

Andrew Browder

# MATHEMATICAL ANALYSIS

An Introduction



Springer

# Undergraduate Texts in Mathematics

---

- Anglin:** Mathematics: A Concise History and Philosophy.  
*Readings in Mathematics.*
- Anglin/Lambek:** The Heritage of Thales.  
*Readings in Mathematics.*
- Apostol:** Introduction to Analytic Number Theory. Second edition.
- Armstrong:** Basic Topology.
- Armstrong:** Groups and Symmetry.
- Axler:** Linear Algebra Done Right.
- Bak/Newman:** Complex Analysis.
- Banchoff/Wermer:** Linear Algebra Through Geometry. Second edition.
- Berberian:** A First Course in Real Analysis.
- Brémaud:** An Introduction to Probabilistic Modeling.
- Bressoud:** Factorization and Primality Testing.
- Bressoud:** Second Year Calculus.  
*Readings in Mathematics.*
- Brickman:** Mathematical Introduction to Linear Programming and Game Theory.
- Browder:** Mathematical Analysis: An Introduction.
- Cederberg:** A Course in Modern Geometries.
- Childs:** A Concrete Introduction to Higher Algebra. Second edition.
- Chung:** Elementary Probability Theory with Stochastic Processes. Third edition.
- Cox/Little/O'Shea:** Ideals, Varieties, and Algorithms.
- Croom:** Basic Concepts of Algebraic Topology.
- Curtis:** Linear Algebra: An Introductory Approach. Fourth edition.
- Devlin:** The Joy of Sets: Fundamentals of Contemporary Set Theory. Second edition.
- Dixmier:** General Topology.
- Driver:** Why Math?
- Ebbinghaus/Flum/Thomas:** Mathematical Logic. Second edition.
- Edgar:** Measure, Topology, and Fractal Geometry.
- Elaydi:** Introduction to Difference Equations.
- Fischer:** Intermediate Real Analysis.
- Flanigan/Kazdan:** Calculus Two: Linear and Nonlinear Functions. Second edition.
- Fleming:** Functions of Several Variables. Second edition.
- Foulds:** Combinatorial Optimization for Undergraduates.
- Foulds:** Optimization Techniques: An Introduction.
- Franklin:** Methods of Mathematical Economics.
- Hairer/Wanner:** Analysis by Its History.  
*Readings in Mathematics.*
- Halmos:** Finite-Dimensional Vector Spaces. Second edition.
- Halmos:** Naive Set Theory.
- Hämmerlin/Hoffmann:** Numerical Mathematics.  
*Readings in Mathematics.*
- Iooss/Joseph:** Elementary Stability and Bifurcation Theory. Second edition.
- Isaac:** The Pleasures of Probability.  
*Readings in Mathematics.*
- James:** Topological and Uniform Spaces.
- Jänich:** Linear Algebra.
- Jänich:** Topology.
- Kemeny/Snell:** Finite Markov Chains.
- Kinsey:** Topology of Surfaces.
- Klambauer:** Aspects of Calculus.
- Lang:** A First Course in Calculus. Fifth edition.
- Lang:** Calculus of Several Variables. Third edition.
- Lang:** Introduction to Linear Algebra.

(continued following index)

Andrew Browder

# Mathematical Analysis

*An Introduction*



Springer

**Andrew Browder**  
**Mathematics Department**  
**Brown University**  
**Providence, RI 02912**  
**USA**

***Editorial Board***

**S. Axler**  
**Department of**  
**Mathematics**  
**Michigan State University**  
**East Lansing, MI 48824**  
**USA**

**F.W. Gehring**  
**Department of**  
**Mathematics**  
**University of Michigan**  
**Ann Arbor, MI 48109**  
**USA**

**P.R. Halmos**  
**Department of**  
**Mathematics**  
**Santa Clara University**  
**Santa Clara, CA 95053**  
**USA**

With five illustrations.

---

Mathematics Subject Classifications (1991): 26-01, 26Axx, 54Cxx, 28Axx, 58Cxx

---

Library of Congress Cataloging-in-Publication Data  
Browder, Andrew.

Mathematical analysis : an introduction / Andrew Browder.

p. cm. – (Undergraduate texts in mathematics)

Includes bibliographical references and index.

ISBN 0-387-94614-4 (hardcover : alk. paper)

1. Mathematical analysis. I. Title. II. Series

QA300.B727 1996

515-dc20

95-44877

Printed on acid-free paper.

© 1996 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Robert Wexler; manufacturing supervised by Jacqui Ashri.

Photocomposed copy prepared from the author's  $\text{\LaTeX}$  files.

Printed and bound by R.R. Donnelley and Sons, Harrisonburg, VA.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-94614-4 Springer-Verlag New York Berlin Heidelberg



*For Anna*



# Preface

This is a textbook suitable for a year-long course in analysis at the advanced undergraduate or possibly beginning-graduate level. It is intended for students with a strong background in calculus and linear algebra, and a strong motivation to learn mathematics for its own sake. At this stage of their education, such students are generally given a course in abstract algebra, and a course in analysis, which give the fundamentals of these two areas, as mathematicians today conceive them.

