Douglas S. Bridges

Foundations of Real and Abstract Analysis

Springer

Douglas S. Bridges
Department of Mathematics
University of Waikato
Hamilton
New Zealand

Mathematics Subject Classification (1991): 26-01, 28-01, 46Axx, 46-01

Library of Congress Cataloging-in-Publication Data
Bridges, D.S. (Douglas S.), 1945-
Foundations of real and abstract analysis / Douglas S. Bridges.
p. cm. — (Graduate texts in mathematics ; 174)
Includes index.
ISBN 0-387-98239-6 (hardcover : alk. paper)
1. Mathematical analysis. I. Title. II. Series.
QA300.B69 1997
515—dc21 97-10649

ISBN 0-387-98239-6 Springer-Verlag New York Berlin Heidelberg SPIN 10524519

Dedicated to the memory of my parents:
Douglas McDonald Bridges and Allison Hogg

Sweet Analytics, 'tis thou hast ravished me.

Faustus (Marlowe)

The stone which the builders refused is become the head stone of the corner.

Psalm cxviii, 22.

...from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.

The origin of species (Darwin)

Preface

The core of this book, Chapters 3 through 5, presents a course on metric, normed, and Hilbert spaces at the senior/graduate level. The motivation for each of these chapters is the generalisation of a particular attribute of the Euclidean space $\mathbf{R}^n$: in Chapter 3, that attribute is *distance*; in Chapter 4, *length*; and in Chapter 5, *inner product.* In addition to the standard topics that, arguably, should form part of the armoury of any graduate student in mathematics, physics, mathematical economics, theoretical statistics,..., this part of the book contains many results and exercises that are seldom found in texts on analysis at this level. Examples of the latter are Wong's Theorem (3.3.12) showing that the Lebesgue covering property is equivalent to the uniform continuity property, and Motzkin's result (5.2.2) that a nonempty closed subset of Euclidean space has the unique closest point property if and only if it is convex.

The sad reality today is that, perceiving them as one of the harder parts of their mathematical studies, students contrive to avoid analysis courses at almost any cost, in particular that of their own educational and technical deprivation. Many universities have at times capitulated to the negative demand of students for analysis courses and have seriously watered down their expectations of students in that area. As a result, mathematics majors are graduating, sometimes with high honours, with little exposure to anything but a rudimentary course or two on real and complex analysis, often without even an introduction to the Lebesgue integral.

For that reason, and also in order to provide a reference for material that is used in later chapters, I chose to begin this book with a long chapter providing a fast–paced course of real analysis, covering convergence of sequences and series, continuity, differentiability, and (Riemann and Riemann–Stieltjes) integration. The inclusion of that chapter means that the prerequisite for the book is reduced to the usual undergraduate sequence of courses on calculus. (One–variable calculus would suffice, in theory, but a lack of exposure to more advanced calculus courses would indicate a lack of the mathematical maturity that is the hidden prerequisite for most senior/graduate courses.)

Chapter 2 is designed to show that the subject of differentiation does not end with the material taught in calculus courses, and to introduce the Lebesgue integral. Starting with the Vitali Covering Theorem, the chapter develops a theory of *differentiation almost everywhere* that underpins a beautiful approach to the Lebesgue integral due to F. Riesz [39]. One minor disadvantage of Riesz's approach is that, in order to handle multivariate integrals, it requires the theory of set–valued derivatives, a topic sufficiently involved and far from my intended route through elementary analysis that I chose to omit it altogether. The only place where this might be regarded as a serious omission is at the end of the chapter on Hilbert space, where I require classical vector integration to investigate the existence of weak solutions to the Dirichlet Problem in three–dimensional Euclidean space; since that investigation is only outlined, it seemed justifiable to rely only on the reader's presumed acquaintance with elementary vector calculus. Certainly, one–dimensional integration is all that is needed for a sound introduction to the L_p spaces of functional analysis, which appear in Chapter 4.

Chapters 1 and 2 form Part I (Real Analysis) of the book; Part II (Abstract Analysis) comprises the remaining chapters and the appendices. I have already summarised the material covered in Chapters 3 through 5. Chapter 6, the final one, introduces functional analysis, starting with the Hahn–Banach Theorem and the consequent separation theorems. As well as the common elementary applications of the Hahn–Banach Theorem, I have included some deeper ones in duality theory. The chapter ends with the Baire Category Theorem, the Open Mapping Theorem, and their consequences. Here most of the applications are standard, although one or two unusual ones are included as exercises.

The book has a preliminary section dealing with background material needed in the main text, and three appendices. The first appendix describes Bishop's construction of the real number line and the subsequent development of its basic algebraic and order properties; the second deals briefly with axioms of choice and Zorn's Lemma; and the third shows how some of the material in the chapters—in particular, Minkowski's Separation Theorem—can be used in the theory of Pareto optimality and competitive equilibria in mathematical economics. Part of my motivation in writing Appendix C was to indicate that "mathematical economics" is a far deeper subject than is suggested by the undergraduate texts on calculus and linear algebra that are published under that title.

I have tried, wherever possible, to present proofs so that they translate *mutatis mutandis* into their counterparts in a more abstract setting, such as that of a metric space (for results in Chapter 1) or a topological space (for results in Chapter 3). On the other hand, some results first appear as exercises in one context before reappearing as theorems in another: one example of this is the Uniform Continuity Theorem, which first appears as[1] Exercise (1.4.8: 8) in the context of a compact interval of **R**, and which is proved later, as Corollary (3.3.13), in the more general setting of a compact metric space. I hope that this procedure of double exposure will enable students to grasp the material more firmly.

The text covers just over 300 pages, but the book is, in a sense, much larger, since it contains nearly 750 exercises, which can be classified into at least the following, not necessarily exclusive, types:

- applications and extensions of the main propositions and theorems;
- results that fill in gaps in proofs or that prepare for proofs later in the book;
- pointers towards new branches of the subject;
- deep and difficult challenges for the very best students.

The instructor will have a wide choice of exercises to set the students as assignments or test questions. Whichever ones are set, as with the learning of any branch of mathematics it is essential that the student attempt as many exercises as the constraints of time, energy, and ability permit.

It is important for the instructor/student to realise that many of the exercises—especially in Chapters 1 and 2—deal with results, sometimes major ones, that are needed later in the book. Such an exercise may not clearly identify itself when it first appears; if it is not attempted then, it will provide revision and reinforcement of that material when the student needs to tackle it later. It would have been unreasonable of me to have included major results as exercises without some guidelines for the solution of the nonroutine ones; in fact, a significant proportion of the exercises of all types come with some such guideline, even if only a hint.

Although Chapters 3 through 6 make numerous references to Chapters 1 and 2, I have tried to make it easy for the reader to tackle the later chapters without ploughing through the first two. In this way the book can be used as a text for a semester course on metric, normed, and Hilbert spaces. (If

[1] A reference of the form Proposition ($a.b.c$) is to Proposition c in Section b of Chapter a; one to Exercise ($a.b.c$: d) is to the dth exercise in the set of exercises with reference number ($a.b.c$); and one to (B3) is to the 3rd result in Appendix B. Within each section, displays that require reference indicators are numbered in sequence: (1), (2), ... The counter for this numbering is reset at the start of a new section.

Chapter 2 is not covered, the instructor may need to omit material that depends on familiarity with the Lebesgue integral—in particular Section 4 of Chapter 4.) Chapter 6 could be included to round off an introductory course on functional analysis.

Chapter 1 could be used on its own as a second course on real analysis (following the typical advanced calculus course that introduces formal notions of convergence and continuity); it could also be used as a first course for senior students who have not previously encountered rigorous analysis. Chapters 1 and 2 together would make a good course on real variables, in preparation for either the material in Chapters 3 through 5 or a course on measure theory. The whole book could be used for a sequence of courses starting with real analysis and culminating in an introduction to functional analysis.

I have drawn on the resource provided by many excellent existing texts cited in the bibliography, as well as some original papers (notably [39], in which Riesz introduced the development of the Lebesgue integral used in Chapter 2). My first drafts were prepared using the T^3 *Scientific Word Processing System*; the final version was produced by converting the drafts to TEX and then using *Scientific Word.* Both T^3 and *Scientific Word* are products of TCI Software Research, Inc.

I am grateful to the following people who have helped me in the preparation of this book:

— Patrick Er, who first suggested that I offer a course in analysis for economists, which mutated into the regular analysis course from which the book eventually emerged;

— the students in my analysis classes from 1990 to 1996, who suffered various slowly improving drafts;

— Cris Calude, Nick Dudley Ward, Mark Schroder, Alfred Seeger, Doru Stefanescu, and Wang Yuchuan, who read and commented on parts of the book;

— the wonderfully patient and cooperative staff at Springer–Verlag;

— my wife and children, for their patience (in more than one sense).

It is right and proper for me here to acknowledge my unspoken debt of gratitude to my parents. This book really began 35 years ago, when, with their somewhat mystified support and encouragement, I was beginning my love affair with mathematics and in particular with analysis. It is sad that they did not live to see its completion.

Douglas Bridges
28 January 1997

Contents

Introduction

We may our ends by our beginnings know.
OF PRUDENCE (Sir John Denham)

What we now call *analysis* grew out of the calculus of Newton and Leibniz, was developed throughout the eighteenth century (notably by Euler), and slowly became logically sound (rigorous) through the work of Gauss, Cauchy, Riemann, Weierstrass, Lebesgue, and many others in the nineteenth and early twentieth centuries.

Roughly, analysis may be characterised as the study of limiting processes within mathematics. These processes traditionally include the convergence of infinite sequences and series, continuity, differentiation, and integration, on the real number line $\mathbf{R}$; but in the last 100 years analysis has moved far from the one– or finite–dimensional setting, to the extent that it now deals largely with limiting processes in infinite–dimensional spaces equipped with structures that produce meaningful abstractions of such notions as *limit* and *continuous*. Far from being merely the fantastical delight of mathematicians, these infinite–dimensional abstractions have served both to clarify phenomena whose true nature is often obscured by the peculiar structure of $\mathbf{R}$, and to provide foundations for quantum physics, equilibrium economics, numerical approximation—indeed, a host of areas of pure and applied mathematics. So important is analysis that it is no exaggeration to describe as seriously deficient any honours graduate in physics, mathematics, or theoretical economics who has not had good exposure to at least the fundamentals of metric, normed, and Hilbert space theory, if not the next step, in which metric notions all but disappear in the further abstraction of topological spaces.

Like many students of mathematics, even very good ones, you may find it hard to see the point of analysis, in which intuition often seems sacrificed to the demon of rigour. Is our intuition—algebraic, arithmetic, and geometric—not a sufficiently good guide to mathematical reality in most cases? Alas, it is not, as is illustrated by considering the differentiability of functions. (We are assuming here that you are familiar with the derivative from elementary calculus courses.)

When you first met the derivative, you probably thought that any continuous (real–valued) function—that is, loosely, one with an unbroken graph—on an interval of **R** has a derivative at all points of its domain; in other words, its graph has a tangent everywhere. Once you came across simple examples, like the absolute value function $x \mapsto |x|$, of functions whose graphs are unbroken but have no tangent at some point, it would have been natural to conjecture that if the graph were unbroken, then it had a tangent at all but a finite number of points. If you were really smart, you might even have produced an example of a continuous function, made up of lots of spikes, which was not differentiable at any of a sequence of points. This is about as far as intuition can go. But, as Weierstrass showed in the last century, and as you are invited to demonstrate in Exercise (1.5.1: 2), there exist continuous functions on **R** whose derivative does not exist anywhere. Even this is not the end of the story: in a technical sense discussed in Chapter 6, most continuous functions on **R** are nowhere differentiable! Here, then, is a dramatic failure of our intuition. We could give examples of many others, all of which highlight the need for the sort of careful analysis that is the subject of this book.

Of course, analysis is not primarily concerned with pathological examples such as Weierstrass's one of a continuous, nowhere differentiable function. Its main aim is to build up a body of concepts, theorems, and proofs that describe a large part of the mathematical world (roughly, the continuous part) and are well suited to the mathematical demands of physicists, economists, statisticians, and others. The central chapters of this book, Chapters 3 through 5, give you an introduction to some of the fundamental concepts and results of modern analysis. The earlier chapters serve either as a background reference for the later ones or, if you have not studied much real analysis before, as a rapid introduction to that topic, in preparation for the rest of the book. The final chapter introduces some of the main themes of functional analysis, the study of continuous linear mappings on infinite–dimensional spaces.

Having understood Chapters 3 through 6, you should be in a position to appreciate such other jewels of modern analysis as

- abstract measure spaces, integration, and probability theory;

- approximation theory, in which complicated types of functions are approximated by more tractable ones such as polynomials of fixed maximum degree;
- spectral theory of linear operators on a Hilbert space, generalising the theory of eigenvalues and eigenvectors of matrices;
- analysis of one and several complex variables;
- duality theory in topological vector spaces;
- Haar measure and duality on locally compact groups, and the associated abstract generalisation of the Fourier transform;
- C^*- and von Neumann algebras of operators on a Hilbert space, providing rigorous foundations for quantum mechanics;
- the theory of partial differential equations and the related potential problems of classical physics;
- the calculus of variations and optimisation theory.

These, however, are the subjects of other books. The time has come to begin this one by outlining the background material needed in the main chapters.

Throughout this book, we assume familiarity with the fundamentals of informal set theory, as found in [20]. We use the following notation for sets of numbers.

The set of natural numbers:	$\mathbf{N}$	$=$	$\{0, 1, 2, \ldots\}$.
The set of positive integers:	$\mathbf{N}^+$	$=$	$\{1, 2, 3, \ldots\}$.
The set of integers:	$\mathbf{Z}$	$=$	$\{0, -1, 1, -2, 2, \ldots\}$.
The set of rational numbers:	$\mathbf{Q}$	$=$	$\left\{\pm\frac{m}{n} : m, n \in \mathbf{N},\ n \neq 0\right\}$.

For the purposes of this preliminary section only, we accept as given the algebraic and order properties of the set $\mathbf{R}$ of real numbers, even though these are not introduced formally until Chapter 1.

When the rule and domain describing a function $f : A \to B$ are known or clearly understood, we may denote f by

$$x \mapsto f(x).$$

Note that we use the arrow $\to$ as in "the function $f : A \to B$", and the barred arrow $\mapsto$ as in "the function $x \mapsto x^3$ on $\mathbf{R}$".

We regard two functions with the same rule but different domains as different functions. In fact, we define two functions f and g to be *equal* if and only if

- they have the same domain and
- $f(x) = g(x)$ for each x in that domain.

Thus the function $x \mapsto x^2$ with domain $\mathbf{N}$ is not the same as the function $x \mapsto x^2$ with domain $\mathbf{R}$. When considering a rule that defines a function, we usually take the domain of the function as the set of all objects x (or at least all x of the type we wish to consider) to which the rule can be applied. For example, if we are working in the context of $\mathbf{R}$, we consider the domain of the function $x \mapsto 1/(x-1)$ to be the set consisting of all real numbers other than 1.

We sometimes give explicit definitions of functions by cases. For example,

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

defines a function $f : \mathbf{R} \to \{0, 1\}$.

A *sequence* is just a special kind of function: namely, one of the form $n \mapsto x_n$ with domain $\mathbf{N}^+$; x_n is then called the nth *term* of the sequence. We denote by $(x_n)_{n=1}^{\infty}$, or $(x_1, x_2, \ldots)$, or even just (x_n), the sequence whose nth term is x_n. (Of course, n is a dummy variable here; so, for example, (x_k) is the same sequence as (x_n).) If all the terms of (x_n) belong to a set X, we refer to (x_n) as a *sequence in* X. We also apply the word "sequence", and notations such as $(x_n)_{n=\nu}^{\infty}$, to a mapping $n \mapsto x_n$ whose domain has the form $\{n \in \mathbf{Z} : n \geq \nu\}$ for some integer ν.

A *subsequence* of (x_n) is a sequence of the form

$$(x_{n_k})_{k=1}^{\infty} = (x_{n_1}, x_{n_2}, x_{n_3}, \ldots),$$

where $n_1 < n_2 < n_3 < \cdots$. More generally, if f is a one–one mapping of $\mathbf{N}^+$ into itself, we write $(x_{f(n)})_{n=1}^{\infty}$, or even just $(x_{f(n)})$, to denote the sequence whose nth term is $x_{f(n)}$. This enables us, in Section 2 of Chapter 1, to make sense of an expression like $\sum_{n=1}^{\infty} x_{f(n)}$, denoting a rearrangement of the infinite series $\sum_{n=1}^{\infty} x_n$.

By a *finite sequence* we mean an ordered n–tuple $(x_1, \ldots, x_n)$, where n is any positive integer.

A nonempty set X is said to be *countable*, or to have *countably many* elements, if it is the range of a sequence. Note that a nonempty finite set is countable according to this definition. An infinite countable set is said to be *countably infinite*. We regard the empty set as being both finite and countable. A set that is not countable is said to be *uncountable,* and to have *uncountably many* elements.

Let f, g be mappings from subsets of a set X into a set Y, where Y is equipped with a binary operation $\diamond$. We introduce the corresponding *pointwise operation* $\diamondsuit$ on f and g by setting

$$(f\diamondsuit g)(x) = f(x)\diamondsuit g(x)$$

whenever $f(x)$ and $g(x)$ are both defined. Thus, taking $Y = \mathbf{R}$, we see that the (pointwise) sum of f and g is given by

$$(f + g)(x) = f(x) + g(x)$$

if $f(x)$ and $g(x)$ are both defined; and that the (pointwise) quotient of f and g is given by

$$(f/g)(x) = f(x)/g(x)$$

if $f(x)$ and $g(x)$ are defined and $g(x) \neq 0$. If $X = \mathbf{N}^+$, so that $f = (x_n)$ and $g = (y_n)$ are sequences, then we also speak of *termwise operations*; for example, the *termwise product* of f and g is the sequence $(x_n y_n)_{n=1}^{\infty}$.

Pointwise operations extend in the obvious ways to finitely many functions. In the case of a sequence $(f_n)_{n=1}^{\infty}$ of functions with values in a normed space (see Chapter 4), once we have introduced the notion of a series in a normed space, we interpret $\sum_{n=1}^{\infty} f_n$ in the obvious way.

By a *family* of elements of a set X we mean a mapping $\lambda \mapsto x_\lambda$ of a set L, called the *index set* for the family, into X. We also denote such a family by $(x_\lambda)_{\lambda\in L}$. A family with index set $\mathbf{N}^+$ is, of course, a sequence. By a *subfamily* of a family $(x_\lambda)_{\lambda\in L}$ we mean a family $(x_\lambda)_{\lambda\in J}$ where $J \subset L$.

If $(S_\lambda)_{\lambda\in L}$ is a family of sets, we write

$$\bigcup_{\lambda\in L} S_\lambda = \left\{x : \exists\lambda \in L \ \left(x \in S_\lambda\right)\right\},$$

$$\bigcap_{\lambda\in L} S_\lambda = \left\{x : \forall\lambda \in L \ \left(x \in S_\lambda\right)\right\},$$

and we call $\bigcup_{\lambda\in L} S_\lambda$ and $\bigcap_{\lambda\in L} S_\lambda$, respectively, the *union* and the *intersection* of the family $(S_\lambda)_{\lambda\in L}$.

We need some information about order relations on a set. (For fuller information about orders in general see Chapter 1 of [9].)

A binary relation R on a set X is said to be

- *reflexive* if
$$\forall a \in X \ (aRa);$$
- *irreflexive* if
$$\forall a \in X \ (\text{not}(aRa));$$
- *symmetric* if
$$\forall a, b \in X \ (aRb \Rightarrow bRa);$$

- *asymmetric* if
$$\forall a, b \in X\ (aRb \Rightarrow \text{not}(bRa));$$

- *antisymmetric* if
$$\forall a, b \in X\ ((aRb \text{ and } bRa) \Rightarrow a = b);$$

- *transitive* if
$$\forall a, b, c \in X\ ((aRa \text{ and } bRc) \Rightarrow aRc);$$

- *total* if
$$\forall a, b \in X\ (aRb \text{ or } bRa).$$

We use $\succcurlyeq$ to represent a reflexive relation, and $\succ$ to represent an irreflexive one. The notation $a \preccurlyeq b$ (respectively, $a \prec b$) is equivalent to $b \succcurlyeq a$ (respectively, $b \succ a$). When dealing with the usual order relations on the real line **R**, we use the standard symbols $\geq, >, \leq, <$ instead of $\succcurlyeq, \succ, \preccurlyeq, \prec$, respectively.

A binary relation R on a set X is said to be

- a *preorder* if it is reflexive and transitive;
- an *equivalence relation* if it is a symmetric preorder (in which case X is partitioned into disjoint *equivalence classes,* each equivalence class consisting of elements that are related under R, and the set of these equivalence classes, written X/R, is called the *quotient set* for R);
- a *partial order* if it is an antisymmetric preorder;
- a *total order* if it is a total partial order;
- a *strict partial order* if it is asymmetric and transitive—or, equivalently, if it is irreflexive and transitive.

If R is a partial order on X, we call the pair (X, R)—or, when there is no risk of confusion, just the set X itself—a *partially ordered set.*

With each preorder $\succcurlyeq$ on X we associate a strict partial order $\succ$ and an equivalence relation $\sim$ on X, defined as follows.

$$\begin{array}{lll} x \succ y & \text{if and only if} & x \succcurlyeq y \text{ and not}(y \succcurlyeq x); \\ x \sim y & \text{if and only if} & x \succcurlyeq y \text{ and } y \succcurlyeq x. \end{array}$$

If $\succcurlyeq$ is a total order, we have the *Law of Trichotomy*:

$$\forall x, y, z \in X\ \left(x \succ y \text{ or } x = y \text{ or } x \prec y\right).$$

Let S be a nonempty subset of a partially ordered set $(X, \succcurlyeq)$. An element $B \in X$ is called an *upper bound,* or *majorant,* of S (relative to $\succcurlyeq$) if $B \succcurlyeq x$ for all $x \in S$. If there exist upper bounds of S, then we say that S is *bounded above,* or *majorised.* An element $B \in X$ is called a *least upper bound,* or *supremum,* of S if the following two conditions are satisfied.

— B is an upper bound of S;

— if B' is an upper bound of S, then $B' \succcurlyeq B$.

Note that S has at most one supremum: for if B, B' are suprema of S, then $B' \succcurlyeq B \succcurlyeq B'$ and so $B' = B$, by the antisymmetry of $\succcurlyeq$. If the supremum of S exists, we denote it by $\sup S$. We also denote it by

$$\sup_{1 \leq i \leq n} x_i, \quad \max S, \quad \max_{1 \leq i \leq n} x_i, \quad \text{or} \quad x_1 \vee x_2 \vee \cdots \vee x_n$$

if $S = \{x_1, \ldots, x_n\}$ is a finite set, and by

$$\sup_{n \geq 1} x_n \quad \text{or} \quad \bigvee_{n=1}^{\infty} x_n$$

if $S = \{x_1, x_2, \ldots\}$ is a countable set; we use similar notations without further comment. An upper bound of S that belongs to S is called a *maximum element* of S, and is then a least upper bound of S. The maximum element, if it exists, of S is also called the *largest*, or *greatest*, element of S.

An element $b \in X$ is called a *lower bound,* or *minorant,* of S (relative to $\succcurlyeq$) if $x \succcurlyeq b$ for all $x \in S$. If there exist lower bounds of S, then we say that S is *bounded below,* or *minorised.* An element $b \in X$ is called a *greatest lower bound,* or *infimum,* of S if the following two conditions are satisfied.

— b is a lower bound of S;

— if b' is a lower bound of S, then $b \succcurlyeq b'$.

S has at most one infimum, which we denote by $\inf S$. When describing infima, we also use such notations as

$$\inf_{1 \leq i \leq n} x_i, \quad \min S, \quad \min_{1 \leq i \leq n} x_i, \quad \text{or} \quad x_1 \wedge x_2 \wedge \cdots \wedge x_n$$

if $S = \{x_1, \ldots, x_n\}$ is a finite set, and

$$\inf_{n \geq 1} x_n \quad \text{or} \quad \bigwedge_{n=1}^{\infty} x_n$$

if $S = \{x_1, x_2, \ldots\}$ is a countable set. A lower bound of S that belongs to S is called a *minimum element* of S, and is a greatest lower bound of S.

The minimum element, if it exists, of S is also called the *smallest*, or *least*, element of S.

The usual partial order $\geq$ on $\mathbf{R}$ gives rise to important operations on functions. If f, g are real–valued functions, we write $f \geq g$ (or $g \leq f$) to indicate that $f(x) \geq g(x)$ for all x common to the domains of f and g. Regarding $\vee$ and $\wedge$ as binary operations on $\mathbf{R}$, we define the corresponding functions $f \vee g$ and $f \wedge g$ as special cases of the notion $f \Diamond g$ previously introduced. By extension of these ideas, if $(f_n)_{n=1}^{\infty}$ is a sequence of real–valued functions, then the functions $\bigvee_{n=1}^{\infty} f_n$ and $\bigwedge_{n=1}^{\infty} f_n$ are defined by

$$\left(\bigvee_{n=1}^{\infty} f_n\right)(x) = \bigvee_{n=1}^{\infty} f_n(x),$$

$$\left(\bigwedge_{n=1}^{\infty} f_n\right)(x) = \bigwedge_{n=1}^{\infty} f_n(x),$$

whenever the right–hand sides of these equations make sense.

Now let f be a mapping of a set X into the partially ordered set $(\mathbf{R}, \geq)$. We say that f is *bounded above on* X if

$$f(X) = \{f(x) : x \in X\}$$

is bounded above as a subset of Y. We call $\sup f(X)$, if it exists, the *supremum of* f on X, and we denote it by $\sup f$, $\sup_{x \in X} f(x)$, or, in the case where X is a finite set, $\max f$. We also use obvious variations on these notations, such as $\sup_{n \geq 1} f(n)$ when $X = \mathbf{N}^{+}$. We adopt analogous definitions and notations for *bounded below on* X, *infimum of* f, $\inf f$, and $\min f$.

Finally, let f be a mapping of a partially ordered set $(X, \succcurlyeq)$ into the partially ordered set $(\mathbf{R}, \geq)$. We say that f is

— *increasing* if $f(x) \geq f(x')$ whenever $x \succcurlyeq x'$;

— *strictly increasing* if $f(x) > f(x')$ whenever $x \succ x'$;

— *decreasing* if $f(x) \leq f(x')$ whenever $x \succcurlyeq x'$; and

— *strictly decreasing* if $f(x) < f(x')$ whenever $x \succ x'$.

Note that we use “increasing” and “strictly increasing” where some authors would use “nondecreasing” and “increasing”, respectively.

Part I

Real Analysis

1

Analysis on the Real Line

...I will a round unvarnish'd tale deliver...
OTHELLO, Act 1, Scene 3

In this chapter we provide a self–contained development of analysis on the real number line. We begin with an axiomatic presentation of **R,** from which we develop the elementary properties of exponential and logarithmic functions. We then discuss the convergence of sequences and series, paying particular attention to applications of the completeness of **R**. Section 3 introduces open and closed sets, and lays the groundwork for later abstraction in the context of a metric space. Section 4 deals with limits and continuity of real–valued functions; the Heine–Borel–Lebesgue and Bolzano–Weierstrass theorems prepare us for the general, and extremely useful, notion of compactness, which is discussed in Chapter 3. The final section deals with the differential and integral calculus, a subject that is reviewed from a more advanced standpoint in Chapter 2.

1.1 The Real Number Line

Although it is possible to construct the real number line **R** from **N** using elementary properties of sets and functions, in order to take us quickly to the heart of real analysis we relegate such a construction to Appendix A and instead present a set of axioms sufficient to characterise **R.** These axioms fall into three categories: the first introduces the algebra of real numbers; the remaining two are concerned with the ordering on **R**.

Axiom R1. **R** *is a field*—that is, there exist

a binary operation $(x, y) \mapsto x + y$ of *addition* on **R**,

a binary operation $(x, y) \mapsto xy$ of *multiplication*[1] on **R,**

distinguished elements 0 (*zero*) and 1 (*one*) of **R,** with $0 \neq 1$,

a unary operation $x \mapsto -x$ (*negation*) on **R,** and

a unary operation $x \mapsto x^{-1}$ of *reciprocation,* or *inversion,* on $\mathbf{R}\setminus\{0\}$

such that for all $x, y, z \in \mathbf{R}$,

$$\begin{aligned} x + y &= y + x, \\ (x + y) + z &= x + (y + z), \\ 0 + x &= x, \\ x + (-x) &= 0, \\ xy &= yx, \\ (xy)z &= x(yz), \\ x(y + z) &= xy + xz, \\ 1x &= x, \text{ and} \\ xx^{-1} &= 1 \text{ if } x \neq 0. \end{aligned}$$

Of course, we also denote x^{-1} by $\frac{1}{x}$ or $1/x$.

Axioms R2. **R** is endowed with a total partial order $\geq$ (*greater than or equal to*), and hence an associated strict partial order $>$ (*greater than*), such that

- if $x \geq y$, then $x + z \geq y + z$, and
- if $x \geq 0$ and $y \geq 0$, then $xy \geq 0$.

Axiom R3. *The least–upper–bound principle:* if a nonempty subset S of **R** is bounded above relative to the relation $\geq$, then it has a (unique) least upper bound.

The elements of **R** are called *real numbers.* We say that a real number x is

- *positive* if $x > 0$,
- *negative* if $-x > 0$, and

[1] For clarity, we sometimes write $x \cdot y$ or $x \times y$ for the product xy.

- *nonnegative* if $x \geq 0$.

We denote the set of positive real numbers by $\mathbf{R}^+$, and the set of nonnegative real numbers by $\mathbf{R}^{0+}$.

Many of the fundamental arithmetic and order properties of **R** are immediate consequences of results in the elementary theories of fields and partial orders, respectively. A number of these, illustrating the interplay between the algebra and the ordering on **R**, are given in the next set of exercises.[2]

(1.1.1) Exercises

Prove each of the following statements, where x, y, x_i, y_i $(1 \leq i \leq n)$ are real numbers.

.1 If $x_i \geq y_i$ for each i, then $\sum_{i=1}^n x_i \geq \sum_{i=1}^n y_i$. If also $x_k > y_k$ for some k, then $\sum_{i=1}^n x_i > \sum_{i=1}^n y_i$.

.2 $x \geq y$ if and only if $x + z \geq y + z$ for all $z \in \mathbf{R}$; this remains true with each instance of $\geq$ replaced by one of $>$.

.3 If $x_i \geq 0$ for each i and $\sum_{i=1}^n x_i = 0$, then $x_1 = x_2 = \cdots = x_n = 0$.

.4 The following are equivalent: $x \geq y$, $x - y \geq 0$, $-y \geq -x$, $0 \geq y - x$; these equivalences also hold with $\geq$ replaced everywhere by $>$.

.5 If $x \geq y$ and $z \geq 0$, then $xz \geq yz$.

.6 If $x > 0$ and $y > 0$, then $xy > 0$; if $x > 0$ and $0 > y$, then $0 > xy$; if $0 > x$ and $0 > y$, then $xy > 0$; and these results hold with $>$ replaced everywhere by $\geq$.

.7 $x^2 \geq 0$, and $x^2 = 0$ if and only if $x = 0$.

.8 If $x > 0$, then $x^{-1} > 0$; and if $x < 0$, then $x^{-1} < 0$.

.9 $x \geq y$ if and only if $xz \geq yz$ for all $z > 0$.

.10 $x > y > 0$ if and only if $y^{-1} > x^{-1} > 0$.

.11 $\max\{x, y\} \geq 0$ if and only if $x \geq 0$ or $y \geq 0$; $\max\{x, y\} > 0$ if and only if $x > 0$ or $y > 0$.

.12 $\min\{x, y\} \geq 0$ if and only if $x \geq 0$ and $y \geq 0$; $\min\{x, y\} > 0$ if and only if $x > 0$ and $y > 0$.

---

[2] If you are comfortable with the elementary field and order properties of **R**, then you can safely omit Exercises (1.1.1) and (1.1.2).

.13 The mapping

$$n \mapsto n1 = \begin{cases} 0 & \text{if } n = 0 \\ \underbrace{1 + 1 + \cdots + 1}_{n \text{ terms}} & \text{if } n \geq 1 \end{cases}$$

from **N** into **R** is one–one and preserves order, addition, and multiplication.

We use this mapping to identify **N** with the subset $\{n1 : n \in \mathbf{N}\}$ of **R.** In turn, we then identify a negative integer n with $-(-n)\,1$, and a rational number m/n with the real number mn^{-1}. We make these identifications without further comment.

.14 If S is a nonempty majorised set of integers, then $m = \sup S$ is an integer. (Assume the contrary and obtain integers n, n' such that $m - 1 < n < n' < m$.)

.15 There exists $n \in \mathbf{Z}$ such that $n - 1 \leq x < n$. (If $x \geq 0$, apply the least–upper–bound principle to $S = \{k \in \mathbf{Z} : k \leq x\}$.)

.16 If $x > 0$ and $y \geq 0$, then there exists $n \in \mathbf{N}^+$ such that $nx > y$. (Consider $\{k \in \mathbf{N} : kx \leq y\}$.)

This important property is sometimes introduced as an axiom, the *Axiom of Archimedes*.

.17 $x > 0$ if and only if there exists a positive integer $n > x^{-1}$.

.18 $x \geq 0$ if and only if $x \geq -1/n$ for all positive integers n.

.19 **Q** is *order dense* in **R**—that is, if $x < y$, then there exists $q \in \mathbf{Q}$ such that $x < q < y$. (Reduce to the case $y > 0$. Choose in turn integers $n > 1/(y - x)$ and $k \geq ny$, and let m be the least integer such that $y \leq m/n$. Show that $x < (m - 1)/n < y$.)

.20 If S and T are nonempty majorised sets of positive numbers, then

$$\sup \{st : s \in S,\ t \in T\} = \sup S \times \sup T.$$

.21 The following are equivalent conditions on nonempty subsets X and Y of **R**.

(i) $x \leq y$ for all $x \in X$ and $y \in Y$.

(ii) There exists $\tau \in \mathbf{R}$ such that $x \leq \tau \leq y$ for all $x \in X$ and $y \in Y$.

Each real number x has a corresponding *absolute value*, defined as

$$|x| = \max\{x, -x\}.$$

(1.1.2) Exercises

Prove each of the following statements about real numbers x, y, ε.

.1 $|x| \geq 0$, and $|x| = 0$ if and only if $x = 0$.

.2 $|x| \leq \varepsilon$ if and only if $-\varepsilon \leq x \leq \varepsilon$.

.3 $|x| < \varepsilon$ if and only if $-\varepsilon < x < \varepsilon$.

.4 $x = 0$ if and only if either $|x| \leq \varepsilon$ for each $\varepsilon > 0$ or else $|x| < \varepsilon$ for each $\varepsilon > 0$.

.5 $|x + y| \leq |x| + |y|$ (*triangle inequality*).

.6 $|x - y| \geq ||x| - |y||$.

.7 $|xy| = |x|\,|y|$.

So far we have not indicated how useful the least–upper–bound principle is. In fact, it is not only useful, but essential: the field **Q** of rational numbers, with its usual ordering $>$, satisfies all the properties listed in axioms **R1** and **R2**, so we need something more to distinguish **R** from **Q**. Moreover, without the least–upper–bound principle or some property equivalent to it, we cannot even prove that a positive real number has a square root.

We now sketch how the least–upper–bound principle enables us to define a^r for any $a > 0$ and any $r \in \mathbf{R}$. When n is an integer, a^n is defined as in elementary algebra. So our first real task is to define $a^{m/n}$ when m and n are nonzero integers; this we do by setting

$$a^{m/n} = \sup\{x \in \mathbf{R} : x^n < a^m\}. \tag{1}$$

Of course, we are using the least–upper–bound principle here, so we must ensure that the set on the right–hand side of (1) is both nonempty and bounded above. To prove that it is nonempty, we use the Axiom of Archime– des (Exercise (1.1.1: 16)) to find a positive integer k such that $ka^m > 1$; then $k^n a^m \geq ka^m > 1$, so $(1/k)^n < a^m$. On the other hand, as

$$\begin{aligned} (1 + a^m)^n \geq 1 + na^m > a^m &\quad \text{if } n \geq 1, \text{ and} \\ \left(\tfrac{1}{1+a^m}\right)^n \geq 1 - na^m > a^m &\quad \text{if } n \leq -1, \end{aligned}$$

the set in question is bounded above (by $1 + a^m$ in the first case, and by $1/(1 + a^m)$ in the second). Hence $a^{m/n}$ exists.

Our first result enables us to prove some basic properties of $a^{m/n}$.

(1.1.3) Lemma. *Let $a > 0$ and s be real numbers, and m, n positive integers such that $s^n < a^m$. Then there exists $t \in \mathbf{R}$ such that $s < t$ and $t^n < a^m$.*

Proof. Using Exercise (1.1.1: 16), choose a positive integer N such that

$$0 < N^{-1} < \min\left\{1,\, 2^{-n}\,(1+|s|)^{-n}\,(a^m - s^n)\right\}.$$

Writing $t = s + N^{-1}$ and using the binomial theorem, we have

$$\begin{aligned} t^n &= \sum_{k=0}^{n} \binom{n}{k} s^{n-k} N^{-k} \\ &\leq s^n + \sum_{k=1}^{n} \binom{n}{k} |s|^{n-k}\, N^{-1} \\ &< s^n + N^{-1} \sum_{k=0}^{n} \binom{n}{k} (1+|s|)^{n-k} \\ &< s^n + (1+|s|)^n\, N^{-1} \sum_{k=0}^{n} \binom{n}{k} \\ &= s^n + 2^n\,(1+|s|)^n\, N^{-1} \\ &< a^m, \end{aligned}$$

as we required. □

Taking $s = 0$ in this lemma, we see that $a^{m/n} > 0$. The lemma also enables us to prove that

$$\left(a^{m/n}\right)^n = a^m. \tag{2}$$

For if $\left(a^{m/n}\right)^n < a^m$, then, by Lemma (1.1.3), there exists $t > a^{m/n}$ such that $t^n < a^m$, which contradicts the definition of $a^{m/n}$; on the other hand, that same definition ensures that $\left(a^{m/n}\right)^n \leq a^m$ and hence that (2) holds.

Using (2) and methods familiar from elementary algebra courses, we can now prove the usual *laws of indices*,

$$\begin{aligned} a^r a^s &= a^{r+s}, \\ (a^r)^s &= a^{rs}, \end{aligned}$$

when the indices r, s are rational.

We next extend the definition of a^r to cover all $r \in \mathbf{R}$. To begin with, we consider the case $a > 1$, when we define

$$a^r = \sup\{a^q : q \in \mathbf{Q},\, q < r\}. \tag{3}$$

It is left as an exercise to show that the set on the right–hand side of (3) is nonempty and bounded above, and that if r is rational, this definition gives a^r the same value as the one given by our earlier definition. We can now prove the laws of indices for arbitrary $r, s \in \mathbf{R}.$ Taking the first law as an illustration, we observe that if u, v are rational numbers with $u < r$ and $v < s$, then $u + v < r + s$, so

$$a^u a^v = a^{u+v} \leq a^{r+s}.$$

By Exercise (1.1.1: 20),

$$\begin{aligned} a^r a^s &= \sup\{a^u : u \in \mathbf{Q},\ u < r\} \times \sup\{a^v : v \in \mathbf{Q},\ v < s\} \\ &= \sup\{a^u a^v : u, v \in \mathbf{Q},\ u < r,\ v < s\} \\ &\leq a^{r+s}. \end{aligned}$$

On the other hand, if $q \in \mathbf{Q}$ and $q < r+s$, then we choose rational numbers u, v with $u < r$, $v < s$, and $q = u + v$: to do so, we use Exercise (1.1.1: 19) to find $u \in \mathbf{Q}$ with $q - s < u < r$ and we then set $v = q - u$. We have

$$a^q = a^{u+v} = a^u a^v \leq a^r a^s.$$

Hence

$$a^{r+s} = \sup\{a^q : q \in \mathbf{Q},\ q < r + s\} \leq a^r a^s,$$

and therefore $a^r a^s = a^{r+s}$.

It remains to define

$$a^r = \begin{cases} \left(a^{-1}\right)^{-r} & \text{if } 0 < a < 1 \\ 1 & \text{if } a = 1 \end{cases}$$

and to verify—routinely—that the laws of indices hold in these cases also.

(1.1.4) Exercises

.1 Let $a > 1$ and let $r \in \mathbf{R}$. Prove that $\{a^q : q \in \mathbf{Q},\ q < r\}$ is nonempty and bounded above. Prove also that if $r = m/n$ for integers m, n with $n \neq 0$, then definitions (1) and (3) give the same value for a^r.

.2 Prove that if $0 < a \neq 1$ and $a^x = 1$, then $x = 0$. (Consider first the case where $a > 1$, and note that if $q \in \mathbf{Q}$ and $a^q \leq 1$, then $q \leq 0$.)

.3 Let $a > 0$ and $x > y$. Prove that if $a > 1$, then $a^x > a^y$; and that if $a < 1$, then $a^x < a^y$.

.4 Prove that if $a > 0$, then for each $x > 0$ there exists a unique $y \in \mathbf{R}$ such that $a^y = x$. (First take $a > 1$ and $x > 1$. Write $a = 1 + t$ and, by expanding $(1 + t)^n$, compute $n \in \mathbf{N}^+$ such that $a^n > x$. Then consider $\{q \in \mathbf{Q} : a^q \leq x\}$.)

.5 Let f be a strictly increasing mapping of $\mathbf{R}$ onto $\mathbf{R}^+$ such that $f(0) = 1$ and $f(x+y) = f(x)f(y)$. Prove that $f(x) = a^x$, where $a = f(1) > 1$. (First prove that $f(q) = a^q$ for all rational q.)

If $a > 0$, Exercise (1.1.4: 4) allows us to define $\log_a$, the *logarithmic function with base* a, as follows. For each $x > 0$,

$$y = \log_a x \quad \text{if and only if} \quad a^y = x.$$

This function has domain $\mathbf{R}^+$ and maps $\mathbf{R}^+$ onto $\mathbf{R}$. From the laws of indices we easily deduce the *laws of logarithms*:

$$\begin{aligned} \log_a xy &= \log_a x + \log_a y, \\ \log_a (x^r) &= r \log_a x, \\ \log_b x &= \log_b a \times \log_a x, \ \text{ where } b > 0. \end{aligned}$$

Anticipating the theory of convergence of series from the next section, we introduce the number

$$\mathrm{e} = \sum_{n=0}^{\infty} \frac{1}{n!}$$

and call $\log_{\mathrm{e}}$ the *natural logarithmic function* on $\mathbf{R}^+$. It is customary to denote $\log_{\mathrm{e}}$ by either log or ln.

(1.1.5) Exercises

.1 Prove the laws of logarithms.

.2 Prove that if $a > 1$, then the function $\log_a$ is strictly increasing; and that if $0 < a < 1$, then $\log_a$ is strictly decreasing.

.3 Let $a > 1$, and let f be an increasing mapping of $\mathbf{R}^+$ into $\mathbf{R}$ such that $f(a) = 1$ and $f(xy) = f(x) + f(y)$. Prove that $f(x) = \log_a x$.

For convenience, we collect here the definitions of the various types of interval in $\mathbf{R}$.

The *open intervals* are the sets of the following forms, where a, b are real numbers with $a < b$:

$$\begin{aligned} (a, b) &= \{x \in \mathbf{R} : a < x < b\}, \\ (a, \infty) &= \{x \in \mathbf{R} : a < x\}, \\ (-\infty, b) &= \{x \in \mathbf{R} : x < b\}, \\ (-\infty, \infty) &= \mathbf{R}. \end{aligned}$$

The *closed intervals* are the sets of the following forms, where a, b are real numbers with $a \leq b$:

$$\begin{aligned} [a, b] &= \{x \in \mathbf{R} : a \leq x \leq b\}, \\ [a, \infty) &= \{x \in \mathbf{R} : a \leq x\}, \\ (-\infty, b] &= \{x \in \mathbf{R} : x \leq b\}. \end{aligned}$$

By convention, $\mathbf{R}$ is regarded as both an open interval and a closed interval. The remaining types of interval are:

$$\begin{array}{lcll} \textit{half open on the left}\textbf{:} & (a, b] & = & \{x \in \mathbf{R} : a < x \leq b\}, \\ \textit{half open on the right}\textbf{:} & [a, b) & = & \{x \in \mathbf{R} : a \leq x < b\}. \end{array}$$

Intervals of the form $[a, b]$, (a, b), $[a, b)$, or $(a, b]$, where $a, b \in \mathbf{R}$, are said to be *finite* or *bounded*, and to have *left endpoint* a, *right endpoint* b, and *length* $b - a$. Intervals of the remaining types are called *infinite* and are said to have length ∞. The length of any interval I is denoted by $|I|$. A bounded closed interval in $\mathbf{R}$ is also called a *compact interval*.

Finally, we define the *complex numbers* to be the elements of the set $\mathbf{C} = \mathbf{R} \times \mathbf{R}$, with the usual equality and with algebraic operations of addition and multiplication defined, respectively, by the equations

$$\begin{aligned} (x, y) + (x', y') &= (x + x', y + y'), \\ (x, y) \times (x', y') &= (xx' - yy', xy' + x'y). \end{aligned}$$

Then $x \mapsto (x, 0)$ is a one–one mapping of $\mathbf{R}$ onto the set $\mathbf{C} \times \{0\}$ and is used to identify $\mathbf{R}$ with that subset of $\mathbf{C}$. With this identification, we have $\mathrm{i}^2 = -1$, where i is the complex number $(0, 1)$; so the complex number (x, y) can be identified with the expression $x + \mathrm{i}y$. The real numbers x and y are then called the *real* and *imaginary parts* of $z = (x, y)$, respectively, and we write

$$\begin{aligned} x &= \mathrm{Re}(x, y), \\ y &= \mathrm{Im}(x, y). \end{aligned}$$

The *conjugate* of z is

$$z^* = (x, -y) = x - \mathrm{i}y,$$

and the *modulus* of z is

$$|z| = \sqrt{x^2 + y^2}.$$

In the remainder of this book we assume the basic properties of the real and complex numbers such as those found in the foregoing exercises.

1.2 Sequences and Series

Although often relegated to a minor role in courses on real analysis, the theory of convergence of sequences and series in $\mathbf{R}$ provides both a model for more abstract convergence theories such as those in our later chapters, and many important examples.

It is convenient to introduce here two useful expressions about properties of positive integers. Let $P(m,n)$ be a property applicable to pairs (m,n) of positive integers. If there exists N such that $P(m,n)$ holds for all $m,n \geq N$, then we say that $P(m,n)$ holds *for all sufficiently large* m *and* n. We interpret similarly the statement $P(n)$ *holds for all sufficiently large* n, where $P(n)$ is a property applicable to positive integers n. On the other hand, if for each positive integer i there exists a positive integer $j > i$ such that $P(j)$ holds, then we say that $P(n)$ holds *for infinitely many values of* n.

We say that a sequence[3] (a_n) of real numbers *converges* to a real number a, called the *limit* of (a_n), if for each $\varepsilon > 0$ there exists a positive integer N, depending on ε, such that $|a - a_n| \leq \varepsilon$ whenever $n \geq N$. Thus (a_n) converges to a if and only if for each $\varepsilon > 0$ we have $|a - a_n| \leq \varepsilon$ for all sufficiently large n. In that case we write

$$\lim_{n\to\infty} a_n = a$$

or

$$a_n \to a \text{ as } n \to \infty,$$

and we also say that a_n *tends to* a as $n \to \infty$.

On the other hand, we say that (a_n) *diverges to* ∞, and we write

$$a_n \to \infty \text{ as } n \to \infty,$$

if for each $K > 0$ we have $a_n > K$ for all sufficiently large n. If for each $K > 0$ we have $a_n < -K$ for all sufficiently large n, then we say that (a_n) *diverges to* $-\infty$, and we write

$$a_n \to -\infty \text{ as } n \to \infty.$$

(1.2.1) Exercises

.1 Prove that if (a_n) converges to both a and a', then $a = a'$. (Show that $|a - a'| < \varepsilon$ for each $\varepsilon > 0$. This exercise justifies the use of the definite article in the phrase "the limit of (a_n)".)

---

[3]We can extend the definitions of convergence and divergence of sequences in the obvious ways to cover families of the form $(a_n)_{n\geq\nu}$, where $\nu \in \mathbf{Z}$; all that matters is that a_n be defined for all sufficiently large positive integers n. This observation makes sense of the last part of Proposition (1.2.2), where we discuss the limit of a quotient of two sequences.

.2 Let $c > 0$. Prove that (a_n) converges to a if and only if for each $\varepsilon > 0$ there exists a positive integer N, depending on ε, such that $|a - a_n| \leq c\varepsilon$ for all $n \geq N$.

.3 Prove that if a sequence (a_n) converges to a limit, then it is *bounded*, in the sense that there exists $c > 0$ such that $|a_n| \leq c$ for all n.

.4 Let $r \in \mathbf{R}$, and let (a_n) be a convergent sequence in $\mathbf{R}$ such that $\lim_{n\to\infty} a_n > r$. Prove that $a_n > r$ for all sufficiently large n.

.5 Let $r \in \mathbf{R}$, and let (a_n) be a convergent sequence in $\mathbf{R}$ such that $a_n \geq r$ for all sufficiently large n. Prove that $\lim_{n\to\infty} a_n \geq r$.

.6 Prove that if (a_n) diverges to infinity and (b_n) converges to a limit $b \in \mathbf{R}$, then the sequence $(a_n + b_n)$ diverges to infinity.

The process of taking limits of sequences preserves the basic operations of arithmetic.

(1.2.2) Proposition. *Let (a_n) and (b_n) be sequences of real numbers converging to limits a and b, respectively. Then as $n \to \infty$,*

$$\begin{aligned} a_n + b_n &\to a + b, \\ a_n - b_n &\to a - b, \\ a_n b_n &\to ab, \\ \max\{a_n, b_n\} &\to \max\{a, b\}, \\ \min\{a_n, b_n\} &\to \min\{a, b\}, \ \text{and} \\ |a_n| &\to |a|. \end{aligned}$$

If also $b \neq 0$, then $b_n \neq 0$ for all sufficiently large n, and $a_n/b_n \to a/b$ as $n \to \infty$.

Proof. We prove only the last statement, leaving the other cases to Exercise (1.2.3: 1).

Assume that $b \neq 0$. Then, by Exercise (1.2.1: 4), there exists N_0 such that $|b_n| > \frac{1}{2}|b|$, and therefore a_n/b_n is defined, for all $n \geq N_0$. Given $\varepsilon > 0$, choose $N \geq N_0$ such that $|a_n - a| < \varepsilon$ and $|b_n - b| < \varepsilon$ for all $n \geq N$. For all such n we have

$$\begin{aligned} \left|\frac{a_n}{b_n} - \frac{a}{b}\right| &= \frac{|ba_n - ab_n|}{|b_n|\,|b|} \\ &\leq \frac{|b(a_n - a) + a(b - b_n)|}{\frac{1}{2}|b|^2} \\ &\leq 2|b|^{-2}\left(|b|\,|a_n - a| + |a|\,|b - b_n|\right) \\ &\leq 2|b|^{-2}\left(|a| + |b|\right)\varepsilon. \end{aligned}$$

The result now follows from Exercise (1.2.1: 2). □

(1.2.3) Exercises

.1 Prove the remaining parts of Proposition (1.2.2).

.2 Prove that if $k \geq 2$ and $\nu \geq 1$ are integers, then

$$\left(1+\tfrac{1}{k}\right)^{k-1} > \tfrac{3}{2} \text{ and } \left(1+\tfrac{1}{k}\right)^{\nu(k-1)} > \nu.$$

Hence prove that if $0 \leq |r| < 1$, then $r^n \to 0$ as $n \to \infty$. (Given $\varepsilon > 0$, first choose ν such that $1/\nu < \varepsilon$. Then choose k such that $|r|^{-1} > 1 + k^{-1}$.)

.3 Prove that if $r > 1$, then $r^n \to \infty$ as $n \to \infty$.

.4 Prove that if $a > 1$, then $\log_a n \to \infty$ as $n \to \infty$.

.5 Prove that if $r = \lim_{n\to\infty} a_n$, then $r = \lim_{k\to\infty} a_{n_k}$ for any subsequence $(a_{n_k})_{k=1}^{\infty}$ of (a_n).

.6 Let (a_n) be a sequence of real numbers such that the subsequences $(a_{2n})_{n=1}^{\infty}$ and $(a_{2n+1})_{n=1}^{\infty}$ both converge to the limit l. Prove that (a_n) converges to l.

.7 Let (a_n) be a sequence in **R**. Prove that if the three subsequences $(a_{2n}), (a_{2n+1})$, and (a_{3n}) are convergent, then so is (a_n).

.8 Give an example of a sequence (a_n) of real numbers with the following properties.

(i) (a_n) is not convergent;

(ii) for each $k \geq 2$ the subsequence $(a_{kn})_{n=1}^{\infty}$ is convergent.

(Split your definition of a_n into two cases—one when n is prime, the other when n is composite.)

When we apply notions such as bounded above, supremum, and infimum to a sequence (s_n) of real numbers, we are really applying them to the set $\{s_n : n \geq 1\}$ of terms of the sequence. Thus the supremum (respectively, infimum) of a majorised (respectively, minorised) sequence (s_n) is denoted by $\sup_{n\geq 1} s_n$, or just $\sup s_n$ (respectively, $\inf_{n\geq 1} s_n$, or just $\inf s_n$).

The next result, known as the *monotone sequence principle*, is a powerful tool for proving the existence of limits.

(1.2.4) Proposition. *An increasing majorised sequence of real numbers converges to its least upper bound; a decreasing minorised sequence of real numbers converges to its greatest lower bound.*

Proof. Let (s_n) be an increasing majorised sequence of real numbers, and s its least upper bound. For each $\varepsilon > 0$, since $s - \varepsilon$ is not an upper bound of (s_n), there exists N such that $s_N > s - \varepsilon$. But (s_n) is both increasing and bounded above by s; so for all $n \geq N$ we have $s - \varepsilon < s_n \leq s$ and therefore $|s - s_n| < \varepsilon$. Since $\varepsilon > 0$ is arbitrary, it follows that $s_n \to s$ as $n \to \infty$.

The case of a decreasing minorised sequence is left as an exercise. □

(1.2.5) Exercises

.1 Prove the second part of the last proposition in two ways.

.2 Prove that an increasing sequence of nonnegative real numbers diverges to infinity if and only if it is not bounded above.

.3 Let $a > 1$ and $x > 0$. Prove that there exists an integer m such that $a^m \leq x < a^{m+1}$. (First take $x \geq 1$, and consider the sequence $(a^n)_{n=0}^{\infty}$.)

.4 Discuss the convergence of the sequence (a_n) defined by $a_{n+1} = \sqrt{ra_n}$, where a_1 and r are positive numbers.

.5 Prove that if $0 < a$ and $k \in \mathbf{N}$, then $\lim_{n\to\infty} \sqrt[n+k]{a} = 1$. (First consider the case where $k = 0$ and $0 < a < 1$. Apply the monotone sequence principle to show that the sequence $\left(\sqrt[n]{a}\right)_{n=1}^{\infty}$ converges to a limit l. By considering the subsequence $\left(\sqrt[2n]{a}\right)$, show that $\sqrt{l} = l$.)

.6 Prove that if (a_n) is a sequence of positive numbers such that

$$l = \lim_{n\to\infty} \frac{a_{n+1}}{a_n}$$

exists, then $\lim_{n\to\infty} \sqrt[n]{a_n} = l$. By considering the sequence

$$1, a, ab, a^2b, a^2b^2, a^3b^2, a^3b^3, \ldots,$$

where a, b are distinct positive numbers, show that the converse is false.

.7 Prove that if $n \geq 2$, then $(n+1)^n \leq n^{n+1}$. Use this to show that $l = \lim_{n\to\infty} \sqrt[n]{n}$ exists. By considering the subsequence $\left(\sqrt[2n]{2n}\right)_{n=1}^{\infty}$, prove that $l = 1$. Hence show that if $a > 1$, then $\lim_{n\to\infty} \left(n^{-1} \log_a n\right) = 0$.

.8 Prove that the sequence $\left(\left(1 + n^{-1}\right)^n\right)_{n=1}^{\infty}$ is convergent. (An interesting proof of this result, based on the well–known inequality involving arithmetic and geometric means, is found in [32].)

.9 Let (a_n) be a sequence of real numbers. If (a_n) is bounded above, then its *upper limit*, or *limit superior*, is defined to be

$$\limsup a_n = \inf_{n\geq 1} \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\}$$

if the infimum on the right exists. Prove that a real number s equals $\limsup a_n$ if and only if for each $\varepsilon > 0$,

— $a_n < s + \varepsilon$ for all sufficiently large n, and

— $a_n > s - \varepsilon$ for infinitely many values of n.

Prove also that

$$\limsup a_n = \lim_{n\to\infty} \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\}.$$

.10 If (a_n) is bounded below, then its *lower limit*, or *limit inferior*, is defined to be

$$\liminf a_n = \sup_{n\geq 1} \inf\{a_n, a_{n+1}, a_{n+2}, \ldots\}$$

if the supremum on the right exists. Establish necessary and sufficient conditions for a real number l to equal $\liminf a_n$.

.11 Prove that $a_n \to a \in \mathbf{R}$ as $n \to \infty$ if and only if

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

A sequence $(S_n)_{n=1}^{\infty}$ of subsets of $\mathbf{R}$ is said to be *nested*, or *descending*, if $S_1 \supset S_2 \supset S_3 \supset \cdots$. We make good use of the following *nested intervals principle*.

(1.2.6) Proposition. *The intersection of a nested sequence of closed intervals in* $\mathbf{R}$ *is nonempty.*

Proof. Let $([a_n, b_n])$ be a nested sequence of closed intervals in $\mathbf{R}$. Then

$$a_1 \leq a_n \leq a_{n+1} \leq b_{n+1} \leq b_n \leq b_1 \tag{1}$$

for each n. By Proposition (1.2.4), (a_n) converges to its least upper bound a, and (b_n) converges to its greatest lower bound b. It follows from the inequalities (1) and Exercise (1.2.1: 5) that $a \leq b$. So for each n, $a_n \leq a \leq b \leq b_n$ and therefore $a \in [a_n, b_n]$. □

The following elementary lemma leads to simple proofs of several important results in analysis.

(1.2.7) Lemma. *If* (a_n) *is a sequence of real numbers, then at least one of the following holds.*

(i) (a_n) *has a constant subsequence;*

(ii) (a_n) *has a strictly increasing subsequence;*

(iii) (a_n) *has a strictly decreasing subsequence.*

Proof. Suppose that (a_n) contains no constant subsequence, and consider the set

$$S = \{n \in \mathbf{N}^+ : \forall k \geq n \ (a_n \geq a_k)\} .$$

If S is bounded, then there exists N such that

$$\forall n \geq N \, \exists k > n \ (a_k > a_n) ,$$

and a simple inductive construction produces positive integers $N \leq n_1 < n_2 < \cdots$ such that $a_{n_{k+1}} > a_{n_k}$ for each k. If, on the other hand, S is unbounded, then we can compute $n_1 < n_2 < \cdots$ such that $a_{n_k} \geq a_{n_{k+1}}$ for each k. In that case, since $(a_{n_k})_{k=1}^{\infty}$ contains no constant subsequence, for each k there exists $j > k$ such that $a_{n_k} > a_{n_j}$; it is now straightforward to construct a strictly decreasing subsequence of (a_{n_k}). □

(1.2.8) Corollary. *A bounded sequence of real numbers has a convergent subsequence.*

Proof. This follows from Lemma (1.2.7) and the monotone sequence principle. □

A sequence (a_n) of real numbers is called a *Cauchy sequence* if for each $\varepsilon > 0$ there exists a positive integer N, depending on ε, such that $|a_m - a_n| \leq \varepsilon$ for all $m, n \geq N$.

(1.2.9) Exercises

.1 Prove that a convergent sequence of real numbers is a Cauchy sequence.

.2 Prove that a Cauchy sequence is bounded.

.3 Prove that if a Cauchy sequence (a_n) has a subsequence that converges to a limit $a \in \mathbf{R}$, then (a_n) converges to a.

.4 Let (a_n) be a bounded sequence each of whose convergent subsequences converges to the same limit. Prove that (a_n) converges to that limit. (cf. Exercises (1.2.3: 6 and 7). By Corollary (1.2.8), there is a subsequence (a_{n_k}) that converges to a limit l. Suppose that (a_n) does not converge to l, and derive a contradiction.)

One of the most important results in convergence theory says that not only does a Cauchy sequence of real numbers *appear* to converge, in that its terms get closer and closer to each other as their indices increase, but it actually does converge.

A subset S of **R** is said to be *complete* if each Cauchy sequence in S converges to a limit that belongs to S.

(1.2.10) Theorem. **R** *is complete.*

Proof. Let (a_n) be a Cauchy sequence in **R**. Then (a_n) is bounded, by Exercise (1.2.9: 2). It follows from Corollary (1.2.8) that (a_n) has a convergent subsequence; so (a_n) converges, by Exercise (1.2.9: 3). □

(1.2.11) Exercises

.1 Find an alternative proof of the completeness of **R**. (Given a Cauchy sequence (a_n) in **R**, consider $\liminf a_n$.)

.2 Show that if, in the system of axioms for **R**, the least–upper–bound principle is replaced by the Axiom of Archimedes (Exercise (1.1.1: 16)), then the nested intervals principle is equivalent to the completeness of **R**. Can you spot where you have used the Axiom of Archimedes?

.3 Under the conditions of the preceding exercise, show that the least–upper–bound principle follows from the completeness of **R**. (Assuming that **R** is complete, consider a nonempty majorised subset S of **R**. Choose $s_1 \in S$ and $b_1 \in B$, where B is the set of upper bounds of S. Construct a sequence (s_n) in S and a sequence (b_n) in B such that

$$s_n \leq s_{n+1} \leq b_{n+1} \leq b_n$$

and

$$0 \leq b_{n+1} - s_{n+1} \leq \tfrac{1}{2}(b_n - s_n).$$

Prove that $0 \leq b_n - b_m \leq 2^{-n+2}(b_1 - s_1)$ whenever $m \geq n$, that (s_n) and (b_n) converge to the same limit b, and that $b = \sup S$.)

.4 Prove *Cantor's Theorem*: if (a_n) is a sequence of real numbers, then in any closed interval of **R** with positive length there exists a real number x such that $x \neq a_n$ for each n. (For each $x \in \mathbf{R}$ and each nonempty $S \subset \mathbf{R}$, define the *distance from x to S* to be the real number

$$\rho(x, S) = \inf\{|x - s| : s \in S\}.$$

First prove the following lemma. If $I = [a, b]$ is a closed interval with positive length, and J_1, J_2, J_3 are the left, middle, and right closed thirds of I, then for each real number x either $\rho(x, J_1) > 0$

or $\rho(x, J_3) > 0$. Use this lemma to construct an appropriate nested sequence of closed intervals.

This argument is a refined version of the "diagonal argument" first used by Cantor. An interesting analysis of Cantor's proof, and of the misinterpretation of that proof over the years, is found in [19].)

.5 Prove that $\mathbf{R}\backslash\mathbf{Q}$ is order dense in $\mathbf{R}$.

The study of infinite series, a major part of analysis in the eighteenth and nineteenth centuries (see [27]), still provides interesting illustrations of the completeness of $\mathbf{R}$.

Let $(a_n)_{n=1}^{\infty}$ be a sequence of real numbers. The real number

$$s_k = \sum_{n=1}^{k} a_n$$

is called the kth *partial sum of the series* $\sum_{n=1}^{\infty} a_n$. Formally, we define the *series* $\sum_{n=1}^{\infty} a_n$ with nth *term* a_n to be the sequence $(s_1, s_2, ...)$ of its partial sums. The *sum* of that series is the limit s of the sequence (s_n), if that limit exists, in which case we say that the series is *convergent,* or that it *converges to* s, and we write

$$\sum_{n=1}^{\infty} a_n = s.$$

We use analogous notations and definitions for the series associated with a family $(a_n)_{n=\nu}^{\infty}$ of real numbers indexed by $\{n \in \mathbf{Z} : n \geq \nu\}$, where ν is an integer, and for the series $\sum_{n=-\infty}^{\infty} a_n$ associated with a family $(a_n)_{n\in\mathbf{Z}}$ indexed by $\mathbf{Z}$. We commonly write $\sum a_n$ for the series $\sum_{n=\nu}^{\infty} a_n$, when it is clear that the indexing of the terms of the series starts with ν.

The completeness of $\mathbf{R}$ is used in the justification of various tests for the convergence of infinite series. These tests are useful because they enable us to prove certain series convergent without finding explicit values for their sums. For example, a number of convergence tests easily show that the series $\sum_{n=1}^{\infty} n^{-2}$ is convergent; but it is considerably harder to show that the sum of this series is actually $\pi^2/6$ (Exercise (5.2.12: 7); see also [31]).

We begin with the *comparison test.*

(1.2.12) Proposition. *If $\sum_{n=1}^{\infty} b_n$ is a convergent series of nonnegative terms, and if $0 \leq a_n \leq b_n$ for each n, then $\sum_{n=1}^{\infty} a_n$ converges.*

Proof. Let b be the sum of the series $\sum_{n=1}^{\infty} b_n$. Then for each N we have

$$\sum_{n=1}^{N} a_n \leq \sum_{n=1}^{N+1} a_n \leq \sum_{n=1}^{N+1} b_n \leq b,$$

so the partial sums of $\sum_{n=1}^{\infty} a_n$ form an increasing majorised sequence. It follows from the monotone sequence principle that $\sum_{n=1}^{\infty} a_n$ converges. □

(1.2.13) Proposition. *If* (a_n) *is a decreasing sequence of positive numbers converging to* 0, *then the* alternating series $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ *converges* (Leibniz's alternating series test).

Proof. For each k let

$$s_k = \sum_{n=1}^{k}(-1)^{n+1}a_n.$$

Then

$$s_{2k+2} - s_{2k} = a_{2k+1} - a_{2k+2} \geq 0$$

and

$$a_1 - s_{2k} = (a_2 - a_3) + ... + (a_{2k-2} - a_{2k-1}) + a_{2k} \geq 0.$$

So the sequence $(s_{2k})_{k=1}^{\infty}$ is increasing and bounded above; whence, by the monotone sequence principle, it converges to its least upper bound s. Now,

$$|s - s_{2m+1}| = |s - s_{2m} - a_{2m+1}| \leq |s - s_{2m}| + a_{2m+1}.$$

Also, both $|s - s_{2m}|$ and a_{2m+1} converge to 0 as $m \to \infty$. It follows that if $\varepsilon > 0$, then $|s - s_{2m}| < \varepsilon$ and $|s - s_{2m+1}| < \varepsilon$ for all sufficiently large m. Hence $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges to s, by Exercise (1.2.3: 6). □

(1.2.14) Exercises

.1 Prove that if the series $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$. By considering $\sum_{n=1}^{\infty} 1/\sqrt{n}$, or otherwise, show that the converse is false.

.2 Prove the comparison test using the completeness of **R,** instead of the least–upper–bound principle.

.3 A series of nonnegative terms is said to *diverge* if the corresponding sequence (s_n) of partial sums diverges to infinity. Prove the *limit comparison test*: If (a_n) and (b_n) are sequences of positive numbers such that

$$\lim_{n\to\infty} \frac{a_n}{b_n} = l > 0,$$

then either $\sum a_n$ and $\sum b_n$ both converge or else they both diverge.

.4 Prove that if $|r| < 1$, then the *geometric series* $\sum_{n=0}^{\infty} r^n$ converges and has sum $1/(1-r)$. What happens to the series if $|r| > 1$?

.5 Let $b \geq 2$ be an integer, and $x \in [0, 1]$. Show that there exists a sequence (a_n) of integers such that

(i) $0 \le a_n < b$ for each n, and

(ii) $x = \sum_{n=1}^{\infty} a_n b^{-n}$.

Show that this sequence (a_n) is uniquely determined by x unless there exist $k, n \in \mathbf{N}$ such that $x = kb^{-n}$, in which case there are exactly two such sequences.

Conversely, show that if (a_n) is a sequence of integers satisfying (i), then $\sum_{n=1}^{\infty} a_n b^{-n}$ converges to a sum x in $[0, 1]$. (The series $\sum_{n=1}^{\infty} a_n b^{-n}$ is called the *b–ary expansion* of x, or the *expansion of x relative to the base b*. If $b = 2$, the series is the *binary expansion* of x, and if $b = 10$, it is the *decimal expansion*.)

.6 Prove that

(i) $\sum_{n=1}^{\infty} 1/n^p$ is divergent if $p \le 1$;

(ii) $\sum_{n=1}^{\infty} (-1)^n/n$ is convergent.

(For (i), first prove the divergence of $\sum_{n=1}^{\infty} 1/n$ by considering the partial sums $\sum_{n=1}^{2^N} 1/n$ for $N = 1, 2, \ldots$.)

.7 Prove that the series

$$\tfrac{1}{9} + \tfrac{1}{19} + \tfrac{1}{29} + \cdots + \tfrac{1}{89} + \tfrac{1}{90} + \tfrac{1}{91} + \cdots + \tfrac{1}{99} + \tfrac{1}{109} + \tfrac{1}{119} + \cdots,$$

where each term contains the digit 9, diverges; and that the series

$$1 + \tfrac{1}{2} + \tfrac{1}{3} + \cdots + \tfrac{1}{8} + \tfrac{1}{10} + \cdots + \tfrac{1}{18} + \tfrac{1}{20} + \cdots,$$

where no term contains the digit 9, converges. (Thus the divergent series $\sum_{n=1}^{\infty} 1/n$ can be turned into a convergent one by weeding out all the terms that contain the digit 9. For a discussion of this and related matters, see [3].)

.8 Prove that if $p \ge 2$, then the series $\sum_{n=1}^{\infty} 1/n^p$ is convergent.

.9 Prove *d'Alembert's ratio test*: let $\sum a_n$ be a series of positive terms such that

$$l = \lim_{n\to\infty} \frac{a_{n+1}}{a_n}$$

exists; then $\sum a_n$ converges if $l < 1$, and diverges if $l > 1$. (In the first case, choose $r \in (l, 1)$ and N such that $0 < a_{n+1} < ra_n$ for all $n \ge N$.)

Give examples where $l = 1$ and (i) $\sum a_n$ converges, (ii) $\sum a_n$ diverges.

.10 Prove that $\sum_{n=0}^{\infty} 1/n!$ converges and has sum < 3. Show also that

$$\sum_{k=0}^{n} \tfrac{1}{k!} - \tfrac{3}{2n} < \left(1 + \tfrac{1}{n}\right)^n < \sum_{k=0}^{n} \tfrac{1}{k!}$$

for all $n \geq 3$, and hence prove that $\lim_{n\to\infty}\left(1 + n^{-1}\right)^n = \sum_{n=0}^{\infty} 1/n!$.

.11 Prove *Cauchy's root test*: let $\sum a_n$ be a series of positive terms, and

$$l = \limsup \sqrt[n]{a_n};$$

then $\sum a_n$ converges if $l < 1$, and diverges if $l > 1$.

.12 Discuss the convergence of the series

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{2^3} + \frac{1}{3^3} + \cdots.$$

What does this series and Exercise (1.2.5: 6) tell you about the relative strengths of the ratio test and the root test?

.13 Let (a_n) be a decreasing sequence of nonnegative real numbers, and for each N let

$$s_N = \sum_{n=1}^{N} a_n, \ t_N = \sum_{n=1}^{N} 2^n a_{2^n}.$$

Show that

(i) if $m \leq 2^N$, then $s_m \leq t_N$, and

(ii) if $m \geq 2^N$, then $s_m \geq \frac{1}{2} t_N$.

Hence prove that $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} 2^n a_{2^n}$ converges.

.14 Use the preceding exercise to show that $\sum_{n=1}^{\infty} 1/n^p$ converges if and only if $p > 1$ (cf. Exercises (1.2.14: 6 and 8)).

.15 Let $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=0}^{\infty}$ be sequences of real numbers, and for each N write $S_N = \sum_{n=0}^{N} a_n$. Show that if $k > j$, then

$$\sum_{n=j}^{k} a_n b_n = \sum_{n=j}^{k-1} S_n (b_n - b_{n+1}) + S_k b_k - S_{j-1} b_j.$$

Now suppose that

(i) there exists $M > 0$ such that $|S_n| \leq M$ for all n,

(ii) $b_n \geq b_{n+1}$ for each n, and

(iii) $\lim_{n\to\infty} b_n = 0$.

Prove that if $k > j$, then $\left|\sum_{n=j}^{k} a_n b_n\right| < 2Mb_j$, and hence that $\sum_{n=0}^{\infty} a_n b_n$ converges. Use this result to give another proof of Leibniz's alternating series test.

A series $\sum a_n$ of real numbers is said to be *absolutely convergent* if $\sum |a_n|$ is convergent.

(1.2.15) Proposition. *An absolutely convergent series is convergent.*

Proof. Let $\sum a_n$ be absolutely convergent. Since the partial sums of $\sum |a_n|$ form a Cauchy sequence, for each $\varepsilon > 0$ there exists N such that

$$\left|\sum_{n=1}^{k} |a_n| - \sum_{n=1}^{j} |a_n|\right| = \sum_{n=j}^{k} |a_n| < \varepsilon$$

whenever $k > j \geq N$. For such j and k we have

$$\left|\sum_{n=1}^{k} a_n - \sum_{n=1}^{j} a_n\right| = \left|\sum_{n=j}^{k} a_n\right| \leq \sum_{n=j}^{k} |a_n| < \varepsilon.$$

Thus the partial sums of $\sum a_n$ form a Cauchy sequence; whence $\sum a_n$ is convergent, by the completeness of $\mathbf{R}$. □

The case $p = 1$ of Example (1.2.14: 6) shows that the converse of Proposition (1.2.15) is false.

By a *power series* we mean a series of the form $\sum_{n=0}^{\infty} a_n x^n$, where the coefficients $a_n \in \mathbf{R}$. Such a series always converges for $x = 0$, but it may converge for nonzero values of x. Its *radius of convergence* is defined to be

$$\sup\left\{r \geq 0 : \sum_{n=0}^{\infty} a_n x^n \text{ converges whenever } |x| \leq r\right\}$$

if this supremum exists, and ∞ otherwise; and its *interval of convergence* is the largest interval I such that the power series converges for all $x \in I$. It is an immediate consequence of Exercise (1.2.16: 10) that every power series has both a radius and an interval of convergence.

(1.2.16) Exercises

.1 Find an alternative proof of Proposition (1.2.15).

.2 Prove that the series $\sum_{n=1}^{\infty} nr^n$ converges absolutely if $-1 < r < 1$.

.3 Let $\sum a_n$, $\sum b_n$ be convergent series of nonnegative terms, with sums a, b, respectively, and let

$$u_n = a_1 b_n + a_2 b_{n-1} + \cdots + a_{n-1} b_2 + a_n b_1.$$

Prove that

$$\sum_{n=1}^{N} u_n \leq \left(\sum_{n=1}^{N} a_n\right)\left(\sum_{n=1}^{N} b_n\right) \leq \sum_{n=1}^{N} u_{2n}$$

and hence that $\sum u_n$ converges to the sum ab (*Cauchy's theorem on the multiplication of series*). Extend this result to the case where the terms a_n may not be nonnegative but $\sum a_n$ is absolutely convergent. (Writing

$$\begin{aligned} b &= \sum_{n=1}^{\infty} b_n, \\ \beta_k &= b - \sum_{n=1}^{k} b_n, \end{aligned}$$

show that

$$\sum_{n=1}^{N} u_n = b \sum_{n=1}^{N} a_n + \sum_{k=0}^{N} a_k \beta_{N-k},$$

and hence that $\sum_{k=0}^{N} a_k \beta_{N-k} \rightarrow 0$ as $N \rightarrow \infty$.)

.4 Show that $\sum_{n=1}^{\infty} (-1)^n / \sqrt{n+1}$ converges, but that the product (as in the preceding exercise) of this series with itself does not converge.

.5 Prove that the *exponential series*

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

converges absolutely for all $x \in \mathbf{R}$. Then prove that

$$\exp(x+y) = \exp(x)\exp(y).$$

.6 Prove that $\exp(x) = \mathrm{e}^x$, where $\mathrm{e} = \exp(1)$. (Use Exercise (1.1.4: 5).) Show that

$$0 < \mathrm{e} - \sum_{n=0}^{N} \frac{1}{n!} < \frac{3}{(N+1)!}$$

for each N, and hence calculate e with an error at most 10^{-6}.

.7 Prove that e is irrational. (Suppose that $\mathrm{e} = p/q$, where p and q are positive integers. Choose $N > \max\{q, 3\}$, show that $N! \sum_{n=N+1}^{\infty} 1/n!$ is an integer, and use the inequality from the preceding exercise to deduce a contradiction.)

.8 Show that

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$$

for each $x \in \mathbf{R}$. (First take $x > 0$. Expand $s_n = (1 + x/n)^n$ using the binomial theorem, and use the monotone sequence principle.)

.9 For each n define

$$\gamma_n = 1 + \tfrac{1}{2} + \tfrac{1}{3} + \cdots + \tfrac{1}{n} - \log n$$

Show that

$$e < \left(1+\tfrac{1}{n}\right)^{n+1} < e^{1+n^{-1}}$$

for each n, and hence that the sequence (γ_n) is decreasing and bounded below. It follows from the monotone sequence principle that *Euler's constant*

$$\gamma = \lim_{n\to\infty} \gamma_n$$

exists. Show that

$$\sum_{n=1}^{N} \frac{(-1)^{n+1}}{n} = \gamma_{2N} - \gamma_N + \log 2$$

and hence that

$$\sum_{n=1}^{N} \frac{(-1)^{n+1}}{n} = \log 2.$$

(cf. Exercise (1.2.14: 6).)

.10 Let $r > 0$. Prove that if $\sum_{n=0}^{\infty} a_n x^n$ converges for $x = r$, then it converges absolutely whenever $|x| < r$; and that if this power series diverges for $x = r$, then it diverges whenever $|x| > r$. (For the first part, show that there exists $M > 0$ such that $|a_n x^n| \leq M\,|x/r|^n$ for all n.)

.11 Find the radius of convergence and the interval of convergence for $\sum_{n=0}^{\infty} x^n$.

.12 Find the radius of convergence and the interval of convergence for $\sum_{n=0}^{\infty}(-1)^n x^n/n$.

.13 Suppose that $a_n \neq 0$ for all n, and that

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = l.$$

Show that if $l = 0$, then $\sum_{n=0}^{\infty} a_n x^n$ converges for all $x \in \mathbf{R}$; and that if $l \neq 0$, then the series has radius of convergence $1/l$.

.14 Let $(a_n)_{n=0}^{\infty}$ be a bounded sequence of real numbers, and let

$$l = \limsup \sqrt[n]{|a_n|}.$$

Show that if $l = 0$, then $\sum_{n=0}^{\infty} a_n x^n$ converges for all $x \in \mathbf{R}$; and that if $l \neq 0$, then the series has radius of convergence $1/l$.

.15 Prove that if $(\sqrt[n]{|a_n|})_{n=0}^{\infty}$ is an unbounded sequence, then the power series $\sum_{n=0}^{\infty} a_n x^n$ only converges for $x = 0$.

.16 Prove that the power series

$$\sum_{n=0}^{\infty} a_n x^n, \ \sum_{n=1}^{\infty} n a_n x^{n-1}, \ \text{and} \ \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$$

have the same radius of convergence. Need they have the same interval of convergence?

By a *rearrangement* of an infinite series $\sum_{n=1}^{\infty} a_n$ we mean a series of the form $\sum_{n=1}^{\infty} a_{f(n)}$ where f is a permutation of $\mathbf{N}^+$ (that is, a one–one mapping of $\mathbf{N}^+$ onto itself). A theorem first proved by Riemann shows that if $\sum a_n$ is a convergent, but not absolutely convergent, series of real numbers, then for each real number s there exists a rearrangement of $\sum a_n$ that converges to s. The second exercise in the next set leads you through a proof of this remarkable result.

(1.2.17) Exercises

.1 Prove that if $\sum a_n$ is absolutely convergent, with sum s, then any rearrangement of $\sum a_n$ converges to s. (Given a permutation f of $\mathbf{N}^+$ and a positive number ε, choose N such that $\sum_{n=N+1}^{\infty} |a_n| < \varepsilon$. Then choose $M \geq N$ such that $\{1, 2, \ldots, N\} \subset \{f(1), f(2), \ldots, f(M)\}$. Show that $\left|\sum_{n=1}^{m} a_{f(n)} - s\right| < 2\varepsilon$ for all $m \geq M$.)

.2 Let $\sum_{n=1}^{\infty} a_n$ be an infinite series of real numbers that converges but is not absolutely convergent. For each n define $a_n^+ = \max\{a_n, 0\}$, $a_n^- = \min\{a_n, 0\}$. Prove that the partial sums of the series $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ form increasing unbounded sequences. Now let s be any real number. Let n_0 and m_0 both equal 0, and let n_1 be the least positive integer such that

$$a_1^+ + a_2^+ + \cdots + a_{n_1}^+ > s.$$

Show how to construct strictly increasing sequences $(n_k)_{k=0}^{\infty}$ and $(m_k)_{k=0}^{\infty}$ of positive integers such that for each $N \geq 0$,

$$\sum_{k=0}^{N-1} \left(\left(a_{n_k+1}^+ + \cdots + a_{n_{k+1}}^+ \right) + \left(a_{m_k+1}^- + \cdots + a_{m_{k+1}}^- \right) \right) < s$$

and

$$\sum_{k=0}^{N-1}\left(\left(a_{n_k+1}^{+}+\cdots+a_{n_{k+1}}^{+}\right)+\left(a_{m_k+1}^{-}+\cdots+a_{m_{k+1}}^{-}\right)\right)$$
$$+\left(a_{n_N+1}^{+}+\cdots+a_{n_{N+1}}^{+}\right)\quad > s.$$

Hence obtain a rearrangement of $\sum_{n=1}^{\infty} a_n$ that converges to s.

.3 Let s be the sum of the series $\sum_{n=1}^{\infty}(-1)^{n-1}/n$. Show that the series

$$1-\tfrac{1}{2}-\tfrac{1}{4}+\tfrac{1}{3}-\tfrac{1}{6}-\tfrac{1}{8}+\tfrac{1}{5}-\tfrac{1}{10}-\tfrac{1}{12}+\cdots$$

converges to $\frac{1}{2}\log 2$.

1.3 Open and Closed Subsets of the Line

In this section we introduce the fundamental topological notions of "open set" and "closed set" in $\mathbf{R}$, notions that readily generalise in later, more abstract contexts.

A subset A of $\mathbf{R}$ is said to be *open* (in $\mathbf{R}$) if to each $x \in A$ there corresponds $r > 0$ such that the open interval $(x-r, x+r)$ is contained in A —or, equivalently, such that $y \in A$ whenever $|x-y| < r$.

(1.3.1) Exercises

.1 Prove that $\mathbf{R}$ itself, the empty set $\emptyset$, and all open intervals are open subsets of $\mathbf{R}$.

.2 Give an example of a sequence of open subsets of $\mathbf{R}$ whose intersection is not open.

The first result in this section describes the two fundamental properties of open sets.

(1.3.2) Proposition. *The union of any family of open sets is open. The intersection of any finite family of open sets is open.*

Proof. Let $(A_i)_{i\in I}$ be any family of open sets. If x belongs to the union U of this family, then $x \in A_i$ for some i. As A_i is open, there exists $r > 0$ such that

$$(x-r, x+r) \subset A_i \subset U.$$

Hence U is open.

Now let $A_1, \ldots, A_n$ be finitely many open sets, and consider any x in their intersection. For each i, since $x \in A_i$ and A_i is open, there exists $r_i > 0$ such that $y \in A_i$ whenever $|x - y| < r_i$. Let

$$r = \min\{r_1, \ldots, r_n\} > 0.$$

If $|x - y| < r$, then $y \in A_i$ for each i, so $y \in \bigcap_{i=1}^n A_i$. Hence $\bigcap_{i=1}^n A_i$ is open. □

In view of Exercise (1.3.1: 2), we cannot drop the word "finite" from the hypothesis of the second part of Proposition (1.3.2).

Our next aim is to characterise open sets in $\mathbf{R}$; to achieve this, we first characterise intervals.

A nonempty subset S of $\mathbf{R}$ is said to have the *intermediate value property* if $(a, b) \subset S$ whenever $a \in S$, $b \in S$, and $a < b$. Of course, as we show in Section 4, this notion is connected with the Intermediate Value Theorem of elementary calculus.

(1.3.3) Proposition. *A subset S of $\mathbf{R}$ has the intermediate value property if and only if it is an interval.*

Proof. It is clear that every interval in $\mathbf{R}$ has the intermediate value property. Conversely, suppose that $S \subset \mathbf{R}$ has that property. Assume, to begin with, that S is bounded, and let a be its infimum and b its supremum. Note that $x \notin S$ if either $x < a$ or $x > b$. If a and b both belong to S, then by the intermediate value property, so does every point of $[a, b]$; whence $S = [a, b]$. If $a \in S$ and $b \notin S$, consider any x such that $a \leq x < b$. By the definition of "supremum", there exists $s \in S$ such that $a \leq x < s$; the intermediate value property now ensures that $x \in S$; whence $S = [a, b)$. Similarly, if $a \notin S$ and $b \in S$, then $S = (a, b]$. The remaining cases are left as exercises. □

(1.3.4) Exercises

.1 Prove that a nonempty open subset of $\mathbf{R}$ with the intermediate value property is an open interval.

.2 Complete the proof of Proposition (1.3.3) in the remaining cases.

.3 Let I, J be open intervals with nonempty intersection. Prove that $I \cup J$ and $I \cap J$ are open intervals.

(1.3.5) Lemma. *A nonempty family of pairwise–disjoint open intervals of $\mathbf{R}$ is countable.*

Proof. Let $\mathcal{F}$ be a nonempty family of pairwise–disjoint open intervals in $\mathbf{R}$, and note that, by Exercise (1.1.1: 19), each of these intervals contains a rational number. The Axiom of Choice (see Appendix B) ensures that there is a function $f : \mathcal{F} \to \mathbf{Q}$ such that $f(I) \in I$ for each $I \in \mathcal{F}$. Since the sets in $\mathcal{F}$ are pairwise disjoint, f is one–one and so has an inverse function g mapping $f(\mathcal{F})$ onto $\mathcal{F}$. As $\mathbf{Q}$ is countable and $f(\mathcal{F}) \subset \mathbf{Q}$, there exists a mapping h of $\mathbf{N}^+$ onto $f(\mathcal{F})$; the composite function $g \circ h$ then maps $\mathbf{N}^+$ onto $\mathcal{F}$, which is therefore countable. □

(1.3.6) Proposition. *A nonempty subset of* $\mathbf{R}$ *is open if and only if it is the union of a sequence of pairwise–disjoint open intervals.*

Proof. It follows from Proposition (1.3.2) and Exercise (1.3.1: 1) that the union of *any* family of open intervals is an open set. Conversely, given a nonempty open subset S of $\mathbf{R}$, define a binary relation $\sim$ on S by setting $x \sim y$ if and only if there exists an open interval $I \subset S$ such that $x, y \in I$. Then $\sim$ is an equivalence relation: it is straightforward to prove the reflexivity and symmetry of $\sim$, and its transitivity follows from Exercise (1.3.4: 3). Clearly, the equivalence class $\dot{x}$ of x under $\sim$ is a union of open intervals and is therefore an open set. Consider points $y, z \in \dot{x}$ and a real number t with $y < t < z$. Choosing open intervals $I_y, I_z \subset S$ such that $x, y \in I_y$ and $x, z \in I_z$, we see from Exercise (1.3.4: 3) that $I_y \cup I_z$ is an open interval; so $t \in (y, z) \subset I_y \cup I_z$, and therefore either $x, t \in I_y \subset S$ or else $x, t \in I_z \subset S$. Hence $\dot{x}$ has the intermediate value property. It follows from Exercise (1.3.4: 1) that $\dot{x}$ is an open interval. Since any two distinct equivalence classes under $\sim$ are disjoint, we now see that

$$S = \bigcup_{x \in S} \dot{x}$$

is a union of pairwise–disjoint open intervals. Reference to Lemma (1.3.5) completes the proof. □

A real number x is an *interior point* of a set $S \subset \mathbf{R}$ if there exists $r > 0$ such that $(x - r,\, x + r) \subset S$. The set of all interior points of S is called the *interior* of S, and is written S°. By a *neighbourhood* of x we mean a set containing x in its interior.

(1.3.7) Exercises

.1 Let S be a nonempty open subset of $\mathbf{R}$, and for each $x \in S$ consider the sets

$$U_x = \{t \in \mathbf{R} : (x, t) \subset S\},$$
$$L_x = \{s \in \mathbf{R} : (s, x) \subset S\}.$$

Let $a = \inf L_x$, $b = \sup U_x$, and $I_x = (a, b)$, where $a = -\infty$ if L_x is not bounded below, and $b = \infty$ if U_x is not bounded above. Give another proof of Proposition (1.3.6) by showing that $(I_x)_{x \in S}$ is a family of disjoint open intervals whose union is S.

.2 Prove that the interior of an open, closed, or half open interval with endpoints a and b, where $a < b$, is the open interval (a, b).

.3 Show that $(S^\circ)^\circ = S^\circ$.

.4 Prove that S° is the largest open set contained in S—in other words, that

(i) S° is open and $S^\circ \subset S$;

(ii) if A is open and $A \subset S$, then $A \subset S^\circ$.

.5 Prove that S is open if and only if $S \subset S^\circ$.

.6 Prove that S° is the union of the open sets contained in S.

.7 Prove that

(i) if $S \subset T$, then $S^\circ \subset T^\circ$;

(ii) $(S \cap T)^\circ = S^\circ \cap T^\circ$.

.8 Prove that U is a neighbourhood of $x \in \mathbf{R}$ if and only if there is an open set A such that $x \in A \subset U$.

Let x be a real number, and S a subset of $\mathbf{R}$. We call x a *cluster point* of S if each neighbourhood of x has a nonempty intersection with S; or, equivalently, if for each $\varepsilon > 0$ there exists $y \in S$ such that $|x - y| < \varepsilon$. The *closure* of S (in $\mathbf{R}$) is the set of all cluster points of S, and is denoted by $\overline{S}$ or S^-. S is said to be *closed* if $S = \overline{S}$.

