+g)(x) = L+M$,
ii) $\lim_{x\to a}\,(f\cdot g)(x) = L\cdot M$, *and*
iii) $\lim_{x\to a}\,(f\,/\,g)(x) = L/M$ *if* $M \neq 0$.

Proof. Left to the reader.

Theorem C.9.3. *If f is defined on A, then f is continuous at $a \in A$ if and only if for every $x \in m(a) \cap {}^*A$, ${}^*f(x) \simeq f(a)$. That is, if and only if for $\Delta x = x - a \simeq 0$, we have $\Delta y = {}^*f(x) - f(a) \simeq 0$.*

Proof. Clear.

Example C.9.1. If $f(x) = x^2$ and $\Delta x = x - a \simeq 0$, then

$$\Delta y = (a+\Delta x)^2 - a^2 = 2a\cdot\Delta x + \Delta x^2 \simeq 0.$$

Corollary C.9.1. *The sum, product, and quotient of functions continuous at a are functions continuous at a, provided that for the quotient, the denominator does not vanish at a.*

Theorem C.9.4. *A function f is uniformly continuous on a set $A \subseteq \mathbb{R}$ if and only if for each x and y in ${}^*A$ with $x \simeq y$, ${}^*f(x) \simeq {}^*f(y)$.*

Proof. Assume f is uniformly continuous on A. Given $\varepsilon > 0$, there is a $\delta > 0$ so that

$$(\forall x \in A)(\forall y \in A)[|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon].$$

Thus if $x \simeq y$ in ${}^*A$, then $|{}^*f(x) - {}^*f(y)| < \varepsilon$. Since ε is arbitrary in $\mathbb{R}^+$, we have ${}^*f(x) \simeq {}^*f(y)$.

Conversely, if for each x and y in ${}^*A$ with $x \simeq y$, ${}^*f(x) \simeq {}^*f(y)$, then given $\varepsilon > 0$ in $\mathbb{R}$, the sentence "$\exists\delta > 0$ such that if $x \in A$ and $y \in A$ and $|x-y| < \delta$, then $|f(x) - f(y)| < \varepsilon$" holds in the nonstandard model, just let δ be a positive infinitesimal, so the sentence holds in the standard model.

Example C.9.2. The function $f(x) = 1/x$ is continuous on $(0,1)$, since for $a \in (0,1)$ and $h \simeq 0$, $1/(a+h) \simeq 1/a$. However, f is not uniformly continuous on $(0,1)$ since for $\eta \in {}^*\mathbb{N}_\infty$, $1/\eta \simeq 1/2\eta \simeq 0$, but ${}^*f(1/\eta) = \eta$ and ${}^*f(1/2\eta) = 2\eta$.

C.10 Differentiation

Theorem C.10.1. *Let f be defined on an interval $[a, a+\delta)$ (or $(a-\delta, a]$) for some positive δ. Then f has a right-hand (left-hand) derivative at a if for all strictly positive (negative) $h \simeq 0$, $({}^*f(a+h) - f(a))/h$ is finite and has a standard part independent of h. The right-hand (left-hand) derivative is that standard part. If the left-hand and right-hand derivatives exist and are equal, then f has a derivative at a.*

Proof. This follows immediately from the nonstandard criterion for a limit.

Example C.10.1. If $f(x) = x^2$, then for any standard x and nonzero $\Delta x \simeq 0$, $\Delta y = (x+\Delta x)^2 - x^2 = 2x\cdot\Delta x + (\Delta x)^2$, so $\Delta y/\Delta x = 2x+\Delta x \simeq 2x$.

Remark C.10.1. If $f'(c)$ exists for $c \in [a,b]$, then when $\Delta x \simeq 0$ and $dy = f'(c)\cdot\Delta x$,

$$\Delta y = {}^*f(c+\Delta x) - f(c) = f'(c)\cdot\Delta x + \varepsilon\cdot\Delta x = dy + \varepsilon\cdot\Delta x$$

where $\varepsilon \simeq 0$. Since $dy \simeq 0$, we have $\Delta y \simeq 0$.

Theorem C.10.2 (Chain Rule). *Let $f : [\alpha,\beta] \mapsto [a,b]$ and $g : [a,b] \mapsto \mathbb{R}$. If $f'(x)$ exists for $x \in [\alpha,\beta]$, while $f(x) = y \in [a,b]$ and $g'(y)$ exists, then $(g\circ f)'(x)$ exists and equals $f'(x)\cdot g'(y)$.*

Proof. Let $z = g(y)$. If $\Delta x \simeq 0$ but $\Delta x \neq 0$, then $\Delta y \simeq 0$, and

$$\Delta y = {}^*f(x+\Delta x) - f(x) = f'(x)\cdot\Delta x + E(x,\Delta x)\cdot\Delta x,$$

where $E(x,\Delta x) \simeq 0$. Moreover,

$$\Delta z = {}^*g(y+\Delta y) - g(y) = g'(y)\cdot\Delta y + F(y,\Delta y)\cdot\Delta y,$$

where $F(y,\Delta y) \simeq 0$ if $\Delta y \simeq 0$ but $\Delta y \neq 0$. In this case,

$$\begin{aligned}\frac{\Delta z}{\Delta x} &= g'(y)\cdot\big(f'(x)+E(x,\Delta x)\big) + F(y,\Delta y)\cdot\big(f'(x)+E(x,\Delta x)\big)\\ &= f'(x)\cdot g'(y) + g'(y)\cdot E(x,\Delta x) + f'(x)\cdot F(y,\Delta y) + E(x,\Delta x)\cdot F(y,\Delta y)\\ &\simeq f'(x)\cdot g'(y).\end{aligned}$$

If $\Delta x \simeq 0$ and $\Delta x \neq 0$ but $\Delta y = 0$, then since one factor in a 0 product must be 0, $f'(x) + E(x,\Delta x) = 0$. In this case, whatever the value of $F(y,\Delta y)$ may be,

$$\frac{\Delta z}{\Delta x} = f'(x)\cdot g'(y) + g'(y)\cdot E(x,\Delta x) + F(y,\Delta y)\cdot 0 \simeq f'(x)\cdot g'(y).$$

Remark C.10.2. For a standard proof of the chain rule, the functions E and F are standard functions with appropriate limit 0. The same observation as given here works for the case that $\Delta y = 0$.

C.11 Riemann Integration

Given a continuous function f on an interval $[a,b]$, we follow Keisler [22] in forming Riemann sums. Each $\Delta x > 0$ corresponds to a unique partition of $[a,b]$ with n subintervals, where n is the first integer such that $a + n\Delta x \geq b$: The partition endpoints are $x_i = a + i\Delta x$ for $0 \leq i \leq n-1$, and $x_n = b$. We let Δx_i denote the length of the i^{th} subinterval of the partition; of course, $\Delta x_i = \Delta x$ except for the last subinterval which may be shorter than Δx. Always evaluating at the left, the Riemann sum $S_a^b(f,\Delta x) = \sum_{i=1}^n f(x_{i-1})\,\Delta x_i$. Letting $M_i = \max_{x\in[x_{i-1},x_i]} f(x)$ and $m_i = \min_{x\in[x_{i-1},x_i]} f(x)$, we set $E_a^b(f,\Delta x) = \max_{1\le i\le n}(M_i - m_i)$. Along with the Riemann sum, we have the upper and lower sums

$$\overline{S}_a^b(f,\Delta x) = \sum_{i=1}^n M_i\,\Delta x_i \ \text{ and } \underline{S}_a^b(f,\Delta x) = \sum_{i=1}^n m_i\,\Delta x_i.$$

For a fixed f, these functions of a, b, and Δx can be extended to ${}^*\mathbb{R}$ and the extensions retain the inequality

$$\underline{S}_a^b(f,\Delta x) \le S_a^b(f,\Delta x) \le \overline{S}_a^b(f,\Delta x).$$

Theorem C.11.1 (Maximum Change Theorem). *If f is continuous on $[a,b]$, then*

$$\lim_{\Delta x\to 0} E_a^b(f,\Delta x) = 0.$$

The Maximum Change Theorem follows from, and indeed is equivalent to, the uniform continuity of f on $[a,b]$. The theorem can be stated and used in an elementary course without introducing uniform continuity, and it can be established there for functions with a bounded derivative. Easy corollaries are the following results.

Theorem C.11.2. *Let f be a continuous function on $[a,b]$. If Δx is a positive infinitesimal, then ${}^*\underline{S}_a^b(f,\Delta x) \simeq {}^*S_a^b(f,\Delta x) \simeq {}^*\overline{S}_a^b(f,\Delta x)$.*

Corollary C.11.1. *If f is continuous on $[a,b]$, then f is Riemann integrable there, and for any positive infinitesimal Δx, $\int_a^b f(x)dx = \mathrm{st}[{}^*S_a^b(f,\Delta x)]$.*

Remark C.11.1. How does one determine in setting up the Riemann integral for an application that the integrand has been correctly chosen? Why, for example, is $f(x)\Delta x$ a good approximation in calculating the area under the graph of f but Δx a bad approximation for the graph's length. An answer more general than bounding the desired quantity with upper and lower Riemann sums uses infinitesimals. It is Duhamel's Principle, which states that the sum of the infinitesimal errors resulting from the infinitesimal approximations should be infinitesimal.

To illuminate the method of nonstandard analysis, we have used examples from analysis on the real line. The method of "internal set theory ", initiated by Edward Nelson [39], works just with the nonstandard model, recognizing some elements of that model as being standard. A good recent development of that framework with applications to higher dimension calculus can be found in Nader Vakil's text [53].

C.12 Metric and Topological Spaces

In applications, one often uses the following property of a nonstandard extension, forming what is called an **enlargement**. That property, already noted, says that for any standard set A in the standard structure, there is a hyperfinite set F with

$$\{{}^*a : a \in A\} \subseteq F \subseteq {}^*A.$$

In a metric space, an open ball with center x and radius $r > 0$ is the set $B(x,r) = \{y \in X : \rho(x,y) < r\}$. In a nonstandard extension of a metric space, we define the monad of a point $x \in X$ by setting

$$\text{monad}(x) = m(x) := \bigcap {}^*B(x,r) = \{y \in {}^*X : {}^*\rho(x,y) \simeq 0\},$$

where the intersection is over all positive standard values of r. We use this monad in the same way we use monads on the real line. For example, a standard set O is called **open** if for each $x \in O$, $m(x) \subset {}^*O$.

There are settings where a metric will not capture the notion we want; we need a topological space. As shown in the chapter on topological spaces, we can use an open base at each point in a topological space in essentially the same way that we use balls in a metric space. Assume in what follows that a space X and an open base $\mathscr{B}_x$ at each point x in X are given.

Definition C.12.1. For each $x \in X$, the **monad** of x is

$$\text{monad}(x) = m(x) := \bigcap_{U \in \mathscr{B}_x} {}^*U.$$

As with balls in a metric space, we indicate that $y \in m(x)$ by writing $y \simeq x$. The **near-standard points** of ${}^*X$ are the points in the monad of some standard point of X.

Remark C.12.1. Since any finite intersection of elements of $\mathscr{B}_x$ contains another element of $\mathscr{B}_x$, there is a $W \in {}^*\mathscr{B}_x$ with $W \subset m(x)$. In a metric space, one can set W equal to a ball of infinitesimal radius.

Example C.12.1. For pointwise convergence on $[0,1]$, the monad of a real-valued function f consists of all internal ${}^*\mathbb{R}$-valued functions g on ${}^*[0,1]$ such that at each standard x, $g(x) \simeq f(x)$.