Mathematics is now a subject splintered into many specialties and sub-specialties, but most of it can be placed roughly into three categories: algebra, geometry, and analysis. In fact, almost all mathematics done today is a mixture of algebra, geometry and analysis, and some of the most interesting results are obtained by the application of analysis to algebra, say, or geometry to analysis, in a fresh and surprising way. What then do these categories signify? Algebra is the mathematics that arises from the ancient experiences of addition and multiplication of whole numbers; it deals with the finite and discrete. Geometry is the mathematics that grows out of spatial experience; it is concerned with shape and form, and with measuring, where algebra deals with counting. Analysis might be described as the mathematics that deals with the ideas of the infinite and the infinitesimal; more specifically, it is the word used to describe the great web of ideas that has grown in the last three centuries from the discovery of the differential and integral calculus. Its basic arena is the system of real numbers, a mathematical construct which combines algebraic concepts of addition, multiplication, etc., with the geometric concept of a line, or continuum.

There is no general agreement on what an introductory analysis course

should include. I have chosen four major topics: the calculus of functions of one variable, treated with modern standards of rigor; an introduction to general topology, focusing on Euclidean space and spaces of functions; the general theory of integration, based on the concept of measure; and the calculus, differential and integral, for functions of several variables, with the inverse and implicit function theorems, and integration over manifolds. Inevitably, much time and effort go into giving definitions and proving technical propositions, building up the basic tools of analysis. I hope the reader will feel this machinery is justified by some of its products displayed here. The theorems of Dirichlet, Liouville, Weyl, Brouwer, and Riemann's Dirichlet principle for harmonic functions, for instance, should need no applications to be appreciated. (In fact, they have a great number of applications.)

An ideal book of mathematics might uphold the standard of economy of expression, but this one does not. The reader will find many repetitions here: where a result might have been proved once and subsequently referred to, I have on occasion simply given the old argument again. My justification is found in communications theory, which has shown mathematically that redundancy is the key to successful communication in a noisy channel. I have also on occasion given more than one proof for a single theorem; this is done not because two proofs are more convincing than one, but because the second proof involves different ideas, which may be useful in some new context.

I have included some brief notes, usually historical, at the end of each chapter. The history is all from secondary sources, and is not to be relied on too much, but it appears that many students find these indications of how things developed to be interesting. A student who wants to learn the material in the early chapters of this book from a historical perspective will find Bressoud's recent book [1] quite interesting.

While this book is meant to be used for a year-long course, I myself have never managed to include everything here in such a course. A year and a half might be reasonable, for students with no previous experience with rigorous analysis. In different years, I have omitted different topics, always regretfully. Every topic treated here meets one of two tests: it is either something that everybody should know, or else it is just too beautiful to leave out. Nevertheless, life is short, and the academic year even shorter, and anyone who teaches with this book should plan on leaving something out. I expect that most teachers who use this book might also be tempted to include some topic that I have not treated, or develop further some theme that is touched on lightly here.

Those students whose previous mathematical experience is mostly with calculus are in for some culture shock. They will notice that this book is only about 30% as large as their calculus text, and contains about one-tenth as many exercises. (So far, so good.) But it will quickly become clear that some of the problems are quite demanding; I believe that (certainly at this

level) more is learned by spending hours, if necessary, on a few problems, sometimes a single problem, than in routinely dispatching a dozen exercises, all following the same pattern. I hope the reader is not discouraged by difficulty, but rewarded by difficulty overcome.

This book consists of theorems, propositions, and lemmas (these words all mean the same thing), along with definitions and examples. Most of these are set off formally as Theorem, Proposition, etc., but some definitions, examples, and theorems are in fact sneaked into the text between the formally announced items. I have been persuaded to number the Theorems, Examples, Definitions, etc. by one sequence. Thus Example 4.4 refers to the fourth item in Chapter 4, where an item could be either a Theorem, Proposition, Lemma, Definition, or Example. It would have been more logical to have it refer to the fourth example in this chapter, but it would have made navigation more difficult.

One of the important things that one learns in a course at this level is how to write a mathematical proof. It is quite difficult to prescribe what constitutes a proper proof. It should be a clear and compelling argument, that forces a reader (who has accepted previous theorems and understands the hypotheses) to accept its assertions. It should be concise, but not cryptic; it should be detailed, but not verbose. We learn to do it by imitating models. Here are two models from ancient Greece. Throughout this book, the symbol ■ will mark the conclusion of a proof.

**Theorem.** *There are infinitely many primes.*

**Proof.** If  $p_1, p_2, \dots, p_n$  are primes, let  $N = p_1 p_2 \cdots p_n + 1$ . Then  $N$  is not divisible by any  $p_j$ ,  $j = 1, 2, \dots, n$ , so either  $N$  is a prime, or  $N$  is divisible by some prime other than  $p_1, p_2, \dots, p_n$ . In either case, there are at least  $n + 1$  primes. ■

Note that this proof assumes a knowledge of what a prime number is, and a previously obtained result that every integer  $N > 1$  is divisible by some prime. Note also that the last sentence of the proof, omitted above, might be either to the effect that the hypothesis that the set of all primes can be listed in a finite sequence  $p_1, p_2, \dots, p_n$  has led to a contradiction, or that we have given the recipe for finding a new prime for each natural number, so that the sequence 2, 3, 7, 43, ..., can be extended indefinitely.