(1.3.8) Exercises

.1 Is $\mathbf{Q}$ closed in $\mathbf{R}$? Is it open in $\mathbf{R}$?

.2 Show that the closure of any interval with endpoints a and b, where $a < b$, is the closed interval $[a, b]$.

.3 Show that $\overline{(\overline{S})} = \overline{S}$.

.4 Prove that $\overline{S}$ is the smallest closed set containing S—in other words, that

(i) $\overline{S}$ is closed and $S \subset \overline{S}$;

(ii) if A is closed and $S \subset A$, then $\overline{S} \subset A$.

.5 Prove that S is closed if and only if $\overline{S} \subset S$.

.6 Prove that $\overline{S}$ is the intersection of the closed sets containing S.

.7 Prove the following.

(i) If $S \subset T$, then $\overline{S} \subset \overline{T}$;

(ii) $\overline{S \cup T} = \overline{S} \cup \overline{T}$.

.8 Prove that

(i) the complement of S° is the closure of $\mathbf{R}\backslash S$;

(ii) the complement of $\overline{S}$ is the interior of $\mathbf{R}\backslash S$.

.9 The *boundary,* or *frontier,* of a set $S \subset \mathbf{R}$ is the intersection of the closures of S and $\mathbf{R}\backslash S$. Describe the boundary of each of the following sets: $\mathbf{R}$, $\emptyset$, $(a, b]$ (where $a < b$), $\mathbf{Q}$.

.10 Prove that a belongs to the boundary of $S\backslash\{a\}$ if and only if $a \in \overline{S\backslash\{a\}}$.

.11 Let C be the *Cantor set*—that is, the subset of $[0, 1]$ consisting of all numbers that have a ternary (base 3) expansion $\sum_{n=1}^{\infty} a_n 3^{-n}$ with $a_n \in \{0, 2\}$ for each n. Prove that

(i) if a, b are two numbers in C that differ in their mth ternary places, then $|a - b| \geq 3^{-m}$;

(ii) C is a closed subset of $\mathbf{R}$;

(iii) C has an empty interior.

What is the boundary of C?

(1.3.9) Proposition. *S is closed if and only if $\mathbf{R}\backslash S$ is open.*

Proof. Suppose that S is closed, and consider any $x \in \mathbf{R}\backslash S$. Since $S = \overline{S}$, x is not a cluster point of S; so there exists a neighbourhood U of x that is disjoint from S. By Exercise (1.3.7: 8), there is an open set A such that $x \in A \subset U$. Then $A \cap S = \emptyset$, so $A \subset \mathbf{R}\backslash S$; whence, by Exercise (1.3.7: 8), $\mathbf{R}\backslash S$ is a neighbourhood of x, and therefore $x \in (\mathbf{R}\backslash S)^\circ$. Since x is any element of $\mathbf{R}\backslash S$, we conclude that $\mathbf{R}\backslash S$ is open.

Conversely, suppose that $\mathbf{R}\backslash S$ is open. Then by Exercise (1.3.7: 8), $\mathbf{R}\backslash S$ is a neighbourhood of each of its points. Since $\mathbf{R}\backslash S$ is disjoint from S, it follows that no point of $\mathbf{R}\backslash S$ is in the closure of S. Thus if $x \in \overline{S}$, then $x \notin \mathbf{R}\backslash S$ and so $x \in S$. Hence $\overline{S} \subset S$, and therefore, by Exercise (1.3.8: 5), S is closed. □

(1.3.10) Proposition. *The intersection of a family of closed sets is closed. The union of a finite family of closed sets is closed.*

Proof. Let $(C_i)_{i\in I}$ be any family of closed sets, and for each i let A_i be the complement of C_i. Then

$$\bigcap_{i\in I} C_i = \mathbf{R}\backslash(\bigcup_{i\in I} A_i).$$

Since, by Proposition (1.3.9), each A_i is open, Proposition (1.3.2) shows that $\bigcup_{i\in I} A_i$ is open; whence, again by Proposition (1.3.9), its complement is closed. This completes the first part of the proof; the second is left as an exercise. □

(1.3.11) Exercises

.1 Complete the proof of Proposition (1.3.10).

.2 Give an example of a sequence of closed sets whose union is not closed.

Which subsets of **R** are both open and closed? Before answering this question, we prove a simple lemma.

(1.3.12) Lemma. *If T is a nonempty open subset of* **R** *that is bounded above (respectively, below), then* $\sup T \notin T$ *(respectively,* $\inf T \notin T$*).*

Proof. Consider, for example, the case where T is bounded above. Suppose that $M = \sup T$ belongs to T. Since T is open, $(M - r, M + r) \subset T$ for some $r > 0$. Hence $M + \frac{1}{2}r \in T$, which is absurd as $M + \frac{1}{2}r > \sup T$. Hence, in fact, $M \notin T$. □

(1.3.13) Proposition. **R** *and* $\emptyset$ *are the only subsets of* **R** *that are both open and closed in* **R**.

Proof. Exercise (1.3.1: 1) and Proposition (1.3.9) show that **R** and $\emptyset$ are both open and closed in **R**. Let S be a nonempty set that is both open and closed, and note that, by Proposition (1.3.9), $\mathbf{R}\backslash S$ is also both open and closed. Suppose $\mathbf{R}\backslash S$ is nonempty. Choosing $a \in S$ and $b \in \mathbf{R}\backslash S$, we have either $a < b$ or $a > b$. Without loss of generality we take the former case, so that

$$T = \{x \in \mathbf{R}\backslash S : x > a\}$$

is nonempty and bounded below. By Proposition (1.3.2), T is also open, being the intersection of the open sets (a, ∞) and $\mathbf{R}\backslash S$. Let $m = \inf T$. Since S is open, there exists $r > 0$ such that $(a - r, a + r) \subset S$; whence $m \geq a + r > a$. Since, by Lemma (1.3.12), $m \notin T$, it follows that $m \notin \mathbf{R}\backslash S$ and therefore that $m \in S$. But S is open, so there exists $\varepsilon > 0$ such that $(m - \varepsilon, m + \varepsilon) \subset S$; this is impossible, since the definition of "infimum" ensures that there exists $t \in \mathbf{R}\backslash S$ such that $t < m + \varepsilon$. This contradiction shows that $\mathbf{R}\backslash S$ is empty; whence $S = \mathbf{R}$. □

1.4 Limits and Continuity

Let I be an interval in **R,** a a point of the closure of I, and f a real–valued function whose domain includes I but not necessarily a. A real number l is called the *limit of* $f(x)$ *as* x *tends to* a *in* I, or the *limit of* f *at* a (*relative to* I), if to each $\varepsilon > 0$ there corresponds $\delta > 0$ such that $|f(x) - l| < \varepsilon$ whenever $x \in I$ and $0 < |x - a| < \delta$. We then write

$$f(x) \to l \text{ as } x \to a, \; x \in I$$

or

$$\lim_{x \to a,\, x \in I} f(x) = l$$

and we say that $f(x)$ *tends to* l *as* x *tends to* a *through values in* I.

The following are the most important cases of this definition.

- $a \in I^\circ$: in this case we use the simpler notations

  $$f(x) \to l \text{ as } x \to a$$

  and

  $$\lim_{x \to a} f(x) = l.$$

- $I = (c, a)$ for some $c < a$ (where c could be $-\infty$): in this case we call l the *left–hand limit* of f as x tends to a; we say that $f(x)$ *tends to* l *as* x *tends to* a *from the left* (or *from below*); and we use the notations

  $$f(x) \to l \text{ as } x \to a^-$$

  and

  $$f(a^-) = \lim_{x \to a^-} f(x) = l.$$

- $I = (a, b)$ for some $b > a$ (where b could be ∞): in this case we call l the *right–hand limit* of f as x tends to a; we say that $f(x)$ *tends to* l *as* x *tends to* a *from the right* (or *from above*); and we use the notations

  $$f(x) \to l \text{ as } x \to a^+$$

  and

  $$f(a^+) = \lim_{x \to a^+} f(x) = l.$$

We stress that although, in our definition of "limit", $f(x)$ is defined for all x in I that are distinct from but sufficiently close to a, $f(a)$ need not be defined. For example, in elementary calculus courses we learn that

$$\lim_{x \to 0} \frac{\sin x}{x} = 1$$

even though $(\sin x)/x$ is not defined at $x = 0$.

(1.4.1) Proposition. *If* $\lim_{x\to a,\, x\in I} f(x) = l$ *and* $\lim_{x\to a,\, x\in I} f(x) = l'$, *then* $l = l'$.

Proof. Given $\varepsilon > 0$, choose $\delta_f,\ \delta_g > 0$ such that

— if $x \in I$ and $0 < |x - a| < \delta_f$, then $|f(x) - l| < \varepsilon/2$, and

— if $x \in I$ and $0 < |x - a| < \delta_g$, then $|f(x) - l'| < \varepsilon/2$.

Setting $\delta = \min\{\delta_f, \delta_g\}$, consider any $x \in I$ such that $0 < |x - a| < \delta$. We have

$$|l - l'| \le |f(x) - l| + |f(x) - l'| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows from Exercise (1.1.2: 4) that $l = l'$. □

(1.4.2) Proposition. *If* $\lim_{x\to a,\, x\in I} f(x) = l$ *and* $\lim_{x\to a,\, x\in I} g(x) = m$, *then as* $x \to a$ *through values in* I,

$$\begin{aligned} f(x) + g(x) &\to l + m, \\ f(x) - g(x) &\to l - m, \\ f(x)g(x) &\to lm, \\ \max\{f(x), g(x)\} &\to \max\{l, m\}, \\ \min\{f(x), g(x)\} &\to \min\{l, m\}, \\ |f(x)| &\to |l|. \end{aligned}$$

If also $m \neq 0$, *then*

$$\lim_{x\to a,\, x\in I} \frac{f(x)}{g(x)} = \frac{l}{m}.$$

Proof. See Exercise (1.4.3: 3). □

(1.4.3) Exercises

.1 Define precisely what it means to say that $f(x)$ *does not converge to any limit as* x *tends to* a *through values in* I. In other words, give the formal negation of the definition of "convergent".

.2 Let the function f be defined in an interval whose interior contains a. Prove that $\lim_{x\to a} f(x) = l$ if and only if $\lim_{x\to a^+} f(x)$ and $\lim_{x\to a^-} f(x)$ exist and equal l.

.3 Prove Proposition (1.4.2).

.4 Use the definition of "limit" to prove that $\lim_{x\to a,\, x\in \mathbf{R}} p(x) = p(a)$ for any polynomial function p and any $a \in \mathbf{R}$.

.5 Let p, q be polynomial functions, and a a real number such that $q(a) \neq 0$. Prove that

$$\lim_{x \to a,\, x \in \mathbf{R}} \frac{p(x)}{q(x)} = \frac{p(a)}{q(a)}.$$

.6 Prove that $\lim_{x \to a,\, x \in I} f(x) = l$ if and only if for each sequence (a_n) of elements of I that converges to a, the sequence $(f(a_n))$ converges to l. (For "only if", use a proof by contradiction.)

.7 Suppose that $\lim_{x \to a,\, x \in I} f(x) > r$. Prove that there exists $\delta > 0$ such that if $x \in I$ and $0 < |x - a| < \delta$, then $f(x) > r$.

.8 Let f be a real–valued function whose domain includes an interval of the form (s, ∞), and let $l \in \mathbf{R}$. We say that $f(x)$ *tends to* l *as* x *tends to* ∞ if to each $\varepsilon > 0$ there corresponds $K > 0$ such that $|f(x) - l| < \varepsilon$ whenever $x > K$; we then write

$$f(x) \to l \text{ as } x \to \infty$$

or

$$\lim_{x \to \infty} f(x) = l.$$

Convince yourself that analogues of Propositions (1.4.1) and (1.4.2) hold for limits as x tends to ∞.

Define the notion $f(x)$ *tends to* l *as* x *tends to* $-\infty$, written

$$f(x) \to l \text{ as } x \to -\infty$$

or

$$\lim_{x \to -\infty} f(x) = l,$$

and convince yourself that analogues of Propositions (1.4.1) and (1.4.2) hold for this notion also.

.9 Formulate definitions of the following notions, where I is an interval.

(i) $f(x) \to \infty$ as $x \to a$ through values in I.

(ii) $f(x) \to -\infty$ as $x \to a$ through values in I.

(iii) $f(x) \to \infty$ as $x \to \infty$.

(iv) $f(x) \to \infty$ as $x \to -\infty$.

(v) $f(x) \to -\infty$ as $x \to \infty$.

(vi) $f(x) \to -\infty$ as $x \to -\infty$.

.10 Prove that if $a > 1$, then $a^x \to 0$ as $x \to -\infty$, and $a^x \to \infty$ as $x \to \infty$. What happens to a^x as $x \to \pm\infty$ when $0 < a < 1$? (Note Exercise (1.2.3: 3).)

.11 Prove that if $a > 1$, then $\log_a x \to 0$ as $x \to -\infty$, and $\log_a x \to \infty$ as $x \to \infty$. What happens to $\log_a x$ as $x \to \pm\infty$ if $0 < a < 1$?

.12 Let f be a real–valued function, and, where appropriate, define

$$\begin{aligned} \lim_{x\to\xi}\sup f(x) &= \inf_{r>0}\sup\{f(x) : 0 < |x-\xi| < r\}, \\ \lim_{x\to\xi+}\sup f(x) &= \inf_{r>0}\sup\{f(x) : 0 < x-\xi < r\}, \\ \lim_{x\to\xi-}\sup f(x) &= \inf_{r>0}\sup\{f(x) : 0 < \xi-x < r\}. \end{aligned}$$

For example, in order that $\lim_{x\to\xi-}\sup f(x)$ be defined, it is necessary that f be defined and bounded[4] on some interval of the form $(\xi-r,\xi)$, where $r > 0$. Prove the following.

(a) $\lim_{x\to\xi}\sup f(x) \le M$ if and only if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $f(x) < M+\varepsilon$ whenever $0 < |x-\xi| < \delta$.

(b) $\lim_{x\to\xi}\sup f(x) \ge M$ if and only if for each pair of positive numbers ε, δ there exists x such that $0 < |x-\xi| < \delta$ and $f(x) > M-\varepsilon$.

Formulate appropriate definitions of the quantities $\lim_{x\to\xi}\inf f(x)$, $\lim_{x\to\xi+}\inf f(x)$, and $\lim_{x\to\xi-}\inf f(x)$. Prove that

(c) $\lim_{x\to\xi}\inf f(x) \le \lim_{x\to\xi}\sup f(x)$, and these two numbers are equal if and only if $l = \lim_{x\to\xi} f(x)$ exists, in which case the numbers equal l.

An important special case of the notion of a limit occurs when the function f is defined at the point a that we are approaching.

A function f defined in some neighbourhood of a is said to be *continuous* at a if $f(x) \to f(a)$ as $x \to a$; in other words, if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ whenever $|x-a| < \delta$. We also say that f is

- *continuous on the left at a* if f is defined on the interval $(a-r,\, a]$ for some $r > 0$ and
$$f(a^-) = \lim_{x\to a^-} f(x) = f(a);$$
- *continuous on the right at a* if f is defined on the interval $[a,\, a+r)$ for some $r > 0$ and
$$f(a^+) = \lim_{x\to a^+} f(x) = f(a);$$

[4] By allowing the lim sup quantities to take the values $\pm\infty$, in a sense that is made precise in Section 3.1, we can remove the restriction that f be bounded near ξ.

- *continuous on the interval* I if $\lim_{x \to t,\, x \in I} f(x) = f(t)$ for each $t \in I$.

Note that the last definition takes care of the one–sided continuity of f at those endpoints of I, if any, that belong to I.

If f is defined in a neighbourhood of a but is not continuous at a, we say that f has a *discontinuity,* or is *discontinuous,* at a.

(1.4.4) Proposition. *Let the real–valued functions f and g be continuous at a. Then $f+g$, $f-g$, fg, $\max\{f,g\}$, $\min\{f,g\}$, and $|f|$ are continuous at a. If also $g(x) \neq 0$ for all x in some neighbourhood of a, then f/g is continuous at a.*

Proof. This is a simple consequence of Proposition (1.4.2). □

(1.4.5) Exercises

.1 Let f be defined on a neighbourhood of a. Prove that f is continuous at a if and only if it is continuous on both the left and the right at a.

.2 Give the details of the proof of Proposition (1.4.4). Extend this result to deal with continuity on an interval I.

.3 Prove that a polynomial function is continuous on $\mathbf{R}$.

.4 Let p, q be polynomial functions, and a a real number such that $q(a) \neq 0$. Prove that the rational function p/q is continuous at a.

.5 Let f be continuous at a, and let g be continuous at $f(a)$. Prove that the composite function $g \circ f$ is continuous at a.

.6 Prove that f is continuous at the point $a \in \mathbf{R}$ if and only if f is *sequentially continuous* at a, in the sense that $f(a_n) \to f(a)$ whenever (a_n) is a sequence of points of the domain of f that converges to a.

.7 Let f be defined in an interval $(a-r,\, a+r)$ where $r > 0$. The *oscillation* of f at a is

$$\omega(f,a) = \lim_{\delta \to 0} \sup\{f(x) - f(y) : x, y \in (a-\delta,\, a+\delta)\}.$$

Prove that f is continuous at a if and only if $\omega(f,a) = 0$.

.8 Let f be an increasing function on $[a,b]$. Prove that $f(\xi^-)$ exists for each $\xi \in (a,b]$, and that $f(\xi^+)$ exists for each $\xi \in [a,b)$. By considering the sets

$$\left\{x \in (a,b) : \left|f(x^+) - f(x^-)\right| > \tfrac{1}{n}\right\},$$

with n a positive integer, prove that the set of points of $[a,b]$ at which f has a discontinuity is either empty or countable.

.9 Let $q_0, q_1, \ldots$ be a one–one enumeration of $\mathbf{Q} \cap [0,1]$, and for each $x \in [0,1]$ define

$$T(x) = \{n \in \mathbf{N} : q_n \le x\}.$$

Define a mapping $f : [0,1] \to \mathbf{R}$ by

$$f(x) = \begin{cases} 0 & \text{if } x = 0 \\ \sum_{n \in T(x)} 2^{-n} & \text{if } 0 < x \le 1. \end{cases}$$

Prove that

(i) f is strictly increasing,

(ii) f is continuous at each irrational point of $[0,1]$, and

(iii) f is discontinuous at each rational point of $[0,1]$.

.10 Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions on an interval I, and suppose that there exists a convergent series $\sum_{n=1}^{\infty} M_n$ of nonnegative terms such that $|f_n(x)| \le M_n$ for each $x \in I$ and each n. Prove that for each $\varepsilon > 0$ there exists N such that $0 \le \sum_{n=j+1}^{k} |f_n(x)| < \varepsilon$ whenever $k > j \ge N$ and $x \in I$. Hence prove that $f(x) = \sum_{n=1}^{\infty} f_n(x)$ defines a function on I (*Weierstrass's M–test*). Prove also that if each f_n is continuous on I, then so is f.

.11 Give two proofs that exp is a continuous function on $\mathbf{R}$.

.12 Prove that if $a > 0$, then the function $x \mapsto a^x$ is continuous on $\mathbf{R}$. (Note that $a^x = \exp(x \log a)$.)

.13 Prove that if $a > 0$, then the function $x \mapsto \log_a x$ is continuous on $\mathbf{R}$. (First take the case $a > 1$. Given $x > 0$ and $\varepsilon > 0$, choose a positive integer $n > 1/\varepsilon$, and then $\delta \in (0,x)$ such that $(x+\delta)/x < a^{1/n}$ and $(x-\delta)/x > a^{-1/n}$.)

.14 Prove that the functions sin and cos, defined by

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!},$$

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

are (well defined and) continuous on $\mathbf{R}$.

.15 Let I be the interval of convergence of the power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$. Prove that f is continuous on I.

.16 Prove that if $\sum_{n=0}^{\infty} a_n$ is a convergent series, then

(i) $\sum_{n=0}^{\infty} a_n x^n$ converges for all $x \in (-1, 1)$, and

(ii) for each $\varepsilon > 0$ there exists $\delta \in (0, 1)$ such that if $1 - \sigma < x < 1$, then

$$\left| \sum_{n=0}^{\infty} a_n - \sum_{n=0}^{\infty} a_n x^n \right| < \varepsilon.$$

Thus $\lim_{x \to 1^-} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n$ (*Abel's Limit Theorem*). (For (ii), note that

$$\left| \sum_{n=0}^{\infty} a_n - \sum_{n=0}^{\infty} a_n x^n \right| \leq \left| \sum_{n=0}^{N} a_n (1 - x^n) \right| + \left| \sum_{n=N+1}^{\infty} a_n \right| + \left| \sum_{n=N+1}^{\infty} a_n x^n \right|$$

for each N. Use Exercise (1.2.14: 15) to handle the last term on the right.)

.17 Let $\sum_{n=0}^{\infty} a_n, \sum_{n=0}^{\infty} b_n$ be convergent series with sums a, b, respectively, and let

$$u_n = a_1 b_n + a_2 b_{n-1} + \cdots + a_{n-1} b_2 + a_n b_1.$$

Prove that if $\sum_{n=0}^{\infty} u_n$ converges, then its sum is ab. (For $-1 < x \leq 1$ set $f(x) = \sum_{n=0}^{\infty} a_n x^n$ and $g(x) = \sum_{n=0}^{\infty} b_n x^n$. Then use Exercises (1.2.16: 10), (1.2.16: 3), and (1.4.5: 16).)

This is the full form of *Cauchy's theorem on the multiplication of series*, and should be compared with Exercise (1.2.16: 3).

Deeper results about continuity—indeed, many results in real–variable theory—depend on two fundamental properties of the real line, described in our next two theorems.

By a *cover* of a subset S of $\mathbf{R}$ we mean a family $\mathcal{U}$ of subsets of $\mathbf{R}$ such that $S \subset \bigcup \mathcal{U}$; we then say that S is *covered* by $\mathcal{U}$ and that $\mathcal{U}$ *covers* S. If also each $U \in \mathcal{U}$ is an open subset of $\mathbf{R}$, we refer to $\mathcal{U}$ as an *open cover* of S (in $\mathbf{R}$). By a *subcover* of a cover $\mathcal{U}$ of S we mean a family $\mathcal{F} \subset \mathcal{U}$ that covers S; if also $\mathcal{F}$ is a finite family, then it is called a *finite subcover* of $\mathcal{U}$.

Although there exist shorter proofs of the next theorem (see the next set of exercises), the one we present is adapted to prove a more general result (Theorem (3.3.9)) in Chapter 3.

(1.4.6) The Heine–Borel–Lebesgue Theorem. *Every open cover of a compact interval I in $\mathbf{R}$ contains a finite subcover of I.*

Proof. Suppose there exists an open cover $\mathcal{U}$ of I that contains no finite subcover of I. Either the closed right half of I or the closed left half (or both) cannot be covered by a finite subfamily of $\mathcal{U}$: otherwise each half, and therefore I itself, would be covered by a finite subfamily. Let I_1 be a closed half of I that is not covered by a finite subfamily of $\mathcal{U}$. In turn, at least one closed half, say I_2, of I_1 cannot be covered by a finite subfamily of $\mathcal{U}$. Carrying on in this way, we construct a nested sequence $I \supset I_1 \supset I_2 \supset \cdots$ of closed subintervals of I such that for each n,

(a) $|I_n| = 2^{-n} |I|$ and

(b) no finite subfamily of $\mathcal{U}$ covers I_n.

By the nested intervals principle (1.2.6), there exists a point $\xi \in \bigcap_{n=1}^{\infty} I_n$. Clearly $\xi \in I$, so there exists $U \in \mathcal{U}$ such that $\xi \in U$. Since U is open, there exists $r > 0$ such that if $|x - \xi| < r$, then $x \in U$. Using (a), we can find N such that if $x \in I_N$, then $|x - \xi| < r$ and therefore $x \in U$; thus $I_N \subset U$. This contradicts (b). □

A real number a is called a *limit point* of a subset S of $\mathbf{R}$ if each neighbourhood of a intersects $S \backslash \{a\}$; or, equivalently, if for each $\varepsilon > 0$ there exists $x \in S$ with $0 < |x - a| < \varepsilon$. By a *limit point of a sequence* (a_n) we mean a limit point of the set $\{a_1, a_2, \ldots\}$ of terms of the sequence.

A nonempty subset A of $\mathbf{R}$ is said to have the *Bolzano–Weierstrass property* if each infinite subset S of A has a limit point belonging to A.

(1.4.7) The Bolzano–Weierstrass Theorem. *Every compact interval in* $\mathbf{R}$ *has the Bolzano–Weierstrass property.*

Proof. Let I be a compact interval, and S an infinite subset of I. By Corollary (1.2.8), any infinite sequence of distinct points of S contains a convergent subsequence; the limit of that subsequence is a limit point of S in the closed set I. □

(1.4.8) Exercises

.1 Let X be a subset of $\mathbf{R}$ with the Bolzano–Weierstrass property, and let (x_n) be a sequence of points in X. Show that there exists a subsequence of (x_n) that converges to a limit in X. (Note Lemma (1.2.7).)

.2 Fill in the details of the following alternative proof of the Heine–Borel–Lebesgue Theorem. Let $\mathcal{U}$ be an open cover of the compact interval $I = [a, b]$, and define

$$A = \{x \in I : [a, x] \text{ is covered by finitely many elements of } \mathcal{U}\}.$$

Then A is nonempty (it contains a) and is bounded above; let $\xi = \sup A$. Suppose that $\xi \neq b$, and derive a contradiction.

.3 Fill in the details of the following alternative proof of the Bolzano–Weierstrass Theorem. Suppose the theorem is false; so there exist a compact interval I and an infinite subset S of I such that no limit point of S belongs to I. Construct a nested sequence $I \supset I_1 \supset I_2 \supset \cdots$ of closed subintervals of I such that for each n,

(a) $|I_n| = 2^{-n} |I|$,

(b) $S \cap I_n$ is an infinite set, and

(c) $S \cap I_n$ has no limit points in I_n.

Let $\xi \in \bigcap_{n=1}^{\infty} I_n$, and show that ξ is a limit point of S. (This is one of the commonest proofs of the Bolzano–Weierstrass Theorem in textbooks.)

.4 Here is a sketch of yet another proof of the Bolzano–Weierstrass Theorem for you to complete. Let I be a compact interval, and S an infinite subset of I; then the supremum of the set

$$A = \{x \in I : S \cap (-\infty, x) \text{ is finite or empty}\}$$

is a limit point of S in I.

.5 Let S be a subset of $\mathbf{R}$ with the Bolzano–Weierstrass property. Prove that S is closed and bounded. (For boundedness, use a proof by contradiction.)

.6 Show that the Bolzano–Weierstrass Theorem can be proved as a consequence of the Heine–Borel–Lebesgue Theorem. (Let I be a compact interval in $\mathbf{R}$, assume the Heine–Borel–Lebesgue Theorem (1.4.6), and suppose that there exists an infinite subset S of I that has no limit point in I. First show that for each $s \in \overline{S}$ there exists $r_s > 0$ such that $S \cap (s - r_s,\, s + r_s) = \{s\}$.)

.7 Let f be a real–valued function defined on an interval I. We say that f is *uniformly continuous* on I if to each $\varepsilon > 0$ there corresponds $\delta > 0$ such that $|f(x) - f(x')| < \varepsilon$ whenever $x, x' \in I$ and $|x - x'| < \delta$. Show that a uniformly continuous function is continuous. Give an example of I and f such that f is continuous, but not uniformly continuous, on I.

.8 Use the Heine–Borel–Lebesgue Theorem to prove the *Uniform Continuity Theorem*: a continuous real–valued function f on a compact interval $I \subset \mathbf{R}$ is uniformly continuous. (For each $\varepsilon > 0$ and each $x \in I$, choose $\delta_x > 0$ such that if $x' \in I$ and $|x - x'| < 2\delta_x$, then $|f(x) - f(x')| < \varepsilon/2$. The intervals $(x - \delta_x, x + \delta_x)$ form an open cover of I.)

.9 Prove the Uniform Continuity Theorem (see the previous exercise) using the Bolzano–Weierstrass Theorem. (If $f : I \to \mathbf{R}$ is not uniformly continuous, then there exists $\alpha > 0$ with the following property: for each $n \in \mathbf{N}^+$ there exist $x_n, y_n \in I$ such that $|x_n - y_n| < 1/n$ and $|f(x_n) - f(y_n)| \geq \alpha$.)

The proof of the following result about boundedness of real–valued functions illustrates well the application of the Heine–Borel–Lebesgue Theorem.

(1.4.9) Theorem. *A continuous real–valued function f on a compact interval I is bounded; moreover, f* attains its bounds *in the sense that there exist points ξ, η of I such that $f(\xi) = \inf f$ and $f(\eta) = \sup f$.*

Proof. For each $x \in I$ choose $\delta_x > 0$ such that if $x' \in I$ and $|x - x'| < \delta_x$, then $|f(x) - f(x')| < 1$. The intervals $(x - \delta_x, x + \delta_x)$, where $x \in I$, form an open cover of I. By Theorem (1.4.6), there exist finitely many points $x_1, \ldots, x_N$ of I such that

$$I \subset \bigcup_{k=1}^{N} (x_k - \delta_{x_k}, x_k + \delta_{x_k}).$$

Let

$$c = 1 + \max\{|f(x_1)|, \ldots, |f(x_N)|\},$$

and consider any point $x \in I$. Choosing k such that $x \in (x_k - \delta_{x_k}, x_k + \delta_{x_k})$, we have

$$\begin{aligned} |f(x)| &\leq |f(x) - f(x_k)| + |f(x_k)| \\ &< 1 + |f(x_k)| \\ &\leq c, \end{aligned}$$

so f is bounded on I.

Now write

$$m = \inf f, \quad M = \sup f.$$

Suppose that $f(x) \neq M$, and therefore $f(x) < M$, for all $x \in I$. Then $x \mapsto 1/(M - f(x))$ is a continuous mapping of I into $\mathbf{R}^+$, by Proposition (1.4.4), and so, by the first part of this proof, has a supremum $G > 0$. For each $x \in I$ we then have $M - f(x) \geq 1/G$ and therefore $f(x) \leq M - 1/G$. This contradicts our choice of M as the supremum of f. □

(1.4.10) Exercises

.1 Prove both parts of Theorem (1.4.9) using the Bolzano–Weierstrass Theorem and contradiction arguments.

.2 Let f be a continuous function on $\mathbf{R}$ such that $f(x) \to \infty$ as $x \to \pm\infty$. Prove that there exists $\xi \in \mathbf{R}$ such that $f(x) \geq f(\xi)$ for all $x \in \mathbf{R}$.

.3 Let f be a continuous function on $\mathbf{R}$ such that $f(x) \to 0$ as $x \to \pm\infty$. Prove that f is both bounded and uniformly continuous.

(1.4.11) The Intermediate Value Theorem. *If f is a continuous real–valued function on an interval I, then $f(I)$ has the intermediate value property* (page 36).

Proof. Let a, b be points of I, and y a real number such that $f(a) < y < f(b)$; without loss of generality assume that $a < b$. Then

$$S = \{x \in [a,b] : f(x) < y\}$$

is nonempty (it contains a) and bounded above by b, so $\xi = \sup S$ exists. Note that $\xi < b$ and that $(\xi, b] \subset I$. We show that $f(\xi) = y$. To this end, suppose first that $f(\xi) < y$. Then, by Exercise (1.4.3: 7), there exists $\delta \in (0,\ b-\xi)$ such that if $x \in I$ and $|x - \xi| < \delta$, then $f(x) < y$; in particular, $f(x) < y$ for all $x \in (\xi,\ \xi + \delta)$, which contradicts the definition of ξ as the supremum of S. Thus $f(\xi) \geq y$.

Now suppose that $f(\xi) > y$; then $\xi > a$. By another application of Exercise (1.4.3: 7), there exists $\delta' \in (0, \xi - a)$ such that if $\xi - \delta' < x < \xi$, then $f(x) > y$. This is impossible, since, by the definition of "supremum", there exist points x of (a, ξ) arbitrarily close to ξ with $f(x) < y$. Hence $f(\xi) \leq y$, and therefore $f(\xi) = y$. □

(1.4.12) Corollary. *Let f be a continuous real–valued function on a compact interval I, and let $m = \inf f$, $M = \sup f$. Then $f(I) = [m, M]$.*

Proof. Use Theorems (1.4.9) and (1.4.11). □

(1.4.13) Exercises

.1 Fill in the details of the following common proof of the Intermediate Value Theorem. Let $f(a) < y < f(b)$, write $a_0 = a$, $b_0 = b$, $c_0 = \frac{1}{2}(a_0 + b_0)$, and assume without loss of generality that $a < b$. If $f(c_0) = y$, there is nothing to prove and we stop our construction. Otherwise, by repeated interval–halving, we construct points $a_0, b_0, c_0, a_1, b_1, c_1, \ldots$ such that

— either $f(c_n) = 0$ for some n and the construction stops,

— or else the construction proceeds *ad infinitum*, $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$, $f(a_n) < y$, $f(b_n) > y$, $c_n = \frac{1}{2}(a_n + b_n)$, and

$$0 < b_n - a_n = \left(\tfrac{1}{2}\right)^n (b - a).$$

Choosing $x \in \bigcap_{n=1}^{\infty}[a_n, b_n]$, we now show that $f(x) = 0$.

.2 Use the Intermediate Value Theorem to prove that if $b > 0$ and n is an odd positive integer, then b has an nth root—that is, there exists $r \in \mathbf{R}$ such that $r^n = b$. (Of course, this result also follows from our definition of a^x in Section 1; but it is instructive to see how it can be derived by other means, such as the Intermediate Value Theorem.)

.3 Show that any polynomial equation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

of odd degree n, with coefficients $a_k \in \mathbf{R}$, has at least one real solution.

.4 What can you say about a function f that is continuous on $[0, 1]$ and assumes only rational values?

.5 Let f, g be continuous functions on $[0, 1]$ such that $f(x) \in [0, 1]$ for all x, $g(0) = 0$, and $g(1) = 1$. Show that $f(x) = g(x)$ for some $x \in [0, 1]$.

.6 Prove that there is no continuous function $f : \mathbf{R} \to \mathbf{R}$ that assumes each real value exactly twice.

.7 Let f be continuous and one–one on an interval I; then $f(I)$ is an interval, by Corollary (1.4.12). Prove that

(i) either f is strictly increasing on I or else f is strictly decreasing on I;

(ii) if $a \in I^\circ$, then $f(a) \in f(I)^\circ$;

(iii) f^{-1} is continuous on $f(I)$.

(For (iii), show that f is sequentially continuous at each point of $f(I)$. You will need Corollary (1.2.8), Exercise (1.4.5: 6), and Exercise (1.2.9: 4).)

Although the Intermediate Value Theorem has many applications, especially in the solution of equations, none of its proofs provides an algorithm for constructing the point x with $f(x) = y$. This claim may come as a surprise: for is not the interval–halving proof in Exercise (1.4.13: 1) algorithmic? Alas, it is not: for, as any good computer scientist knows, there is no algorithm that enables us to decide, for given real numbers y and z, whether $y = z$ or $y \neq z$. (For further discussion of these matters, see the Prolog of [5], and pages 65–66 of [8].)

1.5 Calculus

In this section we cover the fundamentals of the differential and integral calculus of functions of one real variable. We do so rapidly, leaving many details to the exercises, on the assumption that you will have seen much of the material in elementary calculus courses.

Let I be an interval in **R,** x_0 a point of I, and f a real–valued function whose domain includes I. We say that f is

- *differentiable on the left at* x_0 if its *left-hand derivative at* x_0,

$$f'(x_0^-) = \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0-} \frac{f(x_0 + h) - f(x_0)}{h},$$

  exists;

- *differentiable on the right at* x_0 if its *right-hand derivative at* x_0,

$$f'(x_0^+) = \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0+} \frac{f(x_0 + h) - f(x_0)}{h},$$

  exists;

- *differentiable at* x_0 if x_0 is an interior point of I and the *derivative of* f *at* x_0,

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

  exists.

It follows from Exercise (1.4.3: 2) that f is differentiable at an interior point x_0 of its domain if and only if $f'(x_0^-)$ and $f'(x_0^+)$ exist and are equal, in which case their common value is $f'(x_0)$.

We say that f is *differentiable on the interval* I if it is

— differentiable at each interior point of I,

— differentiable on the right at the left endpoint of I if that point belongs to I, and

— differentiable on the left at the right endpoint of I if that point belongs to I.

Higher–order derivatives of f are defined inductively, as follows.

$$\begin{aligned}
f^{(0)} &= f, \\
f^{(1)} &= f', \\
f^{(2)} &= f'' = (f')', \\
f^{(3)} &= f''' = (f'')', \\
f^{(n+1)} &= (f^{(n)})' \quad (n \geq 3).
\end{aligned}$$

If the nth *derivative* $f^{(n)}(x)$ exists, then f is said to be n*–times differentiable* at x; if $f^{(n)}(x)$ exists for each positive integer n, then f is said to be *infinitely differentiable* at x. Definitions of notions such as nth *right–hand derivative,* n*–times differentiable on an interval,* and *infinitely differentiable on an interval* are formulated analogously.

(1.5.1) Exercises

.1 Prove that if f is differentiable at x_0, then it is continuous at x_0. Give an example of a function $f : \mathbf{R} \to \mathbf{R}$ such that $f'(0^-)$ and $f'(0^+)$ both exist but f is not continuous at 0.

.2 For each $x \in \mathbf{R}$ write

$$\rho(x, \mathbf{Z}) = \inf \{|x - n| : n \in \mathbf{Z}\}$$

and

$$f(x) = \sum_{n=0}^{\infty} \frac{\rho(10^n x, \mathbf{Z})}{10^n}.$$

Prove that f is continuous, but nowhere differentiable, on $\mathbf{R}$. (For continuity use Exercise (1.4.5: 10). To show that f is not differentiable at x, it is enough to take $0 \leq x < 1$. Let $0.d_1 d_2 \ldots$ be a decimal expansion of x, the terminating expansion if there is one. Define h_k to be -10^{-k} if $a_k = 4$ or 9, and 10^{-k} otherwise, and consider $h_k^{-1}(f(x + h_k) - f(x))$.)

This example is due to van der Waerden [54]. Weierstrass, in a lecture to the Berlin Academy in 1872, gave the first example of a continuous, nowhere differentiable function: namely,

$$f(x) = \sum_{n=1}^{\infty} a^n \cos(b^n \pi x),$$

where $0 < a < 1$, b is an odd positive integer, and $ab > 1 + 3\pi/2$; for a discussion of a special case of Weierstrass's example, see [28], pages 38–41.

.3 Prove that if f is differentiable at x, then

$$f'(x) = \lim_{h,k \to 0^+} \frac{f(x + h) - f(x - k)}{h + k}.$$

Give an example of a function f where $\lim_{h \to 0} ((f(h) - f(-h))/2h)$ exists but f is not differentiable at 0.

.4 Let $f(x) = x^n$, where n is an integer. Using the definition of "differentiable", prove that $f'(x) = nx^{n-1}$ for all $x \in \mathbf{R}$.

.5 Let f and g be differentiable at a. Prove that $f+g$, $f-g$, cf $(c \in \mathbf{R})$, and fg are differentiable at a, and that

$$\begin{aligned}(f+g)'(a) &= f'(a)+g'(a),\\ (f-g)'(a) &= f'(a)-g'(a),\\ (cf)'(a) &= cf'(a),\\ (fg)'(a) &= f(a)g'(a)+f'(a)g(a).\end{aligned}$$

.6 Under the conditions of the last exercise, suppose also that $g(a) \neq 0$. Give two proofs that f/g is differentiable at a, and that

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a)-f(a)g'(a)}{g(a)^2}.$$

.7 Using the exponential series, prove that $\exp'(0) = 1$. Hence prove that $\exp'(x) = \exp(x)$ for all $x \in \mathbf{R}$.

Our next proposition, the *Chain Rule,* is possibly the most troublesome result of elementary calculus.

(1.5.2) Proposition. *If f is differentiable at a, and g is differentiable at $f(a)$, then $g \circ f$ is differentiable at a, and*

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a).$$

Proof. Setting $b = f(a)$, define

$$h(u) = \begin{cases} \dfrac{g(u)-g(b)}{u-b} & \text{if } u \neq b \\ g'(b) & \text{if } u = b. \end{cases}$$

For all $x \neq a$ in some neighbourhood of a we have

$$\frac{g(f(x))-g(f(a))}{x-a} = (h \circ f)(x) \cdot \frac{f(x)-f(a)}{x-a}. \tag{1}$$

(Note that in verifying this identity, we must consider the possibility that $f(x) = f(a)$.) Since g is differentiable at b, h is continuous at b. Moreover, f is differentiable, and therefore (by Exercise (1.5.1: 1)) continuous, at a; so $h \circ f$ is continuous at a, by Exercise (1.4.5: 5). Hence

$$\begin{aligned}\lim_{x\to a}\left((h \circ f)(x) \cdot \frac{f(x)-f(a)}{x-a}\right) &= \lim_{x\to a}(h \circ f)(x) \cdot \lim_{x\to a}\frac{f(x)-f(a)}{x-a}\\ &= (h \circ f)(a) \cdot f'(a)\\ &= g'(b)f'(a).\end{aligned}$$

The result now follows immediately from (1). □

(1.5.3) Proposition. *Let I be an open interval, f a one–one continuous function on I, and $a \in I$, such that $f'(a)$ exists and is nonzero. Then the inverse function f^{-1} is differentiable at $f(a)$, and*

$$\left(f^{-1}\right)'(f(a)) = \frac{1}{f'(a)}.$$

Proof. Note that $J = f(I)$ is an interval, by Theorem (1.4.11) and Proposition (1.3.3); moreover, by Exercise (1.4.13: 7), $f(a)$ is an interior point of J and f^{-1} is continuous on J. Let (y_n) be any sequence in $J\backslash\{f(a)\}$ that converges to $f(a)$, and write $x_n = f^{-1}(y_n)$; then $x_n \neq a$ and

$$\lim_{n\to\infty} f^{-1}(x_n) = f^{-1}(f(a)) = a.$$

Since f is one–one, it follows that

$$\frac{f(x_n) - f(a)}{x_n - a} \neq 0;$$

whence

$$\lim_{n\to\infty} \frac{f^{-1}(y_n) - f^{-1}(f(a))}{y_n - f(a)} = \lim_{n\to\infty} \frac{x_n - a}{f(x_n) - f(a)} = \frac{1}{f'(a)}.$$

The desired conclusion now follows from Exercise (1.4.3: 6). □

(1.5.4) Exercises

.1 Prove that $\log'(x) = 1/x$ for each $x > 0$.

.2 Let $f(x) = x^r$, where $r \in \mathbf{R}$. Prove that $f'(x) = rx^{r-1}$. (Note that $x^r = \exp(r \log x)$.)

.3 Let f be a strictly increasing function on an interval I, and let a be a point of I such that $f'(a^+)$ exists and is nonzero. Prove that the inverse function f^{-1} is differentiable on the right at $f(a)$, and that

$$\left(f^{-1}\right)'\left(f(a)^+\right) = \frac{1}{f'(a^+)}.$$

.4 Let f be continuous on the compact interval $I = [a, b]$ and differentiable on (a, b). Prove that if $\xi \in (a, b)$ and $f(\xi) = \inf f$, then $f'(\xi) = 0$. Hence prove that if $f(a) = f(b)$, then there exists $\xi \in (a, b)$ such that $f'(\xi) = 0$ (*Rolle's Theorem*).

.5 Let f be continuous on the compact interval $[a,b]$ and n–times differentiable on (a,b). Suppose that there exist $n+1$ distinct points x of (a,b) at which $f(x)=0$. Show that $f^{(n)}(x)=0$ for some $x\in(a,b)$.

.6 Use Rolle's Theorem to prove the *Mean Value Theorem*: if f is continuous on the compact interval $[a,b]$ and differentiable on (a,b), then there exists $\xi\in(a,b)$ such that $f(b)-f(a)=f'(\xi)(b-a)$.

.7 Let f be differentiable on an interval I. Prove that

(i) if $f'(x)\geq 0$ for all $x\in I$, then f is increasing on I;

(ii) if $f'(x)>0$ for all $x\in I$, then f is strictly increasing on I;

(iii) if $f'(x)=0$ for all $x\in I$, then f is constant on I.

.8 Let f be differentiable on an interval I, with $f'(x)\neq 0$ for all $x\in I$. Prove that f is one–one, and that either $f'(x)\geq 0$ for all $x\in I$ or else $f'(x)\leq 0$ for all $x\in I$.

.9 Prove that if f is differentiable on an interval I, then the range of f' has the intermediate value property on I. (Let $f'(x_1)<y<f'(x_2)$, consider $g(x)=f(x)-yx$, and use the preceding exercise.)

.10 Prove *Cauchy's Mean Value Theorem*: if f,g are continuous on $[a,b]$ and differentiable on (a,b), then there exists $\xi\in(a,b)$ such that

$$(f(b)-f(a))\,g'(\xi)=(g(b)-g(a))\,f'(\xi).$$

(Consider the function $x\mapsto(f(b)-f(a))\,g(x)-(g(b)-g(a))\,f(x)$.)

.11 Let f,g be continuous on $[a,b]$ and differentiable on (a,b), let $x_0\in[a,b]$, and suppose that

(i) $g'(x)\neq 0$ for all $x\neq x_0$,

(ii) $f(x_0)=g(x_0)=0$,

(iii) $l=\lim\limits_{x\to x_0}(f'(x)/g'(x))$ exists.

Prove that $\lim_{x\to x_0}(f(x)/g(x))=l$. (*l'Hôpital's Rule*. Use the preceding exercise to show that if (x_n) is any sequence in $[a,b]\setminus\{x_0\}$ that converges to x_0, then $f(x_n)/g(x_n)\to l$ as $n\to\infty$.)

.12 Let g be twice differentiable at 0, with $g(0)=g'(0)=0$. Find $f'(0)$, where f is defined by

$$f(x)=\begin{cases}\dfrac{g(x)}{x} & \text{if } x\neq 0,\\ 0 & \text{if } x=0.\end{cases}$$

The following generalisation of the Mean Value Theorem is one of the most useful results of the differential calculus. Unfortunately, it seems to have no completely transparent, natural proof; all the proofs in the literature use some trick or other to obtain the desired conclusion.

(1.5.5) Taylor's Theorem. *Let f be $(N+1)$–times differentiable on an interval I, and let $a \in I$. Then for each $x \in I$ there exists ξ between a and x such that*

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(\xi)}{N!}(x-\xi)^N(x-a).$$

Proof. Fixing $x \in I$, without loss of generality we take $a < x$. Consider the function $g : [a, x] \to \mathbf{R}$ defined by

$$g(t) = f(x) - \sum_{n=0}^{N} \frac{f^{(n)}(t)}{n!}(x-t)^n.$$

Using Exercises (1.5.1: 4 and 5), we have

$$\begin{aligned} g'(t) &= -f'(t) - \sum_{n=1}^{N} \left(\frac{f^{(n+1)}(t)}{n!}(x-t)^n - \frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1} \right) \\ &= -\frac{f^{(N+1)}(t)}{N!}(x-t)^N. \end{aligned}$$

Applying the Mean Value Theorem (Exercise (1.5.4: 6)), we obtain $\xi \in (a, x)$ such that

$$-\frac{f^{(N+1)}(\xi)}{N!}(x-\xi)^N = g'(\xi) = \frac{g(x)-g(a)}{x-a}.$$

Then, as $g(x) = 0$,

$$g(a) = \frac{f^{(N+1)}(\xi)}{N!}(x-\xi)^N(x-a),$$

which is equivalent to the desired result. □

The expression

$$\sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

is called the *Taylor polynomial of degree* n *at* a, and

$$f(x) - \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

is called the *remainder term of order n* in Taylor's Theorem. The theorem, as stated, has the remainder term in the *Cauchy form*:

$$\frac{f^{(N+1)}(\xi)}{N!}(x-\xi)^N(x-a).$$

The next corollary gives us an alternative form—the *Lagrange form*—of the remainder.

(1.5.6) Corollary. *Under the hypotheses of Taylor's Theorem, for each $x \in I$ there exists t between a and x such that*

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(t)}{(N+1)!}(x-a)^{N+1}.$$

Proof. Again we take $a < x$ and use a trick. With g the function introduced in the preceding proof of Taylor's Theorem, we apply Cauchy's Mean Value Theorem (Exercise (1.5.4: 10)) to g and $t \mapsto (x-t)^{N+1}$. Since $g(x) = 0$, this yields $t \in (a, x)$ such that

$$\frac{g(a)}{(x-a)^{N+1}} = \frac{\frac{f^{(N+1)}(t)}{N!}(x-t)^N}{(N+1)(x-t)^N}.$$

Hence

$$g(a) = \frac{f^{(N+1)}(t)}{(N+1)!}(x-a)^{N+1},$$

from which the required conclusion follows. □

(1.5.7) Proposition. *Let I be the interval of convergence of the power series $\sum_{n=0}^{\infty} a_n x^n$, and let $f(x)$ be the sum of the series on I. Then f is differentiable, and $f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$, at each interior point of I.*

Proof. We first recall from Exercise (1.2.16: 16) that the power series $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=1}^{\infty} n a_n x^{n-1}$ have the same radius of convergence R. Given $x \in I^{\circ}$, let

$$r = \tfrac{1}{2}(|x| + R).$$

Then $\sum_{n=1}^{\infty} n a_n r^{n-1}$ converges absolutely, by Exercise (1.2.16: 10). Given $\varepsilon > 0$, choose N such that $\sum_{n=N+1}^{\infty} n\,|a_n|\,r^{n-1} < \varepsilon$. If $h \neq 0$ and $|x+h| < r$, then, using the Mean Value Theorem, for each $n \geq N+1$ we obtain θ_n between 0 and h such that

$$\frac{(x+h)^n - x^n}{h} = n(x+\theta_n)^{n-1}.$$

Hence

$$\begin{aligned}
&\left|\sum_{n=N+1}^{\infty} a_n\left(\frac{(x+h)^n-x^n}{h}-nx^{n-1}\right)\right| \\
&\leq \sum_{n=N+1}^{\infty} |a_n|\left|\frac{(x+h)^n-x^n}{h}-nx^{n-1}\right| \\
&\leq \sum_{n=N+1}^{\infty} n\,|a_n|\left|(x+\theta_n)^{n-1}-x^{n-1}\right| \\
&\leq \sum_{n=N+1}^{\infty} n\,|a_n|\left(|x+\theta_n|^{n-1}+|x|^{n-1}\right) \\
&\leq 2\sum_{n=N+1}^{\infty} n\,|a_n|\,r^{n-1} \\
&< 2\varepsilon.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\left|\frac{f(x+h)-f(x)}{h}-\sum_{n=1}^{\infty} na_nx^{n-1}\right| \\
&\leq \left|\sum_{n=1}^{N} a_n\left(\frac{(x+h)^n-x^n}{h}-nx^{n-1}\right)\right| \\
&\quad+\left|\sum_{n=N+1}^{\infty} a_n\left(\frac{(x+h)^n-x^n}{h}-nx^{n-1}\right)\right| \\
&< \sum_{n=1}^{N} |a_n|\left|\frac{(x+h)^n-x^n}{h}-nx^{n-1}\right|+2\varepsilon.
\end{aligned}$$

Now, there exists $\delta > 0$ such that if $0 < |h| < \delta$, then $|x+h| < r$ and

$$\left|\frac{(x+h)^n-x^n}{h}-nx^{n-1}\right| < \left(1+\sum_{n=1}^{N}|a_n|\right)^{-1}\varepsilon \quad (1 \leq n \leq N).$$

For such h we then have

$$\left|\frac{f(x+h)-f(x)}{h}-\sum_{n=1}^{\infty} na_nx^{n-1}\right| < 3\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the result follows. □

(1.5.8) Exercises

.1 Let f be infinitely differentiable on an interval I, and suppose that there exists $M > 0$ such that $\left|f^{(n)}(x)\right| \leq M$ for all sufficiently large n and all $x \in I$. Given $a \in I$, prove that the *Taylor expansion*, or *Taylor series, of f about a*,

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

converges to $f(x)$ for each $x \in I$.

.2 Find the Taylor expansion of $\exp(-x^2)$ about 0. For what values of x does this expansion converge to $\exp(-x^2)$?

.3 Prove that $f(x) = \exp(x)$ defines the unique differentiable function such that $f(0) = 1$ and $f'(x) = f(x)$ for all $x \in \mathbf{R}$.

.4 Let $R > 0$, and let $\sum_{n=0}^{\infty} a_n x^n, \sum_{n=0}^{\infty} b_n x^n$ be power series whose intervals of convergence include $(-R, R)$. Suppose that $\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} b_n x^n$ for all $x \in (-R, R)$. Show that $a_n = b_n$ for all n.

.5 Prove that $\sin' x = \cos x$ and that $\cos' x = -\sin x$ for each $x \in \mathbf{R}$. (See Exercise (1.4.5: 14) for the definition of sin and cos .)

.6 Prove the trigonometric addition formulae:

$$\begin{aligned} \cos(a+b) &= \cos a \cos b - \sin a \sin b, \\ \sin(a+b) &= \sin a \cos b + \cos a \sin b. \end{aligned}$$

(Define

$$\begin{aligned} F(x) = {} & (\cos(x+b) - \cos x \cos b + \sin x \sin b)^2 \\ & + (\sin(x+b) - \sin x \cos b - \cos x \sin b)^2 \end{aligned}$$

and consider $F'(x)$.)

.7 Prove that cos is a strictly decreasing function in the interval $[0, 2]$, and that there is a unique p such that $p/2 \in [0, 2]$ and $\cos(p/2) = 0$. Using the addition formulae from the preceding exercise, prove also that $\cos(x + 2p) = \cos x$ and $\sin(x + 2p) = \sin x$ for all $x \in \mathbf{R}$. (Of course, the number p is more usually denoted by π.)

.8 Derive the *binomial series*

$$(1+x)^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$$

for $-1 < x < 1$, where $\binom{\alpha}{0} = 1$ and for $n \geq 1$,

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)}{n!}.$$

(First show that the series in question does converge for $|x| < 1$. Then apply Taylor's Theorem with the Lagrange form of the remainder when $0 \leq x < 1$, and with the Cauchy form when $-1 < x < 0$.)

.9 Let $p \geq 2$, $1/p + 1/q = 1$, and $0 < c < 1$. Prove that

$$(1+c)^q + (1-c)^q - 2\left(1+c^p\right)^{q-1} \geq 0.$$

(Use the binomial series.)

.10 Let $x, y \in \mathbf{R}$, and let p, q be positive numbers with $1/p + 1/q = 1$. Use the preceding exercise to prove that

$$|x+y|^q + |x-y|^q \leq 2\left(|x|^p + |y|^p\right)^{q-1}$$

if $1 < p \leq 2$, and that

$$|x+y|^q + |x-y|^q \geq 2\left(|x|^p + |y|^p\right)^{q-1}$$

if $p \geq 2$.

Although the Riemann integral is taught in elementary calculus courses, for the best part of a century, following the development of a more sophisticated integral by Lebesgue and others, it has had little practical value. Indeed, Dieudonné, in characteristically forthright mood ([13], page 142), claims that the Riemann integral "has at best the importance of a mildly interesting exercise in the general theory of measure and integration. Only the stubborn conservatism of academic tradition could freeze it into a regular part of the curriculum, long after it had outlived its historical importance." We believe, nevertheless, that it is worth presenting a rigorous development of the Riemann integral for both historical and paedagogical reasons; but we skip lightly over this material, leaving much of it to the exercises. We discuss the Lebesgue integral in Chapter 2.

By a *partition* of a compact interval $I = [a, b]$ we mean a finite sequence $P = (x_0, x_1, \ldots, x_n)$ of points of I such that

$$a = x_0 \leq x_1 \leq \cdots \leq x_n = b.$$

The real number

$$\max\{x_{i+1} - x_i : 0 \leq i \leq n-1\}$$

is called the *mesh* of the partition. Loosely, we identify P with the set $\{x_0, \ldots, x_n\}$. A partition Q is called a *refinement* of P if $P \subset Q$.

Now let $f : I \to \mathbf{R}$ a bounded function, and for $0 \leq i \leq n-1$ define

$$m_i(f) = \inf\{f(x) : x_i \leq x \leq x_{i+1}\},$$
$$M_i(f) = \sup\{f(x) : x_i \leq x \leq x_{i+1}\}.$$

The real numbers

$$L(f,P) = \sum_{i=0}^{n-1} m_i(f)\,(x_{i+1} - x_i),$$
$$U(f,P) = \sum_{i=0}^{n-1} M_i(f)\,(x_{i+1} - x_i)$$

are called the *lower sum* and *upper sum*, respectively, for f and P. Since

$$(b-a)\inf f \leq L(f,P) \leq U(f,P) \leq (b-a)\sup f,$$

the *lower integral* of f,

$$\underline{\int_a^b} f = \sup\{L(f,P) : P \text{ is a partition of } [a,b]\},$$

and the *upper integral* of f,

$$\overline{\int_a^b} f = \inf\{U(f,P) : P \text{ is a partition of } [a,b]\},$$

exist.

(1.5.9) Lemma. *Let $f : [a,b] \to \mathbf{R}$ be bounded, and let P, Q be partitions of $[a,b]$. Then $L(f,P) \leq U(f,Q)$.*

Proof. Take $P = (x_0, \ldots, x_n)$, and first consider the case where $Q = P \cup \{\xi\}$ for some point $\xi \notin P$. Choose k such that $x_k < \xi < x_{k+1}$, and write

$$\alpha = \inf\{f(x) : x_k \leq x \leq \xi\},$$
$$\beta = \inf\{f(x) : \xi \leq x \leq x_{k+1}\}.$$

Then $m_k(f) = \min\{\alpha, b\}$, so

$$\begin{aligned} L(f,P) &= \sum_{\substack{i=0,\\ i\neq k}}^{n-1} m_i(f)\,(x_{i+1} - x_i) + m_k(f)(x_{k+1} - x_k) \\ &= \sum_{\substack{i=0,\\ i\neq k}}^{n-1} m_i(f)\,(x_{i+1} - x_i) + m_k(f)(\xi - x_k) + m_k(f)(x_{k+1} - \xi) \end{aligned}$$

$$\leq \sum_{\substack{i=0,\\ i\neq k}}^{n-1} m_i(f)\,(x_{i+1}-x_i) + \alpha(\xi - x_k) + \beta(x_{k+1}-\xi)$$
$$= L(f,Q).$$

Next, if Q is a refinement of P, then $Q = P \cup \{\xi_1, \ldots, \xi_m\}$ for some distinct points $\xi_k \notin P$, so

$$\begin{aligned} L(f,P) &\leq L(f,\, P\cup\{\xi_1\}) \\ &\leq L(f,\, P\cup\{\xi_1,\xi_2\}) \\ &\leq \cdots \\ &\leq L(f,Q). \end{aligned}$$

Similar arguments show that $U(f,Q) \leq U(f,P)$ when Q is a refinement of P.

Now consider any two partitions P, Q of $[a,b]$. Since $P \cup Q$ is a refinement of both P and Q, we have

$$L(f,P) \leq L(f,\, P\cup Q) \leq U(f,\, P\cup Q) \leq U(f,\, Q). \qquad \square$$

It follows from this lemma that, as we might have anticipated,

$$\underline{\int_a^b} f \leq \overline{\int_a^b} f.$$

We say that f is *Riemann integrable over* I if its lower and upper integrals coincide, in which case we define the *Riemann integral of* f *over* I to be

$$\int_a^b f = \underline{\int_a^b} f = \overline{\int_a^b} f.$$

We also define

$$\int_b^a f = -\int_a^b f$$

when f is Riemann integrable over $[a,b]$.

(1.5.10) Exercises

.1 Let $f : [a,b] \to \mathbf{R}$ be bounded. Prove that f is Riemann integrable if and only if for each $\varepsilon > 0$ there exists a partition P of $[a,b]$ such that $U(f,P) - L(f,P) < \varepsilon$.

.2 Let $f(x) = x^2$ on $[0,1]$, and for each positive integer n let P_n be the partition of $[0,1]$ consisting of the points i/n $(0 \leq i \leq n)$. By considering $L(f,P_n)$ and $U(f,P_n)$, show that f is Riemann integrable and that $\int_0^1 f = 1/3$.

.3 Prove that for any $n \in \mathbf{N}$ the function $x \mapsto x^n$ is Riemann integrable over $[a, b]$. (Use the first exercise in this set.)

.4 Prove that an increasing bounded function $f : [a, b] \to \mathbf{R}$ is Riemann integrable.

.5 Prove that a continuous function $f : [a, b] \to \mathbf{R}$ is Riemann integrable. (Use the Uniform Continuity Theorem (Exercise (1.4.8: 8)).)

.6 Define $f : [0, 1] \to \mathbf{R}$ by

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

Prove that f is not Riemann integrable.

.7 Define $f : [0, 1] \to \mathbf{R}$ by

$$f(x) = \begin{cases} 0 & \text{if } x = 0 \text{ or } x \text{ is irrational} \\ \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ for relatively prime positive integers } p, q. \end{cases}$$

Given $\varepsilon > 0$, let $t_0 = 0$; let $t_1, \ldots, t_{m-1}$ be, in increasing order, the points of $(0, 1)$ that have the form p/q where p, q are relatively prime positive integers with $0 < q < 2/\varepsilon$; and let $t_m = 1$. Taking $x_0 = 0$, construct inductively a partition $P = (x_0, x_1, \ldots, x_{2m+1})$ of $[0, 1]$ such that

$$x_{2k} < t_k < x_{2k+1} < \tfrac{1}{2}(t_k + t_{k+1}),$$

$$x_{2k+1} - x_{2k} < \frac{\varepsilon}{2(m+1)},$$

and $x_{2m+1} = 1$. Show that if $x_{2k-1} \leq x \leq x_{2k}$, then $f(x) < \varepsilon/2$; that $U(f, P) < \varepsilon$; and hence that f is Riemann integrable.

.8 Prove that the composition of two Riemann integrable functions need not be Riemann integrable. (Note the preceding two exercises.)

Since, by Exercise (1.5.10: 3), $f(x) = x$ defines a Riemann integrable function over $[a, b]$, the next result generalises Exercise (1.5.10: 5). It should also be compared with Exercise (1.5.10: 8).

(1.5.11) Proposition. *Let $f : [a, b] \to J$ be Riemann integrable, where J is a compact interval, and let $g : J \to \mathbf{R}$ be continuous. Then $g \circ f$ is Riemann integrable.*

Proof. Let

$$K = \sup \{g(y) : y \in J\},$$

which exists by Theorem (1.4.9), and write $I = [a, b]$. According to the Uniform Continuity Theorem (Exercise (1.4.8: 8)), for each $\varepsilon > 0$ there exists δ such that

$$0 < \delta < \frac{\varepsilon}{b - a + 2K}$$

and such that

$$|g(x) - g(y)| < \frac{\varepsilon}{b - a + 2K}$$

whenever $x, y \in J$ and $|x - y| < \delta$. Choose a partition $P = (x_0, \ldots, x_n)$ of I such that

$$U(f, P) - L(f, P) < \delta^2$$

(this is possible in view of Exercise (1.5.10: 1)), and write $P = S \cup T$, where

$$\begin{aligned} S &= \{i : M_i(f) - m_i(f) < \delta\}, \\ T &= \{i : M_i(f) - m_i(f) \geq \delta\}. \end{aligned}$$

If $i \in S$ and $x, x' \in [x_i, x_{i+1}]$, then $|f(x) - f(x')| < \delta$ and so

$$|g \circ f(x) - g \circ f(x')| < \frac{\varepsilon}{b - a + 2K}.$$

Hence

$$M_i(g \circ f) - m_i(g \circ f) \leq \frac{\varepsilon}{b - a + 2K},$$

and therefore

$$\begin{aligned} \sum_{i \in S} (M_i(g \circ f) - m_i(g \circ f))(x_{i+1} - x_i) &\leq \frac{\varepsilon}{b - a + 2K} \sum_{i=0}^{n-1} (x_{i+1} - x_i) \\ &= \frac{\varepsilon(b - a)}{b - a + 2K}. \end{aligned}$$

On the other hand,

$$\begin{aligned} \sum_{i \in T} (x_{i+1} - x_i) &\leq \delta^{-1} \sum_{i \in T} (M_i(f) - m_i(f))(x_{i+1} - x_i) \\ &\leq \delta^{-1} (U(f, P) - L(f, P)) \\ &< \delta, \end{aligned}$$

so

$$\begin{aligned} \sum_{i \in T} (M_i(g \circ f) - m_i(g \circ f))(x_{i+1} - x_i) &\leq 2K \sum_{i \in T} (x_{i+1} - x_i) \\ &< 2K\delta \\ &< \frac{2K\varepsilon}{b - a + 2K}. \end{aligned}$$

It now follows that

$$\begin{aligned}
&U(g\circ f, P) - L(g\circ f, P)\\
&= \sum_{i\in S} (M_i(g\circ f) - m_i(g\circ f))\,(x_{i+1}-x_i)\\
&\qquad + \sum_{i\in T} (M_i(g\circ f) - m_i(g\circ f))\,(x_{i+1}-x_i)\\
&< \frac{\varepsilon(b-a)}{b-a+2K} + \frac{2K\varepsilon}{b-a+2K} = \varepsilon.
\end{aligned}$$

Reference to Exercise (1.5.10: 1) completes the proof. □

Let $f : [a,b] \to \mathbf{R}$ be a bounded function, and $P = (x_0,\dots,x_n)$ a partition of $[a,b]$. Any expression of the form

$$\sum_{i=0}^{n-1} f(\xi_i)(x_{i+1}-x_i),$$

where $\xi_i \in [x_i, x_{i+1}]$ for each i, is called a *Riemann sum* for f (relative to P).

If f is Riemann integrable, geometric arguments like those presented in elementary calculus courses lead us to believe that if the partition P has small mesh, then the corresponding Riemann sums will closely approximate $\int_a^b f$. This expectation is fulfilled in the first of the next set of exercises.

(1.5.12) Exercises

.1 Prove that a bounded function $f : [a,b] \to \mathbf{R}$ is Riemann integrable if and only if there exists a real number Λ with the following property. For each $\varepsilon > 0$ there exists $\delta > 0$ such that $|\Sigma - \Lambda| < \varepsilon$ whenever

— P is a partition of $[a,b]$ with mesh less than δ, and

— Σ is a Riemann sum for f relative to P.

Prove that, in that case, $\Lambda = \int_a^b f$.

.2 Let f, g be Riemann integrable functions on $[a,b]$, and let $\lambda \in \mathbf{R}$. Prove that $f+g$ and λf are Riemann integrable, and that

$$\begin{aligned}
\int_a^b (f+g) &= \int_a^b f + \int_a^b g,\\
\int_a^b (\lambda f) &= \lambda \int_a^b f.
\end{aligned}$$

(Use the preceding exercise.)

.3 Let $f : [a,b] \to \mathbf{R}$ be bounded, and $a \le c \le b$. Prove that f is Riemann integrable over $[a,b]$ if and only if it is Riemann integrable over both $[a,c]$ and $[c,b]$, in which case

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

.4 Let $f : [a,b] \to \mathbf{R}$ be Riemann integrable. Prove that $f^+ = f \vee 0$, $f^- = (-f) \vee 0$, and $|f|$ are Riemann integrable, and that $\left|\int_a^b f\right| \le \int_a^b |f|$. Prove also that if $|f(x)| \le M$ for all $x \in [a,b]$, then

$$\left|\int_x^y f\right| \le M\,|x-y|$$

for all $x, y \in [a,b]$.

.5 Let f and g be Riemann integrable functions on $[a,b]$. Give two proofs that the product function fg is Riemann integrable over $[a,b]$. (For one proof, first take $f \ge 0$ and use Exercise (1.5.10: 1). For a second proof, note that $fg = \frac{1}{4}\left((f+g)^2 - (f-g)^2\right)$.)

.6 Let f be a nonvanishing Riemann integrable function on the compact interval $I = [a,b]$, and suppose that $1/f$ is bounded on I. Show that $1/f$ is Riemann integrable over I.

.7 Prove that if $f : [a,b] \to \mathbf{R}$ is continuous and nonnegative, and $\int_a^b f = 0$, then $f(x) = 0$ for all $x \in [a,b]$.

What makes the calculation of integrals feasible is the connection between integration and differentiation. There are various expressions of this connection, each of which may lay claim to the historic title of *Fundamental Theorem of Calculus*. Here is one strong version of that theorem.

(1.5.13) Theorem. *If F is differentiable on $[a,b]$, and F' is Riemann integrable over $[a,b]$, then*

$$\int_a^b F' = F(b) - F(a).$$

Proof. Let $P = (x_0, \ldots, x_n)$ be any partition of $[a,b]$. By the Mean Value Theorem, for each i there exists $\xi_i \in (x_i, x_{i+1})$ such that

$$F(x_{i+1}) - F(x_i) = F'(\xi_i)(x_{i+1} - x_i).$$

Hence

$$\sum_{i=0}^{n-1} F'(\xi_i)(x_{i+1} - x_i) = \sum_{i=0}^{n-1} (F(x_{i+1}) - F(x_i)) = F(b) - F(a).$$

Since P is *any* partition of $[a,b]$, the result now follows from Exercise (1.5.12: 1). □

Let F, f be two mappings of $[a,b]$ into $\mathbf{R}$. We say that F is a *primitive*, or *antiderivative*, of f on $[a,b]$ if $F'(x) = f(x)$ for all $x \in [a,b]$. In view of the Fundamental Theorem of Calculus (1.5.13), there is an obvious strategy for calculating the Riemann integral of a function $f : [a,b] \to \mathbf{R}$: first find a primitive F of f, and then compute $F(b) - F(a)$. Of course, as any student of calculus quickly learns, finding primitives of Riemann integrable functions is often a (literally) tricky business; moreover, the class of Riemann integrable functions f for which there exist primitives expressible in terms of elementary functions is relatively small [41]. So the Fundamental Theorem of Calculus has severe practical limitations, which have led to the development of highly accurate, fast methods of numerical integration (see, for example, [26]).

(1.5.14) Exercises

.1 Let f be Riemann integrable over $I = [a,b]$, and define

$$F(x) = \int_a^x f \quad (a \leq x \leq b).$$

Prove that F is continuous on I. Prove also that if $x_0 \in I$ and

$$\lim_{x \to x_0,\, x \in I} f(x) = f(x_0),$$

then

$$\lim_{x \to x_0,\, x \in I} \frac{F(x) - F(x_0)}{x - x_0} = f(x_0).$$

In particular, we obtain a result which is also sometimes called the *Fundamental Theorem of Calculus*: if f is continuous on $[a,b]$, then F is differentiable on $[a,b]$, $F'(x) = f(x)$ for all $x \in (a,b)$, $F'(a^+) = f(a)$, and $F'(b^-) = f(b)$.

.2 Let the power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ have radius of convergence R (which could be ∞). Prove that for each $x \in (-R, R)$,

$$\int_0^x f = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}.$$

.3 By considering

$$(1+x)(1-x+x^2-x^3+\cdots+(-1)^n x^n),$$

show that

$$\frac{1}{1+x}=\sum_{n=0}^{\infty}(-1)^n x^n \quad (-1<x<1).$$

Hence find power series expansions for $\log(1-x)$ and $(1-x)^{-1}\log(1-x)$ on $(-1,1)$. Then show that the identity

$$\tfrac{1}{2}\left(\log(1-x)\right)^2=\tfrac{1}{2}x^2+\tfrac{1}{3}\left(1+\tfrac{1}{2}\right)x^3+\tfrac{1}{4}\left(1+\tfrac{1}{2}+\tfrac{1}{3}\right)x^4+\cdots$$

is valid for each $x\in(-1,1]$.

.4 In this exercise you will prove that π^2, and therefore π itself, is irrational. Given a positive integer n, define

$$\phi(x)=\frac{1}{n!}x^n(1-x)^n.$$

Prove that

(i) $\phi^{(k)}(0)=0$ for $k<n$ or $k>2n$;

(ii) $\phi^{(k)}(0)$ and $\phi^{(k)}(1)$ are integers for all $k\in\mathbf{N}$.

Suppose that $\pi^2=p/q$ for some positive integers p,q, and define

$$F=q^n\sum_{k=0}^{n}(-1)^{2n-k}\pi^{2(n-k)}\phi^{(2k)}.$$

Show that $F(0)$ and $F(1)$ are integers, and that

$$F''(x)+\pi^2F(x)=\pi^2p^n\phi(x).$$

Setting

$$G(x)=F'(x)\sin\pi x-\pi F(x)\cos\pi x,$$

show that

$$G'(x)=\pi^2p^n\phi(x)\sin\pi x,$$

and hence that

$$\pi\int_0^1 p^n\phi(x)\sin\pi x\,\mathrm{d}x$$

is an integer. Finally, show that

$$0<\pi\int_0^1 p^n\phi(x)\sin\pi x\,\mathrm{d}x<\frac{\pi p^n}{n!}$$

for all positive integers n, and derive a contradiction.

We end this chapter by sketching the development of a generalisation of the Riemann integral. To do so we must first introduce another important class of functions on intervals.

In the rest of the section, unless we say otherwise, $I = [a, b]$ is a compact interval, and f a mapping of I into $\mathbf{R}$. For all $x, y \in I$ with $x \leq y$ we define the *variation* of f over $[x, y]$ to be

$$T_f(x, y) = \sup \left\{ \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)| : x = x_0 \leq x_1 \leq \cdots \leq x_n = y \right\}$$

if this quantity exists as a real number; we then say that f has *bounded variation* on $[x, y]$.

(1.5.15) Exercises

.1 Let f have bounded variation on I. Prove that

(i) f is bounded on I,

(ii) $T_f(a, b) = T_f(a, x) + T_f(x, b)$ for all $x \in I$, and

(iii) $T_f(a, \cdot)$ is an increasing function on I.

.2 Prove that f has bounded variation on I if and only if there exist increasing functions g, h on I such that $f = g - h$. (For "only if" note part (iii) of the preceding exercise.)

.3 Let f, g be functions of bounded variation on I. Prove that $f + g$, λf (where $\lambda \in \mathbf{R}$), and fg are of bounded variation, and that if $\inf \{|f(x)| : x \in I\} > 0$, then $1/f$ is of bounded variation.

.4 Define $f : [0, 1] \to \mathbf{R}$ by

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x^2} & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0. \end{cases}$$

Prove that f is differentiable at each point of $[0, 1]$ but does not have bounded variation on $[0, 1]$.

.5 Let $f : I \to \mathbf{R}$ have bounded variation. Prove that the one-sided limits $f(x^-)$ and $f(x^+)$ exist at each point of (a, b), as do $f(a^+)$ and $f(b^-)$, and that the set of points of I at which f is discontinuous is either empty or countable. (See Exercise (1.4.5: 8).)

.6 Let (f_n) be a sequence of functions of bounded variation on I such that $f(x) = \lim_{n\to\infty} f_n(x)$ exists for each $x \in I$. Prove that $T_f(a, b) \leq$

$\liminf T_{f_n}(a,b)$. (First show that for any partition $(x_0, x_1, \ldots, x_n)$ of I and any positive integer k,

$$\sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)| \le T_{f_k}(a,b) + \sum_{i=0}^{n-1} |f(x_{i+1}) - f_k(x_{i+1})| + \sum_{i=0}^{n-1} |f(x_i) - f_k(x_i)|.$$

Given $\varepsilon > 0$, then choose k appropriately.)

Now let $\alpha : I \to \mathbf{R}$ be a function with bounded variation on I, $P = (x_0, x_1, \ldots, x_n)$ a partition of I, and $f : I \to \mathbf{R}$ a bounded function. Any expression of the form

$$\sum_{i=0}^{n-1} f(\xi_i) \left(\alpha(x_{i+1}) - \alpha(x_i)\right),$$

where $\xi_i \in [x_i, x_{i+1}]$ for each i, is called a *Riemann–Stieltjes sum* for f (relative to P and α). We say that f is *Riemann–Stieltjes integrable* (over I with respect to α) if there exists a real number Λ with the following property. For each $\varepsilon > 0$ there exists $\delta > 0$ such that $|\Sigma - \Lambda| < \varepsilon$ whenever

— P is a partition of I with mesh less than δ and

— Σ is a Riemann–Stieltjes sum for f relative to P.

In that case, Λ—the *Riemann–Stieltjes integral of f with respect to α*—is the unique real number with this property, and is usually written

$$\int_a^b f(x)\,\mathrm{d}\alpha(x).$$

Exercise (1.5.12: 1) shows that the Riemann integral is just the special case of the Riemann–Stieltjes integral in which $\alpha(x) = x$.

(1.5.16) Exercises

In these exercises, α has bounded variation on I, and f, g are bounded real–valued functions on I.

.1 Why, in the foregoing definitions, do we require the function α to be of bounded variation?

.2 Let α be an increasing function on $I = [a,b]$. Given a partition $P = (x_0, x_1, \ldots, x_n)$ of I, we call the real numbers

$$L(f,P,\alpha) = \sum_{i=0}^{n-1} m_i(f)\left(\alpha(x_{i+1}) - \alpha(x_i)\right),$$
$$U(f,P,\alpha) = \sum_{i=0}^{n-1} M_i(f)\left(\alpha(x_{i+1}) - \alpha(x_i)\right),$$

respectively, the *lower sum* and the *upper sum* for f relative to P and α. Prove that the *lower integral*

$$\underline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) = \sup\left\{L(f,P,\alpha) : \ P \text{ is a partition of } I\right\}$$

and the *upper integral*

$$\overline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) = \inf\left\{U(f,P,\alpha) : \ P \text{ is a partition of } I\right\}$$

of f with respect to α exist. Prove also that f is Riemann–Stieltjes integrable with respect to α if and only if

$$\underline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) = \overline{\int_a^b} f(x)\,\mathrm{d}\alpha(x),$$

in which case their common value is $\int_a^b f(x)\,\mathrm{d}\alpha(x)$.

.3 Prove that if f is continuous, then it is Riemann–Stieltjes integrable with respect to α, and

$$\left|\int_a^b f(x)\,\mathrm{d}\alpha(x)\right| \le \int_a^b |f(x)|\ \mathrm{d}T_\alpha(a,x) \le MT_\alpha(a,b),$$

where $T_\alpha(a,x)$ is the variation of α on the interval $[a,x]$, and $M = \sup\{|f(x)| : a \le x \le b\}$. (Note that $T_\alpha(a,\cdot)$ has bounded variation, by Exercises (1.5.15: 1 and 2).)

.4 Prove that if f, g are Riemann–Stieltjes integrable with respect to α, then so are $f+g$, $f-g$, and λf (where $\lambda \in \mathbf{R}$); in which case,

$$\int_a^b (f(x)+g(x))\ \mathrm{d}\alpha(x) = \int_a^b f(x)\,\mathrm{d}\alpha(x) + \int_a^b g(x)\,\mathrm{d}\alpha(x),$$
$$\int_a^b (f(x)-g(x))\ \mathrm{d}\alpha(x) = \int_a^b f(x)\,\mathrm{d}\alpha(x) - \int_a^b g(x)\,\mathrm{d}\alpha(x),$$

and

$$\int_a^b \lambda f(x)\,\mathrm{d}\alpha(x) = \lambda \int_a^b f(x)\,\mathrm{d}\alpha(x).$$

.5 Let $a \le c \le b$. Prove that f is Riemann–Stieltjes integrable over $[a,b]$ (with respect to α) if and only if it is Riemann–Stieltjes integrable over both $[a,c]$ and $[c,b]$; in which case,

$$\int_a^b f(x)\,\mathrm{d}\alpha(x) = \int_a^c f(x)\,\mathrm{d}\alpha(x) + \int_c^b f(x)\,\mathrm{d}\alpha(x).$$

.6 Prove that if α has a continuous derivative on I, then the Riemann–Stieltjes integral $\int_a^b f(x)\,\mathrm{d}\alpha(x)$ exists and equals the Riemann integral $\int_a^b f\alpha'$.

.7 Let α, β be of bounded variation on I, and suppose that f is Riemann–Stieltjes integrable with respect to both α and β. Prove that f is Riemann–Stieltjes integrable with respect to $\alpha + \beta$, and that

$$\int_a^b f(x)\,\mathrm{d}(\alpha+\beta)(x) = \int_a^b f(x)\,\mathrm{d}\alpha(x) + \int_a^b f(x)\,\mathrm{d}\beta(x).$$

Prove also that for each $\lambda \in \mathbf{R}$, f is Riemann–Stieltjes integrable with respect to $\lambda\alpha$, and

$$\int_a^b f(x)\,\mathrm{d}(\lambda\alpha)(x) = \lambda \int_a^b f(x)\,\mathrm{d}\alpha(x).$$

The next lemma enables us to discuss the continuity of the function $T_f(a,\cdot)$.

(1.5.17) Lemma. *Let f have bounded variation on I. Then for each positive integer n there exists a function $g_n : I \to \mathbf{R}$ such that*

(i) $|f(x^+) - f(x)| = |g_n(x^+) - g_n(x)|$ *whenever* $a \le x < b$,

(ii) $|f(x) - f(x^-)| = |g_n(x) - g_n(x^-)|$ *whenever* $a < x \le b$,

(iii) $T_f(a,\cdot) - g_n$ *is an increasing function, and*

(iv) $0 \le T_f(a,x) - g_n(x) < 1/n$ *for all* $x \in I$.

Moreover, f and g_n are continuous at precisely the same points of I.

Proof. We may assume that $a < b$. Referring to Exercise (1.5.15: 5), choose points $a = x_0 < x_1 < \cdots < x_{m-1} < x_m = b$ such that f is continuous at x_i $(1 \le i \le m-1)$, and

$$\sum_{i=0}^{m-1} |f(x_{i+1}) - f(x_i)| > T_f(a,b) - \tfrac{1}{n}.$$

Setting $g_n(a) = 0$, construct the function g_n on the intervals $[x_i, x_{i+1}]$ inductively, as follows. Assume that $g_n(x)$ has been defined for $a \leq x \leq x_i$, where $i < m$, and consider x with $x_i < x \leq x_{i+1}$. If $f(x_{i+1}) - f(x_i) \geq 0$, set

$$g_n(x) = f(x) + g_n(x_i) - f(x_i);$$

if $f(x_{i+1}) - f(x_i) < 0$, set

$$g_n(x) = -f(x) + g_n(x_i) + f(x_i).$$

This completes the inductive construction.

If $x_i \leq x < x' \leq x_{i+1}$, then (see Exercise (1.5.15: 1))

$$\begin{aligned} T_f(a, x') - T_f(a, x) &= T_f(x, x') \\ &\geq |f(x') - f(x)| \\ &= |g_n(x') - g_n(x)| \\ &\geq g_n(x') - g_n(x). \end{aligned}$$

Thus if $x_{i-1} \leq x < x_i < \cdots < x_j < x' \leq x_{j+1}$, then

$$\begin{aligned} T_f(a, x') - T_f(a, x) &= T_f(a, x') - T_f(a, x_j) \\ &\quad + \sum_{k=i}^{j-1} (T_f(a, x_{k+1}) - T_f(a, x_k)) \\ &\quad + T_f(a, x_i) - T_f(a, x) \\ &\geq g_n(x') - g_n(x_j) + \sum_{k=i}^{j-1} (g_n(x_{k+1}) - g_n(x_k)) \\ &\quad + g_n(x_i) - g_n(x) \\ &= g_n(x') - g_n(x). \end{aligned}$$

It follows that $T_f(a, \cdot) - g_n$ is an increasing function. For all $x \in [a, b]$ we now have

$$\begin{aligned} 0 &= T_f(a, a) - g_n(a) \\ &\leq T_f(a, x) - g_n(x) \\ &\leq T_f(a, b) - g_n(b) \\ &= T_f(a, b) - \sum_{i=0}^{n-1} (g_n(x_{i+1}) - g_n(x_i)) \\ &= T_f(a, b) - \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)| \\ &< \tfrac{1}{n}. \end{aligned}$$

Finally, properties (i) and (ii) hold, since on each interval $[x_i, x_{i+1}]$ either $g_n - f$ or $g_n + f$ is constant; the last part of the statement of the lemma follows immediately. □

(1.5.18) Proposition. *Let f have bounded variation on I. Then*

$$T_f(a, x^+) - T_f(a, x) = \left|f(x^+) - f(x)\right|$$

if $a \leq x < b$, and

$$T_f(a, x) - T_f(a, x^-) = \left|f(x) - f(x^-)\right|$$

if $a < x \leq b$. Hence f and $T_f(a, \cdot)$ are continuous at precisely the same points of I.

Proof. For each positive integer n choose g_n as in the preceding lemma. For $a \leq x < x' < b$ we have

$$\begin{aligned}
&|T_f(a, x') - T_f(a, x) - g_n(x') + g_n(x)| \\
&\leq |T_f(a, x') - g_n(x')| + |T_f(a, x) - g_n(x)| \\
&< \tfrac{2}{n}.
\end{aligned}$$

Letting x' approach x, we see that

$$\left|T_f(a, x^+) - T_f(a, x) - g_n(x^+) + g_n(x)\right| \leq \tfrac{2}{n}.$$

Now noting that

$$g_n(x^+) - g_n(x) = \left|f(x^+) - f(x)\right|,$$

we obtain

$$T_f(a, x^+) - T_f(a, x) = \lim_{n\to\infty} \left(g_n(x^+) - g_n(x)\right) = \left|f(x^+) - f(x)\right|.$$

The rest of the proof is left as an exercise. □

The preceding proposition enables us to prove a result that is used later (in Theorem (6.1.18)) to establish the uniqueness of the representation of certain continuous linear functions.

(1.5.19) Proposition. *Let α be a function of bounded variation on $I = [a, b]$, and let D be the set consisting of a, b, and all points of (a, b) at which α is discontinuous. Then $\int_a^b f(x)\,\mathrm{d}\alpha(x) = 0$ for each continuous function $f : I \to \mathbf{R}$ if and only if $\alpha(x) = \alpha(a)$ for all $x \in I\backslash D$.*

Proof. Note that D is countable, by Exercise (1.5.15: 5). Suppose first that $\int_a^b f(x)\,\mathrm{d}\alpha(x) = 0$ for each continuous f on I, and consider any point

$\xi \in I\backslash D$. There exist arbitrarily small positive numbers $t \in I\backslash D$ such that $\xi < b - t$. For such t let f be the continuous function that equals 1 on the interval $[a, \xi]$, equals 0 on $[\xi + t,\, b]$, and is linear on $[\xi,\, \xi + t]$. Referring to Exercises (1.5.16: 5 and 3), we obtain the estimate

$$\begin{aligned} 0 &= \int_a^b f(x)\,\mathrm{d}\alpha(x) \\ &= \int_a^{\xi} f(x)\,\mathrm{d}\alpha(x) + \int_{\xi}^{\xi+t} f(x)\,\mathrm{d}\alpha(x) + \int_{\xi+t}^{b} f(x)\,\mathrm{d}\alpha(x) \\ &\leq \alpha(\xi) - \alpha(a) + T_{\alpha}\,(\xi,\, \xi + t) + 0. \end{aligned}$$

Letting t tend to 0 and using Proposition (1.5.18), we see that $\alpha(\xi) = \alpha(a)$.

Now suppose, conversely, that $\alpha(x) = \alpha(a)$ for all $x \in I\backslash D$, and let f be any continuous real–valued function on I. Given $\varepsilon > 0$, choose $\delta > 0$ as in the definition of $\int_a^b f(x)\,\mathrm{d}\alpha(x)$. In view of Exercise (1.5.15: 5), we can choose a partition P, with mesh less than δ, consisting of a, b, and points of $I\backslash D$. For any Riemann–Stieltjes sum Σ for f corresponding to P and α, we then have $\Sigma = 0$ and therefore

$$\left| \int_a^b f(x)\,\mathrm{d}\alpha(x) \right| = \left| \Sigma - \int_a^b f(x)\,\mathrm{d}\alpha(x) \right| < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\int_a^b f(x)\,\mathrm{d}\alpha(x) = 0$. □

(1.5.20) Exercises

.1 Complete the proof of Proposition (1.5.18).

.2 Let α be of bounded variation on $I = [a, b]$. Prove that $\int_a^b f(x)\,\mathrm{d}\alpha(x) = 0$ for each continuous $f : I \to \mathbf{R}$ if and only if $\alpha(x) = \alpha(a)$ for all x in a dense subset of I that includes b.

We bring our treatment of the Riemann and Riemann–Stieltjes integrals to an end here. In the next chapter we develop a type of integral, based on a generalisation of the integral as an antiderivative, that is much more powerful than the Riemann integral, and for which it is possible to construct (although we do not do so) a related generalisation analogous to that of Riemann–Stieltjes.

For more on Riemann and Riemann–Stieltjes integration see [17], [42], or [50].

2

Differentiation and the Lebesgue Integral

More matter with less art.
HAMLET, Act 2, Scene 2

In the first section of this chapter we show how the ideas of Chapter 1 can be applied in a theory of the length of a subset of **R**; this leads to the Vitali Covering Theorem, a result with many interesting applications in the theory of differentiation and integration. Building on that material, in the next two sections we describe F. Riesz's development of Lebesgue integration as the inverse process to differentiation "almost everywhere".

2.1 Outer Measure and Vitali's Covering Theorem

Can we assign to a subset A of $\mathbf{R}$ a measure of its length? We have already done this when A is a bounded interval; but what about a more general set A? The answer lies in measure theory, a subject that was pioneered by Lebesgue, Borel, and others at the beginning of this century and which has proved of immense importance in analysis, probability theory, and many other areas of mathematics.

The *outer measure* of A is the quantity

$$\begin{aligned}\mu^*(A) \quad = \quad \inf\Big\{\textstyle\sum_{n=1}^{\infty} |I_n| : (I_n)_{n=1}^{\infty} \text{ is a cover of } A \\ \text{by bounded open intervals}\Big\},\end{aligned}$$

which we take as ∞ if the set on the right–hand side is unbounded.[1] If $\mu^*(A) \in \mathbf{R}$, we say that A *has finite outer measure*. Note that since, for any sequence (I_n) of bounded open intervals that covers A, the terms of the series $\sum_{n=1}^{\infty} |I_n|$ are all positive, the (possibly infinite) sum of the series does not depend on the order of those terms; this is an immediate consequence of Exercise (1.2.17: 1).

If A has outer measure zero, then we say that A is a *set of measure zero*, or that A *has measure zero*. Thus A has measure zero if and only if for each $\varepsilon > 0$ there exists a sequence $(I_n)_{n=1}^{\infty}$ of bounded open intervals such that $A \subset \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} |I_n| < \varepsilon$.