Proposition C.12.1. *A set $O \subseteq X$ is* ***open*** *if and only if for every $x \in O, m(x) \subseteq {}^*O$.*

Proof. To see these are the same thing, we note first that for each $U \in \mathscr{B}_x$, $m(x) \subset {}^*U$. On the other hand, if $m(x) \subseteq {}^*O$, then there is a $W \in {}^*\mathscr{B}_x$ with $W \subseteq m(x) \subseteq {}^*O$, and so "$\exists W \in \mathscr{B}_x$ with $W \subseteq O$" must also be true for the standard structure.

Theorem C.12.1 (Robinson). *A set $A \subseteq X$ is compact if and only if for each $y \in {}^*A$, there is an $x \in A$ with $y \in m(x)$. In particular, X is compact if and only if each point of ${}^*X$ is near-standard.*

Proof. Assume A is compact but there is a $y \in {}^*A$ not in the monad of any $x \in A$. Then each $x \in A$ is contained in an open set O_x with $y \notin {}^*O_x$. The family $\{O_x : x \in A\}$ covers A and therefore has a finite subcover $\{O_1, \cdots, O_n\}$. Now since $A \subseteq \cup_{i=1}^n O_i$, ${}^*A \subseteq \cup_{i=1}^n {}^*O_i$. Since $y \notin {}^*O_i$, for $1 \le i \le n$, $y \notin {}^*A$, which is impossible.

Now assume that A is not compact, i.e., there is a collection $\mathscr{U} = \{O_\alpha : \alpha \in \mathscr{A}\}$ of open sets such that no finite subcollection covers A. Let $\mathscr{S}$ be a hyperfinite collection in ${}^*\mathscr{U}$ with ${}^*O_\alpha \in \mathscr{S}$ for each $\alpha \in \mathscr{A}$. Then there is a $y \in {}^*A$ such that $y \notin V$ for each $V \in \mathscr{S}$. This point y is not in the monad of any standard point since for each $x \in A$, there is an α with $x \in O_\alpha$ but $y \notin {}^*O_\alpha$ since ${}^*O_\alpha \in \mathscr{S}$, whence $y \notin m(x)$.

The topology of pointwise convergence on $[0,1]$ can be generalized as noted in the chapter on topological spaces. Instead of $[0,1]$, we take an arbitrary index set $\mathscr{I}$. Instead of associating the real line with each $\alpha \in \mathscr{I}$, we let X_α be a topological space. Now the point set $\Pi_{\alpha\in\mathscr{I}} X_\alpha$ is the set of all functions f on $\mathscr{I}$ with $f(\alpha) \in X_\alpha$ for each $\alpha \in \mathscr{I}$. The monad of such an element f consists of all internal $g \in {}^*\Pi_{\alpha\in\mathscr{I}} X_\alpha$ with $g(\alpha) \simeq f(\alpha)$ for each standard $\alpha \in \mathscr{I}$. Such a g is a mapping on ${}^*\mathscr{I}$ with $g(\beta) \in X_\beta$ for each $\beta \in {}^*\mathscr{I}$, but the values of g at nonstandard indices are not relevant in determining whether g is in the monad of f. Recall that the set $\Pi_{\alpha\in\mathscr{I}} X_\alpha$ with this topology is called a product space, and the topology is called the product topology.

Theorem C.12.2 (Tychonoff). *The product of compact spaces is compact.*

Proof. If $X = \Pi_{\alpha\in\mathscr{I}} X_\alpha$ and $g \in {}^*X$, then for each standard $\alpha \in \mathscr{I}$, there is an $x_\alpha \in X_\alpha$ with $g(\alpha) \simeq x_\alpha$. (The x_α's are unique if the spaces X_α are Hausdorff.) The element $f \in X$ with $f(\alpha) = x_\alpha$ for each standard $\alpha \in \mathscr{I}$ is in X and $g \in m(f)$.

Theorem C.12.3. *A map f from a set A contained in a metric space (X,d) into a metric space (Y,ρ) is uniformly continuous on A if and only if for every x, $y \in {}^*A$, with $x \simeq y$, ${}^*f(x) \simeq {}^*f(y)$.*

Proof. Assume that $x \simeq y \Rightarrow {}^*f(x) \simeq {}^*f(y)$. Pick $\varepsilon > 0$ in $\mathbb{R}$. Then the sentence

$$(\exists\delta \in \mathbb{R}^+)(\forall x,y \in A)[d(x,y) < \delta \Rightarrow \rho(f(x),f(y)) < \varepsilon]$$

holds for the extension and therefore for the original structure. The converse is similar to the proof for $\mathbb{R}$.

Theorem C.12.4. *A continuous map f from a compact set A contained in a metric space (X,d) into a metric space (Y,ρ) is uniformly continuous on A.*

Proof. Left to the reader.

If one takes the finite (i.e., limited) nonstandard rational numbers modulo the infinitesimal ones, i.e., $[{}^*\mathbb{Q} \cap \mathrm{Fin}({}^*\mathbb{R})]/[{}^*\mathbb{Q} \cap m(0)]$, one obtains the real numbers $\mathbb{R}$. A similar construction, applied to infinite-dimensional spaces, produces spaces that are new, called **nonstandard hulls**. The development of these new spaces was initiated by Luxemburg [37] and extended by Henson and Moore [20]. In many cases, nonstandard hulls simplify the treatment of Banach space ultrapowers [14]. For more information, see Manfred Wolff's chapter in [35].

Definition C.12.2. Let ℓ_∞ denote the set of standard bounded sequences. A linear map $L : \ell_\infty \to \mathbb{R}$ is called a **Banach limit** if for each $\sigma \in \ell_\infty$, $L(\sigma)$ is a value between the $\liminf \sigma$ and the $\limsup \sigma$ and $L(\sigma) = L(T(\sigma))$, where $T(\sigma)(n) = \sigma(n+1)$.

Robinson's nonstandard construction of a Banach limit picks $\eta \in {}^*\mathbb{N}_\infty$ and sets

$$L_\eta(\sigma) = \mathrm{st}\left(\frac{1}{\eta}\sum_{n=1}^{\eta} {}^*\sigma(n)\right).$$

Here, $\Sigma_{n=1}^{\eta}$ is the nonstandard extension, evaluated at η, of the usual summation operator that sums a sequence from 1 to k.

C.13 Measure and Probability Theory

This section gives a brief overview of the construction initiated in [28] of standard measure spaces on nonstandard models (now called "Loeb spaces" in the literature). We will use the present day notation for the measure spaces considered here. The principal device used in nonstandard measure theory is $\aleph_1$-saturation. We will always assume now that we are working with an $\aleph_1$-saturated structure containing the nonstandard real numbers. What this means for the material in this section is that any sequence from an internal set S indexed by the standard natural numbers is just the beginning of an internal sequence in S indexed by the nonstandard natural numbers.

Working in an $\aleph_1$-saturated structure, we can construct a hyperfinite set Λ as the set of elementary outcomes in a conceptual experiment in the "nonstandard world." For coin tossing, for example, Λ can be the set of internal sequences of -1's and 1's of length $\eta \in {}^*\mathbb{N}_\infty$. Given such a hyperfinite Λ, we can let $\mathscr{C}$ consist of all internal subsets of Λ. The collection $\mathscr{C}$ is an internal σ-algebra, but it is also an algebra in the ordinary sense. Suppose P is an internal probability measure on $(\Lambda, \mathscr{C})$. For the coin tossing experiment, for example, each internal set A with internal cardinality $|A|$ would be given the probability $P(A) = |A|/2^\eta$ in ${}^*[0,1]$. In the general case, we can form a finitely additive real-valued measure $\widehat{P}$ on $(\Lambda, \mathscr{C})$ with values in the real interval $[0,1]$ by setting $\widehat{P}(A) = \mathrm{st}(P(A))$. The question is, "Can we extend $\widehat{P}$ to a countably additive measure on $\sigma(\mathscr{C})$, i.e., the σ-algebra generated by $\mathscr{C}$?" This is the question we consider next.

We start with an arbitrary internal measure space $(\Lambda, \mathscr{C}, \mu)$. The internal measure μ does not have to take only limited values, but for simplicity, we will assume here that it does. When Λ is a hyperfinite set, one usually sets $\mathscr{C}$ equal to the collection of all internal subsets of Λ, and μ is usually an internal probability measure. In general, $\mathscr{C}$ is an internal σ-algebra as well as an algebra in the usual sense in Λ. Let $\widehat{\mu}$ be the finitely additive, real-valued measure on $(\Lambda, \mathscr{C})$ defined at each $A \in \mathscr{C}$ by setting $\widehat{\mu}(A) = \mathrm{st}(\mu(A))$. Let $\sigma(\mathscr{C})$ denote the σ-algebra in the ordinary sense generated by $\mathscr{C}$. The collection $\sigma(\mathscr{C})$ is, in general, an external collection of subsets of Λ. We can extend $\widehat{\mu}$ to a measure μ_L defined on the measure completion $L_\mu(\mathscr{C})$ of $\sigma(\mathscr{C})$, and thus obtain a standard measure space $(\Lambda, L_\mu(\mathscr{C}), \mu_L)$ on Λ, by using the Carathéodory Extension Theorem.

Here is where $\aleph_1$-saturation is employed. If a sequence $\langle A_i : i \in \mathbb{N}\rangle$, indexed by the ordinary natural numbers, consists of pairwise disjoint elements of $\mathscr{C}$, and the union A is also in $\mathscr{C}$, then the sequence can be internally extended to a sequence indexed by ${}^*\mathbb{N}$. For any unlimited $\eta \in {}^*\mathbb{N}$, the union up to η contains A, so by spillover, the union up to some standard $n \in \mathbb{N}$ equals A. This means that A is actually a finite union since all but a finite number of the A_i's are empty.

It now follows that $\widehat{\mu}(A) = \sum_{i\in\mathbb{N}} \widehat{\mu}(A_i)$. By the Carathéodory Extension Theorem 10.2.1, the finitely additive measure $\widehat{\mu}$ has a σ-additive extension μ_L defined on the completion $L_\mu(\mathscr{C})$ of the σ-algebra $\sigma(\mathscr{C})$. Moreover, this extension is unique. Now $(\Lambda, L_\mu(\mathscr{C}), \mu_L)$ is an ordinary finite measure space formed on the internal set Λ.

Fix E in $L_\mu(\mathscr{C})$. It is well-known that for any $\varepsilon > 0$ in $\mathbb{R}$, there are sets $A \in \mathscr{C}_\delta$ and $B \in \mathscr{C}_\sigma$ with $A \subseteq E \subseteq B$ such that $\mu_L(B \setminus A) < \varepsilon$. By saturation, we may assume that A and B are actually sets in $\mathscr{C}$. To see this, suppose that $\langle A_n : n \in \mathbb{N}\rangle$ is a decreasing sequence in $\mathscr{C}$ with limit A. We may extend the sequence to an internal decreasing sequence indexed by ${}^*\mathbb{N}$, and choose an unlimited integer γ such that $\mu_L(A_\gamma) = \lim_{n\in\mathbb{N}} \mu_L(A_n)$. A similar proof works for B. We leave it as an exercise to show that by saturation there is a set $C \in \mathscr{C}$ such that the symmetric difference $E \triangle C$ is a null set in $L_\mu(\mathscr{C})$. These set approximation properties characterize $L_\mu(\mathscr{C})$ and have had many important applications. They have also been used in the literature to define $L_\mu(\mathscr{C})$.