**Theorem.** *The square on the hypotenuse of a right triangle is the sum of the squares on the two shorter sides.*

**Proof.** If the sides of the triangle are  $a$ ,  $b$ , and  $c$ , with  $c$  the hypotenuse, then the square of side  $a + b$  can be dissected in two ways, as shown below. Removing the four copies of the triangle present in each dissection, the theorem follows. ■

This is one of the rare occasions when I would accept a picture as a proof. (Having once been shown the proof, by a carefully drawn diagram,

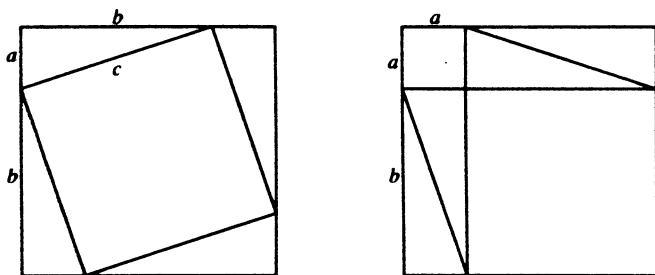


Figure 0.1. The Pythagorean Theorem.

that every triangle is isosceles, I could never again believe that pictures don't lie.) In this argument, the picture is clear and convincing. By the way, there are few pictures in what follows, and they are all simple and hand-drawn, meant to serve as a guide to simple ideas that are (unfortunately but necessarily) being expressed in awkward or complicated notation. I would urge the readers (of this or any other mathematics) to make their own sketches at all times, preferably crude and schematic.

**Acknowledgments.** It is impossible to list all the writers and individuals who have influenced me in the writing of this book, but for me the model of analysis textbooks at this level has always been Rudin's *Principles of Mathematical Analysis* [11]. The second half of this book (which was written first) was greatly influenced by Spivak's *Calculus on Manifolds* [13], the first clear and simple introduction to Stokes' theorem in its modern form. I was fortunate to read in manuscript Munkres' excellent book *Analysis on Manifolds* [10] while I was writing that earlier version, and profited from it.

Many students found typographical errors and other infelicities in the class notes which formed the first version of the first half of the book, and I want to thank especially Andrew Brecher, Greg Friedman, Ezra Miller, and Max Minzner for their detailed examination, and their suggestions for that part. Eva Kallin made many useful criticisms of an earlier version of the notes on which the second half of this book is based. I am grateful to Xiang-Qian Chang, who read the entire manuscript, and pointed out a great number of rough places.

The mistakes that remain are all my own. I would appreciate hearing about them.

# Contents

|  |            |
|--|------------|
| <b>Preface</b>                                     | <b>vii</b> |
| <b>1 Real Numbers</b>                              | <b>1</b>   |
| 1.1 Sets, Relations, Functions . . . . .           | 1          |
| 1.2 Numbers . . . . .                              | 4          |
| 1.3 Infinite Sets . . . . .                        | 6          |
| 1.4 Incommensurability . . . . .                   | 8          |
| 1.5 Ordered Fields . . . . .                       | 10         |
| 1.6 Functions on $\mathbf{R}$ . . . . .            | 16         |
| 1.7 Intervals in $\mathbf{R}$ . . . . .            | 18         |
| 1.8 Algebraic and Transcendental Numbers . . . . . | 20         |
| 1.9 Existence of $\mathbf{R}$ . . . . .            | 21         |
| 1.10 Exercises . . . . .                           | 23         |
| 1.11 Notes . . . . .                               | 25         |
| <b>2 Sequences and Series</b>                      | <b>28</b>  |
| 2.1 Sequences . . . . .                            | 28         |
| 2.2 Continued Fractions . . . . .                  | 36         |
| 2.3 Infinite Series . . . . .                      | 39         |
| 2.4 Rearrangements of Series . . . . .             | 45         |
| 2.5 Unordered Series . . . . .                     | 47         |
| 2.6 Exercises . . . . .                            | 50         |
| 2.7 Notes . . . . .                                | 53         |