(2.1.1) Exercises

.1 Show that for each $A \subset \mathbf{R}$, $\mu^*(A)$ is the infimum of $\sum_{n=1}^{\infty} |I_n|$ taken over all covers of A by sequences $(I_n)_{n=1}^{\infty}$ of bounded, but not necessarily open, intervals.

.2 Prove that if a subset A of $\mathbf{R}$ has finite outer measure, then for each $\varepsilon > 0$ there exists a sequence (I_n) of disjoint bounded open intervals such that $A \subset \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} |I_n| < \mu^*(A) + \varepsilon$. (Use Proposition (1.3.6).)

.3 Show that $\mu^*(\emptyset) = 0$, and that if $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$.

.4 Prove that for each $a \in \mathbf{R}$, $\mu^*(\{a\}) = 0$.

.5 Let A be a subset of $\mathbf{R}$, and $E \subset A$ a set of measure zero. Show that $\mu^*(A \backslash E) = \mu^*(A)$.

.6 Let A be a subset of a compact interval I. Prove that $\mu^*(A) + \mu^*(I \backslash A) \geq |I|$. (It follows from results towards the end of Section 3 of this chapter that, perhaps surprisingly, we cannot replace inequality by equality in this result.)

.7 Let (A_n) be a sequence of subsets of $\mathbf{R}$. Show that

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n),$$

where the right–hand side is taken as ∞ if either any of its terms is ∞ or the series diverges. (If one of the sets A_n has infinite outer measure, then the inequality is trivial. If each A_n has finite outer measure, then for each positive integer n and each $\varepsilon > 0$ there exists a sequence

[1] In Section 1 of Chapter 3 we give a precise meaning to this use of ∞ as an "extended real number".

$(I_{n,k})_{k=1}^{\infty}$ of bounded open intervals such that $A_n \subset \bigcup_{k=1}^{\infty} I_{n,k}$ and $\sum_{k=1}^{\infty} |I_{n,k}| < \mu^*(A_n) + 2^{-n}\varepsilon$.)

Prove that if also the sets A_n are pairwise–disjoint, then

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu^*(A_n).$$

.8 Give two proofs that a countable subset of $\mathbf{R}$ has measure zero. Hence prove that $\mathbf{R}$ is uncountable.

.9 Give two proofs that the union of a sequence of sets of measure zero has measure zero.

.10 Prove that a subset A of $\mathbf{R}$ has finite outer measure if and only if $l = \lim_{n\to\infty} \mu^*(A \cap [-n, n])$ exists, in which case $\mu^*(A) = l$.

.11 Prove that μ^* is *translation invariant*—that is, $\mu^*(A + t) = \mu^*(A)$ for each $A \subset \mathbf{R}$ and each $t \in \mathbf{R}$, where $A + t = \{x + t : x \in A\}$.

(2.1.2) Proposition. *The outer measure of any interval in* $\mathbf{R}$ *equals the length of the interval.*

Proof. Consider, to begin with, a bounded closed interval $[a, b]$. For each $\varepsilon > 0$ we have $[a, b] \subset (a - \varepsilon, b + \varepsilon)$ and therefore

$$\mu^*([a, b]) \leq |(a - \varepsilon, b + \varepsilon)| = b - a + 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\mu^*([a, b]) \leq b - a$. To prove the reverse inequality, let (I_n) be any sequence of bounded open intervals that covers $[a, b]$. Applying the Heine–Borel–Lebesgue Theorem (1.4.6), and re–indexing the terms I_n (which we can do without loss of generality), we may assume that for some N,

$$[a, b] \subset I_1 \cup I_2 \cup \cdots \cup I_N.$$

There exists an interval I_{k_1}, where $1 \leq k_1 \leq N$, that contains a; let this interval be (a_1, b_1). Either $b < b_1$, in which case we stop the procedure, or else $b_1 \leq b$. In the latter case, $b_1 \in [a, b]\backslash(a_1, b_1)$; so there exists an interval I_{k_2}, where $1 \leq k_2 \leq N$ and $k_2 \neq k_1$, that contains b_1; call this interval (a_2, b_2). Repeating this argument, we obtain intervals (a_1, b_1), $(a_2, b_2), \ldots$ in the collection $\{I_1, \ldots, I_N\}$ such that for each i, $a_i < b_{i-1} < b_i$. This procedure must terminate with the construction of (a_j, b_j) for some $j \leq N$. Then $b \in (a_j, b_j)$, so

$$\sum_{n=1}^{N} |I_n| \geq \sum_{i=1}^{j} (b_i - a_i)$$

$$
\begin{aligned}
&= b_j - (a_j - b_{j-1}) - (a_{j-1} - b_{j-2}) \\
&\qquad - \cdots - (a_2 - b_1) - a_1 \\
&> b_j - a_1.
\end{aligned}
$$

It follows that $\sum_{n=1}^{\infty} |I_n| > b-a$ and therefore, since (I_n) was any sequence of bounded open intervals covering $[a, b]$, that $\mu^*([a, b]) \geq b - a$. Coupled with the reverse inequality already established, this proves that $\mu^*([a, b]) = b - a$.

The proof for other types of interval is left as the next exercise. □

(2.1.3) Exercises

.1 Complete the proof of Proposition (2.1.2) in the remaining cases.

.2 Let $\{I_1, \ldots, I_N\}$ be a finite set of bounded open intervals covering $\mathbf{Q} \cap [0, 1]$. Prove that $\sum_{n=1}^{N} |I_n| \geq 1$. (Given $\varepsilon > 0$, extend each I_n, if necessary, to ensure that it has rational endpoints and that the total length of the intervals is increased by at most ε. Then argue as in the proof of Proposition (2.1.2).)

.3 Let X be a subset of $\mathbf{R}$ with finite outer measure. Prove that for each $\varepsilon > 0$ there exists an open set $A \supset X$ with finite outer measure, such that $\mu^*(A) < \mu^*(X) + \varepsilon$. (Use Exercise (2.1.1: 2).) Show that if X is also bounded, then we can choose A to be bounded.

Let X be a subset of $\mathbf{R}$, and $\mathcal{V}$ a family of nondegenerate intervals—that is, intervals each having positive length. We say that $\mathcal{V}$ is a *Vitali covering* of X if for each $\varepsilon > 0$ and each $x \in X$ there exists $I \in \mathcal{V}$ such that $x \in I$ and $|I| < \varepsilon$.

(2.1.4) The Vitali Covering Theorem. *Let $\mathcal{V}$ be a Vitali covering of a set $X \subset \mathbf{R}$ with finite outer measure. Then for each $\varepsilon > 0$ there exists a finite set $\{I_1, \ldots, I_N\}$ of pairwise–disjoint intervals in $\mathcal{V}$ such that*

$$\mu^*\left(X \backslash \bigcup_{n=1}^{N} I_n\right) < \varepsilon.$$

We postpone the proof of this very useful theorem until we have dealt with some auxiliary exercises.

(2.1.5) Exercises

.1 Let $\mathcal{V}$ be a Vitali covering of a subset X of $\mathbf{R}$, x a point of X, and A an open subset of $\mathbf{R}$ containing X. Show that for each $\varepsilon > 0$ there exists $I \in \mathcal{V}$ such that $x \in I$, $I \subset A$, and $|I| < \varepsilon$.

.2 Let $I_1, \ldots, I_N$ be finitely many closed intervals belonging to a Vitali covering $\mathcal{V}$,of a subset X of $\mathbf{R}$ with finite outer measure, and let $x \in X \backslash \bigcup_{n=1}^{N} I_n$. Show that for each $\varepsilon > 0$ there exists $I \in \mathcal{V}$ such that $x \in I$, $|I| < \varepsilon$, and I is disjoint from $\bigcup_{n=1}^{N} I_n$.

Proof of the Vitali Covering Theorem. If necessary replacing the intervals in I by their closures, we may assume that $\mathcal{V}$ consists of closed intervals. Referring to Exercise (2.1.3: 3), choose an open set $A \supset X$ with finite outer measure. In view of Exercise (2.1.5: 1), we may assume without loss of generality that

$$I \subset A \text{ for each } I \in \mathcal{V}. \tag{1}$$

Choosing any interval I_1 in the covering $\mathcal{V}$, we construct pairwise–disjoint intervals $I_1, I_2, \ldots$ in $\mathcal{V}$ inductively as follows. Assume that we have constructed $I_1, \ldots, I_n$ in $\mathcal{V}$. If $X \subset \bigcup_{k=1}^{n} I_k$, then $\mu^* \left(X \backslash \bigcup_{k=1}^{n} I_k\right) = 0$ and we stop the construction. If X is not contained in $\bigcup_{k=1}^{n} I_k$, then Exercise (2.1.5: 2) shows that the set

$$S_n = \left\{ |I| : I \in \mathcal{V},\ I \cap \bigcup_{k=1}^{n} I_k = \emptyset \right\}$$

is nonempty. Since, by (1), S_n is bounded above by $\mu^*(A)$, it follows that

$$s_n = \sup S_n$$

exists; moreover, as each $I \in \mathcal{V}$ is nondegenerate, $s_n > 0$. To complete our inductive construction, we now choose $I_{n+1} \in \mathcal{V}$ such that $I_{n+1} \cap \bigcup_{k=1}^{n} I_k = \emptyset$ and $|I_{n+1}| > \frac{1}{2} s_n$.

We may assume that this construction leads to an infinite sequence $(I_n)_{n=1}^{\infty}$ of pairwise–disjoint elements of $\mathcal{V}$. Since the partial sums of the series $\sum_{n=1}^{\infty} |I_n|$ are bounded by $\mu^*(A)$, the monotone sequence principle (Proposition (1.2.4)) ensures that the series converges. Given $\varepsilon > 0$, we can therefore find N such that

$$\sum_{n=N+1}^{\infty} |I_n| < \frac{\varepsilon}{5}.$$

For each $n > N$ let x_n be the midpoint of I_n, and let J_n be the closed interval with midpoint x_n and length $5\,|I_n|$. It suffices to prove that

$$X \backslash \bigcup_{n=1}^{N} I_n \subset \bigcup_{n=N+1}^{\infty} J_n. \tag{2}$$

For then

$$\mu^*\left(X\backslash\bigcup_{n=1}^{N} I_n\right) \leq \sum_{n=N+1}^{\infty} |J_n| = 5\sum_{n=N+1}^{\infty} |I_n| < \varepsilon.$$

To prove (2), consider any $x \in X\backslash\bigcup_{n=1}^{N} I_n$. By Exercise (2.1.5: 2), there exists $I \in \mathcal{V}$ such that $x \in I$ and $I \cap \bigcup_{n=1}^{N} I_n = \emptyset$. We claim that $I \cap I_m$ is nonempty for some $m > N$. If this were not the case, then for each m we would have $I \cap \bigcup_{n=1}^{m} I_n = \emptyset$ and therefore $|I| \leq s_m < 2|I_{m+1}|$; since $\lim_{m\to\infty} |I_m| = 0$ (by Exercise (1.2.14: 1)), it would follow that $|I| = 0$, which is absurd as $\mathcal{V}$ contains only nondegenerate intervals. Thus

$$\nu = \min\{m > N : I \cap I_m \neq \emptyset\}$$

is well defined, $I \cap \bigcup_{n=1}^{\nu-1} I_n = \emptyset$, and therefore $|I| \leq s_{\nu-1} < 2|I_\nu|$. Since $x \in I$ and $I \cap I_\nu \neq \emptyset$, we see that

$$|x - x_\nu| \leq |I| + \tfrac{1}{2}|I_\nu| < 2|I_\nu| + \tfrac{1}{2}|I_\nu| = \tfrac{5}{2}|I_\nu|.$$

Hence $x \in J_\nu$. This establishes (2) and completes the proof. □

In the remainder of this section we apply the Vitali Covering Theorem in the proofs of some fundamental results in the theory of differentiation and integration.

Let I be an interval in $\mathbf{R}$. We say that a mapping $f : I \to \mathbf{R}$ is *absolutely continuous* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that if $([a_k, b_k])_{k=1}^{n}$ is a finite family of nonoverlapping[2] compact subintervals of I such that $\sum_{k=1}^{n}(b_k - a_k) < \delta$, then $\sum_{k=1}^{n}|f(b_k) - f(a_k)| < \varepsilon$.

(2.1.6) Exercises

.1 Prove that an absolutely continuous function on I is both uniformly continuous and bounded.

.2 Let f, g be absolutely continuous functions on I. Prove that the functions $f+g$, $f-g$, λf (where $\lambda \in \mathbf{R}$), and fg are absolutely continuous, and that if $\inf\{|f(x)| : x \in I\} > 0$, then $1/f$ is absolutely continuous.

.3 Prove that if f is differentiable, with bounded derivative, on an interval I, then f is absolutely continuous.

.4 Let f be absolutely continuous on a compact interval $I = [a, b]$. Prove that f has bounded variation in I, that the variation function $T_f(a, \cdot)$ is absolutely continuous on I, and that f is the difference of two absolutely continuous, increasing functions on I. (See Exercises (1.5.15: 1 and 2).)

---

[2]Two intervals in $\mathbf{R}$ are *nonoverlapping* if their intersection is either empty or contains only endpoints of the intervals.

Let S be a subset of $\mathbf{R}$, and $P(x)$ a statement about real numbers x. If there exists a set E of measure zero such that $P(x)$ holds for all x in $S \backslash E$, then we say that $P(x)$ *holds almost everywhere on* S, or, more loosely, that P *holds almost everywhere on* S; in the case $S = \mathbf{R}$ we say simply that $P(x)$, or P, holds *almost everywhere.*

A simple corollary of the Mean Value Theorem (Exercise (1.5.4: 6)), one that suffices for many applications, states that if f is continuous on $[a, b]$ and $|f'(x)| \leq M$ for all $x \in (a, b)$, then $|f(b) - f(a)| \leq M(b - a)$. Our first application of the Vitali Covering Theorem generalises this corollary, and can be regarded an extension of the Mean Value Theorem itself.

(2.1.7) Proposition. *Let f be an absolutely continuous mapping of a compact interval $I = [a, b]$ into $\mathbf{R}$, and F a differentiable increasing mapping of I into $\mathbf{R}$ such that $|f'(x)| \leq F'(x)$ almost everywhere on I. Then*

$$|f(b) - f(a)| \leq F(b) - F(a). \tag{3}$$

Proof. Let $E \subset I$ be a set of measure zero such that $|f'(x)| \leq F'(x)$ for each $x \in X = I \backslash E$. We may assume without loss of generality that $a, b \in E$. Given $\varepsilon > 0$, choose $\delta > 0$ as in the definition of absolute continuity. For each $x \in X$ there exist arbitrarily small $r > 0$ such that $[x, x + r] \subset (a, b)$,

$$\begin{aligned} |f(x + r) - f(x) - f'(x)r| &< \varepsilon r, \\ |F(x + r) - F(x) - F'(x)r| &< \varepsilon r, \end{aligned}$$

and therefore

$$\begin{aligned} |f(x + r) - f(x)| &\leq |f'(x)|\, r + \varepsilon r \\ &\leq F'(x)r + \varepsilon r \\ &\leq F(x + r) - F(x) + 2\varepsilon r. \end{aligned}$$

The sets of the form $[x, x + r]$, for such $r > 0$, form a Vitali covering of X. By the Vitali Covering Theorem, there exists a finite, pairwise–disjoint collection $([x_k, x_k + r_k])_{k=1}^{N}$ of sets of this type such that

$$\mu^* \left(X \backslash \bigcup_{k=1}^{N} [x_k, x_k + r_k] \right) < \delta.$$

We may assume that $x_k + r_k < x_{k+1}$ for $1 \leq k \leq N - 1$. Thus

$$x_1 - a + \sum_{k=1}^{N-1} (x_{k+1} - x_k - r_k) + b - x_N - r_N < \delta,$$

and therefore

$$|f(x_1) - f(a)| + \sum_{k=1}^{N-1} |f(x_{k+1}) - f(x_k + r_k)| + |f(b) - f(x_N + r_N)| < \varepsilon.$$

It follows that

$$\begin{aligned}
|f(b) - f(a)| &\leq |f(x_1) - f(a)| + \sum_{k=1}^{N-1} |f(x_{k+1}) - f(x_k + r_k)| \\
&\quad + |f(b) - f(x_N + r_N)| + \sum_{k=1}^{N} |f(x_k + r_k) - f(x_k)| \\
&< \varepsilon + \sum_{k=1}^{N} (F(x_k + r_k) - F(x_k) + 2\varepsilon r_k) \\
&< \varepsilon + F(x_1) - F(a) + \sum_{k=1}^{N-1} (F(x_{k+1}) - F(x_k + r_k)) \\
&\quad + \sum_{k=1}^{N} (F(x_k + r_k) - F(x_k) + 2\varepsilon r_k) \\
&\quad + (F(b) - F(x_N + r_N)) \\
&= \varepsilon + F(b) - F(a) + 2\varepsilon \sum_{k=1}^{N} r_k \\
&< F(b) - F(a) + \varepsilon(1 + 2b - 2a).
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that (3) holds. □

(2.1.8) Exercises

.1 Let f be absolutely continuous on $I = [a, b]$, and suppose that for some constant M, $|f'| \leq M$ almost everywhere on I. Prove that $|f(b) - f(a)| \leq M(b - a)$.

.2 Let $f : [a, b] \to \mathbf{R}$ be an absolutely continuous function such that $f'(x) = 0$ almost everywhere on $I = [a, b]$. Give two proofs that f is a constant function. (For one proof use the Vitali Covering Theorem.)

.3 Let f, F be continuous on $I = [a, b]$, and suppose there exists a countable subset D of I such that $|f'(x)| \leq F'(x)$ for all $x \in I \backslash D$. Show that $|f(b) - f(a)| \leq F(b) - F(a)$. (We may assume that D is countably infinite. Let $d_1, d_2, \ldots$ be a one–one mapping of $\mathbf{N}^+$ onto D. Given $\varepsilon > 0$, let X be the set of all points $x \in I$ such that

$$|f(\xi) - f(a)| \leq F(\xi) - F(a) + \varepsilon \left(\xi - a + \sum_{\{n : d_n < \xi\}} 2^{-n} \right)$$

for all $\xi \in [a, x)$, and let $s = \sup X$. Assume that $s < b$, and derive a contradiction.)

.4 Let f be continuous on $I = [a, b]$, and suppose there exists a countable subset D of I such that $f'(x) = 0$ for all $x \in I \backslash D$. Prove that f is constant on I.

.5 Let C be the Cantor set (see Exercise (1.3.8: 11)). Show that $[0, 1] \backslash C$ is a countable union of nonoverlapping open intervals $(J_n)_{n=1}^{\infty}$ whose lengths sum to 1, and that C has measure zero.

For each $x = \sum_{n=1}^{\infty} a_n 3^{-n} \in C$ define $F(x) = \sum_{n=1}^{\infty} a_n 2^{-n-1}$. Show that

(i) if x has two ternary expansions, then they produce the same value for $F(x)$, so that F is a function on C;

(ii) F is a strictly increasing, continuous mapping of C onto $[0, 1]$;

(iii) C is uncountable; and

(iv) F extends to an increasing continuous mapping that is constant on each J_n, equals 0 throughout $(-\infty, 0]$, and equals 1 throughout $[1, \infty)$.

Prove that for each $\delta > 0$ there exist finitely many points

$$a_1 < 0 < b_1 < a_2 < \cdots < b_{N-1} < a_N < 1 < b_N$$

of $[-1, 2]$ such that $C \subset \bigcup_{n=1}^{N} [a_n, b_n]$,

$$\sum_{n=1}^{N} (F(b_n) - F(a_n)) = 1,$$

and $\sum_{n=1}^{N} (b_n - a_n) < \delta$. (Thus F is increasing and continuous, but not absolutely continuous, on $[-1, 2]$.)

Finally, show that $F'(x) = 0$ for all $x \in [0, 1] \backslash C$, but $F(1) > F(0)$.

The last two exercises deserve further comment. Consider a continuous function F on $[0, 1]$ whose derivative exists and vanishes throughout $[0, 1] \backslash E$. If E is countable, then Exercise (2.1.8: 4) shows that F is constant. On the other hand, Exercise (2.1.8: 5) shows that if E is uncountable and of measure zero, then F need not be constant; but if, in that case, F is absolutely continuous, then it follows from Exercise (2.1.8: 2) that it is constant.

Although the derivative of a function f may not exist at a point $x \in \mathbf{R}$, one or more of the following quantities—the *Dini derivates* of f at x—may:

$$D^+f(x) = \lim_{h \to 0^+} \sup \frac{f(x+h) - f(x)}{h},$$
$$D_+f(x) = \lim_{h \to 0^+} \inf \frac{f(x+h) - f(x)}{h},$$
$$D^-f(x) = \lim_{h \to 0^-} \sup \frac{f(x+h) - f(x)}{h},$$
$$D_-f(x) = \lim_{h \to 0^-} \inf \frac{f(x+h) - f(x)}{h}.$$

We consider $D^+f(x)$ to be undefined if

— either there is no $h > 0$ such that f is defined throughout the interval $[x, x+h]$

— or else $(f(x+h) - f(x))/h$ remains unbounded as $h \to 0^+$.

Similar comments apply to the other derivates of f.

(2.1.9) Exercises

.1 Prove that $D^+f(x) \geq D_+f(x)$ and $D^-f(x) \geq D_-f(x)$ whenever the quantities concerned make sense.

.2 Prove that f is differentiable on the right (respectively, left) at x if and only if $D^+f(x) = D_+f(x)$ (respectively, $D^-f(x) = D_-f(x)$).

.3 Let f be a mapping of $\mathbf{R}$ into $\mathbf{R}$, and define $g(x) = -f(-x)$. Prove that for each $x \in \mathbf{R}$, $D^+g(x) = D^-f(-x)$ and $D_-g(x) = D_+f(-x)$.

.4 Let $f : [a, b] \to \mathbf{R}$ be continuous, and suppose that one of the four derivates of f is nonnegative throughout (a, b). Prove that f is an increasing function on $[a, b]$. (Show that $x \mapsto f(x) + \varepsilon x$ is increasing for each $\varepsilon > 0$.)

.5 Consider a function $f : [a, b] \to \mathbf{R}$, and real numbers r, s with $r > s$. Define

$$E = \left\{x \in (a, b) : D^+f(x) > r > s > D_-f(x)\right\}.$$

Let X be an open set such that $E \subset X$ and $\mu^*(X) < \mu^*(E) + \varepsilon$ (see Exercise (2.1.3: 3)). Prove that the intervals of the form $(x - h, x)$ such that $x \in E$, $h > 0$, $[x - h, x] \subset X$, and $f(x) - f(x - h) < sh$ form a Vitali covering of E. Hence prove that for each $\varepsilon > 0$ there exist finitely many points $x_1, \ldots, x_m$ of E, and finitely many positive numbers $h_1, \ldots, h_m$, such that the intervals $J_i = (x_i - h_i, x_i)$ $(1 \leq i \leq m)$ form a pairwise–disjoint collection,

$$\mu^*\left(\bigcup_{i=1}^{m} J_i\right) > \mu^*(E) - \varepsilon$$

and

$$\sum_{i=1}^{m} (f(x_i) - f(x_i - h_i)) < s\,(\mu^*(E) + \varepsilon)\,.$$

Again applying the Vitali Covering Theorem, prove that there exist finitely many points $y_1, \ldots, y_n$ of $E \cap \bigcup_{i=1}^{m} J_i$, and finitely many positive numbers $h'_1, \ldots, h'_n$, such that

$$y_k + h'_k < y_{k+1} \quad (1 \leq k \leq n-1)\,,$$

for each k there exists i such that $(y_k, y_k + h'_k) \subset J_i$, and

$$\sum_{k=1}^{n} (f(y_k + h'_k) - f(y_k)) > r\,(\mu^*(E) - 2\varepsilon)\,.$$

Our next theorem shows, in particular, that the differentiability of the function F can be dropped from the hypotheses of Proposition (2.1.7).

(2.1.10) Theorem. *An increasing function $f : \mathbf{R} \to \mathbf{R}$ is differentiable almost everywhere.*

Proof. It suffices to show that the sets

$$S = \left\{x \in \mathbf{R} : D^+ f(x) \text{ is undefined}\right\},$$
$$T = \left\{x \in \mathbf{R} : D^+ f(x) > D_- f(x)\right\}$$

have measure zero. For, applying this and Exercise (2.1.9: 3) to the increasing function $x \mapsto -f(-x)$, we then see that $D^- f(x) \leq D_+ f(x)$ almost everywhere; whence, by Exercises (2.1.9: 1) and (2.1.1: 9),

$$D^+ f(x) \leq D_- f(x) \leq D^- f(x) \leq D_+ f(x) \leq D^+ f(x) \in \mathbf{R}$$

almost everywhere. (Note that as f is increasing, $D_+ f(x)$ and $D_- f(x)$ are everywhere defined and nonnegative.) Thus the four Dini derivates of f are equal almost everywhere. Reference to Exercise (2.1.9: 2) then completes the proof.

Leaving S to the next set of exercises, we now show that T has measure zero. Since T is the union of a countable family of sets of the form

$$E = \left\{x \in (a, b) : D^+ f(x) > r > s > D_- f(x)\right\},$$

where $a < b$ and r, s are rational numbers with $r > s$, it is enough to prove that such a set E has measure zero. We first use Exercise (2.1.9: 5) to obtain

(i) finitely many points $x_1, \ldots, x_m$ of (a,b), and finitely many positive numbers $h_1, \ldots, h_m$, such that the intervals $J_i = (x_i - h_i, x_i)$ $(1 \leq i \leq m)$ form a pairwise–disjoint collection,

$$\mu^*\left(\bigcup_{i=1}^{m} J_i\right) > \mu^*(E) - \varepsilon,$$

and

$$\sum_{i=1}^{m} (f(x_i) - f(x_i - h_i)) < s\,(\mu^*(E) + \varepsilon)\,;$$

(ii) finitely many points $y_1, \ldots, y_n$ of $E \cap \bigcup_{i=1}^{m} J_i$, and finitely many positive numbers $h_1', \ldots, h_n'$, such that

$$y_k + h_k' < y_{k+1} \quad (1 \leq k \leq n-1)\,, \tag{4}$$

for each k there exists i with $(y_k, y_k + h_k') \subset J_i$, and

$$\sum_{k=1}^{n} (f(y_k + h_k') - f(y_k)) > r\,(\mu^*(E) - 2\varepsilon)\,.$$

For each i with $1 \leq i \leq m$ let

$$S_i = \{k : (y_k, y_k + h_k') \subset J_i\}\,.$$

Since f is increasing, it follows from (4) that

$$\sum_{k \in S_i} (f(y_k + h_k') - f(y_k)) \leq f(x_i) - f(x_i - h_i).$$

Thus, as the intervals J_i are disjoint,

$$\sum_{i=1}^{m} (f(x_i) - f(x_i - h_i)) \geq \sum_{k=1}^{n} (f(y_k + h_k') - f(y_k))\,,$$

so that

$$s\,(\mu^*(E) + \varepsilon) > r\,(\mu^*(E) - 2\varepsilon)\,.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $s\mu^*(E) \geq r\mu^*(E)$. But $r > s$, so we must have $\mu^*(E) = 0$. □

We make good use of the following consequence of Theorem (2.1.10).

(2.1.11) Fubini's Series Theorem. *Let (F_n) be a sequence of increasing continuous functions on $\mathbf{R}$ such that $F(x) = \sum_{n=1}^{\infty} F_n(x)$ converges for all $x \in \mathbf{R}$. Then almost everywhere, F is differentiable, $\sum_{n=1}^{\infty} F_n'(x)$ converges, and $F'(x) = \sum_{n=1}^{\infty} F_n'(x)$.*

Proof. Fix real numbers a, b with $a < b$. It suffices to prove that $F'(x) = \sum_{n=1}^{\infty} F_n'(x)$ almost everywhere on $I = [a, b]$: for then we can apply the result to the intervals $[-n, n]$ as n increases through $\mathbf{N}^+$. If necessary replacing F_n by $F_n - F_n(a)$, we may assume that $F_n(a) = 0$. Write

$$s_n(x) = F_1(x) + \cdots + F_n(x) \quad (x \in I)$$

and note that $F - s_n = \sum_{k=n+1}^{\infty} F_k$ is increasing and nonnegative. By Theorem (2.1.10), s_n is differentiable on $I \backslash A_n$ for some set A_n of measure zero; likewise, F (which is clearly increasing) is differentiable on $I \backslash A_0$ for some set A_0 of measure zero. Then

$$A = \bigcup_{n=0}^{\infty} A_n$$

has measure zero, by Exercise (2.1.1: 9). Since both $F - s_{n+1}$ and $s_{n+1} - s_n$ are increasing functions, for each $x \in I \backslash A$ we have

$$s_n'(x) \leq s_{n+1}'(x) \leq F'(x). \tag{5}$$

It follows from the monotone sequence principle that $\sum_{n=1}^{\infty} F_n'(x)$ converges to a sum $\leq F'(x)$.

Now choose an increasing sequence $(n_k)_{k=1}^{\infty}$ of positive integers such that for each k,

$$0 \leq F(b) - s_{n_k}(b) \leq 2^{-k}.$$

Since $F - s_{n_k}$ is an increasing function, for each $x \in I$ we obtain the inequalities

$$0 \leq F(x) - s_{n_k}(x) \leq 2^{-k}.$$

Hence $\sum_{k=1}^{\infty} (F(x) - s_{n_k}(x))$ converges, by comparison with $\sum_{k=1}^{\infty} 2^{-k}$. Applying the first part of the proof with F_k replaced by $F - s_{n_k}$, we now see that, almost everywhere on I, $\sum_{k=1}^{\infty} \left(F'(x) - s_{n_k}'(x)\right)$ converges and therefore

$$\lim_{k \to \infty} \left(F'(x) - s_{n_k}'(x)\right) = 0.$$

It follows from (5) that

$$F'(x) = \lim_{n \to \infty} s_n(x) = \sum_{n=1}^{\infty} F_n'(x)$$

almost everywhere on I. □

(2.1.12) Exercises

.1 Let f be an increasing function on $[a, b]$, and for each positive integer n define

$$S_n = \left\{x \in (a, b) : D^+ f(x) > n\right\}.$$

Prove that

$$\mu^*(I \backslash S_n) < n^{-1}(f(b) - f(a))$$

and hence that the set of those $x \in (a,b)$ at which $D^+f(x)$ is undefined has measure zero. (Use the Vitali Covering Theorem to show that there exist finitely many points $x_1, x_2, \ldots, x_m$ of (a,b), and positive numbers $h_1, h_2, \ldots, h_m$, such that $x_k + h_k < x_{k+1}$ and $f(x_k + h_k) - f(x_k) > nh_k$.)

.2 Let E be a bounded subset of $\mathbf{R}$ that has measure zero, and let a be a lower bound for E. For each positive integer n choose a bounded open set $A_n \supset E$ such that $\mu^*(A_n) < 2^{-n}$ (this is possible by Exercise (2.1.3: 3)), and define

$$f_n(x) = \begin{cases} 0 & \text{if } x < a \\ \mu^*(A_n \cap [a,x]) & \text{if } x \geq a. \end{cases}$$

Show that

(i) $f = \sum_{n=1}^{\infty} f_n$ is an increasing continuous function on $\mathbf{R}$;

(ii) $D^+f(x)$ is undefined for each $x \in E$.

.3 Prove that if f has bounded variation on $[a,b]$, then it is differentiable almost everywhere on $[a,b]$. (The converse is not true: see Exercise (1.5.15: 4).)

.4 Let f have bounded variation on $[a,b]$. Prove that $T_f'(a,x) = |f'(x)|$ almost everywhere on $[a,b]$. (Using Lemma (1.5.17), construct a sequence (g_n) of functions on I such that for each n, $T_f(a,\cdot) - g_n$ is increasing, $0 \leq T_f(a,\cdot) - g_n \leq 2^{-n}$, and $g_n' = \pm f'$ almost everywhere. Then use Fubini's Series Theorem.)

.5 Prove that if a bounded function f is continuous almost everywhere on a compact interval I, then it is Riemann integrable. (Let M be a bound for $|f|$ on I, let $E \subset I = [a,b]$ be a set of measure zero such that f is continuous on $X = I \backslash E$, and let $\varepsilon > 0$. We may assume that $a, b \in E$. For each $x \in X$ there exist arbitrarily small $r > 0$ such that $[x, x+r] \subset I$ and

$$|f(x') - f(x'')| < \frac{\varepsilon}{2(b-a)} \quad (x \leq x' \leq x'' \leq x+r).$$

The sets $[x, x+r]$ of this type form a Vitali cover of X. With the aid of the Vitali Covering Theorem, construct a partition P of I such that $U(P,f) - L(P,f) < \varepsilon$.)

.6 Prove the converse of the last exercise—namely, if a bounded function $f : [a, b] \to \mathbf{R}$ is Riemann integrable, then it is continuous almost everywhere on $[a, b]$. (For each positive integer n define

$$A_n = \left\{ x \in [a, b] : \omega(f, x) > \tfrac{1}{n} \right\},$$

where $\omega(f, x)$ is the oscillation of f at x; see Exercise (1.4.5: 7). Given $\varepsilon > 0$, choose a partition P of $[a, b]$ such that $U(f, P) - L(f, P) < \varepsilon/2n$. Use this to construct a finite set of intervals that cover A_n and have total length less than ε.)

Re–examine Exercise (1.5.10: 6) in the light of this result.

2.2 The Lebesgue Integral as an Antiderivative

In this section we show how Theorem (2.1.10) and Fubini's Series Theorem (2.1.11) can be used to introduce the Lebesgue integral, a very powerful extension of the Riemann integral, as an antiderivative. Our approach[3] is based on a little–known development by F. Riesz [39].

Let f be a nonnegative real–valued function defined almost everywhere on $\mathbf{R}$. A function $F : \mathbf{R} \to \mathbf{R}$ is called a *Lebesgue primitive* of f if it is increasing, bounded below, and satisfies $F' = f$ almost everywhere.

In order to discuss Lebesgue primitives, we first consider the set $\mathcal{P}_f$ of functions $F : \mathbf{R} \to \mathbf{R}$ that are increasing, bounded below, and satisfy $F' \geq f$ almost everywhere. Note that for such a function,

$$F(-\infty) = \lim_{x \to -\infty} F(x)$$

exists: indeed, the sequence $(F(-n))_{n=1}^{\infty}$, which is decreasing and bounded below, converges to a limit which is easily shown to be $F(-\infty)$.

(2.2.1) Proposition. *If $\mathcal{P}_f$ is nonempty, then there exists an element $F_* \in \mathcal{P}_f$, called an* extremal element *of $\mathcal{P}_f$, such that*

$$F_*(\eta) - F_*(\xi) \leq F(\eta) - F(\xi) \tag{1}$$

whenever $\xi < \eta$ and $F \in \mathcal{P}_f$.

Proof. First note that the set

$$\mathcal{P}_f^0 = \{F \in \mathcal{P}_f : F(-\infty) = 0\}$$

---

[3] It is worth comparing this with the development of the Cauchy integral in [13].

is nonempty: for if $F \in \mathcal{P}_f$, then $F - F(-\infty) \in \mathcal{P}_f^0$. It is now a straightforward exercise to show that

$$F_*(x) = \inf \left\{ F(x) : F \in \mathcal{P}_f^0 \right\}$$

defines an increasing function $F_* : \mathbf{R} \to \mathbf{R}^{0+}$ with $F_*(-\infty) = 0$. Given $\varepsilon > 0$ and real numbers ξ, η with $\xi < \eta$, choose $F_1 \in \mathcal{P}_f^0$ such that $F_1(\xi) < F_*(\xi) + \varepsilon$, and consider any element F of $\mathcal{P}_f$. The function F_2 defined by

$$F_2(x) = \begin{cases} F_1(x) & \text{if } x \leq \xi \\ \\ F(x) + F_1(\xi) - F(\xi) & \text{if } \xi \leq x \end{cases}$$

belongs to $\mathcal{P}_f^0$, so

$$\begin{aligned} F_*(\eta) &\leq F_2(\eta) \\ &= F(\eta) + F_1(\xi) - F(\xi) \\ &< F(\eta) + F_*(\xi) + \varepsilon - F(\xi), \end{aligned}$$

and therefore

$$F_*(\eta) - F_*(\xi) < F(\eta) - F(\xi) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, inequality (1) follows.

Now let N be a positive integer, and choose a sequence (F_n) in $\mathcal{P}_f^0$ such that for each n,

$$0 \leq F_n(N) - F_*(N) \leq 2^{-n}.$$

For each $x \in [-N, N]$ we have

$$F_*(N) - F_*(x) \leq F_n(N) - F_n(x)$$

and therefore

$$0 \leq F_n(x) - F_*(x) \leq F_n(N) - F_*(N) \leq 2^{-n}.$$

Hence the series $\sum_{n=1}^{\infty} (F_n - F_*)$ of increasing functions converges at each point of $[-N, N]$, by comparison with $\sum_{n=1}^{\infty} 2^{-n}$. Fubini's Series Theorem (2.1.11) now shows that $\sum_{n=1}^{\infty} (F_n' - F_*')$ converges almost everywhere on $[-N, N]$. Hence $F_n' - F_*' \to 0$, and therefore $F_*' \geq f$, almost everywhere on $[-N, N]$. Since the union of a sequence of sets of measure zero has measure zero, it follows that $F_*' \geq f$ almost everywhere on $\mathbf{R}$. □

(2.2.2) Corollary. *Under the conditions of* Proposition (2.2.1), *F_1 is an extremal element of $\mathcal{P}_f$ if and only if $F_* - F_1$ is constant on* $\mathbf{R}$.

Proof. If F_1 is an extremal element of $\mathcal{P}_f$, and $\xi \leq \eta$, then

$$F_*(\eta) - F_*(\xi) \leq F_1(\eta) - F_1(\xi) \leq F_*(\eta) - F_*(\xi)$$

and therefore

$$F_*(\eta) - F_*(\xi) = F_1(\eta) - F_1(\xi).$$

It follows that

$$F_1(x) - F_*(x) = F_1(-\infty) - F_*(-\infty)$$

for all $x \in \mathbf{R}$. The converse is trivial. □

(2.2.3) Corollary. *Under the conditions of* Proposition (2.2.1), *if f has a Lebesgue primitive, then F_* is also a Lebesgue primitive.*

Proof. Let F be a Lebesgue primitive of f. Then $F \in \mathcal{P}_f$, so by Proposition (2.2.1),

$$F_*(\eta) - F_*(\xi) \leq F(\eta) - F(\xi) \quad (\xi \leq \eta).$$

Since a finite union of sets of measure zero has measure zero, it follows from this inequality and Theorem (2.1.10) that

$$f \leq F_*' \leq F' = f$$

almost everywhere. Hence $F_*' = f$ almost everywhere, and F_* is a Lebesgue primitive of f. □

We say that a nonnegative function f defined almost everywhere on $\mathbf{R}$ is *Lebesgue integrable* (or simply *integrable*) if there is a bounded Lebesgue primitive of f. In that case we define the *Lebesgue integral* (or simply the *integral*) of f to be

$$\int f = F_*(\infty) - F_*(-\infty),$$

where F_* is an extremal element of $\mathcal{P}_f$ and

$$F_*(\infty) = \lim_{x \to \infty} F_*(x).$$

(The existence of $F_*(\infty)$ is left as an exercise.) Corollary (2.2.2) shows that the value of the integral of f does not depend on the choice of extremal element F_* in $\mathcal{P}_f$. Note that

$$\int f = \sup_{x<y} \left(F_*(y) - F_*(x)\right).$$

We often write

$$\int f = \int f(x)\,\mathrm{d}x = \int f(t)\,\mathrm{d}t = \cdots,$$

as in elementary calculus courses.

(2.2.4) Exercises

.1 In the notation of the proof of Proposition (2.2.1), prove that F_* is an increasing function and that $F_*(-\infty) = 0$.

.2 Prove that if some element of $\mathcal{P}_f$ is bounded above and F_* is an extremal element of $\mathcal{P}_f$, then F_* is bounded above and $F_*(\infty)$ exists.

.3 Let f be an integrable nonnegative function, and F a Lebesgue primitive of f. Prove that if F is absolutely continuous on each compact interval, then it is an extremal element of $\mathcal{P}_f$. (Use Proposition (2.1.7).) Is every bounded Lebesgue primitive of f an extremal element of $\mathcal{P}_f$?

.4 Show that if $f \geq 0$ is Lebesgue integrable, then

$$\int f = \inf \left\{ \sup_{x<y} (F(y) - F(x)) : F \in \mathcal{P}_f \right\}.$$

.5 Let $f(x)$ equal a nonnegative constant c in a bounded interval I, and 0 outside I. Show that f is Lebesgue integrable, with $\int f = c\,|I|$.

.6 Let f be an integrable nonnegative function. Prove that $\int f = 0$ if and only if $f = 0$ almost everywhere.

.7 Let f be an integrable nonnegative function such that $\int f > 0$. Prove that $f(x) > 0$ on some set with positive outer measure. (Suppose that for all positive integers m and n,

$$E_{m,n} = \left\{x \in [-m, m] : f(x) > \tfrac{1}{n}\right\}$$

has measure zero, and use the preceding exercise to obtain a contradiction.)

.8 Let f and g be integrable nonnegative functions such that $f \geq g$ almost everywhere, and let F, G be extremal elements of $\mathcal{P}_f, \mathcal{P}_g$, respectively. Prove that

(i) $F - G \in \mathcal{P}_{f-g}$, and

(ii) $f - g$ is integrable.

(For (ii) note that $F' \geq g$ almost everywhere.)

.9 Let f be an integrable nonnegative function, and F a Lebesgue primitive of f. Show that if $F'(\xi) = f(\xi)$ and $s_n \leq \xi \leq s_n + 2^{-n}$ for each n, then

$$\lim_{n\to\infty} 2^{-n} \int_{s_n}^{s_n+2^{-n}} f = f(\xi).$$

(Note Exercise (1.5.1: 3).)

(2.2.5) Lemma. *Let Φ, Ψ, and $\Phi - \Psi$ be increasing functions on* $\mathbf{R}$ *such that Φ is an extremal element of $\mathcal{P}_{\Phi'}$ and Ψ is bounded. Then Ψ is an extremal element of $\mathcal{P}_{\Psi'}$.*

Proof. Note that Φ' and Ψ' are defined almost everywhere, by Proposition (2.1.10). Hence $\Psi \in \mathcal{P}_{\Psi'}$. By Proposition (2.2.1), $\mathcal{P}_{\Psi'}$ has an extremal element Ψ_*. The function

$$\Theta = \Phi - \Psi + \Psi_*$$

is increasing, bounded below, and has derivative equal to Φ' almost everywhere; so it belongs to $\mathcal{P}_{\Phi'}$. Since Φ is an extremal element of $\mathcal{P}_{\Phi'}$, it follows that $\Psi_* - \Psi = \Theta - \Phi$ is an increasing function. But by our choice of Ψ_*, $\Psi - \Psi_*$ is an increasing function. It follows that $\Psi - \Psi_*$ is constant; whence, by Corollary (2.2.2), Ψ is an extremal element of $\mathcal{P}_{\Psi'}$. □

(2.2.6) Proposition. *If f, g are integrable nonnegative functions defined almost everywhere, and $\lambda \geq 0$, then $f + g$ and λf are integrable,*

$$\int (f+g) = \int f + \int g, \tag{2}$$

and

$$\int \lambda f = \lambda \int f.$$

Proof. Let F_*, G_* be extremal elements of $\mathcal{P}_f$, $\mathcal{P}_g$, respectively. Then $F_* + G_*$ is a bounded Lebesgue primitive of $f + g$, which is therefore integrable; but there is no guarantee that $F_* + G_*$ is an extremal element of $\mathcal{P}_{f+g}$, so we have to work harder to establish the identity (2). To this end, let H_* be an extremal element of $\mathcal{P}_{f+g}$. Then $F_* + G_* - H_*$ is an increasing function. On the other hand, H_* is increasing, and $H_*' = f + g \geq g$ almost everywhere; so by our choice of G_*, $H_* - G_*$ is increasing. Applying Lemma (2.2.5) with $\Phi = H_*$ and $\Psi = H_* - G_*$, we see that $H_* - G_*$ is an extremal element of $\mathcal{P}_f$; whence, by Corollary (2.2.2), $H_* - G_* - F_*$ has a constant value c. It follows that

$$\begin{aligned}
\int (f+g) &= H_*(\infty) - H_*(-\infty) \\
&= (G_*(\infty) + F_*(\infty) + c) - (G_*(-\infty) + F_*(-\infty) + c) \\
&= (F_*(\infty) - F_*(-\infty)) + (G_*(\infty) - G_*(-\infty)) \\
&= \int f + \int g.
\end{aligned}$$

It is left as an exercise to deal with λf. □

(2.2.7) Proposition. *If (f_n) is a sequence of integrable nonnegative functions defined almost everywhere, then $f = \inf f_n$ is integrable.*

Proof. For each n choose an extremal element F_{*n} of $\mathcal{P}_{f_n}$, and note that, by Corollary (2.2.3), F_{*n} is a Lebesgue primitive of f_n. Then $F_{*n} \in \mathcal{P}_f$, so $\mathcal{P}_f$ is nonempty. By Proposition (2.2.1), there exists an extremal element F_* of $\mathcal{P}_f$, and $F_{*n} - F_*$ is increasing; so $(F_{*n} - F_*)' \geq 0$ almost everywhere. Hence, almost everywhere,

$$f_n = F'_{*n} \geq F'_* \geq f,$$

so

$$f = \inf f_n \geq F'_* \geq f,$$

and therefore $F'_* = f$. Moreover, by Exercise (2.2.4: 2), F_{*n}, and therefore F_*, is bounded; so f is integrable. □

(2.2.8) Corollary. *If f, g are integrable nonnegative functions, then so are $f \vee g$ and $f \wedge g$.*

Proof. The integrability of $f \wedge g$ is a special case of Proposition (2.2.7); that of $f \vee g$ then follows from the identity

$$f \vee g = f + g - f \wedge g,$$

Proposition (2.2.6), and Exercise (2.2.4: 8). □

We now extend the Lebesgue integral to functions of variable sign. We say that a real–valued function f defined almost everywhere on $\mathbf{R}$ is (Lebesgue) *integrable* if there exist integrable nonnegative functions f_1, f_2 such that $f = f_1 - f_2$; we then define the (Lebesgue) *integral* of f to be

$$\int f = \int f_1 - \int f_2.$$

(2.2.9) Exercises

.1 Prove that the foregoing is a good definition—in other words, that if f_1, f_2, f_3, f_4 are integrable nonnegative functions such that $f_1 - f_2 = f_3 - f_4$, then $\int f_1 - \int f_2 = \int f_3 - \int f_4$. Prove also that if a nonnegative function f has a bounded Lebesgue primitive, then it is integrable in the new sense, and its integrals in the old and new senses coincide.

.2 Show that f is integrable if and only if $f^+ = f \vee 0$ and $f^- = (-f) \vee 0$ are integrable, in which case $\int f = \int f^+ - \int f^-$. (Choose integrable nonnegative functions f_1, f_2 such that $f = f_1 - f_2$, and note that $f^+ = f_1 - f_1 \wedge f_2$.)

.3 Prove that if f, g are integrable and $\lambda \in \mathbf{R}$, then $f + g$ and λf are integrable, $\int (f + g) = \int f + \int g$, and $\int \lambda f = \lambda \int f$. (For the last part you will first need to complete the proof of Proposition (2.2.6).)

.4 Prove that if f, g are integrable functions such that $f \geq g$ almost everywhere, then $\int f \geq \int g$.

.5 Prove that if f is integrable, then so is $|f|$, and $\left|\int f\right| \leq \int |f|$.

.6 Show that if f and g are integrable, then so are $f \vee g$ and $f \wedge g$. (Reduce to the case where f and g are nonnegative.)

.7 Let $(f_n)_{n=0}^{\infty}$ be a sequence of integrable functions. Prove that

(i) if $f_n \geq f_0$ almost everywhere, then $\inf_{n\geq 1} f_n$ is integrable;

(ii) if $f_n \leq f_0$ almost everywhere, then $\sup_{n\geq 1} f_n$ is integrable.

.8 Let f be a *step function*—that is, a function, defined almost everywhere on **R**, for which there exist points

$$a = x_1 < x_2 < \cdots < x_n = b$$

and real numbers $c_1, \ldots, c_{n-1}$ such that

$$f(x) = \begin{cases} c_i & \text{if } x_i < x < x_{i+1} \\ 0 & \text{if } x < a \text{ or } x > b. \end{cases}$$

Give two proofs that f is integrable and that

$$\int f = \sum_{i=1}^{n-1} c_i(x_{i+1} - x_i).$$

.9 Let f be integrable, t a real number, and $g(x) = f(x+t)$. Prove that g is integrable, with $\int g = \int f$. (*Translation invariance* of the Lebesgue integral. First consider the case where f is nonnegative. Let F_* be an extremal element of $\mathcal{P}_f$, and define $G_*(x) = F_*(x+t)$; prove that G_* is a bounded Lebesgue primitive of g, and that $\int g \leq \int f$.)

Let A be a subset of **R.** The *characteristic function* of A is the mapping $\chi_A : \mathbf{R} \to \mathbf{R}$ defined by

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

A function f defined almost everywhere is said to be *integrable over* A if $f\chi_A$ is integrable, in which case we define

$$\int_A f = \int f\chi_A.$$

If A is a compact interval $[a,b]$, we write $\int_a^b f$ for $\int_A f$. If A is a closed infinite interval, we use the natural analogous notations; for example, if $A = [a,\infty)$, we write $\int_a^\infty f$ for $\int_A f$.

(2.2.10) Proposition. *If f is an integrable function, then f is integrable over any interval. Moreover, if f is nonnegative and F_* is an extremal element of $\mathcal{P}_f$, then*

$$\int_a^b f = F_*(b) - F_*(a)$$

whenever $a \leq b$.

Proof. We only discuss the case where f is nonnegative and the interval is of the form $I = [a,b]$ with $a \leq b$. Accordingly, we define

$$F(x) = \begin{cases} F_*(a) & \text{if } x < a \\ F_*(x) & \text{if } a \leq x \leq b \\ F_*(b) & \text{if } x > b. \end{cases}$$

Then F is a Lebesgue primitive of $f\chi_I$ and so belongs to $\mathcal{P}_{f\chi_I}$. We show that F is an extremal element of $\mathcal{P}_{f\chi_I}$. Let $G \in \mathcal{P}_{f\chi_I}$, and for each pair of real numbers α, β with $\alpha < \beta$ define

$$H_{\alpha,\beta}(x) = \begin{cases} F_*(x) + G(\alpha) - F_*(\alpha) & \text{if } x < \alpha, \\ G(x) & \text{if } \alpha \leq x \leq \beta, \\ F_*(x) + G(\beta) - F_*(\beta) & \text{if } x > \beta. \end{cases}$$

Note that if $a \leq \alpha < \beta \leq b$, then $H_{\alpha,\beta} \in \mathcal{P}_f$. Consider real numbers ξ, η with $\xi < \eta$. If $\eta < a$ or $\xi > b$, then $F(\eta) = F(\xi)$ and so

$$F(\eta) - F(\xi) \leq G(\eta) - G(\xi) \tag{3}$$

holds trivially. If $a \leq \xi < \eta \leq b$, then

$$\begin{aligned} F(\eta) - F(\xi) &= F_*(\eta) - F_*(\xi) \\ &\leq H_{\xi,\eta}(\eta) - H_{\xi,\eta}(\xi) \\ &= G(\eta) - G(\xi). \end{aligned}$$

If $\xi < a$ and $\eta > b$, then, as $H_{a,b} \in \mathcal{P}_f$ and G is increasing,

$$\begin{aligned} F(\eta) - F(\xi) &= F_*(b) - F_*(a) \\ &\leq H_{a,b}(b) - H_{a,b}(a) \\ &= G(b) - G(a) \\ &\leq G(\eta) - G(\xi). \end{aligned}$$

Hence (3) holds in all possible cases, so F is an extremal element of $\mathcal{P}_{f\chi_I}$. Since F is bounded by F_*, f is integrable over I and

$$\int_a^b f = F(\infty) - F(-\infty) = F_*(b) - F_*(a). \qquad \square$$

(2.2.11) Exercises

.1 Let f be an integrable nonnegative function, and F_* an extremal element of $\mathcal{P}_f$. Prove that for each $x \in \mathbf{R},$ f is integrable over $(-\infty, x]$ and

$$\int_{-\infty}^x f = F_*(x) - F_*(-\infty).$$

.2 Complete the proof of Proposition (2.2.10) in the remaining cases.

.3 Let f be a nonnegative integrable function such that $\int_{-\infty}^x f = 0$ for each $x \in \mathbf{R}.$ Prove that $f = 0$ almost everywhere.

.4 Find expressions for $\chi_{A\cap B}$, $\chi_{A\cup B}$, and $\chi_{A\backslash B}$ in terms of χ_A and χ_B. Prove that if f is integrable over both A and B, then it is integrable over $A \cap B, A \cup B$, and $A \backslash B$. Prove also that

(i) if A and B are disjoint, then $\int_{A\cup B} f = \int_A f + \int_B f$;
(ii) if $B \subset A$, then $\int_{A\backslash B} f = \int_A f - \int_B f$.

.5 Let f be a nonnegative integrable function, $[a, b]$ a compact interval, and m a real number such that $f(x) \geq m$ for each $x \in (a, b)$. Give two proofs that $\int_a^b f \geq m(b - a)$. (For one proof use Proposition (2.1.7).)

.6 Let f be a nonnegative integrable function, F a bounded Lebesgue primitive of f, and $[a, b]$ a compact interval. Must we have $\int_a^b f = F(b) - F(a)$?

The power of the Lebesgue integral only appears when we consider the interplay between the operations of integration and of taking limits. There now follows a string of results and exercises that deal with this topic.

A sequence $(f_n)_{n=1}^\infty$ of real-valued functions defined almost everywhere is said to be *increasing* (respectively, *decreasing*) if $f_1 \leq f_2 \leq \cdots$ (respectively, $f_1 \geq f_2 \geq \cdots$) almost everywhere.

(2.2.12) Beppo Levi's Theorem. *Let (f_n) be an increasing sequence of integrable functions such that the corresponding sequence of integrals is bounded above. Then (f_n) converges almost everywhere to an integrable function f, and $\int f = \lim_{n\to\infty} \int f_n$.*

Proof. Replacing f_n by $f_n - f_1$ if necessary, we may assume that $f_n \geq 0$. Choose $M > 0$ such that $\int f_n \leq M$ for each n. By Exercise (2.2.11: 1) and Corollary (2.2.2),

$$F_n(x) = \int_{-\infty}^{x} f_n$$

defines an extremal element F_n of $\mathcal{P}_{f_n}$. Since

$$f_n\chi_{(-\infty,x]} \leq f_{n+1}\chi_{(-\infty,x]} \leq f_{n+1},$$

it follows from Exercise (2.2.9: 4) that $(F_n(x))_{n=1}^{\infty}$ is an increasing sequence that is bounded above by M and therefore converges to a limit $F(x) \leq M$. Since each F_n is an increasing function, so is F; whence, by Theorem (2.1.10), F is differentiable almost everywhere.

If $m > n$, then $F_m' = f_m \geq f_n$ almost everywhere, so $F_m \in \mathcal{P}_{f_n}$. Thus if $x < y$, then

$$F_n(y) - F_n(x) \leq F_m(y) - F_m(x);$$

letting $m \to \infty$, we obtain

$$F_n(y) - F_n(x) \leq F(y) - F(x).$$

It follows that $F' \geq F_n' = f_n$ almost everywhere, which ensures that, almost everywhere, the increasing sequence (f_n) converges to a limit f satisfying

$$f = \sup f_n \leq F'.$$

Since F' is integrable (F is a bounded Lebesgue primitive of F'), it follows from Exercise (2.2.9: 7) that f is integrable. Finally, by Exercises (2.2.4: 4) and (2.2.9: 4),

$$\begin{aligned} F(\infty) - F(-\infty) &\geq \int F' \geq \int f \geq \int f_n \\ &= F_n(\infty) - F_n(-\infty) \\ &\to F(\infty) - F(-\infty) \text{ as } n \to \infty, \end{aligned}$$

so

$$\int f = F(\infty) - F(-\infty) = \lim_{n\to\infty} \int f_n. \qquad \square$$

(2.2.13) Exercises

.1 Let $\alpha \in \mathbf{R}$, and define

$$f(x) = \begin{cases} x^{\alpha} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Prove that f is integrable over $[1, \infty)$ if and only if $\alpha < 1$, and that f is integrable over $[0, 1)$ if and only if $\alpha > 1$. Calculate $\int f$ in each case.

.2 Define $f(x) = \mathrm{e}^{-\alpha x}$, where α is a positive constant. Prove that f is integrable, and calculate $\int f$.

.3 Let f be an integrable function, and I a bounded interval. Use Beppo Levi's Theorem to prove that f is integrable over I. (Consider the sequence $(f \wedge g_n)_{n=1}^{\infty}$, where $g_n(x) = n$ if $x \in I$, and $g_n(x) = 0$ otherwise.) Extend this result to an unbounded interval I. (First take $f \geq 0$. Consider the sequence (f_n), where $f_n(x) = f(x)$ if $x \in I \cap [-n, n]$, and $f_n(x) = 0$ otherwise.)

.4 Prove *Lebesgue's Series Theorem*: if $\sum_{n=1}^{\infty} f_n$ is a series of integrable functions such that the series $\sum_{n=1}^{\infty} \int |f_n|$ converges, then $\sum_{n=1}^{\infty} f_n$ converges almost everywhere to an integrable function, and

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

(Consider the partial sums of the series $\sum_{n=1}^{\infty} f_n^+$ and $\sum_{n=1}^{\infty} f_n^-$.)

.5 Use the preceding exercise to give another proof that if f is a nonnegative integrable function satisfying $\int f = 0$, then $f = 0$ almost everywhere (See also Exercises (2.2.11: 3) and (2.2.4: 6).)

.6 Let (A_n) be a sequence of subsets of $\mathbf{R}$, and f a function that is integrable over each A_n, such that $\sum_{n=1}^{\infty} \int_{A_n} |f|$ converges. Prove that

(i) f is integrable over $A = \bigcup_{n=1}^{\infty} A_n$, and $\int_A |f| \leq \sum_{n=1}^{\infty} \int_{A_n} |f|$;

(ii) if also the sets A_n are pairwise–disjoint, then $\int_A f = \sum_{n=1}^{\infty} \int_{A_n} f$.

.7 Let f be an integrable function, and $\varepsilon > 0$. Show that there exists a bounded interval I such that $\int_{\mathbf{R} \setminus I} |f| < \varepsilon$. (Consider $|f| \chi_n$, where χ_n is the characteristic function of $[-n, n]$.)

.8 Prove that the series $\sum_{n=1}^{\infty} \mathrm{e}^{-n^2 x}$ converges for each $x > 0$. Define

$$f(x) = \begin{cases} \sum_{n=1}^{\infty} \mathrm{e}^{-n^2 x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Prove that f is integrable, and that $\int f = \sum_{n=1}^{\infty} 1/n^2$.

.9 Let (f_n) be a sequence of integrable functions such that $0 \leq f_1 \leq f_2 \leq \cdots$ almost everywhere. Show that $\lim_{n \to \infty} \int f_n = 0$ if and only if $\lim_{n \to \infty} f_n(x) = 0$ almost everywhere. (For "only if" choose a subsequence $(f_{n_k})_{k=1}^{\infty}$ such that $\int f_{n_k} \leq 2^{-k}$ for each k, and use Lebesgue's Series Theorem.)

.10 Let (f_n) be a sequence of step functions such that $0 \leq f_{n+1} \leq f_n$ almost everywhere and $\lim_{n\to\infty} \int f_n = 0$. Without using any of the foregoing theorems or exercises about the convergence of integrals, prove that $\lim_{n\to\infty} f_n(x) = 0$ almost everywhere.[4] (Use the Vitali Covering Theorem.)

.11 Prove *Fatou's Lemma*: if (f_n) is a sequence of nonnegative integrable functions that converges almost everywhere to a function f, and if the sequence $\left(\int f_n\right)_{n=1}^{\infty}$ is bounded above, then f is integrable and

$$\int f \leq \liminf \int f_n.$$

(Apply Beppo Levi's Theorem to the functions $g_n = \inf_{k \geq n} f_k$.)

.12 Let f be defined almost everywhere, and suppose that for each $\varepsilon > 0$ there exist integrable functions g, h such that $g \leq f \leq h$ almost everywhere and $\int (h - g) < \varepsilon$. Prove that f is integrable. (For each n choose integrable functions g_n, h_n such that $g_n \leq f \leq h_n$ almost everywhere and $\int (h_n - g_n) < 2^{-n}$.)

.13 Prove that if E is a set of measure zero, then there exists a nonnegative integrable function f such that $\int f = 0$ and $f(x) = 1$ for all $x \in E$. (For each positive integer n choose a sequence $(I_{n,k})_{k=1}^{\infty}$ of pairwise–disjoint bounded open intervals such that $E \subset A_n = \bigcup_{k=1}^{\infty} I_{n,k}$ and $\sum_{k=1}^{\infty} |I_{n,k}| < 1/n$. Let f be the characteristic function of $\bigcap_{n=1}^{\infty} A_n$.)

Let f, g be functions defined almost everywhere. We say that g *dominates* f if $|f| \leq g$ almost everywhere.

(2.2.14) Lebesgue's Dominated Convergence Theorem. *Let (f_n) be a sequence of integrable functions that converges almost everywhere to a function f, and suppose that there exists an integrable function g that dominates each f_n. Then f is integrable, and $\int f = \lim_{n\to\infty} \int f_n$.*

Proof. The functions

$$g_n = \sup_{k \geq n} f_k$$

are integrable, by Exercise (2.2.9: 7), and form a decreasing sequence converging to f almost everywhere. Noting that $\int (-g_n) \leq \int g$, we now apply

[4] This is the basic result in another approach to Lebesgue integration on **R**, which starts by defining the integral of a step function and then considers the convergence of a sequence (f_n) of step functions when the corresponding sequence of integrals is bounded above; see [40].

Beppo Levi's theorem to the sequence $(-g_n)$ to show that f is integrable and that $\int g_n \to \int f$. Replacing f_n by $-f_n$ in this argument, we see that $\int h_n \to \int f$, where

$$h_n = \inf_{k \geq n} f_k.$$

Finally, $h_n \leq f_n \leq g_n$, so

$$\int h_n \leq \int f_n \leq \int g_n,$$

and therefore $\int f_n \to \int f$. □

(2.2.15) Exercises

.1 Prove that if f is an integrable function, then $\int (f \wedge n) \to \int f$ as $n \to \infty$.

.2 Let f be an integrable function, and for each n define $f_n = (f \wedge n) \vee -n$. Prove that $\int |f - f_n| \to 0$ as $n \to \infty$.

.3 Give two proofs that if f is an integrable function, then $\int \left(|f| \wedge n^{-1}\right) \to 0$ as $n \to \infty$.

.4 Give an example of a sequence (f_n) of integrable functions such that $\lim_{n\to\infty} f_n = 0$ almost everywhere, $\lim_{n\to\infty} \int f_n = 0$, and there is no integrable function that dominates each f_n.

.5 Let (f_n) be a sequence of integrable functions converging almost everywhere to a function f, and let g be an integrable function that dominates f. Show that f is integrable, and that $\int f = \lim_{n\to\infty} \int f_n$. (Consider the functions $(f_n \wedge g) \vee -g$.)

With the help of Lebesgue's Dominated Convergence Theorem we can prove the converse of Exercise (2.2.4: 3), and thereby, for a nonnegative integrable function f, complete the characterisation of the extremal elements of $\mathcal{P}_f$ among the Lebesgue primitives of f.

(2.2.16) Proposition. *If f is a nonnegative integrable function, then each extremal element of $\mathcal{P}_f$ is absolutely continuous on each compact interval.*

Proof. Let $I = [a, b]$ be a compact interval. Given an extremal element F of $\mathcal{P}_f$, consider first the case where f is bounded above almost everywhere by some constant $M > 0$. The function $x \mapsto Mx$ is increasing and has derivative $M \geq f$ almost everywhere. It follows from Proposition (2.2.1) that $F(\eta) - F(\xi) \leq M(\eta - \xi)$ whenever $\xi \leq \eta$. So if $([a_k, b_k])_{k=1}^n$ is a finite sequence of nonoverlapping subintervals of I, then

$$\sum_{k=1}^n |F(b_k) - F(a_k)| \leq M \sum_{k=1}^n (b_k - a_k),$$

from which the absolute continuity of F readily follows.

In the general case we define

$$f_n = (f \wedge n) \vee -n$$

for each positive integer n. Given $\varepsilon > 0$, we see from Exercise (2.2.15: 2) that there exists N such that $\int |f - f_N| < \varepsilon$. Choose an extremal element F_N of $\mathcal{P}_{f_N}$. If $([a_k, b_k])_{k=1}^n$ is a finite sequence of nonoverlapping subintervals of I, then by Proposition (2.2.10),

$$\begin{aligned}
\sum_{k=1}^n |F(b_k) - F(a_k)| &= \sum_{k=1}^n \left| \int_{a_k}^{b_k} f \right| \\
&\leq \sum_{k=1}^n \left| \int_{a_k}^{b_k} f_N \right| + \sum_{k=1}^n \int_{a_k}^{b_k} |f - f_N| \\
&\leq \sum_{k=1}^n |F_N(b_k) - F_N(a_k)| + \int_a^b |f - f_N| \\
&< \sum_{k=1}^n |F_N(b_k) - F_N(a_k)| + \varepsilon.
\end{aligned}$$

Since, by the first part of the proof, F_N is absolutely continuous, it follows that F is absolutely continuous. □

(2.2.17) Exercises

.1 Let f be an integrable function, and F the function defined by

$$F(x) = \int_{-\infty}^x f.$$

Prove that F is absolutely continuous on each compact interval.

.2 Prove that if $G : \mathbf{R} \to \mathbf{R}$ is absolutely continuous on each compact interval, then there exists an integrable function g such that $G' = g$ almost everywhere. (Note that for $a \leq x$, $G(x) = T_G(a, x) - (T_G(a, x) - G(x))$.)

.3 Let f be a nonnegative continuous function on a compact interval $I = [a, b]$, and extend f to $\mathbf{R}$ by setting $f(x) = 0$ for all x outside I.

Prove that f is integrable over $[a,b]$, and that $\int_a^b f = \widehat{\int_a^b} f$, where $\widehat{\int}$ denotes the Riemann integral. (Let

$$F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \widehat{\int_a^x} f & \text{if } a \leq x \leq b \\ \widehat{\int_a^b} f & \text{if } x > b, \end{cases}$$

and show that F is absolutely continuous.)

Two fundamental techniques of calculus are changing the variable in an integral, and integration by parts. We now deal with the former, the latter being left to the next set of exercises.

(2.2.18) Proposition. *Let g be an absolutely continuous, increasing function on $I = [\alpha, \beta]$, $a = g(\alpha)$, $b = g(\beta)$, and f an integrable function on $[a,b]$. Then $(f \circ g)g'$ is integrable, and*

$$\int_a^b f = \int_\alpha^\beta (f \circ g)g'.$$

Proof. We may take $f = g = 0$ outside $[a,b]$. By considering f^+ and f^- separately, we reduce to the case where f is nonnegative. Moreover, we may assume that f is bounded: for if we have proved the proposition in the bounded case, we obtain the desired result in the general case for nonnegative f by considering $f \wedge n$, letting $n \to \infty$, and using Beppo Levi's Theorem.

Note that, by Corollary (1.4.12), g maps $[\alpha, \beta]$ onto $[a,b]$. Choose M such that $0 \leq f \leq M$, and let F be an extremal element of $\mathcal{P}_f$. Then the function

$$G = F \circ g$$

is increasing. Since $t \mapsto Mt$ belongs to $\mathcal{P}_f$, if $\alpha \leq \xi < \eta \leq \beta$, then

$$G(\eta) - G(\xi) = F(g(\eta)) - F(g(\xi)) \leq Mg(\eta) - Mg(\xi); \tag{4}$$

so the function $Mg - G$ is increasing. Since, by Exercise (2.2.4: 3), Mg is an extremal element of $\mathcal{P}_{Mg'}$, we can apply Lemma (2.2.5) with $\Phi = Mg$ and $\Psi = G$, to show that G is an extremal element of $\mathcal{P}_{G'}$; whence

$$\int_a^b f = F(b) - F(a) = G(\beta) - G(\alpha) = \int_\alpha^\beta G'.$$

It therefore remains to prove that

$$G'(t) = f(g(t))g'(t) \tag{5}$$

almost everywhere in I.

Consider the set of those $t \in (\alpha, \beta)$ for which (5) fails to hold. This may be split into five subsets, as follows:

- the set A_1 of measure zero on which $G'(t)$ does not exist;
- the set A_2 of measure zero on which $g'(t)$ does not exist (see Exercises (2.1.6: 4) and (2.1.12: 3));
- the set A_3 of measure zero on which $F'(g(t))$ does not exist;
- the set A_4 of measure zero on which $f(g(t))$ does not exist;
- the set B of those $t \in I \backslash \bigcup_{k=1}^{4} A_k$ such that $F'(g(t)) \neq f(g(t))$.

To complete the proof for bounded nonnegative f, we show that B has measure zero. If $t \in B$ and $g'(t) = 0$, then it follows from (4) that

$$|G'(t)| \leq \lim_{h \to 0} \frac{|Mg(t+h) - Mg(t)|}{|h|} = Mg'(t) = 0,$$

so (5) holds. Let

$$C = \{t \in B : g'(t) \text{ exists and is nonzero, and } F'(g(t)) \neq f(g(t))\}.$$

Since F is a Lebesgue primitive of f, $g(C)$ has measure zero; we must show that C itself has measure zero. To this end, for all positive integers m, n let $C_{m,n}$ be the set of those $t \in C$ such that if $\alpha < t_1 \leq t \leq t_2 < \beta$ and $g(t_2) - g(t_1) \leq 1/m$, then

$$g(t_2) - g(t_1) \geq \frac{t_2 - t_1}{n}.$$

Then (Exercise (2.2.19: 1)) $C = \bigcup_{m,n=1}^{\infty} C_{m,n}$, so we need only prove that for fixed m and n, $C_{m,n}$ has measure zero. Since $C_{m,n} \subset C$, $g(C_{m,n})$ has measure zero; so for each $\varepsilon > 0$ there exists a sequence $([a_k, b_k])_{k=1}^{\infty}$ of compact subintervals of (a, b) such that

(i) $b_k - a_k < 1/m$ for each k,

(ii) $g(C_{m,n}) \subset \bigcup_{k=1}^{\infty} [a_k, b_k]$, and

(iii) $\sum_{k=1}^{\infty} (b_k - a_k) < \varepsilon/n$.

Clearly, we may assume that $g(C_{m,n}) \cap (a_k, b_k)$ is nonempty for each k. Since g is continuous and increasing, it follows from the Intermediate Value Theorem that each $[a_k, b_k]$ is the image under g of a compact subinterval $[\alpha_k, \beta_k]$ of $[\alpha, \beta]$. For each k choose $t \in C_{m,n}$ with $\alpha_k \leq t \leq \beta_k$. Since

$$g(\beta_k) - g(\alpha_k) = b_k - a_k < \tfrac{1}{m},$$

the definition of $C_{m,n}$ ensures that

$$b_k - a_k \geq \frac{\beta_k - \alpha_k}{n}.$$

Thus the intervals $[\alpha_k, \beta_k]$ cover $C_{m,n}$ and have total length

$$\sum_{k=1}^{\infty}(\beta_k - \alpha_k) \leq \sum_{k=1}^{\infty} n(b_k - a_k) < \varepsilon.$$

Since ε is arbitrary, it follows that $C_{m,n}$ has measure zero. □

(2.2.19) Exercises

.1 In the notation of the proof of Proposition (2.2.18), show that $C = \bigcup_{m,n=1}^{\infty} C_{m,n}$.

.2 This exercise deals with *integration by parts.* Let f, g be integrable functions, and $I = [a, b]$ a compact interval. For $a \leq x \leq b$ define

$$F(x) = \int_a^x f, \quad G(x) = \int_a^x g.$$

Prove that the functions Fg and fG, extended to equal 0 outside I, are integrable over I and that

$$\int_a^b Fg + \int_a^b fG = F(b)G(b) - F(a)G(a).$$

The final set of exercises in this section explores further the relation between Riemann and Lebesgue integration. For this purpose, we again denote the Riemann integral by $\widehat{\int}$.

(2.2.20) Exercises

.1 Let the bounded function f be Riemann integrable over the compact interval $I = [a, b]$. Show that for each $\varepsilon > 0$ there exist step functions g, h that vanish outside I, such that $g \leq f \leq h$,

$$\int g \leq \widehat{\int_a^b} f \leq \int h,$$

and $\int (h - g) \leq \varepsilon$. Then use Exercise (2.2.13: 12) to deduce that f is Lebesgue integrable over I and that the Lebesgue and Riemann integrals of f over I are equal.

.2 Define $f : [0,1] \to \mathbf{R}$ by

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

Show that f, which we have already shown is not Riemann integrable (Exercise (1.5.10: 6)), is Lebesgue integrable over $[0,1]$, with $\int_0^1 f = 1$.

.3 Let f be a bounded nonnegative function on $\mathbf{R}$ that is Riemann integrable over each compact interval, such that the *infinite Riemann integral* $J = \lim_{n\to\infty} \widehat{\int_{-n}^{n}} f$ exists. Prove that f is Lebesgue integrable and that its Lebesgue integral equals J. Need this conclusion hold if f is allowed to take negative values?

.4 Let (f_n) be an increasing sequence of Riemann integrable functions over a compact interval $[a,b]$, such that $f(x) = \lim_{n\to\infty} f_n(x)$ defines a Riemann integrable function over $[a,b]$. Prove that

$$\widehat{\int_a^b} f = \lim_{n\to\infty} \widehat{\int_a^b} f_n.$$

2.3 Measurable Sets and Functions

A function f defined almost everywhere on $\mathbf{R}$ is said to be *measurable* if it is the limit almost everywhere of a sequence of integrable functions. Clearly, an integrable function is measurable.

(2.3.1) Proposition. *If a measurable function is dominated by an integrable function, then it is integrable.*

Proof. Let g be an integrable function dominating a measurable function f, and choose a sequence (f_n) of integrable functions converging to f almost everywhere. For each n define

$$g_n = (f_n \wedge g) \vee -g.$$

Then g_n is integrable, by Exercise (2.2.9: 6), and is dominated by g; also,

$$\lim_{n\to\infty} g_n = (f \wedge g) \vee -g = f$$

almost everywhere. It follows from Lebesgue's Dominated Convergence Theorem (2.2.14) that f is integrable. □

(2.3.2) Corollary. *A measurable function f is integrable if and only if $|f|$ is integrable.*

Proof. If $|f|$ is integrable, then, as it dominates f, we see from Proposition (2.3.1) that f is integrable. For the converse we refer to Exercise (2.2.9: 5). □

(2.3.3) Exercises

.1 Prove that if f is a measurable function and I is an interval, then $f\chi_I$ is measurable.

.2 Prove that a continuous function $f : \mathbf{R} \to \mathbf{R}$ is measurable.

.3 Let f, g be measurable functions. Prove that $f + g$, $f - g$, $f \vee g$, and $f \wedge g$ are measurable.

.4 Let (f_n) be a sequence of measurable functions that converges almost everywhere to a function f. Prove that f is measurable. (For each k define the step function g_k by

$$g_k(x) = \begin{cases} k & \text{if } -k \leq x \leq k \\ 0 & \text{otherwise.} \end{cases}$$

First prove that $(f \wedge g_k) \vee -g_k$ is integrable.)

.5 Let f be a measurable function, and p a positive number. Prove that $|f|^p$ is measurable.

.6 Give an example of a measurable function f which is not integrable even though f^2 is.

.7 Give two proofs that the product of two measurable functions is measurable. (For one proof use Exercises (2.3.3: 3 and 5).)

.8 Let the measurable function f be nonzero almost everywhere. Prove that $1/f$ is measurable. (First consider the case where $f \geq c$ almost everywhere for some positive constant c. For general $f \geq 0$ consider $f_n = 1/(f + n^{-1})$.)

.9 Let f be a measurable function, and $\varphi : \mathbf{R} \to \mathbf{R}$ a continuous function. Prove that $\varphi \circ f$ is measurable. (Reduce to the case where f vanishes outside a compact interval $[a, b]$. Then use Exercise (2.2.4: 9) to construct a sequence (f_n) of step functions that vanish outside $[a, b]$ and converge almost everywhere to f.)

.10 Let $-2 < \alpha < -1$, and define

$$f(x) = \begin{cases} x^{\alpha} \sin x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Prove that f is integrable over $(0, \infty)$.

.11 Define

$$f(x) = \begin{cases} \dfrac{\sin x}{x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Prove that f is measurable but not integrable. (For the second part suppose that f is integrable, so $|f|$ is integrable. Use the inequality

$$\int |f| \geq \sum_{n=1}^{N} \int_{2n\pi}^{2n\pi+\pi/3} f \quad (N \in \mathbf{N}^+)$$

to derive a contradiction.)

.12 Let $\alpha > 0$, and define

$$f(x) = \begin{cases} \mathrm{e}^{-x} x^{\alpha-1} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Prove that f is integrable. (Consider the functions $f\chi_{(-\infty,1]}$ and $f\chi_{(1,\infty)}$ separately.)

.13 Give two proofs of the *Riemann–Lebesgue Lemma*: if f is an integrable function, then the functions $x \mapsto f(x)\sin nx$ and $x \mapsto f(x)\cos nx$ are integrable, and

$$\lim_{n\to\infty} \int f(x)\sin nx \,\mathrm{d}x = 0,$$
$$\lim_{n\to\infty} \int f(x)\cos nx \,\mathrm{d}x = 0.$$

(One proof proceeds like this. First reduce to the case where $f \geq 0$ and f vanishes outside a compact interval $I = [-N\pi, N\pi]$ for some positive integer N. Let F be an extremal element of $\mathcal{P}_f$, and carry out integration by parts on $\int_{-N\pi}^{N\pi} F'(x)\sin nx \,\mathrm{d}x$.)

.14 Let φ, ψ, θ be nonnegative bounded integrable functions on $I = [0, c]$ such that

$$\theta(x) \leq \varphi(x) + \int_0^x \psi(t)\theta(t)\,\mathrm{d}t \quad (x \in I).$$

Prove that

$$\theta(x) \leq \varphi(x) + \int_0^x \varphi(t)\psi(t)\exp\left(\int_t^x \psi(s)\,\mathrm{d}s\right)\mathrm{d}t \quad (x \in I).$$

(Define

$$\begin{aligned} \gamma(x) &= \int_0^x \psi(t)\theta(t)\,\mathrm{d}t, \\ \lambda(x) &= \gamma(x)\exp\left(-\int_0^x \psi(t)\,\mathrm{d}t\right). \end{aligned}$$

Show that

$$\lambda'(x) \leq \varphi(x)\psi(x) \exp\left(-\int_0^x \psi(t)\,\mathrm{d}t\right)$$

almost everywhere on I, and then use Proposition (2.1.7).)

A subset A of $\mathbf{R}$ is called a *measurable set* (respectively, *integrable set*) if χ_A is a measurable (respectively, integrable) function. A measurable subset of an integrable set is integrable, by Proposition (2.3.1).

If $A \subset \mathbf{R}$ is integrable, we define its (Lebesgue) *measure* to be $\mu(A) = \int \chi_A$.

(2.3.4) Exercises

.1 Let A, B be measurable sets. Prove that $A \cup B$, $A \cap B$, and $A \backslash B$ are measurable.

.2 Let (A_n) be a sequence of pairwise–disjoint measurable sets. Prove that

(i) $\bigcup_{n=1}^{\infty} A_n$ is measurable;

(ii) if $\sum_{n=1}^{\infty} \mu(A_n)$ is convergent, then $\bigcup_{n=1}^{\infty} A_n$ is integrable, and $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

.3 Prove that any interval in $\mathbf{R}$ is measurable.

.4 Let $\mathcal{B}$ be the smallest collection of subsets of $\mathbf{R}$ that satisfies the following properties.

— Any open subset of $\mathbf{R}$ is in $\mathcal{B}$.
— If $A \in \mathcal{B}$, then $\mathbf{R} \backslash A \in \mathcal{B}$.
— The union of a sequence of elements of $\mathcal{B}$ belongs to $\mathcal{B}$.

The elements of $\mathcal{B}$ are called *Borel sets*. Prove that any Borel set is measurable.

If $\diamond$ is a binary relation on $\mathbf{R}$ and f, g are functions defined almost everywhere on $\mathbf{R}$, we define

$$[\![f \diamond g]\!] = \{x \in \mathbf{R} : f(x) \diamond g(x)\}.$$

So, for example,

$$[\![f > g]\!] = \{x \in \mathbf{R} : f(x) > g(x)\}.$$

We also use analogous notations such as

$$[\![a \leq f < b]\!] = \{x \in \mathbf{R} : a \leq f(x) < b\}.$$

Just as the measurability of a set is related to that of a corresponding (characteristic) function, so the measurability of a function is related to that of certain associated sets.

(2.3.5) Proposition. *Let f be a real–valued function defined almost everywhere. Then f is measurable if and only if $[\![f > r]\!]$ is measurable for each $r \in \mathbf{R}$.*

Proof. Suppose that f is measurable, let $r \in \mathbf{R}$, and for each positive integer n define

$$f_n = \frac{(f-r)^+}{\frac{1}{n} + (f-r)^+}.$$

Since the functions $t \mapsto t^+$ and

$$t \mapsto \frac{t}{\frac{1}{n} + t}$$

are continuous on $\mathbf{R}$ and $\mathbf{R}^{0+}$, respectively, we see from Exercises (2.3.3: 2 and 9) that f_n is measurable. But $\lim_{n\to\infty} f_n = \chi_{[\![f > r]\!]}$ almost everywhere, so $[\![f > r]\!]$ is measurable, by Exercise (2.3.3: 4).

Now assume, conversely, that $[\![f > r]\!]$ is measurable for each $r \in \mathbf{R}$. Given a positive integer n, choose real numbers

$$\ldots, r_{-2}, r_{-1}, r_0, r_1, r_2, \ldots$$

such that $0 < r_{k+1} - r_k < 2^{-n}$ for each k. Then

$$[\![r_{k-1} < f \leq r_k]\!] = [\![f > r_{k-1}]\!] \backslash [\![f > r_k]\!]$$

is measurable, by Exercise (2.3.4: 1); let χ_k denote its characteristic function. The function

$$f_n = \sum_{k=-\infty}^{\infty} r_{k-1}\chi_k$$

is measurable: for it is the limit almost everywhere of the sequence of partial sums of the series on the right–hand side, and Exercises (2.3.3: 3 and 4) apply. To each x in the domain of f there corresponds a unique k such that $r_{k-1} \leq f(x) < r_k$; then

$$0 \leq f(x) - f_k(x) < r_k - r_{k-1} < 2^{-k}.$$

Hence the sequence (f_n) converges almost everywhere to f, which is therefore measurable, again by Exercise (2.3.3: 4). □

The exercises in the next set extend the ideas used in the proof of Proposition (2.3.5). In particular, when taken together with the subsequent discussion of measurability in the sense of Carathéodory (a concept defined shortly), the second and third exercises link our approach to integration with the one originally used by Lebesgue; see [40], pages 94–96.

(2.3.6) Exercises

.1 Let f be a function defined almost everywhere on $\mathbf{R}$. Prove that the following conditions are equivalent.

(i) f is measurable.

(ii) $[\![f \geq r]\!]$ is measurable for each r.

(iii) $[\![f \leq r]\!]$ is measurable for each r.

(iv) $[\![f < r]\!]$ is measurable for each r.

(v) $[\![r \leq f < R]\!]$ is measurable whenever $r < R$.

.2 In the notation of the second part of the proof of Proposition (2.3.5), prove that if f is nonnegative and integrable, then each f_n is integrable and $\lim_{n\to\infty} \int f_n = \int f$.

.3 Let f be a nonnegative measurable function vanishing outside the interval $[a, b]$. For the purpose of this exercise, we call a sequence $(r_n)_{n=0}^{\infty}$ of real numbers *admissible* if $r_0 = 0$ and there exists $\delta > 0$ such that $r_{n+1} - r_n < \delta$ for all n; and we say that the series $\sum_{n=1}^{\infty} r_n\mu(E_n)$ *corresponds* to the admissible sequence, where E_n, whose characteristic function we denote by χ_n, is the measurable set $[\![r_{n-1} \leq f < r_n]\!]$. Suppose that this series converges. Let $(r'_n)_{n=0}^{\infty}$ be any admissible sequence for f, and let χ'_n be the characteristic function of $E'_n = [\![r'_{n-1} \leq f < r'_n]\!]$. Prove that

(i) the series $\sum_{n=1}^{\infty} r'_{n-1}\chi'_n$ and $\sum_{n=1}^{\infty} r_n\chi_n$ converge almost everywhere to integrable functions,

(ii) $\sum_{n=1}^{\infty} r'_{n-1}\chi'_n \leq f \leq \sum_{n=1}^{\infty} r_n\chi_n$ almost everywhere,

(iii) the series $\sum_{n=1}^{\infty} r'_{n-1}\mu(E'_n)$ and $\sum_{n=1}^{\infty} r_n\mu(E_n)$ converge, and

(iv) $\sum_{n=1}^{\infty} r'_{n-1}\mu(E'_n) \leq \sum_{n=1}^{\infty} r_n\mu(E_n)$.

Hence prove that if $\sum_{n=1}^{\infty} r_n\mu(E_n)$ converges for at least one admissible sequence (r_n), then f is integrable, and $\int f$ is both the infimum of the set

$$\left\{\sum_{n=1}^{\infty} r_n\mu(E_n) : (r_n) \text{ is admissible, } \forall n\,(E_n = [\![r_{n-1} \leq f < r_n]\!])\right\}$$

and the supremum of the set

$$\left\{\sum_{n=1}^{\infty} r_{n-1}\mu(E_n) : (r_n) \text{ is admissible, } \forall n\,(E_n = [\![r_{n-1} \leq f < r_n]\!])\right\}.$$

.4 By a *simple function* we mean a finite sum of functions of the form $c\chi$, where $c \in \mathbf{R}$ and χ is the characteristic function of an integrable set. Let f be a nonnegative integrable function. Show that there exists a sequence (f_n) of simple functions such that

(i) $0 \leq f_n \leq f$ for each n,

(ii) $f = \sum_{n=1}^{\infty} f_n$ almost everywhere, and

(iii) $\int f = \sum_{n=1}^{\infty} \int f_n$.

(First reduce to the case where f is nonnegative and vanishes outside a compact interval. Then use the preceding exercise to construct f_k inductively such that $\int \left(f - \sum_{n=1}^{k} f_n \right) < 2^{-k}$.)

This exercise relates our development to axiomatic measure theory, which is based on primitive notions of a "measurable subset" of a set X and the "measure" of such a set, and in which the integral is often built up in the following way. First, define a function $f : X \to \mathbf{R}$ to be measurable if $[\![f < \alpha]\!]$ is a measurable set for each $\alpha \in \mathbf{R}$. Next, define the integral of a simple function $\sum_{n=1}^{N} c_n \chi_{A_n}$, where the measurable sets A_n are pairwise–disjoint, to be $\sum_{n=1}^{N} c_n \mu(A_n)$. If f is a nonnegative measurable function, then define its integral to be the supremum of the integrals of simple functions s which satisfy $0 \leq s \leq f$ on the complement of a set whose measure is 0. For this approach to integration see, for example, [43] or [44].

There is another definition of measurability for sets, due to Carathéodory: we call a set $A \subset \mathbf{R}$ *C–measurable* if

$$\mu^*(A \cap I) + \mu^*(I \backslash A) = |I|$$

for each compact interval I. We prove two lemmas that enable us to show that this notion of measurability is equivalent to our original one.

(2.3.7) Lemma. *Let f be an integrable function. Then there exists a sequence (f_n) of step functions converging almost everywhere to f. Moreover, if f vanishes outside a compact interval I, then*

(i) *f_n can be chosen to vanish outside I; and*

(ii) *if, in addition, f is the characteristic function of an integrable set, then f_n can be taken as the characteristic function of a finite union of subintervals of I.*

Proof. Since $f = \lim_{n \to \infty} f\chi_{[-n,n]}$, it suffices to consider the case where f vanishes outside a compact interval $I = [a, b]$. For each n let

$$a = x_{n,0} < x_{n,1} < \cdots < x_{n,2^n} = b$$

be a partition of $[a,b]$ into 2^{-n} subintervals of equal length. Define a step function f_n by setting

$$f_n(x) = \begin{cases} 2^{-n} \displaystyle\int_{x_{n,j}}^{x_{n,j+1}} f & \text{if } x_{n,j} < x < x_{n,j+1} \\ 0 & \text{otherwise,} \end{cases}$$

and let

$$E = \bigcup_{n=1}^{\infty} \{x_{n,j} : 0 \leq j \leq 2^n\} \cup \{x \in \mathbf{R} : f(x) \text{ is undefined}\}.$$

Then E has measure zero, and, by Exercise (2.2.4: 9), $\lim_{n\to\infty} f_n(x) = f(x)$ for all x in E. This completes the proof of (i).