Given an ${}^*\mathbb{R}$-valued function f on Λ, we set ${}^\circ f(x) = \mathrm{st}(f(x))$ if $f(x)$ is limited in ${}^*\mathbb{R}$. We set ${}^\circ f(x) = +\infty$ if $f(x)$ is positive and unlimited in ${}^*\mathbb{R}$, and we set ${}^\circ f(x) = -\infty$ if $f(x)$ is negative and unlimited in ${}^*\mathbb{R}$. It follows from the set approximation properties that a function $g : X \to \mathbb{R} \cup \{+\infty, -\infty\}$ is measurable with respect to $L_\mu(\mathscr{C})$ if and only if there is an internally $\mathscr{C}$-measurable function f such that ${}^\circ f = g$ μ_L-a.e. The function f is called a *lifting* of g. If such an f takes only limited values, then g is integrable, and by the set approximation properties

$$ {}^\circ\!\!\int f\, d\mu = \int {}^\circ f\, d\mu_L = \int g\, d\mu_L. \tag{C.13.1}$$

It follows from the definition of the standard integral that Equation C.13.1 also holds when

$$ \int [|f| - (|f| \wedge \eta)]\, d\mu \simeq 0 \quad \forall \eta \in {}^*\mathbb{N}_\infty \tag{C.13.2}$$

Condition C.13.2 is a simple expression of the condition called S-integrability in the literature. If we know, on the other hand, that g is integrable, then it is easy to see that for a sufficiently small $\eta \in {}^*\mathbb{N}_\infty$, replacing f with the function $f \cdot \chi_{\{|f| \le \eta\}}$ gives an S-integrable lifting of g. That is, with this replacement, the internal integral of f is limited and Equation C.13.1 holds.

Suppose ${}^*X$ is the nonstandard extension of a compact Hausdorff space X and μ is a limited-valued, internal measure on the internal Baire sets $\mathscr{C}$ in ${}^*X$. Then for each Borel set $B \subseteq X$, the set

$$\tilde{B} = \cup\{\text{monad}(y) : y \in B\}$$

is in $L_\mu(\mathscr{C})$. A Borel measure ν on X can be defined by setting $\nu(B) = \mu_L(\tilde{B})$ for each Borel set $B \subseteq X$. The measure ν restricted to the Baire sets is the standard part of the internal measure μ with respect to the weak* topology on Baire measures. By Robinson's criterion for convergence and clustering, this correspondence between ν and μ yields a nonstandard approach to the weak convergence of measures. See [4] and [30]. In particular, one can obtain Lebesgue measure on an interval $[a,b]$ as follows: Start with internal counting measure on the left endpoints of a hyperfinite interval partition of ${}^*[a,b]$, all intervals having the same size, and note that this is the measure used to produce internal Riemann sums; the standard part of the internal measure is Lebesgue measure.

When used in probability theory, the above general construction allows one to tackle problems of continuous parameter stochastic processes using the combinatorial tools available for discrete parameter processes. Examples are the construction of Poisson processes in [28] and Anderson's representation of Brownian motion and the Itô integral in [3]. The above construction has yielded new standard-analysis results by many researchers in areas such as probability theory, potential theory, number theory, mathematical economics, and mathematical physics. The reader may consult [35, 1], and [48] for further background and many applications. Here are six examples: Keisler's [23] new existence theorem for stochastic differential equations, Perkins' [40] award winning research (Rollo Davidson Prize in Probability Theory) on the theory of local time, Arkeryd's [5] results on gas kinetics, Cutland and Ng's work [13] on the Wiener sphere and Wiener measure, Renling Jin's work on number theory discussed in his chapter in [35], and Sun's work in probability and economics discussed in his contribution to [35]. Another application of the measure theory, taken from [29] and [31], extends Example 11.8.5 to more general boundaries and potential theories. Of particular note is a description in [35] by Yeneng Sun of his application in [51] of nonstandard measure theory and similar rich measure theories to the formation of a long overdue (see [16]), rigorous foundation for dealing with a continuum of independent random variables or traders in an economy. We also note that Fields medalist Terence Tao has used nonstandard measures in his recent multifaceted work [7] and [52].

Appendix D
Answers

Problem 1.8. **i)** For each $E \in \mathscr{C}$, $E \in \{E, X \setminus E, X, \varnothing\}$; the latter collection is the σ-algebra generated by the one element set $\{E\}$. Therefore, $\mathscr{C} \subseteq \mathscr{A}'$. **ii)** If $E \in \mathscr{A}'$, then by definition, $E \in \sigma(\mathscr{C}_0)$ for some countable $\mathscr{C}_0 \subseteq \mathscr{C}$. Since $X \setminus E$ is also in $\sigma(\mathscr{C}_0)$, $\mathscr{A}'$ is stable under complementation. Of course, X and $\varnothing$ also belong to $\sigma(\mathscr{C}_0)$, so they are in $\mathscr{A}'$. **iii)** Let $\{A_n : n \in \mathbb{N}\}$ be a countable collection of sets in $\mathscr{A}'$; each E_n is in $\sigma(\mathscr{C}_n)$ for some countable collection $\mathscr{C}_n \subseteq \mathscr{C}$. Now for each $m \in \mathbb{N}$, $E_m \in \sigma(\cup_n \mathscr{C}_n)$, so $\cup_m E_m \in \sigma(\cup_n \mathscr{C}_n)$. Since $\cup_n \mathscr{C}_n$ is countable, $\cup_m E_m \in \mathscr{A}'$, and we are done.

Problem 1.37. The function f is clearly an injection. It remains to prove it is a surjection. Suppose there exists an $x \notin f[K]$. Since $f[K]$ is compact, it is closed, and so there is an $\varepsilon > 0$ such that $\min_{y \in K} |x - f(y)| \geq \varepsilon$. Again since K is compact, the sequence of iterates $\langle f^n(x) \rangle$ has a convergent subsequence, say $\langle f^{n_i}(x) \rangle$. Then for $i < j$, we have $|f^{n_i}(x) - f^{n_j}(x)| = |x - f^{n_j - n_i}(x)| \geq \varepsilon$, contradicting the fact that every convergent sequence in K is a Cauchy sequence. Thus f is a bijection.

Problem 1.40.

(a) We note that the complement of F is a countable disjoint union of open intervals. On each one of those intervals that is finite in length, we extend with the linear function (i.e., a function for which the graph is a straight line) determined by the value of f at the endpoints. On an infinite interval, we extend with a constant, namely the value at the endpoint. The extended function g is continuous on $\mathbb{R} \setminus F$ since linear functions are continuous, and any point in $\mathbb{R} \setminus F$ is in one of the open intervals forming $\mathbb{R} \setminus F$. Fix $x \in F$ and $\varepsilon > 0$. Fix $\delta > 0$ so that for $\forall y \in F \cap (x - \delta, x + \delta)$, $|f(x) - f(y)| < \varepsilon$. If $\exists y \in F \cap (x, x + \delta)$, let $\delta_1 = |x - y|$. In this case, for all points $z \in [x, x + \delta_1]$, we have $|g(z) - g(x)| < \varepsilon$, since either $z \in F$ or z is in an open interval with endpoints in $F \cap [x, y]$. Here we use the fact that a linear function on an interval takes values between the values at the ends of the interval. If $(x, x + \delta) \cap F = \emptyset$, then the interval $(x, x + \delta)$ is contained in one of the intervals of the complement of F, and $\lim_{y \to x+} g(y) = f(x)$. Therefore, we may still pick $\delta_1 > 0$ so that $\forall z \in [x, x + \delta_1]$, $|g(z) - g(x)| < \varepsilon$. Similarly, we pick δ_2 to work at the left side of x. The minimum of δ_1 and δ_2

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2

works for x and ε. Note that we have extended f to $\mathbb{R}$ without increasing its supremum or decreasing its infimum.

(b) Let the open set be the real line with 0 removed. Let $f(x) = 0$ for $x < 0$ and $f(x) = 1$ for $x > 1$.

Problem 1.41. We partition $[a,b]$ using a $\delta = (b-a)/n$ that works for the uniform continuity of f in terms of $\varepsilon/2$. That is, we find a partition $x_0 = a < x_1 < \cdots < x_n = b$ so that the variation of f on any interval formed by the partition is smaller than $\varepsilon/2$. It follows that the variation of the linear interpolation g on each subinterval is smaller than $\varepsilon/2$. Therefore, if $x_i \leq x \leq x_{i+1}$,

$$\begin{aligned} |f(x) - g(x)| &\leq |f(x) - f(x_i)| + |f(x_i) - g(x_i)| + |g(x_i) - g(x)| \\ &< \varepsilon/2 + 0 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Problem 1.42.

(a) By Theorem 1.11.1, $B = f(A)$ is compact. Thus, $f : A \mapsto B$ and $f^{-1} : B \mapsto A$ are both continuous functions on compact sets. By Theorem 1.11.3, f and f^{-1} are uniformly continuous.

(b) Let $K = [0,\infty) = B$, and let $f : K \mapsto B$ be the square root function. Then, f is a homeomorphism with inverse $f^{-1} : B \mapsto K$ defined by squaring. Moreover, f is uniformly continuous while f^{-1} is not uniformly continuous.

(c) Let $E = (-\pi/2, \pi/2)$, $B = \mathbb{R}$, and $f : E \mapsto B$ by $f(x) = \tan x$. Then, $f^{-1} : B \mapsto E$ is uniformly continuous since it has bounded derivative while f is not uniformly continuous.

Problem 1.43. For each $x \in P_c$ and each $n \in \mathbb{N}$, let $I_{x,n}$ be an open interval centered at x such that if $y \in I_{x,n}$, then $|f(y) - f(x)| < 1/n$. Let $O_n = \cup_{x \in P_c} I_{x,n}$, and let $S = \cap_n O_n$. By definition, $P_c \subseteq S$. Let $z \in S$ and fix $\varepsilon > 0$. Fix n such that $1/n < \varepsilon/2$. There is an $I_{x,n}$ containing z. If $y \in I_{x,n}$, then $|f(y) - f(z)| \leq |f(y) - f(x)| + |f(x) - f(z)| < \varepsilon$. Therefore, we may take δ such that $(z - \delta, z + \delta) \subseteq I_{x,n}$, and this works for z and ε. Thus, z is a point of continuity of f, so $S = P_c$.

Problem 1.44. Recall that a sequence converges if and only if it is Cauchy. Fix $n \in \mathbb{N}$. Given i, j in $\mathbb{N}$, let $A^n_{ij} = \left\{x \in \mathbb{R} : \left|f_i(x) - f_j(x)\right| \leq 1/n\right\}$. Since $\mathbb{R} \backslash A^n_{ij} = \left\{x \in \mathbb{R} : \left|(f_i - f_j)(x)\right| > 1/n\right\}$ is an open set, A^n_{ij} is a closed set. Given $k \in \mathbb{N}$, let $F^n_k = \cap_{i,j \geq k} A^n_{i,j}$. This is the closed set consisting of all points such that for all i, $j \geq k$, $\left|f_i(x) - f_j(x)\right| \leq 1/n$. Let $F_n = \cup_k F^n_k$. This is the set of all points x such that for some $k \in \mathbb{N}$, $x \in F^n_k$; it is an F_σ-set. Now $C \subseteq \cap_{n \in \mathbb{N}} F_n$. Moreover, we have equality since if $x \notin C$, then for some $n \in \mathbb{N}$, $x \notin F_n$.

Problem 2.4. For all $A \subseteq \mathbb{R}$, for every $r \in \mathbb{R}$, translation by r of a countable covering of A by open intervals forms an open interval covering of $A + r$. Since the sum of the length of the intervals in the translated covering is the same as for the original covering, we can only get a smaller infimum by looking at all possible countable, open interval coverings of $A + r$. It follows that $\lambda^*(A + r) \leq \lambda^*(A)$. Since $A = (A + r) + (-r)$, we also have $\lambda^*(A) \leq \lambda^*(A + r)$, whence we have equality.