|          |  |            |
|----------|--|------------|
| <b>3</b> | <b>Continuous Functions on Intervals</b>                   | <b>55</b>  |
| 3.1      | Limits and Continuity . . . . .                            | 55         |
| 3.2      | Two Fundamental Theorems . . . . .                         | 59         |
| 3.3      | Uniform Continuity . . . . .                               | 61         |
| 3.4      | Sequences of Functions . . . . .                           | 62         |
| 3.5      | The Exponential Function . . . . .                         | 65         |
| 3.6      | Trigonometric Functions . . . . .                          | 66         |
| 3.7      | Exercises . . . . .  | 69         |
| 3.8      | Notes . . . . .  | 72         |
| <b>4</b> | <b>Differentiation</b>                                     | <b>74</b>  |
| 4.1      | Derivatives . . . . .                                      | 74         |
| 4.2      | Derivatives of Some Elementary Functions . . . . .         | 76         |
| 4.3      | Convex Functions . . . . .                                 | 78         |
| 4.4      | The Differential Calculus . . . . .                        | 81         |
| 4.5      | L'Hospital's Rule . . . . .                                | 86         |
| 4.6      | Higher Order Derivatives . . . . .                         | 88         |
| 4.7      | Analytic Functions . . . . .                               | 90         |
| 4.8      | Exercises . . . . .  | 93         |
| 4.9      | Notes . . . . .  | 95         |
| <b>5</b> | <b>The Riemann Integral</b>                                | <b>98</b>  |
| 5.1      | Riemann Sums . . . . .                                     | 98         |
| 5.2      | Existence Results . . . . .                                | 102        |
| 5.3      | Properties of the Integral . . . . .                       | 107        |
| 5.4      | Fundamental Theorems of Calculus . . . . .                 | 110        |
| 5.5      | Integrating Sequences and Series . . . . .                 | 113        |
| 5.6      | Improper Integrals . . . . .                               | 114        |
| 5.7      | Exercises . . . . .  | 118        |
| 5.8      | Notes . . . . .  | 121        |
| <b>6</b> | <b>Topology</b>  | <b>123</b> |
| 6.1      | Topological Spaces . . . . .                               | 123        |
| 6.2      | Continuous Mappings . . . . .                              | 126        |
| 6.3      | Metric Spaces . . . . .                                    | 127        |
| 6.4      | Constructing Topological Spaces . . . . .                  | 131        |
| 6.5      | Sequences . . . . .  | 135        |
| 6.6      | Compactness . . . . .                                      | 140        |
| 6.7      | Connectedness . . . . .                                    | 147        |
| 6.8      | Exercises . . . . .  | 150        |
| 6.9      | Notes . . . . .  | 153        |
| <b>7</b> | <b>Function Spaces</b>                                     | <b>155</b> |
| 7.1      | The Weierstrass Polynomial Approximation Theorem . . . . . | 155        |
| 7.2      | Lengths of Paths . . . . .                                 | 159        |



|           |   |            |
|-----------|---|------------|
| 7.3       | Fourier Series . . . . .                  | 161        |
| 7.4       | Weyl's Theorem . . . . .                  | 170        |
| 7.5       | Exercises . . . . .                       | 171        |
| 7.6       | Notes . . . . .                           | 173        |
| <b>8</b>  | <b>Differentiable Maps</b>                | <b>175</b> |
| 8.1       | Linear Algebra . . . . .                  | 176        |
| 8.2       | Differentials . . . . .                   | 182        |
| 8.3       | The Mean Value Theorem . . . . .          | 185        |
| 8.4       | Partial Derivatives . . . . .             | 186        |
| 8.5       | Inverse and Implicit Functions . . . . .  | 191        |
| 8.6       | Exercises . . . . .                       | 196        |
| 8.7       | Notes . . . . .                           | 199        |
| <b>9</b>  | <b>Measures</b>                           | <b>201</b> |
| 9.1       | Additive Set Functions . . . . .          | 202        |
| 9.2       | Countable Additivity . . . . .            | 204        |
| 9.3       | Outer Measures . . . . .                  | 208        |
| 9.4       | Constructing Measures . . . . .           | 211        |
| 9.5       | Metric Outer Measures . . . . .           | 213        |
| 9.6       | Measurable Sets . . . . .                 | 215        |
| 9.7       | Exercises . . . . .                       | 219        |
| 9.8       | Notes . . . . .                           | 221        |
| <b>10</b> | <b>Integration</b>                        | <b>223</b> |
| 10.1      | Measurable Functions . . . . .            | 223        |
| 10.2      | Integration . . . . .                     | 226        |
| 10.3      | Lebesgue and Riemann Integrals . . . . .  | 231        |
| 10.4      | Inequalities for Integrals . . . . .      | 233        |
| 10.5      | Uniqueness Theorems . . . . .             | 237        |
| 10.6      | Linear Transformations . . . . .          | 240        |
| 10.7      | Smooth Transformations . . . . .          | 241        |
| 10.8      | Multiple and Repeated Integrals . . . . . | 244        |
| 10.9      | Exercises . . . . .                       | 247        |
| 10.10     | Notes . . . . .                           | 251        |
| <b>11</b> | <b>Manifolds</b>                          | <b>253</b> |
| 11.1      | Definitions . . . . .                     | 253        |
| 11.2      | Constructing Manifolds . . . . .          | 258        |
| 11.3      | Tangent Spaces . . . . .                  | 260        |
| 11.4      | Orientation . . . . .                     | 262        |
| 11.5      | Exercises . . . . .                       | 265        |
| 11.6      | Notes . . . . .                           | 267        |
| <b>12</b> | <b>Multilinear Algebra</b>                | <b>269</b> |