Now suppose that f is the characteristic function of an integrable set $A \subset I$, and, using the first part of the proof, construct a sequence (ϕ_n) of step functions that vanish outside I and converge almost everywhere to f. Define

$$f_n = \begin{cases} 1 & \text{if } \phi_n(x) > \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Then f_n is the characteristic function of a finite union of subintervals of I, and $f = \lim_{n\to\infty} f_n$ almost everywhere. □

(2.3.8) Lemma. *Let A be an integrable subset of a compact interval I. Then there exists a decreasing sequence (χ_n) of integrable functions converging almost everywhere to χ_A, such that each χ_n is the characteristic function of a countable union of pairwise–disjoint bounded open intervals.*

Proof. Using Lemma (2.3.7), choose a sequence (f_n) of step functions that vanish outside I and converge almost everywhere to χ_A, such that each f_n is the characteristic function of a finite union S_n of pairwise–disjoint bounded open intervals. Then χ_A is also the limit, almost everywhere, of the decreasing sequence (g_n), where

$$g_n = \sup_{k \geq n} f_k.$$

Also, g_n is the characteristic function of $\bigcup_{k=n}^{\infty} S_k$, which is a countable union of bounded open subintervals of I. We now build up a sequence $(T_k)_{k=n}^{\infty}$ of finite collections of pairwise–disjoint bounded open intervals, as follows: taking $T_n = S_n$, suppose we have constructed T_N for some $N \geq n$, and form T_{N+1} by adjoining to T_N all the intervals of the form $J \backslash \bigcup_{k=n}^{N} T_k$ with $J \in S_{N+1}$. Let χ_n be the characteristic function of $\bigcup_{k=n}^{\infty} T_k$, which is a countable union of pairwise–disjoint bounded open intervals. Then $\chi_n = g_n$ almost everywhere, so χ_n converges to χ_A almost everywhere. □

(2.3.9) Proposition. *Let A be a subset of* $\mathbf{R}$. *Then*

(i) *A is measurable if and only if it is C–measurable;*

(ii) *A is integrable if and only if it is measurable and has finite outer measure, in which case $\mu^*(A) = \mu(A)$.*

Proof. Assume, to begin with, that A is C–measurable. Given a compact interval $I = [a, b]$ and $\varepsilon > 0$, choose sequences (I_n) and (J_n) of bounded open intervals such that

$$A \cap I \subset \bigcup_{n=1}^{\infty} I_n,$$

$$I \backslash A \subset \bigcup_{n=1}^{\infty} J_n,$$

$$\sum_{n=1}^{\infty} |I_n| < \mu^*(A \cap I) + \varepsilon/2,$$

$$\sum_{n=1}^{\infty} |J_n| < \mu^*(I \backslash A) + \varepsilon/2.$$

By Lebesgue's Series Theorem (Exercise (2.2.13: 4)), the functions

$$g = \chi_I - \sum_{n=1}^{\infty} \chi_{J_n},$$

$$h = \sum_{n=1}^{\infty} \chi_{I_n}$$

are defined almost everywhere and integrable,

$$\int g = b - a - \sum_{n=1}^{\infty} |J_n|,$$

and

$$\int h = \sum_{n=1}^{\infty} |I_n|.$$

So

$$\int (h - g) < \mu^*(A \cap I) + \mu^*(I \backslash A) - (b - a) + \varepsilon = \varepsilon.$$

Since

$$g \leq \chi_{A \cap I} \leq h$$

almost everywhere and $\varepsilon > 0$ is arbitrary, we see from Exercise (2.2.13: 12) that $\chi_{A\cap I}$ is integrable. Moreover,

$$\begin{aligned} \int \chi_{A\cap I} &\geq \int g \\ &\geq b - a - \mu^*(I\backslash A) - \frac{\varepsilon}{2} \\ &= \mu^*(A\cap I) - \frac{\varepsilon}{2} \end{aligned}$$

and

$$\int \chi_{A\cap I} \leq \int h \leq \mu^*(A\cap I) + \frac{\varepsilon}{2}.$$

Again as $\varepsilon > 0$ is arbitrary, we see that $\int \chi_{A\cap I} = \mu^*(A\cap I)$.

Since χ_A is the limit of the sequence $\left(\chi_{A\cap[-n,n]}\right)_{n=1}^{\infty}$, it follows that A is measurable in our original sense. If also $\mu^*(A)$ is finite, then

$$\lim_{n\to\infty} \mu^*(A\cap[-n,n]) = \mu^*(A),$$

by Exercise (2.1.1: 10); so, applying Beppo Levi's Theorem (2.2.12), we conclude that A is integrable, with $\mu(A) = \mu^*(A)$. On the other hand, if A is integrable, then Lebesgue's Dominated Convergence Theorem (2.2.14) shows that

$$\lim_{n\to\infty} \mu^*(A\cap[-n,n]) = \int \chi_A.$$

It then follows from Exercise (2.1.1: 10) that $\mu^*(A) = \mu(A)$.

It remains to prove that measurability implies C–measurability. Accordingly, let A be measurable in our original sense, and again let $I = [a,b]$ be any compact interval. Using Lemma (2.3.8), construct a decreasing sequence (χ_n) of integrable functions converging almost everywhere to $\chi_{A\cap I}$, such that each χ_n is the characteristic function of the union of a sequence $(I_{n,k})_{k=1}^{\infty}$ of pairwise–disjoint bounded open intervals. Then $\bigcup_{k=1}^{\infty} I_{n,k}$ includes $(A\cap I)\backslash E$, where E is a (possibly empty) set of measure zero; so

$$\mu^*(A\cap I) \leq \sum_{k=1}^{\infty} |I_{n,k}| = \int \chi_n,$$

the last equality being a consequence of Beppo Levi's Theorem (2.2.12). By Lebesgue's Dominated Convergence Theorem, we now have

$$\mu(A\cap I) = \lim_{n\to\infty} \int \chi_n \geq \mu^*(A\cap I).$$

Similarly,

$$b - a - \mu(A\cap I) = \mu(I\backslash A) \geq \mu^*(I\backslash A),$$

so

$$\mu^*(A\cap I) + \mu^*(I\backslash A) \leq b - a.$$

But Exercise (2.1.1: 6) shows that $\mu^*(A \cap I) + \mu^*(I \backslash A) \geq b - a$; so

$$\mu^*(A \cap I) + \mu^*(I \backslash A) = b - a,$$

and therefore A is measurable. □

(2.3.10) Exercise

> Let I be a compact interval, and f an integrable function that vanishes outside I. Prove that there exists a sequence (f_n) of continuous functions, each vanishing outside I, such that $\int |f - f_n| \to 0$. (Reduce to the case where f is bounded. Then use Lemma (2.3.7) to reduce to the case where f is a step function.)

Are all subsets of $\mathbf{R}$ measurable? No: the Axiom of Choice (Appendix B) ensures that nonmeasurable sets exist.[5] To show this, following Zermelo, we define an equivalence relation $\sim$ on $[0, 1)$ by

$$x \sim y \text{ if and only if } x - y \in \mathbf{Q}.$$

Let $\dot{x}$ denote the equivalence class of x under this relation. By the Axiom of Choice, there exists a function ϕ on the set of these equivalence classes such that

$$\phi(\dot{x}) \in \dot{x} \quad (x \in [0, 1)).$$

Let

$$E = \{\phi(\dot{x}) : x \in [0, 1)\}.$$

Now let $r_1, r_2, \ldots$ be a one–one enumeration of $\mathbf{Q} \cap [0, 1)$, and for each n define

$$\begin{aligned}
A_n &= E \cap [0, r_n), \\
B_n &= E \cap [r_n, 1), \\
E_n^0 &= \{x \in [0, 1) : x + r_n - 1 \in A_n\}, \\
E_n^1 &= \{x \in [0, 1) : x + r_n \in B_n\}, \\
E_n &= E_n^0 \cup E_n^1.
\end{aligned}$$

We show that if $r_n < r_m$, then the sets E_m, E_n are disjoint. To this end, first note that

$$E_k = \{x \in [0, 1) : x + r_k - \lfloor x + r_k \rfloor \in E\},$$

[5]Solovay [48] has shown that there is a model of Zermelo-Fraenkel set theory, without the Axiom of Choice, in which every subset of $\mathbf{R}$ is Lebesgue measurable.

where $\lfloor x \rfloor$ denotes the integer part of the real number x. Suppose that $x \in E_m \cap E_n$, so that

$$y_m = x + r_m - \lfloor x + r_m \rfloor \in E$$

and

$$y_n = x + r_n - \lfloor x + r_n \rfloor \in E.$$

Then

$$y_m - y_n = r_m - r_n + \text{ integer}$$

is a rational number. Since E contains exactly one element from each equivalence class under $\sim$, we must have $y_m = y_n$; so $r_m - r_n$ is an integer, which is impossible as $0 \leq r_n < r_m < 1$. Hence, in fact, $E_m \cap E_n$ is empty.

Now suppose that E is measurable; then A_n, B_n are measurable and have finite measure. Since E_n^0 and E_n^1 are translates of A_n and B_n, respectively, it follows from Exercise (2.2.9: 9) that E_n is measurable, with $\mu(E_n) = \mu(E)$. But $\bigcup_{n=1}^{\infty} E_n = [0,1)$, so by Exercise (2.3.4: 2), $\sum_{n=1}^{\infty} \mu(E_n) = 1$. This is absurd, since an infinite series with all terms equal cannot converge unless all its terms are 0. Hence E is not measurable.

For more on nonmeasurable sets, see Chapter 5 of [33].

(2.3.11) Exercises

.1 Let E be a nonmeasurable subset of $\mathbf{R}$, and A a subset of E that is measurable. Prove that $\mu^*(A) = 0$.

.2 Give an example of a nonmeasurable function f such that $|f|$ is integrable.

At first sight it might appear that our approach to the Lebesgue integral cannot be generalised to multiple integrals. However, in the context of $\mathbf{R}^n$ it is relatively straightforward to develop notions of outer measure, set of measure zero, and Dini derivates (of a special, set–based kind), and it is not too hard to prove a version of the Vitali Covering Theorem and hence of Fubini's Series Theorem ([46], Chapter 4). With these at hand, as Riesz has pointed out,[6] it is indeed possible to develop the Lebesgue integral in $\mathbf{R}^n$ by

---

[6] *Il ne s'agira, dans le présent Mémoire, que les fonctions d'une seule variable et il pourrait paraître, à première vue, comme si notre méthode était façonnée entièrement sur ce cas particulier. Dans cet ordre d'idées, il convient d'observer que l'on aurait pu baser les considérations, au lieu de la dérivée au sens ordinaire, sur l'idée moins exigeante de dérivée par rapport à un réseau, comme s'en sert M. de la Vallée Poussin pour l'étude de la dérivation des fonctions d'ensemble* [53]. *Non seulement que la démonstration de l'existence presque partout de cette sorte de dérivée d'une fonction monotone est presque immédiate, mais en outre on ne rencontre aucune nouvelle difficulté quand on veut passer au cas de plusieurs variables et les considérations concernant l'intégrale s'étendent à ce cas général avec des modifications évidentes.* ([39], pages 192–193)

methods akin to those we have used for one–dimensional integration. But as there are more illuminating approaches to integration on $\mathbf{R}^n$, especially once a general theory of measures has been developed (see [44] or [43]), we do not discuss the theory of multivariate integrals in this book.

Part II

Abstract Analysis

3

Analysis in Metric Spaces

...an excellent play; well digested in the scenes,
set down with as much modesty as cunning.
HAMLET, Act 2, Scene 2

In Section 1 we abstract many of the ideas from Chapter 1 to the context of a metric space, a set in which we can measure the distance between two points. In Section 2 we discuss limits and continuity in that context. Section 3 deals with compactness, which, as a substitute for finiteness, is perhaps the single most useful concept in analysis. The next section covers connectedness and lifts the Intermediate Value Theorem into its proper context. Finally, in Section 5, we study the product of a family of metric spaces, thereby enabling us to deal with analysis in $\mathbf{R}^n$ and $\mathbf{C}^n$.

3.1 Metric and Topological Spaces

The notion of a metric space generalises the properties of $\mathbf{R}$ that are associated with the distance given by the function $(x, y) \mapsto |x - y|$. A further generalisation, which we touch on at the end of this section, is a topological space, in which, since there may be no analogue of distance, the concept of open set plays a primary role.

A *metric*, or *distance function*, on a set X is a mapping ρ of $X \times X$ into $\mathbf{R}$ such that the following properties hold for all x, y, z in X.

M1 $\rho(x, y) \geq 0$.

M2 $\rho(x,y) = 0$ if and only if $x = y$.

M3 $\rho(x,y) = \rho(y,x)$.

M4 $\rho(x,y) \leq \rho(x,z) + \rho(z,y)$ (*triangle inequality*).

A *metric space* is a pair (X, ρ) consisting of a set X and a metric ρ on X; when the identity of the metric is clear from the context, we simply refer to X itself as a metric space. We use the letter ρ to denote the metric on any metric space, except where it might be confusing to do so.

The standard example of a metric space is, of course, the real line $\mathbf{R}$ taken with the metric $(x,y) \mapsto |x-y|$. More generally, if S is a subset of $\mathbf{R}$, then the restriction of this metric to a function on $S \times S$ is a metric on S. Unless we say otherwise, whenever we consider $S \subset \mathbf{R}$ as a metric space, we assume that it carries this canonical metric.

(3.1.1) Exercises

.1 Let $x_1, \ldots, x_n$ be elements of a metric space X. Prove the *generalised triangle inequality*:

$$\rho(x_1, x_n) \leq \rho(x_1, x_2) + \rho(x_2, x_3) + \cdots + \rho(x_{n-1}, x_n).$$

.2 Let X be a set. Prove that the mapping $\rho : X \times X \to \mathbf{R}$, defined by

$$\rho(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

is a metric on X.

This metric is called the *discrete metric,* and X, taken with the discrete metric, is called a *discrete space*.

.3 Prove that each of the following mappings from $\mathbf{R}^n \times \mathbf{R}^n$ to $\mathbf{R}$ is a metric on $\mathbf{R}^n$.

(i) $(x,y) \mapsto \sum_{i=1}^n |x_i - y_i|$ (*taxicab metric*).
(ii) $(x,y) \mapsto \max\{|x_i - y_i| : 1 \leq i \leq n\}$.

Here, and in the next two exercises, $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

.4 Prove the *Cauchy–Schwarz inequality,*

$$\sum_{i=1}^n x_i y_i \leq \left(\sum_{i=1}^n x_i^2\right)^{1/2} \left(\sum_{i=1}^n y_i^2\right)^{1/2}.$$

Hence prove *Minkowski's inequality,*

$$\left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2} \leq \left(\sum_{i=1}^n x_i^2\right)^{1/2} + \left(\sum_{i=1}^n y_i^2\right)^{1/2}.$$

.5 Show that the mapping

$$(x, y) \mapsto \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

is a metric on $\mathbf{R}^n$. (Note the preceding exercise.)

This metric is known as the *Euclidean metric*, and $\mathbf{R}^n$, taken with the Euclidean metric, is known as *Euclidean n–space* or *n–dimensional Euclidean space.*

.6 Let p be a prime number. For each positive integer n define $v_p(n)$ to be the exponent of p in the prime factorisation of n. For each rational number $r = \pm m/n$, where m, n are positive integers, define

$$v_p(r) = v_p(m) - v_p(n).$$

Show that this definition does not depend on the particular representation of r as a quotient of integers, and that if r' is also rational, then

$$v_p(rr') = v_p(r) + v_p(r').$$

Finally, show that

$$\rho(x, y) = \begin{cases} p^{-v_p(x-y)} & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

defines a metric ρ —which we call the *p–adic metric*—on $\mathbf{Q}$, such that

$$\rho(x, z) \leq \max \{\rho(x, y),\ \rho(y, z)\}.$$

On any set X a metric ρ that satisfies this last property is called an *ultrametric,* and (X, ρ) is called an *ultrametric space*; clearly, X is then a metric space.

.7 Let X be a nonempty set, and denote by $\mathcal{B}(X, \mathbf{R})$ the set of all bounded mappings of X into $\mathbf{R}$. Show that

$$\rho(f, g) = \sup \{|f(x) - g(x)| : x \in X\}$$

defines a metric on $\mathcal{B}(X, \mathbf{R})$.

From now on, when we refer to $\mathcal{B}(X, \mathbf{R})$ as a metric space, it is understood that the metric is the one defined in this exercise.

.8 A *pseudometric* on a set X is a mapping $\rho : X \times X \to \mathbf{R}$ that satisfies **M1, M3, M4** and the following weakening of **M2**: if $x = y$, then $\rho(x, y) = 0$. The pair (X, ρ), or, loosely, X itself, is then called a *pseudometric space.* Prove that in that case,

$$x \sim y \text{ if and only if } \rho(x, y) = 0$$

defines an equivalence relation on X, and that

$$\rho(\overline{x}, \overline{y}) = \rho(x, y)$$

defines a metric ρ on the quotient set $X/\sim$, where $\overline{x}$ is the corresponding equivalence class of x.

In practice, we often identify $X/\sim$ with X, and thereby turn X into a metric space, by calling two elements x, y of X *equal* if $\rho(x, y) = 0$ or, equivalently, if $\overline{x} = \overline{y}$.

.9 Prove that

$$\rho(f, g) = \widehat{\int_a^b} |f - g|$$

defines a metric on the set of continuous real–valued mappings on the compact interval $[a, b]$, where $\widehat{\int_a^b}$ denotes the Riemann integral.

.10 Prove that

$$\rho(f, g) = \int |f - g|$$

defines a pseudometric on the set of Lebesgue integrable functions on **R**.

The corresponding metric space (see Exercise (3.1.1: 8)) is denoted by $L_1(\mathbf{R})$. We see from Exercise (2.2.4: 6) that two elements of $L_1(\mathbf{R})$ are equal if and only if, as functions, they are equal almost everywhere.

Let X and Y be metric spaces. A bijection f of X onto Y is called an *isometry* if

$$\rho(f(x), f(y)) = \rho(x, y)$$

for all x, y in X, in which case the inverse mapping f^{-1} is an isometry of Y onto X, and the spaces X and Y are said to be *isometric* (under f). Two isometric spaces can be regarded as indistinguishable for all practical purposes that involve only distance.

Now let X be a metric space, and Y a set in one–one correspondence with X. With any bijection f of X onto Y there is associated a natural metric ρ_Y on Y, defined by setting

$$\rho_Y(f(x), f(y)) = \rho(x, y).$$

We say that the metric ρ has been *transported* from X to Y by f. The mapping f is then an isometry from (X, ρ) onto (Y, ρ_Y).

An important example of the transport of a metric occurs in connection with the real line $\mathbf{R}$, and enables us, in Section 3.2, to discuss the convergence of sequences in a metric space as a special case of the convergence of functions. The mapping f defined on $\mathbf{R}$ by

$$f(x) = \frac{x}{1+|x|} \qquad (x \in \mathbf{R})$$

is an order–preserving bijection of $\mathbf{R}$ onto the open interval $(-1,1)$, with inverse mapping g defined by

$$g(y) = \frac{y}{1-|y|} \qquad (|y| < 1).$$

Let $\bar{\mathbf{R}}$ be obtained from $\mathbf{R}$ by adjoining two new elements $-\infty$ and ∞, called the *points at infinity*. (Note that $-\infty$ and ∞ are not real numbers, and that the real numbers are often referred to as the *finite* elements of $\bar{\mathbf{R}}$.) Extend f to a bijection of $\bar{\mathbf{R}}$ onto $[-1,1]$ by setting

$$f(-\infty) = -1 \text{ and } f(\infty) = 1.$$

Then g extends to a bijection of $[-1,1]$ onto $\bar{\mathbf{R}}$, such that the extended mapping g is the inverse of the extended mapping f. Now transport (by g) the standard metric $(s,t) \mapsto |s-t|$ from $[-1,1]$ to $\bar{\mathbf{R}}$; that is, define

$$\rho_{\bar{\mathbf{R}}}(x,y) = |f(x) - f(y)| \qquad (x, y \in \bar{\mathbf{R}}).$$

Taken with the metric $\rho_{\bar{\mathbf{R}}}$, the set $\bar{\mathbf{R}}$ becomes a metric space, called the *extended real line*. Note that $\rho_{\bar{\mathbf{R}}}$, restricted to $\mathbf{R}$, is different from the standard metric $(x,y) \mapsto |x-y|$ on $\mathbf{R}$.

We introduce the order relations $>$, $\geq$ on $\bar{\mathbf{R}}$ (and hence the opposite relations $<$, $\leq$) by setting

$$x > y \text{ if and only if } f(x) > f(y),$$
$$x \geq y \text{ if and only if } f(x) \geq f(y).$$

On $\mathbf{R}$ these relations coincide with the respective standard inequality relations.

(3.1.2) Exercises

.1 Prove that the function $\rho_{\bar{\mathbf{R}}}$ is a metric on $\bar{\mathbf{R}}$.

.2 Show that the relations $>$ and $\geq$ on $\bar{\mathbf{R}}$ have the properties that you would expect. In particular, prove that

(i) $-\infty < x < \infty$ for all $x \in \mathbf{R}$;

(ii) a nonempty subset S of $\mathbf{R}$ is bounded, and has a supremum and infimum, relative to the order $\geq$ on $\bar{\mathbf{R}}$ (where $\sup S$ and $\inf S$ may equal ∞ or $-\infty$);

(iii) when restricted to $\mathbf{R}$, the order relations $>$ and $\geq$ on $\bar{\mathbf{R}}$ coincide with the standard order relations $>$ and $\geq$.

Let (X, ρ) be a metric space, $a \in X$, and $r > 0$. We define the *open ball with centre a and radius r* to be

$$B(a, r) = \{x \in X : \rho(a, x) < r\},$$

and the *closed ball with center a and radius r* to be

$$\overline{B}(a, r) = \{x \in X : \rho(a, x) \leq r\}.$$

For example, the open and closed balls with centre a and radius r in $\mathbf{R}$ are the intervals $(a - r, a + r)$ and $[a - r, a + r]$, respectively; and the open ball with centre ∞ and radius $r \in (0, 1)$ in $\bar{\mathbf{R}}$ is $\left(r^{-1} - 1, \infty\right) \cup \{\infty\}$.

In order to define the notions of *open set, interior point, interior of a set, neighbourhood, cluster point, closure,* and *closed set* for a metric space X, in the corresponding definition for subsets of $\mathbf{R}$ we replace

— the open interval $(x - r, x + r)$ by its analogue, the open ball $B(x, r)$ in X, and

— the inequality $|x - y| < r$ by the inequality $\rho(x, y) < r$.

For example, a subset A of X is said to be *open* (in X) if for each $x \in A$ there exists $r > 0$ such that $B(x, r) \subset A$.

Propositions (1.3.2), (1.3.9), and (1.3.10), and the applicable parts of Exercises (1.3.7) and (1.3.8), carry over unchanged into the context of a metric space. When we mention those results in future, it is assumed that we are referring to their metric space versions.

(3.1.3) Exercises

.1 Prove that X itself, the empty set $\emptyset \subset X$, and the open balls in X are open sets; and that X, $\emptyset$, and the closed balls in X are closed sets.

.2 Give proofs of the metric space analogues of Proposition (1.3.2), Exercises (1.3.7: 3–8), and Exercises (1.3.8: 3–8).

.3 Prove that a subset of X is closed if and only if $X \backslash S$ is open (cf. Proposition (1.3.9)).

.4 Prove that the intersection of a family of closed sets is closed, and that the union of a finite family of closed sets is closed (cf. Proposition (1.3.10)).

.5 Suppose that ρ is an ultrametric on X (see Exercise (3.1.1:6)). Prove the following statements.

(i) If $\rho(x,y) \neq \rho(y,z)$, then $\rho(x,z) = \max\{\rho(x,y),\, \rho(y,z)\}$.

(ii) If $y \in B(x,r)$, then $B(y,r) = B(x,r)$.

(iii) Every open ball in X is a closed set.

(iv) If two open balls in X have a nonempty intersection, then one of them is a subset of the other.

Is every closed ball in X an open set? Does (iv) hold with "open ball" replaced by "ball"?

.6 Two metrics on a set are said to be *equivalent* if they give rise to the same class of open sets. Prove that the Euclidean metric is equivalent to each of the metrics in Exercise (3.1.1:3).

.7 Prove that ∞ is a cluster point of $\mathbf{R}$, considered as a subset of the metric space $\bar{\mathbf{R}}$.

If S is a subset of a metric space X, then the restriction to $S \times S$ of the metric ρ on X is a metric—also denoted by ρ—on S, and is said to be *induced* on S by ρ. The set S, taken with that induced metric, is called a (metric) *subspace* of X.

(3.1.4) Exercise

Prove that if $x \in S$ and $r > 0$, then $S \cap B(x,r)$ is the open ball, and $S \cap \overline{B}(x,r)$ is the closed ball, with centre x and radius r in the subspace S.

(3.1.5) Proposition. *Let S be a subspace of the metric space (X,ρ), and A a subset of S. Then A is open in S if and only if $A = S \cap E$ for some open set E in X; and A is closed in S if and only if $A = S \cap E$ for some closed set E in X*

Proof. We prove only the part dealing with open sets, since the other part then follows by considering complements. Accordingly, suppose that $A = S \cap E$ for some open set E in X, and let $x \in A$. Choosing $r > 0$ such that $B(x,r) \subset E$, we see that

$$x \in S \cap B(x,r) \subset S \cap E.$$

Since, by Exercise (3.1.4), $S \cap B(x,r)$ is the open ball with centre x and radius r in S, it follows that x is an interior point of $S \cap E$ in the subspace S. Hence $A = S \cap E$ is open in S.

Conversely, suppose that A is open in S. Then, by Exercise (3.1.4), for each $x \in A$ there exists $r_x > 0$ such that $S \cap B(x, r_x) \subset A$. So

$$A = \bigcup_{x \in A} (S \cap B(x, r_x)) = S \cap \bigcup_{x \in A} B(x, r_x),$$

where the set $\bigcup_{x \in A} B(x, r_x)$ is open in X, by (the metric space analogue of) Proposition (1.3.2). □

(3.1.6) Exercises

In each of these exercises S is a subspace of (X, ρ).

.1 Complete the proof of Proposition (3.1.5).

.2 Prove that the following conditions are equivalent.

(i) Every subset of S that is open in S is open in X.
(ii) S is open in X.

.3 Prove that the following conditions are equivalent.

(i) Every subset of S that is closed in S is closed in X.
(ii) S is closed in X.

.4 Let $x \in S$ and $U \subset S$. Show that U is a neighbourhood of x in S if and only if $U = S \cap V$ for some neighbourhood V of x in X.

.5 Let $x \in S$. Show that the following conditions are equivalent.

(i) Every neighbourhood of x in S is a neighbourhood of x in X.
(ii) S is a neighbourhood of x in X.

Let A and B be subsets of X. We say that A is

- *dense with respect to* B if $B \subset \overline{A}$, and
- *dense in* X, or *everywhere dense,* if $\overline{A} = X$.

The space X is called *separable* if it contains a countable dense subset. For example, $\mathbf{Q}$ and $\mathbf{R}\backslash\mathbf{Q}$ are dense in $\mathbf{R}$, by Exercises (1.1.1: 19) and (1.2.11: 5). Thus $\mathbf{R}$ is separable, as $\mathbf{Q}$ is countable.

(3.1.7) Proposition. *If A is dense with respect to B, and B is dense with respect to C, then A is dense with respect to C.*

Proof. We have $B \subset \overline{A}$ and $C \subset \overline{B}$. By Exercises (1.3.8: 7 and 3), $\overline{B} \subset \overline{(\overline{A})} = \overline{A}$; whence $C \subset \overline{A}$. □

(3.1.8) Exercises

.1 Show that A is dense in X if and only if each nonempty open set in X contains a point of A.

.2 Prove that $(-\infty, -1) \cup (-1, 1) \cup (1, \infty)$ is dense in **R**.

.3 Prove that a nonempty subspace S of a separable metric space X is separable. (Let (x_n) be a dense sequence in X. For each positive integer m consider the set $\{n : \rho(x_n, S) < 1/m\}$.)

.4 Prove that the union of a countable family of separable subspaces of X is separable. What about the union of an uncountable family of separable subspaces?

.5 A point x of a metric space is said to be *isolated* if there exists $r > 0$ such that $B(x, r) = \{x\}$. Prove that the set of isolated points of a separable metric space is either empty or countable.

.6 Prove that a nonempty family of pairwise–disjoint, nonempty open subsets of a separable metric space is countable. (Use the preceding exercise.)

If S is a nonempty subset of X and $x \in X$, then we define the *distance from x to S* to be the real number

$$\rho(x, S) = \inf\{\rho(x, s) : s \in S\}.$$

More generally, if T is also a nonempty subset of X, then we define

$$\rho(S, T) = \inf\{\rho(s, t) : s \in S,\, t \in T\}.$$

(3.1.9) Proposition. *If S is a nonempty subset of X, and x, y are two points of X, then*

$$|\rho(x, S) - \rho(y, S)| \leq \rho(x, y).$$

Proof. For each $s \in S$ we have

$$\rho(x, S) \leq \rho(x, s) \leq \rho(x, y) + \rho(y, s).$$

It follows that

$$\rho(x, S) \leq \rho(x, y) + \inf\{\rho(y, s) : s \in S\} = \rho(x, y) + \rho(y, S)$$

and therefore that

$$\rho(x, S) - \rho(y, S) \leq \rho(x, y).$$

Similarly,

$$\rho(y, S) - \rho(x, S) \leq \rho(x, y).$$

The result follows immediately. □

The *diameter* of a nonempty subset S of a metric space X is defined as

$$\operatorname{diam}(S) = \sup\{\rho(x, y) : x \in S,\ y \in S\}$$

and is either a nonnegative real number or ∞. Clearly, if $S \subset T$, then $\operatorname{diam}(S) \leq \operatorname{diam}(T)$; and $\operatorname{diam}(S) = 0$ if and only if S contains exactly one point.

A subset S of X is said to be *bounded* if its diameter is finite—that is, if $\operatorname{diam}(S) \in \mathbf{R}$.

(3.1.10) Exercises

.1 For nonempty subsets S, T of X, prove that $\rho(S, S) = 0$ and $\rho(S, T) = \rho(T, S)$.

.2 Is it true that if S, T are closed subsets of $\mathbf{R}$ such that $\rho(S, T) = 0$, then $S \cap T$ is nonempty?

.3 Prove that a nonempty subset S of X is closed if and only if $\rho(x, S) > 0$ for each $x \in X \backslash S$.

.4 Let X be an ultrametric space, and B, B' distinct open balls of radius r in X both of which are contained in a closed ball of radius r. Compute $\rho(B, B')$. (Note Exercise (3.1.3: 5).)

.5 Is it true that

$$\operatorname{diam}(B(x, r)) = \operatorname{diam}(\overline{B}(x, r)) = 2r$$

for any metric space X, $x \in X$, and $r > 0$?

.6 Prove that

(i) the union of two bounded subsets of a metric space is bounded;

(ii) the union of finitely many bounded subsets of a metric space is bounded.

Is the union of an infinite family of bounded subsets necessarily bounded?

Although the notion of a metric space is sufficiently strong to underpin a large amount of analysis, the following more general notion is needed in more advanced work.[1]

A *topological space* (X, τ) consists of a set X and a family τ of subsets of X satisfying the following conditions.

TO1 $X \in \tau$ and $\emptyset \in \tau$.

TO2 If $A_i \in \tau$ for each $i \in I$, then $\bigcup_{i \in I} A_i \in \tau$.

[1] As we do not use the notion of a topology, other than a metric one, in the remainder of this book, this part of the section can be skipped without penalty.

TO3 If $A_1 \in \mathcal{T}$ and $A_2 \in \mathcal{T}$, then $A_1 \cap A_2 \in \mathcal{T}$.

$\mathcal{T}$ is called the *topology* of the space, and the elements of $\mathcal{T}$ the *open sets* of that topology. When the topology $\mathcal{T}$ is clear from the context, we speak loosely of X as a *topological space* and of the elements of $\mathcal{T}$ as *open sets* in X.

A metric space (X, ρ) is associated with a topological space $(X, \mathcal{T})$ in the obvious way: the open sets of $\mathcal{T}$ are precisely those subsets of X that are open relative to the metric ρ. In such a case we say that the metric ρ *defines* the topology $\mathcal{T}$, and we identify the metric space (X, ρ) with the associated topological space $(X, \mathcal{T})$.

A topological space $(X, \mathcal{T})$ is said to be *metrisable* if there is a metric ρ on X that defines the topology $\mathcal{T}$. Not every topological space is metrisable. For example, if $X = \{0, 1\}$ is given the topology $\mathcal{T}$ consisting of $\emptyset$ and X itself, then every neighbourhood of 0 intersects every neighbourhood of 1; if $\mathcal{T}$ were metrisable, then the distinct points $0, 1$ of X would have disjoint neighbourhoods—namely, $B(0, \frac{1}{2})$ and $B(1, \frac{1}{2})$. For characterisations of metrisable topological spaces see [25].

Let S be a subset of a topological space X, and $x \in X$. We say that x is an *interior point* of S if there is an open set A such that $x \in A \subset S$; and we define the *interior* of S to be the set of all interior points of S. By a *neighbourhood* of x we mean a set $U \subset X$ containing x in its interior. On the other hand, x is called a *cluster point* of S if each neighbourhood of x has a nonempty intersection with S; and we define the closure of S (in X) to be the set $\bar{S}$ of all cluster points of S. A subset C of X is said to be *closed* (in X) if it equals its closure.

Propositions (1.3.2), (1.3.9), and (1.3.10), and the applicable parts of Exercises (1.3.7) and (1.3.8), all hold in the context of a topological space.

(3.1.11) Exercises

.1 Prove that the standard metric on $\mathbf{R}$, and the metric induced on $\mathbf{R}$ as a subset of the extended real line $\bar{\mathbf{R}}$, give rise to the same topology on $\mathbf{R}$.

.2 Prove the statement immediately preceding this set of exercises.

3.2 Continuity, Convergence, and Completeness

In contrast to our approach to limits in Chapter 1, in the context of a metric space we first introduce the notion of continuity. The following definition is intended to capture formally the idea that $f(x)$ is close to $f(a)$ whenever x is close to a.

Let X, Y be metric spaces, and f a mapping of X into Y. We say that f is

- *continuous at the point* $a \in X$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(a), f(x)) < \varepsilon$ whenever $x \in X$ and $\rho(a, x) < \delta$;

- *continuous on* X, or simply *continuous,* if it is continuous at each point of X.

If f is not continuous at $a \in X$, we say that f has a *discontinuity* at a, or that f is *discontinuous at* a.

(3.2.1) Exercises

.1 Prove that the *identity mapping* $i_X : X \to X$, defined on the metric space X by $i_X(x) = x$, is continuous.

.2 Prove that any constant mapping between metric spaces is continuous.

.3 A mapping $f : X \to Y$ between metric spaces is said to be *contractive* if $\rho(f(x), f(y)) < \rho(x, y)$ whenever x, y are distinct points of X. Prove that a contractive mapping is continuous.

.4 Let X be a metric space, $a \in X$, and f, g two functions from X into $\mathbf{R}$ that are continuous at a. Prove that the functions $f + g$, $f - g$, $\max\{f, g\}$, $\min\{f, g\}$, $|f|$, and fg are continuous at a. Prove also that if $g(a) \neq 0$, then f/g is defined in a neighbourhood of a and is continuous at a.

.5 Let Y be a closed subset of a metric space X, and $f : Y \to \mathbf{R}$ a bounded continuous mapping. Prove that

$$x \mapsto \inf\{f(y)\rho(x, y) : y \in Y\}$$

is continuous on $X \backslash Y$. (Note Exercise (3.1.10: 3).)

.6 Let h be a mapping of $\mathbf{R}^{0+}$ into itself such that

(i) $h(t) = 0$ if and only if $t = 0$,

(ii) $h(s + t) \leq h(s) + h(t)$ for all s, t.

Let ρ be a metric on a set X. Prove that $d = h \circ \rho$ is a metric on X, and that if h is continuous at 0, then d is equivalent to ρ (see Exercise (3.1.3: 6)). Prove, conversely, that if X contains a point that is not isolated relative to ρ (see Exercise (3.1.8: 5), and if ρ and $h \circ \rho$ are equivalent metrics, then h is continuous at 0.

Taking $h(t) = \min\{t, 1\}$ in the first part of this exercise, we obtain a bounded metric equivalent to the given metric on X.

(3.2.2) Proposition. *The following are equivalent conditions on a mapping $f : X \to Y$, where X, Y are metric spaces.*

(i) *f is continuous.*

(ii) *For each open set $A \subset Y$, $f^{-1}(A)$ is open in X.*

(iii) *For each closed set $A \subset Y$, $f^{-1}(A)$ is closed in X.*

Proof. Suppose that f is continuous, let $A \subset Y$ be an open set, and consider any a in $f^{-1}(A)$. Since $f(a) \in A$ and A is open, there exists $\varepsilon > 0$ such that $B(f(a), \varepsilon) \subset A$. Choose $\delta > 0$ such that if $\rho(a, x) < \delta$, then $\rho(f(a), f(x)) < \varepsilon$ and therefore $f(x) \in A$. Then $B(a, \delta) \subset f^{-1}(A)$. Hence $f^{-1}(A)$ is open in X, and therefore (i) implies (ii).

Since a set is open if and only if its complement is closed, it readily follows that (ii) is equivalent to (iii). Finally, assume (ii), let $a \in X$ and $\varepsilon > 0$, and set $A = B(f(a), \varepsilon) \subset Y$. Then A is open in Y, so $f^{-1}(A)$ is open in X. Since $a \in f^{-1}(A)$, there exists $\delta > 0$ such that $B(a, \delta) \subset f^{-1}(A)$; so if $\rho(a, x) < \delta$, then $f(x) \in A$ and therefore $\rho(f(a), f(x)) < \varepsilon$. Hence f is continuous at a. Since $a \in X$ is arbitrary, f is continuous on X. Thus (ii) implies (i). □

The preceding result says that a mapping between metric spaces is continuous if and only if the *inverse image* of each open set is open. But the *image* of an open set under a continuous mapping need not be open: the continuous function $x \mapsto 0$ maps each nonempty open subset of $\mathbf{R}$ onto the closed set $\{0\}$. Likewise, although the inverse image of a closed set under a continuous mapping is closed, the image of a closed set need not be: the mapping $(x, y) \mapsto x$ on the Euclidean space $\mathbf{R}^2$ takes the hyperbola $\{(x, y) : xy = 1\}$, a closed set, onto the open set $\mathbf{R} \backslash \{0\}$.

(3.2.3) Proposition. *Let X, Y, Z be metric spaces. If $f : X \to Y$ is continuous at $a \in X$, and $g : Y \to Z$ is continuous at $f(a)$, then the composite mapping $g \circ f : X \to Z$ is continuous at a. If f is continuous on X and g is continuous on Y, then $g \circ f$ is continuous on X.*

Proof. Suppose that f is continuous at a and that g is continuous at $b = f(a)$. Let $\varepsilon > 0$. The continuity of g at b ensures that there exists $\delta' > 0$ such that if $\rho(b, y) < \delta'$, then $\rho(g(b), g(y)) < \varepsilon$. In turn, as f is continuous at a, there exists $\delta > 0$ such that if $\rho(a, x) < \delta$, then $\rho(f(a), f(x)) < \delta'$. So if $\rho(a, x) < \delta$, then $\rho(b, f(x)) < \delta'$ and therefore $\rho(g(b), g(f(x))) < \varepsilon$; that is,

$$\rho\left(g \circ f(a),\, g \circ f(x)\right) < \varepsilon.$$

Hence $g \circ f$ is continuous at a.

The second conclusion of the proposition follows immediately from the first. □

Let S be a subset of a metric space X, and a a *limit point* of S—that is, a point of the closure of $S\setminus\{a\}$. Let f be a mapping of $S\setminus\{a\}$ into a metric space Y, and l a point of Y. We say that $f(x)$ has a *limit l as x tends to a in S* if the mapping $F : S \cup \{a\} \to Y$ defined by

$$F(x) = \begin{cases} f(x) & \text{if } x \in S\setminus\{a\} \\ l & \text{if } x = a \end{cases}$$

is continuous at a relative to the subspace $S \cup \{a\}$ of X. We then also use such expressions as *l is a limit of the mapping f at a with respect to S*, or *$f(x)$ converges to l as x tends to a in S*, or *$f(x)$ tends to l as $x \in S$ tends to a*. In that case we write

$$l = \lim_{x\to a,\, x\in S} f(x)$$

or

$$f(x) \to l \text{ as } x \to a,\ x \in S,$$

Note that in this definition it is not required either that $a \in S$ or that $f(x)$ be defined at $x = a$.

In the special case where $S = X$ we often write $\lim_{x\to a} f(x)$, rather than $\lim_{x\to a,\, x\in X} f(x)$.

(3.2.4) Exercises

.1 Prove that the following condition is both necessary and sufficient for $l \in Y$ to be a limit of $f(x)$ as $x \in S$ tends to a: for each $\varepsilon > 0$ there exists $\delta > 0$ such that if $x \in S$ and $0 < \rho(a,x) < \delta$, then $\rho(l, f(x)) < \varepsilon$.

.2 Prove that a mapping f has at most one limit at $a \in \overline{(S\setminus\{a\})}$ with respect to the subset S of X. (Thus we are safe in referring to "the" limit of f at a.)

.3 Let $a \in X$ be a limit point of X. Prove that $f : X \to Y$ is continuous at a if and only if $f(a) = \lim_{x\to a,\, x\in X} f(x)$.

.4 Show that if $l = \lim_{x\to a,\, x\in S} f(x)$, then for each subset A of S such that $a \in \overline{A\setminus\{a\}}$, l is the limit of f at a with respect to A.

.5 Show that if $l = \lim_{x\to a,\, x\in X} f(x)$ and the mapping $g : Y \to Z$ is continuous at l, then $g(l) = \lim_{x\to a,\, x\in X} g(f(x))$.

.6 Prove that if $l = \lim_{x\to a,\, x\in S} f(x)$, then $l \in \overline{f(S)}$.

Using the metric on the extended real line $\bar{\mathbf{R}}$ introduced in Section 1 of this chapter, we can handle the convergence of sequences in a metric space X as a special case of the convergence of functions. To this end, recall that a sequence $(x_n)_{n=1}^{\infty}$ in X is really a mapping $n \mapsto x_n$ of $\mathbf{N}^+$ into X, and note that ∞ is a limit point of $\mathbf{N}^+$ in $\bar{\mathbf{R}}$ (see Exercise (3.1.3: 7)). If the mapping $n \mapsto x_n$ has a limit l at the point $\infty \in \bar{\mathbf{R}}$ with respect to $\mathbf{N}^+$, we call l the *limit of the sequence* (x_n), we say that *the sequence* (x_n) *converges to* l *as* n *tends to* ∞, and we write

$$l = \lim_{n\to\infty} x_n$$

or

$$x_n \to l \text{ as } n \to \infty.$$

The next proposition shows, in particular, that on $\mathbf{R}$ our current notion of convergence of sequences coincides with the one introduced in Section 1.2.

(3.2.5) Proposition. *In order that* $l = \lim_{n\to\infty} x_n$, *it is necessary and sufficient that for each* $\varepsilon > 0$ *there exist a positive integer* N *such that* $\rho(l, x_n) < \varepsilon$ *whenever* $n \geq N$.

Proof. By Exercise (3.2.4: 1), in order that $a = \lim_{n\to\infty} x_n$, it is necessary and sufficient that for each $\varepsilon > 0$ there exist $\delta > 0$ such that if $n \in \mathbf{N}^+$ and $0 < \rho_{\bar{\mathbf{R}}}(\infty, n) < \delta$, then $\rho(a, x_n) < \varepsilon$. But

$$\rho_{\bar{\mathbf{R}}}(\infty, n) = (n+1)^{-1} > 0,$$

so $\rho_{\bar{\mathbf{R}}}(\infty, n) < \delta$ if and only if $n \geq N$, where N is the smallest positive integer $> \delta^{-1} - 1$. The desired conclusion now follows. □

In view of Proposition (3.2.5), we can easily adapt to the context of a metric space many of the elementary results about limits of sequences that were proved in the context of $\mathbf{R}$ in Chapter 1. We frequently do this without further comment.

(3.2.6) Proposition. *Let* S *be a subset of the metric space* X, *and* $a \in X$. *In order that* $a \in \overline{S}$, *it is necessary and sufficient that* a *be the limit of a sequence of points of* S.

Proof. To prove the necessity of the stated condition, assume that $a \in \overline{S}$. Then for each positive integer n there exists a point x_n in $S \cap B(a, n^{-1})$. Since $\rho(x_n, a) < 1/N$ whenever $n \geq N$, the sequence (x_n) converges to a.

The sufficiency part of the proposition is left as an exercise. □

(3.2.7) Proposition. *Let* (x_n) *be a sequence in* X, *and* $a \in X$. *In order that there exist a subsequence of* (x_n) *converging to* a, *it is necessary and* *sufficient that for each neighbourhood U of a, $x_n \in U$ for infinitely many values of n.*

Proof. The condition is clearly necessary. Conversely, if it is satisfied, then we can construct, inductively, a strictly increasing sequence (n_k) of positive integers such that $x_{n_k} \in B(a, k^{-1})$ for each k. Since $x_{n_j} \in B(a, k^{-1})$ whenever $j \geq k$, the subsequence (x_{n_k}) of (x_n) converges to the limit a. □

(3.2.8) Exercises

.1 Prove the sufficiency of the condition in Proposition (3.2.6).

.2 Prove that the subset A is dense in the metric space X if and only if for each $x \in X$ there exists a sequence (x_n) of points of A that converges to X.

.3 Let A be a dense subset of X, and let f, g be continuous functions from X into a metric space Y such that $f(x) = g(x)$ for all x in A. Prove that $f(x) = g(x)$ for all x in X.

.4 Let f be a mapping between metric spaces X and Y, and let $a \in X$. Prove that f is continuous at a if and only if it is *sequentially continuous* at a, in the sense that $f(x_n) \to f(a)$ whenever (x_n) is a sequence in X that converges to a.

.5 Let X be a separable metric space, and f a mapping of X into $\mathbf{R}$. For each pair of rational numbers q, q' let $X_{q,q'}$ be the set of $t \in X$ such that $\lim_{x \to t,\, x \in X} f(x)$ exists and

$$f(t) \leq q < q' \leq \lim_{x \to t,\, x \in X} f(x).$$

Show that $X_{q,q'}$ is either empty or countable. (Use Exercise (3.1.8: 5).) Hence prove that the set of points $t \in X$ such that $\lim_{x \to t,\, x \in X} f(x)$ exists and does not equal $f(t)$ is empty or countable.

A sequence (x_n) in a metric space X is called a *Cauchy sequence* if for each $\varepsilon > 0$ there exists a positive integer N such that $\rho(x_m, x_n) < \varepsilon$ whenever $m, n \geq N$. Any convergent sequence is a Cauchy sequence: for if (x_n) converges to a limit l, then, given $\varepsilon > 0$ and choosing N such that $\rho(x_n, l) < \varepsilon/2$ for all $n \geq N$, we use the triangle inequality to show that $\rho(x_m, x_n) < \varepsilon$ whenever $m, n \geq N$.

We say that X is *complete* if each Cauchy sequence in X has a limit in X. We have already seen that $\mathbf{R}$ is complete (Theorem (1.2.10)).

(3.2.9) Proposition. *A complete subspace of a metric space is closed. A closed subspace of a complete metric space is complete.*

Proof. Let S be a subspace of the metric space X. If $x \in \overline{S}$, then by Proposition (3.2.6), there exists a sequence (x_n) in S that converges to x. Being convergent, (x_n) is a Cauchy sequence in S. So if S is complete, then (x_n) converges to a limit s in S. By Exercise (3.2.4: 2), we then have $x = s$, so $x \in S$. Hence $\overline{S} = S$ —that is, S is closed in X.

Conversely, suppose that X is complete and S is closed in X. If (x_n) is a Cauchy sequence in S, then it converges to a limit x in X. By Proposition (3.2.6), $x \in \overline{S} = S$. Hence S is complete. □

(3.2.10) Exercises

.1 Prove that a sequence (x_n) in an ultrametric space is a Cauchy sequence if and only if $\lim_{n\to\infty} \rho(x_n, x_{n+1}) = 0$. Give an example to show that this is not the case in a general metric space.

.2 Show that a Cauchy sequence (x_n) in X is *bounded*, in the sense that $\{x_n : n \geq 1\}$ is a bounded subset of X.

.3 Prove that if a Cauchy sequence (x_n) has a subsequence that converges to a limit a, then $x_n \to a$ as $n \to \infty$.

.4 Prove that the interval $I = (0, 1]$ is not complete with respect to the metric ρ induced by the usual metric on $\mathbf{R}$. Define a mapping $\rho' : I \times I \to \mathbf{R}$ by

$$\rho'(x, y) = \left| \frac{1}{x} - \frac{1}{y} \right|.$$

Show that ρ' is a metric on I, that ρ and ρ' are equivalent metrics on I, and that (I, ρ') is complete.

.5 Let A and B be complete subsets of a metric space. Give at least two proofs that $A \cup B$ and $A \cap B$ are complete.

.6 Suppose that $\rho(S, T) > 0$ for any two disjoint closed subsets S, T of X. Prove that X is complete. (Suppose there exists a Cauchy sequence (x_n) that does not converge to a limit in X. First reduce to the case where $x_m \neq x_n$ whenever $m \neq n$. Then consider the sets $\{x_{2n} : n \geq 1\}$ and $\{x_{2n-1} : n \geq 1\}$.)

.7 Prove that if X is a nonempty set, then the metric space $\mathcal{B}(X, \mathbf{R})$ is complete. (See Exercise (3.1.1: 7). Given a Cauchy sequence (f_n) in $\mathcal{B}(X, \mathbf{R})$ and a positive number ε, first show that for each $x \in X$, $(f_n(x))_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbf{R}$ and therefore converges to a limit $f(x) \in \mathbf{R}$. Then prove that the function f so defined is bounded, and that (f_n) converges to f in the metric on $\mathcal{B}(X, \mathbf{R})$.)

.8 Let X be a metric space, $a \in X$, and for all $x, y \in X$ define

$$\phi_x(y) = \rho(x, y)$$

and

$$Y = \{\phi_a + f : f \in \mathcal{B}(X, \mathbf{R})\}.$$

Prove that

(i) $\phi_x \in Y$,

(ii) the equation

$$d(F, G) = \sup\{|F(x) - G(x)| : x \in X\}$$

defines a metric on Y,

(iii) $x \mapsto \phi_x$ is an isometric mapping of X into Y, and

(iv) the closure $\widehat{X}$ of $\{\phi_x : x \in X\}$ in Y is a complete metric space.

We call the metric space $(\widehat{X}, d)$ *the completion* of X. More generally, we say that a complete metric space X' is *a completion* of X if there is an isometry of X onto a dense subspace of X'; but as two completions of the same metric space X are isometric (why?), we commonly refer to any completion of X as "the" completion of X.

We now arrive at the notion of uniform continuity, a natural strengthening of continuity that, as we show in Theorem (3.3.12), turns out to be equivalent to continuity for certain very important spaces.

We say that a mapping $f : X \to Y$ between metric spaces is *uniformly continuous* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(x), f(y)) < \varepsilon$ whenever $x, y \in X$ and $\rho(x, y) < \delta$.

(3.2.11) Exercises

.1 Prove that a uniformly continuous mapping is continuous. Give an example of a continuous mapping on $(0, 1]$ that is not uniformly continuous.

.2 Let f, g be uniformly continuous mappings of X into $\mathbf{R}$. Show that $f+g$, $f-g$, and fg are uniformly continuous on X. Show that if also $\inf_{x\in X} |f(x)| > 0$, then $1/f$ is uniformly continuous on X.

.3 Let $f : X \to Y$ and $g : Y \to Z$ be uniformly continuous mappings between metric spaces. Show that $g \circ f$ is uniformly continuous on X.

.4 Let S be a nonempty subset of X. Show that the mapping $x \mapsto \rho(x, S)$ is uniformly continuous on X.

.5 Let (a_n) be a sequence in X. Prove that the function

$$x \mapsto \inf_{n \geq 1} \rho(x, a_n)$$

is uniformly continuous on X.

.6 Let α be a positive number. A mapping f between metric spaces X and Y is said to satisfy a *Lipschitz condition of order* α, or to be *Lipschitz of order* α, if

$$\rho(f(x), f(y)) \leq (\rho(x,y))^{\alpha} \quad (x, y \in X).$$

Prove that such a mapping is uniformly continuous.

.7 Prove that a mapping f between metric spaces X, Y is uniformly continuous if and only if $\rho(f(S), f(T)) = 0$ whenever $S, T \subset X$ and $\rho(S,T) = 0$.

.8 Prove that if X is not complete, then there exists a uniformly continuous mapping of X into $\mathbf{R}^+$ with infimum 0. (See Exercise (3.2.10: 8).)

.9 Prove that if X is not complete, then there exists an unbounded continuous mapping of X into $\mathbf{R}$.

.10 Suppose that every continuous mapping of X into $\mathbf{R}$ is uniformly continuous. Prove that X is complete. (Assume that X is a dense subset of its completion $\widehat{X}$, as defined in Exercise (3.2.10: 8), and that there exists a Cauchy sequence of elements of X converging to $x_\infty \in \widehat{X} \backslash X$. Consider the function $x \mapsto 1/\rho(x, x_\infty)$ on X.)

(3.2.12) Proposition. *Let D be a dense subset of a metric space X, and f a uniformly continuous mapping of D into a complete metric space Y. Then there exists a unique continuous mapping F of X into Y such that $F(x) = f(x)$ for all x in D; moreover, F is uniformly continuous on X.*

Proof. For each $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(x), f(x')) < \varepsilon$ whenever $\rho(x, x') < \delta$. Given x in X, let (x_n) be a sequence in D converging to x. Since for each $\varepsilon > 0$ there exists N such that $\rho(x_m, x_n) < \delta$, and therefore $\rho(f(x_m), f(x_n)) < \varepsilon$, whenever $m, n \geq N$, we see that $(f(x_n))$ is a Cauchy sequence in Y. As the latter space is complete, $(f(x_n))$ converges to a limit ξ in Y. Moreover, if (x'_n) is another sequence in D converging to x, then $\lim_{n\to\infty} f(x'_n) = \xi$: for, replacing (x_n) by the sequence $(x_1, x'_1, x_2, x'_2, \ldots)$ in the foregoing argument, we can show that $(f(x_1), f(x'_1), f(x_2), f(x'_2), \ldots)$ is a Cauchy sequence; since the subsequence $(f(x_n))$ converges to ξ, we conclude from Exercise (3.2.10: 3) that the sequence $(f(x_1), f(x'_1), f(x_2), f(x'_2), \ldots)$, and hence the subsequence $(f(x'_n))$, converges to ξ. Thus

$$F(x) = \xi = \lim_{n\to\infty} f(x_n)$$

is an unambiguous definition of a function F from X into Y. If $x \in D$, then $(x, x, \ldots)$ is a sequence in D converging to x, so $F(x) = f(x)$.

To prove that F is uniformly continuous, consider x, x' in X such that $\rho(x, x') < \delta$, and let (x_n) and (x'_n) be sequences in D converging to x and x', respectively. Then $(f(x_n))$ and $(f(x'_n))$ converge to $F(x)$ and $F(x')$, respectively. So for all sufficiently large n we have $\rho(F(x), f(x_n)) < \varepsilon$, $\rho(F(x'), f(x'_n)) < \varepsilon$, and $\rho(x_n, x'_n) < \delta$; whence $\rho(f(x_n), f(x'_n)) < \varepsilon$, and therefore, by the triangle inequality, $\rho(F(x), F(x')) < 3\varepsilon$. Thus F is uniformly continuous on X. Finally, the uniqueness of F is an immediate consequence of Exercise (3.2.8: 3). □

The foregoing result enables us to extend uniformly continuous functions from dense subsets to the whole space. We close this section with a famous theorem that enables us to extend continuous real–valued functions from closed subspaces to the whole space.

(3.2.13) The Tietze Extension Theorem. *Let X be a metric space, Y a closed subspace of X, and f a bounded continuous mapping of Y into* $\mathbf{R}$. *Then there exists a bounded continuous mapping $F : X \to \mathbf{R}$ such that*

(i) $F(y) = f(y)$ *for all* $y \in Y$,

(ii) $\inf_{x\in X} F(x) = \inf_{y\in Y} f(y)$, *and*

(iii) $\sup_{x\in X} F(x) = \sup_{y\in Y} f(y)$.

Proof. We may assume that f is not constant. Let h be an increasing function of the form $x \mapsto ax + b$ mapping the interval $[\inf f, \sup f]$ onto $[1, 2]$; replacing f by $h \circ f$, if necessary, we reduce to the case where $\inf f = 1$ and $\sup f = 2$. Since Y is closed, $\rho(x, Y) > 0$ for all $x \in X\backslash Y$ (Exercise (3.1.10: 3)), and so

$$F(x) = \begin{cases} f(x) & \text{if } x \in Y \\ \dfrac{\inf_{y\in Y} f(y)\rho(x,y)}{\rho(x,Y)} & \text{if } x \in X\backslash Y \end{cases}$$

defines a function $F : X \to \mathbf{R}$ that coincides with f on Y. To prove that F satisfies (ii) and (iii), we need only show that $1 \leq F(x) \leq 2$ for all $x \in X\backslash Y$. For such x and all $y \in Y$ we have

$$F(x) \leq \frac{2\rho(x,y)}{\rho(x,Y)}.$$

So, given $\varepsilon > 0$ and choosing $y \in Y$ such that

$$\rho(x,y) \leq \left(1 + \frac{\varepsilon}{2}\right)\rho(x,Y),$$

we obtain $F(x) \leq 2 + \varepsilon$. On the other hand, choosing $y' \in Y$ such that

$$1 \leq \frac{\rho(x,y')}{\rho(x,Y)} \leq \frac{f(y')\rho(x,y')}{\rho(x,Y)} < F(x) + \varepsilon,$$

we see that $F(x) > 1-\varepsilon$. As $\varepsilon > 0$ is arbitrary, it follows that $1 \leq F(x) \leq 2$.

Since f is continuous on Y°, so is F. Also, the function $x \mapsto \rho(x,Y)$ is uniformly continuous on $X\backslash Y$, by Exercise (3.2.11: 4); so, by Exercises (3.2.1: 5 and 4), F is continuous on $X\backslash Y$. It therefore remains to prove the continuity of F at any $\xi \in Y \cap \overline{X\backslash Y}$. Given $\varepsilon > 0$, choose $r > 0$ such that if $y \in Y$ and $\rho(\xi,y) < r$, then $|f(\xi) - f(y)| < \varepsilon$. It suffices to prove that if $x \in X\backslash Y$ and $\rho(x,\xi) < r/4$, then

$$F(\xi) - \varepsilon \leq F(x) \leq F(\xi) + \varepsilon. \tag{1}$$

To this end, observe that for each $y \in Y\backslash B(\xi,r)$,

$$\rho(x,y) \geq \rho(\xi,y) - \rho(x,\xi) > \tfrac{3r}{4} > 2\rho(x,\xi) \geq \rho\left(x, Y \cap B(\xi,r)\right),$$

so

$$f(y)\rho(x,y) > \tfrac{3r}{4} > f(\xi)\rho(x,\xi) \geq \inf_{\eta \in Y \cap B(\xi,r)} f(\eta)\rho(x,\eta).$$

It follows that

$$\rho(x,Y) = \rho\left(x,\ Y \cap B(\xi,r)\right) \tag{2}$$

and that

$$\inf_{y \in Y} f(y)\rho(x,y) = \inf_{y \in Y \cap B(\xi,r)} f(y)\rho(x,y). \tag{3}$$

For each $y \in Y \cap B(\xi,r)$ we have

$$f(\xi) - \varepsilon < f(y) < f(\xi) + \varepsilon$$

and therefore

$$\left(f(\xi) - \varepsilon\right)\rho(x,Y) \leq f(y)\rho(x,y) \leq \left(f(\xi) + \varepsilon\right)\rho(x,y).$$

Hence

$$\left(f(\xi) - \varepsilon\right)\rho(x,Y) \leq \inf_{y \in Y \cap B(\xi,r)} f(y)\rho(x,y) \leq \left(f(\xi) + \varepsilon\right)\rho(x, Y \cap B(\xi,r)),$$

and so, by (2) and (3),

$$\left(f(\xi) - \varepsilon\right)\rho(x,Y) \leq \inf_{y \in Y} f(y)\rho(x,y) \leq \left(f(\xi) + \varepsilon\right)\rho(x,Y).$$

Dividing through by $\rho(x,Y)$, we obtain the desired inequalities (1). □

The mapping F in Theorem (3.2.13) is called a *continuous extension of* f *to* X.

(3.2.14) Exercises

.1 Give two proofs of *Urysohn's Lemma*: if S, T are nonempty disjoint closed subspaces of a metric space X, then there exists a continuous mapping $f : X \to [0,1]$ such that $f(S) = \{0\}$ and $f(T) = \{1\}$. (For one proof, note that $\rho(x,S) + \rho(x,T) > 0$ for all $x \in X$.)

.2 Let Y be a closed subspace of a metric space X, and f a continuous mapping of Y into $\mathbf{R}$. Prove that there exists a continuous extension $F : X \to \mathbf{R}$ of f. (First apply Theorem (3.2.13) to $g \circ f$ for some suitable function g.)

.3 Suppose that for each pair S, T of nonempty disjoint closed subsets of X there exists a uniformly continuous mapping $f : X \to [0,1]$ such that $f(S) = \{0\}$ and $f(T) = \{1\}$. Prove that X is complete.

.4 Show that the following are equivalent conditions on X.

(i) Every continuous function $f : X \to \mathbf{R}$ is uniformly continuous.

(ii) $\rho(S,T) > 0$ for all nonempty disjoint closed subsets S, T of X.

(To prove that (ii) implies (i), suppose that $f : X \to \mathbf{R}$ is continuous but not uniformly continuous. Then there exist sequences $(x_n), (y_n)$ in X and a positive number α such that $\lim_{n\to\infty} \rho(x_n, y_n) = 0$ and $|f(x_n) - f(y_n)| \geq \alpha$ for all n. Consider the sets $S = \{x_n : n \geq 1\}$ and $T = \{y_n : n \geq 1\}$.)

3.3 Compactness

In the context of a metric space, the various notions associated with the word *compactness* represent different generalisations of, and approximations to, finiteness.

Let S be a subset of a metric space (X, ρ). By a *cover* of S we mean a family $\mathcal{U}$ of subsets of X such that $S \subset \bigcup \mathcal{U}$; we then say that S is *covered* by $\mathcal{U}$, and that $\mathcal{U}$ *covers* S. If also each $U \in \mathcal{U}$ is an open subset of X, we refer to $\mathcal{U}$ as an *open cover* of S. On the other hand, if $\mathcal{U}$ is a finite set, we call it a *finite cover* of S. By a *subcover* of $\mathcal{U}$ we mean a subfamily $\mathcal{F}$ of $\mathcal{U}$ that covers S.

A metric space X is called *compact,* or a *compact space*, if every open cover of X contains a finite subcover. By a *compact set* in a metric space X we mean a subset of X that is compact when considered as a metric subspace of X.

Note that we can apply our definition of compactness to a topological space X, even if the topology of X is not metrisable.

The Heine–Borel–Lebesgue Theorem (1.4.6) shows that a bounded closed interval in $\mathbf{R}$ is compact.

(3.3.1) Proposition. *A compact subset of a metric space is separable and bounded.*

Proof. Let S be a compact subset of a metric space X. We may assume that S is nonempty. For each positive integer n the family $(B(s, n^{-1}))_{s \in S}$ of open balls is an open cover of S, so there exists a finite subset F_n of S such that S is covered by the balls $B(s, n^{-1})$ with $s \in F_n$. It follows that the countable set $\bigcup_{n=1}^{\infty} F_n$ is dense in S, which is therefore separable.

Now fix $s_1 \in F_1$, and define the nonnegative number

$$R = \max\{\rho(s, s_1) : s \in F_1\}.$$

For each $x \in S$ choose $s \in F_1$ such that $\rho(x, s) < 1$; then

$$\rho(x, s_1) \leq \rho(x, s) + \rho(s, s_1) < 1 + R.$$

Hence S is bounded. $\square$

(3.3.2) Proposition. *A compact set in a metric space is closed.*

Proof. Let S be a compact subset of a metric space X. We may assume that $X \backslash S$ is nonempty. If $a \in X \backslash S$, then for each $s \in S$,

$$0 < r_s = \rho(a, s).$$

The open balls $B(s, \frac{1}{2} r_s)$, with $s \in S$, form an open cover of S, so there exists a finite subset F of S such that S is covered by the balls $B(s, \frac{1}{2} r_s)$ with $s \in F$. Define the positive number

$$r = \min\{r_s : s \in F\}.$$

For each $x \in S$ choose $s \in F$ such that $x \in B(s, \frac{1}{2} r_s)$; then

$$\begin{aligned} \rho(a, x) &\geq \rho(a, s) - \rho(x, s) \\ &\geq r_s - \tfrac{1}{2} r_s \\ &\geq \tfrac{1}{2} r. \end{aligned}$$

It follows that $B(a, \frac{1}{2} r) \subset X \backslash S$ and therefore that a is an interior point of $X \backslash S$. Since a is any point of $X \backslash S$, we conclude that $X \backslash S$ is open and therefore that S is closed. $\square$

(3.3.3) Proposition. *A compact metric space is complete.*

Proof. Let X be a compact metric space, and $\widehat{X}$ its completion (Exercise (3.2.10: 8)). By Proposition (3.3.2), X is a closed subspace of $\widehat{X}$. Since $\widehat{X}$ is complete, it follows from Proposition (3.2.9) that X is complete. □

(3.3.4) Proposition. *A closed subset of a compact metric space is compact.*

Proof. Let S be a closed subset of a compact metric space X, and let $\mathcal{U}$ be an open cover of S. By Proposition (3.1.5), for each $U \in \mathcal{U}$ there exists an open set V_U in X such that $U = S \cap V_U$. Then $X \backslash S$ and the sets V_U, with $U \in \mathcal{U}$, form an open cover of X. Since X is compact, there exist finitely many sets $U_1, \ldots, U_n$ in $\mathcal{U}$ such that

$$\{X \backslash S\} \cup \{V_{U_1}, \ldots, V_{U_n}\}$$

is an open cover of X. Clearly, $\{U_1, \ldots, U_n\}$ covers S and so is a finite subcover of $\mathcal{U}$; whence S is compact. □

(3.3.5) Exercises

.1 Find an alternative proof of Proposition (3.3.2).

.2 Find an alternative proof of Proposition (3.3.3). (Suppose that X is compact but not complete, and let (x_n) be a Cauchy sequence in X that does not converge and therefore has no convergent subsequence. Then for each $x \in X$ there exist $r_x > 0$ and $N_x \in \mathbf{N}^+$ such that $\rho(x_n, x) > r_x$ for all $n \geq N_x$. Cover X by finitely many of the balls $B(x, \frac{1}{2}r_x)$.)

.3 Prove that a subset of the Euclidean space $\mathbf{R}^n$ is compact if and only if it is bounded and closed.

.4 A family $\mathcal{F}$ of subsets of a set X is said to have the *finite intersection property* if every finite subfamily of $\mathcal{F}$ has a nonempty intersection. Prove that a metric space X is compact if and only if every family of closed subsets of X with the finite intersection property has a nonempty intersection.

.5 Let K be a compact subset of an open set $U \subset X$. Prove that there exists $r > 0$ such that if $\rho(x, K) \leq r$, then $x \in U$.

.6 Prove that any open cover of a separable metric space has a countable subcover. (This is a special case of *Lindelöf's Theorem*; see page 72 of [47].)

(3.3.6) Proposition. *If f is a continuous mapping of a compact metric space X into a metric space Y, then $f(X)$ is a compact set.*

Proof. Let $\mathcal{U}$ be an open cover of $f(X)$. By Proposition (3.2.2), the family $\left(f^{-1}(U)\right)_{U\in\mathcal{U}}$ is an open cover of X. Since X is compact, there is a finite set $\mathcal{F}\subset\mathcal{U}$ such that $\left(f^{-1}(U)\right)_{U\in\mathcal{F}}$ is an open cover of X. Then $\mathcal{F}$ is an open cover of $f(X)$, which is therefore compact. □

(3.3.7) Exercises

.1 Prove that a continuous mapping f of a compact metric space X into $\mathbf{R}$ is bounded. Prove also that f *attains its bounds*, in the sense that there exist points a,b in X such that $f(a)=\inf f$ and $f(b)=\sup f$.

.2 Prove that a continuous mapping of a compact space X into $\mathbf{R}^+$ has a positive infimum.

.3 Prove that if f is a continuous one–one mapping of a compact metric space X onto a metric space Y, then the inverse mapping $f^{-1}:Y\rightarrow X$ is continuous. (Use Proposition (3.2.2).)

.4 A mapping f of a set X into itself is called a *self–map* of X. By a *fixed point* of such a mapping we mean a point $x\in X$ such that $f(x)=x$. Let f be a contractive self–map of a compact metric space X (see Exercise (3.2.1: 3)). Prove that the mapping $x\mapsto\rho(x,f(x))$ of X into $\mathbf{R}$ is continuous. Applying Exercise (3.3.7: 1) to this mapping, deduce that f has a fixed point (*Edelstein's Theorem*). Prove that there is no other fixed point of f.

There are other properties of a metric space X that capture the idea of approximate finiteness and are intimately related to compactness. We say that X is

- *sequentially compact* if every sequence in X has a convergent subsequence;
- *totally bounded*, or *precompact*, if for each $\varepsilon>0$ there exists a finite cover of X by subsets of diameter $<\varepsilon$.

Sequential compactness, like compactness, is a topological concept, whereas total boundedness is a metric notion. An analogue of sequential compactness can be defined for a general topological space; see under "filters" in [7] or [47]. For a nonmetric analogue of total boundedness we need the context of a uniform space, which is also discussed in [7] and [47].

The total boundedness of a metric space X can be expressed differently. By an *ε–approximation* to X we mean a subset S of X such that $\rho(x,S)<\varepsilon$ for each $x\in X$. It is easy to show that X is totally bounded if and only if for each $\varepsilon>0$ it contains a finite ε–approximation. Note that since the empty set is regarded as finite, it is also totally bounded.

Corollary (1.2.8) shows that a bounded closed subset of **R** is sequentially compact.

(3.3.8) Exercises

.1 Prove that a bounded interval in **R** is totally bounded.

.2 Prove that a subset of a totally bounded metric space is totally bounded.

.3 Prove that if a metric space is either sequentially compact or totally bounded, then it is bounded.

.4 Show that a totally bounded metric space is separable.

.5 Let f be a uniformly continuous mapping of a totally bounded metric space into a metric space. Prove that the range of f is totally bounded.

.6 Let X be a metric space that is not totally bounded. Prove that there exist a sequence (x_n) in X and a positive number α such that $\rho(x_m, x_n) \geq \alpha$ whenever $m \neq n$.

.7 Let f be a function of bounded variation on a compact interval $I \subset \mathbf{R}$. Prove that $f(I)$ is totally bounded. (Use the preceding exercise.)

.8 Let X be a metric space that is not totally bounded, and choose (x_n) and α as in Exercise (3.3.8: 6). For each n construct a uniformly continuous function $\phi_n : X \to [0,1]$ such that (i) $\phi_n(x_n) = 1$ and (ii) $\phi_n(x) = 0$ if $\rho(x, x_n) \geq \alpha/3$. Given any sequence (c_n) of real numbers, show that $f = \sum_{n=1}^{\infty} c_n \phi_n$ is a well–defined continuous function on X, and that if (c_n) is bounded, then f is uniformly continuous on X.

.9 Let (X, ρ) be a separable metric space. Show that there exists on X a metric d equivalent to ρ, such that (X, d) is totally bounded. (Let (x_n) be a dense sequence in X, and use Exercise (3.2.1: 6) to reduce to the case where $\rho < 1$. Define

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \left| \rho(x, x_n) - \rho(y, x_n) \right|$$

for all $x, y \in X$.)

We now arrive at a fundamental theorem linking compactness, sequential compactness, and total boundedness.

(3.3.9) Theorem. *The following are equivalent conditions on a metric space* (X, ρ).

(i) *X is compact.*

(ii) *X is sequentially compact.*

(iii) *X is totally bounded and complete.*

Proof. First, let X be a compact metric space, and (x_n) a sequence in X. For each n let F_n be the closure of $\{x_n, x_{n+1}, x_{n+2}, \ldots\}$ in X. It is easy to show that $(F_n)_{n=1}^{\infty}$ has the finite intersection property. By Exercise (3.3.5: 4), $\bigcap_{n=1}^{\infty} F_n$ contains a point a. Consider any neighbourhood U of a. For each n, since $a \in F_n$, there exists $m \geq n$ such that $x_m \in U$. It follows that U contains x_k for infinitely many values of k; whence, by Proposition (3.2.7), there exists a subsequence of (x_n) converging to a. Thus (i) implies (ii).

Next, let X satisfy (ii). Then any Cauchy sequence in X has a convergent subsequence and so converges to a limit in X, by Exercise (3.2.10: 3); whence X is complete. Suppose that X is not totally bounded. Then, by Exercise (3.3.8: 6), there exist a sequence (x_n) in X and a positive number α such that $\rho(x_m, x_n) \geq \alpha$ whenever $m \neq n$. Clearly, (x_n) has no Cauchy subsequences and therefore no convergent subsequences. This contradicts our assumption (ii); so, in fact, X is totally bounded. Thus (ii) implies (iii).

It remains to prove that (iii) implies (i). Accordingly, let X be totally bounded and complete, and suppose that there exists an open cover $\mathcal{U}$ of X that contains no finite subcover. With $B_0 = X$, we construct a sequence $(B_n)_{n=1}^{\infty}$ of closed balls in X such that for each $n \geq 1$,

(a) B_n has radius 2^{-n},

(b) B_n has a nonempty intersection with B_{n-1}, and

(c) no finite subfamily of $\mathcal{U}$ is a cover of B_{n-1}.

Having constructed $B_0, \ldots, B_{n-1}$ with the applicable properties, let $(V_j)_{j=1}^{m}$ be a finite cover of B_{n-1} by balls in B_{n-1} of radius 2^{-n}. (Note that B_{n-1} is totally bounded, by Exercise (3.3.8: 2).) Amongst the sets V_j there exists at least one—call it B_n—that is not covered by finitely many of the sets in $\mathcal{U}$: otherwise each of the finitely many sets V_j, and therefore B_{n-1}, would be covered by finitely many elements of $\mathcal{U}$, thereby contradicting (c). This completes the inductive construction of B_n.

For each $n \geq 1$ let x_n be the centre of B_n. Since $B_n \cap B_{n-1}$ is nonempty, it follows from the triangle inequality that for $n \geq 2$,

$$\rho(x_n, x_{n-1}) \leq 2^{-n} + 2^{-n+1} < 2^{-n+2}.$$

So if $j > i \geq N \geq 1$, then

$$\rho(x_i, x_j) \leq \sum_{k=i+1}^{j} \rho(x_k, x_{k-1})$$

$$< \sum_{k=i+1}^{j} 2^{-k+2}$$

$$< 2^{-i+1} \sum_{k=0}^{\infty} 2^{-k} = 2^{-i+2} \leq 2^{-N+2}.$$

Hence (x_n) is a Cauchy sequence in X and so, as X is complete, converges to a limit x_∞ in X. Now pick $U \in \mathcal{U}$ such that $x_\infty \in U$. Since U is open, there exists $r > 0$ such that $B(x_\infty, r) \subset U$. Choosing $N > 1$ such that $\rho(x_\infty, x_N) < r/2$ and $2^{-N} < r/2$, we see that for each $x \in B_N$,

$$\rho(x, x_\infty) \leq \rho(x, x_N) + \rho(x_\infty, x_N) < 2^{-N} + r/2 < r,$$

so $x \in B(x_\infty, r)$. Hence $B_N \subset B(x_\infty, r) \subset U$, which contradicts (c). It follows that our initial assumption about the open cover $\mathcal{U}$ is false; whence X is compact, and therefore (iii) implies (i). □

The proof that (iii) implies (i) in Theorem (3.3.9) is a generalisation of the argument we used to prove the Heine–Borel–Lebesgue Theorem (1.4.6).

(3.3.10) Exercises

.1 Use sequential compactness arguments to show that a compact subset of a metric space is both bounded and closed.

.2 Show that if X is compact, then there exist points a, b of X such that $\rho(a, b) = \operatorname{diam}(X)$.

.3 Let A, B be nonempty disjoint subsets of a metric space X with A closed and B compact. Give two proofs that $\rho(A, B) > 0$.

.4 Let (S_n) be a descending sequence of compact sets in a metric space X (so $S_1 \supset S_2 \supset \cdots$). Prove, in at least two different ways, that if $S_n \neq \emptyset$ for all n, then $\bigcap_{n=1}^{\infty} S_n \neq \emptyset$.

.5 Let X be a compact space in which each point x is isolated (see Exercise (3.1.8: 5)). Give at least two proofs that X is finite.

.6 Prove that if every continuous mapping of X into $\mathbf{R}$ is bounded, then X is compact. (First suppose that X is not totally bounded, and use Exercise (3.3.8: 8) to construct an unbounded continuous mapping of X into $\mathbf{R}$. Then use Exercise (3.2.11: 9).)

Is this true if "continuous" is replaced by "uniformly continuous" in the hypothesis?

.7 Prove that if every uniformly continuous mapping of X into $\mathbf{R}^+$ has a positive infimum, then X is compact. (cf. Exercise (3.3.7: 2). Use Exercises (3.3.8: 8), (3.2.11: 5), and (3.2.11: 8).)

.8 Let (X, ρ) be a metric space, and suppose that X is complete with respect to every metric equivalent to ρ (see Exercise (3.1.3: 6)). Prove that X is compact. (Suppose that X is not totally bounded. By Exercise (3.3.8: 6), there exist a sequence (x_n) in X and a positive number α such that $\rho(x_m, x_n) \geq \alpha$ whenever $m \neq n$. Show that

$$d(x, y) = \min \left\{ \rho(x, y), \inf_{m,n \geq 1} \left\{ \rho(x, x_m) + \left| \tfrac{1}{m} - \tfrac{1}{n} \right| \alpha + \rho(y, x_n) \right\} \right\}$$

defines a metric equivalent to ρ with respect to which X is not complete.)

.9 Prove that the following are equivalent conditions on a metric space (X, ρ).

(i) If d is a metric equivalent to ρ, and S, T are disjoint closed subsets of (X, d), then $d(S, T) > 0$.

(ii) X is compact.

(Use Exercises (3.3.10: 3), (3.2.10: 6), and (3.3.10: 8); also, note the guide to the solution of Exercise (3.3.5: 2).)

The following property of a metric space X is known as the *Lebesgue covering property*.

> For each open cover $\mathcal{U}$ of X there exists $r > 0$ such that any open ball of radius r in X is contained in some $U \in \mathcal{U}$.

The positive number r associated with the open cover $\mathcal{U}$ in this way is called a *Lebesgue number* for $\mathcal{U}$.

(3.3.11) Proposition. *A compact metric space has the Lebesgue covering property.*

Proof. Let X be a compact metric space, and $\mathcal{U}$ an open cover of X. For each $x \in X$ choose $r_x > 0$ such that $B(x, 2r_x) \subset U$ for some $U \in \mathcal{U}$. The balls $B(x, r_x)$, with $x \in X$, form an open cover of X, from which we can extract a finite subcover, say

$$\{B(x_i, r_{x_i}) : 1 \leq i \leq n\}.$$

Then

$$0 < r = \min\{r_{x_1}, \ldots, r_{x_n}\}.$$

Given $x \in X$, choose i such that $x \in B(x_i, r_{x_i})$. Then for each $y \in B(x, r)$ we have

$$\rho(y, x_i) \leq \rho(x, y) + \rho(x, x_i) < r + r_{x_i} \leq 2r_{x_i}.$$

So

$$B(x,r) \subset B(x_i, 2r_{x_i}) \subset U$$

for some $U \in \mathcal{U}$. $\square$

The implications (i) ⇒ (ii) ⇒ (iii) of the first part of the next result—a general version of the *Uniform Continuity Theorem* for metric spaces—are well known, in contrast to the implication (iii) ⇒ (i), which is due to Wong [55].

(3.3.12) Theorem. *The following are equivalent conditions on a metric space X.*

(i) *X has the Lebesgue covering property.*

(ii) *Every continuous mapping of X into a metric space is uniformly continuous.*

(iii) *Every continuous mapping of X into $\mathbf{R}$ is uniformly continuous.*

Proof. Assuming (i), let f be a continuous mapping of X into a metric space, and let $\varepsilon > 0$. For each $t \in X$ there exists $\delta_t > 0$ such that if $\rho(x,t) < \delta_t$, then $\rho(f(x), f(t)) < \varepsilon/2$. It follows from the triangle inequality that if x and y belong to $B(t, \delta_t)$, then $\rho(f(x), f(y)) < \varepsilon$. Let $\delta > 0$ be a Lebesgue number for the open cover $(B(t,\delta_t))_{t \in X}$ of X. If x and y are points of X such that $\rho(x,y) < \delta$, then both x and y belong to $B(x,\delta)$, which is a subset of $B(t,\delta_t)$ for some t; so $\rho(f(x), f(y)) < \varepsilon$. Thus f is uniformly continuous, and therefore (i) implies (ii).

It is trivial that (ii) implies (iii). To complete the proof, suppose that X does not have the Lebesgue covering property; so there exists an open cover $\mathcal{U}$ of X for which there is no Lebesgue number. For each positive integer n we can therefore construct $x_n \in X$ such that $B(x_n, n^{-1}) \backslash U$ is nonempty for each $U \in \mathcal{U}$. Then there exists $y_n \in B(x_n, n^{-1}) \backslash \{x_n\}$: for otherwise we would have

$$B(x_n, n^{-1}) = \{x_n\} \subset U$$

for some $U \in \mathcal{U}$. We show that

$$\text{neither } (x_n) \text{ nor } (y_n) \text{ has a convergent subsequence.} \tag{1}$$

Indeed, if (x_n) had a subsequence that converged to a limit $\xi \in X$, then, choosing $U \in \mathcal{U}$ such that $\xi \in U$, we would have $B(x_n, n^{-1}) \subset U$ for some n, a contradiction. On the other hand, if $(y_{n_k})_{k=1}^{\infty}$ were a convergent subsequence of (y_n), then the subsequence (x_{n_k}) of (x_n) would converge to the same limit, which contradicts what we have just proved.

Setting $n_1 = 1$, suppose we have constructed $n_1 < n_2 < \cdots < n_k$ such that the sets

$$S_k = \{x_{n_1}, \ldots, x_{n_k}\},$$
$$T_k = \{y_{n_1}, \ldots, y_{n_k}\}$$

are disjoint. There exists $n_{k+1} > n_k$ such that $x_{n_{k+1}} \notin S_k$ and $y_{n_{k+1}} \notin T_k$: otherwise we would have either $x_j \in S_k$ for infinitely many j or else $y_j \in T_k$ for infinitely many j; since S_k and T_k are finite, this would imply that either (x_n) or (y_n) had a convergent subsequence, thereby contradicting (1). Thus we have inductively constructed a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of positive integers such that the sets

$$S = \{x_{n_k} : k \geq 1\},$$
$$T = \{y_{n_k} : k \geq 1\}$$

are disjoint. These sets are both closed in X: for example, any point of $\overline{S} \backslash S$ would be the limit of some subsequence of (x_n), which contradicts (1). Applying Urysohn's Lemma (Exercise (3.2.14: 1)), we now construct a continuous function $f : X \to [0,1]$ such that $f(S) = \{0\}$ and $f(T) = \{1\}$. Since $\rho(x_{n_k}, y_{n_k}) < 1/n_k$ but $|f(x_{n_k}) - f(y_{n_k})| = 1$, the function f is not uniformly continuous. Hence (iii) implies (i). □

(3.3.13) Corollary—The Uniform Continuity Theorem. *Every continuous mapping of a compact metric space into a metric space is uniformly continuous.*

Proof. This follows from the preceding two results. □

The converse of Corollary (3.3.13) is not true, since every function from the discrete metric space **N** to **R** is uniformly continuous but **N**, being unbounded, is not compact. However, there is an interesting partial converse to Corollary (3.3.13), which we discuss in Section 4.

(3.3.14) Exercises

.1 Use a sequential compactness argument to prove that a compact metric space has the Lebesgue covering property.

.2 Give an example of a totally bounded metric space for which the Lebesgue covering property does not hold.

.3 Prove that X has the Lebesgue covering property if and only if for each nonempty closed set $S \subset X$ and each open set U containing S, there exists $r > 0$ such that the r*–enlargement* of S,

$$B(S, r) = \{x \in X : \rho(x, S) < r\},$$

is contained in U. (For "only if", consider the open cover $\{X\backslash S, U\}$ of X. For "if", suppose that X does not have the Lebesgue covering property and, as in the second part of the proof of Theorem (3.3.12), construct disjoint nonempty closed subsets S, T of X such that $\rho(S,T) = 0$; then show that there exists $r > 0$ such that $B(S,r) \subset X\backslash T$.)

.4 Prove that a metric space with the Lebesgue covering property is complete. Need it be totally bounded?

.5 Let X have the Lebesgue covering property, and let Y be a closed subset of X. Give two proofs that Y has the Lebesgue covering property. (For one proof, use the Tietze Extension Theorem; for another, work directly with an open cover of Y.)

.6 Prove the Uniform Continuity Theorem using sequential compactness without the Lebesgue covering property.

.7 Let X be a metric space, and h a mapping of X into a compact metric space Y. Suppose that $f \circ h$ is uniformly continuous for each continuous (and therefore uniformly continuous) mapping $f : Y \to \mathbf{R}$. Give at least two proofs that h is uniformly continuous.

The notion of compactness can be generalised in a number of ways. The one we deal with is typical of topology, in that it replaces a global property (one that holds for the whole space) by a local one (one that holds in some neighbourhood of any given point).

A metric space X is said to be *locally compact*, or a *locally compact space*, if each point in X has a compact neighbourhood in X. For example, although (in view of Proposition (3.3.1)) $\mathbf{R}$ is not compact, it is locally compact: if $x \in \mathbf{R}$, then $[x-1, x+1]$ is a compact neighbourhood of x in $\mathbf{R}$. Of course, a compact metric space is locally compact.

(3.3.15) Proposition. *Let X be a locally compact space, and S a subset of X. If either S is open or S is closed, then S is locally compact.*

Proof. Let $a \in S$, and choose a compact neighbourhood K of a in X. If S is open, then

$$a \in (K \cap S)^\circ = K^\circ \cap S^\circ,$$

so there exists $r > 0$ such that $\overline{B}(a,r) \subset K$ and $\overline{B}(a,r) \subset S$. As $\overline{B}(a,r)$ is closed in X, it is closed in K (by Proposition (3.1.5)) and therefore compact (by Proposition (3.3.4)). Hence a has a compact neighbourhood in S, and so S is locally compact.

Now suppose that S is closed in X. Since K is a neighbourhood of a in X, $K \cap S$ is a neighbourhood of a in S (Exercise (3.1.6: 4)). Also, $K \cap S$ is closed in K, by Proposition (3.1.5), and therefore compact, by Proposition (3.3.4). Hence S is locally compact. □

(3.3.16) Exercises

.1 Let S and T be locally compact subspaces of a locally compact metric space X. Prove that $S \cap T$ is locally compact. Need $S \cup T$ be locally compact?

.2 Is every locally compact space complete?

.3 Let X be a metric space in which every bounded set is contained in a compact set. Prove that X is locally compact and separable.

.4 Let X be locally compact, and K a compact subset of X. Prove that for some $r > 0$ the closure of the r–enlargement of K is compact. (See Exercise (3.3.14: 3).)

.5 Let X be a separable locally compact metric space. Show that there exists a sequence (V_n) of open subsets of X, each of which has compact closure, with the property that for each $x \in X$ and each neighbourhood U of x there exists n such that $x \in V_n \subset U$. Hence prove that there exists a sequence (U_n) of open subsets of X with the following properties.

(i) $\overline{U_n}$ is compact;

(ii) $\overline{U_n} \subset U_{n+1}$;

(iii) $X = \bigcup_{n=1}^{\infty} U_n$.

(Set $U_1 = V_1$ and $U_{n+1} = V_{n+1} \cup B(\overline{U_n}, r)$, where, using Exercise (3.3.16: 4), $r > 0$ is chosen to make the closure of $B(\overline{U_n}, r)$ compact.)

.6 Let X be a separable locally compact metric space that is not compact, and let (U_n) be as in the preceding exercise. Use Urysohn's Lemma (Exercise (3.2.14: 1)) to show that there exists a continuous function $f : X \to \mathbf{R}$ such that $f(x) \leq n$ for all $x \in \overline{U_n}$, and $f(x) \geq n$ for all $x \in X \backslash \overline{U_n}$. Then show that

$$d(x, y) = \rho(x, y) + |f(x) - f(y)|$$

defines a metric d equivalent to ρ, and that in the space (X, d) any bounded set is contained in a compact set.

3.4 Connectedness

In analysis there are many situations where progress is made by restricting attention to parts of a metric space that cannot be split into smaller, separated parts. Our next definition captures this imprecise idea formally.

A metric space is said to be *connected*, or a *connected space,* if it can *not* be expressed as a union of two disjoint nonempty open subsets. So if X is connected, and if S, T are nonempty open subsets of X such that $S \cup T = X$, then $S \cap T \neq \emptyset$. A subspace that is connected is called a *connected set* in the metric space. Clearly, the empty subset of any metric space is connected.

(3.4.1) Proposition. *The following are equivalent conditions on a metric space X.*

(i) *X is connected.*

(ii) *X is not a union of two disjoint nonempty closed subsets.*

(iv) *The only subsets of X that are both open and closed in X are X and the empty subset.*

Proof. The straightforward proof is left as the next exercise. □

(3.4.2) Exercises

.1 Prove Proposition (3.4.1).

.2 Prove that a metric space X is connected if and only if there is no continuous mapping of X onto $\{0, 1\}$.

We showed in Proposition (1.3.13) that the only subsets of $\mathbf{R}$ that are both open and closed are $\mathbf{R}$ and $\emptyset$. It follows from Proposition (3.4.1) that $\mathbf{R}$ is connected. In fact, we can say more.

(3.4.3) Proposition. *A nonempty subset of* $\mathbf{R}$ *is connected if and only if it is an interval.*

Proof. Let S be a nonempty subset of $\mathbf{R}$, and suppose first that S is connected. Let a, b be points of S with $a \leq b$, and consider any x such that $a \leq x \leq b$. If $x \notin S$, then S is the union of the disjoint subsets $S \cap (-\infty, x)$ and $S \cap (x, \infty)$, each of which is open in S, by Proposition (3.1.5). This contradicts the assumption that S is connected. So $x \in S$, and therefore S has the intermediate value property. Hence, by Proposition (1.3.3), S is an interval.