Problem 2.5. Fix a finite covering $\{I_k\}$ of A by open intervals. Let P be the set of points in $[0,1]$ that are not covered. For each $x \in P$, the minimum distance to the intervals in the covering is 0; otherwise, a rational in A is not covered. Thus, P is contained in the finite set of endpoints of the intervals forming the covering. Of course, we may have $P = \varnothing$. Therefore, P is empty or at most a finite set, and therefore has outer measure 0. It follows by subadditivity that

$$1 = \lambda^*([0,1]) \le \lambda^*(P) + \lambda^*([0,1]\backslash P) = \lambda^*([0,1]\backslash P) \le \sum l(I_k).$$

Problem 2.22. If $E \in \mathscr{M}$, $A \subseteq \mathbb{R}$, and $r \in \mathbb{R}$, then $A \cap (E+r) = [(A-r) \cap E] + r$. Since $x \mapsto x + r$ is a bijection of $\mathbb{R}$ onto $\mathbb{R}$, $\widetilde{E} + r = \complement(E+r)$. We have previously seen that for any set $A \subseteq \mathbb{R}$, $\lambda^* A = \lambda^*(A-r)$. It now follows since E is measurable that

$$\begin{aligned}\lambda^*(A) &= \lambda^*(A-r) = \lambda^*([A-r] \cap E) + m^*([A-r] \cap \widetilde{E}) \\ &= \lambda^*(([A-r] \cap E) + r) + \lambda^*(([A-r] \cap \widetilde{E}) + r) \\ &= \lambda^*(A \cap (E+r)) + \lambda^*(A \cap (\widetilde{E}+r)) \\ &= \lambda^*(A \cap (E+r)) + \lambda^*(A \cap \complement(E+r)).\end{aligned}$$

The result is not true for outer measures derived from arbitrary integrators F. For example, a set contained in an interval on which F is constant will have outer measure 0, and therefore be measurable. Translation of the set can move the set to an interval where the integrator is not constant. Then measurability is not automatic.

Problem 2.24. If $A \notin \mathscr{M}$, then it must follow that $m^*(A) > 0$. Let

$$\alpha = \sup\{m(F) : F \text{ closed, } F \subset A\},\ \beta = \inf\{m(G) : G \text{ open, } A \subset G\}.$$

If A is not measurable, we must have $\alpha < \beta$. By removing an F_σ set from A, we may assume that $\alpha = 0$ and still $\beta > 0$. Of course, $\beta = m^*(A) = m(S)$ for some G_δ set S containing A. Also, $m^*(S \setminus A) > 0$, since otherwise $A = S \cap \complement(S \setminus A)$ would be measurable. Therefore,

$$m^*(S \cap A) + m^*(S \cap \widetilde{A}) = m^*(A) + m^*(S \setminus A) = m(S) + m^*(S \setminus A) \neq m(S).$$

Problem 2.31. Let E be a Lebesgue measurable subset of the non-measurable set P. Let $\langle r_i \rangle$ be an enumeration of the rational numbers in $[0,1)$ with $r_0 = 0$. For each i, let $E_i = E +' r_i$. Then since the P_i's are disjoint, the subsets E_i are disjoint. By invariance of Lebesgue measure with respect to translation using the operation $+'$, $\lambda(E_i) = \lambda(E)$. Therefore, for any $n \in \mathbb{N}$,

$$0 \le n \cdot \lambda(E) \le \sum_{i=1}^{\infty} \lambda(E_i) = \lambda(\cup_{i=1}^{\infty} E_i) \le \lambda([0,1)) = 1.$$

It follows that $\lambda(E) = 0$.

Problem 2.32. Given the set A with $\lambda^*(A) > 0$,

$$A = \bigcup_{i=0}^{+\infty} [(A \cap [-i-1, -i)) \cup (A \cap [i, i+1))].$$

It follows from subadditivity that there is an interval $I_k = [k, k+1)$ with integer endpoints such that if $A_k = A \cap I_k$, then $\lambda^*(A_k) > 0$. Since $\lambda^*(A_k - k) > 0$, and $A_k - k \subseteq [0,1)$, we may assume that $A \subseteq [0,1)$. (Once we have proved the result for such an A, the general case then follows by translating a non-measurable $E \subseteq A_k - k \subseteq [0,1)$ to $E + k \subseteq A_k \subseteq A$.) Let P be the non-measurable subset of $[0,1)$ constructed in Section 2.7, and let P_i be the corresponding sets $P +' r_i$ formed using rationals r_i. Note that each set P_i is non-measurable. If for some integer $i \geq 0$, $A \cap P_i$ is measurable, then by Problem 2.31 and the invariance of Lebesgue measure with respect to translation using the operation $+'$, $\lambda(A \cap P_i) = 0$. Since $0 < \lambda(A) \leq \sum \lambda^*(A \cap P_i)$, there is an integer i with $\lambda^*(A \cap P_i) \neq 0$, whence $A \cap P_i \notin \mathscr{M}$.

Problem 2.34. The open set O_α of measure $\alpha < 1$ that is removed to form the generalized Cantor set is dense in $[0,1]$, therefore every point of $[0,1] \diagdown O_\alpha$ is a boundary point of O_α and that boundary has measure $1 - \alpha$.

Problem 3.4. Suppose f is an extended-real valued, measurable function on A. Recall that the set where f takes the value $+\infty$ is measurable and the set where f takes the value $-\infty$ is measurable. If f is restricted to a measurable set $B \subseteq A$ where f takes only one value in $[-\infty, +\infty]$, then for any $\alpha \in \mathbb{R}$, $\{x \in B : f(x) > \alpha\}$ is either $\varnothing$ or B, so f is measurable on B. If $c = 0$, cf is defined and equal to 0 exactly on the measurable set where f is finite. If $c > 0$ in $\mathbb{R}$, cf is measurable where f is finite by the same argument as for the finite case. Moreover, $cf = +\infty$ where $f = +\infty$, and $cf = -\infty$ where $f = -\infty$, so cf is measurable on A. A similar argument shows that cf is measurable on A if $c < 0$ in $\mathbb{R}$. If g is also a measurable, extended-real valued function on A, then $f \wedge g$ and $f \vee g$ are measurable on A by the same argument as for the finite case. The measurability of $f + g$ follows by arguments already given for the measurable set where both functions are finite. The rest of the argument for $f + g$ follows from the fact that the sum is defined and equal to $+\infty$ on the measurable set

$$\left[f^{-1}[\{+\infty\}] \bigcap g^{-1}[(-\infty, +\infty]]\right] \bigcup \left[g^{-1}[\{+\infty\}] \bigcap f^{-1}[(-\infty, +\infty]]\right],$$

and the sum is defined and equal to $-\infty$ on the measurable set

$$\left[f^{-1}[\{-\infty\}] \bigcap g^{-1}[[-\infty, +\infty)]\right] \bigcup \left[g^{-1}[\{-\infty\}] \bigcap f^{-1}[[-\infty, +\infty)]\right].$$

The measurability of $f \cdot g$ follows by arguments already given for the measurable set where both functions are finite. The rest of the argument for $f \cdot g$ follows from the fact that the product is defined and equal to $+\infty$ on the measurable set

$$\left[f^{-1}[\{+\infty\}] \bigcap g^{-1}[(0, +\infty]]\right] \bigcup \left[g^{-1}[\{+\infty\}] \bigcap f^{-1}[(0, +\infty]]\right]$$
$$\bigcup \left[f^{-1}[\{-\infty\}] \bigcap g^{-1}[[-\infty, 0)]\right] \bigcup \left[g^{-1}[\{-\infty\}] \bigcap f^{-1}[[-\infty, 0)]\right],$$

and the product is defined and equal to $-\infty$ on the measurable set

$$\left[f^{-1}[\{-\infty\}] \bigcap g^{-1}[(0, +\infty]]\right] \bigcup \left[g^{-1}[\{-\infty\}] \bigcap f^{-1}[(0, +\infty]]\right].$$

Problem 3.18. If we know that the function in the hint is measurable, then since f is the pointwise limit as $n \to \infty$, we know that f is measurable. Recall that the defining property of an interval is that it is not empty, and if two points are in the set, then every point between them is in the set. This means that for a strictly increasing function or even an increasing function, the inverse image of an open interval is either empty or an interval, and therefore Borel measurable.

Problem 4.3. Restricting f to A, the integral $\int_A f = \int_A f + \int_{E\setminus A} 0 = \int_E f \cdot \chi_A$.

Problem 4.11. We are given that $\int f^+ = +\infty$, and f^- and h are integrable. It follows that $\int f + \int h = +\infty$. We want to show that $\int (f+h) = +\infty$. Now $(f+h)^+ = (f+h^+ - h^-)^+ \geq (f^+ - f^- - h^-)^+$. Since $(f+h)^- \leq (f^- + h^-)$, and $(f^- + h^-)$ is integrable, it is sufficient to show that $\int (f^+ - f^- - h^-) = +\infty$. This follows from the fact that if $f^+ - f^- - h^-$ equaled an integrable function g, we would have $f^+ = g + f^- + h^-$, which is integrable, but f^+ is not integrable.

Problem 4.22. Fix a measurable set E, and let $\langle f_{n_k} \rangle$ be any subsequence of $\langle f_n \rangle$. By Fatou's Lemma

$$\begin{aligned} \int_{\mathbb{R}} f = \int_E f + \int_{\mathbb{R}\setminus E} f &\leq \underline{\lim}_k \int_E f_{n_k} + \underline{\lim}_k \int_{\mathbb{R}\setminus E} f_{n_k} \\ &\leq \underline{\lim}_k \left(\int_E f_{n_k} + \int_{\mathbb{R}\setminus E} f_{n_k} \right) = \int_{\mathbb{R}} f. \end{aligned}$$

Now $\int_E f \leq \underline{\lim}_k \int_E f_{n_k}$ and the same is true for $\mathbb{R} \setminus E$, but the sums of the left and right side for E and $\mathbb{R} \setminus E$ are equal and finite. It follows that $\int_E f = \underline{\lim}_k \int_E f_{n_k}$. Since this is true for any subsequence, $\int_E f_n$ exists and equals $\int_E f$.

Problem 4.31. Let $\langle I_n \rangle$ be a sequence of finite closed subintervals on which the proper Riemann integral of f exists (so f will be bounded on each I_n) with $I_n \nearrow I$. On each I_n, $\int_{I_n} f = R\int_{I_n} f$. Moreover, $|f \cdot \chi_{I_n}| \leq |f|$. Therefore, by the Lebesgue Dominated Convergence Theorem,

$$R\int_I f = \lim_n R\int_{I_n} f = \lim_n \int_{I_n} f = \int f.$$

Problem 4.32. Suppose $\int |f_n| \to \int |f|$. Now $|f_n - f| \to 0$ a.e., $|f_n - f| \leq g_n := |f_n| + |f|$, $g_n \to g := 2|f|$ a.e., and we are assuming that $\int g_n \to \int g$. Therefore, $\int |f_n - f| \to 0$. On the other hand, if we know that $\int |f_n - f| \to 0$, then by the triangle inequality,

$$0 \leq \left| \int |f_n| - \int |f| \right| \leq \int ||f_n| - |f|| \leq \int |f_n - f| \to 0.$$

Problem 4.44. This is the Lebesgue Dominated Convergence Theorem with the parameter $n \to \infty$ replaced with $y \searrow 0$. In fact the Lebesgue Dominated Convergence Theorem tells us that the result is true for any sequence $y_n \searrow 0$, and so the result is true for the more general limit process $y \searrow 0$. Note that it is not enough to do this for just one sequence such as $y_n = 1/n$. You must work with an arbitrary sequence y_n converging to 0.