|   |            |
|---|------------|
| 12.1 Vectors and Tensors . . . . .                    | 269        |
| 12.2 Alternating Tensors . . . . .                    | 272        |
| 12.3 The Exterior Product . . . . .                   | 277        |
| 12.4 Change of Coordinates . . . . .                  | 280        |
| 12.5 Exercises . . . . .                              | 282        |
| 12.6 Notes . . . . .                                  | 283        |
| <b>13 Differential Forms</b>                          | <b>285</b> |
| 13.1 Tensor Fields . . . . .                          | 285        |
| 13.2 The Calculus of Forms . . . . .                  | 286        |
| 13.3 Forms and Vector Fields . . . . .                | 288        |
| 13.4 Induced Mappings . . . . .                       | 290        |
| 13.5 Closed and Exact Forms . . . . .                 | 291        |
| 13.6 Tensor Fields on Manifolds . . . . .             | 293        |
| 13.7 Integration of Forms in $\mathbf{R}^n$ . . . . . | 294        |
| 13.8 Exercises . . . . .                              | 295        |
| 13.9 Notes . . . . .                                  | 296        |
| <b>14 Integration on Manifolds</b>                    | <b>297</b> |
| 14.1 Partitions of Unity . . . . .                    | 297        |
| 14.2 Integrating $k$ -Forms . . . . .                 | 300        |
| 14.3 The Brouwer Fixed Point Theorem . . . . .        | 305        |
| 14.4 Integrating Functions on a Manifold . . . . .    | 307        |
| 14.5 Vector Analysis . . . . .                        | 312        |
| 14.6 Harmonic Functions . . . . .                     | 314        |
| 14.7 Exercises . . . . .                              | 318        |
| 14.8 Notes . . . . .                                  | 321        |
| <b>References</b>                                     | <b>323</b> |
| <b>Index</b>  | <b>325</b> |

# Real Numbers

In this chapter, we describe the system of real numbers, deducing some of their essential properties from the axioms for a complete ordered field. Before doing so, we take a quick look at the ideas and notations of sets, relations, and functions, sketch the construction of the integers and the rational numbers (starting from the natural numbers), and indicate the need for a field larger than the rational numbers. At the end of the chapter, we sketch the proof of the existence and (essential) uniqueness of a complete ordered field.

## 1.1 Sets, Relations, Functions

We assume no knowledge of formal set theory, but do assume familiarity with the basic notations and elementary calculations. Thus,  $x \in A$  means that  $x$  is an element of  $A$ ,  $A \subset B$  (also written  $B \supset A$ ) means that  $x \in B$  whenever  $x \in A$ . The sets  $A$  and  $B$  are equal if and only if  $A \subset B$  and  $B \subset A$ . For any sets  $A, B$ , the intersection  $A \cap B$  and the union  $A \cup B$  are given by

$$A \cap B = \{x : x \in A \text{ and } x \in B\}, \quad A \cup B = \{x : x \in A \text{ or } x \in B\},$$

where the “or” above is the nonexclusive “or,” that is, is understood to mean “and/or.” The *relative complement*  $A \setminus B$  is defined by

$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}.$$

Here,  $x \notin B$  means that  $x$  is not an element of  $B$ ; similar notations, such as  $x \neq y$  or  $x \not\leq y$ , will be used later without comment. More generally, if  $\mathcal{A}$  is a collection of sets, i.e., a set whose elements are sets, we write

$$\bigcap_{A \in \mathcal{A}} A = \{x : x \in A \text{ for every } A \in \mathcal{A}\},$$

$$\bigcup_{A \in \mathcal{A}} A = \{x : x \in A \text{ for some } A \in \mathcal{A}\}.$$

The *empty set*  $\emptyset$  has no elements; it enjoys the property  $\emptyset \subset A$  for every set  $A$ . When all the sets currently under consideration are subsets of some set  $X$  (which will always be the case in what follows), we often write  $A^C$  to stand for  $X \setminus A$ , and call it the *complement* of  $A$ . We will often make use of the formulas

$$\left(\bigcup_{\alpha \in A} E_\alpha\right)^C = \bigcap_{\alpha \in A} E_\alpha^C, \quad \left(\bigcap_{\alpha \in A} E_\alpha\right)^C = \bigcup_{\alpha \in A} E_\alpha^C,$$

known as DeMorgan's laws.

If  $X$  and  $Y$  are sets, their *Cartesian product*  $X \times Y$  is the set of all ordered pairs  $(x, y)$  with  $x \in X$  and  $y \in Y$ . The set of all subsets of a set  $X$  is called the *power set* of  $X$ , and denoted by  $\mathcal{P}(X)$ . We observe that  $\mathcal{P}(\emptyset)$  is not empty; it has exactly one element,  $\emptyset$ , and has two distinct subsets, namely,  $\emptyset$  and  $\{\emptyset\}$ .

While the reader is undoubtedly familiar with the real and complex numbers, in this book we assume only the existence, and the familiar properties, of the set  $\mathbf{N} = \{1, 2, \dots\}$  of natural numbers. We will show, at least in outline, how the more complicated number systems arise from  $\mathbf{N}$ .

Given sets  $X$  and  $Y$ , a *relation* from  $X$  to  $Y$  is a subset  $R$  of  $X \times Y$ . We say  $R$  is a relation on  $X$  if  $Y = X$ . We write  $xRy$  if  $(x, y) \in R$ . Three kinds of relations have frequent applications.

1. A relation  $\sim$  on  $X$  is called an *equivalence relation* if it satisfies the three conditions:

- (a) for all  $x \in X$ ,  $x \sim x$ ;
- (b) if  $x \sim y$ , then  $y \sim x$ ; and
- (c) if  $x \sim y$  and  $y \sim z$ , then  $x \sim z$ .