Now let S be an interval in $\mathbf{R}$, and suppose that S is not connected. Then there exist nonempty open subsets A, B of the subspace S such that

$S = A \cup B$ and $A \cap B = \emptyset$. We may assume that there exist $a \in A$ and $b \in B$ such that $a < b$. Let x be the supremum of the nonempty bounded set $A \cap [a, b)$, and suppose that $x \in A$. Then $a \le x < b$, as $b \notin A$. Since A is open in S, there exists $r > 0$ such that $S \cap [x, x + r] \subset A \cap [a, b)$. Being an interval, S has the intermediate value property (Proposition (1.3.3)), so $[a, b] \subset S$, and therefore $[x, x + r] \subset S$. Hence $x + r \in A \cap [a, b)$, which contradicts the definition of x. Thus, in fact, $x \notin A$. A similar argument shows that $x \notin B$, which is absurd since, as we have already observed, $[a, b] \subset S$. This contradiction shows that S is connected. □

(3.4.4) Exercise

Let S, T be nonempty closed subsets of a metric space X such that $S \cup T$ and $S \cap T$ are connected. Prove that S and T are connected. Give an example to show that the conclusion no longer holds if we remove the hypothesis that S and T are closed.

We now prove some general results about connected spaces.

(3.4.5) Proposition. *If S, T are subsets of a metric space X such that S is connected and $S \subset T \subset \overline{S}$, then T is connected. In particular, $\overline{S}$ is connected.*

Proof. Suppose that A, B are nonempty open sets in the subspace T such that $T = A \cup B$ and $A \cap B = \emptyset$. As S is dense in T, both $S \cap A$ and $S \cap B$ are nonempty. They are clearly disjoint, and, by Proposition (3.1.5), they are open in S. Since $S = (S \cap A) \cup (S \cap B)$, we have contradicted the fact that S is connected. □

(3.4.6) Proposition. *If $\mathcal{F}$ is a family of connected sets in a metric space X such that $\bigcap \mathcal{F}$ is nonempty, then $\bigcup \mathcal{F}$ is connected.*

Proof. Let $S = \bigcup \mathcal{F}$ and $a \in \bigcap \mathcal{F}$. Suppose that $S = A \cup B$, where A, B are nonempty disjoint open sets in S. Consider, for example, the case where $a \in A$. Choose $F \in \mathcal{F}$ such that $B \cap F$ is nonempty, and note that $a \in A \cap F$. Then $A \cap F$ and $B \cap F$ are open in F (by Proposition (3.1.5)), have union F, are disjoint, and are nonempty. This contradicts the fact that F is connected. □

(3.4.7) Exercises

.1 Let S, T be connected subsets of a metric space X such that $\overline{S} \cap T$ is nonempty. Prove that $S \cup T$ is connected.

.2 Let (S_n) be a sequence of connected subsets of a metric space X such that $S_n \cap S_{n+1}$ is nonempty for each n. Prove that $\bigcup S_n$ is connected.

.3 A metric space X is said to be *chain connected* if for each pair a, b of points of X, and each $\varepsilon > 0$, there exist finitely many points $a = x_0, x_1, \dots, x_n = b$ such that $\rho(x_i, x_{i+1}) < \varepsilon$ for $i = 0, \dots, n-1$. Prove that a compact, chain connected metric space is connected.

.4 If X is a metric space, then it follows from Proposition (3.4.6) that for each $x \in X$,

$$C_x = \bigcup \{S \subset X : S \text{ is connected and } x \in S\}$$

is connected. C_x is called the *connected component* of x in X. Prove the following statements.

(i) C_x is closed in X.

(ii) C_x is the largest connected subset of X that contains x.

(iii) If $y \in C_x$, then $C_y = C_x$.

(iv) If $y \notin C_x$, then $C_y \cap C_x = \emptyset$.

.5 A subset S of a metric space X is said to be *totally disconnected* if for each $x \in X$ the connected component of x in S is $\{x\}$. Prove that

(i) every countable subset of $\mathbf{R}$ is totally disconnected;

(ii) the irrational numbers form a totally disconnected set in $\mathbf{R}$.

.6 A metric space X is said to be *locally connected* if for each $x \in X$ and each neighbourhood U of X there exists a connected neighbourhood V of x with $V \subset U$. Prove that X is locally connected if and only if the following property holds: for each open subset S of X, and each $x \in S$, the connected component of x in the subspace S is an open subset of X.

.7 Use Proposition (3.4.3) and the previous exercise to give another proof of Proposition (1.3.6).

.8 Let X be a connected space, and S a nonempty subset of X such that $X \backslash S$ is also nonempty. Show that the boundary of S is nonempty. (Suppose the contrary.)

(3.4.8) Proposition. *The range of a continuous mapping from a connected metric space into a metric space is connected.*

Proof. Let X be a connected space, and f a continuous mapping of X into a metric space Y. Suppose that $f(X) = S \cup T$, where S, T are nonempty disjoint open sets in the subspace $f(X)$ of Y. By Proposition

(3.2.2), the nonempty disjoint sets $f^{-1}(S)$ and $f^{-1}(T)$ are open in X. Since

$$X = f^{-1}(f(X)) = f^{-1}(S \cup T) = f^{-1}(S) \cup f^{-1}(T),$$

it follows that X is not connected, a contradiction. □

A very important consequence of Proposition (3.4.8) is the following generalised *Intermediate Value Theorem.*

(3.4.9) Theorem. *Let f be a continuous mapping of a connected metric space X into $\mathbf{R}$, and a, b points of $f(X)$ such that $a < b$. Then for each $y \in (a, b)$ there exists $x \in X$ such that $f(x) = y$.*

Proof. By Propositions (3.4.8) and (3.4.3), $f(X)$ is an interval. The result follows immediately. □

(3.4.10) Exercises

.1 Let X be an unbounded connected metric space. Prove that for each $x \in X$ and each $r > 0$ there exists $y \in X$ such that $\rho(x, y) = r$.

.2 Let S be a connected subset of the Euclidean space $\mathbf{R}^n$. Prove that for each $r > 0$ the set

$$\{x \in \mathbf{R}^n : \rho(x, S) \leq r\}$$

is also connected.

.3 Let X be a compact metric space, and suppose that the closure of any open ball $B(a, r)$ in X is the closed ball $\overline{B}(a, r)$. Show that any open or closed ball in X is connected. (Suppose that $B(a, r) = S \cup T$, where S, T are nonempty, disjoint, and closed in the subspace $B(a, r)$. Without loss of generality take a in A. Show that

$$C = \{x \in X \backslash S : \rho(a, x) \geq \rho(a, T)\}$$

is compact, and hence that there exists $t_0 \in T$ such that $\rho(a, t_0) = \rho(a, T) > 0$. Then consider $\overline{B}(a, \rho(a, T))$.)

Show by an example that we cannot remove the compactness of X from the hypotheses of this result.

We now prove the partial converse to Corollary (3.3.13) that was postponed from Section 3.

(3.4.11) Proposition. *Let X be a connected metric space such that every continuous function from X to $\mathbf{R}$ is uniformly continuous. Then X is compact.*

Proof. Suppose that X is not totally bounded. By Exercise (3.3.8: 6), there exist a sequence (x_n) in X and a positive number α such that $\rho(x_m, x_n) \geq \alpha$ whenever $m \neq n$. Using Exercise (3.3.8: 8), we can construct, for each k, a uniformly continuous function $\phi_k : X \to [0,1]$ such that $\phi_k(x_k) = 1$, $\phi_k(x) = 0$ if $\rho(x, x_k) \geq \alpha/3$, and $f = \sum_{n=1}^{\infty} n\phi_n$ is a well–defined continuous function on X; to be precise, we set

$$\phi_k(x) = \max\left\{0,\ 1 - 3\alpha^{-1}\rho(x, x_k)\right\}.$$

Our hypotheses ensure that f is uniformly continuous. Now, X is connected, the mapping $x \mapsto \rho(x, x_n)$ is continuous on X, $\rho(x_n, x_n) = 0$, and $\rho(x_{n+1}, x_n) \geq \alpha$. It follows from Theorem (3.4.9) that there exists $x \in X$ such that $\rho(x, x_n) = \alpha/3n$. Then

$$f(x_n) - f(x) = n - (n-1) = 1.$$

Since $n > 1$ is arbitrary, f is not uniformly continuous. This contradiction shows that X is totally bounded.

Now suppose that X is not complete; so there exists a Cauchy sequence (x_n) in X that does not converge to a limit in X. Without loss of generality we may assume that X is a dense subset of its completion $(\widehat{X}, \rho)$. So (x_n) converges to a limit $x_\infty \in \widehat{X} \backslash X$. The function $x \mapsto \rho(x, x_\infty)$ is (uniformly) continuous and positive–valued on X, so

$$f(x) = \frac{1}{\rho(x, x_\infty)}$$

defines a continuous mapping $f : X \to \mathbf{R}^+$. By our hypotheses, f is uniformly continuous on X, so there exists $\delta > 0$ such that if $x, y \in X$ and $\rho(x, y) < \delta$, then $|f(x) - f(y)| < 1$. Choose N such that $\rho(x_m, x_n) < \delta$ for all $n \geq N$. Since $\rho(x_N, x_\infty) > 0$, there exist positive integers k, m such that $m > N$ and

$$\rho(x_m, x_\infty) < \frac{1}{k+1} < \frac{1}{k} < \rho(x_N, x_\infty).$$

Then $\rho(x_m, x_N) < \delta$ but

$$f(x_m) - f(x_N) > (k+1) - k = 1,$$

contrary to our choice of δ. Hence, in fact, X is complete and therefore, by Theorem (3.3.9), compact. □

There is another type of connectedness of importance in analysis and topology, one that generalises the informal idea that a subset X of the Euclidean plane is in one piece if any two points of X can be joined by a path that lies wholly in X. (In spite of this correct claim about the importance of this type of connectedness, we do not actually use it later in the book; so you can ignore the rest of this section with impunity.)

Let X be a metric space. A continuous mapping $f:[0,1]\to X$ such that $f(0)=a$ and $f(1)=b$ is called a *path* in X with *endpoints* a and b, or a *path in* X *from* a *to* b; the path f is also said to *join* a *to* b. We say that X is *path connected,* or a *path connected space,* if for each pair a,b of points of X there is a path in X from a to b. By a *path connected subset* of X we mean a subset of X that is path connected as a subspace of X.

A subset S of $\mathbf{R}^n$ is said to be *convex* if $tx+(1-t)y\in S$ whenever $x,y\in S$ and $0\leq t\leq 1$. A convex subset S of $\mathbf{R}^n$ is path connected: for if $a,b\in S$, then

$$f(t)=(1-t)a+tb\quad(0\leq t\leq 1)$$

defines a path in S from a to b. In particular, an interval in $\mathbf{R}$ is path connected.

(3.4.12) Proposition. *A path connected space is connected.*

Proof. Let X be a path connected space; we may assume that X is nonempty. Let $a\in X$, and for each $x\in X$ let f_x be a path in X joining a to x; for convenience, let $I=[0,1]$. Then $f_x(I)$ is connected, by Propositions (3.4.3) and (3.4.8), and $a\in f_x(I)$. Hence, by Proposition (3.4.6), $X=\bigcup_{x\in X}f_x(I)$ is connected. □

Propositions (3.4.3) and (3.4.12) show that path connectedness and connectedness are equivalent properties of a nonempty subset S of $\mathbf{R}$, and hold precisely when S is an interval. In $\mathbf{R}^2$, however, there are subsets that are connected but not path connected; see Exercise (3.4.16: 1). Our next result is therefore substantial.

(3.4.13) Proposition. *A connected open subset of* $\mathbf{R}^n$ *is path connected.*

In order to prove Proposition (3.4.13) we need some simple consequences of the following *Glueing Lemma.*

(3.4.14) Lemma. *Let X,Y be metric spaces, and let A,B be closed subsets of X whose union is X. Let $f:A\to Y$ and $g:B\to Y$ be continuous functions such that $f(x)=g(x)$ for all $x\in A\cap B$. Then the function $h:X\to Y$ defined by*

$$h(x)=\begin{cases} f(x) & \text{if } x\in A\\ g(x) & \text{if } x\in B\end{cases}$$

is continuous.

Proof. Let C be a closed subset of Y. Then, by Proposition (3.2.2), $f^{-1}(C)$ is closed in the subspace A of X,and hence, by Exercise (3.1.6: 3), in X. Similarly, $g^{-1}(C)$ is closed in X. Hence

$$h^{-1}(C)=f^{-1}(C)\cup g^{-1}(C)$$

is closed in X. It follows from Proposition (3.2.2) that h is continuous. □

Now consider two paths f, g in a metric space X such that $f(1) = g(0)$. We define the *product of the paths* f and g to be the path gf, where

$$gf(t) = \begin{cases} f(2t) & \text{if } 0 \leq t \leq \frac{1}{2} \\ g(2t-1) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

It follows from Proposition (3.4.14) that gf is a path in X joining $f(0)$ to $g(1)$.

The product gf of two paths must not be confused with the composite $g \circ f$ of two mappings. Indeed, unless f is a path in $[0,1]$, the composite of the paths f and g is undefined.

(3.4.15) Exercise

We define the *path component* of a point x in a metric space X to be

$$P_x = \{y \in X : \text{there exists a path in } X \text{ from } x \text{ to } y\}.$$

Prove that P_x is the union of the path connected subsets of X that contain x, and that it is the largest path connected subset of X containing x. Prove also that if $x, y \in X$, then either $P_x = P_y$ or $P_x \cap P_y = \emptyset$.

Proof of Proposition (3.4.13). Let U be a connected open subset of $\mathbf{R}^n$. For each x in U let U_x be the path component of x in U; we first show that U_x is open in U. Given y in U_x, choose a path f in U joining x to y; choose also $r > 0$ such that $B(y,r) \subset U$. Since $B(y,r)$ is convex, for each $z \in B(y,r)$ there exists a path g in $B(y,r)$ joining y to z; then gf is a path in U joining x to z. Hence $B(y,r) \subset U_x$, ,and therefore U_x is open in U.

Now suppose that U is not path connected. Then there exist distinct points of U that cannot be joined by a path in U. Let a be one of these points. By the foregoing, U_a is nonempty and open in U, as is

$$V = \bigcup \{U_x : x \in U \backslash U_a\}.$$

Moreover, $U = U_a \cup V$. Since U is connected, $U_a \cap V$ is nonempty, so there exists $b \in U \backslash U_a$ such that $U_a \cap U_b \neq \emptyset$. Exercise (3.4.15) shows that $U_a = U_b$; whence $b \in U_a$, a contradiction. Thus U is path connected. □

(3.4.16) Exercises

.1 Let

$$\begin{aligned} A &= \left\{(0,y) \in \mathbf{R}^2 : -1 \leq y \leq 1\right\}, \\ B &= \left\{(x,y) \in \mathbf{R}^2 : 0 < x \leq 1,\ y = \sin\tfrac{\pi}{x}\right\}, \end{aligned}$$

and $X = A \cup B$. Prove that any connected subset of X that intersects both A and B has diameter greater than 2. Then prove that X is not path connected. (Suppose there exists a path $f : [0,1] \to X$ with $f(0) \in A$ and $f(1) \in B$. Let

$$\tau = \sup \{t \in [0,1] : f([0,t]) \subset A\},$$

and show that there exists $\tau' > \tau$ such that $f(\tau') \in B$ and the diameter of $f([\tau, \tau'])$ is less than 1.)

.2 Let $\mathcal{F}$ be a family of path connected subsets of a metric space X such that $\bigcap \mathcal{F} \neq \emptyset$. Prove that $\bigcup \mathcal{F}$ is path connected.

.3 Let $(S_n)_{n=1}^{\infty}$ be a sequence of path connected subsets of a metric space X such that for each $n \geq 1$,

$$S_n \cap \bigcup_{i=1}^{n-1} S_i \neq \emptyset.$$

Prove that $\bigcup_{n=1}^{\infty} S_n$ is path connected.

3.5 Product Metric Spaces

Let (X_1, ρ_1) and (X_2, ρ_2) be nonempty[2] metric spaces, and X their Cartesian product $X_1 \times X_2$. Throughout this section we use such notations as $x = (x_1, x_2)$, $x' = (x_1', x_2')$, and $a = (a_1, a_2)$ for points of X; we write $B_k(a_k, r)$ (respectively, $\overline{B}_k(a_k, r)$) for the open (respectively, closed) ball in X_k with centre a_k and radius r.

It is a simple exercise to show that the mapping $\rho : X \times X \to \mathbf{R}$ defined by

$$\rho(x, y) = \max\{\rho_1(x_1, y_1), \rho_2(x_2, y_2)\}$$

is a metric—called the *product metric*—on X; taken with this metric, X is called the *product* of the metric spaces X_1 and X_2. We assume that X carries this metric in the remainder of this section.

There are at least two other natural metrics on the set X: namely, the metrics ρ' and ρ'' defined by

$$\rho'(x, y) = \sqrt{\rho_1(x_1, y_1)^2 + \rho_2(x_2, y_2)^2}$$

and

$$\rho''(x, y) = \rho_1(x_1, y_1) + \rho_2(x_2, y_2).$$

[2] The requirement that X_1 and X_2 be nonempty enables us to avoid some minor complications.

Since
$$\rho(x, y) \leq \rho'(x, y) \leq \rho''(x, y) \leq 2\rho(x, y),$$
the identity mapping i_X (see Exercise (3.2.1: 1)) is uniformly continuous when its domain and range are given any of the metrics ρ, ρ', ρ''. Hence, in particular, each of these three metrics gives rise to the same topology (family of open sets) on X; that is, the metrics are equivalent (see Exercise (3.1.3: 6)).

(3.5.1) Lemma. *The open ball with centre a and radius r in the product space X is $B_1(a_1, r) \times B_2(a_2, r)$, and the closed ball with centre a and radius r in X is $\overline{B}_1(a_1, r) \times \overline{B}_2(a_2, r)$.*

Proof. For example, we have

$$\begin{aligned} \rho(a, x) < r \quad &\Leftrightarrow \quad \max\{\rho_1(a_1, x_1), \rho_2(a_2, x_2)\} < r \\ &\Leftrightarrow \quad \rho_1(a_1, x_1) < r \text{ and } \rho_2(a_2, x_2) < r, \end{aligned}$$

so $B(a, r) = B_1(a_1, r) \times B_2(a_2, r)$. □

(3.5.2) Proposition. *If A_1 is open in X_1, and A_2 is open in X_2, then $A_1 \times A_2$ is open in X.*

Proof. Let
$$a \in A = A_1 \times A_2.$$
Then $a_1 \in A_1$ and $a_2 \in A_2$; so there exist $r_1, r_2 > 0$ such that $B_1(a_1, r_1) \subset A_1$ and $B_2(a_2, r_2) \subset A_2$. Let $r = \min\{r_1, r_2\}$; then by Lemma (3.5.1),

$$B(a, r) \subset B_1(a_1, r_1) \times B_2(a_2, r_2) \subset A.$$

Hence $a \in A^\circ$, and so A is open in X. □

(3.5.3) Corollary. *If U_k is a neighbourhood of x_k in X_k, then $U_1 \times U_2$ is a neighbourhood of x in X.*

Proof. Choose an open set A_k in X_k such that $x_k \in A_k \subset U_k$. Then

$$(x_1, x_2) \in A_1 \times A_2 \subset U_1 \times U_2,$$

where, by the previous proposition, $A_1 \times A_2$ is an open subset of X. □

The mapping $\mathrm{pr}_k : X \to X_k$ defined by

$$\mathrm{pr}_k(x_1, x_2) = x_k$$

is called the *projection* of X onto X_k.

(3.5.4) Proposition. *If A is an open set in X, then $\mathrm{pr}_k(A)$ is open in the space X_k.*

Proof. Consider any $x_1 \in X_1$. Either

$$A(x_1) = \{x_2 \in X_2 : (x_1, x_2) \in A\}$$

is empty and therefore open, or else there exists $x_2 \in A(x_1)$. In the latter case, since A is open, we can choose $r > 0$ such that $B(x, r) \subset A$, where $x = (x_1, x_2)$. If $x_2' \in X_2$ and $\rho_2(x_2, x_2') < r$, then

$$\rho(x, (x_1, x_2')) = \rho_2(x_2, x_2') < r,$$

so $(x_1, x_2') \in A$. Hence $A(x_1)$ is open in this case also. Since

$$\mathrm{pr}_2(A) = \bigcup_{x_1 \in X_1} A(x_1),$$

a union of open sets, it follows that $\mathrm{pr}_2(A)$ is open in X_2. A similar argument shows that $\mathrm{pr}_1(A)$ is open in X_1. □

Note that the projections of a closed subset of X need not be closed; see the remarks following the proof of Proposition (3.2.2) on page 137.

(3.5.5) Proposition. *If $A_1 \subset X_1$ and $A_2 \subset X_2$, then*

$$\overline{A_1 \times A_2} = \overline{A_1} \times \overline{A_2}.$$

Proof. Let $a \in \overline{A_1} \times \overline{A_2}$. Then for each $\varepsilon > 0$ there exist $x_1 \in A_1$ and $x_2 \in A_2$ such that $\rho_1(a_1, x_1) < \varepsilon$ and $\rho_2(a_2, x_2) < \varepsilon$; whence $\rho(a, x) < \varepsilon$, where

$$x = (x_1, x_2) \in A_1 \times A_2.$$

Thus $\overline{A_1} \times \overline{A_2} \subset \overline{A_1 \times A_2}$.

On the other hand, if $a \notin \overline{A_1} \times \overline{A_2}$, then either $a_1 \notin \overline{A_1}$ or $a_2 \notin \overline{A_2}$. Taking, for example, the first alternative, we see from Exercise (3.1.3: 3) and Corollary (3.5.3) that the set $(X_1 \backslash \overline{A_1}) \times X_2$, which is clearly disjoint from $A_1 \times A_2$, is a neighbourhood of a; thus $a \notin \overline{A_1 \times A_2}$. Hence

$$X \backslash \left(\overline{A_1} \times \overline{A_2}\right) \subset X \backslash \overline{A_1 \times A_2},$$

so $\overline{A_1 \times A_2} \subset \overline{A_1} \times \overline{A_2}$, and therefore $\overline{A_1 \times A_2} = \overline{A_1} \times \overline{A_2}$. □

(3.5.6) Corollary. *$A_1 \times A_2$ is closed in X if and only if A_k is closed in X_k for each k.*

Proof. This follows immediately from the last proposition. □

A mapping f from a set E into $X = X_1 \times X_2$ can be identified with the ordered pair $(\mathrm{pr}_1 \circ f, \mathrm{pr}_2 \circ f)$; where there is no risk of confusion, we write f_k for the mapping $\mathrm{pr}_k \circ f$ of E into X_k, so that $f = (f_1, f_2)$.

(3.5.7) Proposition. *Let f be a mapping of a metric space (E, d) into X. Then f is continuous at $a \in E$ if and only if both f_1 and f_2 are continuous at a.*

Proof. Suppose that for each k, f_k is continuous at a_k. Given $\varepsilon > 0$, choose $\delta_k > 0$ such that if $d(a, x) < \delta_k$, then $\rho_k(f_k(a), f_k(x)) < \varepsilon$. If $d(a, x) < \min\{\delta_1, \delta_2\}$, then

$$\rho(f(a), f(x)) = \max\{\rho_1(f_1(a), f_1(x)),\ \rho_2(f_2(a), f_2(x))\} < \varepsilon.$$

Thus f is continuous at a.

To prove the converse, first note that, trivially, pr_k is continuous on X; so if f is continuous at a, then so is $\mathrm{pr}_k \circ f$, by Proposition (3.2.3). □

(3.5.8) Proposition. *Let f be a mapping of a metric space E into X. Then f is uniformly continuous if and only if both f_1 and f_2 are uniformly continuous.*

Proof. This is left as an exercise. □

(3.5.9) Exercises

.1 Prove that if a mapping f of X into a metric space Y is continuous at (a, b), then the mappings $x_1 \mapsto f(x_1, b)$ and $x_2 \mapsto f(a, x_2)$ are continuous at a and b, respectively.

.2 Prove Proposition (3.5.8).

.3 Let E be a metric space, $A \subset E$, and $a \in \overline{A \backslash \{a\}}$. Prove that a mapping $f : E \to X$ has a limit at a with respect to A if and only if both $b_1 = \lim_{t \to a,\, t \in A} f_1(t)$ and $b_2 = \lim_{t \to a,\, t \in A} f_2(t)$ exist, in which case $\lim_{t \to a,\, t \in A} f(t) = (b_1, b_2)$.

.4 Prove that a sequence (x_n) in X converges to a limit in X if and only if both $\xi_1 = \lim_{n \to \infty} \mathrm{pr}_1(x_n)$ and $\xi_2 = \lim_{n \to \infty} \mathrm{pr}_2(x_n)$ exist, in which case $\lim_{n \to \infty} x_n = (\xi_1, \xi_2)$.

.5 Prove that a sequence (x_n) in X is a Cauchy sequence if and only if $(\mathrm{pr}_1(x_n))$ is a Cauchy sequence in X_1 and $(\mathrm{pr}_2(x_n))$ is a Cauchy sequence in X_2.

.6 For $i = 1, 2$ let X_i, Y_i be metric spaces, and f_i a mapping of X_i into Y_i. Prove that the mapping

$$(x_1, x_2) \mapsto (f_1(x_1), f_2(x_2))$$

of $X_1 \times X_2$ into $Y_1 \times Y_2$ is continuous if and only if both f_1 and f_2 are continuous.

We have now reached the main result of this section.

(3.5.10) Proposition. *Let T be any one of the following types of metric space: complete, totally bounded, compact. Then the product $X = X_1 \times X_2$ of two nonempty metric spaces X_1 and X_2 is of type T if and only if both X_1 and X_2 are of type T.*

Proof. Leaving the necessity of the stated conditions as an exercise, we prove their sufficiency. To this end, assume that X_1 and X_2 are complete, and consider a Cauchy sequence (x_n) in X. By Exercise (3.5.9: 5), $(\mathrm{pr}_k(x_n))_{n=1}^{\infty}$ is a Cauchy sequence in X_k; since X_k is complete,

$$\xi_k = \lim_{n \to \infty} \mathrm{pr}_k(x_n)$$

exists. Reference to Exercise (3.5.9: 4) shows that (x_n) converges to the point (ξ_1, ξ_2) of X. Hence X is complete.

It is easy to see that if $\varepsilon > 0$ and F_k is a finite ε–approximation to X_k, then $F_1 \times F_2$ is a finite ε–approximation to X. It follows that if X_1 and X_2 are totally bounded, so is X.

The first two parts of the proof, and Theorem (3.3.9), show that if X_1 and X_2 are compact, then so is X. □

(3.5.11) Exercises

.1 Prove that the product of two discrete metric spaces is discrete.

.2 In the notation of Proposition (3.5.10), prove that if X is of type T, then so are X_1 and X_2.

.3 Prove that X is separable if and only if both X_1 and X_2 are separable.

.4 Prove that X is locally compact if and only if both X_1 and X_2 are locally compact.

.5 Prove that X is connected (respectively, path connected) if and only if both X_1 and X_2 are connected (respectively, path connected).

.6 Prove that a subset of the product space $\mathbf{R}^2$ or $\mathbf{C}^2$ is compact if and only if it is closed and bounded.

.7 Prove that the Euclidean spaces $\mathbf{R}^2$ and $\mathbf{C}^2$ are complete.

.8 Show that in the product space $\mathbf{R}^2$ the set

$$X = (\{0\} \times [0,1]) \cup ([0,1] \times \{0\})$$

is compact, that every ball in X is connected, but that the closure of an open ball in X need not be the corresponding closed ball (cf. Exercise (3.4.10: 3)).

We define the *product of a finite family* $(X_1, \rho_1), \ldots, (X_n, \rho_n)$ *of metric spaces* to be the metric space (X, ρ), where

$$X = X_1 \times \cdots \times X_n$$

and

$$\rho((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \max\{\rho_i(x_i, y_i) : i = 1, \ldots, n\}.$$

The results proved so far in this section extend in the obvious ways to a product of more than two, but finitely many, metric spaces. The final set of exercises in this chapter shows how we can handle the product of a sequence of metric spaces.

(3.5.12) Exercises

.1 Let $((X_n, \rho_n))_{n=1}^{\infty}$ be a sequence of nonempty metric spaces such that $\operatorname{diam}(X_n) \leq 1$ for each n. Let X be the set of all sequences $(x_n)_{n=1}^{\infty}$ such that $x_n \in X_n$ for each n, and define a mapping $\rho : X \times X \to \mathbf{R}$ by

$$\rho((x_n), (y_n)) = \sum_{n=1}^{\infty} 2^{-n} \rho_n(x_n, y_n).$$

Prove that ρ is a metric on X.

The metric space (X, ρ) is called the *product of the sequence* (X_n) *of metric spaces* and is usually denoted by $\prod_{n=1}^{\infty} X_n$.

The next four exercises use the notation of Exercise (3.5.12: 1).

.2 Let $x = (x_n)_{n=1}^{\infty}$ be a point of X. Prove that $U \subset X$ is a neighbourhood of x in X if and only if for some positive integer m and some $r > 0$, U contains a set of the form

$$U_m(x, r) = \{(y_n)_{n=1}^{\infty} \in X : \rho_i(x_i, y_i) \leq r \text{ for } 1 \leq i \leq m\}.$$

.3 For each n let A_n be a subset of X_n. Prove that the closure of $\prod_{n=1}^{\infty} A_n$ in X is $\prod_{n=1}^{\infty} \overline{A_n}$.

.4 For each k let $x_k = (x_{k,n})_{n=1}^{\infty}$ be a point of X. Prove that the sequence $(x_k)_{k=1}^{\infty}$ converges in X to a limit $a = (a_n)_{n=1}^{\infty}$ if and only if for each n the sequence $(x_{k,n})_{k=1}^{\infty}$ converges to a_n in X_n. Prove also that $(x_k)_{k=1}^{\infty}$ is a Cauchy sequence in X if and only if for each n the sequence $(x_{k,n})_{k=1}^{\infty}$ is a Cauchy sequence in X_n.

.5 With T as in Proposition (3.5.10), prove that X is of type T if and only if X_n is of type T for each n.

.6 Let $(X_n)_{n=1}^{\infty}$ be a sequence of discrete metric spaces, each having positive diameter ≤ 1. Prove that the product space $\prod_{n=1}^{\infty} X_n$ is not discrete.

4

Analysis in Normed Linear Spaces

...I could be bounded in a nutshell, and
count myself a king of infinite space...
HAMLET, Act 2, Scene 2

Many significant applications of analysis are the fruit of cross–fertilisation between metric structure and algebraic structure. In this chapter we discuss such a cross–breed: a normed (linear) space. Section 1 introduces these objects and deals with their elementary analytic and geometric properties. In Section 2 we discuss linear mappings between normed spaces, paying particular attention to bounded linear functionals—continuous linear mappings into **R** and **C**. Although many of the most important normed spaces of analysis are infinite–dimensional, finite–dimensional ones remain significant in many ways; they are dealt with in Section 3. The next two sections deal with two fundamental classes of infinite–dimensional complete normed spaces: the L_p integration spaces and the space $\mathcal{C}(X)$ of continuous functions from a compact metric space X into **R**. They also characterise the associated bounded linear functionals. Two of the most important results about $\mathcal{C}(X)$—Ascoli's Theorem and the Stone–Weierstrass Theorem (a far–reaching generalisation of the classical Weierstrass Approximation Theorem)—are proved in Sections 5 and 6. Both of these theorems reappear in the final section of the chapter, where they are applied to the concrete classical problem of solving ordinary differential equations.

4.1 Normed Linear Spaces

Metric spaces offer one context within which the analytic and topological properties of $\mathbf{R}$ can be generalised, but they do not provide a natural framework for a generalisation of the algebraic properties of $\mathbf{R}$. A framework of the latter sort is made available by the notion of a normed linear space.

Let $\mathbf{F}$ stand for either $\mathbf{R}$ or $\mathbf{C}$, and let X be a linear space (vector space) over $\mathbf{F}$. A *norm* on X is a mapping $x \mapsto \|x\|$ of X into $\mathbf{R}$ such that the following properties hold for all $x, y \in X$ and $\lambda \in \mathbf{F}$.

N1 $\|x\| \geq 0.$

N2 $\|x\| = 0$ if and only if $x = 0.$

N3 $\|\lambda x\| = |\lambda| \; \|x\|$.

N4 $\|x + y\| \leq \|x\| + \|y\|$ (*triangle inequality*).

A *normed linear space*, or *normed space*, over $\mathbf{F}$ is a pair $(X, \|\cdot\|)$ consisting of a linear space X over $\mathbf{F}$ and a norm $\|\cdot\|$ on X; by abuse of language, we refer to the linear space X itself as a normed space if it is clear from the context which norm is under consideration. We say that the normed space X is *real* or *complex*, depending on whether $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$. A vector with norm 1 is called a *unit vector*.

The simplest example of a norm is, of course, the mapping $x \mapsto |x|$ on $\mathbf{F}$.

If X is a normed space, then the mapping $(x, y) \mapsto \|x - y\|$ of $X \times X$ into $\mathbf{R}$ is a metric on X (Exercise (4.1.1: 1)), and is said to be *associated* with the norm on X. When we consider X as a metric space, it is understood that we are referring to the metric associated with the given norm on X. By the *unit ball* of X we mean the closed ball with centre 0 and radius 1,

$$\overline{B}(0,1) = \{x \in X : \|x\| \leq 1\},$$

relative to the metric associated with the norm on X.

(4.1.1) Exercises

.1 Prove that $\rho(x, y) = \|x - y\|$ defines a metric on a normed space X, such that

$$\rho(x + z,\, y + z) = \rho(x, y),$$
$$\rho(\lambda x,\, \lambda y) = |\lambda| \; \rho(x, y)$$

for all $x, y, z \in X$ and $\lambda \in \mathbf{F}$.

.2 Show that

$$|\,\|x\| - \|y\|\,| \leq \|x - y\|$$

for all vectors x, y in a normed space X. Hence prove that if a sequence (x_n) converges to a limit x in X, then $\|x\| = \lim_{n\to\infty} \|x_n\|$.

.3 Prove that

$$\|x\| = \inf\left\{|t|^{-1} : t \in \mathbf{F}, t \neq 0,\ \|tx\| \leq 1\right\}$$

for each element x of a normed space X.

.4 Prove that for each positive integer n the mappings

$$\begin{aligned}(x_1, \ldots, x_n) &\mapsto \max\{|x_1|, \ldots, |x_n|\},\\ (x_1, \ldots, x_n) &\mapsto \sqrt{x_1^2 + \cdots + x_n^2},\\ (x_1, \ldots, x_n) &\mapsto |x_1| + \cdots + |x_n|\end{aligned}$$

are norms on $\mathbf{F}^n$. In the case $\mathbf{F} = \mathbf{R}$ the second of these norms is called the *Euclidean norm* on $\mathbf{R}^n$, and the associated metric is the Euclidean metric (see Exercise (3.1.1: 5)).

.5 Let X be a nonempty set, and denote by $\mathcal{B}(X, \mathbf{F})$ the set of all bounded mappings of X into $\mathbf{F}$, taken with the pointwise operations of addition and multiplication–by–scalars:

$$\begin{aligned}(f+g)(x) &= f(x) + g(x),\\ (\lambda f)(x) &= \lambda f(x).\end{aligned}$$

The *supremum norm,* or *sup norm,* on $\mathcal{B}(X, \mathbf{F})$ is defined by

$$\|f\| = \sup\{|f(x)| : x \in X\}.$$

Verify that the sup norm is a norm on $\mathcal{B}(X, \mathbf{F})$.

.6 Prove that $\|f\|_1 = \int |f|$ defines a norm on the set $L_1(\mathbf{R})$ of all Lebesgue integrable functions (defined almost everywhere) on $\mathbf{R}$, where two such functions are taken as equal if and only if they are equal almost everywhere.

.7 Let X_1, X_2 be normed spaces over $\mathbf{F}$, and recall that the standard operations of addition and multiplication–by–scalars on the product vector space $X = X_1 \times X_2$ are given by

$$\begin{aligned}(x_1, x_2) + (x_1', x_2') &= (x_1 + x_1',\ x_2 + x_2'),\\ \lambda(x_1, x_2) &= (\lambda x_1, \lambda x_2).\end{aligned}$$

Verify that the mapping $(x_1, x_2) \mapsto \max\{\|x_1\|, \|x_2\|\}$ is a norm on X, and that the metric associated with this norm is the product metric on X (considered as the product of the metric spaces X_1 and X_2).

Taken with this norm, which we call the *product norm*, X is known as the *product* of the normed spaces X_1 and X_2. The product norm and the product space for a finite number of normed spaces are defined analogously.

(4.1.2) Proposition. *Let X be a normed space over $\mathbf{F}$. Then*

(i) *the mapping $(x, y) \mapsto x + y$ is uniformly continuous on $X \times X$;*

(ii) *for each $\lambda \in \mathbf{F}$ the mapping $x \mapsto \lambda x$ is uniformly continuous on X;*

(iii) *for each $x \in X$ the mapping $\lambda \mapsto \lambda x$ is uniformly continuous on $\mathbf{F}$;*

(iv) *the mapping $(\lambda, x) \mapsto \lambda x$ is continuous on $\mathbf{F} \times X$.*

Proof. The uniform continuity of the first three mappings follows from the relations

$$\begin{aligned} \|(x+y) - (x'+y')\| &\le \|x - x'\| + \|y - y'\|, \\ \|\lambda x - \lambda y\| &= |\lambda| \, \|x - y\|, \\ \|\lambda x - \lambda' x\| &= |\lambda - \lambda'| \, \|x\|. \end{aligned}$$

On the other hand, the relations

$$\begin{aligned} \|\lambda x - \lambda_0 x_0\| &= \|\lambda_0(x - x_0) + (\lambda - \lambda_0)x_0 + (\lambda - \lambda_0)(x - x_0)\| \\ &\le |\lambda_0| \, \|x - x_0\| + |\lambda - \lambda_0| \, \|x_0\| + |\lambda - \lambda_0| \, \|x - x_0\| \end{aligned}$$

easily lead to the continuity of $(\lambda, x) \mapsto \lambda x$ at (λ_0, x_0). □

If X is a normed space and S is a linear subset of X, then the restriction to S of the norm on X is a norm on S; taken with this norm, S is called a *normed linear subspace*, or simply a *subspace*, of the normed space X.

(4.1.3) Proposition. *If S is a subspace of a normed space X, then the closure of S in X is a subspace of X.*

Proof. Let f be the mapping $(x, y) \mapsto x + y$ of $X \times X$ into X. As S is a subspace, f maps $S \times S$ into S, so

$$S \times S \subset f^{-1}(S) \subset f^{-1}(\overline{S}),$$

and therefore $\overline{S \times S}$ is a subset of the closure of $f^{-1}(\overline{S})$. Since, by Proposition (4.1.2), f is continuous on $X \times X$, it follows from Proposition (3.2.2) that $f^{-1}(\overline{S})$ is a closed subset of X; whence $\overline{S \times S} \subset f^{-1}(\overline{S})$. But $\overline{S \times S} = \overline{S} \times \overline{S}$, by Proposition (3.5.5); so if $x \in \overline{S}$ and $y \in \overline{S}$, then $x+y \in \overline{S}$. A similar argument, using the continuity of the mapping $(\lambda, x) \mapsto \lambda x$, shows that if $\lambda \in \mathbf{F}$ and $x \in \overline{S}$, then $\lambda x \in \overline{S}$. □

(4.1.4) Lemma. *If S is a closed subspace of a normed space X, and $a \in X$, then*

$$a + S = \{a + x : x \in S\}$$

is closed in X.

Proof. Let f be the mapping $z \mapsto z - a$ of X into itself. Since f is the composition of the mappings $z \mapsto (z, -a)$ and $(x, y) \mapsto x + y$, it follows from Exercise (3.5.9: 1), Proposition (4.1.2), and Proposition (3.2.3) that f is continuous on X. But $a + S = f^{-1}(S)$; so, by Proposition (3.2.2), $a + S$ is closed in X. □

(4.1.5) Exercises

.1 Explain why, in Proposition (4.1.2), the mapping $(\lambda, x) \mapsto \lambda x$ is not uniformly continuous on $\mathbf{F} \times X$.

.2 Complete the proof of Proposition (4.1.3).

.3 Let $(\widehat{X}, \rho)$ be a metric space, and X a normed space such that

(i) $\rho(x, y) = \|x - y\|$ for all $x, y \in X$;

(ii) X is dense in $\widehat{X}$.

Show that the operations of addition and multiplication-by–scalars can be extended uniquely to make $\widehat{X}$ a normed space with associated metric the given metric ρ. (Use Propositions (4.1.2) and (3.2.12).)

.4 Let A and B be nonempty subsets of a normed space X, and define

$$A + B = \{x + y : x \in A,\ y \in B\}.$$

Prove that

(i) if A is open, then $A + B$ is open;

(ii) if A is compact and B is closed, then $A + B$ is closed.

Need $A + B$ be closed when A and B are closed?

.5 Recall that a subset C of a linear space is said to be *convex* if $tx + (1-t)y \in C$ whenever $x, y \in C$ and $0 \leq t \leq 1$. Prove that the closure of a convex subset of a normed space is convex.

.6 Let C be a nonempty closed convex subset of a normed space X, x_0 a point of $X \backslash C$, and r a positive number such that $C \cap B(x_0, r)$ is empty. Prove that $C + B(0, r)$ is open and convex, and that $x_0 \notin C + B(0, r)$.

.7 Let C be a nonempty convex subset of $\mathbf{R}^n$, and x_0 a point of $\overline{C} \backslash C$. Prove that each open ball with centre x_0 intersects the complement of $\overline{C}$. (First consider the case where C has a nonempty interior.) Does the conclusion hold if we drop the hypothesis that C is convex?

In a tribute to one of the founders of functional analysis, the Polish mathematician Stefan Banach (1892–1945), a complete normed linear space is called a *Banach space*. Among examples of Banach spaces are

- Euclidean n–space $\mathbf{R}^n$ (Exercise (4.1.1: 4));
- $\mathcal{B}(X, \mathbf{F})$ where X is a nonempty set and the norm is the sup norm (Exercise (4.1.1: 5));
- certain spaces of continuous or integrable functions that we consider in later sections of this chapter.

(4.1.6) Exercises

.1 Let c_0 be the real vector space (with termwise algebraic operations) consisting of all infinite sequences $x = (x_n)_{n=1}^\infty$ in $\mathbf{R}$ that converge to 0. For each $x \in c_0$ write

$$\|x\| = \sup_{n \geq 1} |x_n| .$$

Prove that this defines a norm on c_0 with respect to which c_0 is a separable Banach space. (For the second part, consider a Cauchy sequence $(x_k)_{k=1}^\infty$ in c_0, where for each k, $x_k = (x_{k,n})_{n=1}^\infty$. Show that for each n, $(x_{k,n})_{k=1}^\infty$ is a Cauchy sequence in $\mathbf{R}$. Denoting its limit by ξ_n, show that $(\xi_n)_{n=1}^\infty$ belongs to c_0 and is the limit of the sequence $(x_k)_{k=1}^\infty$ in the space c_0.)

.2 Let l_1 denote the space of all sequences of real numbers such that the corresponding series is absolutely convergent, and for each $x = (x_n)_{n=1}^\infty \in l_1$ write

$$\|x\|_1 = \sum_{n=1}^{\infty} |x_n| .$$

Prove that this defines a norm on l_1 with respect to which l_1 is a separable Banach space.

.3 Let l_∞ denote the space of all bounded sequences of real numbers, and for each $x = (x_n)_{n=1}^\infty \in l_\infty$ write

$$\|x\|_\infty = \sup_{n \geq 1} |x_n| \,.$$

Prove that this defines a norm on l_∞ with respect to which l_∞ is a Banach space.

.4 Prove that if X is a nonempty set, then $\mathcal{B}(X, \mathbf{F})$, taken with the supremum norm, is a Banach space.

We now sketch how any normed space X can be embedded as a dense subspace of a Banach space. Defining

$$\begin{aligned} \phi_x(y) &= \|x - y\| \quad (x, y \in X), \\ Y &= \{\phi_0 + f : f \in \mathcal{B}(X, \mathbf{R})\}, \\ d(F, G) &= \sup\{|F(x) - G(x)| : x \in X\} \quad (F, G \in Y), \end{aligned}$$

we recall from Exercise (3.2.10: 8) that (Y, d) is a complete metric space, that $x \mapsto \phi_x$ is an isometric mapping of X onto a subset Z of Y, and that the closure $\widehat{X}$ of Z is a complete subspace of Y. We transport the algebraic structure from X to Z by defining

$$\begin{aligned} \phi_x + \phi_y &= \phi_{x+y}, \\ \lambda\phi_x &= \phi_{\lambda x} \end{aligned}$$

for all $x, y \in X$ and $\lambda \in \mathbf{F}$. Then

$$\|\phi_x\| = d(\phi_x, \phi_0) = \|x\|$$

defines a norm on Z whose associated metric is the one induced by d. Using Exercise (4.1.5: 3), we can extend the operations of addition and multiplication–by–scalars uniquely from X (identified with its image under the mapping $x \mapsto \phi_x$) to $\widehat{X}$, thereby making $\widehat{X}$ a Banach space in which there is a dense linear subspace isometric and algebraically isomorphic to X.

In practice, we normally forget about the mapping $x \mapsto \phi_x$ and regard X simply as a dense subspace of $\widehat{X}$, which we call the *completion* of the normed space X.

(4.1.7) Exercise

Fill in the details of the proof that the foregoing constructions provide $\widehat{X}$ with the structure of a Banach space and that $x \mapsto \phi_x$ is a norm–preserving algebraic isomorphism of X with a dense subspace of $\widehat{X}$.

Banach spaces form the natural abstract context for the notion of convergence of series. Given a sequence (x_n) of elements of a normed space X, we define the corresponding *series* $\sum_{n=1}^{\infty} x_n$ to be the sequence (s_n), where $s_n = \sum_{k=1}^{n} x_k$ is the nth *partial sum* of the series. The series $\sum_{n=1}^{\infty} x_n$ is said to be

- *convergent* if the sequence (s_n) converges to a limit s in X, called the *sum* of the series,
- *absolutely convergent* if the series $\sum_{n=1}^{\infty} \|x_n\|$ is convergent in $\mathbf{R}$,
- *unconditionally convergent* if $\sum_{n=1}^{\infty} x_{f(n)}$ converges for each permutation f of $\mathbf{N}^+$.

In the first case we write $\sum_{n=1}^{\infty} x_n = s$.

(4.1.8) Exercises

.1 Prove that a series $\sum_{n=1}^{\infty} x_n$ in a Banach space X converges if and only if for each $\varepsilon > 0$ there exists a positive integer N such that $\| \sum_{n=i+1}^{j} x_n \| < \varepsilon$ whenever $j > i \geq N$.

.2 Prove that an absolutely convergent series in a Banach space is unconditionally convergent. (See Exercise (1.2.17: 1).)

.3 Let X be a normed linear space, and suppose that each absolutely convergent series in X is convergent. Prove that X is a Banach space. (Given a Cauchy sequence (x_n) in X, choose $n_1 < n_2 < \cdots$ such that $\|x_i - x_j\| < 2^{-k}$ for all $i, j \geq n_k$. Then consider the series $\sum_{k=1}^{\infty} \left(x_{n_{k+1}} - x_{n_k} \right)$.)

.4 In the Banach space c_0 of Exercise (4.1.6: 1), for each positive integer n let x_n be the element with nth term $1/n$ and all other terms 0. Prove that the series $\sum_{n=1}^{\infty} x_n$ is unconditionally convergent but not absolutely convergent.

Exercises (1.2.17: 1 and 2) show that a series in $\mathbf{R}$ is unconditionally convergent if and only if it is absolutely convergent. Exercise (4.1.8: 4) shows that this need not be true if $\mathbf{R}$ is replaced by an infinite–dimensional Banach space. In fact, if every unconditionally convergent series in a Banach space X is absolutely convergent, then X is finite–dimensional; this is the *Dvoretsky–Rogers Theorem* (see [12], Chapter VI).

Let X be a normed space over $\mathbf{F}$, and S a linear subspace of X. Then

$$x \sim y \text{ if and only if } x - y \in S$$

defines an equivalence relation on X. The set of equivalence classes under this relation is written X/S and is called the *quotient space* of X by S. The *canonical map* φ of X onto X/S is defined by

$$\varphi(x) = \{x + s : s \in S\},$$

and maps each element of X to its equivalence class under $\sim$. We define operations of addition and multiplication–by–scalars on X/S by

$$\begin{aligned}\varphi(x) + \varphi(y) &= \varphi(x + y), \\ \varphi(\lambda x) &= \lambda\varphi(x).\end{aligned}$$

These definitions are sound: for if $x \sim x'$ and $y \sim y'$, then $x + y \sim x' + y'$ and $\lambda x \sim \lambda x'$. If S is a *closed* linear subspace of X, then

$$\|\varphi(x)\| = \rho\,(x, S) = \inf\{\|x - s\| : s \in S\}$$

defines a norm, called the *quotient norm*, on X/S. In that case we assume that X/S is equipped with the foregoing algebraic operations and with the quotient norm.

(4.1.9) Exercises

.1 Verify the claims made without proof in the preceding paragraph.

.2 Prove that if S is closed in X, then the canonical map $\varphi : X \to X/S$ is uniformly continuous on S.

(4.1.10) Proposition. *If S is a closed linear subspace of a Banach space X, then the quotient space X/S is a Banach space.*

Proof. Let φ be the canonical map of X onto X/S, and consider a sequence (x_n) in X such that $(\varphi(x_n))$ is a Cauchy sequence in X/S. Choose a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ of positive integers such that

$$\left\|\varphi(x_{n_{k+1}}) - \varphi(x_{n_k})\right\| < 2^{-k} \quad (k \geq 1).$$

Setting $s_1 = 0$, we construct inductively a sequence (s_k) in S such that for each k,

$$\left\|(x_{n_{k+1}} - s_{k+1}) - (x_{n_k} - s_k)\right\| < 2^{-k}. \tag{1}$$

Indeed, having constructed elements $s_1, \ldots, s_k$ of S with the applicable properties, we have

$$\begin{aligned}&\inf\left\{\left\|x_{n_{k+1}} - (x_{n_k} - s_k) - s\right\| : s \in S\right\} \\ &= \inf\left\{\left\|x_{n_{k+1}} - x_{n_k} - s\right\| : s \in S\right\} \\ &= \left\|\varphi(x_{n_{k+1}} - x_{n_k})\right\| \\ &= \left\|\varphi(x_{n_{k+1}}) - \varphi(x_{n_k})\right\| < 2^{-k},\end{aligned}$$

so there exists $s_{k+1} \in S$ such that (1) holds.

We now see from (1) that $(x_{n_k} - s_k)_{k=1}^{\infty}$ is a Cauchy sequence in the Banach space X; whence it converges to a limit z in X. By Exercise (4.1.9: 2),

$$\varphi(x_{n_k}) = \varphi(x_{n_k}) - \varphi(s_k) = \varphi(x_{n_k} - s_k) \rightarrow \varphi(z) \text{ as } k \rightarrow \infty.$$

Thus the Cauchy sequence $(\varphi(x_n))$ has a convergent subsequence. It follows from Exercise (3.2.10: 3) that $(\varphi(x_n))$ itself converges in $X \backslash S$. □

4.2 Linear Mappings and Hyperplanes

In the context of normed spaces, the important mappings are not just continuous but also preserve the algebraic structure.

Recall that a mapping u between vector spaces X, Y is *linear* if

$$u(x+y) = u(x) + u(y)$$

and

$$u(\lambda x) = \lambda u(x)$$

whenever $x, y \in X$ and $\lambda \in \mathbf{F}$. If $Y = \mathbf{F}$, then u is called a *linear functional* on X. Examples of linear mappings are

- the mapping $x \mapsto Ax$ on $\mathbf{F}^n$, where A is an n–by–n matrix over $\mathbf{F}$;
- the Lebesgue integral, regarded as a mapping of $L_1(\mathbf{R})$ into $\mathbf{R}$ (see Exercise (4.1.1: 6));
- the mapping $(x_n)_{n=1}^{\infty} \mapsto x_1$ of c_0 into $\mathbf{R}$ (see Exercise (4.1.6: 1));
- the canonical mapping of a normed space X onto the quotient space X/S, where S is a closed subspace of X;
- the mapping $x \mapsto \phi_x$ of a normed space onto a dense subspace of its completion (see page 179).

Here is the fundamental result about the continuity of linear mappings between normed spaces.

(4.2.1) Theorem. *The following are equivalent conditions on a linear mapping of a normed space X into a normed space Y.*

(i) *u is continuous at* 0.

(ii) *u is continuous on X.*

(iii) *u is uniformly continuous on X.*

(iv) *u is bounded on the unit ball of X.*

(v) *u is bounded on each bounded subset of X.*

(vi) *There exists a positive number c, called a* bound *for u, such that $\|u(x)\| \leq c\,\|x\|$ for all $x \in X$.*

Proof. Suppose that u is continuous at 0. Then there exists $r > 0$ such that

$$\|u(x)\| = \|u(x) - u(0)\| \leq 1$$

whenever $\|x\| \leq r$. For each nonzero $t \in \mathbf{F}$ with $\|tx\| \leq 1$ we have $\|rtx\| \leq r$ and therefore

$$\|u(x)\| = r^{-1}\,|t|^{-1}\,\|u(rtx)\| \leq r^{-1}\,|t|^{-1}.$$

It follows from Exercise (4.1.1: 3) that $\|u(x)\| \leq r^{-1}\,\|x\|$ for all $x \in X$. Hence (i) implies (vi).

It is clear that (vi) $\Rightarrow$ (v) $\Rightarrow$ (iv). Next, suppose that there exists $c > 0$ such that $\|u(x)\| \leq c$ whenever $\|x\| \leq 1$. Since

$$\|u(x)\| = \|x\|\left\|u\left(\|x\|^{-1}\,x\right)\right\| \leq c\,\|x\| \quad (x \neq 0)$$

and $u(0) = 0$, we see that (vi) holds, with c a bound for u. We now have

$$\|u(x-y)\| \leq c\,\|x-y\| \quad (x, y \in X),$$

from which it follows that u is uniformly continuous on X. Thus (iv) $\Rightarrow$ (vi) $\Rightarrow$ (iii).

Finally, it is obvious that (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). □

In view of property (v) of Proposition (4.2.1), we commonly refer to a continuous linear mapping between normed spaces X, Y as a *bounded linear mapping* on X. We define the *norm* of such a mapping by

$$\|u\| = \sup\{\|u(x)\| : x \in X,\ \|x\| \leq 1\}. \tag{1}$$

The argument used to prove that (iv) $\Rightarrow$ (vi) in the last proof shows that

$$\|u(x)\| \leq \|u\|\ \|x\| \quad (x \in X).$$

In Exercise (4.2.2: 11) you will prove that equation (1) defines a norm on the linear space $L(X, Y)$ of all bounded linear mappings $u : X \to Y$, and that if Y is a Banach space, then so is $L(X, Y)$. The Banach space $L(X, \mathbf{F})$, consisting of all bounded linear functionals from X into its ground field $\mathbf{F}$, is called the *dual space*, or simply the *dual*, of X, and is denoted by X^*. The interplay between a Banach space and its dual is one of the most significant themes of functional analysis, so we spend some time later in this chapter and in Chapter 6 identifying the duals of certain important Banach spaces.

Two norms $\|\cdot\|$, $\|\cdot\|'$ on a vector space X are said to be *equivalent* if both the identity mapping from $(X, \|\cdot\|)$ onto $(X, \|\cdot\|')$ and its inverse are continuous; since those mappings are linear, it follows from Proposition (4.2.1) that $\|\cdot\|$ and $\|\cdot\|'$ are equivalent norms on X if and only if there exist positive constants a, b such that $a\|x\| \leq \|x\|' \leq b\|x\|$ for all $x \in X$.

(4.2.2) Exercises

.1 Prove that a linear mapping $u : X \to Y$ between normed spaces is bounded if and only if there exists $c > 0$ such that $\|u(x)\| \leq c$ for all $x \in X$ with $\|x\| = 1$, and that we then have

$$\|u\| = \sup\{\|u(x)\| : x \in X, \|x\| = 1\}.$$

.2 Let u be a bounded linear mapping on a normed space X. Prove that

$$\|u\| = \inf\{c \geq 0 : \|u(x)\| \leq c\|x\| \text{ for all } x \in X\}.$$

.3 Show that any two of the three norms on $\mathbf{R}^n$ introduced in Exercise (4.1.1: 4) are equivalent.

.4 Let $\|\cdot\|$, $\|\cdot\|'$ be equivalent norms on a linear space X. Prove that if X is complete with respect to $\|\cdot\|$, then it is complete with respect to $\|\cdot\|'$.

.5 Let $X_1, \ldots, X_n$ be normed spaces, $X = X_1 \times \cdots \times X_n$, and Y a normed space. Let u be a *multilinear* mapping of X into Y—that is, a mapping linear in each of its n variables. Prove that u is continuous if and only if there exists a constant $c > 0$ (which we call a *bound* for u) such that

$$\|u(x_1, \ldots, x_n)\| \leq c\|x_1\|\|x_2\| \cdots \|x_n\|$$

for all $(x_1, \ldots, x_n) \in X$.

.6 Let X, Y be normed spaces, and $u : X \to Y$ a linear mapping such that for each sequence (x_n) in X converging to 0, the sequence $(u(x_n))$ is bounded in Y. Prove that u is continuous. (Let (x_n) be a sequence converging to 0 in X, reduce to the case where $\|x_n\| < 1/n^2$ for each n, and then consider the sequence (nx_n).)

.7 Prove that a linear mapping $u : X \to Y$ between normed spaces is bounded if and only if for each Cauchy sequence (x_n) in X, $(u(x_n))$ is a Cauchy sequence in Y.

.8 Let u be a continuous linear mapping of a normed space X into a Banach space Y, and let $\sum_{n=1}^{\infty} x_n$ be an absolutely convergent series in X. Prove that the series $\sum_{n=1}^{\infty} u(x_n)$ converges absolutely in Y.

.9 Recalling the Banach space l_1 of Exercise (4.1.6: 2), for each n let e_n be the element of l_1 with nth term 1 and all other terms 0. Show that to each bounded sequence (x_n) in a Banach space X there corresponds a unique bounded linear mapping $u : l_1 \to X$ such that $u(e_n) = x_n$ for each n.

Now let X be a separable Banach space, and (x_n) a dense sequence in the unit ball B of X. Define the bounded linear mapping $u : l_1 \to X$ as previously. Prove that u maps l_1 onto X. (Given $x \in B$, construct inductively $n_1 < n_2 < \cdots$ such that

$$\left\| 2^{k-1}(x - x_{n_1}) - \sum_{j=2}^{k} 2^{k-j} x_{n_j} \right\| < 2^{-k}$$

for each k.)

Thus every separable Banach space is the range of a bounded linear mapping on l_1. For further results of this type see [12].

.10 Let D be a dense linear subspace of a normed space X, and u a bounded linear mapping from D into a normed space Y. Prove that u extends to a bounded linear mapping, with the same norm, from X into Y. (First use Proposition (3.2.12).)

.11 Prove that if X, Y are normed spaces, then

$$\|u\| = \sup\{\|u(x)\| : x \in X,\ \|x\| \leq 1\}$$

defines a norm on $L(X, Y)$, and that if Y is complete, then $L(X, Y)$ is a Banach space with respect to this norm. (To establish the completeness, let (u_n) be a Cauchy sequence in $L(X, Y)$, and show that

$$u(x) = \lim_{n \to \infty} u_n(x)$$

defines an element u of $L(X, Y)$ such that $\|u - u_n\| \to 0$ as $n \to \infty$.)

.12 Let c_0 be the Banach space of Exercise (4.1.6: 1). For each positive integer n let e_n be the sequence whose nth term is 1 and which has all other terms equal to 0. Let u be a bounded linear functional on c_0. Prove that the series $\sum_{n=1}^{\infty} u(e_n)$ is absolutely convergent, and that the norm of u is $\sum_{n=1}^{\infty} |u(e_n)|$. Conversely, prove that if $\sum_{n=1}^{\infty} t_n$ is an absolutely convergent series of real numbers, then there is a unique bounded linear functional u on c_0 such that $u(e_n) = t_n$ for each n. Describe $u(x)$, where $x = (x_n)_{n=1}^{\infty} \in c_0$.

This example shows that the dual space c_0^* can be identified with the Banach space l_1 of Exercise (4.1.6: 2).

.13 Prove that l_1^* can be identified with the Banach space l_∞ of Exercise (4.1.6: 3).

.14 Prove the *Uniform Boundedness Theorem*: let $(T_i)_{i\in I}$ be a family of bounded linear mappings from a Banach space X into a normed space Y, such that $\{\|T_i x\| : i \in I\}$ is bounded for each $x \in X$; then $\{\|T_i\| : i \in I\}$ is bounded. (Suppose the contrary. Then construct sequences $(x_n)_{n=1}^\infty$ in X and $(i_n)_{n=1}^\infty$ in I such that for each n,

$$\begin{aligned}\|x_n\| &= 4^{-n},\\ \|T_{i_n} x_n\| &> \tfrac{2}{3}\|T_{i_n}\|\,\|x_n\|,\end{aligned}$$

and

$$\|T_{i_n}\| > 3 \times 4^n \left(n + \sup_{i\in I}\{\|T_i(x_1 + \cdots + x_{n-1})\|\}\right).$$

Taking $x = \sum_{n=1}^\infty x_n$, deduce the contradiction that $\|T_{i_n} x\| > n$ for each n.

This proof was published in [22]. A less elementary, but more standard, approach to the Uniform Boundedness Theorem is based on Baire's Theorem (6.3.1) and is discussed in Chapter 6.)

.15 A normed space X is said to be *uniformly convex* if it has the following property: for each $\varepsilon > 0$ there exists $\delta \in (0,1)$ such that $\|x - y\| < \varepsilon$ whenever $\|x\| \leq 1$, $\|y\| \leq 1$, and $\left\|\frac{1}{2}(x+y)\right\| > 1 - \delta$. Prove that if u is a bounded linear functional on a uniformly convex Banach space X, then there exists a unit vector $x \in X$ such that $|u(x)| = \|u\|$.

Recall that the *kernel*, or *null space*, of a linear mapping $u : X \to Y$ between vector spaces is the subspace

$$\ker(u) = u^{-1}(0) = \{x \in X : u(x) = 0\}$$

of X. We say that u is *nonzero* if $\ker(u) \neq X$—that is, if there exists $x \in X$ such that $u(x) \neq 0$; otherwise, u is said to be *zero*.

(4.2.3) Proposition. *A linear functional on a normed space X is continuous if and only if its kernel is closed in X.*

Proof. Let u be a linear functional on X, and $S = \ker(u)$. As $\{0\}$ is a closed subset of X, Proposition (3.2.2) shows that if u is continuous, then S is closed in X. Suppose, conversely, that S is closed in X. Since the zero linear functional is certainly continuous, we may assume that there exists $a \in X$ such that $u(a) = 1$. Then $0 \notin a + S$. On the other hand, by Lemma (4.1.4), $a + S$ is closed in X, so its complement is open. Hence there exists $r > 0$ such that $x \notin a + S$ whenever $\|x\| \leq r$. Suppose that $\|x\| \leq r$ and $|u(x)| > 1$, and let $y = u(x)^{-1}x$. Then $\|y\| \leq r$, so $y \notin a + S$. On the other hand,

$$u(y - a) = u(x)^{-1}u(x) - 1 = 0,$$

so $y - a \in S$, and therefore

$$y = a + (y - a) \in a + S.$$

This contradiction shows that $|u(x)| \leq 1$ whenever $\|x\| \leq r$. It follows from Proposition (4.2.1) that u is continuous. □

As we show in a moment, nonzero linear functionals on a normed space X are associated with certain subspaces of X which we now define.

A subspace H of a vector space X is called a *hyperplane* if

— $X \backslash H$ is nonempty, and

— for each $a \in X \backslash H$ and each $x \in X$ there exists a unique pair $(t, y) \in \mathbf{F} \times H$ such that $x = ta + y$.

This expression of the element x is called its *representation* relative to the pair (H, a) consisting of the hyperplane H and the element a of $X \backslash H$.

(4.2.4) Proposition. *The kernel of a nonzero linear functional on a normed space X is a hyperplane in X. Conversely, if H is a hyperplane in X and $a \notin H$, then there exists a unique linear functional u on X such that* $\ker(u) = H$ *and* $u(a) = 1$.

Proof. First let u be a nonzero linear functional on X. If $a \notin \ker(u)$ and $x \in X$, then, using the linearity of u, we easily verify that $x = ta + y$, with $t \in \mathbf{F}$ and $y \in \ker(u)$, if and only if $t = u(x)/u(a)$. Hence $\ker(u)$ is a hyperplane.

Conversely, let H be a hyperplane in X, and let $a \notin H$. For each $x \in X$ there exists a unique pair (t, y) in $\mathbf{F} \times H$ such that $x = ta + y$. Setting $u(x) = t$ and $f(x) = y$, we define functions $u : X \to \mathbf{F}$ and $f : X \to H$. If also $x' \in X$, then

$$x + x' = (u(x) + u(x'))\, a + f(x) + f(x'),$$

where $u(x) + u(x') \in \mathbf{F}$ and (as H is a linear subset of X) $f(x) + f(x') \in H$; the uniqueness of the representation of a given element of X relative to (H, a) ensures that $u(x + x') = u(x) + u(x')$. Similar uniqueness arguments show that $u(\lambda x) = \lambda u(x)$ whenever $\lambda \in \mathbf{F}$ and $x \in X$, and that $u(a) = 1$. In particular, it follows that u is a linear functional on X. Moreover, $u(x) = 0$ if and only if $x = f(x) \in H$; so $\ker(u) = H$. It remains to prove that u is the unique linear functional on X which takes the value 1 at a and has kernel H. But if v is another such linear functional on X, then for each $x \in X$ we have

$$\begin{aligned} v(x) &= v\left(u(x)a + f(x)\right) \\ &= u(x)v(a) + v(f(x)) \\ &= u(x)\,1 + 0 \\ &= u(x). \quad \square \end{aligned}$$

(4.2.5) Exercises

.1 Let H be a hyperplane in a normed space X, $a \in X \backslash H$, and $\alpha \in \mathbf{R}$. Prove that there exists a unique linear functional u on X such that $a + H = \{x \in X : u(x) = \alpha\}$.

.2 Let u be a nonzero bounded linear functional on a normed space X, and $H = \ker(u)$. Show that $\rho(x, H) = \|u\|^{-1} |u(x)|$ for each $x \in X$.

.3 A *translated hyperplane*[1] in a normed space X is a subset of the form $v + H$ where H is a hyperplane in X and $v \in X$. Prove that a translated hyperplane is closed if and only if its complement has a nonempty interior.

.4 Let K be a subset of a normed space X, and u a linear functional on X. For each $\alpha \in \mathbf{R}$ the translated hyperplane

$$H_\alpha = \{x \in X : u(x) = \alpha\}$$

is called a *hyperplane of support* for K if

— there exists $x_0 \in K$ such that $u(x_0) = \alpha$, and

— either $u(x) \geq \alpha$ for all $x \in K$ or $u(x) \leq \alpha$ for all $x \in K$.

Prove that if K is compact and is not contained in any H_α, then for exactly two real numbers α, H_α is a hyperplane of support for K. (Consider the set $\left\{t \in \mathbf{R} : u^{-1}(t) \cap K \neq \emptyset\right\}$.)

[1] Some authors use the term “hyperplane” for a translated hyperplane.

4.3 Finite–Dimensional Normed Spaces

Before studying some of the more important infinite–dimensional spaces in analysis, we devote a section to the major analytic properties of finite–dimensional spaces. We begin by showing that for any positive integer n, any normed space of dimension n over $\mathbf{F}$ can be identified with the product space $\mathbf{F}^n$.

(4.3.1) Proposition. *If X is an n–dimensional normed space with basis $\{e_1, \ldots, e_n\}$, then*

$$(\xi_1, \ldots, \xi_n) \mapsto \sum_{i=1}^{n} \xi_i e_i$$

is a one–one bounded linear mapping of the product space $\mathbf{F}^n$ onto X with a bounded linear inverse.

Proof. Let f denote the mapping in question. It is easy to verify that f is one–one and maps $\mathbf{F}^n$ onto X, and that both f and f^{-1} are linear. Let

$$c = \max_{1 \leq i \leq n} \|e_i\|.$$

The inequalities

$$\left\| \sum_{i=1}^{n} \xi_i e_i \right\| \leq \sum_{i=1}^{n} |\xi_i| \, \|e_i\| \leq c \sum_{i=1}^{n} |\xi_i| \leq nc \max_{1 \leq i \leq n} |\xi_i|$$

show that f is bounded and therefore continuous. Let

$$S = \left\{ (\xi_1, \ldots, \xi_n) \in \mathbf{F}^n : \max_{1 \leq i \leq n} |\xi_i| = 1 \right\}.$$

Then S is closed (Exercise (4.3.2: 1)) and bounded, and is therefore compact (see Exercise (3.5.11: 6)). Now, the mapping $\xi \mapsto \|f(\xi)\|$ is continuous and (as $\{e_1, \ldots, e_n\}$ is a basis) maps S into $\mathbf{R}^+$; so, by Exercise (3.3.7: 2),

$$0 < r = \inf \{\|f(\xi)\| : \xi \in S\}.$$

If ξ is any nonzero element of $\mathbf{F}^n$, then, setting $\eta = \|\xi\|^{-1} \xi$, we have $\eta \in S$ and therefore

$$r \leq \|f(\eta)\| = \|\xi\|^{-1} \|f(\xi)\|.$$

Hence $\|\xi\| \leq r^{-1} \|f(\xi)\|$. Since this holds trivially when $\xi = 0$, we see that r^{-1} is a bound for the linear mapping f^{-1}. □

(4.3.2) Exercises

.1 Prove that the set S in the preceding proof is closed.

.2 Show that if X is n–dimensional with basis $\{e_1, \ldots, e_n\}$, then the mapping

$$\sum_{i=1}^{n} \xi_i e_i \mapsto \max_{1 \leq i \leq n} |\xi_i|$$

is a norm on X, and that X is complete with respect to this norm.

.3 Find an alternative proof that the mapping f^{-1} in the proof of Proposition (4.3.1) is continuous.

.4 Prove that any linear mapping from a finite–dimensional normed space into a normed space is bounded. Hence prove that any two norms on a given finite–dimensional linear space are equivalent.

(4.3.3) Proposition. *A finite–dimensional normed space is complete.*

Proof. Let X be a finite–dimensional normed space. We may assume that $X \neq \{0\}$, so that X has a basis $\{e_1, \ldots, e_n\}$. Let f be the mapping in Proposition (4.3.1), and let (x_n) be a Cauchy sequence in X. Then $\left(f^{-1}(x_n)\right)$ is a Cauchy sequence in $\mathbf{F}^n$ and therefore (see Exercise (3.5.11: 7)), converges to a limit $y \in \mathbf{F}^n$. Since f is continuous, (x_n) converges to $f(y) \in X$. □

(4.3.4) Corollary. *A finite–dimensional subspace of a normed space X is closed in X.*

Proof. This is an immediate consequence of Propositions (4.3.3) and (3.2.9). □

Our next result is surprisingly useful. We use it to simplify the proof of Theorem (4.3.6).

(4.3.5) Riesz's Lemma. *Let S be a closed subspace with a nonempty complement in a normed space X, and let $0 < \theta < 1$. Then there exists a unit vector $x \in X$ such that $\|x - y\| > \theta$ for each $y \in S$.*

Proof. Fix $x_0 \in X \backslash S$. By Exercise (3.1.10: 3),

$$0 < r = \rho(x_0, S) < \theta^{-1} r.$$

Choosing $s_0 \in S$ such that

$$r \leq \|x_0 - s_0\| < \theta^{-1} r,$$

let

$$x = \|x_0 - s_0\|^{-1} (x_0 - s_0).$$

Then $\|x\| = 1$. Also, for each $s \in S$,

$$s_0 + \|x_0 - s_0\| \, s \in S,$$

so

$$\|x_0 - s_0\| \, \|x - s\| = \|x_0 - (s_0 + \|x_0 - s_0\| \, s)\| \geq \rho(x_0, S) = r,$$

and therefore

$$\|x - s\| \geq \frac{r}{\|x_0 - s_0\|} > \theta. \quad \square$$

It follows from Riesz's Lemma that in an infinite–dimensional normed space X, if $0 < \theta < 1$, then there exists a sequence (x_n) of unit vectors such that $\|x_m - x_n\| > \theta$ whenever $m \neq n$ (see Exercise (4.3.7: 4)). This result can be improved in various ways. For example, in Chapter 6 we prove that in any infinite–dimensional normed space there exists a sequence (x_n) of unit vectors such that $\|x_m - x_n\| > 1$ whenever $m \neq n$. A much deeper result, due to Elton and Odell, says that if X is an infinite–dimensional normed space, then there exist $\varepsilon > 0$ and a sequence (x_n) of unit vectors in X such that $\|x_m - x_n\| \geq 1 + \varepsilon$ whenever $m \neq n$; see Chapter XIV of [12].

We now use Riesz's Lemma to provide a topological characterisation of finite–dimensional normed spaces.

(4.3.6) Theorem. *A normed space is finite–dimensional if and only if its unit ball is totally bounded, in which case that ball is compact.*

Proof. For simplicity, we take the case $\mathbf{F} = \mathbf{R}$. Let X be a normed space, and B its (closed) unit ball; we may assume that $X \neq \{0\}$. Suppose that X is finite–dimensional with basis $\{e_1, \ldots, e_n\}$, and let u be the one–one linear mapping $\sum_{i=1}^n \xi_i e_i \mapsto (\xi_1, \ldots, \xi_n)$ of X onto the product metric space $\mathbf{R}^n$. By Propositions (4.3.1) and (4.2.1), there exists $R > 0$ such that if $\|\sum_{i=1}^n \xi_i e_i\| \leq 1$, then

$$\|(\xi_1, \ldots, \xi_n)\| = \max_{1 \leq i \leq n} |\xi_i| \leq R. \tag{1}$$

By the Heine–Borel–Lebesgue Theorem (1.4.6) and Proposition (3.5.10), $[-R,R]^n$ is a compact subset of $\mathbf{R}^n$; it follows from Propositions (4.3.1) and (3.3.6) that $u^{-1}([-R,R]^n)$ is a compact subset of X. Since B is closed and, by (1), a subset of $u^{-1}([-R,R]^n)$, we see from Proposition (3.3.4) that B is compact.

Assume, conversely, that B is totally bounded. Construct a finite $\frac{1}{2}$–approximation F to B, let S be the finite–dimensional subspace of X generated by F, and suppose that $X \neq S$. Then, by Riesz's Lemma (4.3.5), there exists a unit vector $x \in X$ such that $\|x - s\| > \frac{1}{2}$ for all $s \in S$; but this is absurd, as $\|x - s\| < \frac{1}{2}$ for some $s \in F$. Hence, in fact, $X = S$. □

(4.3.7) Exercises

.1 Show that if a normed space X contains a totally bounded ball, then every closed ball in X is compact.

.2 Prove that a normed space X is finite–dimensional if and only if $\{x \in X : \|x\| = 1\}$ is compact.

.3 Prove that a normed space is locally compact if and only if it is finite–dimensional.

.4 Let X be an infinite–dimensional normed space. Use Riesz's Lemma to construct, inductively, a sequence (x_n) of unit vectors in X such that for each n,

(i) $x_1, \ldots, x_n$ are linearly independent, and

(ii) $\rho(x_{n+1}, X_n) \geq \frac{1}{2}$, where $X_n = \operatorname{span}\{x_1, \ldots, x_n\}$.

Hence prove that the unit ball of X is not compact.

This provides us with another proof that if the unit ball of a normed space is compact, then the space is finite–dimensional.

.5 Let X be a metric space, $x \in X$, and S a nonempty subset of X. A point $b \in S$ is called a *best approximation*, or a *closest point*, to x in S if $\rho(x, b) = \rho(x, S)$. Prove the *Fundamental Theorem of Approximation Theory*: if X is a finite–dimensional subspace of a normed space X, then each point of X has a best approximation in X. (See [10], [38], or [52] for further information about approximation theory, a major branch of analysis with many important practical applications.)

.6 Prove that any hyperplane in a finite–dimensional normed space is closed.

Now let X be the subspace of c_0 consisting of all sequences $(x_n)_{n=1}^{\infty}$ of real numbers such that $x_n = 0$ for all sufficiently large n. Show that

$$f((x_n)_{n=1}^{\infty}) = \sum_{n=1}^{\infty} n x_n$$

defines a linear functional $f : X \to \mathbf{R}$ whose kernel is not closed in X.

.7 Let S be a nonempty closed subset of $\mathbf{R}^N$, and K, B closed balls in $\mathbf{R}^N$ such that (i) $B \subset K$ and (ii) K intersects S in a single point ζ on the boundary of K. If B does not intersect the boundary of K, let ξ be the centre of B; otherwise, B must intersect the boundary of K in a single point, which we denote by ξ. For each positive integer n let

$$K_n = \tfrac{1}{n}(\xi - \zeta) + K.$$

Prove that for all sufficiently large n we have $B \subset K_n$ and $K_n \cap S = \emptyset$. Hence prove that there exists a ball K' that is concentric with K, has radius greater than that of K, and is disjoint from S. (For the first part, begin by showing that there exists a positive integer ν such that $B \subset K_n$ for all $n \geq \nu$. Then suppose that for each $n \geq \nu$ there exists $s_n \in K_n \cap S$. Show that there exists a subsequence $(s_{n_k})_{k=1}^{\infty}$ converging to ζ, and hence find k such that $s_{n_k} \in K \cap S$, a contradiction.)

A sequence in a normed linear space X is said to be *total* if it generates a dense linear subspace of X—that is, if the linear space consisting of all finite linear combinations of terms of the sequence is dense in X. In that case X is separable. To see this, let (a_n) be a total sequence in X, and let S be the set of all finite linear combinations $r_1 a_1 + \cdots + r_n a_n$ with each coefficient r_k rational. (By a *rational complex number* we mean a complex number whose real and imaginary parts are rational.) If $\lambda_1, \ldots, \lambda_n$ are in $\mathbf{F}$, then there exist rational elements $r_1, \ldots, r_n$ of $\mathbf{F}$ such that $\sum_{k=1}^{n} |\lambda_k - r_k| \, \|a_k\|$ is arbitrarily small; since

$$\left\| \sum_{k=1}^{n} \lambda_k a_k - \sum_{k=1}^{n} r_k a_k \right\| \leq \sum_{k=1}^{n} |\lambda_k - r_k| \, \|a_k\|$$

and (a_n) is total, it follows that S is dense in X; but S is countable.

We have the following converse.

(4.3.8) Proposition. *If X is an infinite–dimensional separable normed space, then it has a total sequence of linearly independent vectors.*

Proof. Let (a_n) be a dense sequence in X, and assume without loss of generality that $a_1 \neq 0$. We construct inductively a strictly increasing sequence $1 = n_1 < n_2 < \cdots$ of positive integers such that for each k,

(i) the vectors $a_{n_1}, \ldots, a_{n_k}$ are linearly independent, and

(ii) for $1 \leq m \leq n_k$, a_m is a linear combination of $a_{n_1}, \ldots, a_{n_k}$.

Indeed, if $a_{n_1}, \ldots, a_{n_k}$ have been constructed with properties (i) and (ii), we take n_{k+1} to be the smallest integer $m > n_k$ such that a_m does not belong to the subspace X_k of X generated by $\{a_{n_1}, \ldots, a_{n_k}\}$. (If no such integer exists, then, being closed by Proposition (4.3.4), X_k contains the closure of the subspace of X generated by the dense sequence (a_n), so $X = X_k$ is finite–dimensional—a contradiction.) It now follows from (ii) that the sequence $(a_{n_k})_{k=1}^{\infty}$ is total in X. □

(4.3.9) Exercises

.1 Prove that the Banach spaces c_0 and l_1 are separable.

.2 Show that the Banach space l_∞ is not separable. (Consider the set of elements of l_∞ whose terms belong to $\{0, 1\}$.)

4.4 The L_p Spaces

In this section we introduce certain infinite–dimensional Banach spaces of integrable functions that appear very frequently in many areas of pure and applied mathematics. For convenience, we call real numbers p, q *conjugate exponents* if $p > 1$, $q > 1$, and $1/p + 1/q = 1$.

We begin our discussion with an elementary lemma.

(4.4.1) Lemma. *If x, y are positive numbers and $0 < \alpha < 1$, then*

$$x^{\alpha} y^{1-\alpha} \leq \alpha x + (1 - \alpha) y.$$

Proof. Taking $u = x/y$, consider

$$f(u) = u^{\alpha} - \alpha u - 1 + \alpha.$$

We have $f'(u) = \alpha(u^{\alpha-1} - 1)$, which is positive if $0 < u < 1$ and negative if $u > 1$. Since $f(1) = 0$, it follows from Exercise (1.5.4: 7) that $f(u) \leq 0$ for all $u > 0$. This immediately leads to the desired inequality. □

(4.4.2) Proposition. *Let p, q be conjugate exponents, and f, g measurable functions on* **R** *such that $|f|^p$ and $|g|^q$ are integrable. Then fg is integrable, and* Hölder's inequality

$$\left| \int fg \right| \leq \left(\int |f|^p \right)^{1/p} \left(\int |g|^q \right)^{1/q} \tag{1}$$

holds.

Proof. We first note that if $\int |f|^p = 0$, then $|f|^p = 0$ almost everywhere; so $f = 0$, and therefore $fg = 0$, almost everywhere. Then fg is integrable, $\int fg = 0$, and (1) holds trivially, as it does also in the case where $\int |g|^q = 0$.

Thus we may assume that $\int |f|^p > 0$ and $\int |g|^q > 0$. We then have, almost everywhere,

$$\frac{|fg|}{\left(\int |f|^p\right)^{1/p} \left(\int |g|^q\right)^{1/q}} = \left(\frac{|f|^p}{\int |f|^p}\right)^{1/p} \left(\frac{|g|^q}{\int |g|^q}\right)^{1/q} \leq \frac{|f|^p}{p\int |f|^p} + \frac{|g|^q}{q\int |g|^q}$$

(where the last step uses Lemma (4.4.1)), so

$$|fg| \leq \left(\int |f|^p\right)^{1/p} \left(\int |g|^q\right)^{1/q} \left(\frac{|f|^p}{p\int |f|^p} + \frac{|g|^q}{q\int |g|^q}\right). \tag{2}$$

Now, fg is measurable and the right–hand side of (2) is integrable. Hence, by Proposition (2.3.1), fg is integrable and

$$\left|\int fg\right| \leq \int |fg| \leq \left(\int |f|^p\right)^{1/p} \left(\int |g|^q\right)^{1/q} \left(\frac{1}{p} + \frac{1}{q}\right),$$

from which (1) follows. □

(4.4.3) Proposition. *Let $p \geq 1$, and let f, g be measurable functions on $\mathbf{R}$ such that $|f|^p$ and $|g|^p$ are integrable. Then $|f+g|^p$ is integrable, and* Minkowski's inequality

$$\left(\int |f+g|^p\right)^{1/p} \leq \left(\int |f|^p\right)^{1/p} + \left(\int |g|^p\right)^{1/p}$$

holds.

Proof. Clearly, we may assume that $p > 1$. Now, $|f+g|^p$ is measurable, by Exercise (2.3.3: 5). Since

$$|f+g|^p \leq (2\max\{|f|, |g|\})^p \leq 2^p \left(|f|^p + |g|^p\right)$$

and the last function is integrable, it follows from Proposition (2.3.1) that $|f+g|^p$ is integrable. The functions $|f|$ and $|f+g|^{p-1}$ are measurable, by Exercise (2.3.3: 5), and

$$\left(|f+g|^{p-1}\right)^q = |f+g|^p \in L_1(\mathbf{R}).$$

Thus, by Proposition (4.4.2), $|f+g|^{p-1}|f|$ is integrable and

$$\int |f+g|^{p-1}|f| \leq \left(\int |f+g|^p\right)^{1-p^{-1}} \left(\int |f|^p\right)^{1/p}.$$

Similarly, $|f+g|^{p-1}|g|$ is integrable and

$$\int |f+g|^{p-1}|g| \leq \left(\int |f+g|^p\right)^{1-p^{-1}} \left(\int |g|^p\right)^{1/p}.$$

It follows that

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

from which we easily obtain Minkowski's inequality. □

(4.4.4) Exercises

.1 Prove Hölder's inequality

$$\left|\sum_{n=1}^{N} x_n y_n\right| \leq \left(\sum_{n=1}^{N} |x_n|^p\right)^{1/p} \left(\sum_{n=1}^{N} |y_n|^q\right)^{1/q}$$

and Minkowski's inequality

$$\left(\sum_{n=1}^{N} |x_n+y_n|^p\right)^{1/p} \leq \left(\sum_{n=1}^{N} |x_n|^p\right)^{1/p} + \left(\sum_{n=1}^{N} |y_n|^p\right)^{1/p}$$

for finite sequences $x_1, \dots, x_N$ and $y_1, \dots, y_N$ of real numbers.

.2 A sequence (x_n) of real numbers is called *p–power summable* if the series $\sum_{n=1}^{\infty} |x_n|^p$ converges. Prove that if (x_n) is p–power summable and (y_n) is q–power summable, where p, q are conjugate exponents, then

(i) $\sum_{n=1}^{\infty} x_n y_n$ is absolutely convergent, and

(ii) Hölder's inequality holds in the form

$$\left|\sum_{n=1}^{\infty} x_n y_n\right| \leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} \left(\sum_{n=1}^{\infty} |y_n|^q\right)^{1/q}.$$

Prove also that if (x_n) and (y_n) are both p–power summable, then so is (x_n+y_n), and Minkowski's inequality

$$\left(\sum_{n=1}^{\infty} |x_n+y_n|^p\right)^{1/p} \leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} + \left(\sum_{n=1}^{\infty} |y_n|^p\right)^{1/p}$$

holds.

.3 Let $p \geq 1$, and let l_p denote the set of all p–power summable sequences, taken with termwise addition and multiplication–by–scalars. Prove that

$$\left\|(x_n)_{n=1}^{\infty}\right\|_p = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}$$

defines a norm on l_p. (We define the normed space $l_p(\mathbf{C})$ of p–power summable sequences of complex numbers in the obvious analogous way.)

Let X be a measurable subset of $\mathbf{R}$, and $p \geq 1$. We define $L_p(X)$ to be the set of all functions f, defined almost everywhere on $\mathbf{R}$, such that f is measurable, f vanishes almost everywhere on $\mathbf{R}\backslash X$, and $|f|^p$ is integrable. Taken with the pointwise operations of addition and multiplication–by–scalars, $L_p(X)$ becomes a linear space. If we follow the usual practice of identifying two measurable functions that are equal almost everywhere, then

$$\|f\|_p = \left(\int |f|^p\right)^{1/p}$$

is a norm, called the *L_p–norm*, on $L_p(X)$. (We met the normed space $L_1(\mathbf{R})$ in Exercise (4.1.1: 6).)

When $X = [a, b]$ is a compact interval, we write $L_p[a, b]$ rather than $L_p([a, b])$.

(4.4.5) Exercises

In these exercises, X is a measurable subset of $\mathbf{R}$.

.1 Let X be integrable and $1 \leq r < s$. Prove the following.

(i) $L_s(X) \subset L_r(X)$. (Note that if $f \in L_s(X)$, then $|f|^r \in L_{s/r}(X)$.)

(ii) The linear mapping $f \mapsto f$ of $L_s(X)$ into $L_r(X)$ is bounded and has norm $\leq \mu(X)^{r^{-1}-s^{-1}}$.

.2 Let $1 \leq r \leq t \leq s < \infty$, $r \neq s$,

$$\alpha = \frac{t^{-1} - s^{-1}}{r^{-1} - s^{-1}}, \quad \beta = \frac{r^{-1} - t^{-1}}{r^{-1} - s^{-1}}.$$

and $f \in L_r(X) \cap L_s(X)$. Prove that $f \in L_t(X)$ and

$$\|f\|_t \leq \|f\|_r^{\alpha} \|f\|_s^{\beta}.$$

(Consider $|f|^{\alpha t} |f|^{\beta t}$.)

.3 Prove that the step functions that vanish outside X form a dense subspace of $L_p(X)$ for $p \geq 1$. (First consider the case where X is a compact interval.)

.4 Let p, q be conjugate exponents, and let $f, g \in L_p(X)$. Prove that if $1 < p < 2$, then

$$2\left(\|f\|_p^p + \|g\|_p^p\right)^{q-1} \geq \|f+g\|_p^q + \|f-g\|_p^q$$

and

$$\|f+g\|_p^p + \|f-g\|_p^p \geq 2\left(\|f\|_p^q + \|g\|_p^q\right)^{p-1},$$

and that the reverse inequalities hold if $p \geq 2$. (*Clarkson's inequalities.* Use Exercise (1.5.8: 10).)

.5 Use the preceding exercise to prove that if $p > 1$, then $L_p(X)$ is uniformly convex. (See Exercise (4.2.2: 15).)

(4.4.6) The Riesz–Fischer Theorem. *$L_p(X)$ is a Banach space for all $p \geq 1$. More precisely, if (f_n) is a Cauchy sequence in $L_p(X)$, then there exist $f \in L_p(X)$ and a subsequence $(f_{n_k})_{k=1}^{\infty}$ of (f_n) such that*

(i) $\lim_{n\to\infty} \|f - f_n\|_p = 0$, *and*

(ii) $f_{n_k} \to f$ *almost everywhere on X as $k \to \infty$.*

Proof. We illustrate the proof with the case $X = \mathbf{R}$ and $p > 1$. Given a Cauchy sequence (f_n) in $L_p(\mathbf{R})$, choose a subsequence $(f_{n_k})_{k=1}^{\infty}$ such that

$$\|f_m - f_n\|_p \leq 2^{-k} \quad (m, n \geq n_k).$$

Then

$$\left\|f_{n_{k+1}} - f_{n_k}\right\|_p \leq 2^{-k}.$$

Writing $q = p/(p-1)$, we see from Proposition (4.4.2) that for each positive integer N, $\left|f_{n_{k+1}} - f_{n_k}\right|$ is integrable over $[-N, N]$, and

$$\begin{aligned}\int \left|f_{n_{k+1}} - f_{n_k}\right| \chi_{[-N,N]} &\leq \left\|f_{n_{k+1}} - f_{n_k}\right\|_p \left(\int \chi_{[-N,N]}\right)^{1/q} \\ &\leq 2^{-k}(2N)^{1/q},\end{aligned}$$

so the series

$$\sum_{k=1}^{\infty} \int \left|f_{n_{k+1}} - f_{n_k}\right| \chi_{[-N,N]}$$

converges. It follows from Lebesgue's Series Theorem (Exercise (2.2.13: 4)) that there exists a set E_N of measure zero such that the series

$$\sum_{k=1}^{\infty} \left| f_{n_{k+1}}(x) - f_{n_k}(x) \right| \chi_{[-N,N]}(x)$$

converges for all $x \in \mathbf{R} \backslash E_N$, and the function $\sum_{k=1}^{\infty} \left| f_{n_{k+1}} - f_{n_k} \right| \chi_{[-N,N]}$ is integrable. Then

$$E = \bigcup_{N=1}^{\infty} E_N$$

is a set of measure zero, and

$$f(x) = \lim_{k \to \infty} f_{n_k}(x) = f_{n_1}(x) + \sum_{k=1}^{\infty} \left(f_{n_{k+1}}(x) - f_{n_k}(x) \right)$$

exists for all $x \in \mathbf{R} \backslash E$. The function f so defined is measurable, by Exercise (2.3.3: 4). Since

$$\|f_{n_k}\|_p \le \|f_{n_1}\|_p + \|f_{n_k} - f_{n_1}\|_p \le \|f_{n_1}\|_p + \tfrac{1}{2}$$

for all k, we see from Fatou's Lemma (Exercise (2.2.13: 11)) that $|f|^p$ is integrable and hence that $f \in L_p(\mathbf{R})$. Moreover, if $n \ge n_i$, then by applying Fatou's Lemma to the sequence $(|f_{n_k} - f_n|)_{k=i}^{\infty}$ we see that

$$\|f - f_n\|_p = \lim_{k \to \infty} \|f_{n_k} - f_n\|_p \le 2^{-i}.$$

Hence $\lim_{n \to \infty} \|f - f_n\|_p = 0$. □

(4.4.7) Exercises

.1 Prove the Riesz–Fischer Theorem for a general measurable set $X \subset \mathbf{R}$. Prove it also in the case $p = 1$.