Problem 5.12. Order the rational numbers in $(0,1]$, and at the n'th rational q_n put the weight 2^{-n}. Let $F(0)=0$, and for $x>0$, let $F(x)$ denote the total of the weights up to x plus 1 to get a jump of 1 at 0. That is, $F(x)=1+\Sigma_{0<q_n\leq x}2^{-n}$. Then F is increasing and has jump 2^{-n} at q_n. To see that F is continuous at each irrational number x, we fix $\varepsilon>0$ and $N\in\mathbb{N}$ such that $\Sigma_{n=N}^{\infty}2^{-n}<\varepsilon$. Choose $\delta>0$ so that the interval $[x-\delta,x+\delta]$ contains no rational q_k for $k<N$ and is contained in $(0,1)$. Then $F(x+\delta)-F(x-\delta)<\varepsilon$, so $\forall y$ with $|x-y|<\delta$, $|F(x)-F(y)|\leq F(x+\delta)-F(x-\delta)<\varepsilon$.

Problem 5.17. The only discontinuities of an increasing function are at the jumps, and there are at most a countable number of these. (For example, there can be at most $(F(K)-F(-K))\cdot n$ jumps of size $1/n$.) There is no derivative at a jump. Moreover, at a point x that is not a jump point for F, $G(x)=F(x)$. Suppose x_0 is a point where a derivative $F'(x_0)$ exists, and x_n is a sequence of points where a jump of F may or may not occur with $x_n\searrow x_0$. Since a countable set cannot form an open interval, there is a sequence of non-jump points $y_n\searrow x_0$ and a sequence of non-jump points $z_n\searrow x_0$ with $y_n\leq x_n\leq z_n$ for each n. Then

$$\frac{F(y_n)-F(x_0)}{y_n-x_0}\cdot\left(\frac{y_n-x_0}{x_n-x_0}\right)\leq\frac{G(x_n)-G(x_0)}{x_n-x_0}\leq\frac{F(z_n)-F(x_0)}{z_n-x_0}\cdot\left(\frac{z_n-x_0}{x_n-x_0}\right).$$

We may choose each z_n and y_n close enough to x_n so that $1-\frac{1}{n}\leq\left(\frac{y_n-x_0}{x_n-x_0}\right)\leq 1$, and $1\leq\left(\frac{z_n-x_0}{x_n-x_0}\right)\leq 1+\frac{1}{n}$. With this choice,

$$\frac{F(y_n)-F(x_0)}{y_n-x_0}\cdot\left(1-\frac{1}{n}\right)\leq\frac{G(x_n)-G(x_0)}{x_n-x_0}\leq\frac{F(z_n)-F(x_0)}{z_n-x_0}\cdot\left(1+\frac{1}{n}\right).$$

The first and last terms of the inequality both converge to $F'(x_0)$, and so the same is true of the middle sequence of ratios. A similar proof works for jump points $x_n\nearrow x_0$, or we may apply the previous result to $G(x)=-F(-x)$. The proof also shows that $F'(x_0)=G'(x_0)$.

Problem 5.19. Assume g is increasing. Otherwise, work with $-g$. Also assume that $E\subseteq(0,1)$ since $\lambda(\{g(0),g(1)\})=0$. Fix $\varepsilon>0$, and let $\delta>0$ corresponding to ε in the definition of absolute continuity. Let O be an open set containing E and contained in $(0,1)$ with $\lambda(O)<\delta$. The open set O is the union $\cup_{i=1}^{\infty}(a_i,b_i)$ of disjoint open intervals; of course, $\Sigma_{i=1}^{\infty}(b_i-a_i)<\delta$. Now $g[E]$ is contained in the union of the sets $g[(a_i,b_i)]$. These are intervals since g is monotone increasing and continuous. Moreover, since δ corresponds to ε with respect to absolute continuity, any finite union has measure $<\varepsilon$, so the countable union has measure $\leq\varepsilon$. Thus $\lambda(g[E])\leq\varepsilon$. Since ε is arbitrary, $\lambda(g[E])=0$.

Problem 5.21. Part a) is left to the reader. b) Assume f is absolutely continuous and $|f'|\leq B$ where it exists. Then for any interval $[x,y]$,

$$|f(y)-f(x)|=\left|\left(f(x)+\int_x^y f'd\lambda\right)-f(x)\right|=\left|\int_x^y f'd\lambda\right|\leq\int_x^y|f'|\,d\lambda\leq B\cdot|y-x|.$$

Now assume f is an *arbitrary* real-valued function that satisfies a Lipschitz condition with constant M. Then f is absolutely continuous by Part a. Moreover, every difference quotient $\frac{f(y)-f(x)}{y-x}$ lies in the interval $[-M,M]$, so the derivative also lies in $[-M,M]$ wherever it exists.

Note: Suppose $F(x)=\int_0^x f(t)d\lambda$ where $f\geq 0$ is increasing, continuous, and integrable on $[0,1)$, but f is unbounded on $[0,1)$. Then F is absolutely continuous but $F'=f$ is not bounded. We see that there is no hope of getting a Lipschitz condition here since for any interval $[x,x+\Delta x]\subseteq[0,1)$, we have

$$F(x+\Delta x)-F(x)\geq f(x)\cdot\Delta x$$

Problem 7.16. Assume that $E=\cup E_n$, where for each n, the closure of E_n in X, $\overline{E}_n$, contains no open subsets of X. Then $A=\cup_n(A\cap E_n)$, and for each n, $\overline{(A\cap E_n)}\subseteq\overline{E}_n$ and thus contains no open subsets of X.

Problem 7.20. In a metric space, a singleton set $\{x\}$ is closed. If $\{x\}$ is not isolated, that is, if $\{x\}$ is not open, then $\{x\}$ is nowhere dense since it is already closed and contains no open sets. A complete metric space X cannot be the countable union of such singleton sets. One can also note that if $\langle x_n\rangle$ is a sequence of non-isolated points in a metric space, then $\bigcap_n(X\setminus\{x_n\})$ is the countable intersection of dense open sets.

Problem 7.35.

a) Assume $f_m\to f$ in L^∞. For each $n,m\in\mathbb{N}$, let $E_n^m:=\{|f_m-f|\geq 1/n\}$. If $f_m\to f$ in L^∞, then for each n, there is a $k_n\in\mathbb{N}$ such that $\cup_{m\geq k_n}E_n^m$ is a set of measure 0. Therefore, $E:=\cup_n\cup_{m\geq k_n}E_n^m$ is a set of measure 0, and f_n converges to f uniformly on the complement of E. Now assume that f_n converges to f uniformly on the complement of a set E of measure 0. For any $n\in\mathbb{N}$, there is a k_n such that for all $m\geq k_n$, $|f_m-f|\leq 1/n$ on the complement of E, whence $\|f_m-f\|_\infty\leq 1/n$ for all $m\geq k_n$.

b) Let $\langle f_n\rangle$ be a Cauchy sequence in L^∞. For each $n,m,j\in\mathbb{N}$, let $E_{m,n}^j=\{|f_m-f_n|\geq 1/j\}$. For each j, there is a k_j such that $\cup_{m,n\geq k_j}E_{m,n}^j$ is a set of measure 0. Therefore, $E:=\cup_j\cup_{m,n\geq k_j}E_{m,n}^j$ is a set of measure 0. For each $x\notin E$, $\langle f_n(x)\rangle$ is Cauchy. Let $f(x)$ be the limit on the complement of E and 0 on E. By Part a, we need to show that f_n converges to f uniformly on $X\setminus E$. For any $j\in\mathbb{N}$, there is a $k_j\in\mathbb{N}$ such that $\forall n,\ m\geq k_j$ and $\forall x\in X\setminus E$, $|f_n(x)-f_m(x)|<1/j$, whence $|f_n(x)-f(x)|<2/j$. It follows that f_n converges to f uniformly on $X\setminus E$.

Problem 7.39. We need only show that c is closed with respect to the topology given by the ℓ^∞-norm, since a closed subset of a complete space is complete. For each $n\in\mathbb{N}$, let $i\mapsto f_n(i)$ be a convergent sequence, so the sequence of sequences, $\langle f_n:n\in\mathbb{N}\rangle$, is a sequence in c. Let f be a bounded sequence, i.e., in $\ell^\infty(\mathbb{N})$, with $\|f_n-f\|_\infty\to 0$. We will have shown that c is closed if we show that $f\in c$; that is, that f is Cauchy. Fix $\varepsilon>0$ and $n\in\mathbb{N}$ such that $\|f_n-f\|_\infty<\varepsilon/3$. Since for this n, f_n is Cauchy, there is a k such that for all $i,j\geq k$, $|f_n(i)-f_n(j)|<\varepsilon/3$, whence

$$|f(i)-f(j)|\leq|f(i)-f_n(i)|+|f_n(i)-f_n(j)|+|f_n(j)-f(j)|<\varepsilon.$$

Problem 7.41. Working with just the real part of f, we may assume that $f \geq 0$, since if the result is true for f^+ approximating from below with $\varphi^+ \geq 0$ and the result is true for f^- approximating from below with $\varphi^- \geq 0$, then in general,

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

We may now take a sequence of simple functions φ_n increasing up to f. Since f^p is integrable, each φ_n^p is integrable, so since each φ_n takes only a finite number of values, each φ_n vanishes off of a set of finite measure. Since $|f-\varphi_n|^p \to 0$ at each point where f is finite, that is, almost everywhere, and $|f-\varphi_n|^p \leq |f|^p$, we have by the Lebesgue Dominated Convergence Theorem $\int |f-\varphi_n|^p \, d\mu \to 0$. We choose $N \in \mathbb{N}$ so that for all $n \geq N$, $\int |f-\varphi_n|^p \, d\mu < \varepsilon^p$.

Problem 7.43. First we show that $\|f\|_\infty \leq \liminf_{p\to\infty} \|f\|_p$. We may assume that $\|f\|_\infty > 0$ (otherwise the inequality is trivial). For every strictly positive $r < \|f\|_\infty$, the set $A_r = \{x \in X : |f(x)| > r\}$ has strictly positive but finite measure. For all finite $p \geq 1$,

$$\|f\|_p = \left(\int |f(x)|^p\right)^{1/p} \geq \left(\int (r \cdot \chi_{A_r})^p\right)^{1/p} = r\mu(A_r)^{1/p},$$

whence $r \leq \liminf_{p\to\infty} \|f\|_p$. Since $r \in (0, \|f\|_\infty)$ is arbitrary, we conclude that $\|f\|_\infty \leq \liminf_{p\to\infty} \|f\|_p$. On the other hand,

$$\begin{aligned} \|f\|_p &= \left(\int |f(x)|^p\right)^{1/p} = \left(\int |f(x)| \cdot |f(x)|^{p-1}\right)^{1/p} \\ &\leq \left(\int |f(x)| \cdot \|f\|_\infty^{p-1}\right)^{1/p} = \left(\int |f(x)|\right)^{1/p} \cdot \|f\|_\infty^{1-1/p} = \|f\|_1^{1/p} \|f\|_\infty^{1-1/p}. \end{aligned}$$

Therefore, $\limsup_{p\to\infty} \|f\|_p \leq \|f\|_\infty$, whence, $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$.