These properties are called *reflexivity*, *symmetry*, and *transitivity*, in that order. The simplest example is of course the relation of equality. If  $\sim$  is an equivalence relation on  $X$ , it partitions  $X$  into equivalence classes: for each  $x \in X$ , let  $C_x = \{y \in X : y \sim x\}$ . We call  $C_x$  the equivalence class of  $x$ . Since  $x \in C_x$  by (a), we see that  $X$  is the union of all the equivalence classes. If there exists  $z \in C_x \cap C_y$ , then  $z \sim x$  and  $z \sim y$ ; since this implies  $y \sim z$  by (b) and hence  $y \sim x$

by (c), we obtain  $y \in C_x$ . Another application of (c) shows that if  $u \sim y$ , then  $u \sim x$ , i.e., that  $C_y \subset C_x$ . In the same way,  $C_x \subset C_y$ , so  $C_x = C_y$ . We summarize: either  $C_x$  and  $C_y$  are disjoint, or  $C_x = C_y$ . Thus the distinct equivalence classes form a disjoint family of sets whose union is  $X$ . It is common practice to denote the equivalence class of an element  $x$  by  $[x]$ , rather than  $C_x$ .

2. A relation  $<$  on  $X$  is called a *partial order* if it satisfies two conditions:

- (a) it is transitive, i.e.,  $x < y$  and  $y < z$  implies  $x < z$ ; and
- (b) for all  $x \in X$ ,  $x \not< x$ .

When (a) holds, it is easily seen that (b) is equivalent to *antisymmetry*: if  $x < y$ , then  $y \not< x$ . A relation  $<$  on  $X$  is called a *total order* if it is a partial order, and satisfies the further condition: for any  $x$  and  $y$  in  $X$ , either  $x < y$  or  $y < x$  or  $x = y$  (trichotomy). In view of (a) and (b), these alternatives are mutually exclusive, so exactly one of them holds. An example of a partial order is the relation of proper inclusion on any family of sets. When  $<$  is a partial order on  $X$ , we write  $x \leq y$  to mean that either  $x < y$  or  $x = y$ . Some writers define a partial order to be a relation  $\leq$  which is transitive and has the property that  $x \leq y$  and  $y \leq x$  together imply that  $x = y$ ; given such a relation, one can define  $x < y$  to mean that  $x \leq y$  and  $x \neq y$ , and find that  $<$  is a partial order in our sense.

3. A relation  $f$  from  $X$  to  $Y$  is called a *function* if for each  $x \in X$  there is exactly one  $y \in Y$  such that  $xfy$ . We write  $y = f(x)$ , or sometimes  $y = f_x$ , to mean  $xfy$ . This definition of a function identifies a function with its graph, and so furthers the program, popular among mathematicians in this century, that all mathematical objects should be sets. In practice, of course, almost everybody thinks of a function from  $X$  to  $Y$  as a rule  $f$  which assigns to each  $x \in X$  an element  $f(x)$  of  $Y$ . One never uses the relation notation  $xfy$  in respectable society when  $f$  is a function. We write  $f : X \rightarrow Y$  to mean that  $f$  is a function from  $X$  to  $Y$ ; here,  $X$  is called the *domain* of  $f$ , and  $\{f(x) : x \in X\}$  is called the *range* of  $f$ , or the *image* of  $f$ . We note that the range of  $f$  is to be distinguished from the target space  $Y$  of  $f$ . We say that  $f$  is *surjective*, or *onto*, if the image of  $f$  is all of  $Y$ , and we say that  $f$  is *injective*, or *one-one*, if  $f(x) = f(y)$  only when  $x = y$ . A function which is both surjective and injective is called *bijective*. A bijective function is also referred to as a *one-to-one correspondence*. The words *map* and *mapping* are synonyms for function. We often use the notation  $f : x \mapsto f(x)$ . For instance,  $f : n \mapsto n^2$  is another way of saying that  $f$  is the function for which  $f(n) = n^2$ ; the notation here suggests that the domain of  $f$  is either the set of integers, or some subset of that set. A function  $x : \mathbb{N} \rightarrow X$

is also called a *sequence* in  $X$ ; here, it is standard practice to use the notation  $x_n$  instead of  $x(n)$ . If  $E : A \rightarrow \mathcal{P}(X)$  is a function, we speak of an indexed family of subsets of  $X$ ; we usually write  $\alpha \mapsto E_\alpha$  in this situation. When  $A = \mathbf{N}$ , this means a sequence of subsets of  $X$ . The symbols  $\bigcup_{\alpha \in A} E_\alpha$  and  $\bigcap_{\alpha \in A} E_\alpha$  have the obvious meanings; when  $A = \mathbf{N}$ , it is customary to write  $\bigcup_{n=1}^{\infty} E_n$  and  $\bigcap_{n=1}^{\infty} E_n$  instead.