.2 Prove that the space l_p is complete for $p \ge 1$.

In order to establish an elegant characterisation of bounded linear functionals on $L_p(X)$, we first discuss those functions whose derivatives almost everywhere belong to $L_q(\mathbf{R})$.

(4.4.8) Lemma. *Let $I = [a, b]$ be a compact interval, $q > 1$, and G a real–valued function defined almost everywhere on $\mathbf{R}$ and vanishing outside I. Then the following conditions are equivalent.*

(i) *There exists $g \in L_q(\mathbf{R})$ such that $G'(x) = g(x)$ almost everywhere.*

(ii) *There exists* $M > 0$ *such that*

$$\sum_{k=1}^{n-1} \frac{|G(x_{k+1}) - G(x_k)|^q}{(x_{k+1} - x_k)^{q-1}} \leq M$$

whenever the points $x_k \in I$ *and* $x_1 < x_2 < \cdots < x_n$.

In that case, the smallest such M *is* $\int |g|^q$.

Proof. Writing $p = q/(q-1)$, suppose that (i) holds, and let $a \leq x_1 < x_2 < \cdots < x_n \leq b$. Applying Proposition (4.4.2) to the functions χ_I and $\chi_I g$, we have

$$\begin{aligned} |G(x_{k+1}) - G(x_k)| &= \left| \int_{x_k}^{x_{k+1}} g \right| \\ &\leq \left(\int_{x_k}^{x_{k+1}} \chi_I \right)^{1/p} \left(\int_{x_k}^{x_{k+1}} |g|^q \right)^{1/q} \\ &= (x_{k+1} - x_k)^{1/p} \left(\int_{x_k}^{x_{k+1}} |g|^q \right)^{1/q}. \end{aligned}$$

Hence

$$|G(x_{k+1}) - G(x_k)|^q \leq (x_{k+1} - x_k)^{q-1} \int_{x_k}^{x_{k+1}} |g|^q,$$

and therefore

$$\sum_{k=1}^{n-1} \frac{|G(x_{k+1}) - G(x_k)|^q}{(x_{k+1} - x_k)^{q-1}} \leq \sum_{k=1}^{n} \int_{x_k}^{x_{k+1}} |g|^q \leq \int |g|^q.$$

Thus (ii) holds, and the smallest M for which (ii) holds is at most $\int |g|^q$.

Now suppose that (ii) holds, and let $((a_k, b_k))_{k=1}^N$ be a finite sequence of nonoverlapping open subintervals of I. Applying Exercise (4.4.4: 1), we obtain

$$\begin{aligned} \sum_{k=1}^{N} |G(b_k) - G(a_k)| &= \sum_{k=1}^{N} \left(\frac{|G(b_k) - G(a_k)|}{(b_k - a_k)^{1/p}} \right) (b_k - a_k)^{1/p} \\ &\leq \left(\sum_{k=1}^{N} \frac{|G(b_k) - G(a_k)|^q}{(b_k - a_k)^{q-1}} \right)^{1/q} \left(\sum_{k=1}^{N} (b_k - a_k) \right)^{1/p} \\ &\leq M \left(\sum_{k=1}^{N} (b_k - a_k) \right)^{1/p}. \end{aligned}$$

Hence G is absolutely continuous. It follows from Exercise (2.2.17: 2) that there exists an integrable function g such that $G'(x) = g(x)$ almost everywhere.

For each positive integer n let

$$x_{n,k} = a + \frac{k}{2^n}(b-a) \quad (0 \le k \le 2^n)$$

and define a step function g_n by setting

$$g_n(x) = \begin{cases} \dfrac{G(x_{n,k+1}) - G(x_{n,k})}{x_{n,k+1} - x_{n,k}} & \text{if } x_{n,k} < x < x_{n,k+1} \\ 0 & \text{otherwise.} \end{cases}$$

Then $g(x) = \lim_{n\to\infty} g_n(x)$ almost everywhere—to be precise, on the complement of the union of

$$\{x_{n,k} : n \ge 1,\ 0 \le k \le 2^n\}$$

and the set of measure zero on which $G' = g$. Also,

$$\int |g_n|^q = \sum_{k=0}^{2^n} \frac{|G(x_{n,k+1}) - G(x_{n,k})|^q}{(x_{n,k+1} - x_{n,k})^{q-1}} \le M.$$

Applying Fatou's Lemma, we now see that $|g|^q$ is integrable and $\int |g|^q \le M$. Hence (ii) implies (i). Referring to the last sentence of the first part of the proof, we also see that $\int |g|^q$ is the smallest M for which (ii) holds. □

(4.4.9) Lemma. *Let u be a bounded linear functional on $L_p[a,b]$, and define*

$$G(x) = \begin{cases} u\left(\chi_{[a,x]}\right) & \text{if } a \le x \le b \\ 0 & \text{otherwise.} \end{cases}$$

Let $a \le x_1 < x_2 < \cdots < x_n \le b$, and let $f = \sum_{k=1}^{n-1} c_k \chi_{[x_k, x_{k+1}]}$. Then

$$u(f) = \sum_{k=1}^{n-1} c_k \left(G(x_{k+1}) - G(x_k)\right).$$

Moreover, if there exists $g \in L_q[a,b]$ such that $G' = g$ almost everywhere, then $u(f) = \int fg$.

Proof. We have

$$\begin{aligned} u(f) &= \sum_{k=1}^{n-1} c_k u(\chi_{[x_k, x_{k+1}]}) \\ &= \sum_{k=1}^{n-1} c_k \left(u\left(\chi_{[a,x_{k+1}]} - \chi_{[a,x_k]}\right)\right) \\ &= \sum_{k=1}^{n-1} c_k \left(G(x_{k+1}) - G(x_k)\right). \end{aligned}$$

Now suppose that $G' = g$ almost everywhere for some $g \in L_q[a,b]$. Then

$$u(f) = \sum_{k=1}^{n-1} c_k \int_{x_k}^{x_{k+1}} g = \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} c_k g = \int fg. \quad \square$$

We now show that if p, q are conjugate exponents, then the dual space L_p^* can be identified with L_q.

(4.4.10) Theorem. *Let p, q be conjugate exponents. Then for each $g \in L_q(X)$,*

$$u_g(f) = \int fg$$

defines a bounded linear functional on $L_p(X)$ with norm equal to $\|g\|_q$. Conversely, to each bounded linear functional u on $L_p(X)$ there corresponds a unique $g \in L_q(X)$ such that $u = u_g$.

Proof. If $g \in L_q(X)$, then by Lemma (4.4.2), u_g is well defined on $L_p(X)$. It is trivial that u_g is linear, and Hölder's inequality shows that $\|g\|_q$ is a bound for u_g. On the other hand, taking

$$f = (g^+)^{q/p} - (g^-)^{q/p},$$

we see that $f \in L_p(X)$ and

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

Hence $\|u_g\| = \|g\|_q$.

Now consider any bounded linear functional u on $L_p(X)$. To begin with, take the case where X is a compact interval $[a,b]$. If $u = u_g$ for some $g \in L_q(X)$, then $u\left(\chi_{[a,x]}\right) = \int_a^x g$ for each $x \in X$. This suggests that we define

$$G(x) = \begin{cases} u\left(\chi_{[a,x]}\right) & \text{if } x \in X \\ 0 & \text{if } x \in \mathbf{R}\backslash X \end{cases}$$

and try to show that $G' \in L_q(X)$ and that $u(f) = \int fG'$ for all $f \in L_p(X)$. To this end, let $a \le x_1 < x_2 < \cdots < x_n \le b$. Let ϕ be the step function that vanishes outside $[x_1, x_n]$ and at each x_i, and that takes the constant value

$$c_k = \frac{|G(x_{k+1}) - G(x_k)|^{q-1} \operatorname{sgn}(G(x_{k+1}) - G(x_k))}{(x_{k+1} - x_k)^{q-1}}$$

on (x_k, x_{k+1}), where

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Then, by Lemma (4.4.9),

$$\begin{aligned} \sum_{k=1}^{n-1} \frac{|G(x_{k+1}) - G(x_k)|^q}{(x_{k+1} - x_k)^{q-1}} &= u(\phi) \\ &\leq \|u\| \, \|\phi\|_p \\ &= \|u\| \left(\sum_{k=1}^{n-1} |c_k|^p \, (x_{k+1} - x_k) \right)^{1/p} \\ &= \|u\| \left(\sum_{k=1}^{n-1} \frac{|G(x_{k+1}) - G(x_k)|^q}{(x_{k+1} - x_k)^{q-1}} \right)^{1/p}, \end{aligned}$$

and therefore

$$\sum_{k=1}^{n-1} \frac{|G(x_{k+1}) - G(x_k)|^q}{(x_{k+1} - x_k)^{q-1}} \leq \|u\|^q .$$

Thus, by Lemma (4.4.8), there exists $g \in L_q(X)$ such that $G' = g$ almost everywhere and $\|g\|_q \leq \|u\|$. It follows from Lemma (4.4.9) that $u(f) = \int fg$ for each step function f that vanishes outside X. The set of such step functions is dense in the space $L_p(X)$, by Exercise (4.4.5: 3); moreover, the linear functionals u and $f \mapsto \int fg$ are bounded, and therefore uniformly continuous, on $L_p(X)$. Referring to Proposition (3.2.12), we conclude that $u = u_g$.

It remains to remove the restriction that X be a compact interval and to prove the uniqueness of g for a given u. This is left as an exercise. □

(4.4.11) Exercises

.1 Complete the proof of Theorem (4.4.10) by removing the restriction that X be a compact interval, and by proving the uniqueness of the function g for a given u.

.2 A measurable function f on $\mathbf{R}$ is said to be *essentially bounded* if there exists $M > 0$ such that $|f(x)| \leq M$ almost everywhere. Prove that

$$\|f\|_\infty = \inf \{M > 0 : |f(x)| \leq M \text{ almost everywhere}\}$$

defines a norm on the vector space L_∞ of all essentially bounded functions under pointwise operations, and that L_∞ is a Banach space with respect to this norm (where, as usual, we identify measurable functions that are equal almost everywhere).

The real number $\|f\|_\infty$ is called the *essential supremum* of the element f of L_∞.

.3 Let K be a compact subset of $\mathbf{R}$, and $f : K \to \mathbf{R}$ a continuous function. Extend f to $\mathbf{R}$ by setting $f(x) = 0$ if $x \in \mathbf{R}\backslash K$. Prove that $f \in L_\infty$ and that $\|f\|_\infty = \sup_{x \in K} |f(x)|$.

.4 Prove that if $f \in L_1$ and $g \in L_\infty$, then $fg \in L_1$ and Hölder's inequality holds in the form

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

.5 Prove that for each $g \in L_\infty$,

$$u_g(f) = \int fg$$

defines a bounded linear functional on L_1 with norm equal to $\|g\|_\infty$, and that every bounded linear functional on L_1 has the form u_g for a unique corresponding $g \in L_\infty$.

.6 Let $0 < p < 1$, and let L_p consist of all measurable functions f on $\mathbf{R}$ such that $|f|^p$ is integrable. Show that when we identify functions that are equal almost everywhere,

$$\rho_p(f, g) = \int |f - g|^p$$

defines a metric on L_p, and that (L_p, ρ_p) is a complete metric space. Show also that the only continuous linear mapping from L_p (with pointwise operations) to $\mathbf{R}$ is the zero mapping.

4.5 Function Spaces

Among the most important examples of Banach spaces are certain subsets of the space $\mathcal{B}(X, Y)$ of bounded functions from a nonempty set X into a Banach space Y, where the norm on $\mathcal{B}(X, Y)$ is the *sup norm*

$$\|f\| = \sup\{\|f(x)\| : x \in X\}.$$

Note that when X is a compact interval $[a,b]$, we usually write $\mathcal{B}[a,b]$ rather than $\mathcal{B}([a,b])$; we use similar notations without further comment in related situations.

A special case of the next result has appeared already (Exercise (4.1.6: 4)).

(4.5.1) Proposition. *If Y is a Banach space, then $\mathcal{B}(X,Y)$ is a Banach space.*

Proof. Let (f_n) be a Cauchy sequence in $\mathcal{B}(X,Y)$, and $\varepsilon > 0$. There exists N such that $\|f_m - f_n\| < \varepsilon$ for all $m, n \geq N$. For each $x \in X$ we have

$$\|f_m(x) - f_n(x)\| \leq \|f_m - f_n\| < \varepsilon$$

whenever $m, n \geq N$; so $(f_n(x))_{n=1}^{\infty}$ is a Cauchy sequence in Y. Since Y is complete,

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists; also, for all $m \geq N$,

$$\|f_m(x) - f(x)\| = \lim_{n\to\infty} \|f_m(x) - f_n(x)\| \leq \varepsilon. \tag{1}$$

Hence

$$\|f(x)\| \leq \|f_N(x)\| + \|f_N(x) - f(x)\| \leq \|f_N\| + \varepsilon.$$

Since $x \in X$ is arbitrary, we see that $f \in \mathcal{B}(X,Y)$. Also, it follows from (1) that $\|f_m - f\| \leq \varepsilon$ for all $m \geq N$. Since $\varepsilon > 0$ is arbitrary, (f_n) converges to f in $\mathcal{B}(X,Y)$. Hence $\mathcal{B}(X,Y)$ is complete. □

(4.5.2) Exercises

.1 Let Y be a finite–dimensional Banach space, and $\{e_1, \ldots, e_n\}$ a basis of Y. Prove that each $f \in \mathcal{B}(X,Y)$ can be written uniquely in the form $x \mapsto \sum_{k=1}^{n} f_k(x)e_k$ with each $f_k \in \mathcal{B}(X,\mathbf{F})$. Prove also that for each k, $f \mapsto f_k$ is a bounded linear mapping of $\mathcal{B}(X,Y)$ into $\mathcal{B}(X,\mathbf{F})$.

.2 Prove that the mapping $f \mapsto \sup_{t\in X} f(x)$ of $\mathcal{B}(X,\mathbf{R})$ into $\mathbf{R}$ is continuous.

.3 Let Y be a Banach space, and $\sum_{n=1}^{\infty} f_n$ a series in $\mathcal{B}(X,Y)$. Let $\sum_{n=1}^{\infty} c_n$ be a convergent series of nonnegative real numbers such that $\|f_n\| \leq c_n$ for each n. Show that $\sum_{n=1}^{\infty} f_n$ converges in the Banach space $\mathcal{B}(X,Y)$.

.4 Let $I = [a,b]$ be a compact interval, and $\mathcal{BV}(I)$ the linear space of all real–valued functions of bounded variation on I, with pointwise operations. Show that

$$\|f\|_{\mathrm{bv}} = |f(a)| + T_f(a,b)$$

defines a norm on $\mathcal{BV}(I)$, and that $\mathcal{BV}(I)$ is complete with respect to this norm. (For the second part recall Exercise (1.5.15: 6).)

Let $f, f_1, f_2, \dots$ be mappings of a nonempty set X into a metric space (Y, ρ). We say that the sequence (f_n)

- *converges simply* to f on X if for each $x \in X$ the sequence $(f_n(x))$ converges to $f(x)$ in Y;
- *converges uniformly* to f on X if

$$\sup_{x \in X} \rho(f_n(x), f(x)) \to 0 \text{ as } n \to \infty.$$

Clearly, uniform convergence implies simple convergence; but, as the next exercise shows, the converse is false.

(4.5.3) Exercises

.1 Give an example of a sequence (f_n) of continuous mappings from $[0,1]$ into $[0,1]$ that converges to 0 simply, but not uniformly, on $[0,1]$. (Consider a spike of height 1 travelling along the x–axis towards 0.)

.2 Let Y be a normed space. Prove that a sequence (f_n) in $\mathcal{B}(X, Y)$ converges to a limit f in the normed space $\mathcal{B}(X, Y)$ if and only if (f_n) converges uniformly to f on X.

.3 Let X be a compact metric space, and (f_n), (g_n) strictly increasing sequences of real–valued functions on X that converge simply to the same bounded function $f : X \to \mathbf{R}$. Show that for each m there exists n such that $f_m < g_n$ (that is, $f_m(x) < g_n(x)$ for all $x \in X$). Show also that we cannot omit "compact" from the hypotheses.

Now let (X, ρ) be a metric space, and Y a normed space. The set of all continuous mappings of X into Y is denoted by $\mathcal{C}(X, Y)$ or $\mathcal{C}_Y(X)$, and the set of all bounded continuous mappings of X into Y by $\mathcal{C}^\infty(X, Y)$ or $\mathcal{C}_Y^\infty(X)$; so

$$\mathcal{C}^\infty(X, Y) = \mathcal{B}(X, Y) \cap \mathcal{C}(X, Y).$$

If X is compact, then $\mathcal{C}^\infty(X, Y) = \mathcal{C}(X, Y)$, by Exercise (3.3.7: 1). In general, $\mathcal{C}^\infty(X, Y)$ is a linear subspace of $\mathcal{B}(X, Y)$; we consider it as a normed space, taken with the sup norm.

We usually write $\mathcal{C}^\infty(X)$ and $\mathcal{C}(X)$, respectively, instead of $\mathcal{C}^\infty(X, \mathbf{R})$ and $\mathcal{C}(X, \mathbf{R})$.

(4.5.4) Proposition. *If X is a metric space and Y a Banach space, then $\mathcal{C}^\infty(X, Y)$ is a closed, and therefore complete, subspace of $\mathcal{B}(X, Y)$.*

Proof. Let (f_n) be a sequence of elements of $\mathcal{C}(X,Y)$ converging to a limit f in $\mathcal{B}(X,Y)$. For each $\varepsilon > 0$ there exists N such that $\|f - f_n\| \leq \varepsilon/3$ whenever $n \geq N$. Given x_0 in X, construct a neighbourhood U of x_0 such that if $x \in U$, then $\|f_N(x) - f_N(x_0)\| \leq \varepsilon/3$. For each $x \in U$ we then have

$$\begin{aligned} \|f(x) - f(x_0)\| &\leq \|f(x) - f_N(x)\| + \|f_N(x) - f_N(x_0)\| \\ &\qquad + \|f_N(x_0) - f(x_0)\| \\ &\leq \|f - f_N\| + \tfrac{\varepsilon}{3} + \|f_N - f\| \\ &= \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ and $x_0 \in X$ are arbitrary, it follows that f is continuous on X. Thus $\mathcal{C}^\infty(X,Y)$ is closed in $\mathcal{B}(X,Y)$; whence, by Propositions (4.5.1) and (3.2.9), $\mathcal{C}^\infty(X,Y)$ is complete. □

Proposition (4.5.4) shows that *a uniform limit of bounded continuous functions is continuous.* Taken with Exercise (4.5.3: 1), this observation highlights the significance of the next theorem.

(4.5.5) Dini's Theorem. *Let X be a compact metric space, and (f_n) an increasing sequence in $\mathcal{C}(X)$ that converges simply to a continuous function f. Then (f_n) converges to f uniformly.*

Proof. Let $\varepsilon > 0$. For each $x \in X$ there exists N_x such that if $n \geq N_x$, then $0 \leq f(x) - f_n(x) \leq \varepsilon/3$. Since f and f_{N_x} are continuous, there exists an open neighbourhood U_x of x such that if $x' \in U_x$, then $|f(x) - f(x')| \leq \varepsilon/3$ and $|f_{N_x}(x) - f_{N_x}(x')| \leq \varepsilon/3$; whence

$$\begin{aligned} 0 &\leq f(x') - f_{N_x}(x') \\ &\leq |f(x) - f(x')| + f(x) - f_{N_x}(x) + |f_{N_x}(x) - f_{N_x}(x')| \\ &\leq \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} \\ &= \varepsilon. \end{aligned}$$

Since X is compact, there are finitely many points $x_1, \ldots, x_\nu$ of X such that the neighbourhoods U_{x_i} cover X. Setting

$$n_\varepsilon = \max\{N_{x_i} : 1 \leq i \leq \nu\},$$

consider $n \geq n_\varepsilon$. Given $x \in X$, choose i such that $x \in U_{x_i}$; then

$$0 \leq f(x) - f_n(x) \leq f(x) - f_{n_\varepsilon}(x) \leq f(x) - f_{N_{x_i}}(x) \leq \varepsilon.$$

Since $\varepsilon > 0$ and $x \in X$ are arbitrary, we conclude that (f_n) converges to f uniformly. □

(4.5.6) Exercises

.1 Show that "increasing" can be replaced by "decreasing" in Dini's Theorem.

.2 Give an alternative proof of Dini's Theorem using the sequential compactness of X.

.3 Let X be a metric space, Y a Banach space, and D a dense subset of X. Let (f_n) be a sequence of bounded continuous mappings of X into Y such that the restrictions of the functions f_n to D form a uniformly convergent sequence. Prove that (f_n) is uniformly convergent on X.

.4 Let X be a metric space, and Y a normed space. Prove that the mapping $(x, f) \mapsto f(x)$ is continuous on $X \times \mathcal{C}^\infty(X, Y)$.

.5 Let I be a compact interval in $\mathbf{R}$, and (f_n) a sequence of increasing real functions on I that converges simply in I to a continuous function f. Prove that f is increasing and that (f_n) converges to f uniformly on I.

.6 Let a, b be real numbers with $b > 0$, and let X be the set of all continuous mappings $f : [0, b] \to \mathbf{R}$ such that $f(0) = a$. Prove that X is complete with respect to the sup norm.

.7 Let I be a compact interval in $\mathbf{R}$, $x_0 \in I$, and $\alpha > 0$. Show that

$$\|f\|' = \sup \left\{ \mathrm{e}^{-\alpha|x - x_0|} |f(x)| : x \in I \right\}$$

defines a norm on $\mathcal{C}(I)$, and that $\mathcal{C}(I)$ is complete with respect to this norm.

.8 In the notation of Exercise (4.5.2: 4) prove that if a sequence (f_n) converges to a limit f with respect to the norm $\|\cdot\|_{\mathrm{bv}}$ on $\mathcal{BV}(I)$, then it converges to f with respect to the sup norm on $\mathcal{B}(I)$.

With $I = [0, 1]$ find a sequence in $\mathcal{BV}(I) \cap \mathcal{C}(I)$ that

(i) converges to a limit $f \in \mathcal{C}(I)$ with respect to the sup norm, and

(ii) is not a Cauchy sequence with respect to $\|\cdot\|_{\mathrm{bv}}$.

(Note Exercise (1.5.15: 4).)

Let X be a metric space, Y a normed space, and $\mathcal{F}$ a subset of $\mathcal{B}(X, Y)$. We say that $\mathcal{F}$ is

- *equicontinuous at* $a \in X$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\|f(x) - f(a)\| < \varepsilon$ whenever $f \in \mathcal{F}$ and $\rho(x, a) < \delta$,
- *equicontinuous* (*on* X) if it is equicontinuous at each point of X;

- *uniformly equicontinuous* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\|f(x) - f(y)\| < \varepsilon$ whenever $f \in \mathcal{F}$, $x \in X$, $y \in X$, and $\rho(x, y) < \delta$.

Clearly, uniform equicontinuity implies equicontinuity, and if $\mathcal{F}$ is equicontinuous at a, then each $f \in \mathcal{F}$ is continuous at a.

(4.5.7) Exercises

In these exercises, X, Y, and $\mathcal{F}$ are as in the first sentence of the last paragraph.

.1 Suppose that there exist constants $c > 0$ and $\lambda \geq 1$ such that

$$\|f(x) - f(y)\| \leq c\,\rho(x, y)^{\lambda}$$

for all $f \in \mathcal{F}$ and all $x, y \in X$. Show that $\mathcal{F}$ is uniformly equicontinuous.

.2 Let $\alpha > 0$, and let $\mathcal{F}$ be the set of all mappings $f : [0, 1] \to \mathbf{R}$ such that f' exists, is continuous, and has sup norm at most α. Show that $\mathcal{F}$ is uniformly equicontinuous.

.3 Show that $\{x^n : n \in \mathbf{N}\}$ is not equicontinuous at 1.

.4 Let (f_n) be an equicontinuous sequence of real–valued functions on X. Prove that the sequence

$$(f_1 \vee f_2 \vee \cdots \vee f_n)_{n=1}^{\infty}$$

is also equicontinuous.

.5 For each $\lambda \in L$ let $\mathcal{F}_\lambda \subset \mathcal{B}(X, Y)$ be equicontinuous at a. Prove that if L is a finite set, then $\bigcup_{\lambda \in L} \mathcal{F}_\lambda$ is equicontinuous at a. Give an example where L is an infinite set and $\bigcup_{\lambda \in L} \mathcal{F}_\lambda$ is not equicontinuous at a.

.6 Let (f_n) be a sequence of functions in $\mathcal{B}(X, Y)$ that converges simply to a function f and is equicontinuous at $a \in X$. Prove that f is continuous at a. Hence prove that the closure of an equicontinuous set in $\mathcal{C}^{\infty}(X, Y)$ is equicontinuous.

.7 Prove that if X is compact and $\mathcal{F} \subset \mathcal{C}(X, Y)$ is equicontinuous, then $\mathcal{F}$ is uniformly equicontinuous.

.8 Suppose that X is compact, and let (f_n) be a convergent sequence in $\mathcal{C}(X, Y)$. Prove that (f_n) is uniformly equicontinuous. (Let $f = \lim_{n\to\infty} f_n$. Given $\varepsilon > 0$, choose N such that $\|f - f_n\| < \varepsilon$ for all $n \geq N$. First find $\delta_1 > 0$ such that $\|f_n(x) - f_n(y)\| < 3\varepsilon$ whenever $\rho(x, y) < \delta_1$ and $n \geq N$.)

.9 Suppose that X is compact, and that (f_n) is an equicontinuous sequence in $\mathcal{C}(X,Y)$ that converges simply to a function $f : X \to Y$. Then f is continuous on X, by Exercise (4.5.7: 6). Show that (f_n) converges uniformly to f. (Given $\varepsilon > 0$, use Exercise (4.5.7: 7) to obtain δ as in the definition of "uniformly equicontinuous". Then cover X by finitely many balls of the form $B(x, \delta)$.)

.10 Let (f_n) be a sequence of continuous real–valued mappings on a compact interval I.

(i) Prove that if (f_n) is a Cauchy sequence in $\mathcal{C}(I)$, then it is a Cauchy sequence in $L_2(I)$.

(ii) Prove that if (f_n) is both equicontinuous and a Cauchy sequence relative to the L_2–norm, then (f_n) converges in $\mathcal{C}(I)$.

(For (ii) fix $t_0 \in I$ and $\varepsilon > 0$. Choose $\delta > 0$ such that if $t \in I$ and $|t - t_0| < \delta$, then $|f_n(t) - f_n(t_0)| < \varepsilon$ for all n. Let χ be the characteristic function of $I \cap [t_0 - \delta, t_0 + \delta]$, and show that

$$\int \chi(t) \left| f_m(t_0) - f_n(t_0) \right|^2 \, dt < c\delta\varepsilon^2$$

for some constant $c > 0$ and all sufficiently large m and n. Deduce that $(f_n(t_0))$ is a Cauchy sequence in I.)

Show that the equicontinuity hypothesis cannot be dropped in (ii).

If X is compact and Y is a Banach space, then we have a powerful characterisation of totally bounded subsets of $\mathcal{C}(X,Y)$.

(4.5.8) Ascoli's Theorem.[2] *Let X be a compact metric space, Y a normed space, and $\mathcal{F}$ a subset of $\mathcal{C}(X,Y)$. Then $\mathcal{F}$ is totally bounded if and only if*

(i) *$\mathcal{F}$ is equicontinuous and*

(ii) *for each $x \in X$,*

$$\mathcal{F}_x = \{f(x) : f \in \mathcal{F}\}$$

is a totally bounded subset of Y.

Proof. Assume first that $\mathcal{F}$ is totally bounded, and let $\varepsilon > 0$. Construct a finite ε–approximation $\{f_1, \ldots, f_N\}$ to $\mathcal{F}$. Then for each f in $\mathcal{F}$ there exists i such that $\|f - f_i\| \leq \varepsilon$. So for each $x \in X$ we have $\|f(x) - f_i(x)\| \leq \varepsilon$, from which it follows that $\{f_1(x), \ldots, f_N(x)\}$ is an ε–approximation to $\mathcal{F}_x$.

[2] This is also known as the *Ascoli–Arzelà Theorem.*

Hence $\mathcal{F}_x$ is totally bounded. On the other hand, choose $\delta > 0$ such that if $\rho(x,y) < \delta$, then $\|f_k(x) - f_k(y)\| \leq \varepsilon$ for each k. With f and f_i as in the foregoing, we have

$$\begin{aligned}\|f(x) - f(y)\| &\leq \|f(x) - f_i(x)\| + \|f_i(x) - f_i(y)\| \\ &\qquad + \|f_i(y) - f(y)\| \\ &\leq \varepsilon + \varepsilon + \varepsilon \\ &= \varepsilon.\end{aligned}$$

Hence $\mathcal{F}$ is equicontinuous.

Now assume, conversely, that conditions (i) and (ii) hold, and let ε be any positive number. For each $x \in X$ choose an open neighbourhood U_x of x such that $\|f(y) - f(x)\| < \varepsilon$ for each $f \in \mathcal{F}$ and each $y \in U_x$. Since X is compact, it can be covered by a finite family $\{U_{x_1}, \ldots, U_{x_m}\}$ of such neighbourhoods. Now, the sets $\mathcal{F}_{x_i}$ $(1 \leq i \leq m)$ are totally bounded, as is therefore their union K. Let $\{\xi_1, \ldots, \xi_n\}$ be a finite ε–approximation to K. On the other hand, let Φ be the finite set of all mappings of $\{1, \ldots, m\}$ into $\{1, \ldots, n\}$, and for each $\varphi \in \Phi$ let

$$S_\varphi = \left\{ f \in \mathcal{F} : \left\| f(x_i) - \xi_{\varphi(i)} \right\| \leq \varepsilon \ \ (1 \leq i \leq m) \right\}.$$

Then for each $f \in \mathcal{F}$ there exists $\varphi \in \Phi$ such that $f \in S_\varphi$. Since there are only finitely many of the sets S_φ (some of which may be empty), to complete the proof that $\mathcal{F}$ is totally bounded it suffices to prove that the diameter of each S_φ is at most 4ε. To this end, consider any $\varphi \in \Phi$ and any two elements f, g of S_φ. Given $x \in X$, choose i such that $x \in U_{x_i}$. Then $\|f(x) - f(x_i)\| \leq \varepsilon$ and $\|g(x) - g(x_i)\| \leq \varepsilon$. But $\left\|f(x_i) - \xi_{\varphi(i)}\right\| \leq \varepsilon$ and $\left\|g(x_i) - \xi_{\varphi(i)}\right\| \leq \varepsilon$; two applications of the triangle inequality show, in turn, that $\|f(x_i) - g(x_i)\| \leq 2\varepsilon$ and $\|f(x) - g(x)\| \leq 4\varepsilon$. Since $x \in X$ is arbitrary, it follows that $\|f - g\| \leq 4\varepsilon$; whence $\operatorname{diam}(S_\varphi) \leq 4\varepsilon$. □

(4.5.9) Exercises

.1 Let X be compact, and let (f_n) be a bounded equicontinuous sequence of mappings of X into Y. Prove that there exists a subsequence $(f_{n_k})_{k=1}^{\infty}$ such that $(f_{n_k}(x))_{k=1}^{\infty}$ converges for each $x \in X$. (Let (x_n) be a dense sequence in X. Setting $f_{0,n} = f_n$, construct sequences $(f_{i,n})_{n=1}^{\infty}$ $(i = 0, 1, \ldots)$ such that for all i and n,

(i) $(f_{i+1,n})$ is a subsequence of $(f_{i,n})$ and

(ii) $(f_{i,n}(x_i))_{n=1}^{\infty}$ converges in Y.

Then show that $(f_{n,n}(x))$ converges in Y for each $x \in X$.)

Use this result to give another proof of the "if" part of Ascoli's Theorem.

.2 Let $c_0, c_1 > 0$, and let S consist of all differentiable functions $f : [0,1] \to \mathbf{R}$ such that $\|f\| \leq c_0$ and $\|f'\| \leq c_1$. Prove that S is a compact subset of $\mathcal{C}[0,1]$.

.3 For each positive integer n and each $x \geq 0$ let

$$f_n(x) = \sin \sqrt{x + 4n^2\pi^2}.$$

Prove that

(i) (f_n) is equicontinuous on $\mathbf{R}^{0+}$;

(ii) (f_n) converges simply to 0 on $\mathbf{R}^{0+}$;

(iii) (f_n) is not totally bounded in $\mathcal{C}^\infty(\mathbf{R}^{0+})$.

(For the last part, show that if (f_n) were totally bounded, then it would converge to 0 uniformly on $\mathbf{R}^{0+}$.)

4.6 The Theorems of Weierstrass and Stone

In this section we follow a path from the famous, and widely applicable, approximation theorem of Weierstrass to its remarkable generalisation by Stone.

(4.6.1) The Weierstrass Approximation Theorem. *If I is a compact interval in $\mathbf{R}$, then the set of polynomial functions on I is dense in $\mathcal{C}(I)$.*

Thus for each $f \in \mathcal{C}(I)$ and each $\varepsilon > 0$ there exists a polynomial function p on I such that

$$\|f - p\| = \sup\{|f(x) - p(x)| : x \in I\} < \varepsilon.$$

In other words, each element of $\mathcal{C}(I)$ can be *uniformly approximated*, to any degree of accuracy, by polynomial functions.

We derive the Weierstrass Approximation Theorem as a simple consequence of a more general theorem about linear operators on $\mathcal{C}(I)$.

By a *positive linear operator* on $\mathcal{C}(X)$, where X is any metric space, we mean a linear mapping $L : \mathcal{C}(X) \to \mathcal{C}(X)$ such that $Lf \geq 0$ whenever $f \geq 0$.

In the remainder of this section we let p_k denote the monomial function $x \mapsto x^k$ on $\mathbf{R}$.

(4.6.2) Korovkin's Theorem. *Let I be a compact interval in $\mathbf{R}$, and (L_n) a sequence of positive linear operators on $\mathcal{C}(I)$ such that*

$$\lim_{n\to\infty} L_n p_k = p_k \quad (k = 0, 1, 2).$$

Then $L_n f \to f$ for all $f \in \mathcal{C}(I)$.

Proof. For each t in I let g_t be the element of $\mathcal{C}(I)$ defined by

$$g_t(x) = (t-x)^2 = t^2 p_0(x) - 2tp_1(x) + p_2(x).$$

The linearity of L_n implies that

$$L_n g_t = t^2 L_n p_0 - 2t L_n p_1 + L_n p_2\,;$$

whence

$$\begin{aligned} 0 &\leq (L_n g_t)(t) \\ &= t^2\left((L_n p_0)(t) - 1\right) - 2t\left((L_n p_1)(t) - t\right) + \left((L_n p_2)(t) - t^2\right) \\ &\leq t^2 \left\| L_n p_0 - p_0 \right\| + |2t| \left\| L_n p_1 - p_1 \right\| + \left\| L_n p_2 - p_2 \right\|. \end{aligned}$$

Since t^2 and $|2t|$ are bounded on I, our hypotheses ensure that $(L_n g_t)(t) \to 0$ uniformly on I as $n \to \infty$. We use this observation shortly.

Given $f \in \mathcal{C}(I)$ and $\varepsilon > 0$, and noting the Uniform Continuity Theorem (Corollary (3.3.13)), choose $\delta > 0$ such that if $x, y \in I$ and $|x-y| < \delta$, then $|f(x) - f(y)| < \varepsilon$. Fix t in I, and consider any $x \in I$. If $|t-x| \geq \delta$, then

$$|f(t) - f(x)| \leq 2\|f\| \leq 2\|f\| \frac{(t-x)^2}{\delta^2} = \frac{2}{\delta^2}\|f\|\, g_t(x).$$

It follows from this and our choice of δ that

$$|f(t) - f(x)| \leq \frac{2}{\delta^2}\|f\|\, g_t(x) + \varepsilon$$

for all x in I. Hence

$$-\varepsilon p_0 - \frac{2}{\delta^2}\|f\|\, g_t \leq f(t)p_0 - f \leq \varepsilon p_0 + \frac{2}{\delta^2}\|f\|\, g_t.$$

Since L_n is linear and positive, we have

$$-\varepsilon L_n p_0 - \frac{2}{\delta^2}\|f\|\ L_n g_t \leq f(t)\, L_n p_0 - L_n f \leq \varepsilon L_n p_0 + \frac{2}{\delta^2}\|f\|\ L_n g_t.$$

Hence

$$|f(t)(L_n p_0)(t) - (L_n f)(t)| \leq \varepsilon \|L_n p_0\| + \frac{2}{\delta^2}\|f\|\,(L_n g_t)(t).$$

Thus

$$\begin{aligned} &|f(t) - (L_n f)(t)| \\ &\leq |f(t) - f(t)\,(L_n p_0)(t)| + |f(t)\,(L_n p_0)(t) - (L_n f)(t)| \\ &\leq |f(t)|\ |1 - (L_n p_0)(t)| + \varepsilon \|L_n p_0\| + \frac{2}{\delta^2}\|f\|\ (L_n g_t)(t) \\ &\leq \|f\|\ \|p_0 - L_n p_0\| + \varepsilon\left(\|p_0\| + \|p_0 - L_n p_0\|\right) + \frac{2}{\delta^2}\|f\|\ (L_n g_t)(t). \end{aligned}$$

It now follows from our hypotheses, and the observation in the first paragraph of the proof, that $\|f - L_n f\| \to 0$ as $n \to \infty$. □

Proof of the Weierstrass Approximation Theorem. Without loss of generality, take $I = [0,1]$. For each $f \in \mathcal{C}(I)$ and each positive integer n define the corresponding *Bernstein polynomial* $B_n f$ by

$$(B_n f)(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f(k/n).$$

Then B_n is a positive linear operator on $\mathcal{C}(I)$. Routine calculations (with reference to the binomial theorem) show that $B_n p_0 = p_0$, that $B_n p_1 = p_1$, and that

$$(B_n p_2)(x) = \frac{n-1}{n} x^2 + \frac{1}{n} x \to x^2 \text{ as } n \to \infty.$$

It follows from Korovkin's theorem that $\lim_{n\to\infty} \|f - B_n f\| = 0$ for each $f \in \mathcal{C}(I)$. □

(4.6.3) Exercises

.1 Show that there is no loss of generality in our taking $I = [0,1]$ in the proof of the Weierstrass Approximation Theorem.

.2 Prove that $B_n p_0 = p_0$, that $B_n p_1 = p_1$, and that

$$(B_n p_2)(x) = \frac{n-1}{n} x^2 + \frac{1}{n} x \to x^2 \text{ as } n \to \infty.$$

.3 Let $f(x) = x^3$. Calculate $B_n(f)$, and hence prove that $B_n(f) \to f$ as $n \to \infty$.

.4 Prove that if p is a polynomial function of degree at most k on $[0,1]$, then so is $B_n(p)$ for each n. (Use induction on k.)

.5 Prove that there is only one positive linear operator L on $\mathcal{C}[0,1]$ such that $L(f) = f$ for all quadratic polynomial functions. (Use Korovkin's Theorem.)

.6 Suppose that f and f' belong to $\mathcal{C}(I)$, where I is a compact interval. Prove that for each $\varepsilon > 0$ there exists a polynomial function p such that $\|f - p\| < \varepsilon$ and $\|f' - p'\| < \varepsilon$. (Reduce to the case $I = [0,1]$. First find a polynomial q such that $\|f' - q\| < \varepsilon$.)

.7 Let I be a compact interval contained in $(0,1)$. For each $f \in \mathcal{C}(I)$ and each $n \in \mathbf{N}$ define $Q_n f$ on I by

$$(Q_n f)(x) = \sum_{k=0}^{n} \left\lfloor \binom{n}{k} f(k/n) \right\rfloor x^k (1-x)^{n-k},$$

where $\lfloor t \rfloor$ denotes the integer part of t. Prove that $\|B_n f - Q_n f\| \to 0$ and hence that, on I, f is the uniform limit of a sequence of polynomials with integer coefficients.

.8 A function $f : \mathbf{R} \to \mathbf{C}$ is said to be *periodic* if

$$\alpha = \min \{\tau > 0 : \forall t \in \mathbf{R}\ (f(t+\tau) = f(t))\}$$

exists and is positive, in which case α is called the *period* of f and f is also said to be α*–periodic*.

Prove *Korovkin's Theorem for* 2π*–periodic functions*: let $I = [-\pi, \pi]$, let

$$\mathcal{P}(I) = \{f \in \mathcal{C}(I) : f(-\pi) = f(\pi)\},$$

and let (L_n) be a sequence of positive linear operators on $\mathcal{P}(I)$ such that $L_n f \to f$ uniformly as $n \to \infty$ for $f = 1$, cos, and sin; then $L_n f \to f$ uniformly for all $f \in \mathcal{P}(I)$. (Write $z = \cos x$ and apply Theorem (4.6.2).)

.9 Although this exercise mentions Fourier series, it does not require any knowledge of Fourier analysis. Let $I = [-\pi, \pi]$. For each $f \in \mathcal{P}(I)$ the kth *partial sum of the Fourier series of* f is

$$(S_k f)(x) = \frac{a_0}{2} + \sum_{n=1}^{k} (a_n \cos nx + b_n \sin nx),$$

where for $n \geq 1$,

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos t \, \mathrm{d}t,$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin t \, \mathrm{d}t.$$

The nth *Cesàro mean of the Fourier series of* f is

$$G_n f = \frac{1}{n} \sum_{k=0}^{n-1} S_k f.$$

Prove that

$$(G_n f)(x) = \frac{1}{2n\pi} \int_{-\pi}^{\pi} f(t+x) \left(\frac{\sin \frac{1}{2}nt}{\sin \frac{1}{2}t} \right)^2 \mathrm{d}t,$$

and hence that G_n is a positive linear operator on $\mathcal{P}(I)$. Then prove that for each $f \in \mathcal{P}(I)$, $(G_n f)$ converges to f uniformly on I. (Use the preceding exercise.)

The following result, which was proved by Müntz in 1914, is an interesting generalisation of the Weierstrass Approximation Theorem.

> *Let* $(\lambda_n)_{n=1}^{\infty}$ *be a sequence in* $[1,\infty)$ *that diverges to* ∞. *Then* $\operatorname{span}\left\{1, x^{\lambda_1}, x^{\lambda_2}, \ldots\right\}$ *is dense in* $\mathcal{C}[0,1]$ *if and only if the series* $\sum_{n=1}^{\infty} 1/\lambda_n$ *diverges to* ∞.

An elementary proof of Müntz's Theorem can be found on pages 193–198 of [10]. Here is a very recent generalisation of Müntz's Theorem, due to P. Borwein and T. Erdélyi [6].

> *Let* $(\lambda_n)_{n=1}^{\infty}$ *be a sequence of distinct positive real numbers. Then* $\operatorname{span}\left\{1, x^{\lambda_1}, x^{\lambda_2}, \ldots\right\}$ *is dense in* $\mathcal{C}[0,1]$ *if and only if* $\sum_{n=1}^{\infty} \lambda_n \big/ \left(\lambda_n^2+1\right)$ *diverges to* ∞.

A different, more abstract, generalisation of Theorem (4.6.1) was given by Stone in 1937. In this generalisation we let X be a compact metric space; we consider $\mathcal{C}(X)$ as an algebra, with the pointwise operations of addition, multiplication, and multiplication–by–scalars; and we are interested in dense subalgebras of $\mathcal{C}(X)$. The property introduced in the next definition plays a key role in the proof of the Stone–Weierstrass theorem.

We say that a set $\mathcal{A}$ of real–valued functions on a metric space X *separates the points* of X if for each pair x, y of distinct points of X there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

(4.6.4) The Stone–Weierstrass Theorem. *Let* X *be a compact metric space, and* $\mathcal{A}$ *a subalgebra of* $\mathcal{C}(X)$ *that contains the constant functions and separates the points of* X. *Then* $\mathcal{A}$ *is dense in the Banach space* $\mathcal{C}(X)$.

The next two lemmas lead us to the proof of this theorem.

(4.6.5) Lemma. *Under the hypotheses of* Theorem (4.6.4), *if* $\varphi, \psi \in \mathcal{A}$, *then* $\varphi \wedge \psi$ *and* $\varphi \vee \psi$ *belong to the closure of* $\mathcal{A}$ *in* $\mathcal{C}(X)$.

Proof. Given $f \in \mathcal{A}$ and $\varepsilon > 0$, first apply the Weierstrass Approximation Theorem (4.6.1) to construct a polynomial function p such that

$$\left| \, |t| - p(t) \right| < \varepsilon \quad (0 \leq t \leq \|f\|).$$

Then

$$\left| \, |f(x)| - p \circ f(x) \right| < \varepsilon \quad (x \in X).$$

Since ε is arbitrary, we see that $|f| \in \overline{\mathcal{A}}$. The desired conclusion now follows by taking $f = |\varphi - \psi|$ and noting the identities

$$\varphi \wedge \psi = \tfrac{1}{2}\left(\varphi + \psi - |\varphi - \psi|\right),$$
$$\varphi \vee \psi = \tfrac{1}{2}\left(\varphi + \psi + |\varphi - \psi|\right). \quad \square$$

(4.6.6) Lemma. *Under the hypotheses of* Theorem (4.6.4), *for each pair x, y of distinct points of X and each pair a, b of real numbers, there exists $g \in \mathcal{A}$ such that $g(x) = a$ and $g(y) = b$.*

Proof. Since $\mathcal{A}$ separates the points of X, there exists $h \in \mathcal{A}$ such that $h(x) \neq h(y)$. Define

$$g(t) = a + (b - a)\,\frac{h(t) - h(x)}{h(y) - h(x)}.$$

Since $\mathcal{A}$ contains the constant functions and is an algebra, $g \in \mathcal{A}$. Clearly, $g(x) = a$ and $g(y) = b$. □

Proof of the Stone–Weierstrass Theorem. Given $f \in \mathcal{C}(X)$ and $\varepsilon > 0$, we need only show that there exists $h \in \overline{\mathcal{A}}$ such that $\|f - h\| < \varepsilon$. To this end, for each g in $\overline{\mathcal{A}}$ define

$$U(g) = \{x \in X : g(x) < f(x) + \varepsilon\},$$
$$L(g) = \{x \in X : g(x) > f(x) - \varepsilon\},$$

and note that, as g is continuous, $U(g)$ and $L(g)$ are open sets. It follows from Lemma (4.6.6) that for each $t \in X$ the sets $U(g)$, with $g \in \mathcal{A}$ and $g(t) = f(t)$, form an open cover of X. Since X is compact, we can extract a finite subcover $\{U(g_1), \ldots, U(g_n)\}$ of X. Define

$$h_t = g_1 \wedge g_2 \wedge \cdots \wedge g_n.$$

Then $h_t \in \overline{\mathcal{A}}$, by Lemma (4.6.5); $h_t(x) < f(x) + \varepsilon$ for each $x \in X$; and $h_t(t) = f(t)$, so $t \in L(h_t)$. Thus $(L(h_t))_{t \in X}$ is an open cover of X, from which we can extract a finite subcover, say $\{L(h_{t_1}), \ldots, L(h_{t_m})\}$. Then the function

$$h = h_{t_1} \vee h_{t_2} \vee \cdots \vee h_{t_m}$$

belongs to $\overline{\mathcal{A}}$, by Lemma (4.6.5); also,

$$f(x) - \varepsilon < h(x) < f(x) + \varepsilon$$

for each $x \in X$, so $\|f - h\| < \varepsilon$. □

It is simple to verify that the Weierstrass Approximation Theorem is the special case of the Stone–Weierstrass Theorem in which the algebra $\mathcal{A}$ consists of all polynomial functions on the compact interval I. Since the polynomial functions on I with rational coefficients form a countable dense set in this algebra $\mathcal{A}$, we see that $\mathcal{C}(I)$ is a separable metric space; this is a special case of the following more general corollary of the Stone–Weierstrass Theorem.

(4.6.7) Corollary. *If X is a compact metric space, then the Banach space $\mathcal{C}(X)$ is separable.*

Proof. Let (x_n) be a dense sequence in X, and for each positive integer k write

$$f_{n,k}(t) = \rho(t,\ X \backslash B(x_n, k^{-1})).$$

The set $\mathcal{S}$ of all functions of the form

$$f_{n_1,k_1}^{\alpha_1} f_{n_2,k_2}^{\alpha_2} \cdots f_{n_i,k_i}^{\alpha_i},$$

with each α_k a nonnegative integer, is countable. Hence the subspace $\mathcal{A}$ of $\mathcal{C}(X)$ generated by $\mathcal{S}$ is separable (see the paragraph immediately preceding Proposition (4.3.8)). So to complete the proof we need only show that $\mathcal{A}$ is dense in $\mathcal{C}(X)$. Since $\mathcal{A}$ is a subalgebra of $\mathcal{C}(X)$, if $\mathcal{S}$ separates the points of X we can invoke the Stone–Weierstrass Theorem. But for each pair x, y of distinct points of X we can choose n, k such that $x \in B(x_n, k^{-1})$ and $y \in X \backslash B(x_n, k^{-1})$. We then have $f_{n,k}(x) \neq 0$ (as $X \backslash B(x_n, k^{-1})$ is closed) and $f_{n,k}(y) = 0$. □

(4.6.8) Exercises

.1 Let f be a strictly increasing continuous function on $I = [0, 1]$. Prove that the subalgebra of $\mathcal{C}(I)$ generated by $\{1, f\}$ is dense in $\mathcal{C}(I)$.

.2 Let X be a compact metric space containing at least two points, and let $\mathcal{A}$ be the subalgebra of $\mathcal{C}(X)$ generated by the family

$$(t \mapsto \rho(t, x))_{x \in X}.$$

Prove that $\mathcal{A}$ is dense in $\mathcal{C}(X)$.

.3 Define a sequence (u_n) of polynomial functions on $\mathbf{R}$ inductively, as follows.

$$\begin{aligned} u_1(t) &= 0, \\ u_{n+1}(t) &= u_n(t) + \tfrac{1}{2}\left(t - u_n(t)^2\right). \end{aligned}$$

Prove that u_n maps $[0, 1]$ into $[0, 1]$, and that the sequence $(u_n(t))_{n=1}^{\infty}$ converges uniformly to $\sqrt{t}$ on $[0, 1]$. Hence prove that if $\mathcal{A}$ is a subalgebra of $\mathcal{C}[0, 1]$ and $f \in \mathcal{A}$, then $|f| \in \overline{\mathcal{A}}$.

This proof can be used to eliminate the reference to the Weierstrass Approximation Theorem from the proof of the Stone–Weierstrass Theorem, thereby making the former a genuine corollary of the latter.

.4 Let I be a compact interval in $\mathbf{R}$, and f a continuous mapping of the rectangle $I \times I$ into $\mathbf{R}$. Prove that for each $\varepsilon > 0$ there exists a polynomial

$$p(x, y) = \sum_{j,k=0}^{n} a_{j,k} x^j y^k$$

such that

$$\sup_{x,y\in I} |f(x,y) - p(x,y)| < \varepsilon.$$

.5 Prove the *Complex Stone–Weierstrass Theorem*: let X be a compact metric space, and $\mathcal{A}$ a subalgebra of $\mathcal{C}(X,\mathbf{C})$ that contains the constant functions, separates the points of X, and is closed under complex conjugation (so that $f^* \in \mathcal{A}$ whenever $f \in \mathcal{A}$, where $f^*(x) = f(x)^*$); then $\mathcal{A}$ is dense in $\mathcal{C}(X,\mathbf{C})$.

Can we remove the hypothesis that $\mathcal{A}$ is closed under complex conjugation?

.6 Use the Stone–Weierstrass Theorem to prove that each 2π–periodic continuous function $f : \mathbf{R} \to \mathbf{C}$ is a uniform limit of a sequence of *trigonometric polynomials* of the form

$$t \mapsto \sum_{n=-N}^{N} (a_n \sin nt + b_n \cos nt),$$

where the coefficients a_n, b_n belong to $\mathbf{C}$ (cf. Exercise (4.6.3: 9). Let $\mathcal{S}$ be the set of 2π-periodic elements of $\mathcal{C}^\infty(\mathbf{R},\mathbf{C})$. First note that

$$F(e^{it}) = f(t)$$

defines an isometric isomorphism of $\mathcal{S}$ with $\mathcal{C}(\mathbf{T},\mathbf{C})$, where

$$\mathbf{T} = \{z \in \mathbf{C} : |z| = 1\}$$

is the unit circle in the complex plane.)

.7 Let I be a compact interval, and $p \geq 1$. Prove that the Banach space $L_p(I)$ is separable. Prove also that $L_p(\mathbf{R})$ is separable. (First use Exercise (2.3.10) to prove that $\mathcal{C}(I)$ is dense in $L_p(I)$.)

4.7 Fixed Points and Differential Equations

In this final section of the chapter we show how various ideas that have appeared in the earlier sections are used to establish the existence of a solution φ of the first–order ordinary differential equation $\varphi'(x) = f(x,\varphi(x))$ on a compact interval. In order to do this, we first introduce a fundamental fixed–point theorem.

Let X and Y be metric spaces, and f a mapping of X into Y. We say that f satisfies a *Lipschitz condition*, or is a *Lipschitz mapping,* if there exists a constant $c > 0$ such that $\rho(f(x), f(y)) \leq c\,\rho(x,y)$ for all x, y in X; c is then called a *Lipschitz constant* for f, and f is said to be *Lipschitz* *of order c*. In the special case where $0 < c < 1$, f is called a *contraction mapping* of X into Y.

A Lipschitz map is uniformly continuous (Exercise (3.2.11: 6)).

A mapping of a metric space X into itself is called a *self–map*. By a *fixed point* of a self–map $f : X \to X$ we mean a point $\xi \in X$ such that $f(\xi) = \xi$.

(4.7.1) Exercises

.1 Let

$$p(x, y) = \sum_{j,k=0}^{n} a_{j,k} x^j y^k$$

be a polynomial function of two variables x, y. Prove that p satisfies a Lipschitz condition on any bounded subset of $\mathbf{R}^2$.

.2 Let f be a mapping of a metric space X into itself, and define the *iterates* of f inductively: for each $x \in X$,

$$f^n(x) = \begin{cases} x & \text{if } n = 0 \\ f(f^{n-1}(x)) & \text{if } n \in \mathbf{N}^+. \end{cases}$$

Prove that if, for some positive integer N, f^N has a unique fixed point ξ, then ξ is a fixed point of f, and f has no other fixed point.

Fixed points play an important role in many applications of mathematics, including the solution of differential equations and the existence of economic equilibria [51]. Many of these applications depend on our next result, *Banach's Contraction Mapping Theorem*.

(4.7.2) Theorem. *A contraction mapping of a nonempty complete metric space into itself has a unique fixed point.*

Proof. Let X be a nonempty complete metric space, f a contraction mapping of X into itself, and $c \in (0, 1)$ a Lipschitz constant for f. Choose x_0 in X, and define a sequence $(x_n)_{n=1}^{\infty}$ inductively by setting $x_n = f(x_{n-1})$. For each $k \geq 1$ we have

$$\begin{aligned} \rho(x_k, x_{k+1}) &= \rho(f(x_{k-1}), f(x_k)) \\ &\leq c\,\rho(x_{k-1}, x_k) \\ &\leq \cdots \\ &\leq c^k \rho(x_0, x_1). \end{aligned}$$

So if $m > n \geq 1$, then

$$\begin{aligned}\rho(x_n, x_m) &\leq \sum_{k=n}^{m-1} \rho(x_k, x_{k+1}) \\ &\leq \sum_{k=n}^{m-1} c^k \rho(x_0, x_1) \\ &\leq \rho(x_0, x_1) \sum_{k=n}^{\infty} c^k \\ &= \rho(x_0, x_1) \frac{c^n}{1-c} \to 0 \text{ as } n \to \infty.\end{aligned}$$

Hence (x_n) is a Cauchy sequence in the complete space X. Let ξ be its limit in X; then

$$\xi = \lim_{n\to\infty} x_n = \lim_{n\to\infty} f(x_{n-1}) = f(\xi).$$

Thus ξ is a fixed point of f.

Now suppose that η is a fixed point of f distinct from ξ. Then

$$\rho(\xi, \eta) = \rho(f(\xi), f(\eta)) \leq c\,\rho(\xi, \eta) < \rho(\xi, \eta),$$

which is absurd. Hence ξ is the unique fixed point of f in X. □

Recall that a self–map f of a metric space X is said to be *contractive* if $\rho(f(x), f(y)) < \rho(x, y)$ for all distinct x, y in X; and that, according to *Edelstein's Theorem* (Exercise (3.3.7: 4)), a contractive self–map of a compact metric space has a unique fixed point. The next exercise shows that we can neither remove the compactness hypothesis from Edelstein's Theorem nor replace the word "contraction" by "contractive" in the hypotheses of Banach's Contraction Mapping Theorem.

(4.7.3) Exercises

.1 Let B be the unit ball in the Banach space c_0, and for each positive integer n let e_n be the element of c_0 whose nth term is 1 and all of whose other terms are 0. Show that there is a unique linear mapping $u : c_0 \to c_0$ such that

$$u(e_n) = \left(1 - \frac{1}{2^n}\right) e_{n+1}$$

for each n. Then show that

$$v(x) = \tfrac{1}{2}(1 + \|x\|)\, e_1 + u(x)$$

defines a contractive map of B into itself such that $v(x) \neq x$ for each $x \in B$. (For the last part note that $\prod_{k=1}^{n}\left(1 - 2^{-k}\right) \geq 1 - \sum_{k=1}^{n} 2^{-k}$.)

.2 Let X, Y be Banach spaces over $\mathbf{F}$, U the open ball in X with centre 0 and radius a, and V the open ball in Y with centre 0 and radius b. Let $0 \le c < 1$, and let $\varphi : U \times V \to Y$ be a continuous mapping such that for all $x \in U$,

(i) $\|\varphi(x, y_1) - \varphi(x, y_2)\| \le c\,\|y_1 - y_2\|$ for all $y_1, y_2 \in V$, and

(ii) $\|\varphi(x, 0)\| < b\,(1 - c)$.

Show that there exists a unique mapping $f : U \to V$ such that $f(x) = \varphi(x, f(x))$ for all $x \in U$, and that f is continuous on U. (For each $x \in U$ define

$$\begin{aligned} f_0(x) &= 0, \\ f_{n+1}(x) &= \varphi(x, f_n(x)). \end{aligned}$$

Show that f_n is a continuous mapping of U into V, that the series $\sum_{n=1}^{\infty} (f_n - f_{n-1})$ converges absolutely in the Banach space $\mathcal{B}(U, \mathbf{F})$, and that its sum is the required function f.)

.3 Let Y be a Banach space, $y_0 \in Y$, $V = B(y_0, b) \subset Y$, and $0 \le c < 1$. Let v be a mapping of V into Y such that

(i) $\|v(y_1) - v(y_2)\| \le c\,\|y_1 - y_2\|$ for all $y_1, y_2 \in V$, and

(ii) $\|v(y_0) - y_0\| < b(1 - c)$.

Prove that v has a unique fixed point in V.

.4 Let X be a metric space such that each continuous self–map of a closed subset of S has a fixed point. Prove that X is complete. (Suppose the contrary, and choose a Cauchy sequence (x_n) in X that does not converge in X. Assuming, without loss of generality, that $x_i \ne x_j$ whenever $i \ne j$, for each $x \in X$ let

$$\alpha_x = \inf\{\rho(x, x_n) : x \ne x_n\}.$$

Show that $\alpha_x > 0$. Next, let $0 < r < 1$, set $\sigma(0) = 0$, and define $\sigma(n)$ inductively such that $\sigma(n) > \sigma(n-1)$ and

$$\rho(x_i, x_j) \le r\alpha_{x_{\sigma(n-1)}} \quad (i, j \ge \sigma(n)).$$

Let $S = \left\{x_{\sigma(n)} : n \ge 1\right\}$ and $f(x_{\sigma(n)}) = x_{\sigma(n+1)}$.)

.5 Let a, b be real numbers with $0 < b < 1$, and let X be the set of all continuous mappings $f : [0, b] \to \mathbf{R}$ such that $f(0) = a$ (so, according to Exercise (4.5.6: 6), X is a Banach space relative to the sup norm). Define a mapping T on X by

$$(Tf)(t) = a + \int_0^t |f(x)|\,\mathrm{d}x \quad (0 \le t \le b).$$

Prove that T is a contraction mapping of X into itself, and hence that there exists a unique $f \in X$ that is differentiable and satisfies $f' = |f|$ on the interval $(0, b)$.

A function φ is said to be *continuously differentiable* on an interval I of $\mathbf{R}$ if φ' exists and is continuous on I.

We now use Theorem (4.7.2) to prove the first of two theorems about the existence of solutions of ordinary differential equations, thereby generalising the work of Exercise (4.7.3: 5).

(4.7.4) Picard's Theorem. *Let K be the rectangle*

$$\left\{(x, y) \in \mathbf{R}^2 : |x - x_0| \leq a,\ |y - y_0| \leq b\right\}$$

where $a, b > 0$. Let $f : K \to \mathbf{R}$ be a continuous mapping such that there exists $c > 0$ with

$$|f(x, y_1) - f(x, y_2)| \leq c\,|y_1 - y_2|$$

for all applicable x, y_1, y_2 (in other words, f satisfies a Lipschitz condition in its second variable). *Let*

$$M = \sup_{(x,y) \in K} |f(x, y)|$$

and

$$h = \begin{cases} \min\left\{a, \frac{b}{M}\right\} & \text{if } M > 0 \\ a & \text{if } M = 0. \end{cases}$$

Then there exists a unique continuously differentiable mapping φ on the interval $I = [x_0 - h, x_0 + h]$, such that

$$\varphi(x_0) = y_0$$

and

$$\varphi'(x) = f(x, \varphi(x)) \textit{ for all } x \in I.$$

Proof. In view of the version of the Fundamental Theorem of Calculus in Exercise (1.5.14: 1), it suffices to find a continuous mapping $\varphi : I \to \mathbf{R}$ satisfying

$$\varphi(x) = y_0 + \int_{x_0}^{x} f(t, \varphi(t))\,\mathrm{d}t \qquad (1)$$

for all $x \in I$. Let V denote the closed ball with centre $y \mapsto y_0$ and radius b in the Banach space $(\mathcal{C}(I), \|\cdot\|)$, where $\|\cdot\|$ denotes the sup norm. If $y \in V$, then for all $t \in I$ we have $|y(t) - y_0| \leq b$ and therefore $(t, y(t)) \in K$; so

$$F_y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))\,\mathrm{d}t$$

defines a mapping $F_y : I \to \mathbf{R}.$ We see from Exercise (1.5.12: 4) that F_y satisfies the Lipschitz condition

$$|F_y(x) - F_y(x')| = \left| \int_{x'}^{x} f(t, y(t))\,\mathrm{d}t \right| \leq M\,|x - x'|$$

and is therefore uniformly continuous on I. Moreover,

$$|F_y(x) - y_0| \leq M\,|x - x_0| \leq Mh \leq b$$

for all $x \in I$, so $y \mapsto F_y$ maps V into V.

We now endow $\mathcal{C}(I)$ not with its usual norm, but with the norm defined by

$$\|f\|' = \sup\left\{ \mathrm{e}^{-2c|x-x_0|}\,|f(x)| : x \in I \right\}.$$

Recall from Exercise (4.5.6: 7) that $\mathcal{C}(I)$, and hence V, is complete with respect to the metric ρ' associated with this norm. We prove that $y \mapsto F_y$ is a contraction mapping on (V, ρ'). To this end, consider $y_1, y_2 \in V$ and $x \in I$. Taking, for example, the case where $x \geq x_0$, we have

$$\begin{aligned} |F_{y_1}(x) - F_{y_2}(x)| &\leq \int_{x_0}^{x} |f(t, y_1(t)) - f(t, y_2(t))|\,\mathrm{d}t \\ &\leq c \int_{x_0}^{x} |y_1(t) - y_2(t)|\,\mathrm{d}t \\ &\leq c\,\|y_1 - y_2\|' \int_{x_0}^{x} \mathrm{e}^{2c|t-x_0|}\,\mathrm{d}t \\ &< \frac{1}{2}\mathrm{e}^{2c|x-x_0|}\,\|y_1 - y_2\|', \end{aligned}$$

since

$$\int_{x_0}^{x} \mathrm{e}^{2c(t-x_0)}\,\mathrm{d}t = \frac{1}{2c}\left(\mathrm{e}^{2c(x-x_0)} - 1\right).$$

It follows that

$$\|F_{y_1} - F_{y_2}\|' < \frac{1}{2}\,\|y_1 - y_2\|' \quad (y_1, y_2 \in \mathcal{C}(I)).$$

Applying Banach's Contraction Mapping Theorem (4.7.2), we now obtain a unique element φ of V satisfying equation (1). □

A restricted version of Picard's Theorem can be proved by applying the Contraction Mapping Theorem to a certain complete subset of $\mathcal{C}(I)$, taken with the usual sup norm; this produces a positive number δ, which may be smaller than h, and a solution of the differential equation on the interval $[x_0 - \delta, x_0 + \delta]$. (See Chapter X of [13].) With a bit more work, it can then be shown that that solution extends to I (Exercise (4.7.5: 4)). The introduction of the norm $\|\cdot\|'$—a device due to Bielicki [4]—both simplifies the proof and provides, at a stroke, the solution over the whole interval I. (Note that when f is only known to be defined on K, I is the largest interval on which it makes sense to talk about a solution of the differential equation $y' = f(x, y)$.)

By examining closely the proofs of Theorems (4.7.2) and (4.7.4), we obtain the following iteration scheme for a sequence (y_n) of functions converging to a solution of the differential equation in the preceding theorem.

$$\begin{aligned} y_0(x) &= y_0, \\ y_n(x) &= y_0 + \int_{x_0}^{x} f(t, y_{n-1}(t))\, \mathrm{d}t \quad (n \geq 1). \end{aligned}$$

This scheme can be used in practice, although there are better methods of finding solutions of first–order differential equations of special types.

(4.7.5) Exercises

.1 Apply the foregoing iteration scheme to solve the differential equation $y' = y$ on $\mathbf{R}$ with initial condition $y(0) = 3$.

.2 Let
$$K = \left\{(x, y) \in \mathbf{R}^2 : |x| \leq a,\ |y| \leq b\right\},$$
where a, b are positive constants. Let f be a continuous mapping of K into $\mathbf{R}$ such that $f(x, y) < 0$ if $xy > 0$, and $f(x, y) > 0$ if $xy < 0$. Prove that $x \mapsto 0$ is the unique solution of the differential equation $y' = f(x, y)$ defined in a neighbourhood of 0 and such that $y(0) = 0$. (Assume the contrary, and consider, in a compact interval containing 0, the points where a solution attains its maximum or minimum.)

.3 Define $f : \mathbf{R}^2 \to \mathbf{R}$ by
$$f(x, y) = \begin{cases} -2x & \text{if } y \geq x^2 \\ -\frac{2y}{x} & \text{if } |y| < x^2 \\ 2x & \text{if } y \leq -x^2. \end{cases}$$
Define a sequence of functions by setting $y_0(x) = x^2$ and
$$y_{n+1}(x) = \int_0^x f(t, y_n(t))\, \mathrm{d}t.$$
Show that for each $x \neq 0$ the sequence $(y_n(x))_{n=0}^{\infty}$ is not convergent. Comment on this, in the light of Exercise (4.7.5: 2) and the paragraph immediately preceding this set of exercises.

.4 Let x_0, y_0, I, and K be as in Theorem (4.7.4), and let $f : K \to \mathbf{R}$ be a continuous function with the following property: for each $(\xi, \eta) \in K^\circ$ there exist $\delta > 0$ and a unique continuously differentiable function $y : [\xi - \delta, \xi + \delta] \to \mathbf{R}$ such that $y(\xi) = \eta$ and $y'(x) = f(x, y(x))$ whenever $|x - \xi| \leq \delta$. Show that there exists a unique continuously differentiable function $\varphi : I \to \mathbf{R}$ such that $\varphi(x_0) = y_0$ and $\varphi'(x) = f(x, \varphi(x))$ for all $x \in I$. (Let S be the set of all positive numbers $\delta \leq h$ with the property that there exists a continuously differentiable function $y : [x_0 - \delta, x_0 + \delta] \to \mathbf{R}$ such that $y(x_0) = y_0$ and $y'(x) = f(x, y(x))$ whenever $|x - x_0| \leq \delta$. Let $\sigma = \sup S$, suppose that $\sigma < h$, and derive a contradiction.)

.5 Let I be the closed interval $[a, b]$ in $\mathbf{R}$, and

$$A = \left\{(x, y) \in \mathbf{R}^2 : a \leq x \leq y \leq b\right\}.$$

Let the function $k : I \times I \to \mathbf{R}$ be continuous on A and vanish everywhere on $(I \times I) \backslash A$, and for each $f \in \mathcal{C}(I)$ define $Tf : I \to \mathbf{R}$ by

$$Tf(t) = \int_a^t k(s, t) f(s) \, \mathrm{d}s \quad (t \in I).$$

Show that for all sufficiently large n, T^n is a contraction mapping of $\mathcal{C}(I)$ into itself, and hence that the integral equation

$$f(t) = g(t) + \int_a^t k(s, t) f(s) \, \mathrm{d}s,$$

has a unique solution f in $\mathcal{C}(I)$ for each given $g \in \mathcal{C}(I)$. (For the contraction mapping part, show that

$$|T^n f(x) - T^n g(x)| \leq \frac{M^n}{n!} (x - a)^n \, \|f - g\|$$

for all $x \in I$ and $f, g \in \mathcal{C}(I)$.)

.6 Taking $I = [0, 1]$, use the preceding exercise to find the solution of the integral equation

$$f(t) = g(t) + c \int_0^t (t - s)^3 f(s) \, \mathrm{d}s \quad (t \in I),$$

where c is a positive constant and $g \in \mathcal{C}(I)$.

.7 Let $c > 0$, let f be a continuous real–valued mapping that satisfies the condition

$$|f(x, y_1) - f(x, y_2)| \leq c \, |y_1 - y_2|$$

on the strip $[a, b] \times \mathbf{R}$ in $\mathbf{R}^2$, and let (x_0, y_0) be any point of that strip. Prove that the differential equation $y' = f(x, y)$ has a unique solution $y : [a, b] \to \mathbf{R}$ such that $y(x_0) = y_0$. (For each $x \in [a, b]$ define $y_0(x) = y_0$ and

$$y_{n+1}(x) = y_0 + \int_{x_0}^{x} f(t, y_n(t))\,\mathrm{d}t.$$

Let

$$M = |y_0| + \max\{|y_1(x)| : a \leq x \leq b\}.$$

Show that the series

$$y_0 + \sum_{n=1}^{\infty} (y_n(x) - y_{n-1}(x))$$

converges uniformly on $[a, b]$ to a sum $y(x)$, by comparison with the series

$$M + \sum_{n=1}^{\infty} c^{n-1} M \frac{(b-a)^{n-1}}{(n-1)!}.$$

Then show that y is the desired unique solution.)

.8 Let K be the compact set

$$\{(x, y) \in \mathbf{R} \times \mathbf{R}^n : |x - x_0| \leq a,\ \|y - y_0\| \leq b\},$$

where $a, b > 0$. Let $f = (f_1, \ldots, f_n)$ be a continuous mapping of K into $\mathbf{R}^n$ such that there exists $c > 0$ with

$$|f(x, y_1) - f(x, y_2)| \leq c\,\|y_1 - y_2\|$$

for all applicable x, y_1, y_2. Let

$$M = \sup_{(x,y) \in K} \|f(x, y)\|$$

and

$$h = \begin{cases} \min\left\{a, \frac{b}{M}\right\} & \text{if } M > 0 \\ a & \text{if } M = 0. \end{cases}$$

Prove that there exists a unique mapping $\varphi = (\varphi_1, \ldots, \varphi_n)$ of the interval $I = [x_0 - h, x_0 + h]$ into $\mathbf{R}^n$, such that

(i) $\varphi(x_0) = y_0$, and

(ii) for each k the component mapping φ_k is continuously differentiable and satisfies $\varphi_k'(x) = f_k(x, \varphi(x))$ on I.

.9 Let p, q, and r be continuous real–valued functions on the interval $[a, b]$, let $x_0 \in [a, b]$, and let y_0, y_0' be real numbers. Use the preceding exercise to prove that there exists a unique function $y : [a, b] \to \mathbf{R}$ satisfying the differential equation

$$y'' + p(x)y' + q(x) = r(x)$$

on $[a, b]$, with initial conditions $y(x_0) = y_0$ and $y'(x_0) = y_0'$.

Although Picard's Theorem enables us to solve, both in principle and in practice, a large class of differential equations, there are simple examples of differential equations to which it does not apply and yet for which solutions can easily be found. One such example is the equation $y' = y^{1/3}$ with initial condition $y(0) = 0$: this equation has two solutions—namely, $y = 0$ and $y = (2x/3)^{3/2}$ —but the function $(x, y) \mapsto y^{1/3}$ does not satisfy a Lipschitz condition at $(0, 0)$. The final theorem of this chapter covers cases such as this, and provides us with a good application of Ascoli's Theorem and the Stone–Weierstrass Theorem.

(4.7.6) Peano's Theorem. *Let K be the rectangle*

$$\left\{(x, y) \in \mathbf{R}^2 : |x - x_0| \leq a,\ |y - y_0| \leq b\right\},$$

where $a, b > 0$. Let $f : K \to \mathbf{R}$ be a continuous mapping,

$$M = \sup_{(x,y)\in K} |f(x, y)|$$

and

$$h = \begin{cases} \min\left\{a, \frac{b}{M}\right\} & \text{if } M > 0 \\ a & \text{if } M = 0. \end{cases}$$

Then there exists a continuously differentiable mapping φ on the interval $I = [x_0 - h, x_0 + h]$, such that

$$\varphi(x_0) = y_0$$

and

$$\varphi'(x) = f(x, \varphi(x)) \text{ for all } x \in I.$$

Proof. Using Exercise (4.6.8: 4), construct a sequence (p_n) of polynomial functions of two variables such that $\|f - p_n\| \leq 2^{-n}$ for each n, where $\|\cdot\|$ denotes the sup norm on $\mathcal{C}(K)$. We may assume that $|p_n| \leq 2M$ for each n. By Exercise (4.7.1: 1), Picard's Theorem, and the Fundamental Theorem of Calculus, the integral equation

$$y(x) = y_0 + \int_{x_0}^{x} p_n(t, y(t))\,\mathrm{d}t$$

has a unique solution φ_n on the interval I. Exercise (1.5.12: 4) shows that for all $x_1, x_2 \in I$,

$$|\varphi_n(x_2) - \varphi_n(x_1)| \leq \left| \int_{x_1}^{x_2} p_n(t, \varphi_n(t))\, \mathrm{d}t \right| \leq 2M\, |x_2 - x_1|\,.$$

It follows that (φ_n) is an equicontinuous sequence in $\mathcal{C}(I)$. Also, for each $x \in I$,

$$|\varphi_n(x)| \leq |y_0| + \left| \int_{x_0}^{x} p_n(t, \varphi_n(t))\, \mathrm{d}t \right| \leq |y_0| + 2M\, |I|\ ;$$

so (φ_n) is a bounded sequence in $\mathcal{C}(I)$. Applying Ascoli's Theorem (4.5.8), and, if necessary, passing to a subsequence of (φ_n), we may now assume that (φ_n) converges uniformly on I to an element φ of $\mathcal{C}(I)$. Since f is uniformly continuous on K, for each $\varepsilon > 0$ there exists $t > 0$ such that if $(x_i, y_i) \in K$ and

$$\max\{|x_1 - x_2|\,, |y_1 - y_2|\} < t,$$

then

$$|f(x_1, y_1) - f(x_2, y_2)| < \varepsilon.$$

Choose N such that for all $n \geq N$,

$$\|\varphi - \varphi_n\| < \min\{t, \varepsilon\}$$

and $2^{-n} < \varepsilon$. Consider any $x \in I$ and any $n \geq N$. Note that for each $t \in I$, $(t, \varphi(t))$ belongs to the closed set K and

$$|f(t, \varphi(t)) - f(t, \varphi_n(t))| < \varepsilon.$$

We now have

$$\left| \int_{x_0}^{x} f(t, \varphi(t))\, \mathrm{d}t - \int_{x_0}^{x} f(t, \varphi_n(t))\, \mathrm{d}t \right| \leq \varepsilon\, |x - x_0| < |I|\, \varepsilon$$

and therefore

$$\begin{aligned} \left| \varphi(x) - y_0 - \int_{x_0}^{x} f(t, \varphi(t))\, \mathrm{d}t \right| &\leq |\varphi(x) - \varphi_n(x)| \\ &\quad + \left| \varphi_n(x) - y_0 - \int_{x_0}^{x} p_n(t, \varphi_n(t))\, \mathrm{d}t \right| \\ &\quad + \left| \int_{x_0}^{x} \left(p_n(t, \varphi_n(t)) - f(t, \varphi_n(t)) \right)\, \mathrm{d}t \right| \\ &\quad + \left| \int_{x_0}^{x} \left(f(t, \varphi_n(t)) - f(t, \varphi(t)) \right)\, \mathrm{d}t \right| \\ &< \varepsilon + 0 + |I|\, \|f - p_n\| + |I|\, \varepsilon \\ &= (1 + 2\,|I|)\, \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

$$\varphi(x) = y_0 + \int_{x_0}^{x} f(t, \varphi(t))\,\mathrm{d}t \quad (x \in I).$$

A final application of the Fundamental Theorem of Calculus (see Exercise (1.5.14: 1)) shows that φ is continuously differentiable and satisfies the desired conditions. □

There are two fundamental differences between Picard's Theorem and Peano's:

- in the former the solution is unique, whereas in the latter it need not be;
- the proof of Picard's Theorem embodies an algorithm for computing the solution, but Peano's Theorem uses the highly nonconstructive property of sequential compactness and is an intrinsically nonalgorithmic theorem.

By an *ε–approximate solution* to the differential equation

$$y' = f(x, y), \quad y(x_0) = y_0 \tag{2}$$

in an interval J containing x_0 we mean a mapping $y : J \to \mathbf{R}$ with the following properties.

— There exists a partition $(x_1, x_2, \dots, x_n)$ of J such that y is continuously differentiable on each of the intervals $[x_i, x_{i+1}]$;

— $|y'(x) - f(x, y(x))| \leq \varepsilon$ for all $x \in \bigcup_{i=1}^{n-1} (x_i, x_{i+1})$;

— $y(x_0) = y_0$.

(4.7.7) Exercises

.1 Under the hypotheses of Theorem (4.7.6), but without invoking that theorem, show that for each $\varepsilon > 0$ there exists an ε-approximate solution of (2). (Choose $\delta > 0$ such that $|f(x_1, y_1) - f(x_2, y_2)| \leq \varepsilon$ whenever $(x_i, y_i) \in K$ and $\|(x_1, y_1) - (x_2, y_2)\| \leq \delta$. Take points $x_0 < x_1 < \cdots < x_n = x_0 + h$ such that $x_{i+1} - x_i \leq \min\{\delta, \delta/M\}$, and construct an ε–approximate solution of (2) on $[x_0, x_0 + h]$ that is linear on each of the intervals $[x_i, x_{i+1}]$; then deal with the interval $[x_0 - h, x_0]$. This technique is known as the *Cauchy–Euler method.*)

.2 Under the hypotheses of Theorem (4.7.6), let (ε_n) be a sequence of positive numbers converging to 0, and for each n let φ_n be an ε_n–approximate solution of the differential equation (2) on I. Suppose that (φ_n) converges uniformly to a continuous function φ on I. Prove that

(i) $(t, \varphi(t)) \in K$ for each $t \in I$;

(ii) $\int_{x_0}^{x} f(t, \varphi_n(t))\, dt \to \int_{x_0}^{x} f(t, \varphi(t))\, dt$ uniformly on I as $n \to \infty$;

(iii) φ is a solution of the differential equation (2) on I.

.3 Use the preceding two exercises to give an alternative proof of Peano's Theorem.

.4 Let I, K, f, M, and c be as in the hypotheses of Picard's Theorem. Let $\varepsilon_1, \varepsilon_2 > 0$, and let φ_i be an ε_i–approximate solution to the differential equation on I. Show that

$$|\varphi_1(x) - \varphi_2(x)| \leq |\varphi_1(x_0) - \varphi_2(x_0)|\, \mathrm{e}^{c|x-x_0|} + (\varepsilon_1 + \varepsilon_2)\, \frac{\mathrm{e}^{c|x-x_0|} - 1}{c}$$

for each $x \in I$. (Use Exercise (2.3.3: 14).) Hence find an alternative proof of Picard's Theorem.

5
Hilbert Spaces

When shall we three meet again...?
MACBETH, Act 1, Scene 1

This chapter explores the elementary theory of Hilbert spaces. In Section 1 we introduce the notion of an inner product, with its associated norm, on a linear space, and prove some fundamental inequalities. The next section deals with orthogonality, projections, and orthonormal bases in a Hilbert space, and with their use in approximation theory. In Section 3 we derive Riesz's characterisation of the bounded linear functionals on a Hilbert space, and show how this can be applied both in the theory of operators and to prove the existence of weak solutions of the Dirichlet Problem.

5.1 Inner Products

So far we have shown how to abstract the notions of distance and length from Euclidean space to the abstract contexts of a metric space and a normed space, respectively. In this chapter we show how to abstract the notion of the inner product in $\mathbf{R}^n$ to the context of a linear space. The resulting combination of distance, length, and inner product provides the space with an extremely rich structure that turns out to have many significant applications in pure and applied mathematics. In particular—although we are not able to explore that subject in this book—certain linear self–maps of such a space are the mathematical analogues of quantum–mechanical operations.

By an *inner product* on a linear space X over $\mathbf{F}$ we mean a mapping $(x,y) \mapsto \langle x,y\rangle$ of $X \times X$ into $\mathbf{F}$ such that the following hold for all x, y, z in X and all λ, μ in $\mathbf{F}$.

IP1 $\langle x,y\rangle = \langle y,x\rangle^*$.

IP2 $\langle \lambda x + \mu y, z\rangle = \lambda\langle x,z\rangle + \mu\langle y,z\rangle$.

IP3 $\langle x,x\rangle \geq 0$, and $\langle x,x\rangle = 0$ if and only if $x = 0$.

The element $\langle x,y\rangle$ of $\mathbf{F}$ is then called the *inner product of the vectors x and y.* Note that by IP2, the inner product is linear in the first variable; and that by IP2 and IP1, it is *conjugate linear* in the second—that is,

$$\langle x, \lambda y + \mu z\rangle = \lambda^*\langle x,y\rangle + \mu^*\langle x,z\rangle.$$

We define an *inner product space,* or a *prehilbert space,* to be a pair $(X, \langle\cdot,\cdot\rangle)$ consisting of a linear space X over $\mathbf{F}$ and an inner product $\langle\cdot,\cdot\rangle$ on X. When there is no confusion over the inner product, we refer to X itself as an inner product space. By a *subspace of an inner product space* X we mean a linear subset S of X, taken with the inner product induced on S by that on X; thus the inner product on S is the restriction to $S \times S$ of the inner product on X.

The simplest example of an inner product space is the *Euclidean space* $\mathbf{F}^n$, with the inner product of vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ defined by

$$\langle x,y\rangle = \sum_{k=1}^{n} x_k y_k^*.$$

For another example consider the linear space $l_2(\mathbf{C})$ of square–summable sequences in $\mathbf{C}$, introduced in Exercise (4.4.4: 3), where the inner product of two elements $x = (x_k)$ and $y = (y_k)$ is defined as

$$\langle x,y\rangle = \sum_{k=1}^{\infty} x_k y_k^*.$$

This can be regarded as a generalisation of the first example, since the one–one mapping $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_n, 0, 0, \ldots)$ of $\mathbf{C}^n$ into $l_2(\mathbf{C})$ preserves the value of the inner product.

Before discussing a third example, in Exercise (5.1.1: 2), let us agree to call a complex–valued function f on a subset X of $\mathbf{R}$ *integrable* if its real and imaginary parts are integrable over X, in which case we define

$$\int_X f = \int_X \mathrm{Re}(f) + \mathrm{i}\int_X \mathrm{Im}(f).$$

The complex integration spaces $L_p(X, \mathbf{C})$ are then defined in the obvious way, and we use $L_p(X, \mathbf{F})$ to denote either $L_p(I)$ or $L_p(X, \mathbf{C})$, depending on whether $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$.

(5.1.1) Exercises

.1 Prove that the equation $\langle x, y \rangle = \sum_{n=1}^{\infty} x_n y_n^*$ does define an inner product on $l_2(\mathbf{C})$. (You must first prove that the series $\sum_{n=1}^{\infty} x_n y_n^*$ is convergent when (x_n) and (y_n) are elements of $l_2(\mathbf{C})$.)

.2 By a *weight function* on a compact interval $I = [a, b]$ we mean a nonnegative continuous function w on I such that if $f \in \mathcal{C}(I)$ and $\int_I w(t)f(t)\,\mathrm{d}t = 0$, then $f = 0$. Prove that

$$\langle f, g \rangle = \int_a^b w(t)f(t)g(t)^*\,\mathrm{d}t$$

defines an inner product on $L_2(I, \mathbf{F})$ (where, as always, we identify two elements of $L_2(I, \mathbf{F})$ that are equal almost everywhere). We denote the corresponding inner product space by $L_{2,w}(I, \mathbf{F})$.

(5.1.2) Proposition. *Let X be an inner product space. Then*

$$\|x\| = \langle x, x \rangle^{1/2}$$

defines a norm on X. Moreover, the inner product and this norm satisfy the Cauchy–Schwarz inequality

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|$$

and Minkowski's inequality

$$\langle x + y, \, x + y \rangle^{1/2} \leq \langle x, x \rangle^{1/2} + \langle y, y \rangle^{1/2}.$$

Proof. We first prove the two inequalities. For any $x, y \in X$ and any $\lambda \in \mathbf{F}$ we have, by IP1 through IP3,

$$0 \leq \langle x + \lambda y, \, x + \lambda y \rangle = \langle x, x \rangle + \langle x, \lambda y \rangle + \langle \lambda y, x \rangle + \langle \lambda y, \lambda y \rangle,$$

so

$$\|x\|^2 + \lambda^* \langle x, y \rangle + \lambda \langle x, y \rangle^* + \lambda\lambda^* \|y\|^2 \geq 0, \tag{1}$$

with equality if and only if $x + \lambda y = 0$. If $\|y\| \neq 0$, the Cauchy–Schwarz inequality is obtained by taking $\lambda = -\langle x, y \rangle / \|y\|^2$; if $\|x\| \neq 0$, the equality is obtained by taking $\lambda = -\langle x, y \rangle / \|x\|^2$; if $\|x\| = \|y\| = 0$, then IP3 shows that $x = 0 = y$ and hence that $\langle x, y \rangle = 0$, so the Cauchy–Schwarz inequality holds trivially.

Taking $\lambda = 1$ in (1) and using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
\langle x + y, \, x + y \rangle &= \langle x, x \rangle + 2 \operatorname{Re} \langle x, y \rangle + \langle y, y \rangle \\
&\leq \langle x, x \rangle + 2 \, |\langle x, y \rangle| + \langle y, y \rangle \\
&\leq \langle x, x \rangle + 2 \, \|x\| \, \|y\| + \langle y, y \rangle \\
&= \left(\langle x, x \rangle^{1/2} + \langle y, y \rangle^{1/2} \right)^2,
\end{aligned}$$

which immediately yields Minkowski's inequality. It is now a simple exercise, involving this inequality and the defining properties of an inner product, to show that $x \mapsto \langle x, x\rangle^{1/2}$ is a norm on X. $\square$

When we refer to the norm or the metric structure on an inner product space X, we always have in mind the norm, and the corresponding metric structure, associated with the inner product as in Proposition (5.1.2).

(5.1.3) Exercises

.1 Complete the details of the proof that if $\langle \cdot, \cdot\rangle$ is an inner product on a linear space X, then $\|x\| = \langle x, x\rangle^{1/2}$ defines a norm on X.

.2 Prove that an inner product on a linear space X is continuous, and that it is uniformly continuous on bounded sets, with respect to the corresponding product norm on $X \times X$.

.3 Prove the *parallelogram law* for vectors x, y in an inner product space:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 .$$

Interpreting a norm as a length, we see that this law generalises the plane geometry theorem that the sum of the squares of the diagonals of a parallelogram equals the sum of the squares of its sides.

.4 Use the parallelogram law to show that a Hilbert space is uniformly convex (see Exercise (4.2.2: 15)).

.5 Let X be a normed space whose norm satisfies the parallelogram law (see the exercise before last). Show that if $\mathbf{F} = \mathbf{R}$, then

$$\langle x, y\rangle = \tfrac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right)$$

defines an inner product on X such that $\|x\| = \langle x, x\rangle^{1/2}$ for each $x \in X$. Then show that if $\mathbf{F} = \mathbf{C}$, there is a unique inner product on X related to the norm in this way.