Problem 7.47. By assumption, each open L^∞-ball about f contains a continuous function. That is, for each $n \in \mathbb{N}$, there is a continuous function g_n such that $\|f-g_n\|_\infty \leq 1/n$. At each $x \in [0,1]$, the sequence $\langle g_n(x)\rangle$ is a Cauchy sequence. To see this, fix $x \in [0,1]$ and $\varepsilon > 0$. There is an n_0 such that for all $n > n_0$, $\|f-g_n\|_\infty < \varepsilon/4$. Since the g's are continuous, for each $n, m > n_0$, there is an open interval U centered at x such that for all $y \in U \cap [0,1]$, $|g_n(x) - g_n(y)| < \varepsilon/4$ and $|g_m(x) - g_m(y)| < \varepsilon/4$. Let f_0 denote a function representing the L^∞-equivalence class f. We may have to ignore the behavior of f_0 on a set of measure 0 in U, but for any other point $y \in U \cap [0,1]$, we have $|f_0(y) - g_n(y)| < \varepsilon/4$ and $|f_0(y) - g_m(y)| < \varepsilon/4$. Using such a point y as a "catalyst", we have

$$|g_n(x) - g_m(x)| \leq |g_n(x) - g_n(y)| + |g_n(y) - f_0(y)| + |f_0(y) - g_m(y)| + |g_m(y) - g_m(x)| < \varepsilon.$$

Let $h(x) = \lim_n g_n(x)$ for each $x \in [0,1]$. We have actually shown that the g_n's form a uniform Cauchy sequence, so they converge to h uniformly. It follows that h is continuous. Now, $h = f_0$ except on a set of measure 0, namely the union over n of the exceptional sets for the inequalities $|f_0 - g_n| < 1/n$. Therefore, the equivalence class containing f_0 contains the continuous function h.

Problem 8.16. The result follows from the fact that there are at most a finite number of terms x_α with $\|x_\alpha\|$ greater or equal to $1/n$ for each $n \in \mathbb{N}$. This follows, since if it is not true for some $n \in \mathbb{N}$, then for $\varepsilon = 1/(3n)$ and any finite sum $\sum_{x_\alpha \in F} x_\alpha$ that is within $1/(3n)$ of the limit L, there will still be a term x_β, with $\beta \notin F$ and $\|x_\beta\| \geq \frac{1}{n}$. When x_β is added to the old sum, the new sum will be farther than $1/(3n)$ from L. That is,

$$\left\| L - \left(\sum_{x_\alpha \in F} x_\alpha + x_\beta \right) \right\| \geq \left| \left\| L - \sum_{x_\alpha \in F} x_\alpha \right\| - \|x_\beta\| \right|$$
$$= \|x_\beta\| - \left\| L - \sum_{x_\alpha \in F} x_\alpha \right\| \geq \frac{1}{n} - \frac{1}{3n} > \frac{1}{3n}.$$

Problem 9.9. By replacing each $y \in S$ with the set of all ordinals less than or equal to y and taking the union of this countable collection of countable sets, we may assume that S is a countable **initial segment** of the ordinals. That is, if $\alpha \in S$, then every element of Y smaller than α is also in S. Now Y is uncountable, so $Y \neq S$. Every element of $Y \setminus S$ must be an upper bound of S, and there is a first element of $Y \setminus S$, which is the least upper bound of S. Note that there may be no greatest element of S. For example, S may correspond to the set of natural numbers.

Problem 9.10. Suppose U, V formed a disconnection of the union. Since neither U nor V can have empty intersection with the union, we may fix A_α and A_β with $A_\alpha \cap U \neq \varnothing$ and $A_\beta \cap V \neq \varnothing$. Since A_α is connected and contained in $U \cup V$, the pair U, V cannot be a disconnection of A_α. On the other hand, since U intersected with the union is disjoint from V intersected with the union, $(A_\alpha \cap U) \cap (A_\alpha \cap V) = \varnothing$. Therefore, we must have $A_\alpha \cap V = \varnothing$. That is, $A_a \subseteq U$. Similarly, $A_\beta \subseteq V$. This contradicts the assumption that $A_a \cap A_\beta \neq \varnothing$.

Problem 9.13. If we assume that $\overline{A}$ is pathwise connected, we can obtain a contradiction to the fact that the unit interval has finite length. Here is a proof: Assume there is a path $f(t)$, $0 \leq t \leq 1$, in the closure of the graph of $\sin(1/x)$ such that $f(0) = (1/\pi, 0)$ and $f(1)$ is on the y-axis. Then f is a uniformly continuous function from $[0,1]$ to $\mathbb{R}^2$. In particular, there is a $\delta > 0$ such that if $0 < a < b < 1$, and $b - a < \delta$, then $|f(b) - f(a)| < 1/4$. For any point p in $\mathbb{R}^2$, let $Y(p)$ be the y-coordinate. Let $b_1 = \max\{t \in [0,1] : Y(f(t)) = -1/2\}$ and let $a_1 = \max\{t < b_1 : Y(t) = 1/2\}$. Given a pair a_i, b_i, let $b_{i+1} = \max\{t < a_i : Y(f(t)) = -1/2\}$ and let $a_{i+1} = \max\{t < b_{i+1} : Y(f(t)) = 1/2\}$. In this way, we obtain an infinite number of disjoint intervals $[a_i, b_i]$ in $(0,1)$ with $b_i - a_i \geq \delta$ for all i. This is a contradiction to the fact that the unit interval has finite length.

Problem 9.30. Assume $(X, \mathscr{T})$ is locally compact. Fix $x \in X$ and a $\mathscr{T}$-open neighborhood U of x such that the $\mathscr{T}$-closure $\overline{U}$ in X is compact with respect to $\mathscr{T}$. By Problem 9.29, $\overline{U}$ is compact, and therefore closed with respect to $\mathscr{S}$. Now $U = X \cap V$, where V is $\mathscr{S}$-open. Moreover, $V \diagdown \overline{U} \subseteq Z \diagdown X$, and $V \diagdown \overline{U}$ is $\mathscr{S}$-open. Since X is dense in Z, $V \diagdown \overline{U} = \varnothing$, so $V = U$. It follows that X is $\mathscr{S}$-open in Z.

Now assume that $Z \setminus X$ is $\mathscr{S}$-closed. Fix $x \in X$. By Proposition 9.8.2, Z is normal, so there are disjoint $\mathscr{S}$-open sets U and V, with $x \in U$ and $Z \setminus X \subset V$. It follows that the $\mathscr{S}$-closure of U is an $\mathscr{S}$-compact subset of X, and it is therefore $\mathscr{T}$-compact.

Problem 9.32. Fix $\varepsilon > 0$, $x \in X$, and a neighborhood U of x such that for any $y \in U$ and any $f \in \mathscr{F}$, $\rho(f(y), f(x)) < \varepsilon/3$. Given $g \in \overline{\mathscr{F}}$, we will show that for all $y \in U$, $\rho(g(y), g(x)) < \varepsilon$. Given $y \in U$, since $g \in \overline{\mathscr{F}}$, there is an $f \in \mathscr{F}$ with $\rho(f(y), g(y)) < \varepsilon/3$ and $\rho(f(x), g(x)) < \varepsilon/3$, whence

$$\rho(g(y), g(x)) < \rho(g(y), f(y)) + \rho(f(y), f(x)) + \rho(f(x), g(x)) < \varepsilon.$$

Thus, U works for x and ε in terms of the family $\overline{\mathscr{F}}$.

Problem 9.34. Since $C(X)$ is a metric space with the metric obtained from the uniform norm, a function is in the closure of $\mathscr{A}$ if and only if it is the uniform limit of a sequence from $\mathscr{A}$. Fix $f, g \in \overline{\mathscr{A}}$ and sequences $\langle f_n \rangle$ and $\langle g_n \rangle$ in $\mathscr{A}$ with $f_n \to f$ and $g_n \to g$. Then $f_n + g_n \to f + g$, $f_n \cdot g_n \to f \cdot g$, and for any $\alpha \in \mathbb{R}$, $\alpha f_n \to \alpha f$.

Problem 9.36. First, assume there is a function $f \in \mathscr{A}$ that is nowhere equal to 0. The algebra generated by 1 and $\mathscr{A}$ consists of functions of the form $\alpha + h$ where $h \in \mathscr{A}$ and $\alpha \in \mathbb{R}$. Since this algebra is dense in $C(X)$, there is a sequence h_n in $\mathscr{A}$ and a sequence of real numbers α_n such that $\alpha_n + h_n \to 1/f$. Multiplying by f, we see that $\alpha_n f + h_n f \to 1$ since for each n, $\alpha_n f + h_n f \in \mathscr{A}$, $1 \in \overline{\mathscr{A}}$, and so $\overline{\mathscr{A}} = C(X)$.

If for each $x \in X$, there is an $f \in \mathscr{A}$ with $f(x) \neq 0$, then the nonnegative function f^2 is in $\mathscr{A}$, and a suitable finite sum of such functions is positive everywhere on X. Therefore, if $\overline{\mathscr{A}} \neq C(X)$, there must be a unique (since $\mathscr{A}$ separates points) point p such that $f(p) = 0$ for all $f \in \mathscr{A}$. Fix $g \in C(X)$ such that $g(p) = 0$. Then there is a sequence $\langle f_n + \alpha_n \rangle$ with $f_n \in \mathscr{A}$ and $\alpha_n \in \mathbb{R}$ such that $f_n + \alpha_n$ converges uniformly to g on X. Since $f_n(p) = 0$ for each n and $g(p) = 0$, $\alpha_n \to 0$. Therefore $g \in \overline{\mathscr{A}}$.

Problem 10.4. Fix a closed and bounded interval $[a_k, b_k]$ in the k^{th} coordinate axis of X for each $k \leq n$ so that $\overline{R}$ is the product of these closed intervals. Again, $\{P_i : i = 1, \cdots, m\}$ is a finite covering of $\overline{R}$ by general open rectangles. There is a $J \in \mathbb{N}$ such that if each of the intervals $[a_i, b_i]$ is partitioned into J closed subintervals of size $(b_i - a_i)/J$, then each product of each of the resulting closed subintervals is contained in at least one of the P_k's. To see this, assume no such J exists, and when $J = n$, let x_n be a vertex closest to the origin of one of the products of the closed subintervals for which the result fails. By the Bolzano-Weierstrass Theorem, there is a cluster point x_0 of the sequence $\langle x_n : n \in \mathbb{N} \rangle$. One of the general open rectangles P_{k_0} contains x_0. A subsequence $\langle x_{n_k} \rangle$ of the sequence $\langle x_n \rangle$ converges to x_0. For each k, the other vertices of the closed product for which x_{n_k} is a vertex also converge to x_0. Therefore, for some $n \in \mathbb{N}$, all of the vertices including x_n of one of the products of closed subintervals is in the open general rectangle P_{k_0}, contradicting the definition of x_n.

It now follows that there is a partition of $\overline{R}$ into a finite number of closed general rectangles Q_j, $1 \leq j \leq k$, so that each Q_j is contained in one of the open rectangles P_i. The volume of each P_i is greater than or equal to the sum of the volumes of the rectangles Q_j that are contained in P_i. Therefore,

$$V(\overline{R}) = \sum_{j=1}^{k} V(Q_j) \leq \sum_{i=1}^{n} V(P_i).$$

Problem 11.2. This is clear for a one-dimensional subspace generated by $x_1 \in X$, since when $a_n x_1 \to x_0$, the coefficients a_n form a Cauchy sequence of scalars. Assume that any $(n-1)$-dimensional manifold in X must be closed. Fix linearly independent elements $x_1, \cdots, x_n$. By assumption, no x_i is a limit of linear combinations of the other x_i's, for if it were, the $(n-1)$-dimensional manifold spanned by the others would not be closed. Let y be a limit point of a sequence $y_i = \sum_{j=1}^{n} \alpha_j^i x_j$. We want to show that for each j, the sequence $\left\langle \alpha_j^i \right\rangle$ is bounded, since then, a diagonalization argument allows us to assume that for each j, $\alpha_j^i \to \alpha_j \in \mathbb{R}$, and then $y_i \to \sum_{j=1}^{n} \alpha_j x_j$. Since we also have $y_i \to y$, $y = \sum_{j=1}^{n} \alpha_j x_j$ is in the span of $x_1, \cdots, x_n$. Now assume that for some j, after reordering let it be n, and for some subsequence of the original sequence y_i, we have $\lim \left|\alpha_n^i\right| = +\infty$. For that subsequence, $y_i = \left(\sum_{j=1}^{n-1} \alpha_j^i x_j\right) + \alpha_n^i x_n$, so $x_n = \left(1/\alpha_n^i\right) \cdot y_i + z_i$, where z_i is in the span of the other x_j's. Since $y_i \to y$, $|y_i|$ is a bounded sequence, whence $\left|\left(1/\alpha_n^i\right) \cdot y_i\right| \to 0$. This means that x_n is in the closure of the span of the other x_j's violating our assumption. Therefore, we are done.