## 1.2 Numbers

We are familiar with the set of natural numbers  $\mathbf{N}$ . This set comes with an algebraic structure and an order relation. That is, there are two operations that assign a natural number to any given pair of natural numbers (i.e., maps of  $\mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$ ), called addition and multiplication, and they have familiar formal properties such as  $m + n = n + m$  for every  $m, n \in \mathbf{N}$ , etc. There is also an order relation on  $\mathbf{N}$ : we say  $m < n$  if there exists  $k \in \mathbf{N}$  such that  $m + k = n$ . We emphasize the following property of this order relation:

*if  $A$  is a nonempty subset of  $\mathbf{N}$ , then  $A$  has a least element; i.e., there exists  $a \in A$  such that  $a \leq n$  for all  $n \in A$ .*

This fact, described technically by the phrase “ $\mathbf{N}$  is well-ordered,” and which is easily seen to imply that  $\mathbf{N}$  is totally ordered by  $<$ , is easily seen to be equivalent to the following, known as the *principle of finite induction*:

*Let  $A$  be a subset of  $\mathbf{N}$ , satisfying the two conditions: (a)  $1 \in A$ ; and (b) if  $n \in A$ , then  $n + 1 \in A$ . Then  $A = \mathbf{N}$ .*

A variant of this is equivalent, and sometimes convenient:

*Let  $A$  be a subset of  $\mathbf{N}$ , satisfying the two conditions: (a)  $1 \in A$ ; and (b') if  $k \in A$  for all  $k \in \mathbf{N}$  with  $k < n$ , then  $n \in A$ . Then  $A = \mathbf{N}$ .*

The principle of induction is essential in proving many theorems, and we assume at least some familiarity with this procedure.

The set  $\mathbf{Z}$  of integers is sometimes obtained from  $\mathbf{N}$  by the following procedure. Let  $\mathcal{Z}$  be the set of all ordered pairs  $(m, n)$  of natural numbers, i.e., let  $\mathcal{Z} = \mathbf{N} \times \mathbf{N}$ . Define a relation  $\sim$  on  $\mathcal{Z}$  by declaring  $(m, n) \sim (j, k)$  if and only if  $m + k = n + j$ . This is easily seen to be an equivalence relation. We define  $[m, n]$  to be the equivalence class of  $(m, n)$  for each  $(m, n) \in \mathcal{Z}$ , and denote by  $\mathbf{Z}$  the set of all such equivalence classes. We give the name *integers* to the elements of  $\mathbf{Z}$ . We define addition in  $\mathbf{Z}$  by  $[m, n] + [j, k] = [m + j, n + k]$ ; to see that this makes sense, it must be verified that if  $(m, n) \sim (m', n')$  and  $(j, k) \sim (j', k')$ , then  $(m + j, n + k) \sim (m' + j', n' + k')$ , but this is very easy. We define multiplication in  $\mathbf{Z}$  by

the rule:  $[m, n][j, k] = [mj + nk, nj + mk]$ . We define an order relation in  $\mathbf{Z}$  by  $[m, n] < [j, k]$  if and only if  $m + k < j + n$ . (We use the same symbol here for the order relation in  $\mathbf{N}$  and the order relation in  $\mathbf{Z}$ , even though this is not really correct; correctness can carry a high price in notation, and finally even in comprehension.) Again, for the multiplication and for the order relation to be well-defined, it must be verified that the definitions yield the same result independent of the choice of representative for the equivalence class. For any  $m \in \mathbf{N}$ ,  $(m, m) \sim (1, 1)$ ; we denote  $[1, 1]$  by 0. We observe that for any  $m, n \in \mathbf{N}$ ,  $[m, n] + [n, m] = [m + n, n + m] = [1, 1] = 0$ . If  $m > n$  there is some  $k \in \mathbf{N}$  with  $m = k + n$ , and we see that  $(m, n) \sim (k + 1, 1)$ ; similarly, if  $n > m$ , there is some  $k \in \mathbf{N}$  such that  $(m, n) \sim (1, k + 1)$ . The map  $\phi : k \mapsto [k + 1, 1]$  of  $\mathbf{N} \rightarrow \mathbf{Z}$  is injective, and preserves all the structure of  $\mathbf{N}$ . By this we mean that for all  $j, k \in \mathbf{N}$ , we have  $\phi(j + k) = \phi(j) + \phi(k)$ ,  $\phi(jk) = \phi(j)\phi(k)$ , and that if  $j < k$ , then  $\phi(j) < \phi(k)$ . In other words,  $\mathbf{N}' = \{[k + 1, 1] : k \in \mathbf{N}\}$  is an exact replica of  $\mathbf{N}$ ; we have  $\mathbf{N}' = \{z \in \mathbf{Z} : z > 0\}$ , the set of all positive integers. We say that the map  $\phi$  is an *isomorphism* of  $\mathbf{N}$  onto  $\mathbf{N}'$ . Henceforth, we shall identify  $\mathbf{N}$  with  $\mathbf{N}'$ , i.e., regard  $\mathbf{N}$  as a subset of  $\mathbf{Z}$ . We write  $k$  to mean  $[k + 1, 1]$ , and  $-k$  for  $[1, k + 1]$ .