.6 Prove that there is no inner product on $\mathcal{C}[0,1]$ such that $\langle f, f\rangle^{1/2} = \|f\|$ (the supremum norm). (Show that the supremum norm does not obey the parallelogram law.)

.7 Prove that the inner product space $L_{2,w}(I, \mathbf{F})$, introduced in Exercise (5.1.1: 2), is complete. Prove also that $\mathcal{C}(I, \mathbf{F})$ is not a complete subspace of this inner product space.

An inner product space that is complete with respect to its norm is called a *Hilbert space.* For example, the Euclidean space $\mathbf{F}^n$ is a Hilbert space, as is $l_2(\mathbf{C})$. On the other hand, if w is a nonnegative weight function on a compact interval I, then Exercise (5.1.3: 7) shows that $L_{2,w}(I, \mathbf{F})$ is complete, but $\mathcal{C}(I, \mathbf{F})$ is not complete, with respect to the inner product

$$\langle f, g \rangle = \int_I w(t) f(t) g(t)^* \, dt.$$

(5.1.4) Exercises

.1 Let X be an inner product space, X_0 a closed linear subspace of X, and φ the canonical mapping of X onto the quotient space X/X_0. Prove that

$$\langle \varphi(x), \varphi(y) \rangle = \langle x, y \rangle$$

unambiguously defines an inner product on X/X_0, and that the corresponding norm is the quotient norm on X/X_0.

.2 Show that each inner product space X can be embedded as a dense subset of a Hilbert space H. (Extend the inner product by continuity to the completion of X, as defined on page 179.) H is then known as the (*Hilbert space*) *completion* of X.

.3 Prove that any two completions H and H' of an inner product space X are isomorphic, in the sense that there exists a one–one linear mapping u of H onto H' such that $\langle u(x), u(y) \rangle = \langle x, y \rangle$ for all $x, y \in H$.

.4 Let I be a compact interval. Show that $L_2(I, \mathbf{F})$ is the completion of the Hilbert space $\mathcal{C}(I, \mathbf{F})$ with respect to the inner product $\langle f, g \rangle = \int_I f(t) g(t)^* \, dt$.

5.2 Orthogonality and Projections

Two elements x, y of an inner product space X are said to be *orthogonal* if $\langle x, y \rangle = 0$, in which case we write $x \perp y$. In view of IP1, the relation $\perp$ is symmetric: $x \perp y$ if and only if $y \perp x$. A vector x is said to be *orthogonal to the subset* S of X if $x \perp s$ for each $s \in S$; we then write $x \perp S$. The set of all vectors orthogonal to S is called the *orthogonal complement* of S, and is written $S^\perp$ (pronounced "S perp"). It follows from IP2 that $S^\perp$ is a (linear) subspace of X; and from IP3 that $S \cap S^\perp$ is nonempty if and only if $0 \in S$, in which case $S \cap S^\perp = \{0\}$. Moreover, $S^\perp$ is orthogonal to $\overline{S}$, in the sense that every element of $S^\perp$ is orthogonal to $\overline{S}$: for, by Exercise (5.1.3: 2), if (s_n) is a sequence of elements of S converging to $s \in \overline{S}$, then for each $x \in S^\perp$,

$$\langle x, s \rangle = \lim_{n \to \infty} \langle x, s_n \rangle = 0.$$

For each $x \in X$, $\{x\}^{\perp}$ is the kernel of the continuous linear functional $z \mapsto \langle z, x\rangle$ on X, and so, by Proposition (4.2.3), is a closed subspace of X. Hence for each subset S of X,

$$S^{\perp} = \bigcap_{s \in S} \{s\}^{\perp}$$

is closed in X.

If x and y are orthogonal vectors, then, expanding $\langle x+y, x+y\rangle$, we obtain *Pythagoras's Theorem*:

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 .$$

(5.2.1) Proposition. *Let S be a nonempty complete convex subset of an inner product space X, and let $a \in X$. Then there exists a unique vector s in S such that $\|a-s\| = \rho(a,S)$.*

Proof. Let $d = \rho(a,S)$, and choose a sequence (s_n) in S such that $\rho(a,s_n) \to d$. Using the parallelogram law (Exercise (5.1.3: 3)) and the convexity of S, for all m and n we compute

$$\begin{aligned}
\|s_m - s_n\|^2 &= \|s_m - a - (s_n - a)\|^2 \\
&= 2\|s_m - a\|^2 + 2\|s_n - a\|^2 - \|s_m - a + (s_n - a)\|^2 \\
&= 2\|s_m - a\|^2 + 2\|s_n - a\|^2 - 4\left\|\tfrac{1}{2}(s_m + s_n) - a\right\|^2 \\
&\leq 2\|s_m - a\|^2 + 2\|s_n - a\|^2 - 4d^2 \\
&= 2\left(\|s_m - a\|^2 - d^2\right) + 2\left(\|s_n - a\|^2 - d^2\right) \\
&\to 0 \text{ as } m, n \to \infty.
\end{aligned}$$

Hence (s_n) is a Cauchy sequence. Since S is complete, (s_n) converges to a limit s in S; then

$$\|a - s\| = \lim_{n \to \infty} \|a - s_n\| = \rho(a,S).$$

On the other hand, if $s' \in S$ and $\|a - s'\| = \rho(a,S)$, then a computation similar to the one used at the start of the proof shows that

$$\begin{aligned}
\|s - s'\|^2 &= 2\|s - a\|^2 + 2\|s' - a\|^2 - 4\left\|\tfrac{1}{2}(s + s') - a\right\|^2 \\
&= 4\left(d^2 - \left\|\tfrac{1}{2}(s + s') - a\right\|^2\right) \\
&\leq 0,
\end{aligned}$$

so that $s = s'$. □

It is worth digressing here to prove a converse of the foregoing result.

(5.2.2) Proposition. *Let S be a nonempty closed subset of the Euclidean space $\mathbf{R}^N$ such that each point of $\mathbf{R}^N$ has a unique closest point in S. Then S is convex.*

Proof. Supposing that S is not convex, we can find $a, b \in S$ and $\lambda \in (0,1)$ such that

$$z = \lambda a + (1-\lambda)\, b \notin S.$$

Since $X \backslash S$ is open, there exists $r > 0$ such that $\overline{B}(z,r) \cap S = \emptyset$. Let $\mathcal{F}$ be the set of all closed balls B such that $\overline{B}(z,r) \subset B$ and $S \cap B^\circ = \emptyset$; then $\overline{B}(z,r) \in \mathcal{F}$. The radii of the balls belonging to $\mathcal{F}$ are bounded above, since any ball containing B and having sufficiently large radius will meet S. Let r_∞ be the supremum of the radii of the members of $\mathcal{F}$, and let $(\overline{B}(x_n, r_n))_{n=1}^\infty$ be a sequence of elements of $\mathcal{F}$ such that $r_n \to r_\infty$. Then $x_n \in \overline{B}(z, r_\infty)$ for each n. Since $\overline{B}(z, r_\infty)$ is compact (Theorem (4.3.6)) and therefore sequentially compact (Theorem (3.3.9)), we may assume without loss of generality that (x_n) converges to a limit x_∞. Let $K = \overline{B}(x_\infty, r_\infty)$; we prove that $K \in \mathcal{F}$.

First we consider any $x \in \overline{B}(z,r)$ and any $\varepsilon > 0$. Choosing m such that $\|x_m - x_\infty\| < \varepsilon$, and noting that $\overline{B}(z,r) \subset \overline{B}(x_m, r_m)$, we have

$$\begin{aligned} \|x - x_\infty\| &\leq \|x - x_m\| + \|x_m - x_\infty\| \\ &< r_m + \varepsilon \\ &\leq r_\infty + \varepsilon. \end{aligned}$$

Since ε is arbitrary, we conclude that $\|x - x_\infty\| \leq r_\infty$; whence $\overline{B}(z,r) \subset K$. On the other hand, supposing that there exists $s \in S \cap B(x_\infty, r_\infty)$, choose $\delta > 0$ such that $\|s - x_\infty\| < r_\infty - \delta$, and then n such that $0 \leq r_\infty - r_n < \delta/2$ and $\|x_n - x_\infty\| < \delta/2$. We have

$$\begin{aligned} \|s - x_n\| &\leq \|s - x_\infty\| + \|x_\infty - x_n\| \\ &< r_\infty - \delta + \tfrac{\delta}{2} \\ &= r_\infty - \tfrac{\delta}{2} \\ &< r_n, \end{aligned}$$

so $s \in S \cap B(x_n, r_n)$. This is absurd, as $B(x_n, r_n) \in \mathcal{F}$; hence $S \cap B(x_\infty, r_\infty)$ is empty, and therefore $K \in \mathcal{F}$.

Now, the centre x_∞ of K has a unique closest point p in S. This point cannot belong to K°, as $K \in \mathcal{F}$; nor can it lie outside K, as r_∞ is the supremum of the radii of the balls in $\mathcal{F}$. Therefore p must lie on the boundary of K. The unique closest point property of S ensures that the boundary of K intersects S in the single point p. It now follows from Exercise (4.3.7: 7) that there exists a ball K' that is concentric with K, has radius greater than r_∞, and is disjoint from S. This ball must contain $\overline{B}(z,r)$ and so belongs to $\mathcal{F}$. Since this contradicts our choice of r_∞, we conclude that S is, in fact, convex. □

(5.2.3) Exercises

.1 Give an example of a norm $\|\cdot\|'$ on $\mathbf{R}^2$, a closed convex subset S of $\mathbf{R}^2$, and a point $x \in \mathbf{R}^2$ such that x has infinitely many closest points in S relative to $\|\cdot\|'$.

.2 Let S be the subset of c_0 consisting of all elements (x_n) such that $\sum_{n=1}^{\infty} 2^{-n} x_n = 0$. Show that S is a closed subspace of c_0 and that no point of $c_0 \backslash S$ has a closest point in S. (Given $a = (a_n) \in c_0 \backslash S$, set $\alpha = \sum_{n=1}^{\infty} 2^{-n} a_n$ and show that $\rho(a, S) \leq |\alpha|$. Let $x = (x_n) \in S$, suppose that $\|a - x\| \leq |\alpha|$, and obtain a contradiction.)

.3 Let S be a closed convex set in a uniformly convex Banach space X. (See Exercise (4.2.2: 15).) Show that to each point a of X there corresponds a unique closest point in S. (To prove existence, reduce to the case where $a = 0$ and $\rho(0, S) = 1$. Then choose a sequence (s_n) in S such that $\|s_n\| \to 1$, and show that $\left(\|s_n\|^{-1} s_n\right)$ is a Cauchy sequence in X.)

.4 Give two proofs that c_0 is not uniformly convex.

.5 Let S be a bounded closed subset of $\mathbf{R}^N$ with the property that to each $x \in \mathbf{R}^N$ there corresponds a unique *farthest point* of S—that is, a point s_0 of S such that

$$\|x - s_0\| = \sup\{\|x - s\| : s \in S\}.$$

Show that S consists of a single point. (First show that S is bounded. Then choose $r > 0$ such that $S \subset \overline{B}(0, r/2)$, and consider the family $\mathcal{F}$ of all closed balls B such that $S \subset B \subset \overline{B}(0, r)$. Show that $\mathcal{F}$ contains a ball with minimum radius, and then show that that radius is 0.)

Proposition (5.2.2) was first proved by Motzkin in 1935, and the result in Exercise (5.2.3: 5) by Motzkin, Straus, and Valentine in 1953.

We now turn from our digression to the subject of projections. In the special case of Proposition (5.2.1) where S is a complete subspace of X, the unique point of S closest to a given vector $x \in X$ is called the *projection of the vector x on S*, and the mapping that carries each vector in X to its projection on S is called the *projection of X on S*. For example,

- the projection of X on X is the *identity operator* $I : X \to X$ defined by $Ix = x$;

- projections on finite–dimensional subspaces of X are always defined, since finite–dimensional normed spaces are complete, by Proposition (4.3.3);
- the projection of a Hilbert space on any closed subspace is defined, in view of Proposition (3.2.9).

The next result enables us to show that projections are bounded linear mappings.

(5.2.4) Proposition. *Let S be a complete subspace of an inner product space X, and P the projection of X on S. Then for each $x \in X$, Px is the unique vector $s \in S$ such that $x - s$ is orthogonal to S.*

Proof. Given x in X, let $d = \rho(x, S)$. For all y in S and λ in $\mathbf{F}$, we have $Px - \lambda y \in S$, so that

$$\langle x - Px + \lambda y,\, x - Px + \lambda y \rangle \geq d^2 = \langle x - Px,\, x - Px \rangle\,,$$

and therefore

$$|\lambda|^2 \|y\|^2 + 2\,\mathrm{Re}\,(\lambda^* \langle x - Px,\, y \rangle) \geq 0.$$

Suppose that $\mathrm{Re}\,\langle x - Px,\, y \rangle \neq 0$; then by the Cauchy–Schwarz inequality, $y \neq 0$. Taking

$$\lambda = -\frac{\langle x - Px,\, y \rangle}{\|y\|^2},$$

we obtain the contradiction

$$|\lambda|^2 \|y\|^2 + 2\,\mathrm{Re}\,(\lambda^* \langle x - Px,\, y \rangle) < 0.$$

Thus $\mathrm{Re}\,\langle x - Px,\, y \rangle = 0$. Likewise, $\mathrm{Im}\,\langle x - Px,\, y \rangle = 0$, so $\langle x - Px,\, y \rangle = 0$.

If, conversely, s is any vector in S such that $x - s$ is orthogonal to S, then $s - Px$ is in S, and so

$$\langle s - Px,\, s - Px \rangle = \langle x - Px,\, s - Px \rangle - \langle x - s,\, s - Px \rangle = 0;$$

whence $s = Px$, by IP3. □

(5.2.5) Exercises

.1 Prove that if S is a complete subspace of an inner product space X, then $(S^\perp)^\perp = S$.

.2 Let P be the projection of a Hilbert space H onto a complete subspace S. Use Proposition (5.2.4) to show that P is a linear mapping, and that

$$\langle Px, Py \rangle = \langle Px, y \rangle = \langle x, Py \rangle$$

for all $x, y \in H$. Show also that $\|Px\| \leq \|x\|$ for all $x \in H$, and that if $S \neq \{0\}$, then $\|P\| = 1$.

.3 In the notation of the preceding exercise prove that each vector $x \in H$ has a unique representation in the form $x = y + z$ with $y \in S$ and $z \perp S$, and that $I - P$ is the projection of H on $S^{\perp}$.

.4 To each vector a in an inner product space X there corresponds a linear functional u_a defined on X by

$$u_a(x) = \langle x, a \rangle .$$

Prove that u_a is bounded and has norm $\|a\|$; and that if $a \neq 0$, then $z = \|a\|^{-2} a$ is in $\ker(u_a)^{\perp}$, $u_a(z) = 1$, and $a = \|z\|^{-2} z$.

.5 Let f be a nonzero linear functional on the Euclidean space $\mathbf{R}^n$. Prove that there exists a nonzero vector p orthogonal to the hyperplane $\ker(f)$, such that $f(x) = \langle x, p \rangle$ for each $x \in \mathbf{R}^n$. (Choose $a \in \mathbf{R}^n \backslash \ker(f)$ such that $f(a) = 1$. Let b be the foot of the perpendicular from a to $\ker(f)$, and let $p = \lambda(a - b)$ for an appropriate value of λ. Note that each $x \in \mathbf{R}^n$ can be written uniquely in the form $x = f(x)a + y$ with $y \in \ker(f)$.)

A family $(e_i)_{i \in I}$ of elements of an inner product space X is said to be *orthogonal* if $\langle e_i, e_j \rangle = 0$ whenever i, j are distinct indices in I. If, in addition, $\|e_i\| = 1$ for each i, then (e_i) is called an *orthonormal family*; in that case we call $\langle x, e_i \rangle$ the corresponding ith *coordinate* of the element x of X.

For example, in the space $L_2([-\pi, \pi], \mathbf{C})$, taken with the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) g(t)^* \, \mathrm{d}t,$$

the functions e_n $(n = 0, \pm 1, \pm 2, \ldots)$ form an orthonormal sequence, where

$$e_n(t) = \mathrm{e}^{\mathrm{i}nt}.$$

The corresponding nth coordinate of $f \in L_2([-\pi, \pi], \mathbf{C})$ is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \mathrm{e}^{-\mathrm{i}nt} \, \mathrm{d}t,$$

which is better known as the nth *Fourier coefficient* of f.

(5.2.6) Exercise

Verify the mathematical claims made in the last paragraph.

If $(e_i)_{i\in I}$ is an orthonormal family in an inner product space X, then for any finite index set $J \subset I$ the vectors $e_j \quad (j \in J)$ are linearly independent: for if $\sum_{j\in J} \lambda_j e_j = 0$, where each $\lambda_j \in \mathbf{F}$, then for each $i \in J$,

$$0 = \left\langle \sum_{j\in J} \lambda_j e_j,\, e_i \right\rangle = \sum_{j\in J} \lambda_j \langle e_j, e_i \rangle = \lambda_i.$$

Thus the vectors $e_j \quad (j \in J)$ form a basis for a finite–dimensional subspace of X.

(5.2.7) Lemma. *Let $(e_n)_{n=1}^N$ be a finite orthonormal family in an inner product space X. Then for each $x \in X$,*

$$\left\| x - \sum_{n=1}^N \langle x, e_n \rangle e_n \right\|^2 = \|x\|^2 - \sum_{n=1}^N |\langle x, e_n \rangle|^2,$$

$$\left\| \sum_{n=1}^N \langle x, e_n \rangle e_n \right\|^2 = \sum_{n=1}^N |\langle x, e_n \rangle|^2 \le \|x\|^2,$$

and $x - \sum_{n=1}^N \langle x, e_n \rangle e_n$ is orthogonal to each e_k.

Proof. For each n write $\lambda_n = \langle x, e_n \rangle$. Then

$$\begin{aligned} \left\| \sum_{n=1}^N \langle x, e_n \rangle e_n \right\|^2 &= \left\langle \sum_{m=1}^N \lambda_m e_m, \sum_{n=1}^N \lambda_n e_n \right\rangle \\ &= \sum_{m,n=1}^N \lambda_m \lambda_n^* \langle e_m, e_n \rangle = \sum_{n=1}^N |\lambda_n|^2 . \end{aligned}$$

So

$$\begin{aligned} 0 &\le \left\| x - \sum_{n=1}^N \lambda_n e_n \right\|^2 \\ &= \|x\|^2 - \left\langle x, \sum_{n=1}^N \lambda_n e_n \right\rangle - \left\langle \sum_{n=1}^N \lambda_n e_n, x \right\rangle + \left\| \sum_{n=1}^N \lambda_n e_n \right\|^2 \\ &= \|x\|^2 - \sum_{n=1}^N \lambda_n^* \langle x, e_n \rangle - \sum_{n=1}^N \lambda_n \langle x, e_n \rangle^* + \sum_{n=1}^N |\lambda_n|^2 \end{aligned}$$

$$= \|x\|^2 - 2\sum_{n=1}^{N} |\lambda_n|^2 + \sum_{n=1}^{N} |\lambda_n|^2$$
$$= \|x\|^2 - \sum_{n=1}^{N} |\lambda_n|^2 .$$

The first two of the desired conclusions now follow. On the other hand,

$$\left\langle x - \sum_{n=1}^{N} \lambda_n e_n,\, e_k \right\rangle = \langle x, e_k\rangle - \sum_{n=1}^{N} \lambda_n \langle e_n, e_k\rangle = \langle x, e_k\rangle - \lambda_k = 0. \quad \square$$

(5.2.8) Proposition. *If $(e_i)_{i\in I}$ is an orthonormal family in an inner product space X, then for each $x \in X$,*

$$I_x = \{i \in I : \langle x, e_i\rangle \neq 0\}$$

is either empty or countable.

Proof. Lemma (5.2.7) shows that for each finite subset J of I we have

$$\sum_{i\in J} |\langle x, e_i\rangle|^2 \leq \|x\|^2 .$$

Hence the set

$$I_{x,n} = \left\{ i \in I : |\langle x, e_i\rangle|^2 > n^{-1} \|x\|^2 \right\}$$

has at most $n-1$ elements. Since $I_x = \bigcup_{n=1}^{\infty} I_{x,n}$, we conclude that if I_x is nonempty, then it is countable. $\square$

When $(e_i)_{i\in I}$ is an orthonormal family in an inner product space X, Proposition (5.2.8) enables us to make sense of certain summations, such as $\sum_{i\in I} |\langle x, e_i\rangle|^2$, over possibly uncountable index sets. If I_x is empty, we define $\sum_{i\in I} |\langle x, e_i\rangle|^2 = 0$. If I_x is nonempty, it is either finite or countably infinite; taking, for example, the latter case (the former is even easier to handle), we define

$$\sum_{i\in I} |\langle x, e_i\rangle|^2 = \sum_{n=1}^{\infty} |\langle x, f_n\rangle|^2 , \tag{1}$$

where $f_1, f_2, \ldots$ is a one–one enumeration of I_x. Note that the series on the right–hand side converges, since its terms are nonnegative and (by Lemma (5.2.7)) its partial sums are bounded by $\|x\|^2$; it follows from Exercise (1.2.17: 1) that the value of the expression on the left–hand side of (1) is independent of our choice of the one–one enumeration $f_1, f_2, \ldots$ of I_x. Moreover, we have *Bessel's inequality*

$$\sum_{i\in I} |\langle x, e_i\rangle|^2 \leq \|x\|^2 .$$

In turn, when X is a Hilbert space, we can give meaning to another important type of series. Writing

$$s_k = \sum_{n=1}^{k} \langle x, f_n\rangle f_n$$

and using Lemma (5.2.7), we see that if $k > j$, then

$$\|s_j - s_k\|^2 = \left\| \sum_{n=j+1}^{k} \langle x, f_n\rangle f_n \right\|^2 = \sum_{n=j+1}^{k} |\langle x, f_n\rangle|^2 .$$

Since $\sum_{n=1}^{\infty} |\langle x, f_n\rangle|^2$ converges, (s_n) is a Cauchy sequence in X; so, as X is complete, $\sum_{n=1}^{\infty} \langle x, f_n\rangle f_n$ converges to a sum $s \in X$. Likewise, if $f_1', f_2', \ldots$ is another one–one enumeration of I_x, then $\sum_{n=1}^{\infty} \langle x, f_n'\rangle f_n'$ converges to a sum $s' \in X$. We show that $s = s'$. Given $\varepsilon > 0$, we choose N such that if $k \geq N$, then

$$\|s - s_k\| < \varepsilon, \;\; \|s' - s_k'\| < \varepsilon, \text{ and } \sum_{n=k+1}^{\infty} |\langle x, f_n\rangle|^2 < \varepsilon^2,$$

where

$$s_k' = \sum_{n=1}^{k} \langle x, f_n'\rangle f_n'.$$

Taking

$$m = \max\{k : f_k' = f_n \text{ for some } n \leq N\},$$

we see that $m \geq N$ and

$$\|s_m' - s_N\|^2 \leq \sum_{n=N+1}^{\infty} |\langle x, f_n\rangle|^2 < \varepsilon^2.$$

Hence

$$\|s - s'\| \leq \|s - s_N\| + \|s_N - s_m'\| + \|s_m' - s'\| < 3\varepsilon.$$

Since ε is arbitrary, it follows that $s = s'$. Hence the value of

$$\sum_{i\in I} \langle x, e_i\rangle e_i = \sum_{n=1}^{\infty} \langle x, f_n\rangle f_n$$

is independent of the choice of the one–one enumeration $f_1, f_2, \ldots$ of I_x.

(5.2.9) Exercise

Let $(e_i)_{i\in I}$ be an orthonormal family in a Hilbert space H, and x, y elements of H such that I_x is countably infinite. Show that the value of the expression

$$\sum_{i\in I} \langle x, e_i \rangle \langle e_i, y \rangle = \sum_{n=1}^{\infty} \langle x, f_n \rangle \langle f_n, y \rangle$$

is independent of the one–one enumeration $f_1, f_2, \ldots$ of I_x. Show also that if $f_1', f_2', \ldots$ is a (possibly finite) one–one enumeration of I_y, then

$$\sum_{i\in I} \langle x, e_i \rangle \langle e_i, y \rangle = \sum_{n=1}^{\infty} \langle x, f_n' \rangle \langle f_n', y \rangle .$$

(5.2.10) Proposition. *Let $(e_i)_{i\in I}$ be an orthonormal family in a Hilbert space H, let S be the closure in H of the subspace of H generated by (e_i), and let P be the projection of H on S. Then for all x, y in H,*

$$\begin{aligned} Px &= \sum_{i\in I} \langle x, e_i \rangle e_i, \\ \|Px\|^2 &= \sum_{i\in I} |\langle x, e_i \rangle|^2 , \\ \|x - Px\|^2 &= \|x\|^2 - \sum_{i\in I} |\langle x, e_i \rangle|^2 , \\ \langle Px,\ Py \rangle &= \sum_{i\in I} \langle x, e_i \rangle \langle e_i, y \rangle . \end{aligned}$$

Proof. Consider, for example, the case where I_x is countably infinite. Let $f_1, f_2, \ldots$ be a one–one enumeration of I_x. Lemma (5.2.7) shows that

$$\begin{aligned} \left\| x - \sum_{n=1}^{N} \langle x, f_n \rangle f_n \right\|^2 &= \|x\|^2 - \sum_{n=1}^{N} |\langle x, f_n \rangle|^2 , \\ \left\| \sum_{n=1}^{N} \langle x, f_n \rangle f_n \right\|^2 &= \sum_{n=1}^{N} |\langle x, f_n \rangle|^2 , \end{aligned}$$

and $x - \sum_{n=1}^{N} \langle x, f_n \rangle f_n$ is orthogonal to $f_1, \ldots, f_N$. Letting $N \to \infty$, we see that

$$\begin{aligned} \left\| x - \sum_{i\in I} \langle x, e_i \rangle e_i \right\|^2 &= \|x\|^2 - \sum_{i\in I} |\langle x, e_i \rangle|^2 , \\ \left\| \sum_{i\in I} \langle x, e_i \rangle e_i \right\|^2 &= \sum_{i\in I} |\langle x, e_i \rangle|^2 , \end{aligned}$$

and $z = x - \sum_{i\in I} \langle x, e_i\rangle e_i$ is orthogonal to each f_n. For each $i \in I$ either $e_i = f_n$ for some n, and therefore $z \perp e_i$, or else $i \notin I_x$; in the latter case, using the continuity of the inner product, we have

$$\begin{aligned} \langle z, e_i\rangle &= \langle x, e_i\rangle - \left\langle \sum_{n=1}^{\infty} \langle x, f_n\rangle f_n, e_i \right\rangle \\ &= 0 - \sum_{n=1}^{\infty} \langle \langle x, f_n\rangle f_n, e_i\rangle \\ &= -\sum_{n=1}^{\infty} \langle x, f_n\rangle \langle f_n, e_i\rangle \\ &= 0, \end{aligned}$$

as e_i is orthogonal to each f_n. It now follows that z is orthogonal to each vector in S, and hence, by Proposition (5.2.4), that $Px = \sum_{i\in I} \langle x, e_i\rangle e_i$.

Using Exercise (5.2.5: 2), the continuity of the inner product, and Exercise (5.2.9), we now obtain

$$\begin{aligned} \langle Px,\, Py\rangle &= \langle Px, y\rangle \\ &= \left\langle \sum_{n=1}^{\infty} \langle x, f_n\rangle f_n, y \right\rangle \\ &= \sum_{n=1}^{\infty} \langle x, f_n\rangle \langle f_n, y\rangle \\ &= \sum_{i\in I} \langle x, e_i\rangle \langle e_i, y\rangle . \quad \square \end{aligned}$$

By an *orthonormal basis* of a Hilbert space H we mean an orthonormal family that generates a dense linear subspace of H. The following is a more or less immediate consequence of Proposition (5.2.10).

(5.2.11) Proposition. *The following are equivalent conditions on an orthonormal family $(e_i)_{i\in I}$ in a Hilbert space H.*

(i) *(e_i) is an orthonormal basis of H.*

(ii) *$x = \sum_{i\in I} \langle x, e_i\rangle e_i$ for each $x \in H$.*

(iii) *$\sum_{i\in I} |\langle x, e_i\rangle|^2 = \|x\|^2$ for each $x \in H$.*

(iv) *$\langle x, y\rangle = \sum_{i\in I} \langle x, e_i\rangle \langle e_i, y\rangle$ for all $x, y \in H$.*

The identity in condition (iv) of this proposition is known as *Parseval's identity.*

(5.2.12) Exercises

.1 Prove Proposition (5.2.11).

.2 Use Zorn's Lemma (Appendix B) to prove that every nonzero Hilbert space has an orthonormal basis.

.3 Let $(e_i)_{i\in I}$ be an orthonormal basis in a separable Hilbert space H. By considering the balls $B(e_i, 1/\sqrt{2})$, or otherwise, show that I is a countable set.

.4 Let H be an infinite–dimensional inner product space, and $(e_n)_{n=1}^{\infty}$ an infinite orthonormal sequence of vectors in H. By considering (e_n), and without invoking either Theorem (4.3.6) or Exercise (4.3.7: 4), prove that the unit ball of H is not sequentially compact.

.5 Let $(e_n)_{n=1}^{\infty}$ be an orthonormal basis of a separable Hilbert space H, and $(a_n)_{n=1}^{\infty}$ an element of $l_2(\mathbf{C})$. Show that there exists a unique element a of H such that $\langle a, e_n \rangle = a_n$ for each n. (Show that the partial sums of the series $\sum_{n=1}^{\infty} a_n e_n$ form a Cauchy sequence.)

.6 Prove that the functions

$$t \mapsto e_n(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}nt} \quad (n \in \mathbf{Z})$$

form an orthonormal basis of $L_2([-\pi, \pi], \mathbf{C})$. (Noting Exercise (5.2.6), show that the linear space S generated by $\{e_n : n \in \mathbf{Z}\}$ is dense in $L_2([-\pi, \pi], \mathbf{C})$. To do this, first consider $f \in \mathcal{C}([-\pi, \pi], \mathbf{C})$. Construct a continuous function g on $\mathbf{R}$, with period 2π, such that $\|f - g\|_2$ is arbitrarily small. Then use Exercise (4.6.8: 6) to approximate g, and therefore f, by an element of S.)

It follows from this exercise and Proposition (5.2.11) that for each $f \in L_2([-\pi, \pi], \mathbf{C})$ the corresponding *Fourier expansion*

$$x \mapsto \sum_{n=-\infty}^{\infty} \widehat{f}(n) \mathrm{e}^{\mathrm{i}nx}$$

converges to f in the L_2 norm, where

$$\widehat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \mathrm{e}^{-\mathrm{i}nt} \, \mathrm{d}t.$$

In this case Parseval's identity takes the form

$$\int_{-\pi}^{\pi} |f(x)|^2 \, \mathrm{d}x = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \left| \widehat{f}(n) \right|^2.$$

.7 Take $f(x) = x$ in the preceding exercise and apply Parseval's identity, to show that $\sum_{n=1}^{\infty} n^{-2} = \pi^2/6$.

.8 Show that $\sum_{n=1}^{\infty} n^{-4} = \pi^4/90$. (Consider $f(x) = \frac{1}{2}\left(x^2 - \pi^2\right)$.)

Although Zorn's Lemma guarantees the existence of an orthonormal basis in any Hilbert space (Exercise (5.2.12: 2)), it does not enable us to construct orthonormal bases. The following *Gram–Schmidt orthonormalisation process* enables us to construct orthonormal bases when the Hilbert space H is separable.

Using Proposition (4.3.8), first construct a (possibly finite) total sequence $(a_1, a_2, \ldots)$ of linearly independent vectors in H. For each n let H_n be the n–dimensional subspace of H spanned by $\{a_1, \ldots, a_n\}$; since this subspace is complete (by Proposition (4.3.3)), the projection P_n of H onto it is defined. Suppose we have found orthogonal vectors $b_1, \ldots, b_n$ generating H_n. If $H = H_n$, stop the construction. Otherwise, $a_{n+1} \notin H_n$; so by Proposition (5.2.4),

$$b_{n+1} = a_{n+1} - P_n a_{n+1}$$

is orthogonal to H_n, and therefore

$$\langle b_{n+1}, b_k \rangle = 0 \quad (1 \leq k \leq n).$$

Elementary linear algebra shows that $\{b_1, \ldots, b_{n+1}\}$ is a basis of H_{n+1}. This completes the inductive construction of a (possibly finite) orthogonal sequence $(b_1, b_2, \ldots)$ in H such that for each n, $\{b_1, \ldots, b_n\}$ is a basis of H_n. Setting $e_n = \|b_n\|^{-1} b_n$ and noting that $\bigcup_n H_n$ is dense in H, we see that (e_n) is an orthonormal basis of H.

The Gram–Schmidt orthonormalisation process has a very important application in approximation theory, which we now describe.

Let w be a nonnegative continuous weight function on a compact interval $I = [a, b]$. Define the inner product

$$\langle f, g \rangle_w = \int_a^b w(t) f(t) g(t) \, \mathrm{d}t$$

on $L_2(I)$, and the corresponding *weighted least squares norm* by

$$\|f\|_{2,w} = \left(\int_a^b w(t) f(t)^2 \, \mathrm{d}t \right)^{1/2}.$$

Given an element f of $L_2(I)$ and a natural number N, we have the approximation problem:

Find the polynomial function p *of degree at most* N *that minimises the value of*

$$\|f-p\|_{2,w}^2=\int_a^b w(t)\,(f(t)-p(t))^2\,\mathrm{d}t.$$

This polynomial is called the *least squares approximation to* f *of degree at most* N.

Now, the set $\mathcal{P}_N$ of polynomials of degree $\leq N$ is a finite–dimensional subspace of $\mathcal{C}(I)$; so the projection P_N of $\mathcal{C}(I)$ on $\mathcal{P}_N$ exists, and the unique least squares approximation to f of degree at most N is given by $p_N=P_Nf$. To compute the coefficients of p_N, we can use elementary multivariate calculus to calculate the values of $\lambda_0,\ldots,\lambda_N$ that minimise

$$\int_a^b w(t)\left(f(t)-\sum_{n=0}^N \lambda_n t^n\right)^2\,\mathrm{d}t;$$

see [29]. However, this procedure is computationally inefficient if we are looking for least squares approximations to several functions in $\mathcal{C}(I)$. In that case a better procedure is to apply the Gram–Schmidt process to the total sequence consisting of the monomials $1,t,t^2,\ldots$, to compute orthonormal polynomials $q_0,q_1,\ldots$, where $q_n(t)$ has degree n and $\{q_0,\ldots,q_n\}$ is a basis for $\mathcal{P}_n$; then

$$P_Nf=\sum_{n=0}^N \langle f,q_n\rangle_w\ q_n,$$

by Proposition (5.2.10). One advantage of this method is that, having found the least squares approximation p_n to f of degree at most n, in order to find the least squares approximation of degree at most $n+1$ we simply add to p_n the single term $\langle f,q_{n+1}\rangle_w\ q_{n+1}$.

(5.2.13) Exercises

.1 In the notation of the preceding paragraphs, take $I=[-1,1]$ and $w(t)=1$, and compute q_0,q_1, and q_2. Hence find the quadratic least squares approximation to e^x in $\mathcal{C}[-1,1]$.

.2 Let w be a nonnegative continuous weight function on $I=[a,b]$, let $f\in\mathcal{C}(I)$, and for each n let p_n denote the least squares approximation to f of degree at most n. Prove that

$$\lim_{n\to\infty}\|f-p_n\|_{2,w}=0.$$

.3 In the notation of the last exercise, let (q_n) be a sequence of polynomial functions that is orthogonal relative to $\langle\cdot,\cdot\rangle_w$, such that q_n has degree n. Prove that each polynomial p of degree n can be written uniquely as a linear combination of $q_0, \ldots, q_n$, and find the coefficient of q_k in this linear combination.

.4 Continuing Exercise (5.2.13: 3), prove that $q_n(t)$ has n distinct real zeroes, and that those zeroes lie in the open interval (a, b). (Let

$$p(t) = (t - t_1) \cdots (t - t_m),$$

where $t_1, \ldots, t_m$ are the zeroes of $q_n(t)$ in (a, b) at which $q_n(t)$ changes sign. Assume that $m < n$, show that $\int_a^b w(t)p(t)q_n(t)\,\mathrm{d}t \neq 0$, and deduce a contradiction.)

.5 Continuing Exercise (5.2.13: 4), write

$$\begin{aligned} q_n(t) &= A_n t^n + B_n t^{n-1} + \ldots, \\ c_n &= \langle q_n, q_n \rangle_w, \\ \alpha_n &= \frac{A_{n+1}}{A_n}, \\ \beta_n &= \alpha_n \left(\frac{B_{n+1}}{A_{n+1}} - \frac{B_n}{A_n} \right), \end{aligned}$$

and, for $n \geq 1$,

$$\gamma_n = \frac{A_{n+1}A_{n-1}}{A_n^2} \cdot \frac{\langle q_n, q_n \rangle_w}{\langle q_{n-1}, q_{n-1} \rangle_w}.$$

Prove the *triple recursion formula*:

$$q_{n+1}(t) = (\alpha_n t + \beta_n) q_n(t) - \gamma_n q_{n-1}(t).$$

(Consider $p(t) = q_{n+1}(t) - \alpha_n t q_n(t)$.)

.6 Let $I = [a, b]$, let $w \in \mathcal{C}(I)$, and let p be a polynomial function. Prove the equivalence of the following conditions.

(i) $\int_a^b w(t)p(t)q(t)\,\mathrm{d}t = 0$ for all polynomial functions q of degree at most n.

(ii) There exists an $(n+1)$-times differentiable function u on I such that

$$w(x)p(x) = u^{(n+1)}(x) \quad (x \in I)$$

and

$$u^{(k)}(a^+) = u^{(k)}(b^-) = 0 \quad (k = 0, 1, \ldots, n).$$

.7 Let $I = [-1, 1]$, let $\alpha, \beta \in (-1, \infty)$, and let

$$w(x) = (1 - x)^{\alpha} (1 + x)^{\beta} \quad (x \in I).$$

For each $n \in \mathbf{N}$ define the *Jacobi polynomial of degree* n by *Rodrigues's formula*:

$$\phi_n(x) = (1 - x)^{-\alpha} (1 + x)^{-\beta} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left((1 - x)^{\alpha+n} (1 + x)^{\beta+n} \right)$$

(where, of course, $\mathrm{d}^n/\mathrm{d}x^n$ denotes the nth derivative). Use the preceding exercise to prove that $(\phi_n)_{n=0}^{\infty}$ is an orthogonal sequence in $L_{2,w}(I, \mathbf{C})$.

.8 In the special case $\alpha = \beta = 0$ of the last exercise, the Jacobi polynomial ϕ_n is known as a *Legendre polynomial* and is usually denoted by P_n. Prove that the Legendre polynomials satisfy the recurrence relation

$$P_{n+1}(x) = (4n + 2)xP_n(x) - 4n^2 P_{n-1}(x)$$

on $[-1, 1]$. Use this and Exercise (5.2.13: 1) to find $P_3(x)$ and $P_4(x)$. (To establish the recurrence relation, write each term in the form

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}} \left(\left(x^2 - 1\right)^{n-1} q(x) \right),$$

where $q(x)$ is a quadratic polynomial.)

5.3 The Dual of a Hilbert Space

We saw in Exercise (5.2.5: 4) that for each vector a in an inner product space X the mapping $x \mapsto \langle x, a \rangle$ is a bounded linear functional on X. We now show that the dual of a Hilbert space consists precisely of bounded linear functionals of this form (cf. Exercise (5.2.5: 5)).

(5.3.1) The Riesz Representation Theorem. *If u is a bounded linear functional on a Hilbert space H, then there exists a unique vector $a \in H$ such that $u(x) = \langle x, a \rangle$ for each $x \in H$. In that case $\|u\| = \|a\|$.*

Proof. We first dispose of the uniqueness[1] of a: indeed, if

$$\langle x, a \rangle = u(x) = \langle x, a' \rangle \quad (x \in H),$$

[1] This uniqueness argument applies to a linear functional of the form $x \mapsto \langle x, a \rangle$ on an inner product space.

then, taking $x = a - a'$, we obtain

$$\|a - a'\|^2 = \langle a - a', a - a' \rangle = 0,$$

so $a = a'$.

To establish the existence of a, we may assume that $u \neq 0$. As u is linear and continuous, $\ker(u)$ is a closed subspace of H (Proposition (4.2.3)). Let P be the projection of H on $\ker(u)$, and choose $y \in H$ such that $u(y) \neq 0$. Setting

$$z = u(y)^{-1}(y - Py),$$

we see that $z \in \ker(u)^{\perp}$, by Proposition (5.2.4), and that

$$u(z) = u(y)^{-1}(u(y) - u(Py)) = 1.$$

So for each x in H we have

$$x - u(x)z \in \ker(u)$$

and therefore

$$0 = \langle x - u(x)z, z \rangle = \langle x, z \rangle - u(x) \langle z, z \rangle = \langle x, z \rangle - u(x) \|z\|^2.$$

Thus $u(x) = \langle x, a \rangle$, where $a = \|z\|^{-2} z$. The Cauchy–Schwarz inequality shows that $|u(x)| \leq \|a\| \|x\|$. Since also $\left| u \left(\|a\|^{-1} a \right) \right| = \|a\|$, we see that $\|u\| = \|a\|$. □

(5.3.2) Exercises

.1 Find an alternative proof of the existence part of the Riesz Representation Theorem (5.3.1) for a separable Hilbert space H. (Let $(e_n)_{n=1}^{\infty}$ be an orthonormal basis of H, and u a bounded linear functional on H. Show that $\sum_{n=1}^{\infty} u(e_n)^* e_n$ converges to the desired element $a \in H$.)

.2 Use the Riesz Representation Theorem to give another solution to Exercise (5.2.12: 5). (In the notation of that exercise, show that $u(x) = \sum_{n=1}^{\infty} a_n^* \langle x, e_n \rangle$ defines a bounded linear functional on H.)

.3 By the *second dual* of a normed space X we mean the dual space $X^{**} = (X^*)^*$ of X^*. We say that X is *reflexive* if for each $u \in X^{**}$ there exists $x_u \in X$ such that $u(f) = f(x_u)$ for each $f \in X^*$. Prove that any Hilbert space is reflexive.

By an *operator* on a normed space X we mean a bounded linear mapping from X into itself; the set of operators on X is written $L(X)$. (Strictly speaking, we have here defined a *bounded* operator; since we do not consider unbounded operators in this book, it is convenient for us to use the term "operator" to mean "bounded operator".)

It is common practice to denote the composition of operators by juxtaposition; thus if S, T are operators on X, then $T \circ S$ is usually written TS; moreover, we write T^2 for TT, T^3 for $T(TT)$, and so on.

For a first application of the Riesz Representation Theorem, let T be an operator on a Hilbert space H, and for each $a \in X$ consider the linear functional $x \mapsto \langle Tx, a\rangle$ on H. The inequality

$$|\langle Tx, a\rangle| \leq \|Tx\| \, \|a\| \leq \|T\| \, \|a\| \, \|x\|$$

shows that this functional is bounded and has norm at most $\|T\| \, \|a\|$. By the Riesz Representation Theorem, there exists a unique vector T^*a such that

$$\langle Tx, a\rangle = \langle x, T^*a\rangle \quad (x \in H);$$

moreover,

$$\|T^*a\| \leq \|T\| \, \|a\|. \tag{1}$$

The mapping $T^* : H \to H$ so defined is called the *adjoint* of T, and is an operator on H. To justify this last claim, consider a, b in H and λ, μ in $\mathbf{F}$. Since

$$\begin{aligned}
\langle Tx, \lambda a + \mu b\rangle &= \lambda^* \langle Tx, a\rangle + \mu^* \langle Tx, b\rangle \\
&= \lambda^* \langle x, T^*a\rangle + \mu^* \langle x, T^*b\rangle \\
&= \langle x, \lambda T^*a + \mu T^*b\rangle
\end{aligned}$$

for all $x \in H$, we see that

$$T^*(\lambda a + \mu b) = \lambda T^*a + \mu T^*b.$$

So T^* is linear. Inequality (1) shows that T^* is bounded and has norm at most $\|T\|$. Since

$$\langle T^*x, y\rangle = \langle y, T^*x\rangle^* = \langle Ty, x\rangle^* = \langle x, Ty\rangle,$$

the uniqueness of the adjoint of T^* shows that $(T^*)^* = T$. So $\|T\| = \|(T^*)^*\| \leq \|T^*\|$ and therefore $\|T^*\| = \|T\|$.

An operator T on H is said to be

- *selfadjoint*, or *Hermitian*, if $T^* = T$;
- *normal* if $T^*T = TT^*$.

Selfadjoint and normal operators have particularly amenable properties and are among the most important objects in Hilbert space theory. (See [24], [44], and other books that deal with such topics as spectral theory.)

(5.3.3) Exercises

In all the exercises of this set except the first, H is a complex Hilbert space, S and T are operators on H, and $\operatorname{ran}(T)$ denotes the range of T.

.1 Let $(e_1, e_2, \ldots, e_n)$ be an orthonormal basis of the Euclidean Hilbert space $\mathbf{F}^n$, and T an operator on $\mathbf{F}^n$. Show that

$$Tx = \sum_{j,k=1}^{n} \langle x, e_j \rangle \langle Te_j, e_k \rangle e_k,$$

and hence that T can be associated with the n–by–n matrix whose (j,k)th entry is $\langle Te_j, e_k \rangle$. With what matrix is T^* associated in this way?

.2 By a *bounded conjugate–bilinear functional* on H we mean a mapping $u : H \times H \to \mathbf{C}$ that is linear in the first variable, conjugate linear in the second, and bounded, in the sense that there exists $c > 0$ such that $|u(x,y)| \leq c \|x\| \|y\|$ for all $x, y \in H$. The least such c is the number written

$$\|u\| = \sup \{|u(x,y)| : x, y \in H, \|x\| \leq 1, \|y\| \leq 1\}.$$

Show that the mapping $u : H \times H \to \mathbf{C}$ defined by

$$u(x,y) = \langle Tx, y \rangle \tag{2}$$

is a bounded conjugate–linear functional on H such that $\|u\| = \|T\|$. Show also that each bounded conjugate–linear functional u on H is related to a unique corresponding operator T as in equation (2). (For the second part, show that for each $x \in H$ the mapping $y \mapsto u(x,y)^*$ is a bounded linear functional on H.)

.3 Verify the *polarisation identity*:

$$\begin{aligned} 4\langle Tx, y \rangle &= \langle T(x+y), x+y \rangle - \langle T(x-y), x-y \rangle \\ &\quad + \mathrm{i} \langle T(x+\mathrm{i}y), x+\mathrm{i}y \rangle - \mathrm{i} \langle T(x-\mathrm{i}y), x-\mathrm{i}y \rangle. \end{aligned}$$

Show that if $\langle Sx, x \rangle = \langle Tx, x \rangle$ for all $x \in H$, then $S = T$.

Give an example of a nonzero operator T on the real Hilbert space $\mathbf{R}^2$ such that $\langle Tx, x \rangle = 0$ for all $x \in \mathbf{R}^2$.

.4 Let λ, μ be complex numbers. Show that $(\lambda S + \mu T)^* = \lambda^* S^* + \mu^* T^*$ and $(ST)^* = T^* S^*$.

.5 Prove that T^*T and TT^* are selfadjoint.

.6 Prove each of the following statements.

(i) $\ker(T^*) = \operatorname{ran}(T)^\perp$

(ii) $\overline{\operatorname{ran}(T^*)} = \ker(T)^\perp$.

(iii) $\ker(T) = \ker(T^*T)$.

(iv) $\operatorname{ran}(TT^*)$ is dense in $\operatorname{ran}(T)$.

.7 Show that

(i) T is selfadjoint if and only if $\langle Tx, x\rangle \in \mathbf{R}$ for all $x \in H$.

(ii) T is normal if and only if $\|Tx\| = \|T^*x\|$ for each $x \in H$.

(For part (i), consider $\langle Tx, x\rangle - \langle T^*x, x\rangle$, and note Exercise (5.3.3: 3).)

.8 Prove that T is a projection if and only if $T^*T = T$, in which case T is *idempotent*—that is, $T^2 = T$. (For "if", show first that T is selfadjoint, and then that $(x - Tx) \perp Ty$ for all $x, y \in H$.)

We close this chapter by sketching how the techniques of Hilbert space theory can be applied to prove the existence of a type of solution for one of the fundamental problems of potential theory. (For more information on this topic, see, for example, pages 117–122 of [23].)

For the rest of this chapter only, we follow the usual notational conventions of applied mathematicians. Thus we denote three–dimensional vectors by boldface letters, the element of volume in $\mathbf{R}^3$ by $\mathrm{d}V$, the element of surface area by $\mathrm{d}S$, the unit outward normal to a surface by $\mathbf{n}$, and the inner product of two vectors $\mathbf{u}, \mathbf{v}$ in $\mathbf{R}^3$ by $\mathbf{u}\cdot\mathbf{v}$. We assume familiarity with calculus in $\mathbf{R}^3$, including the elementary vector analysis of the *gradient operator* ∇ and the *divergence operator* div, defined, respectively, by

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$$

for a real–valued function f, and

$$\operatorname{div}\mathbf{u} = \left(\frac{\partial u_x}{\partial x}, \frac{\partial u_y}{\partial y}, \frac{\partial u_z}{\partial z}\right)$$

for a vector $\mathbf{u} = (u_x, u_y, u_z)$. We also assume the fundamentals of the theory of $L_2(\Omega)$ when Ω is a Lebesgue measurable subset of $\mathbf{R}^3$.

Let Ω be a bounded open set in $\mathbf{R}^3$ for which *Gauss's Divergence Theorem* holds:

$$\int_\Omega \operatorname{div}\mathbf{u}\,\mathrm{d}V = \int_{\partial\Omega} \mathbf{u}\cdot\mathbf{n}\,\mathrm{d}S,$$

where $\partial\Omega$ is the boundary surface of Ω and $\mathbf{u} : \overline{\Omega} \to \mathbf{R}^3$ is continuously differentiable on Ω. It follows that *Green's Theorem* holds in the form

$$\int_\Omega \left(u\,\nabla^2 v - v\,\nabla^2 u\right)\,\mathrm{d}V = \int_{\partial\Omega}\left(u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n}\right)\,\mathrm{d}S,$$

where u, v are twice continuously differentiable mappings of Ω into $\mathbf{R}$, $\partial/\partial n$ denotes differentiation along the outward normal to $\partial\Omega$, and ∇^2 is the *Laplacian operator,*

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

We assume the following result, embodying *Poincaré's inequality.*

> *There exists a constant $c > 0$ such that if $v : \overline{\Omega} \to \mathbf{R}$ is differentiable on Ω and vanishes on the boundary of Ω, then*
>
> $$\left(\int_\Omega v^2 \, \mathrm{d}V \right)^{1/2} \leq c \left(\int_\Omega \|\nabla v\|^2 \, \mathrm{d}V \right)^{1/2}.$$

For a proof of this inequality under reasonable conditions on Ω we refer to [37], Chapter 5, Theorem 1.

Given a bounded continuous function $f : \Omega \to \mathbf{R}$, we consider the corresponding *Dirichlet Problem*:

> *Find a function $u : \overline{\Omega} \to \mathbf{R}$ that is twice differentiable on Ω, satisfies $\nabla^2 u = f$ on Ω, and vanishes on the boundary of Ω.*

Suppose we have found a solution u of this Dirichlet Problem. Let $v : \Omega \to \mathbf{R}$ be twice differentiable and have compact support in Ω —that is, $v = 0$ outside some compact subset of Ω. Then it follows from Green's Theorem that

$$\int_\Omega u \, \nabla^2 v \, \mathrm{d}V = \int_\Omega v f \, \mathrm{d}V, \tag{3}$$

since both u and v vanish on $\partial\Omega$.

Now, it may not be possible to solve the Dirichlet Problem on Ω; but, as we now show, we can find a function u on $\overline{\Omega}$ that behaves appropriately on $\partial\Omega$ and that satisfies (3) for all $v : \Omega \to \mathbf{R}$ that are twice differentiable and have compact support in Ω. More advanced theory of partial differential equations then provides conditions on Ω under which this so–called *weak solution* u of the Dirichlet Problem can be identified with a solution of the standard type.

Let $\mathcal{C}_0^1(\overline{\Omega})$ be the space of functions $u : \overline{\Omega} \to \mathbf{R}$ that have compact support in Ω and are differentiable on Ω; and let $\mathcal{C}^1(\overline{\Omega})$ be the space of functions $u : \overline{\Omega} \to \mathbf{R}$ such that u is differentiable on Ω and ∇u extends to a continuous function on $\overline{\Omega}$. Let $\widetilde{\mathcal{C}}^1(\overline{\Omega})$ be the space consisting of all elements of $\mathcal{C}^1(\overline{\Omega})$ that vanish on $\partial\Omega$, $H_0^1(\overline{\Omega})$ the completion of $\widetilde{\mathcal{C}}^1(\overline{\Omega})$ with respect to the inner product defined by

$$\langle u, v \rangle = \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}V,$$

and $\|\cdot\|_H$ the corresponding norm on $H_0^1(\overline{\Omega})$. It is not hard to show that $\mathcal{C}_0^1(\overline{\Omega})$ is dense in $H_0^1(\overline{\Omega})$ with respect to this norm, and that $H_0^1(\overline{\Omega})$ can be identified with a certain set of Lebesgue integrable real–valued functions u on $\overline{\Omega}$.

Now define a linear functional φ_f on $\widetilde{\mathcal{C}}^1(\overline{\Omega})$ by

$$\varphi_f(v) = \int_\Omega vf \, \mathrm{d}V.$$

Applying the Cauchy–Schwarz inequality in the Hilbert space $L_2(\Omega)$, we obtain

$$|\varphi_f(v)| \leq \left(\int_\Omega v^2 \, \mathrm{d}V\right)^{1/2} \left(\int_\Omega f^2 \, \mathrm{d}V\right)^{1/2}.$$

Hence, by Poincaré's inequality,

$$\begin{aligned} |\varphi_f(v)| &\leq c\left(\int_\Omega f^2 \, \mathrm{d}V\right)^{1/2} \left(\int_\Omega \|\nabla v\|^2 \, \mathrm{d}V\right)^{1/2} \\ &= c\left(\int_\Omega f^2 \, \mathrm{d}V\right)^{1/2} \|v\|_H , \end{aligned}$$

where the constant c is independent of v. Thus the linear functional φ_f is bounded. It therefore extends by continuity to a bounded linear functional φ_f on $H_0^1(\overline{\Omega})$; see Exercise (4.2.2: 10). Thus, by the Riesz Representation Theorem (5.3.1), there exists a unique element u of $H_0^1(\overline{\Omega})$ such that

$$\varphi_f(v) = -\langle v, u\rangle \quad \left(v \in H_0^1(\overline{\Omega})\right).$$

For each v that has compact support in Ω and is twice differentiable on Ω, we now use the elementary vector identity

$$\operatorname{div}(u\nabla v) = \nabla u \cdot \nabla v + u\,\nabla^2 v$$

and Gauss's Divergence Theorem to show that

$$\begin{aligned} \int_\Omega u\,\nabla^2 v \, \mathrm{d}V &= -\langle v, u\rangle + \int_\Omega \operatorname{div}(u\nabla v) \, \mathrm{d}V \\ &= \varphi_f(v) + \int_{\partial\Omega} u\nabla v \cdot \mathbf{n} \, \mathrm{d}S \\ &= \int_\Omega vf \, \mathrm{d}V. \end{aligned}$$

(Recall that $v = 0$ on the boundary of Ω). This completes the proof that u is the weak solution that we wanted.

6

An Introduction to Functional Analysis

...a wonderful piece of work; which not to have been blessed withal would have discredited your travel.
ANTONY AND CLEOPATRA, Act 1, Scene 2

In this chapter we first discuss the Hahn–Banach Theorem, the most famous case of which provides conditions under which a bounded linear functional on a subspace of a normed space X can be extended, with preservation of its norm, to a bounded linear functional on the whole of X. We then present several applications of this theorem, some of which illustrate the interplay between a normed space and its dual. In Section 2 we use the Hahn–Banach Theorem to obtain results about the separation of convex sets by hyperplanes. The last section of the chapter introduces the Baire Category Theorem, and includes some of its many applications in classical and functional analysis.

6.1 The Hahn–Banach Theorem

Let X be a linear space over $\mathbf{F}$. If $\mathbf{F} = \mathbf{C}$, then by a *complex–linear functional* on X we mean a mapping $f : X \to \mathbf{C}$ such that

$$f(x + y) = f(x) + f(y)$$

and

$$f(\lambda x) = \lambda f(x)$$

for all $x, y \in X$ and all $\lambda \in \mathbf{C}$. If f maps X into $\mathbf{R}$ and satisfies these equations for all real numbers λ, then f is called a *real–linear functional* on X.

According to our first lemma, real–linear functionals can be characterised as the real parts of associated complex–linear functionals.

(6.1.1) Lemma. *Let X be a complex normed linear space. If f is a complex–linear functional on X and u is the real part of f, then u is a real–linear functional on X and*

$$f(x) = u(x) - \mathrm{i}u(\mathrm{i}x) \quad (x \in X). \tag{1}$$

If u is a real–linear functional on X and f is defined by equation (1)*, then f is a complex–linear functional on X. Moreover, if f and u are related as in equation* (1) *and either f or u is bounded, then both functionals are bounded and $\|f\| = \|u\|$.*

Proof. If f is a complex–linear functional on X and $u = \mathrm{Re}\,(f)$, then it is easy to show that u is real–linear; moreover, equation (1) follows from the fact that $z = \mathrm{Re}\,(z) - \mathrm{i}\,\mathrm{Re}(\mathrm{i}z)$ for any complex number z. On the other hand, if u is a real–linear functional on X, and f is defined as in (1), then it is clear that $f(x+y) = f(x) + f(y)$, and that $f(\lambda x) = \lambda f(x)$ for all *real* λ. Also,

$$\begin{aligned} f(\mathrm{i}x) &= u(\mathrm{i}x) - \mathrm{i}u(\mathrm{i}^2 x) \\ &= u(\mathrm{i}x) - \mathrm{i}u(-x) \\ &= u(\mathrm{i}x) + \mathrm{i}u(x) \\ &= \mathrm{i}f(x), \end{aligned}$$

from which it follows that f is complex–linear.

If f is bounded, then as $|u(x)| \leq |f(x)|$ for all $x \in X$, u is bounded and $\|u\| \leq \|f\|$. For each $x \in X$ there exists $\lambda \in \mathbf{C}$ such that $|\lambda| = 1$ and $f(\lambda x) = \lambda f(x) = |f(x)|$; then $f(\lambda x) \in \mathbf{R}$, so

$$\begin{aligned} |f(x)| &= f(\lambda x) \\ &= \mathrm{Re}\,(f(\lambda x)) \\ &= u(\lambda x) \\ &\leq \|u\| \, \|\lambda x\| = \|u\| \, \|x\| \, . \end{aligned}$$

Hence $\|f\| \leq \|u\|$, and therefore $\|f\| = \|u\|$. Finally, if u is bounded, then for all $x \in X$ with $\|x\| \leq 1$ we have

$$\begin{aligned} |f(x)| &\leq |u(x)| + |u(\mathrm{i}x)| \\ &\leq \|u\| \left(\|x\| + \|\mathrm{i}x\|\right) \\ &\leq 2\,\|u\| \, , \end{aligned}$$

so f is bounded. By the foregoing, $\|f\| = \|u\|$. □

Let X be a vector space over $\mathbf{F}$, and p a mapping of X into $\mathbf{R}$. We say that p is

- *subadditive* if $p(x+y) \leq p(x) + p(y)$ for all $x, y \in X$;
- *positively homogeneous* if $p(\lambda x) = \lambda p(x)$ for all $x \in X$ and $\lambda \geq 0$;
- a *sublinear functional* if it is subadditive and positively homogeneous;
- a *seminorm* if it is nonnegative and subadditive, and if $p(\lambda x) = |\lambda|\, p(x)$ for all $x \in X$ and $\lambda \in \mathbf{F}$.

For example, if $c \geq 0$, then $p(x) = c\,\|x\|$ defines a sublinear functional on X.

Now let X_0 be a subspace of X, f_0 a linear functional on X_0, and f a linear functional on X. We say that f *extends* f_0 to X, or that f is an *extension* of f_0 to X, if $f(x) = f_0(x)$ for all $x \in X_0$. If also f is bounded and $\|f\| = \|f_0\|$, we say that f is a *norm–preserving extension* of f_0 to X.

We now prove a preliminary version of the extension theorem for linear functionals.

(6.1.2) Proposition. *Let X be a real normed space, X_0 a subspace of X, x_1 a point of $X \backslash X_0$, and X_1 the subspace of X spanned by $X_0 \cup \{x_1\}$. Let p be a sublinear functional on X, and f_0 a linear functional on X_0 such that $f_0(x) \leq p(x)$ for all $x \in X_0$. Then there exists a linear functional f that extends f_0 to X_1 and satisfies $f(x) \leq p(x)$ for all $x \in X_1$.*

Proof. Since $x_1 \notin X_0$, each element of X_1 can be written uniquely in the form $x + \lambda x_1$ with $x \in X_0$ and $\lambda \in \mathbf{R}$. Let τ be any real number, and provisionally define

$$f(x + \lambda x_1) = f_0(x) + \lambda\tau.$$

It is easily shown that f is a linear extension of f_0 to X_1; hence it remains to choose τ so that

$$f_0(x) + \lambda\tau \leq p(x + \lambda x_1) \quad (x \in X_0, \lambda \in \mathbf{R}\backslash\{0\}). \tag{2}$$

To this end, replacing x by λx, using the positive homogeneity of p, and then dividing both sides of (2) by $|\lambda|$, we observe that (2) is equivalent to the two conditions

$$\begin{array}{rcll} f_0(x) + \tau & \leq & p(x + x_1) & \text{if } x \in X_0 \text{ and } \lambda > 0, \\ -f_0(x) - \tau & \leq & p(-x - x_1) & \text{if } x \in X_0 \text{ and } \lambda < 0. \end{array}$$

In turn, these two conditions can be gathered together in one:

$$-p(-x' - x_1) - f_0(x') \leq \tau \leq p(x + x_1) - f_0(x) \quad (x, x' \in X_0).$$

But for all $x, x' \in X_0$ we have

$$\begin{aligned} f_0(x) - f_0(x') &= f_0(x - x') \\ &\leq p(x - x') \\ &= p(x + x_1 - x' - x_1) \\ &\leq p(x + x_1) + p(-x' - x_1) \end{aligned}$$

and therefore

$$-p(-x' - x_1) - f_0(x') \leq p(x + x_1) - f_0(x).$$

Thus in order to satisfy (2), and thereby complete the proof, we need only invoke Exercise (1.1.1: 21). □

This brings us to the *Hahn–Banach Theorem.*

(6.1.3) Theorem. *Let X_0 be a subspace of a real normed space X, p a sublinear functional on X, and f_0 a linear functional on X_0 such that $f_0(x) \leq p(x)$ for all x in X_0. Then there exists a linear functional f that extends f_0 to X and satisfies $f(x) \leq p(x)$ for all $x \in X$.*

Proof. Let $\mathcal{F}$ denote the set of all linear functionals f that are defined on subspaces of X containing X_0 and that have the following properties.

(i) $f = f_0$ on X_0 and

(ii) $f(x) \leq p(x)$ for all x in the domain of f.

Define the binary relation $\preccurlyeq$ on $\mathcal{F}$ by inclusion:

$$f \preccurlyeq g \text{ if and only if } f \subset g.$$

Then $\preccurlyeq$ is a partial order on $\mathcal{F}$. Let $\mathcal{C}$ be a chain in $\mathcal{F}$ (that is, a nonempty totally ordered subset of $\mathcal{F}$), and define

$$G = \bigcup_{g \in \mathcal{C}} g = \{(x, y) : \exists g \in \mathcal{C}\ (y = g(x))\}.$$

If $(x, y_1) \in G$ and $(x, y_2) \in G$, then there exist $g_1, g_2 \in \mathcal{C}$ such that $(x, y_1) \in g_1$ and $(x, y_2) \in g_2$. Since $\mathcal{C}$ is a chain, either $g_1 \subset g_2$ or else, as we may assume, $g_2 \subset g_1$; then $(x, y_2) \in g_1$ and therefore, as g_1 is a function, $y_2 = y_1$. It follows that G is a function on X; and that if x is in the domain of some $g \in \mathcal{C}$, then x is in the domain of G, $G(x) = g(x)$, and therefore $G(x) \leq p(x)$. It is easy to show that the domain of G contains X_0, and that $G = f_0$ on X_0. To complete the proof that $G \in \mathcal{F}$, we must show that G is linear on X. To this end, given x, x' in the domain of G, choose $g, g' \in \mathcal{C}$ such that $(x, G(x)) \in g$ and $(x', G(x')) \in g'$. As $\mathcal{C}$ is a chain, we may assume that $g' \subset g$, so that $(x', G(x')) \in g$; as g is linear, it follows that $x + x'$ is in the domain of g and therefore in the domain of G, and that

$$\begin{aligned} G(x+x') &= g(x+x') \\ &= g(x) + g(x') \\ &= G(x) + G(x'). \end{aligned}$$

Similarly, for each $\lambda \in \mathbf{R}$, $G(\lambda x) = \lambda G(x)$. Hence $G \in \mathcal{F}$. It is trivial to verify that G is an upper bound of $\mathcal{C}$ in $\mathcal{F}$.

We can now apply Zorn's Lemma (see Appendix B) to produce a maximal element f of $\mathcal{F}$. It only remains to show that f is defined throughout X. But if f is not defined at some point x_0 of X, then, using Proposition (6.1.2), we can find an element g of $\mathcal{F}$ such that $f \preccurlyeq g$ and g is defined at x_0. Since f is maximal in $\mathcal{F}$, it follows that $f = g$, a contradiction. □

The name "Hahn–Banach Theorem" is often applied to the following corollary.

(6.1.4) Corollary. *Let X_0 be a subspace of a normed space X, and f_0 a bounded linear functional on X_0. Then there exists a norm–preserving extension of f_0 to X.*

Proof. First consider the case where f_0 is a real–linear functional on X_0. Applying Theorem (6.1.3) with $p(x) = \|f_0\|\,\|x\|$, we obtain a real–linear functional f that extends f_0 to X and satisfies $f(x) \leq \|f_0\|\,\|x\|$ for all $x \in X$. Replacing x by $-x$ in this last inequality, we see that

$$|f(x)| = \max\{f(x), -f(x)\} \leq \|f_0\|\,\|x\|$$

for all $x \in X$; whence f is bounded, and $\|f\| \leq \|f_0\|$. But f extends f_0, so $\|f\| \geq \|f_0\|$ and therefore $\|f\| = \|f_0\|$.

When f is a complex–linear functional, we apply the foregoing argument to construct a norm–preserving extension u of the real–linear functional $\mathrm{Re}(f_0)$ to X. Lemma (6.1.1) then shows us that

$$f(x) = u(x) - \mathrm{i}u(\mathrm{i}x)$$

defines a norm–preserving extension of f_0 to X. □

(6.1.5) Exercises

.1 Prove the *complex Hahn–Banach Theorem*: let X_0 be a subspace of a complex normed space X, p a seminorm on X, and f_0 a linear functional on X_0 such that $|f_0(x)| \leq p(x)$ for all $x \in X_0$; then there exists a linear functional f that extends f_0 to X and satisfies $|f(x)| \leq p(x)$ for all $x \in X$. (First apply the Hahn–Banach Theorem to the real–linear functional $\mathrm{Re}(f_0)$.)

.2 Let X be a separable normed space. Prove Theorem (6.1.3) without using Zorn's Lemma. (Let (x_n) be a dense sequence in X, and p a sublinear functional on X. Starting with a given linear functional f_0 on a subspace X_0 of X, extend f_0 inductively to the subspace X_n of X spanned by $X_{n-1} \cup \{x_n\}$, such that the linear extension f_n to X_n satisfies $f_n(x) \leq p(x)$ for all $x \in X_n$. Then consider $f = \bigcup_{n=0}^{\infty} f_n$.)

The Hahn–Banach Theorem—especially in the form of Corollary (6.1.4)—has many interesting applications. We begin with some of the simpler ones.

(6.1.6) Proposition. *Let S be a closed subspace of a normed space X, and let $x_0 \in X \backslash S$. Then there exists a bounded linear functional f on X such that*

(i) $f(x_0) = 1$ *and*

(ii) $f(x) = 0$ *for all* $x \in S$.

Proof. Let X_0 be the subspace of X spanned by $S \cup \{x_0\}$, and define

$$f_0(x + \lambda x_0) = \lambda \quad (x \in S,\ \lambda \in \mathbf{F}).$$

(This is a good definition: for, as $x_0 \notin S$, the representation of a given element of X_0 in the form $x + \lambda x_0$, with $x \in S$ and $\lambda \in \mathbf{F}$, is unique.) Then f_0 is a linear functional on X_0, $f_0(x) = 0$ if $x \in S$, and $f_0(x_0) = 1$. Now, as S is closed, we see from Exercise (3.1.10: 3) that $\rho(x_0, S) > 0$. So for all x in S and all nonzero $\lambda \in \mathbf{F}$,

$$\|x + \lambda x_0\| = |\lambda| \left\|\lambda^{-1} x + x_0\right\| \geq |\lambda|\ \rho(x_0, S).$$

Hence

$$|f(x + \lambda x_0)| = |\lambda| \leq \rho(x_0, S)^{-1} \|x + \lambda x_0\|,$$

and therefore f_0 has the bound $\rho(x_0, S)^{-1}$. Applying Corollary (6.1.4) to f_0, we obtain the desired linear functional f on X. □

(6.1.7) Proposition. *If x_0 is a nonzero element of a normed space X, then there exists a bounded linear functional f on X such that $f(x_0) = \|x_0\|$ and $\|f\| = 1$.*

Proof. Let X_0 be the subspace of X generated by $\{x_0\}$, define a linear functional f_0 on X_0 by $f_0(\lambda x_0) = \lambda \|x_0\|$, and apply Corollary (6.1.4) to f_0. □

(6.1.8) Corollary. *For each x in a normed space X,*

$$\|x\| = \sup\{|f(x)| : f \in X^*,\ \|f\| = 1\}.$$

Proof. If $x = 0$, the conclusion is trivial. If $x \neq 0$, then for all $f \in X^*$ with $\|f\| = 1$ we have

$$|f(x)| \leq \|f\| \, \|x\| = \|x\| \, .$$

Since, by Proposition (6.1.7), there exists $f \in X^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$, the result follows. □

The remaining results and exercises in this section illustrate the interaction between a normed space X and its dual X^*, one of the most fascinating and beautiful aspects of modern analysis, in which the Hahn–Banach Theorem plays a fundamental part.

(6.1.9) Exercises

.1 Show that if X is a finite–dimensional Banach space, then X^* is finite–dimensional and $\dim(X^*) = \dim(X)$. (Reduce to the case where X is n–dimensional Euclidean space.)

.2 Let S be a closed subspace of a Banach space X, and define

$$S^\perp = \{f \in X^* : f(x) = 0 \text{ for all } x \in S\} \, .$$

Prove that $S^\perp$ is a closed linear subspace of X^*. Show that the following procedure yields a well–defined mapping T of S^* into $X^*/S^\perp$: given f in S^*, choose a norm–preserving extension F of f to X, and set Tf equal to the element of $X^*/S^\perp$ that contains F. Prove that T is a norm–preserving linear isomorphism of S^* onto $X^*/S^\perp$. Hence prove that for each $F \in X^*$,

$$\sup \{|F(x)| : x \in S, \, \|x\| \leq 1\} = \inf \left\{\|F - f\| : f \in S^\perp\right\} .$$

.3 Let S be a closed linear subspace of a Banach space X, let φ be the canonical map of X onto X/S, and for each f in $(X/S)^*$ define $Tf = f \circ \varphi$. Prove that T is an isometric linear isomorphism of $(X/S)^*$ onto $S^\perp$. Hence prove that for each $x \in X$,

$$\inf \{\|x - s\| : s \in S\} = \sup \left\{|f(x)| : f \in S^\perp, \, \|f\| \leq 1\right\} .$$

.4 Let $x_1, \ldots, x_n$ be elements of a real normed space X, and $c_1, \ldots, c_n$ real numbers. Prove the equivalence of the following conditions.

(i) There exists $f \in X^*$ with $\|f\| = 1$ and $f(x_i) = c_i$ for each i.

(ii) $|\lambda_1 c_1 + \cdots + \lambda_n c_n| \leq \|\lambda_1 x_1 + \cdots + \lambda_n x_n\| \, .$

.5 Let X be a normed space, and define

$$\widehat{x}(f) = f(x) \quad (x \in X,\ f \in X^*).$$

Prove that

(i) the mapping $x \mapsto \widehat{x}$ is a linear isometry of X into its second dual X^{**};

(ii) X is reflexive (see Exercise (5.3.2: 3)) if and only if this mapping has range X^{**};

(iii) if X is reflexive, then it is a Banach space.

.6 Prove that if X is an infinite–dimensional normed space, then X^* is infinite–dimensional. (cf. Exercise (6.1.9: 1). Suppose that X^* is finite–dimensional, and consider the mapping $x \mapsto \widehat{x}$ defined in the preceding exercise.)

.7 Let X be a Banach space. Prove that

(i) if X^* is separable, then so is X;

(ii) if X is separable and reflexive, then X^* is separable.

(For (i), let $\{f_1, f_2, \ldots\}$ be dense in X^*, and for each n choose a unit vector x_n such that $|f_n(x_n)| \geq \frac{1}{2}\|x_n\|$. Let Y be the closure of the subspace generated by $\{x_1, x_2, \ldots\}$, suppose that $Y \neq X$, and use Proposition (6.1.6) to deduce a contradiction.)

.8 Prove that a closed subspace Y of a reflexive Banach space X is reflexive. (For each $f \in X^*$ let f_Y denote the restriction of f to Y. Given $u \in Y^{**}$, choose $\xi \in X$ such that $u(f_Y) = f(\xi)$ for all $f \in X^*$. Then use Propositions (6.1.6) and (6.1.4).)

.9 We saw in Exercise (4.4.11: 5) that the space L_∞, introduced in Exercise (4.4.11: 2), can be identified with the dual space of $L_1 = L_1(\mathbf{R})$. In this exercise we show that L_1 can be identified with a subset of the dual of L_∞ but is not the whole of that dual.

Prove that for each $g \in L_1$,

$$u_g(f) = \int fg$$

defines an element of the dual of L_∞. Let X_0 be the set of all continuous functions $f : \mathbf{R} \to \mathbf{R}$ that vanish outside some compact set, and define a bounded linear functional u_0 on X_0 by $u_0(f) = f(0)$.

Using Corollary (6.1.4), construct a norm–preserving extension u of u_0 to L_∞. By considering $u(f_n)$, where

$$f_n(x) = \begin{cases} (1-|x|)^n & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1, \end{cases}$$

show that there is no element g of L_1 such that $u = u_g$.

It follows from this exercise that, in contrast to L_p for $1 < p < \infty$ (see Theorem (4.4.10)), L_1 is not reflexive.

The next three lemmas, together with our work on the Hahn–Banach Theorem, enable us to produce a substantial strengthening of the following consequence of Riesz's Lemma (4.3.5): in an infinite–dimensional normed space, if $0 < \theta < 1$, then there exists a sequence (x_n) of unit vectors such that $\|x_m - x_n\| > \theta$ whenever $m \neq n$.

(6.1.10) Lemma. *If $f, f_1, \ldots, f_n$ are linear functionals on a linear space X over $\mathbf{F}$ such that $\ker(f) \supset \bigcap_{i=1}^n \ker(f_i)$, then $f, f_1, \ldots, f_n$ are linearly dependent.*

Proof. We may assume that none of the functions under consideration is identically zero. We proceed by induction on n. In the case $n = 1$, choose $a \in X$ such that $f_1(a) = 1$. Then for each $x \in X$,

$$(x - f_1(x)a) \in \ker(f_1),$$

so

$$0 = f(x - f_1(x)a) = f(x) - f(a)f_1(x).$$

Hence $f = f(a)f_1$, and therefore f and f_1 are linearly dependent.

Now suppose that the lemma holds for $n = k$, and consider the case $n = k+1$. Let g be the restriction of f to $\ker(f_{k+1})$, and for $i = 1, \ldots, k$ let g_i be the restriction of f_i to $\ker(f_{k+1})$. Then $\ker(g) \supset \bigcap_{i=1}^k \ker(g_i)$, so $g = \sum_{i=1}^k \lambda_i g_i$ for some elements λ_i of $\mathbf{F}$, by our induction hypothesis. Thus $f - \sum_{i=1}^k \lambda_i f_i$ vanishes on $\ker(f_{k+1})$. By the case $n = 1$ that we have already proved, $f - \sum_{i=1}^k \lambda_i f_i$ and f_{k+1} are linearly dependent; so $f, f_1, \ldots, f_{k+1}$ are linearly dependent and the induction is complete. □

(6.1.11) Lemma. *Let X be an infinite–dimensional normed space, and $f_1, \ldots, f_n$ elements of X^*. Then $\bigcap_{i=1}^n \ker(f_i) \neq \{0\}$.*

Proof. First assume that the f_i are linearly independent. By Exercise (6.1.9: 6), there exists an element f of X^* such that $f, f_1, \ldots, f_n$ are linearly independent. Lemma (6.1.10) now shows that $\bigcap_{i=1}^n \ker(f_i)$ is not contained in $\ker(f)$, from which the desired conclusion follows immediately.

Now consider the case where the f_i are linearly dependent. Without loss of generality, we may assume that for some $m \leq n$, $\{f_1, \ldots, f_m\}$ is a basis for the linear space generated by all the f_i. By the first part of the proof, there exists a nonzero element ξ in $\bigcap_{i=1}^{m} \ker(f_i)$; clearly, $\xi \in \bigcap_{i=1}^{n} \ker(f_i)$. □

(6.1.12) Lemma. *Let X be an infinite–dimensional normed space, and $f_1, \ldots, f_n$ linearly independent elements of X^*. Then there exist nonzero elements ξ, η of X such that $f_i(\eta) < 0 = f_i(\xi)$ for each i.*

Proof. The existence of ξ follows from Lemma (6.1.11). On the other hand, Lemma (6.1.10) shows that for each i there exists $x_i \in X$ such that $f_i(x_i) = 1$ and $f_j(x_i) = 0$ when $j \neq i$. Setting $\eta = -\sum_{i=1}^{n} x_i$, we see that $f_i(\eta) = -1$ for each i. □

(6.1.13) Proposition. *If X is an infinite-dimensional normed space, then there exists a sequence (x_n) of unit vectors in X such that $\|x_m - x_n\| > 1$ whenever $m \neq n$.*

Proof. We construct the required vectors inductively as follows. Choosing a unit vector $x_1 \in X$, apply Proposition (6.1.7) to obtain $f_1 \in X^*$ such that $\|f_1\| = 1 = f_1(x_1)$. Now suppose that we have constructed unit vectors $x_1, \ldots, x_n$ in X, and linearly independent unit vectors $f_1, \ldots, f_n$ in X^*, such that $f_i(x_i) = 1 = \|f_i\|$ for each i. By Lemma (6.1.12), there exist nonzero elements ξ, η of X such that $f_i(\eta) < 0 = f_i(\xi)$ for each i. Choose $c > 0$ such that $\|\eta\| < \|\eta + c\xi\|$. Setting

$$x_{n+1} = \|\eta + c\xi\|^{-1} (\eta + c\xi),$$

note that $f_i(x_{n+1}) < 0$ for $1 \leq i \leq n$. Now use Proposition (6.1.7) to obtain an element f_{n+1} of X^* such that $\|f_{n+1}\| = 1 = f_{n+1}(x_{n+1})$. Suppose that $f_{n+1} = \sum_{i=1}^{n} \lambda_i f_i$ for some elements λ_i of $\mathbf{F}$. Then

$$\begin{aligned}
\|\eta + c\xi\| &= f_{n+1}(\eta + c\xi) \\
&= \left| \sum_{i=1}^{n} \lambda_i f_i(\eta + c\xi) \right| \\
&= \left| \sum_{i=1}^{n} \lambda_i f_i(\eta) \right| \\
&= |f_{n+1}(\eta)| \\
&\leq \|\eta\| \\
&< \|\eta + c\xi\|,
\end{aligned}$$

a contradiction. Hence the linear functionals $f_1, \ldots, f_n, f_{n+1}$ are linearly independent. Moreover, if $1 \leq i \leq n$, then

$$\|x_{n+1} - x_i\| \geq |f_i(x_{n+1} - x_i)| = |f_i(x_{n+1}) - f_i(x_i)| > 1,$$

since $f_i(x_i) = 1$ and $f_i(x_{n+1}) < 0$. This completes our inductive construction. □

A sequence (x_n) in a Banach space X is called a *Schauder basis* if for each $x \in X$ there exists a unique sequence (λ_n) in $\mathbf{F}$ such that $x = \sum_{n=1}^{\infty} \lambda_n x_n$. In that case, X is separable, and the mapping $x \mapsto (\lambda_n)_{n=1}^{\infty}$ can be used to identify X with a sequence space.

The notion of a Schauder basis generalises that of a basis in a finite–dimensional space. In the spaces c_0 and l_p $(1 \leq p < \infty)$ let e_n be the vector with nth term equal to 1 and all other terms 0; then $\{e_1, e_2, \ldots\}$ is a Schauder basis. Schauder bases for other separable Banach spaces, such as $\mathcal{C}[0,1]$, are not so easy to construct, and Enflo [15] has shown that there exist separable Banach subspaces of c_0 that do not have a Schauder basis. We can, however, prove the following theorem.

(6.1.14) Theorem. *Every infinite–dimensional Banach space contains an infinite–dimensional closed subspace with a Schauder basis.*

The next two lemmas make this possible.

(6.1.15) Lemma. *Let (x_n) be a total sequence of nonzero elements of a Banach space X, and c a positive number such that if $\lambda_1, \ldots, \lambda_n$ belong to* $\mathbf{F}$, *and $m < n$, then*

$$\left\| \sum_{i=1}^{m} \lambda_i x_i \right\| \leq c \left\| \sum_{i=1}^{n} \lambda_i x_i \right\|. \tag{3}$$

Then (x_i) is a Schauder basis for X.

Proof. Consider any sequence (λ_n) in $\mathbf{F}$ such that $\sum_{i=1}^{\infty} \lambda_i x_i$ converges in X. If $n > k$, then

$$\begin{aligned} |\lambda_k| &= \|x_k\|^{-1} \|\lambda_k x_k\| \\ &\leq c \|x_k\|^{-1} \left\| \sum_{i=k}^{n} \lambda_i x_i \right\|. \end{aligned}$$

Letting $n \to \infty$, we see that

$$|\lambda_k| \leq c \|x_k\|^{-1} \left\| \sum_{i=k}^{\infty} \lambda_i x_i \right\|.$$

A simple induction argument now enables us to prove that if $\sum_{i=1}^{\infty} \lambda_i x_i = 0$, then $\lambda_i = 0$ for each i. Thus a given element of X has at most one representation in the form $\sum_{i=1}^{\infty} \lambda_i x_i$ with each λ_i in $\mathbf{F}$. It remains to show that such a representation exists.

Let X_∞ be the subspace of X generated by $\{x_1, x_2, \ldots\}$, and for each n let X_n be the subspace generated by $\{x_1, \ldots, x_n\}$. Define a (clearly linear) mapping P_n of X_∞ onto X_n by

$$P_n\left(\sum_{i=1}^{\infty} \lambda_i x_i\right) = \sum_{i=1}^{n} \lambda_i x_i.$$

It follows from (3) that c is a bound for P_n on X_∞. But X_∞ is dense in X, so, by Exercise (4.2.2: 10), P_n extends to a bounded linear mapping P_n on X with bound c. By Corollary (6.1.4), the mapping $\sum_{i=1}^{\infty} \lambda_i x_i \mapsto \lambda_n$ extends to a bounded linear functional f_n on X such that

$$f_n(x)x_n = P_n(x) - P_{n-1}(x),$$

where, for convenience, we set $P_0(x) = 0$.

We prove that $x = \sum_{n=1}^{\infty} f_n(x)x_n$ for each $x \in X$. To this end, let $\varepsilon > 0$ and, using the fact that the sequence (x_n) is total, choose $\lambda_1, \ldots, \lambda_N$ in $\mathbf{F}$ such that

$$\left\| x - \sum_{n=1}^{N} \lambda_n x_n \right\| < \varepsilon.$$

For each $k \geq N$ we have

$$\begin{aligned}
\|x - P_k(x)\| &\leq \left\| x - \sum_{n=1}^{N} \lambda_n x_n \right\| + \left\| \sum_{n=1}^{N} \lambda_n x_n - P_k(x) \right\| \\
&< \varepsilon + \left\| P_k\left(\sum_{n=1}^{N} \lambda_n x_n - x\right) \right\| \\
&\leq \varepsilon + \|P_k\| \left\| \sum_{n=1}^{N} \lambda_n x_n - x \right\| \\
&\leq \varepsilon + \|P_k\|\, \varepsilon \\
&\leq (1 + c)\varepsilon.
\end{aligned}$$

Hence

$$x = \lim_{k\to\infty} P_k x = \lim_{k\to\infty} \sum_{n=1}^{k} f_n(x)x_n = \sum_{n=1}^{\infty} f_n(x)x_n. \quad \square$$

(6.1.16) S. Mazur's Lemma. *Let Y be a finite–dimensional subspace of an infinite–dimensional Banach space X. Then for each $\varepsilon > 0$ there exists a unit vector $\xi \in X$ such that*

$$\|y\| \leq (1 + \varepsilon)\, \|y + \lambda\xi\| \tag{4}$$

for all $y \in Y$ and $\lambda \in \mathbf{F}$.

Proof. Without loss of generality we may take $\varepsilon < 1$. Let $\{y_1, \ldots, y_n\}$ be an $\varepsilon/2$–approximation to the set

$$S = \{y \in Y : \|y\| = 1\}$$

(which is compact, by Exercise (4.3.7: 2)). Using Proposition (6.1.7), for $i = 1, \ldots, n$ construct $f_i \in X^*$ with norm 1 such that $f_i(y_i) = 1$. By Lemma (6.1.11), there exists a unit vector $\xi \in \bigcap_{i=1}^n \ker(f_i)$. Consider any vector $y \in Y$ and any $\lambda \in \mathbf{F}$. If $y = 0$, then (4) is trivial. If $y \neq 0$, then we may assume that $\|y\| = 1$: otherwise, we just consider $\|y\|^{-1} y$. Choosing i such that $\|y - y_i\| < \varepsilon/2$, we have

$$\begin{aligned} \|y + \lambda\xi\| &\geq \|y_i + \lambda\xi\| - \|y - y_i\| \\ &\geq f_i(y_i + \lambda\xi) - \tfrac{\varepsilon}{2} \\ &= 1 - \tfrac{\varepsilon}{2} \\ &> \frac{1}{1+\varepsilon}, \end{aligned}$$

since $\varepsilon < 1$. Hence (4) obtains. □

Proof of Theorem (6.1.14). Let X be an infinite–dimensional Banach space, and $\varepsilon > 0$. Choose positive numbers ε_n such that

$$\ln(1 + \varepsilon_n) < 2^{-n-2} \ln(1 + \varepsilon)$$

for each n. Then

$$\prod_{n=1}^{\infty} (1 + \varepsilon_n) = \lim_{N\to\infty} \prod_{n=1}^{N} (1 + \varepsilon_n) \leq \sqrt{1 + \varepsilon} < 1 + \varepsilon.$$

Let x_1 be a unit vector in X. By Mazur's Lemma, there exists a unit vector $x_2 \in X$ such that

$$\|y\| \leq (1 + \varepsilon_1) \|y + \lambda x_2\|$$

for all y in the subspace generated by x_1 and for all $\lambda \in \mathbf{F}$. By the same lemma, there exists a unit vector $x_3 \in X$ such that

$$\|y\| \leq (1 + \varepsilon_2) \|y + \lambda x_3\|$$

for all y in the subspace generated by $\{x_1, x_2\}$ and for all $\lambda \in \mathbf{F}$. Carrying on in this way, we construct an infinite sequence (x_n) of unit vectors in X such that

$$\|y\| \leq (1 + \varepsilon_n) \|y + \lambda x_{n+1}\|$$

for all y in the subspace generated by $\{x_1, \ldots, x_n\}$ and for all $\lambda \in \mathbf{F}$. It follows that if $\lambda_1, \ldots, \lambda_n \in \mathbf{F}$ and $m < n$, then

$$\left\| \sum_{i=1}^{m} \lambda_i x_i \right\| \leq (1 + \varepsilon_m) \left\| \sum_{i=1}^{m} \lambda_i x_i + \lambda_{m+1} x_{m+1} \right\|$$

$$
\begin{aligned}
&\leq (1+\varepsilon_m)(1+\varepsilon_{m+1}) \left\| \sum_{i=1}^{m+1} \lambda_i x_i + \lambda_{m+2} x_{m+2} \right\| \\
&\leq \cdots \\
&\leq (1+\varepsilon_m)(1+\varepsilon_{m+1}) \cdots (1+\varepsilon_{n-1}) \left\| \sum_{i=1}^{n} \lambda_i x_i \right\| \\
&\leq (1+\varepsilon) \left\| \sum_{i=1}^{n} \lambda_i x_i \right\| .
\end{aligned}
$$

Hence, by Lemma (6.1.15), $(x_n)_{n=1}^{\infty}$ is a Schauder basis of the closure of the subspace of X that it generates. □

For our last application of the Hahn–Banach Theorem in this section, we show that if I is a compact interval, then the dual space $\mathcal{C}(I)^*$ can be isometrically embedded in the Banach space $\left(\mathcal{BV}(I), \|\cdot\|_{\mathrm{bv}}\right)$ of functions of bounded variation on I (introduced in Exercise (4.5.2: 4)). To this end, for convenience we say that a bounded function $f : I \to \mathbf{R}$ is *representable* if there exists an increasing sequence (f_n) of elements of $\mathcal{C}(I)$ that converges simply to f. We denote by $\mathcal{R}(I)$ the subspace of $\mathcal{B}(I)$ consisting of all bounded real–valued functions on I that can be written as the difference of two representable functions. Note that $\mathcal{C}(I) \subset \mathcal{R}(I)$.