Problem 11.3. We may assume that $f \neq 0$. The kernel K is $f^{-1}[\{0\}]$. By Theorem 11.1.1, f is bounded if and only if it is continuous, and so the inverse image of the closed set $\{0\}$ is closed. Now we will assume that K is closed and show that f is continuous at 0. It then follows by Theorem 11.1.1 that f is bounded. Fix $x \in X$ with $f(x) = 1$. If $f(y) \neq 0$, then $f((f(y) \cdot x) - y) = 1 \cdot f(y) - f(y) = 0$, so $f(y) \cdot x - y = k$, where $k \in K$. It follows that for each $y \in X$, there is a scalar α and a $k \in K$ such that $\alpha x + k = y$. That is, there is only one dimension left after K. Now fix a sequence $\langle z_n \rangle$ in X converging to 0. We must show that $f(z_n) \to 0$. If $|f(z_n)| > 1$, we may replace z_n with $z_n / f(z_n)$, so without loss of generality we may assume that $|f(z_n)| \leq 1$. For each n, there is a scalar α and a $k_n \in K$ with $\alpha_n x + k_n = z_n$. Since $f(x) = 1$, for each n, $|\alpha_n| \leq 1$. Given any subsequence of the sequence $\langle \alpha_n \rangle$, we choose a further subsequence that converges to a scalar a. Since $z_n \to 0$, we then have $k_n \to -ax$. Since the kernel is a closed subspace not containing x, $a = 0$, and so $f(\alpha_n x + k_n) = \alpha_n \to 0$. It follows that for the original sequence we have $f(z_n) = f(\alpha_n x + k_n) \to 0$.

Problem 11.8. If $\|F\| \leq 1$, then for all $x \in X$, $|F(x)| / \|x\| \leq 1$, so $F(x) \leq |F(x)| \leq \|x\|$. On the other hand, if for all $x \in X$, $F(x) \leq \|x\|$, then for all $x \in X$, $-F(x) = F(-x) \leq \|-x\| = \|x\|$, so $-\|x\| \leq F(x) \leq \|x\|$, whence $|F(x)| \leq \|x\|$, and the result follows.

Problem 11.9. We must show that the natural imbedding $\widetilde{\varphi}$ of X^* into the third dual X^{***} is a surjection if and only if the natural imbedding φ of X into X^{**} is a surjection. Given any $F \in X^{***}$, we consider the restriction of F to $\varphi[X]$, and define an element $f_F \in X^*$ at each $x \in X$ by setting $f_F(x) := F(\varphi(x))$. This is linear since given a linear combination $\alpha x + \beta y$ in X,

$$f_F(\alpha x + \beta y) = F(\varphi(\alpha x + \beta y)) = \alpha F(\varphi(x)) + \beta F(\varphi(y)) = \alpha f_F(x) + \beta f_F(y).$$

Moreover, for any $x \in X$ with $\|x\| \leq 1$, $\|\varphi(x)\| \leq 1$, so

$$\|f_F(x)\| = \|F(\varphi(x))\| \leq \|F\| \, \|\varphi(x)\| \leq \|F\|,$$

so f_F is bounded.

Assume that X is reflexive, that is, $X^{**} = \varphi[X]$. Given an arbitrary $F \in X^{***}$ we want to show that $F = \widetilde{\varphi}(f_F)$; this will show that $\widetilde{\varphi}$ is surjective. Given an arbitrary $g \in X^{**}$, by assumption, $g = \varphi(x_g)$ for some element $x_g \in X$. Therefore,

$$F(g) = F(\varphi(x_g)) = f_F(x_g) = \varphi(x_g)(f_F) = g(f_F) = \widetilde{\varphi}(f_F)(g).$$

Since g is arbitrary in X^{**}, $F = \widetilde{\varphi}(f_F)$.

Suppose now that there is a $g \in X^{**} \setminus \varphi[X]$, that is, X is not reflexive. By Theorem 11.5.1, $\varphi[X]$ is closed in X^{**}. By Proposition 11.4.2, we may fix $F \in X^{***}$ with $F(h) = 0$ for all $h \in \varphi[X]$ and $F(g) > 0$. If we assume for the sake of reaching a contradiction that X^* is reflexive, then $F = \widetilde{\varphi}(f_0)$ for some $f_0 \in X^*$. Now for all $x \in X$,

$$f_0(x) = \varphi(x)(f_0) = \widetilde{\varphi}(f_0)(\varphi(x)) = F(\varphi(x)) = 0,$$

since $\varphi(x) \in \varphi[X]$. That is, f_0 is the 0 functional in X^*, whence $F = \widetilde{\varphi}(f_0)$ is the 0 functional in X^{***}. This contradicts the assumption that $F(g) > 0$.

Problem 11.13. As suggested, we may consider A as a subset of X^{**} under the canonical injection φ. That is, we will look at the set $\{\varphi(x) : x \in A\}$. Now φ is an isometry, so we only have to find a common bound for the values of $\|\varphi(x)\|$, $x \in A$. By definition, each $f \in X^*$ is continuous with respect to the weak topology on X. Since A is a compact set with respect to the weak topology, there is a constant M_f such that for all $x \in A$, $|\varphi(x)(f)| = |f(x)| \leq M_f$. By the Uniform Boundedness Principle 11.6.1, there is a constant M such that for all $x \in A$, $\|x\| = \|\varphi(x)\| \leq M$.

Problem 11.16. By the Alaoglu Theorem 11.7.1, the unit ball of the dual space of $C(X)$ is compact in the weak* topology. We have a directed set consisting of all finite subsets of the indices, directed by containment. For each finite set of indices S, there is a measure ν_S that works. That is, $\int f_{\alpha_i}\, d\nu_S = c_{\alpha_i}$ for all $i \in S$. This net has a cluster point given by a Radon measure ν. Fix an index α. We want to show that $\int f_\alpha d\nu = c_\alpha$. Let U be the weak* neighborhood of ν determined by f_α and an $\varepsilon > 0$. That is, U consists of all finite signed Radon measures μ such that $|\int f_\alpha d\mu - \int f_\alpha d\nu| < \varepsilon$. Since our net is frequently in U, there is a finite set S containing α such that $\nu_S \in U$. Since ν_S works for $\alpha \in S$, $\int f_\alpha d\nu_S = c_\alpha$, so $|c_\alpha - \int f_\alpha d\nu| < \varepsilon$. Since ε is arbitrary, $\int f_\alpha d\nu = c_\alpha$.

References

1. Albeverio, S., Fenstad, J.E., Høegh–Krohn, R., Lindstrøm, T.: Nonstandard Methods in Stochastic Analysis and Mathematical Physics. Academic, Orlando (1986)
2. Aldaz, J.M.: A general covering lemma for the real line. Real Anal. Exch. **17**, 394–398 (1991/92)
3. Anderson, R.M.: A nonstandard representation of Brownian motion and Itô integration. Isr. J. Math. **25**, 15–46 (1976)
4. Anderson, R.M., Rashid, S.: A nonstandard characterization of weak convergence. Proc. Am. Math. Soc. **69**, 327–332 (1978)
5. Arkeryd, L.: Loeb solutions of the Boltzmann equation. Arch. Ration. Mech. Anal. **86**, 85–97 (1984)
6. Bateman, P.T., Erdős, P.: Geometrical extrema suggested by a lemma of Besicovitch. Am. Math. Mon. **58**(5), 306–314 (1951)
7. Bergelson, V., Tao, T.: Multiple recurrence in quasirandom groups. Geom. Funct. Anal. **24**, 1–48 (2014)
8. Besicovitch, A.S.: A general form of the covering principle and relative differentiation of additive functions (I), (II). Proc. Camb. Philos. Soc. **41**, 103–110 (1945); **42**, 1–10 (1946)
9. Bliedtner, J., Loeb, P.A.: A reduction technique for limit theorems in analysis and probability theory. Ark. Mat. **30**, 25–43 (1992)
10. Bliedtner, J., Loeb, P.A.: The optimal differentiation basis and liftings of L^∞. Trans. Am. Math. Soc. **352**, 4693–4710 (2000)
11. Bliedner, J., Loeb, P.A.: A local maximal function simplifying measure differentiation. MAA Mon. **114**, 532–536 (2007)
12. Constantinescu, C., Cornea, A.: Ideale Ränder Riemannscher Flächen. Springer, Heidelberg (1963)
13. Cutland, N.J., Ng, S.-A.: The wiener sphere and wiener measure. Ann. Probab. **21**(1), 1–13 (1993)
14. Dacunha-Castelle, D., Krivine, J.L.: Application des ultraproducts a l'etude des espaces et des algebres de Banach. Stud. Math. **41**, 315–334 (1995)
15. Davis, M.: Applied Nonstandard Analysis. Wiley, New York (1977)
16. Doob, J.L.: Stochastic processes depending on a continuous parameter. Trans. Am. Math. Soc. **42**, 107–140 (1937)
17. Dunford, N., Schwartz, J.: Linear Operators Part I. Interscience, New York (1958)
18. Egoroff, D.: Sur les suites des fonctions mesurables. C. R. Hebd. Séances Acad. Sci. **152**, 244–246 (1911)
19. Füredi, Z., Loeb, P.A.: On the best constant for the Besicovitch covering theorem. Proc. Am. Math. Soc. **121**(4), 1063–1073 (1994)

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2

20. Henson, C.W., Moore, L.C.: Nonstandard analysis and the theory of Banach spaces. In: Hurd, A.E. (ed.) Nonstandard Analysis - Recent Developments, pp. 27–112. Springer, Berlin (1983)
21. Insall, M., Loeb, P.A., Marciniak, M.: End compactifications and general compactifications. J. Log. Anal. **6**(7), 1–16 (2014)
22. Keisler, H.J.: Elementary Calculus, An Infinitesimal Approach, 1st edn. Prindle, Weber & Smith, Boston (1976); 2nd edn. Prindle, Weber & Smith, Boston (1986)
23. Keisler, H.J.: An infinitesimal approach to stochastic analysis. Mem. Am. Math. Soc. **48**(297) (1984)
24. Kelley, J.L.: The Tychonoff product theorem implies the axiom of choice. Fundam. Math. **37**, 75–76 (1950)
25. Littlewood, J.E.: Lectures on the Theory of Functions. Oxford University Press, Oxford (1944)
26. Loeb, P.A.: A new proof of the Tychonoff Theorem. Am. Math. Mon. **72**(7), 711–717 (1965)
27. Loeb, P.A.: Compactifications of Hausdorff spaces. Proc. Am. Math. Soc. **22**, 627–634 (1969)
28. Loeb, P.A.: Conversion from nonstandard to standard measure spaces and applications in probability theory. Trans. Am. Math. Soc. **211**, 113–122 (1975)
29. Loeb, P.A.: Applications of nonstandard analysis to ideal boundaries in potential theory. Isr. J. Math. **25**, 154–187 (1976)
30. Loeb, P.A.: Weak limits of measures and the standard part map. Proc. Am. Math. Soc. **77**(1), 128–135 (1979)
31. Loeb, P.A.: A construction of representing measures for elliptic and parabolic differential equations. Math. Ann. **260**, 51–56 (1982)
32. Loeb, P.A.: A functional approach to nonstandard measure theory. In: Proceedings of the S. Kakutani Retirement Conference. Contemp. Math. **26**, 251–261 (1984)
33. Loeb, P.A., Talvila, E.: Covering lemmas and Lebesgue integration. Sci. Math. Jpn. **53**, 209–221 (2001)
34. Loeb, P.A., Talvila, E.: Lusin's theorem and Bochner integration. Sci. Math. Jpn. **60**, 113–120 (2004)
35. Loeb, P.A., Wolff, M. (eds.): Nonstandard Analysis for the Working Mathematician, 2nd edn. Springer, Heidelberg (2015)
36. Lusin, N.: Sur les propriétés des fonctions mesurables. C. R. Acad. Sci. Paris **154**, 1688–1690 (1912)
37. Luxemburg, W.A.J.: A general theory of monads. In: Luxemburg, W.A.J. (ed.) Applications of Model Theory to Algebra, Analysis, and Probability. Holt, Rinehart, and Winston, New York (1969)
38. Morse, A.P.: Perfect blankets. Trans. Am. Math. Soc. **61**, 418–442 (1947)
39. Nelson, E.: Internal set theory: a new approach to nonstandard analysis. Bull. Am. Math. Soc. **83**, 1165–1198 (1977)
40. Perkins, E.A.: A global intrinsic characterization of Brownian local time. Ann. Probab. **9**, 800–817 (1981)
41. Radó, T.: Sur un problème relatif à un théorème de Vitali". Fundam. Math. **11**, 228–229 (1928)
42. Raman-Sundström, M.: A pedagogical history of compactness. MAA Mon. **122**, 619–635 (2015)
43. Reifenberg, E.F.: A problem on circles. Math. Gaz. **32**, 290–292 (1948)
44. Robinson, A.: Non-standard Analysis. North-Holland, Amsterdam (1966)
45. Royden, H.L.: Real Analysis, 3rd edn. Prentice Hall, New Jersey (1987)
46. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill, New York (1987)
47. Solovay, R.M.: A model of set-theory in which every set of reals is Lebesgue measurable. Ann. Math. **92**(1), 1–56 (1970)
48. Stroyan, K.D., Bayod, J.M.: Foundations of Infinitesimal Stochastic Analysis. North-Holland, Amsterdam (1986)
49. Stroyan, K.D., Luxemburg, W.A.J.: Introduction to the Theory of Infinitesimals. Academic, New York (1976)
50. Sullivan, J.M.: An explicit bound for the Besicovitch covering theorem. Geom. Anal. **4**(2), 219–231 (1993)

51. Sun, Y.N.: A theory of hyperfinite processes: the complete removal of individual uncertainty via exact LLN. J. Math. Econ. **29**, 419–503 (1998)
52. Tao, T.: Hilbert's Fifth Problem and Related Topics. Graduate Studies in Mathematics, vol. 153. American Mathematical Society, Providence (2014)
53. Vakil, N.: Real Analysis Through Modern Infinitesimals. Cambridge University Press, Cambridge (2011)

Index

absolutely continuous function, 86
absolutely continuous measure, 88
accumulation point, 12
Alaoglu Theorem, 206
algebra, 4
algebra of functions, 172
almost everywhere, 49
antisymmetric, 6, 221
arcwise connected, 156
Ascoli-Arzela Theorem, 171
Axiom of Choice, 4, 221

Baire Category Theorem, 113
Baire first category, 113
Baire second category, 113
Baire sets, 213
Banach limit, 251
Banach space, 119
Banach spaces, 191
Besicovitch Covering Theorem, 230
Bessel's Inequality, 140
bilinear property, 128
bisection partition, 27
Bolzano-Weierstrass property, 158
Borel field, 5
Borel measurable, 50
Borel measure, 190
Borel sets, 33, 190
Borel-Cantelli, 37
Bounded Convergence Theorem, 65
bounded function, 3
bounded linear functional, 120
bounded linear map, 192
bounded variation, 85

canonical form of simple functions, 61, 97
Cantelli, 37
Cantor set, 39
Cantor-Lebesgue function, 40
Caratheodory Extension Theorem, 181
Cauchy sequence, 7
CBS Inequality, 128
chain, 221
chain rule, 247
characteristic function, 50
Chebyschev inequality, 74
Chebyshev's Inequality, 107
clopen sets, 190
closed ball, 112
closed interval, 1
closed set, 10, 109, 150
closure, 110, 150
closure point, 150
closure points, 110
cluster point, 7, 110, 153
cluster point of a net, 153
compact, 15, 157
compact support, 168
complete measure, 49, 95
Complete orthonormal family, 137
completely regular space, 161
completion, 49
component, 156
conjugate exponents, 116
conjugate space, 192
connected set, 155
continuous, 16
continuous at a point, 152
continuous function, 111
convergence in measure, 72, 102
convex, 105
convex set, 130
countable additivity, 32

P.A. Loeb, *Real Analysis*, DOI 10.1007/978-3-319-30744-2

countable subadditivity, 184
countably additive measure, 25

Daniell Integral, 189
De Morgan's Laws, 4
degenerate interval, 1
dense, 112
dense set, 46
dense subset, 149
Dini's Theorem, 22
Dirac measure, 42
directed set, 13, 153
disconnection, 155
domination, 59
doubling measure, 190
dual space, 192

Egoroff's Theorem, 54
enlargement, 249
equicontinuous, 169
equivalent functions, 115
equivalent norms, 191
essential supremum, 115
essentially bounded, 116
expectation, 101
extended real numbers, 7

factor of product space, 163
fat Cantor set, 39
Fatou's Lemma, 66
filter, 224
filter base, 225
finite intersection property, 14, 157
finite measure, 95
finite measure space, 95
finite measure support, 60
finite subcover, 157
finite subcovering, 14
finitely additive measure, 25
first axiom of countability, 149
First General Fatou Lemma, 70
first uncountable ordinal, 153
Fourier coefficient, 136
Frechet filter, 224
free ultrafilter, 236
function
  nonstandard extension, 240
Fundamental Theorem of Calculus, 83

Gelfand Integral, 205
generalized Cantor set, 39
generalized sequence, 13
greatest element, 6

Hahn decomposition, 103
harmonic function, 217, 226
Hausdorff Maximal Principle, 137
Hausdorff Maximal Principle for Total Orderings, 223
Hausdorff space, 154, 161
Hilbert space, 127
Holder's Inequality, 117
homeomorphic, 152
homeomorphism, 16, 152

imbedding, 166
index set, 3
indexed family, 135
indicator function, 50
infimum, 8
inner product, 127
inner product space, 127
integrable function, 99
integral, 63, 98
integral of complex-valued function, 102
integral of simple function, 97
integral of step function, 58
integrand, 98
integrator, 29
interior, 13, 151
internal set theory, 248
interval, 1
interval partition, 84
isolated point, 122
isometric, 152
isometric isomorphism, 196
isometry, 140, 152

Jordan decomposition, 104

least element, 6
least upper bound, 6
Lebesgue Decomposition, 135
Lebesgue Density Theorem, 84
Lebesgue Differentiation Theorem, 82, 226
Lebesgue Dominated Convergence Theorem, 70
Lebesgue integral, 63
Lebesgue measurable function, 46
Lebesgue measurable sets, 184
Lebesgue measure, 33, 184
Lebesgue outer measure, 26
limit inferior, 9
limit of a sequence, 7, 110
limit superior, 9
Lindelof, 162
linear functional, 120
linear manifold, 191

linear ordering, 6, 221
linear space, 114
linear transformation, 191
linearly independent, 136
Lipschitz function, 56
local filter base, 147
local maximal function, 81
locally connected, 156
locally pathwise connected, 177
lower Riemann integral, 58
Lusin's Theorem, 53

maximal element, 6, 221
maximal orthonormal family, 137
meager, 113
measurable function, 46, 96
measurable rectangle, 185
measurable set, 31, 95, 179
measurable space, 95
measure, 32, 95, 102
measure on an algebra, 181
measure space, 32, 95
metric, 109
metric space, 109
metrizable, 164
minimal element, 6
minimal partition, 58
minimal representation, 58
Minkowski's Inequality, 118
Monotone Convergence Theorem, 67
Morse Covering Theorem, 230
mutually singular measures, 104

near-standard point, 249
negative set, 102
negative variation, 85
neighborhood, 16, 148
net, 13
net convergence, 153
nonstandard hulls, 251
norm, 115
norm of a linear functional, 120
norm of a linear transformation, 192
normal space, 161
normed space, 115
nowhere dense, 113
null set, 95, 103

one point compactification, 168
open base, 148
open covering, 14, 157
open interval, 1
open mapping, 203
open set, 10, 109, 148, 249
order topology, 153
ordinal numbers, 153
orthogonal elements, 130
orthogonal family, 136
orthonormal family, 136
outer measure, 179

parallelogram law, 129
partial ordering, 6, 221
partition, 57
partition for a step function, 58
Partition of Unity, 208
partition refinement, 61, 97
pathwise connected, 156
Perfect set, 43
point of closure, 12, 150
pointwise closure, 169
pointwise convergence, 18
Poisson Kernel, 217
polygonal function, 19
positive linear functional, 64, 189, 208
positive set, 102
positive variation, 85
power set, 1
probability measure, 95
probability space, 95
product measure, 185
product space, 163
product topology, 163
projection, 163

Q-compactification, 167
quantifiers, 1

Radon measure, 208
Radon-Nikodym derivative, 134
Radon-Nikodym Theorem, 89
random variable, 96
real Hilbert space, 128
refinement, 85
reflexive, 202
regular Borel measure, 213
regular space, 161
relative topology, 154
relatively open, 154
remainder, 167
Riemann integrable, 58
Riemann integral, 58
Riemann sum, 57
Riemann-Lebesgue Lemma, 76
Riesz Representation Theorem, 208

second axiom of countability, 149
second dual, 202

Second General Fatou Lemma, 70
separable, 112, 149
separates points, 172
separates points and closed sets, 161
sequence converges, 153
sigma-algebra, 5
sigma-finite measure, 95
signed measure, 102
simple function, 50, 97
singular measure, 88
Span, 136
stable, 4
step function, 50
Stone-Cech compactification, 168, 169
strictly convex, 105
strong topology, 205
subsequence, 9
superset, 1
supremum, 8
symmetric difference, 51

tags, 14
topological properties, 152
topological space, 148
topology, 148
topology of pointwise convergence, 163
topology of uniform convergence on compact sets, 170
total ordering, 6, 221
total variation, 85
total variation of a measure, 104
translate, 130
translation invariant, 26
triangle inequality, 109, 129

ucc topology, 170
ultrafilter, 224
uniform boundedness principle, 113
uniform closure, 172
uniform convergence, 18
uniform convergence on compact sets, 149
uniformly continuous, 17
uniformly continuous function, 111
unit mass, 42
unordered sum, 14, 139
upper bound, 6
upper Riemann integral, 58
Urysohn Lemma, 207
Urysohn Metrization Theorem, 166
Urysohn's lemma, 161

vague topology, 205
vanishing, 62
vector lattice, 172

weak convergence of measures, 206
weak topology, 205
weak* topology, 205
well-ordering, 7