(6.1.17) Exercises

1. Prove that if J is a compact subinterval of I, then $-\chi_J$ is representable.

.2 Let $f \in \mathcal{C}(I)$, where $I = [a, b]$, let $P = (x_0, \ldots, x_n)$ be a partition of I, and for each k $(0 \leq k \leq n-1)$ let ξ_k be any point of $[x_k, x_{k+1}]$. Define $\psi \in \mathcal{B}(I)$ by

$$\psi = \sum_{k=0}^{n-1} f(\xi_k) \left(\chi_{[a, x_{k+1}]} - \chi_{[a, x_k]} \right) .$$

Show that $\|f - \psi\| \to 0$ as the mesh of P tends to 0.

(6.1.18) Theorem. *Let $I = [a, b]$ be a compact interval. Then for each real–valued function α of bounded variation on I,*

$$u_\alpha = \int_a^b f(x)\, \mathrm{d}\alpha(x)$$

defines a bounded linear functional, with norm $T_\alpha(a, b)$, on the Banach space $\mathcal{C}(I)$. Moreover, each bounded linear functional on $\mathcal{C}(I)$ is of the form

u_α, where α is a function of bounded variation on I that is unique up to an additive constant.

Proof. Throughout this proof, $P = (x_0, x_1, \ldots, x_n)$ is a partition of I, and for each i, ξ_i is any point of the interval $[x_i, x_{i+1}]$. Consider first a real–valued function α of bounded variation on I. The linearity of u_α follows from Exercise (1.5.16: 4). For each $f \in \mathcal{C}(I)$ we have the following inequality for Riemann–Stieltjes sums:

$$\begin{aligned} \left| \sum_{i=0}^{n-1} f(\xi_i)(\alpha(x_{i+1}) - \alpha(x_i)) \right| &\le \|f\| \sum_{i=0}^{n-1} |\alpha(x_{i+1}) - \alpha(x_i)| \\ &\le \|f\| \, T_\alpha(a,b). \end{aligned}$$

In the limit as the mesh of the partition tends to 0 we obtain the inequality

$$|u_\alpha(f)| \le \|f\| \, T_\alpha(a,b),$$

which shows that the linear functional u_α has bound $T_\alpha(a,b)$.

Now consider any bounded linear functional u on $\mathcal{C}(I)$. By Corollary (6.1.4), there exists a norm–preserving extension $u^\sharp$ of u to $\mathcal{R}(I)$. Referring to Exercise (6.1.17: 1), define a function $\alpha : I \to \mathbf{R}$ by

$$\alpha(x) = u^\sharp(\chi_{[a,x]}) \quad (x \in I)\,.$$

To show that α is of bounded variation on I, let P be as in the foregoing, and for each k $(0 \le k \le n-1)$ let

$$\sigma_k = \operatorname{sgn}\left(\alpha(x_{k+1}) - \alpha(x_k)\right).$$

Then

$$\phi = \sum_{k=0}^{n-1} \sigma_k \left(\chi_{[a,x_{k+1}]} - \chi_{[a,x_k]} \right) \in \mathcal{R}(I),$$

$\|\phi\| \le 1$, and

$$\sum_{k=0}^{n-1} |\alpha(x_{k+1}) - \alpha(x_k)| = u^\sharp(\phi) \le \left\| u^\sharp \right\| = \|u\|\,.$$

Hence α is of bounded variation on I, and

$$T_\alpha(a,b) \le \|u\|\,. \tag{5}$$

If, now, f is any element of $\mathcal{C}(I)$, consider the function

$$\psi = \sum_{k=0}^{n-1} f(\xi_k) \left(\chi_{[a,x_{k+1}]} - \chi_{[a,x_k]} \right),$$

which, again by Exercise (6.1.17: 1), belongs to $\mathcal{R}(I)$. We have

$$\left| u(f) - \sum_{k=0}^{n-1} f(\xi_k)\left(\alpha(x_{k+1}) - \alpha(x_k)\right) \right| = \left| u^\sharp(f) - u^\sharp(\psi) \right| \leq \|u\| \, \|f - \psi\| \, .$$

Letting the mesh of the partition P tend to 0, we see from Exercise (6.1.17: 2) that $\|f - \psi\| \to 0$; also,

$$\sum_{k=0}^{n-1} f(\xi_k)\left(\alpha(x_{k+1}) - \alpha(x_k)\right) \to \int_a^b f(x)\, \mathrm{d}\alpha(x).$$

Hence $u(f) = u_\alpha(f)$. Moreover, from (5) and the first part of the proof, $\|u\| = T_\alpha(a, b)$.

Finally, the uniqueness, up to an additive constant, of the function α corresponding to the given bounded linear functional u on $\mathcal{C}(I)$ follows from Proposition (1.5.19). □

The full power of the Hahn–Banach Theorem is not needed to prove Theorem (6.1.18): for, as is shown on pages 106–110 of [40], it is possible to construct an extension of u to $\mathcal{R}(I)$ by elementary means.

We say that a function $f : I \to \mathbf{R}$ of bounded variation on $I = [a, b]$ is *normalised* if $f(a) = 0$. It is easy to show that the normalised elements form a closed, and therefore complete, linear subspace of the Banach space $\left(\mathcal{BV}(I), \|\cdot\|_{\mathrm{bv}}\right)$.

(6.1.19) Corollary. *Under the hypotheses of* Theorem (6.1.18), $\mathcal{C}(I)^*$ *is isometrically isomorphic to the Banach space of normalised functions of bounded variation on* I.

(6.1.20) Exercises

.1 Let $I = [a, b]$ be a compact interval. Prove that the normalised elements of $\mathcal{BV}(I)$ form a Banach space relative to the norm $\|\cdot\|_{\mathrm{bv}}$. Then prove Corollary (6.1.19).

.2 Compute the norm of the bounded linear functional u defined on $\mathcal{C}[-1, 1]$ by

$$u(f) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} f(1/n).$$

.3 Let X be a compact metric space, and u a linear functional on $\mathcal{C}(X)$ that is *positive,* in the sense that $u(f) \geq 0$ for all nonnegative $f \in \mathcal{C}(X)$. Prove that u is bounded and has norm equal to $u(\mathbf{1})$, where $\mathbf{1}$ is the constant function $x \mapsto 1$ on X.

.4 Let u be a bounded linear functional on $\mathcal{C}(X)$, where X is a compact metric space. Prove that there exist positive linear functionals v, w on $\mathcal{C}(X)$ such that $u = v - w$. (For $f \geq 0$ in $\mathcal{C}(X)$ let $v(f) = \sup\{u(g) : g \in \mathcal{C}(X),\ 0 \leq g \leq f\}$.)

6.2 Separation Theorems

In this section we use the Hahn–Banach Theorem to establish a number of geometric results about the separation of convex sets by a hyperplane. These results have many applications, including some significant ones in mathematical economics (see Appendix C).

If A is a subset of a vector space, and $t \in \mathbf{F}$, we define

$$tA = \{tx : x \in A\}.$$

(6.2.1) Lemma. *Let X be a normed space, and A a convex subset of X containing 0 in its interior. Then the* Minkowski functional $p : X \to \mathbf{R}$, *defined by*

$$p(x) = \inf\{t > 0 : x \in tA\},$$

is a sublinear functional on X. If $p(x) < 1$, then $x \in A$; and if A is open, then

$$A = \{x \in X : p(x) < 1\}.$$

Proof. Choose $r > 0$ such that $B(0, r) \subset A$. If $x \neq 0$, then

$$x \in \frac{2\|x\|}{r} B(0, r) \subset \frac{2\|x\|}{r} A.$$

It follows that p is defined throughout X. Let α, β be positive numbers such that $x \in \alpha A$ and $y \in \beta A$; then

$$x + y = (\alpha + \beta)\left(\frac{\alpha}{\alpha+\beta}\alpha^{-1}x + \frac{\beta}{\alpha+\beta}\beta^{-1}y\right),$$

where, by convexity,

$$\frac{\alpha}{\alpha+\beta}\alpha^{-1}x + \frac{\beta}{\alpha+\beta}\beta^{-1}y \in A.$$

So $x + y \in (\alpha + \beta)A$. It now follows that $p(x + y) \leq p(x) + p(y)$. On the other hand, if $\lambda > 0$, then for all positive t we have

$$\lambda x \in tA \Leftrightarrow x \in \left(\lambda^{-1}t\right) A$$

and therefore

$$t \geq p(\lambda x) \Leftrightarrow \lambda^{-1}t \geq p(x),$$

so $p(\lambda x) = \lambda p(x)$. This last equation also holds when $\lambda = 0$, since $p(0) = 0$. Thus p is a sublinear functional on X.

If $p(x) < 1$, then there exists $t \in (0,1)$ such that $t^{-1}x \in A$; by the convexity of A, $x = (1-t)0 + t(t^{-1}x)$ belongs to A.

Finally, suppose that A is open, and consider any $x \in A$. Since $p(0) = 0$, to prove that $p(x) < 1$ we may assume that $x \neq 0$. Choose $s > 0$ such that $\overline{B}(x, s\|x\|) \subset A$; then $(1+s)x \in A$, so $p(x) \leq (1+s)^{-1} < 1$. □

(6.2.2) Lemma. *Let A be a nonempty open convex subset of a normed space X, and x_0 a point of $X \backslash A$. Then there exists a bounded real–linear functional f on X such that $f(x) < f(x_0)$ for all x in A.*

Proof. By translation, we may assume that $0 \in A$; so, by Lemma (6.2.1),

$$p(x) = \inf\left\{t > 0 : t^{-1}x \in A\right\}$$

defines a sublinear functional on X, and $p(x) < 1$ if and only if $x \in A$. Hence $p(x_0) \geq 1$. Let X_0 be the real linear subspace of X generated by $\{x_0\}$, and define a bounded real–linear functional f_0 on X_0 by

$$f_0(\lambda x_0) = \lambda \quad (\lambda \in \mathbf{R}).$$

If $\lambda \geq 0$, then

$$f_0(\lambda x_0) = \lambda \leq \lambda p(x_0) = p(\lambda x_0);$$

if $\lambda < 0$, then

$$f_0(\lambda x_0) = \lambda < 0 \leq p(\lambda x_0).$$

Thus $f_0(x) \leq p(x)$ for all $x \in X_0$. By the Hahn–Banach Theorem (6.1.3), there exists a real–linear functional f on X such that

- $f(x) = f_0(x)$ for all $x \in X_0$, and
- $f(x) \leq p(x)$ for all $x \in X$.

For all $x \in A$,

$$f(x) \leq p(x) < 1 = f(x_0).$$

It follows that the nonempty open set A is contained in the complement of the translated hyperplane $x_0 + \ker(f)$; whence, by Exercise (4.2.5: 3) and Lemma (4.1.4), the hyperplane $\ker(f)$ is closed in X. It follows from Proposition (4.2.3) that f is bounded. □

(6.2.3) Proposition. *Let C be a nonempty closed convex subset of a normed space X, and x_0 a point of $X \backslash C$. Then there exist a bounded real–linear functional f on X, and a real number α, such that $f(x) < \alpha < f(x_0)$ for all $x \in C$.*

Proof. Choose $r > 0$ such that $B(x_0, r) \cap C = \emptyset$. By Exercise (4.1.5: 6),

$$A = \{x + y : x \in C,\ y \in B(0, r)\}$$

is open and convex; also, $x_0 \notin A$. By Lemma (6.2.2), there exists a bounded real–linear functional f on X such that $f(x) < f(x_0)$ for all x in A. Since f is not identically 0, $f(b) > 0$ for some $b \in B(0, r)$. Taking $\alpha = f(x_0) - f(b)$, we see that for all $x \in C$,

$$f(x) = f(x + b) - f(b) < \alpha < f(x_0). \quad \square$$

(6.2.4) Exercises

.1 Let A be a compact convex subset of a real normed space X, and B a closed convex subset of X. Prove that there exist $f \in X^*$ and $\alpha, \beta \in \mathbf{R}$ such that $f(x) \leq \alpha < \beta \leq f(y)$ for all $x \in A$ and $y \in B$.

.2 Prove *Helly's Theorem*: let $\mathcal{F}$ be a finite family of convex subsets of $\mathbf{R}^n$ with the property that the intersection of any $n + 1$ sets in $\mathcal{F}$ is nonempty; then $\bigcap \mathcal{F}$ is nonempty. (First use induction on the number of sets in $\mathcal{F}$; then use induction on the dimension n.)

.3 Let K be a convex subset of a normed space X, and $S \subset K$. We say that S is an *extreme subset* of K if, for any distinct points x, y of K such that $\frac{1}{2}(x + y) \in S$, we have $x \in S$ and $y \in S$. If also S contains only one element, then that element is called an *extreme point* of K.

Prove that the intersection of any family of extreme subsets of K is either empty or an extreme subset.

Now suppose that K is also compact, and let $\mathcal{E}$ be the family of all extreme subsets of K, partially ordered by inclusion. Prove that $\mathcal{E}$ has a minimal element S_0. (Use the finite intersection property and Zorn's Lemma.) Then prove that S_0 consists of a single point. (Suppose that S_0 contains two distinct points ξ, η. Choose $f \in X^*$ such that $f(\xi) < f(\eta)$, and let $\alpha = \sup_{x \in K} f(x)$. Show that $S_1 = \{x \in S_0 : f(x) = \alpha\}$ is an extreme subset of K such that $S_0 \backslash S_1 \neq \emptyset$.)

Finally, prove that K has at least one extreme point.

.4 By the *convex hull* of a subset K of a normed space we mean the set of all elements of the form $\sum_{n=1}^{N} \lambda_i x_i$, where the x_i are elements of K and the λ_i are nonnegative real numbers such that $\sum_{i=1}^{N} \lambda_i = 1$. Prove the *Krein–Milman Theorem*: a compact convex subset of a normed space X is the closure of the convex hull of the set of its extreme points. (Let C be the convex hull of the set of extreme points of the compact convex set, let $x_0 \in X \backslash \overline{C}$, and apply Proposition (6.2.3).)

When $X = \mathbf{R}^n$, there is a weak extension of Proposition (6.2.3) to the case where C need not be closed.

(6.2.5) Proposition. *Let C be a nonempty convex subset of the Euclidean space $\mathbf{R}^n$, and x_0 a point of $\mathbf{R}^n \backslash C$. Then there exists a bounded real–linear functional f on $\mathbf{R}^n$ such that $f(x) \leq f(x_0)$ for all $x \in C$.*

Proof. Since $\overline{C}$ is nonempty, closed, and convex, Proposition (6.2.3) allows us to assume that $x_0 \in \overline{C} \backslash C$. Then, by Exercise (4.1.5: 7), each open ball with centre x_0 contains some point of the complement of $\overline{C}$. Choose a sequence (x_k) in $\mathbf{R}^n$ that converges to x_0, such that $x_k \notin \overline{C}$ for each k. By Proposition (6.2.3) and Theorem (5.3.1), for each k there exist $p_k \in \mathbf{R}^n$ and $\alpha_k \in \mathbf{R}$ such that

- $\langle x, p_k \rangle < \alpha_k$ for all $x \in \overline{C}$, and
- $\langle x_k, p_k \rangle = \alpha_k$.

Replacing p_k by $\|p_k\|^{-1} p_k$, we may assume that $\|p_k\| = 1$ for each k. Since the unit ball of $\mathbf{R}^n$ is compact (Theorem (4.3.6)), we may pass to a subsequence and assume that (p_k) converges to a limit p in $\mathbf{R}^n$; then $\|p\| = 1$. Also, as

$$|\alpha_k| \leq \|p_k\| \, \|x_k\| = \|x_k\|$$

and the sequence $(\|x_k\|)$, being convergent, is bounded, (α_k) is a bounded sequence in $\mathbf{R}$. Passing to another subsequence, we may further assume that (α_k) converges to a limit α in $\mathbf{R}$. By continuity, for all $x \in \overline{C}$ we have

$$\langle x, p \rangle \leq \alpha = \langle x_0, p \rangle \, .$$

It remains to take $f(x) = \langle x, p \rangle$. □

Let H be a hyperplane in the normed space X, and a an element of $X \backslash H$. By Propositions (4.2.4) and (6.1.1), for each $\alpha \in \mathbf{R}$ there exists a unique real–linear functional f on X such that

$$a + H = \{x \in X : f(x) = \alpha\} \, .$$

We say that the translated hyperplane $a + H$ *separates* the nonempty subsets A and B of X if $f(x) \leq \alpha$ for all $x \in A$, and $f(x) \geq \alpha$ for all $x \in B$.

(6.2.6) Minkowski's Separation Theorem. *Let A and B be disjoint nonempty convex subsets of $\mathbf{R}^n$. Then there exists a closed translated hyperplane that separates A and B.*

Proof. The nonempty set

$$C = B - A = \{x - y : x \in B, \, y \in A\}$$

is convex, and $0 \notin C$. By Proposition (6.2.5), there exists a bounded real–linear functional f on $\mathbf{R}^n$ such that $f(z) \geq f(0) = 0$ for all $z \in C$. Hence $f(x) \geq f(y)$ for all $x \in A$ and $y \in B$, and we need only apply Exercise (1.1.1: 21) to obtain the required real number α. The corresponding hyperplane $f^{-1}(\{\alpha\})$ then separates A and B. □

(6.2.7) Exercise

> Let A, B be disjoint nonempty convex subsets of a normed space X such that A is compact and B is closed. Prove that there exist a bounded real–linear functional f on X, and a real number α, such that $f(x) > \alpha$ for all $x \in A$, and $f(x) < \alpha$ for all $x \in B$. (Reduce to the case where $A = \{0\}$, note Exercise (4.1.5: 6), and apply Proposition (6.2.3).)

6.3 Baire's Theorem and Beyond

In this section we prove one of the most useful theorems about complete metric spaces, Baire's Theorem, and then study several of its many interesting consequences. Among these are the existence of uncountably many continuous, nowhere differentiable functions on $[0, 1]$, and the Open Mapping Theorem for bounded linear mappings between Banach spaces.

(6.3.1) Baire's Theorem. *The intersection of a sequence of dense open sets in a complete metric space is dense.*

Proof. Let X be a complete metric space, (U_n) a sequence of dense open subsets of X, and

$$U = \bigcap_{n=1}^{\infty} U_n.$$

We need only prove that for $x_0 \in X$ and $r_0 > 0$, the set $U \cap \overline{B}(x_0, r_0)$ is nonempty. To this end, since U_1 is dense in X, we can find x_1 in $U_1 \cap B(x_0, r_0)$. Moreover, since both U_1 and $B(x_0, r_0)$ are open, so is their intersection; whence there exists r_1 such that $0 < r_1 < 1$ and

$$\overline{B}(x_1, r_1) \subset U_1 \cap B(x_0, r_0).$$

Since U_2 is dense in X, we can now find x_2 in $U_2 \cap B(x_1, r_1)$; but $U_2 \cap B(x_1, r_1)$ is open, so there exists r_2 such that $0 < r_2 < 1/2$ and

$$\overline{B}(x_2, r_2) \subset U_2 \cap B(x_1, r_1).$$

Carrying on in this way, we construct a sequence (x_n) of points of X, and a sequence (r_n) of positive numbers, such that for each $n \geq 1$, $0 < r_n < 1/n$ and

$$\overline{B}(x_n, r_n) \subset U_n \cap B(x_{n-1}, r_{n-1}).$$

By induction, if $m \geq n$, then $x_m \in B(x_n, r_n)$; whence

$$\rho(x_m, x_n) < r_n < \tfrac{1}{n} \quad (m \geq n). \tag{1}$$

Thus (x_n) is a Cauchy sequence in X. Since X is complete, (x_n) converges to a limit x_∞ in X. Letting m tend to ∞ in inequality (1), we have $\rho(x_\infty, x_n) \leq r_n$, and therefore $x_\infty \in \overline{B}(x_n, r_n)$, for each n. Taking $n = 0$, we see that $x_\infty \in \overline{B}(x_0, r_0)$; taking $n \geq 1$, we see that $x_\infty \in U_n$. □

The alternative name *Baire Category Theorem* for Theorem (6.3.1) originates from the following definitions (due to Baire).

A subset S of a metric space X is said to be

- *nowhere dense* in X if the interior of $\overline{S}$ is empty;
- *of the first category* if it is a countable union of nowhere dense subsets;
- *of the second category* if it is not of the first category.

Baire's Theorem is equivalent to the statement *a nonempty complete metric space is of the second category.*

(6.3.2) Exercises

.1 Prove the last statement; more precisely, prove that if a nonempty complete metric space is the union of a sequence of closed sets, then at least one of those closed sets has a nonempty interior.

.2 Prove the extended version of Cantor's theorem on the uncountability of $\mathbf{R}$ (Exercise (1.2.11: 4)): if (x_n) is a sequence of real numbers, then $\{x \in \mathbf{R} : \forall n\, (x \neq x_n)\}$ is dense in $\mathbf{R}$.

.3 Prove that a nonempty complete metric space without isolated points is uncountable.

We now show how Baire's Theorem can be used to prove the existence of continuous functions on $I = [0, 1]$ that are nowhere differentiable on I.

For each positive integer n let E_n be the set of all $f \in \mathcal{C}(I)$ with the property:

> there exists $t \in \left[0, 1 - n^{-1}\right]$ such that $|f(t+h) - f(t)| \leq nh$ whenever $0 < h < 1 - t$.

Note that $\bigcup_{n=1}^{\infty} E_n$ contains any $f \in \mathcal{C}(I)$ such that for some $t \in [0, 1)$ the right–hand derivative of f at t,

$$f'(t^+) = \lim_{h \to 0, h > 0} \frac{f(t+h) - f(t)}{h},$$

exists. To see this, consider such f and t. Choose a positive integer n_1 such that $t \in \left[0, 1 - n_1^{-1}\right]$ and $|f'(t^+)| < n_1$. Next choose $h_0 > 0$ such that if $0 < h < h_0$, then $|f(t+h) - f(t)| \leq n_1 h$. If $h_0 = 1 - t$, set $n = n_1$. If $h_0 < 1 - t$, then for $h_0 \leq h < 1 - t$ we have

$$|f(t+h) - f(t)| \leq \frac{2\|f\|}{h_0} h,$$

where $\|\cdot\|$ denotes the sup norm on $\mathcal{C}(I)$; so, taking $n = \max\{n_1, n_2\}$, where the positive integer $n_2 > 2\|f\|/h_0$, we have $f \in E_n$.

We prove that $\mathcal{C}(I) \backslash E_n$ is dense and open in I. To this end, first let $(f_k)_{k=1}^{\infty}$ be a sequence in E_n that converges to a limit f in $\mathcal{C}(I)$. Then there exists a sequence (t_k) in $\left[0, 1 - n^{-1}\right]$ such that

$$|f_k(t_k + h) - f_k(t_k)| \leq nh$$

whenever $k \geq 1$ and $0 < h < 1 - t_k$. Since $\left[0, 1 - n^{-1}\right]$ is sequentially compact, we may assume without loss of generality that (t_k) converges to a limit $t \in \left[0, 1 - n^{-1}\right]$. If $0 < h < 1 - t$, then for all sufficiently large k we have $0 < h < 1 - t_k$ and therefore

$$\begin{aligned}
|f(t+h) - f(t)| &\leq |f(t+h) - f(t_k+h)| + |f(t_k+h) - f_k(t_k+h)| \\
&\quad + |f_k(t_k+h) - f_k(t_k)| + |f_k(t_k) - f(t_k)| \\
&\quad + |f(t_k) - f(t)| \\
&\leq |f(t+h) - f(t_k+h)| + \|f - f_k\| + nh \\
&\quad + \|f - f_k\| + |f(t_k) - f(t)|.
\end{aligned}$$

Letting $k \to \infty$ and using the continuity of f, we obtain

$$|f(t+h) - f(t)| \leq nh.$$

Hence $f \in E_n$, and therefore E_n is closed in $\mathcal{C}(I)$. Thus $\mathcal{C}(I) \backslash E_n$ is open in $\mathcal{C}(I)$.

Given $f \in \mathcal{C}(I)$ and $\varepsilon > 0$, we now use the Weierstrass Approximation Theorem (4.6.1) to construct a polynomial function p such that $\|f - p\| < \varepsilon/2$. Choosing a positive integer

$$N > \varepsilon^{-1}(n + \|p'\|),$$

define a continuous function $q : [0,1] \to \mathbf{R}$ such that for $0 \leq k \leq N - 1$,

$$q\left(\frac{k}{N}\right) = 0,$$

$$q\left(\frac{k + \frac{1}{2}}{N}\right) = \varepsilon/2,$$

and q is linear on each of the intervals

$$\left[\frac{k}{N}, \frac{k+\frac{1}{2}}{N}\right], \left[\frac{k+\frac{1}{2}}{N}, \frac{k+1}{N}\right].$$

Let $g = p + q \in \mathcal{C}(I)$. For each $t \in [0,1)$ we have

$$\left|g'(t^+)\right| \geq \left|q'(t^+)\right| - \left|p'(t)\right| \geq N\varepsilon - \|p'\| > n,$$

so $g \notin E_n$. Since

$$\|f - g\| \leq \|f - p\| + \|q\| < \varepsilon,$$

we conclude that $\mathcal{C}(I)\backslash E_n$ is dense in $\mathcal{C}(I)$.

Now let F_n be the set of all $f \in \mathcal{C}(I)$ with the property:

> there exists $t \in \left[n^{-1}, 1\right]$ such that $|f(t+h) - f(t)| \leq nh$ whenever $0 < h < t$.

Arguments similar to those just used show that $\mathcal{C}(I)\backslash F_n$ is dense and open in $\mathcal{C}(I)$, and that it contains any $f \in \mathcal{C}(I)$ such that for some $t \in (0,1]$ the left–hand derivative of f at t,

$$f'(t^-) = \lim_{h \to 0, h<0} \frac{f(t+h) - f(t)}{h},$$

exists. Let

$$S = \bigcup_{n=1}^{\infty} E_n \cup \bigcup_{n=1}^{\infty} F_n.$$

Since $\mathcal{C}(I)$ is complete (Proposition (4.5.4)), we see from Baire's Theorem that

$$\mathcal{C}(I)\backslash S = \bigcap_{n=1}^{\infty} (\mathcal{C}(I)\backslash E_n) \cap \bigcap_{n=1}^{\infty} (\mathcal{C}(I)\backslash F_n)$$

is dense in $\mathcal{C}(I)$. Clearly, $\mathcal{C}(I)\backslash S$ consists of continuous, nowhere differentiable functions on I.

(6.3.3) Exercises

.1 Prove that $[0,1]$ cannot be written as the union of a sequence of pairwise–disjoint closed sets. (Suppose that there exists a sequence (F_n) of pairwise–disjoint closed sets whose union is $[0,1]$. Show that the union of the boundaries of the sets F_n is closed and has an empty interior.)

.2 Let X be a Banach space, and C a closed convex subset of X that is *absorbing*—that is, for each $x \in X$ there exists $t > 0$ such that $tx \in C$. Prove that 0 does not belong to the closure of $X\backslash C$. (Suppose the contrary, and show that for each positive integer n the complement of nC is dense and open in X.)

.3 Prove that if a Banach space is generated by a compact set, then it is finite–dimensional.

.4 Let X be a complete metric space, and $(f_i)_{i\in I}$ a family of continuous mappings of X into $\mathbf{R}$. Suppose that for each $x \in X$ there exists $M_x > 0$ such that $|f_i(x)| \leq M_x$ for all $i \in I$. Prove that there exist a nonempty open set $E \subset X$ and a positive integer N such that $|f_i(x)| \leq N$ for all $i \in I$ and all $x \in E$. (Let

$$\begin{aligned} C_{n,i} &= \{x \in X : |f_i(x)| \leq n\}, \\ C_n &= \textstyle\bigcap_{i\in I} C_{n,i}, \end{aligned}$$

and use Baire's Theorem.)

A mapping f between metric spaces X and Y is called an *open mapping* if $f(S)$ is an open subset of Y whenever S is an open subset of X.

(6.3.4) Exercise

Prove that a linear mapping T between normed spaces X, Y is open if and only if there exists $r > 0$ such that $B(0,r) \subset T(\overline{B}(0,1))$.

We now aim to apply Baire's Theorem to prove the following fundamental result on linear mappings between Banach spaces.

(6.3.5) The Open Mapping Theorem. *A bounded linear mapping of a Banach space onto a Banach space is open.*

The next lemma prepares us for the proof of this theorem.

(6.3.6) Lemma. *Let T be a linear mapping of a Banach space X into a normed space Y. Then T is open if and only if there exists $r > 0$ such that $B(0,r) \subset \overline{T\left(\overline{B}(0,1)\right)}$.*

Proof. Suppose that such a real number r exists. In view of the preceding exercise, it suffices to prove that $B(0,r/2) \subset T(\overline{B}(0,1))$. Given $y \in B(0,r/2)$, since $2y \in B(0,r)$, we can find an element x_1 of the unit ball of X such that

$$\|2y - Tx_1\| < \tfrac{r}{2}.$$

So $2^2y - 2Tx_1 \in B(0,r)$, and therefore there exists x_2 in the unit ball of X such that

$$\left\|2^2y - 2Tx_1 - Tx_2\right\| < \tfrac{r}{2}.$$

Carrying on in this way, we construct a sequence (x_n) of elements of the unit ball of X such that

$$\left\|2^N y - 2^{N-1}Tx_1 - 2^{N-2}Tx_2 - \cdots - Tx_N\right\| < \tfrac{r}{2}$$

for each N. Thus

$$\left\| y - \sum_{n=1}^{N} 2^{-n} T x_n \right\| < 2^{-N-1} r,$$

and therefore the series $\sum_{n=1}^{\infty} 2^{-n} T x_n$ converges to y. Since

$$\sum_{n=j}^{k} 2^{-n} \left\| x_n \right\| \leq \sum_{n=j}^{k} 2^{-n}$$

whenever $k > j$, we see from Exercise (4.1.8: 2) that $\sum_{n=1}^{\infty} 2^{-n} x_n$ converges to an element x in the unit ball of X. The boundedness of T now ensures that

$$Tx = \sum_{n=1}^{\infty} 2^{-n} T x_n = y.$$

Hence $B(0, r/2) \subset T(\overline{B}(0,1))$, and therefore T is open.

The converse is trivial. □

Proof of the Open Mapping Theorem. Let T be a bounded linear mapping of a Banach space X onto a Banach space Y. Then

$$Y = T(X) = \bigcup_{n=1}^{\infty} \overline{T(\overline{B}\,(0,n))},$$

where each of the sets $\overline{T\left(\overline{B}\,(0,n)\right)}$ is closed in Y. By Exercise (6.3.2: 1), there exists a positive integer N such that $\overline{T(\overline{B}\,(0,N))}$ has a nonempty interior; so there exist $y_1 \in Y$ and $R > 0$ such that

$$B(y_1, R) \subset \overline{T(\overline{B}\,(0,N))}.$$

Setting $z = N^{-1} y_1$ and $r = N^{-1} R$, we easily see that

$$B(z, r) \subset \overline{T(\overline{B}\,(0,1))}.$$

So if $y \in Y$ and $\|y\| < r$, then

$$z \pm y \in \overline{T(\overline{B}\,(0,1))}$$

and therefore

$$y = \tfrac{1}{2}\left((z+y) - (z-y)\right) \in \overline{T(\overline{B}\,(0,1))}.$$

Hence

$$B(0, r) \subset \overline{T(\overline{B}\,(0,1))},$$

and therefore, by Lemma (6.3.6), T is open. □

The Open Mapping Theorem is one of a number of closely interrelated results.

(6.3.7) Banach's Inverse Mapping Theorem. *A one–one bounded linear mapping of a Banach space onto a Banach space has a bounded linear inverse.*

Proof. Let T be a one–one bounded linear mapping of a Banach space X onto a Banach space Y. It is routine to prove that T^{-1} is a linear mapping from Y onto X. By the Open Mapping Theorem (6.3.5), if U is an open subset of X, then

$$(T^{-1})^{-1}(U) = T(U)$$

is open in Y; so T^{-1} is continuous, by Proposition (3.2.2), and is therefore a bounded linear mapping, by Proposition (4.2.1). □

By the *graph* of a mapping $f : X \to Y$ we mean the subset

$$\mathcal{G}(f) = \{(x, f(x)) : x \in X\}$$

of $X \times Y$. (The graph of f is really the same as the function f itself, regarded as a set of ordered pairs.)

(6.3.8) The Closed Graph Theorem. *A linear mapping of a Banach space X into a Banach space Y is bounded if and only if its graph is a closed subset of $X \times Y$.*

Proof. Let T be a linear mapping of X into Y. It is a simple exercise to show that if T is bounded, then its graph is a closed subset of $X \times Y$.

Suppose, conversely, that $\mathcal{G}(T)$ is closed in $X \times Y$. Since $X \times Y$, a product of complete metric spaces, is complete (by Proposition (3.5.10)), we see from Proposition (3.2.9) that $\mathcal{G}(T)$, which is clearly a subspace of $X \times Y$, is a Banach space. Define a mapping H of $\mathcal{G}(T)$ onto X by

$$H(x, Tx) = x \quad (x \in X).$$

It is straightforward to show that H is one–one and linear. Also,

$$\begin{aligned} \|H(x, Tx)\| &\leq \|x\| + \|Tx\| \\ &\leq 2 \max\{\|x\|, \|Tx\|\} \\ &= 2\|(x, Tx)\|, \end{aligned}$$

so H is bounded. It follows from Banach's Inverse Mapping Theorem (6.3.7) that H^{-1} is a bounded linear mapping of X onto $\mathcal{G}(T)$; but

$$\|Tx\| \leq \|(x, Tx)\| = \left\|H^{-1}x\right\| \leq \left\|H^{-1}\right\| \|x\|$$

for all $x \in X$, and so T is bounded. □

We met the following result—the *Uniform Boundedness Theorem*—in Exercise (4.2.2: 14), where you were asked to fill in the details of a relatively little known elementary proof. We now place the Uniform Boundedness Theorem in its normal context, with its standard proof.

(6.3.9) Theorem. *Let $(T_i)_{i \in I}$ be a family of bounded linear mappings from a Banach space X into a Banach space Y, such that $\{\|T_i x\| : i \in I\}$ is bounded for each $x \in X$. Then $\{\|T_i\| : i \in I\}$ is bounded.*

Proof. Our hypotheses ensure that for each $x \in X$,

$$u_x(i) = T_i x \quad (i \in I)$$

defines an element u_x of $\mathcal{B}(I, Y)$. Clearly, the mapping $x \mapsto u_x$ of X into $\mathcal{B}(I, Y)$ is linear. We prove that its graph is closed in $X \times \mathcal{B}(I, Y)$. Indeed, if (x_n) is a sequence converging to a limit x_∞ in X, such that the sequence (u_{x_n}) converges to a limit f in $\mathcal{B}(I, Y)$, then for each $i \in I$ we have

$$\begin{aligned} \|f(i) - u_{x_\infty}(i)\| &\leq \|f(i) - u_{x_n}(i)\| + \|u_{x_n}(i) - u_{x_\infty}(i)\| \\ &\leq \|f - u_{x_n}\| + \|T_i(x_n - x_\infty)\| \\ &\leq \|f - u_{x_n}\| + \|T_i\| \ \|x_n - x_\infty\| \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Hence $f = u_{x_\infty}$, and so the linear mapping $x \mapsto u_x$ has a closed graph. By Proposition (4.5.1) and the Closed Graph Theorem (6.3.8), this mapping is bounded. Let

$$c = \sup \{\|u_x\| : x \in X, \ \|x\| \leq 1\}.$$

Then for all $i \in I$ and all x in the unit ball of X,

$$\|T_i x\| = \|u_x(i)\| \leq \|u_x\| \leq c. \quad \square$$

(6.3.10) Exercises

.1 Prove that if T is an open bounded linear mapping of a Banach space X into a normed space Y, then the range of T is complete.

.2 Let X be a separable real Banach space with a Schauder basis (x_n), and let S be the linear space consisting of all sequences $(\lambda_n)_{n=1}^\infty$ of real numbers such that the series $\sum_{n=1}^\infty \lambda_n x_n$ converges in X. Show that

$$\|(\lambda_n)_{n=1}^\infty\| = \sup_{N \geq 1} \left\| \sum_{n=1}^N \lambda_n x_n \right\|$$

defines a norm on S. Prove that S is a Banach space with respect to this norm. Then show that the mapping

$$(\lambda_n)_{n=1}^{\infty} \mapsto \sum_{n=1}^{\infty} \lambda_n x_n$$

is a bounded linear isomorphism of S onto X with a continuous inverse. Deduce that for each positive integer N the *coordinate functional* $\sum_{n=1}^{\infty} \lambda_n x_n \mapsto \lambda_N$ belongs to the dual space X^*.

.3 Use an argument like that of Exercise (6.3.3: 4) to give another proof of the Uniform Boundedness Theorem.

.4 Let X, Y be Banach spaces, and suppose that for all distinct y, y' in Y there exists a bounded linear functional f on Y such that $f(y) \neq f(y')$. Let $T : X \to Y$ be a linear mapping such that if (x_n) is a sequence in X converging to 0, then $(f \circ T)(x_n)$ converges to 0 for each bounded linear functional f on Y. Prove that T is bounded. (Use the Closed Graph Theorem.)

.5 Let (T_n) be a sequence of bounded linear mappings of a Banach space X into a Banach space Y, such that the sequence $(T_n x)$ converges in Y for each $x \in X$. Prove that

$$Tx = \lim_{n\to\infty} T_n x$$

defines a bounded linear mapping $T : X \to Y$. (Use the Uniform Boundedness Theorem.)

.6 Let S, T be mappings of a Hilbert space H into itself such that $\langle Sx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in H$. Prove that S and T are linear mappings. Then give two proofs that both S and T are bounded. (For one proof use the Closed Graph Theorem.)

.7 Let A, B be disjoint subspaces of a Banach space X such that each element x of X can be written uniquely in the form

$$x = P_A x + P_B x$$

with $P_A x \in A$ and $P_B x \in B$. Prove that the *oblique projection mappings* $P_A : X \to A$ and $P_B : X \to B$ so defined are linear, and that they are bounded if and only if A and B are closed in X.

.8 Prove *Landau's Theorem*: if (a_n) is a sequence of complex numbers such that $\sum_{n=1}^{\infty} a_n x_n$ converges for each $(x_n) \in l_2(\mathbf{C})$, then $(a_n) \in l_2(\mathbf{C})$. (For each $x = (x_n) \in l_2(\mathbf{C})$ and each k, define $s_k(x) = \sum_{n=1}^{k} a_n x_n$. Apply the Uniform Boundedness Theorem to the sequence $(s_k)_{k=1}^{\infty}$ of linear functionals on $l_2(\mathbf{C})$, to show that the partial sums of $\sum_{n=1}^{\infty} |a_n|^2$ are bounded.)

.9 Let w be a weight function on the compact interval $I = [a, b]$. For each positive integer n let $(x_{n,0}, x_{n,1}, \ldots, x_{n,n})$ be a partition of I, and define a linear functional L_n on $\mathcal{C}(I)$ by

$$L_n f = \sum_{k=0}^{n} c_{n,k} f(x_{n,k}),$$

where each $c_{n,k} \in \mathbf{R}$. Prove *Polya's Theorem on approximate quadrature*: in order that $\lim_{n\to\infty} L_n f = \int_a^b f(x)w(x)\,\mathrm{d}x$ for all $f \in \mathcal{C}(I)$, it is necessary and sufficient that $\sup_{n\geq 1} \sum_{k=0}^{n} |c_{n,k}| < \infty$.

.10 Let T be an operator on a Hilbert space. Prove that $\operatorname{ran}(T)$ is closed if and only if $\operatorname{ran}(T^*)$ is closed. (Suppose that $\operatorname{ran}(T)$ is closed, note Exercise (5.3.3: 6), and show that $\operatorname{ran}(T^*T)$ is closed. To do so, let $T^*Tx_n \to \xi$, and use the Uniform Boundedness Theorem to show that the linear functional $Tx \mapsto \langle x, \xi \rangle$ is bounded on the Hilbert space $\operatorname{ran}(T)$.)

Perhaps the standard illustration of the Uniform Boundedness Theorem in action is the proof that there exists a 2π–periodic continuous function $f : \mathbf{R} \to \mathbf{C}$ whose Fourier series does not converge at 0.

Let $\mathcal{S}$ denote the subspace of $\mathcal{C}^\infty(\mathbf{R}, \mathbf{C})$ consisting of all 2π–periodic continuous mappings of $\mathbf{R}$ into $\mathbf{C}$. Recall that the *Fourier series,* or *Fourier expansion,* of $f \in \mathcal{S}$ at x is defined to be

$$s(f, x) = \sum_{n=-\infty}^{\infty} \widehat{f}(n)\mathrm{e}^{\mathrm{i}nx},$$

where the *Fourier coefficients* are given by

$$\widehat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\mathrm{e}^{-\mathrm{i}nt}\,\mathrm{d}t \quad (n \in \mathbf{Z}).$$

For each positive integer N let

$$s_N(f, x) = \sum_{n=-N}^{N} \widehat{f}(n)\mathrm{e}^{\mathrm{i}nx}.$$

Then

$$s_N(f, x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_N(x-t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(-t)D_N(t)\,\mathrm{d}t,$$

where the *Dirichlet kernel* D_N is defined by

$$D_N(t) = \sum_{n=-N}^{N} \mathrm{e}^{\mathrm{i}nt}.$$

Define a linear mapping $u_N : \mathcal{S} \to \mathbf{R}$ by

$$u_N(f) = s_N(f, 0).$$

Then

$$|u_N(f)| \leq \frac{1}{2\pi} \|f\| \int_{-\pi}^{\pi} |D_N(t)| \, \mathrm{d}t,$$

where $\|f\|$ is the sup norm of f. Thus u_N is bounded, and

$$\|u_N\| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |D_N(t)| \, \mathrm{d}t. \tag{2}$$

On the other hand, there exists a sequence (f_n) of elements of $\mathcal{S}$ such that

- $-1 \leq f_n \leq 1$ for each n, and
- $f_n(t) \to \operatorname{sgn}(D_N(t))$ for each $t \in \mathbf{R}$;

see Exercise (6.3.11: 2). Using Lebesgue's Dominated Convergence Theorem (2.2.14), we now obtain

$$u_N(f_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_n(-t) \, D_N(t) \, \mathrm{d}t \to \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{sgn}(D_N(-t)) \, D_N(t) \, \mathrm{d}t$$

as $n \to \infty$. But

$$D_N(t) = \frac{\sin\left(N + \frac{1}{2}\right) t}{\sin\left(\frac{t}{2}\right)} = D_N(-t),$$

so

$$\lim_{n \to \infty} u_N(f_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |D_N(t)| \, \mathrm{d}t,$$

and therefore, in view of (2),

$$\|u_N\| = \frac{1}{2\pi} \int_{-\pi}^{\pi} |D_N(t)| \, \mathrm{d}t.$$

Next, noting that $|\sin(t/2)| \leq t/2$ for all $t > 0$, we have

$$\begin{aligned}
\|u_N\| &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |D_N(t)| \, \mathrm{d}t \\
&= \frac{1}{\pi} \int_0^{\pi} \left| \frac{\sin\left(N + \frac{1}{2}\right) t}{\sin\left(\frac{t}{2}\right)} \right| \mathrm{d}t \\
&\geq \frac{2}{\pi} \int_0^{\pi} \frac{1}{t} \left|\sin\left(N + \tfrac{1}{2}\right) t\right| \, \mathrm{d}t \\
&= \frac{2}{\pi} \int_0^{\left(N + \frac{1}{2}\right)\pi} \frac{1}{t} |\sin t| \, \mathrm{d}t
\end{aligned}$$

$$
\begin{aligned}
&> \frac{2}{\pi}\sum_{n=1}^{N}\frac{1}{n\pi}\int_{(n-1)\pi}^{n\pi}|\sin t|\ \mathrm{d}t \\
&= \frac{4}{\pi^2}\sum_{n=1}^{N}\frac{1}{n}
\end{aligned}
$$

and therefore $\|u_N\| \to \infty$ as $N \to \infty$. By the Uniform Boundedness Theorem (6.3.9), there exists $f \in \mathcal{S}$ such that the set $\{|u_N(f)| : N \geq 1\}$ is unbounded. Hence the Fourier series of f cannot converge at 0.

(6.3.11) Exercises

.1 Prove that

$$\sum_{n=-N}^{N} \mathrm{e}^{\mathrm{i}nt} = \frac{\sin\left(N+\frac{1}{2}\right)t}{\sin\left(\frac{t}{2}\right)}$$

for each natural number N.

.2 Prove that, in the notation of the preceding paragraphs, there exists a sequence (f_n) of elements of $\mathcal{S}$ such that $-1 \leq f_n \leq 1$ for each n, and such that $f_n(t) \to \operatorname{sgn}(D_N(t))$ for each $t \in \mathbf{R}$.

.3 In view of the Riemann–Lebesgue Lemma (Exercise (2.3.3: 13)),

$$Tf = \left(\widehat{f}(n)\right)_{n=1}^{\infty}$$

defines a mapping $T : L_1[-\pi,\pi] \to c_0$. Prove that T is one–one but not onto c_0. (Show that there exists $\alpha > 0$ such that $\|T(D_n)\| \geq \alpha\,\|D_n\|_1$ for each n.)

Appendix A
What Is a Real Number?

In this appendix we sketch Bishop's adaptation of Cauchy's construction of the set **R,** based on the idea that a real number is an object that can be approximated arbitrarily closely by rational numbers.

Passing over the standard construction of the set **Z** of integers, we define a *rational number* to be an ordered pair (m, n) of integers, usually written m/n or $\frac{m}{n}$, such that $n \neq 0$. Two rational numbers m/n and m'/n' are said to be *equal*, and we write $m/n = m'/n'$, if mn' and $m'n$ are equal integers; this relation of equality is an equivalence relation.

We should really define a rational number to be an equivalence class of ordered pairs of integers relative to the equivalence relation of equality that we have just introduced. In that case two rational numbers (equivalence classes) would be equal if and only if they were one and the same. However, it more closely reflects common practice if we follow the approach in which rational numbers are given by the integer pairs themselves, and equality of rational numbers is a defined notion (given by a certain condition on the integer pairs) rather than the logical one of identity.[1] We follow a similar approach to the equality of real numbers in due course. In every case we use without further mention the standard symbol = to denote equality.

[1]For example, from childhood we are led to consider the rational numbers $\frac{1}{2}$, $\frac{2}{4}$, and $\frac{3}{6}$ as equal, not as representatives of some equivalence class. For another example, we consider the numbers $1, 0 \cdot 999 \cdots$, and $\frac{5}{5}$ to be equal although they are not logically identical (they are presented to us in different ways).

We omit the details of the familiar algebraic operations and the order relations $>, \geq$ on the set $\mathbf{Q}$ of rational numbers. We identify the integer n with the rational number $n/1$.

By a *real number* we mean a sequence $x = (x_n)_{n=1}^{\infty}$ of rational numbers that is *regular* in the sense that

$$|x_m - x_n| \leq \tfrac{1}{m} + \tfrac{1}{n} \quad (m, n \in \mathbf{N}^+).$$

The term x_n is called the nth *rational approximation* to the real number x. The set of real numbers is, of course, denoted by $\mathbf{R}$.

We identify a rational number r with the real number $(r, r, r, ...)$; with that identification, $\mathbf{Q}$, $\mathbf{N}$, and $\mathbf{Z}$ become subsets of $\mathbf{R}$.

To specify completely the set $\mathbf{R}$ of real numbers, we must equip it with an appropriate notion of equality. Two real numbers $x = (x_n)$ and $y = (y_n)$ are said to be *equal* if

$$|x_n - y_n| \leq \tfrac{2}{n} \quad (n \in \mathbf{N}^+).$$

Note that this notion of equality is an equivalence relation: it is clearly reflexive and symmetric; its transitivity is a simple consequence of the following result.

(A.1) Lemma. *Two real numbers $x = (x_n)$ and $y = (y_n)$ are equal if and only if for each positive integer k there exists a positive integer N_k such that $|x_n - y_n| \leq 1/k$ whenever $n \geq N_k$.*

Proof. If $x = y$, then for each k we need only take $N_k = 2k$. Conversely, suppose that for each k there exists N_k with the stated property, and consider any positive integers n and k. Setting $m = k + N_k$, we have

$$\begin{aligned} |x_n - y_n| &\leq |x_n - x_m| + |x_m - y_m| + |y_m - y_n| \\ &\leq \left(\tfrac{1}{n} + \tfrac{1}{m}\right) + \tfrac{1}{k} + \left(\tfrac{1}{n} + \tfrac{1}{m}\right) \\ &< \tfrac{2}{n} + \tfrac{3}{k}. \end{aligned}$$

Since this holds for all positive integers k, we see that $|x_n - y_n| \leq 2/n$. But n is arbitrary, so $x = y$. □

(A.2) Exercises

.1 Complete the proof that equality of real numbers is an equivalence relation.

.2 Let k be any positive integer. Show that the operation which assigns to each real number $(x_n)_{n=1}^{\infty}$ its kth rational approximation x_k does not preserve equality.

.3 Prove that two real numbers $x = (x_n)$ and $y = (y_n)$ are equal if and only if for each $c > 0$ and each positive integer k there exists N_k such that $|x_n - y_n| \leq c/k$ for all $n \geq N_k$.

To introduce the algebraic operations on $\mathbf{R}$ we need a special bound for the terms of a regular sequence $x = (x_n)$ of rational numbers. We define the *canonical bound* K_x of x to be the least positive integer greater than $|x_1| + 2$. It is easy to show that $|x_n| < K_x$ for all n.

The arithmetic operations on real numbers $x = (x_n)$ and $y = (y_n)$ are defined in terms of the rational approximations to those numbers as follows.

$$\begin{aligned}
(x+y)_n &= x_{2n} + y_{2n}, \\
(x-y)_n &= x_{2n} - y_{2n}, \\
(xy)_n &= x_{2\kappa n} y_{2\kappa n}, \quad \text{where } \kappa = \max\{K_x, K_y\}, \\
\max\{x,y\}_n &= \max\{x_n, y_n\}, \\
\min\{x,y\}_n &= \min\{x_n, y_n\}, \\
|x|_n &= |x_n|.
\end{aligned}$$

Here, for example, $(x+y)_n$ denotes the nth rational approximation to the real number $x+y$, and $\max\{x_n, y_n\}$ is the maximum, computed in the usual way, of the rational numbers x_n and y_n. Of course, we must verify that the foregoing definitions yield real numbers; we illustrate this verification with the case of the product xy. Writing $z_n = x_{2\kappa n} y_{2\kappa n}$, so that $xy = (z_n)$, for all positive integers m and n we have

$$\begin{aligned}
|z_m - z_n| &= |x_{2\kappa m}(y_{2\kappa m} - y_{2\kappa n}) + y_{2\kappa n}(x_{2\kappa m} - x_{2\kappa n})| \\
&\leq |x_{2\kappa m}|\,|y_{2\kappa m} - y_{2\kappa n}| + |y_{2\kappa n}|\,|x_{2\kappa m} - x_{2\kappa n}| \\
&\leq \kappa\left(\tfrac{1}{2\kappa m} + \tfrac{1}{2\kappa n}\right) + \kappa\left(\tfrac{1}{2\kappa m} + \tfrac{1}{2\kappa n}\right) \\
&= \tfrac{1}{m} + \tfrac{1}{n}.
\end{aligned}$$

Thus xy is a regular sequence of rational numbers—that is, a real number.

In the rest of this appendix, $x = (x_n)$, $y = (y_n)$, and $z = (z_n)$ are real numbers.

(A.3) Exercises

.1 Prove that $|x_n| < K_x$ for each n.

.2 Prove that $x + y$, $x - y$, $\max\{x, y\}$, $\min\{x, y\}$, and $|x|$ are real numbers.

.3 Let x' and y' be real numbers such that $x = x'$ and $y = y'$. Prove that $x + y = x' + y'$ and $xy = x'y'$.

Thus the operations of addition and multiplication arise from *functions* on the Cartesian product $\mathbf{R} \times \mathbf{R}$ when the relation of equality on that set is defined in the natural way: $(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$.

.4 Sums, differences, products, maxima, and minima of finitely many real numbers are defined inductively: for example, we define

$$\max\{x_1, \ldots, x_{n+1}\} = \max\{\max\{x_1, \ldots, x_n\}, x_{n+1}\}.$$

Show that if σ is a permutation of $\{1, \ldots, n\}$, then

$$\max\{x_{\sigma(1)}, \ldots, x_{\sigma(n)}\} = \max\{x_1, \ldots, x_n\}.$$

.5 Prove each of the following identities.

(i) $x + y = y + x$

(ii) $x + (y + z) = (x + y) + z$

(iii) $xy = yx$

(iv) $0 + x = x + 0 = x$

(v) $1x = x1 = x$.

(These should serve to convince you that addition and multiplication, as defined previously, have the properties that we expect from elementary school.)

.6 Prove that for each m the mth rational approximation to $1/n - |x - x_n|$ is $1/n - |x_{4m} - x_n|$.

The real number $x = (x_n)$ is said to be *positive* if there exists n such that $x_n > 1/n$. We define $x > y$ to mean that $x - y$ is positive; thus $x > 0$ if and only if x is positive. On the other hand, we say that x is

- *negative* if $-x$ is positive, and
- *nonnegative* if $x_n \geq -1/n$ for all n.

We write $x \geq y$ to denote that $x - y$ is nonnegative, and we define $x < y$ and $x \leq y$ to have the usual meanings relative to the relations $>, \geq$.

(A.4) Exercise

Prove that if $x > 0$, then $x \geq 0$.

(A.5) Proposition. *A real number $x = (x_n)$ is positive if and only if there exists a positive integer N such that $x_m \geq 1/N$ for all $m \geq N$. On the other hand, x is nonnegative if and only if for each positive integer k there exists a positive integer N_k such that $x_m \geq -1/k$ for all $m \geq N_k$.*

Proof. If x is positive, then $x_n > 1/n$ for some n. Choosing the positive integer N so that $2/N \leq x_n - 1/n$, for each $m \geq N$ we have

$$\begin{aligned} x_m &\geq x_n - |x_m - x_n| \\ &\geq x_n - \tfrac{1}{m} - \tfrac{1}{n} \\ &\geq x_n - \tfrac{1}{N} - \tfrac{1}{n} \\ &> \tfrac{1}{N}. \end{aligned}$$

So the required property holds. If, conversely, that property holds, then $x_{N+1} > 1/(N+1)$, so $x > 0$.

The proof of the second part of the proposition is left as an exercise. □

(A.6) Exercises

.1 Prove the second part of the preceding proposition.

.2 Prove that if $x = x'$, $y = y'$, and $x > y$ (respectively, $x \geq y$), then $x' > y'$ (respectively, $x' \geq y'$).

.3 Prove the *Axiom of Archimedes*: if $x > 0$ and $y \geq 0$, then there exists $n \in \mathbf{N}^+$ such that $nx > y$.

.4 Prove that on $\mathbf{Q}$ the relations $>$ and $\geq$, defined as for real numbers, coincide with the standard elementary order relations between rational numbers.

.5 Prove the triangle inequality for real numbers: $|x + y| \leq |x| + |y|$.

It is left as a relatively straightforward exercise to prove most of the elementary properties of the partial orders $>, \geq$ on $\mathbf{R}$. However, we need to tie up a few loose ends, the first of which concerns the order density of $\mathbf{Q}$ in $\mathbf{R}$ and requires a simple lemma.

(A.7) Lemma. $|x - x_n| \leq 1/n$ *for each* n.

Proof. Fix the positive integer n. By Exercise (A.3: 6), for each m the mth rational approximation to $1/n - |x - x_n|$ is

$$\tfrac{1}{n} - |x_{4m} - x_n| \geq \tfrac{1}{n} - \left(\tfrac{1}{4m} + \tfrac{1}{n}\right) = -\tfrac{1}{4m} > -\tfrac{1}{m}.$$

Hence $1/n - |x - x_n| \geq 0$, and therefore $|x - x_n| \leq 1/n$. □

(A.8) Proposition. $\mathbf{Q}$ *is* order dense *in* $\mathbf{R}$*—that is, for all* x *and* y *in* $\mathbf{R}$ *with* $x < y$, *there exists* $r \in \mathbf{Q}$ *such that* $x < r < y$.

Proof. Since

$$0 < y - x = (y_{2n} - x_{2n})_{n=1}^{\infty},$$

there exists N such that $y_{2N} - x_{2N} > 1/N$. Writing

$$r = \tfrac{1}{2}\left(x_{2N} + y_{2N}\right)$$

and using Lemma (A.7), we have

$$\begin{aligned} r - x &\geq r - x_{2N} - |x_{2N} - x| \\ &\geq \tfrac{1}{2}\left(y_{2N} - x_{2N}\right) - \tfrac{1}{2N} > 0, \end{aligned}$$

and similarly $y - r > 0$. Hence $x < r < y$. □

Here is a good application of Proposition (A.8).

(A.9) Proposition. *If $x + y > 0$, then either $x > 0$ or $y > 0$.*

Proof. Let $x + y > 0$. By Proposition (A.8), there exists a rational number α such that $0 < \alpha < x + y$. Using Exercise (A.6: 3), choose a positive integer $n > 4/\alpha$. Let $r = x_n$ and $s = y_n$. Then r and s are rational; also, by Lemma (A.7), $|x - r| < \alpha/4$ and $|y - s| < \alpha/4$. Using the triangle inequality, we now see that

$$\begin{aligned} r + s &\geq (x + y) - (|x - r| + |y - s|) \\ &> \alpha - \left(\tfrac{\alpha}{4} + \tfrac{\alpha}{4}\right) \\ &= \tfrac{\alpha}{2}. \end{aligned}$$

Since r and s are rational numbers, either $r > \alpha/4$ or $s > \alpha/4$. In the first case, $x \geq r - |x - r| > 0$; in the second, $y > 0$. □

For each nonzero real number x the *reciprocal*, or *inverse*, of x is the real number $\frac{1}{x}$ (also written $1/x$ or x^{-1}) defined as follows. Choose a positive integer N such that $|x_n| \geq 1/N$ for all $n \geq N$, and set

$$\left(\frac{1}{x}\right)_n = \begin{cases} 1/x_{N^3} & \text{if } n < N \\ 1/x_{nN^2} & \text{if } n \geq N. \end{cases}$$

The last set of exercises in this appendix shows that this is a good definition of $1/x$.

(A.10) Exercises

.1 Let x be a nonzero real number, and $1/x$ the reciprocal of x as just defined. Prove that $1/x$ is a real number, and that it is the unique real number t such that $xt = 1$.

.2 Let x be a nonzero real number, and let N be as in the definition of $1/x$. Let M be a positive integer such that $|x_n| \geq 1/M$ for all $n \geq M$, and define a real number $y = (y_n)$ by

$$y_n = \begin{cases} 1/x_{M^3} & \text{if } n < M \\ 1/x_{nM^2} & \text{if } n \geq M. \end{cases}$$

Give two proofs that $y = 1/x$.

.3 Prove that the operation that assigns $1/x$ to the nonzero real number x is a function (respects equality) and maps the set of nonzero real numbers onto itself.

Appendix B

Axioms of Choice and Zorn's Lemma

In the early years of this century it was recognised that the following principle, the *Axiom of Choice,* was necessary for the proofs of several important theorems in mathematics.

AC *If $\mathcal{F}$ is a nonempty family of pairwise–disjoint nonempty sets, then there exists a set that intersects each member of $\mathcal{F}$ in exactly one element.*

In particular, Zermelo used this axiom explicitly in his proof that every set S can be well–ordered—that is, there is a total partial order $\geq$ on S with respect to which every nonempty subset of X has a least element [57]. It was shown by Gödel [18] in 1939 that the Axiom of Choice is consistent with the axioms of Zermelo–Fraenkel set theory (ZF), in the sense that the axiom can be added to ZF without leading to a contradiction, and by Cohen [11] in 1963 that the negation of the Axiom of Choice is also consistent with ZF. Thus the Axiom of Choice is *independent* of ZF: it can be neither proved nor disproved without adding some extra principles to ZF.

The Axiom of Choice is commonly used in an equivalent form (the one we used in the proof of Lemma (1.3.5)):

AC′ *If A and B are nonempty sets, $S \subset A \times B$, and for each $x \in A$ there exists $y \in B$ such that $(x, y) \in S$, then there exists a function $f : A \to B$ —called a* choice function *for S —such that $(x, f(x)) \in S$ for each $x \in A$.*

To prove the equivalence of these two forms of the Axiom of Choice, first assume that the original version AC of the axiom holds, and consider nonempty sets A, B and a subset S of $A \times B$ such that for each $x \in A$ there exists $y \in B$ with $(x, y) \in S$. For each $x \in A$ let

$$F_x = \{x\} \times \{y \in B : (x, y) \in S\}.$$

Then $\mathcal{F} = (F_x)_{x \in A}$ is a nonempty family of pairwise–disjoint sets, so, by AC, there exists a set C that has exactly one element in common with each F_x. We now define the required choice function $f : A \to B$ by setting

$$(x, f(x)) = \text{ the unique element of } C \cap F_x$$

for each $x \in A$.

Now assume that the alternative form AC$'$ of the Axiom of Choice holds, and consider a nonempty family $\mathcal{F}$ of pairwise–disjoint nonempty sets. Taking

$$\begin{aligned} A &= \mathcal{F}, \\ B &= \bigcup_{X \in \mathcal{F}} X, \\ S &= \{(X, x) : X \in \mathcal{F},\, x \in X\} \end{aligned}$$

in AC$'$, we obtain a function

$$f : \mathcal{F} \to \bigcup_{X \in \mathcal{F}} X$$

such that $f(X) \in X$ for each $X \in \mathcal{F}$. The range of f is then a set that has exactly one element in common with each member of $\mathcal{F}$.

There are two other choice principles that are widely used in analysis. The first of these, the *Principle of Countable Choice,* is the case $A = \mathbf{N}$ of AC$'$. The second is the *Principle of Dependent Choice*:

> *If $a \in A$, $S \subset A \times A$, and for each $x \in A$ there exists $y \in A$ such that $(x, y) \in S$, then there exists a sequence $(a_n)_{n=1}^{\infty}$ in A such that $a_1 = a$ and $(a_n, a_{n+1}) \in S$ for each n.*

It is a good exercise to show that the Axiom of Choice entails the Principle of Dependent Choice, and that the Principle of Dependent Choice entails the Principle of Countable Choice. Since the last two principles can be derived as consequences of the axioms of ZF, they are definitely weaker than the Axiom of Choice.

There are many principles that are equivalent to the Axiom of Choice. One of those, Zorn's Lemma, is needed for our proof of the Hahn–Banach Theorem in Chapter 6.

A nonempty subset C of a partially ordered set $(A, \succcurlyeq)$ is called a *chain* if for all $x, y \in C$ either $x \succcurlyeq y$ or $y \succcurlyeq x$. *Zorn's Lemma* states that

> *If every chain in a partially ordered set A has an upper bound in A, then A has a maximal element.*

For a fuller discussion of axioms of choice, Zorn's Lemma, and related matters, see the article by Jech on pages 345–370 of [2].

Appendix C
Pareto Optimality

In this appendix we show how some of the results and ideas in our main chapters can be applied within theoretical economics.

We assume that there are a finite number m of *consumers* and a finite number n of *producers*. Consumer i has a *consumption set* $X_i \subset \mathbf{R}^N$, where a *consumption bundle* $x_i = (x_{i_1}, \ldots, x_{i_N}) \in X_i$ is interpreted as follows: x_{i_k} is the quantity of the kth commodity (a good or a service) taken by consumer i when he chooses the consumption bundle x_i. Producer j has a *production set* $Y_j \subset \mathbf{R}^N$, where the kth entry in the *production vector* $y_j = (y_{j_1}, \ldots, y_{j_N}) \in Y_j$ is interpreted as the amount of the kth commodity produced by producer j under her adopted production schedule. Other important sets in this context are the *aggregate consumption set*

$$X = X_1 + \cdots + X_m$$

and the *aggregate production set*

$$Y = Y_1 + \cdots + Y_n.$$

A *price vector* is simply an element p of $\mathbf{R}^N$; the kth component p_k of p is the price of one unit of the kth commodity. Thus the total cost to consumer i of the consumption bundle x_i is $\langle p, x_i \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbf{R}^N$; and the profit to producer j of the production vector y_j is $\langle p, y_j \rangle$.

We assume that the preferences of consumer i are represented by a reflexive, transitive total partial order $\succcurlyeq_i$ on X_i, called the *preference relation* of consumer i. The corresponding relations $\succ_i$ of *strict preference*, and $\sim_i$

of *preference–indifference*, are defined on X_i as follows.

$$\begin{array}{lll} x \succ_i y & \text{if and only if} & x \succcurlyeq_i y \text{ and not } (y \succcurlyeq_i x)\,; \\ x \sim_i y & \text{if and only if} & x \succcurlyeq_i y \text{ and } y \succcurlyeq_i x. \end{array}$$

Routine arguments show that $\succ_i$ and $\sim_i$ are transitive; that $x \succ_i x$ is contradictory; that if either $x \succ_i y$ or $x \sim_i y$, then $x \succcurlyeq_i y$; and that if either $x \succ_i y \succcurlyeq_i z$ or $x \succcurlyeq_i y \succ_i z$, then $x \succ_i z$. The informal meaning of $x \succcurlyeq_i y$ is that consumer i finds x at least as attractive as y; $x \succ_i y$ means that he strictly prefers x to y; and $x \sim_i y$ signifies that he does not mind which of x or y he obtains.

It is convenient to introduce consumer i's *upper contour set at* x,

$$[x, \rightarrow) = \{\xi \in X_i : \xi \succcurlyeq_i x\}\,,$$

and his *strict upper contour set at* x,

$$(x, \rightarrow) = \{\xi \in X_i : \xi \succ_i x\}\,.$$

The preference relation $\succ_i$ is said to be *locally nonsatiated at* $x_i \in X_i$ if for each $\varepsilon > 0$, $B(x_i, \varepsilon) \cap (x_i, \rightarrow)$ is nonempty—that is, there exists $x_i' \in X_i$ such that $\|x_i - x_i'\| < \varepsilon$ and $x_i' \succ_i x_i$.

By a *chosen point for consumer* i *under the price vector* p we mean a point $\xi_i \in X_i$ such that for all $x_i \in X_i$,

$$\langle p, \xi_i \rangle \geq \langle p, x_i \rangle \Rightarrow \xi_i \succcurlyeq_i x_i$$

or, equivalently,

$$x_i \succ_i \xi_i \Rightarrow \langle p, x_i \rangle > \langle p, \xi_i \rangle\,.$$

(C.1) Lemma. *If* $\xi_i \in X_i$ *is a chosen point for consumer* i *under the price vector* p*, and* $x_i \sim \xi_i$ *is a point of* X_i *at which* $\succ_i$ *is locally nonsatiated, then* $\langle p, x_i \rangle \geq \langle p, \xi_i \rangle$.

Proof. Suppose that $\langle p, x_i \rangle < \langle p, \xi_i \rangle$. By the continuity of the mapping $x \mapsto \langle p, x \rangle$ on $\mathbf{R}^N$, there exists $r > 0$ such that if $x_i' \in X_i$ and $\|x_i' - x_i\| < r$, then $\langle p, x_i' \rangle < \langle p, \xi_i \rangle$. As $\succcurlyeq_i$ is locally nonsatiated at x_i, there exists $x_i' \in X_i$ such that $x_i' \succ_i x_i$ and $\|x_i' - x_i\| < r$. Then $\langle p, \xi_i \rangle > \langle p, x_i' \rangle$; so $\xi_i \succcurlyeq_i x_i'$, as ξ_i is a chosen point. But we also have $x_i' \succ_i x_i \sim_i \xi_i$ and therefore $x_i' \succ_i \xi_i$, a contradiction. □

We now assume that consumer i has an initial endowment of commodities, represented by the vector $\bar{x}_i = (\bar{x}_{i_1}, \ldots, \bar{x}_{i_N})$. The total initial endowment of all consumers is then

$$\bar{x} = \bar{x}_1 + \cdots + \bar{x}_m \in X.$$

We say that an element $(y_1, \ldots, y_n)$ of $Y_1 \times \cdots \times Y_n$ is an *admissible array of production vectors*; and that an element $(x_1, \ldots, x_m)$ of $X_1 \times \cdots \times X_m$ is a *feasible array of consumption bundles* if there exists an admissible array $(y_1, \ldots, y_n)$ of production vectors such that

$$\sum_{i=1}^{m} x_i = \sum_{j=1}^{n} y_j + \bar{x}.$$

Intuitively, a feasible array is one that can be obtained by a distribution of the total initial endowment and the total of the production vectors under some production schedule.

An array $(\xi_1, \ldots, \xi_m) \in X_1 \times \cdots \times X_m$ of consumption bundles is said to be *Pareto optimal*, or a *Pareto optimum*, if it is feasible and if the following condition holds.

PO If $(x_1, \ldots, x_m)$ is a feasible array such that $x_i \succ_i \xi_i$ for some i, then there exists k such that $\xi_k \succ_k x_k$.

Equivalently, the array is Pareto optimal if there is no feasible array $(x_1, \ldots, x_m)$ such that $x_i \succcurlyeq_i \xi_i$ for all i, and such that $x_i \succ_i \xi_i$ for at least one i.

By a *competitive equilibrium* we mean a triple consisting of a price vector p, an array $(\xi_1, \ldots, \xi_m)$ of consumption bundles, and an admissible array $(\eta_1, \ldots, \eta_n)$ of production vectors, satisfying the following conditions.

CE1 For $1 \le i \le m$, ξ_i is a chosen point for consumer i under the price vector p.

CE2 For $1 \le j \le n$, if $y_j \in Y_j$, then $\langle p, \eta_j \rangle \ge \langle p, y_j \rangle$.

CE3 $\sum_{i=1}^{m} \xi_i = \sum_{j=1}^{n} \eta_j + \bar{x}$.

Condition CE1 expresses consumer satisfaction; CE2, profit maximisation; and CE3, feasibility.

(C.2) Proposition. *Assume that each $\succcurlyeq_i$ is locally nonsatiated, and let*

$$(p,\ (\xi_1, \ldots, \xi_m),\ (\eta_1, \ldots, \eta_n))$$

be a competitive equilibrium. Then $(\xi_1, \ldots, \xi_m)$ is a Pareto optimum.

Proof. Condition **CE3** ensures that $(\xi_1, \ldots, \xi_m)$ is a feasible array of consumption bundles. Suppose that $(\xi_1, \ldots, \xi_m)$ is not a Pareto optimum. Then there exist an array $(x_1, \ldots, x_m)$ of consumption bundles and an admissible array $(y_1, \ldots, y_n)$ of production vectors such that

$$\sum_{i=1}^{m} x_i = \sum_{j=1}^{n} y_j + \bar{x}, \tag{1}$$

$x_i \succcurlyeq_i \xi_i$ for all i, and $x_k \succ_k \xi_k$ for some k. By CE1, if $x_i \succ_i \xi_i$, then $\langle p, x_i \rangle > \langle p, \xi_i \rangle$; in particular, $\langle p, x_k \rangle > \langle p, \xi_k \rangle$. If $\xi_i \succcurlyeq_i x_i$, then $x_i \sim_i \xi_i$ and so, by Lemma (C.1), $\langle p, x_i \rangle \geq \langle p, \xi_i \rangle$. Thus

$$\begin{aligned}\sum_{i=1}^{m} \langle p, x_i \rangle &> \sum_{i=1}^{m} \langle p, \xi_i \rangle \\ &= \sum_{j=1}^{n} \langle p, \eta_j \rangle + \langle p, \bar{x} \rangle \quad \text{(by CE3)} \\ &\geq \sum_{j=1}^{n} \langle p, y_j \rangle + \langle p, \bar{x} \rangle \quad \text{(by CE2).}\end{aligned}$$

Hence

$$\left\langle p, \left(\sum_{i=1}^{m} x_i - \sum_{j=1}^{n} y_j - \bar{x} \right) \right\rangle > 0,$$

and therefore, by the Cauchy–Schwarz inequality in $\mathbf{R}^N$,

$$\sum_{i=1}^{m} x_i \neq \sum_{j=1}^{n} y_j + \bar{x}.$$

This contradicts (1). □

Our next aim is to establish a partial converse of Proposition (C.2), providing conditions under which a Pareto optimum gives rise to a competitive equilibrium. We first introduce some more definitions.

The preference relation $\succcurlyeq_i$ on X_i is said to be *convex* if

- X_i is convex,
- $x \succ_i x' \Rightarrow tx + (1-t)x' \succ_i x'$ whenever $0 < t < 1$, and
- $x \sim_i x' \Rightarrow tx + (1-t)x' \succcurlyeq_i x'$ whenever $0 < t < 1$.

In that case the sets $[x, \rightarrow)$ and $(x, \rightarrow)$ are convex.

We say that consumer i is *nonsatiated* at $\xi_i \in X_i$ if there exists $x \in X_i$ such that $x \succ_i \xi_i$; otherwise, we say that he is *satiated* at ξ_i.

(C.3) Proposition. *Let $(\xi_1, \ldots, \xi_m)$ be a Pareto optimum such that for at least one value of i, consumer i is nonsatiated at ξ_i, and let $(\eta_1, \ldots, \eta_n)$ be an admissible array of production vectors. Suppose that $\succcurlyeq_i$ is convex for each i, and that the aggregate production set Y is convex. Then there exists a nonzero price vector p such that*

(i) *for each i, if $x_i \in X_i$ and $x_i \succcurlyeq_i \xi_i$, then $\langle p, x_i \rangle \geq \langle p, \xi_i \rangle$;*

(ii) *for each j, if $y_j \in Y_j$, then $\langle p, \eta_j \rangle \geq \langle p, y_j \rangle$.*

Proof. We may assume that consumer 1 is nonsatiated at ξ_1. Choose an admissible array $(\eta_1, \ldots, \eta_n)$ of production vectors such that

$$\xi = \sum_{i=1}^{m} \xi_i = \sum_{j=1}^{n} \eta_j + \bar{x}.$$

Let A be the algebraic sum of the sets $(\xi_1, \rightarrow)$ and $\sum_{i=2}^{m} [\xi_i, \rightarrow)$,

$$A = \left\{ \sum_{i=1}^{N} x_i \in \mathbf{R}^N : x_1 \succ_1 \xi_1 \text{ and } \forall i \geq 2 \ (x_i \succcurlyeq_i \xi_i) \right\},$$

and let

$$B = \left\{ x \in \mathbf{R}^N : \exists y \in Y \ (x = y + \bar{x}) \right\}.$$

Clearly, B is convex; by our convexity hypotheses, A is convex. If $A \cap B$ is nonempty, then there exist $x_1 \succ_1 \xi_1$, $x_i \succcurlyeq_i \xi_i$ $(2 \leq i \leq m)$, and $y_j \in Y_j$ $(1 \leq j \leq n)$, such that

$$\sum_{i=1}^{m} x_i = \sum_{j=1}^{n} y_j + \bar{x}.$$

This contradicts the hypothesis that $(\xi_1, \ldots, \xi_m)$ is a Pareto optimum. Hence A and B are disjoint subsets of $\mathbf{R}^N$. Since these sets are clearly nonempty, it follows from Minkowski's Separation Theorem (6.2.6) and the Riesz Representation Theorem (5.3.1) that there exist a nonzero vector $p \in \mathbf{R}^N$ and a real number α such that $\langle p, x \rangle \geq \alpha$ for all $x \in A$, and $\langle p, x \rangle \leq \alpha$ for all $x \in B$. Since $\xi \in B$, we have $\langle p, \xi \rangle \leq \alpha$. We now show that $\langle p, \xi \rangle = \alpha$.

To this end, consider $\sum_{i=1}^{m} x_i$,with $x_1 \succ_1 \xi_1$ and $x_i \succcurlyeq_i \xi_i$ $(2 \leq i \leq m)$. For $0 < t < 1$ define

$$z_i(t) = t x_i + (1 - t) \xi_i \quad (1 \leq i \leq m)$$

and

$$z(t) = \sum_{i=1}^{m} z_i(t).$$

Since $\succcurlyeq_i$ is convex for each i,

$$z_1(t) \in (\xi_1, \rightarrow),$$
$$z_i(t) \in [\xi_i, \rightarrow) \quad (2 \leq i \leq m);$$

whence $z(t) \in A$ and therefore $\langle p, z(t) \rangle \geq \alpha$. Letting $t \rightarrow 0$ and using the continuity of the mapping $x \mapsto \langle p, x \rangle$ on $\mathbf{R}^N$, we see that $\langle p, \xi \rangle \geq \alpha$ and therefore that $\langle p, \xi \rangle = \alpha$, as we wanted to show.

It now follows that $\langle p, x\rangle \geq \langle p, \xi\rangle$ for all $x \in A$, and that $\langle p, x\rangle \leq \langle p, \xi\rangle$ for all $x \in B$. Thus if $(y_1, \ldots, y_n)$ is an admissible array of production vectors, then

$$\left\langle p, \sum_{j=1}^{n} y_j + \bar{x} \right\rangle \leq \langle p, \xi\rangle = \left\langle p, \sum_{j=1}^{n} \eta_j + \bar{x} \right\rangle$$

and therefore

$$\sum_{j=1}^{n} \langle p, y_j\rangle \leq \sum_{j=1}^{n} \langle p, \eta_j\rangle .$$

Given $j \in \{1, \ldots, n\}$, and taking $y_j \in Y_j$ and $y_k = \eta_k$ for all $k \neq j$ $(1 \leq k \leq n)$, we now obtain $\langle p, \eta_j\rangle \geq \langle p, y_j\rangle$. This completes the proof of (ii).

A similar argument, using the fact that $\langle p, x\rangle \geq \langle p, \xi\rangle$ for all $x \in A$, shows that

$$\langle p, x_1\rangle \geq \langle p, \xi_1\rangle \text{ for all } x_1 \in (\xi_1, \rightarrow) \tag{2}$$

and that $\langle p, x_i\rangle \geq \langle p, \xi_i\rangle$ for all $x_i \in [\xi_i, \rightarrow)$ $(2 \leq i \leq m)$. To complete the proof of (i), we show that if $x_1 \sim_1 \xi_1$, then $\langle p, x_1\rangle \geq \langle p, \xi_1\rangle$. To this end, we recall that consumer 1 is nonsatiated at ξ_1, so there exists $x_1' \in X_1$ with

$$x_1' \succ_1 \xi_1 \sim_1 x_1.$$

It follows from this and the convexity of $\succsim_1$ that for each $t \in (0,1)$,

$$x_1'(t) = tx_1' + (1-t)x_1 \succ_1 \xi_1;$$

whence $\langle p, x_1'(t)\rangle \geq \langle p, \xi_1\rangle$, by (2). The continuity of the function $x \mapsto \langle p, x\rangle$ on $\mathbf{R}^N$ now ensures that $\langle p, x_1\rangle \geq \langle p, \xi_1\rangle$, as we required. This completes the proof of (i). □

(C.4) Corollary. *Under the hypotheses of* Proposition (C.3), *suppose also that the following conditions hold.*

(i) *For each price vector p and each i $(1 \leq i \leq m)$, there exists $\xi_i' \in X_i$ such that $\langle p, \xi_i'\rangle < \langle p, \xi_i\rangle$* (cheaper point condition).

(ii) *For each i $(1 \leq i \leq m)$, $(\xi_i, \rightarrow)$ is open in X_i.*

Then $(p, (\xi_1, \ldots, \xi_m), (\eta_1, \ldots, \eta_n))$ is a competitive equilibrium.

Proof. In view of Proposition (C.3), we need only prove that CE1 holds. To this end, let $x_i \succ_i \xi_i$, and choose $\xi_i' \in X_i$ as in hypothesis (i). Then, by Proposition (C.3), $\xi_i \succ_i \xi_i'$. For each $t \in (0,1)$ define

$$x_i(t) = t\xi_i' + (1-t)x_i.$$

As $(\xi_i, \rightarrow)$ is open in X_i, we can choose $t \in (0,1)$ so small that $x_i(t) \succ_i \xi_i$. Then, by Proposition (C.3),

$$\begin{aligned} t\langle p, \xi_i'\rangle + (1-t)\langle p, x_i\rangle &= \langle p, x_i(t)\rangle \\ &\geq \langle p, \xi_i\rangle \\ &= t\langle p, \xi_i\rangle + (1-t)\langle p, \xi_i\rangle \\ &> t\langle p, \xi_i'\rangle + (1-t)\langle p, \xi_i\rangle . \end{aligned}$$

Hence

$$(1-t)\langle p, x_i\rangle > (1-t)\langle p, \xi_i\rangle$$

and therefore $\langle p, x_i\rangle > \langle p, \xi_i\rangle$. Thus ξ_i is a chosen point. □

The cheaper point assumption cannot be omitted from the hypotheses of Corollary (C.4); see pages 198–201 of [51].

References

The following list contains both works that were consulted during the writing of this book and suggestions for further reading.

University libraries usually have lots of older books, such as [36], dealing with classical real analysis at the level of Chapter 1; a good modern reference for this material is [16]. Excellent references for the abstract theory of measure and integration, following on from the material in Chapter 2, are [21], [44], and [43]. (Note, incidentally, the advocacy of a Riemann–like integral by some authors [1].) Dieudonné's book [13], the first of a series in which he covers a large part of modern analysis, is outstanding and was a source of much inspiration in my writing of Chapters 3 through 5. An excellent text for a general course on functional analysis is [45]. This could be followed by, or taken in conjunction with, material from the two volumes by Kadison and Ringrose [24] on operator algebra theory, currently one of the most active and important branches of analysis. Two other excellent books, each of which overlaps our book in some areas but goes beyond it in others, are [34], which includes such topics as spectral theory and abstract integration, and [14], which extends measure theory into a rigorous development of probability. More specialised books expanding material covered in Chapter 6 are the one by Oxtoby [33] on the interplay between Baire category and measure, and Diestel's absorbing text [12] on sequences and series in Banach spaces.

A wonderful book, written in a more discursive style than most others at this level, is the classic by Riesz and Nagý [40]; although more old–fashioned in its approach (it was first published in 1955), it is a source of much valuable material on Lebesgue integration and the theory of operators on Hilbert space. A relatively unusual approach to analysis, in which all concepts and proofs must be fully constructive, is followed in [5]; see also Chapter 4 of [8].

For general applications of functional analysis see Zeidler's two volumes [56]. Applications of analysis in mathematical economics can be found in [9], [30], and [51].

[1] R. G. Bartle: Return to the Riemann integral, *Amer. Math. Monthly* 103 (1996), 625–632.

[2] J. Barwise: *Handbook of Mathematical Logic,* North–Holland, Amsterdam, 1977.

[3] G. H. Behforooz: Thinning out the harmonic series, *Math. Mag.* 68(4), 289–293, 1985.

[4] A. Bielicki: Une remarque sur la méthode de Banach–Cacciopoli–Tikhonov, *Bull. Acad. Polon. Sci.* IV (1956), 261–268.

[5] E.A. Bishop and D.S. Bridges: *Constructive Analysis,* Grundlehren der math. Wissenschaften 279, Springer–Verlag, Berlin–Heidelberg–New York, 1985.

[6] P. Borwein and T. Erdélyi: The full Müntz theorem in $\mathcal{C}[0,1]$ and $L_1[0,1]$, *J. London Math. Soc.* (2), 54 (1996), 102–110.

[7] N. Bourbaki: *Eléments de Mathématique, Livre III: Topologie Générale,* Hermann, Paris, 1958.

[8] D.S. Bridges: *Computability: A Mathematical Sketchbook,* Graduate Texts in Mathematics 146, Springer–Verlag, Berlin–Heidelberg–New York, 1994.

[9] D.S. and G.B. Mehta: *Representations of Preference Orderings,* Lecture Notes in Economics and Mathematical Systems 422, Springer–Verlag, Berlin–Heidelberg–New York, 1995.

[10] E.W. Cheney: *Introduction to Approximation Theory,* McGraw–Hill, New York, 1966.

[11] P.J. Cohen: *Set Theory and the Continuum Hypothesis,* W.A. Benjamin, Inc., New York, 1966.

[12] J. Diestel: *Sequences and Series in Banach Spaces,* Graduate Texts in Mathematics 92, Springer–Verlag, Berlin–Heidelberg–New York, 1984.

[13] J. Dieudonné: *Foundations of Modern Analysis,* Academic Press, New York, 1960.

[14] R.M. Dudley, *Real Analysis and Probability,* Chapman & Hall, New York, 1989.

[15] P. Enflo: A counterexample to the approximation property in Banach spaces, *Acta Math.* 130 (1973), 309–317.

[16] E. Gaughan: *Introduction to Analysis* (4th Edn), Brooks/Cole, Pacific Grove, CA, 1993.

[17] R.P. Gillespie: *Integration,* Oliver & Boyd, Edinburgh, 1959.

[18] K. Gödel: *The Consistency of the Axiom of Choice and the Generalized Continuum Hypothesis with the Axioms of Set Theory,* Annals of Mathematics Studies, Vol. 3, Princeton University Press, Princeton, NJ, 1940.

[19] R. Gray: Georg Cantor and transcendental numbers, *Amer. Math. Monthly* 101 (1994), 819–832.

[20] P.R. Halmos: *Naive Set Theory,* van Nostrand, Princeton, NJ, 1960; reprinted as Undergraduate Texts in Mathematics, Springer–Verlag, Berlin–Heidelberg–New York, 1974.

[21] P.R. Halmos: *Measure Theory*, van Nostrand, Princeton, NJ, 1950; reprinted as Graduate Texts in Mathematics 18, Springer–Verlag, Berlin–Heidelberg–New York, 1975.

[22] J. Hennefeld: A nontopological proof of the uniform boundedness theorem, *Amer. Math. Monthly* 87 (1980), 217.

[23] F. John: *Partial Differential Equations* (4th Edn), Applied Mathematical Sciences 1, Springer–Verlag, Berlin–Heidelberg–New York, 1982.

[24] R.V. Kadison and J.R. Ringrose: *Fundamentals of the Theory of Operator Algebras,* Academic Press, New York, 1983 (Vol. 1) and 1986 (Vol. 2).

[25] J.L. Kelley: *General Topology,* van Nostrand, Princeton, NJ, 1955; reprinted as Graduate Texts in Mathematics 27, Springer–Verlag, Berlin–Heidelberg–New York, 1975.

[26] D. Kincaid and E.W. Cheney: *Numerical Analysis* (2nd Edn), Brooks/Cole Publishing Co., Pacific Grove, CA, 1996.

[27] M. Kline: *Mathematical Thought from Ancient to Modern Times,* Oxford University Press, Oxford, 1972.

[28] T.W. Körner: *Fourier Analysis,* Cambridge University Press, Cambridge, 1988.

[29] J. Marsden and A. Tromba: *Vector Calculus* (3rd Edn), W.H. Freeman & Co., New York, 1988.

[30] A. Mas–Colell, M.D. Whinston, J.R. Green: *Microeconomic Theory,* Oxford University Press, Oxford, 1995.

[31] Y. Matsuoka: An elementary proof of the formula $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$, *Amer. Math. Monthly* 68 (1961), 485–487.

[32] N.S. Mendelsohn: An application of a famous inequality, *Amer. Math. Monthly* 58 (1951), 563.

[33] J.C. Oxtoby: *Measure and Category,* Graduate Texts in Mathematics 2, Springer–Verlag, Berlin–Heidelberg–New York, 1971.

[34] G.K. Pedersen: *Analysis Now,* Graduate Texts in Mathematics 118, Springer–Verlag, Berlin–Heidelberg–New York, 1991.

[35] W.E. Pfaffenberger: A converse to a completeness theorem, *Amer. Math. Monthly* 87 (1980), 216.

[36] E.G. Phillips: *A Course of Analysis* (2nd Edn), Cambridge Univ. Press, Cambridge, 1939.

[37] J. Rauch: *Partial Differential Equations,* Graduate Texts in Mathematics 128, Springer–Verlag, Berlin–Heidelberg–New York, 1991.

[38] J.R. Rice: *The Approximation of Functions* (Vol. 1), Addison–Wesley, Reading, MA, 1964.

[39] F. Riesz: Sur l'intégrale de Lebesgue comme l'opération inverse de la dérivation, *Ann. Scuola Norm. Sup. Pisa* (2) 5, 191–212 (1936).

[40] F. Riesz and B. Sz–Nagy: *Functional Analysis,* Frederic Ungar Publishing Co., New York, 1955. Republished by Dover Publications Inc., New York, 1990.

[41] J. Ritt: *Integration in Finite Terms,* Columbia University Press, New York, 1948.

[42] W.W. Rogosinski: *Volume and Integral,* Oliver & Boyd, Edinburgh, 1962.

[43] H. Royden: *Real Analysis* (3rd Edn), Macmillan, New York, 1988.

[44] W. Rudin: *Real and Complex Analysis,* McGraw–Hill, New York, 1970.

[45] W. Rudin: *Functional Analysis* (2nd Edn), McGraw–Hill, New York, 1991.

[46] S. Saks: *Theory of the Integral* (2nd Edn), Dover Publishing, Inc., New York, 1964.

[47] H. Schubert: *Topology* (S. Moran, transl.), Macdonald Technical & Scientific, London, 1968.

[48] R.M. Solovay: A model of set theory in which every set of reals is Lebesgue measurable, *Ann. Math.* (Ser. 2) 92, 1–56 (1970).

[49] M. Spivak: *Calculus,* W.A. Benjamin, London, 1967.

[50] B. Sz–Nagy: *Introduction to Real Functions and Orthogonal Expansions,* Oxford University Press, New York, 1965.

[51] A. Takayama: *Mathematical Economics,* The Dryden Press, Hinsdale IL., 1974.

[52] J.A. Todd: *Introduction to the Constructive Theory of Functions,* Birkhäuser Verlag, Basel, 1963.

[53] C. de la Vallée Poussin: *Intégrales de Lebesgue, fonctions d'ensemble, classes de Baire,* Gauthier–Villars, Paris, 1916.

[54] B.L. van der Waerden: Ein einfaches Beispiel einer nichtdifferenzierbaren stetigen Funktion, *Math. Zeitschr.* 32, 474–475, 1930.

[55] Y.M. Wong: The Lebesgue covering property and uniform continuity, *Bull. London Math. Soc.* 4, 184–186, 1972.

[56] E. Zeidler: *Applied Functional Analysis* (2 Vols), Applied Mathematical Sciences 108–109, Springer–Verlag, Berlin–Heidelberg–New York,1995.

[57] E. Zermelo: Beweis, dass jede Menge wohlgeordnet werden kann, *Math. Annalen* 59 (1904) 514–516.

Index