This somewhat elaborate construction of  $\mathbf{Z}$  was geared toward the result: for any  $a, b \in \mathbf{Z}$ , there exists a unique  $c \in \mathbf{Z}$  such that  $a = b + c$ . An analogous path lets us construct the rational numbers. We want to expand the integers to a system which not only has addition, subtraction, and multiplication, but division as well, i.e., where the equation  $ax = b$  has a solution  $x$  for any given  $a$  and  $b$ , provided  $a \neq 0$ . We let  $\mathbf{Q}$  denote the set of all ordered pairs  $(m, n)$  of integers such that  $n \neq 0$ , i.e.,  $\mathbf{Q} = \mathbf{Z} \times (\mathbf{Z} \setminus \{0\})$ . We define a relation  $\sim$  on  $\mathbf{Q}$  by the rule  $(m, n) \sim (j, k)$  means  $mk = nj$ . We verify this is an equivalence relation, and let  $\mathbf{Q}$  be the set of equivalence classes. We call  $\mathbf{Q}$  the set of rational numbers. We write  $m/n$  or  $\frac{m}{n}$  to denote the equivalence class of  $(m, n)$ . We make the definitions:  $(m/n)(j/k) = (mj/nk)$ ,  $(m/n) + (j/k) = (mk + nj)/nk$ , and  $(m/n) < (j/k)$  if and only if  $mk < nj$  when  $n$  and  $k$  are positive. We can verify that these definitions make sense: they are independent of the choice of  $(m, n)$  in the equivalence class  $m/n$ , etc., and for the order relation, we note we can always choose  $n > 0$  in writing an element of  $\mathbf{Q}$  as  $m/n$ . Again, we observe that the map  $m \mapsto m/1$  is an injective mapping of  $\mathbf{Z}$  into  $\mathbf{Q}$ , which preserves all the structure of  $\mathbf{Z}$ . We henceforth identify  $\mathbf{Z}$  with the subset  $\{(m, 1) : m \in \mathbf{Z}\}$  of  $\mathbf{Q}$ . The construction of  $\mathbf{Q}$  was engineered in view of the goal: for any  $a, b \in \mathbf{Q}$ , with  $a \neq 0$ , there exists a unique  $c \in \mathbf{Q}$  with  $ac = b$ .

It is frequently useful to use the "best" representation of a rational number, i.e., to write it as a quotient of integers without common factor. We illustrate induction arguments by proving this can be done. Given  $q \in \mathbf{Q}$ , consider

$$\{k \in \mathbf{N} : q = m/k \text{ for some } m \in \mathbf{Z}\} = \{k \in \mathbf{N} : kq \in \mathbf{Z}\}.$$

This is a nonempty subset of  $\mathbf{N}$ , hence contains a smallest element  $n$ . In the representation  $q = m/n$ ,  $m$  and  $n$  have no common factor greater than 1; for if  $m = ij$  and  $n = kj$ , with  $j \in \mathbf{N}$ ,  $j > 1$ , we would have  $q = i/k$ , though  $k < n$ , which is impossible. If also  $m \in \mathbf{N}$  and  $mq \in \mathbf{Z}$ , it is easy to see that  $m$  is an integer multiple of  $n$ .

## 1.3 Infinite Sets

A set  $F$  is said to be *finite* if for some  $n \in \mathbf{N}$  there exists a bijective mapping  $\phi$  from  $\{1, 2, \dots, n\}$  to  $F$ . The natural number  $n$  here is uniquely determined, and we call it the *cardinality* of  $F$ , written either as  $\text{card } F$  or as  $\#F$ . If  $\phi: j \mapsto x_j$ , we have  $F = \{x_1, x_2, \dots, x_n\}$ . We also call the empty set finite, and assign it cardinality 0. If a set is not finite, it is *infinite*. If  $X$  is an infinite set, there exists an injective mapping of the natural numbers  $\mathbf{N}$  into  $X$ . If there exists a bijective map of  $\mathbf{N}$  onto  $X$ , we say that  $X$  is *countably infinite*. Thus,  $X$  is countably infinite if and only if its elements can be listed in an infinite sequence:  $X = \{x_1, x_2, \dots\}$ . We call a set  $E$  *countable* if it is either finite or countably infinite; if  $X$  is infinite but not countable, we say  $X$  is *uncountable*. If  $X$  is countably infinite, we also write  $\text{card } X = \aleph_0$ , pronounced "aleph naught." We give just a few results about countability.

**1.1 Proposition.** *Every subset of  $\mathbf{N}$  is countable.*

**Proof.** Suppose  $A$  is an infinite subset of  $\mathbf{N}$ . Define  $\phi(1)$  to be the smallest element of  $A$ ; since  $A$  is infinite, it is nonempty, so  $\phi(1)$  is well-defined by the fundamental property of  $\mathbf{N}$ . Having defined  $\phi(n-1)$ , set  $\phi(n)$  to be the smallest element of  $\{k \in A : k > \phi(n-1)\}$ . Again, the set in brackets is nonempty since  $A$  is infinite, so  $\phi(n)$  is well-defined. In this way, we construct a map  $\phi: \mathbf{N} \rightarrow A$ , and it is easy to check that this map is bijective. ■

**1.2 Corollary.** *Every subset of a countable set is countable.*

**1.3 Proposition.** *A set  $A$  is countable if and only if there exists an injective map of  $A$  into  $\mathbf{N}$ , if and only if there exists a surjective map of  $\mathbf{N}$  onto  $A$ .*

**Proof.** It is trivial that if  $A$  is finite, there exists an injective map of  $A$  into  $\mathbf{N}$  and a surjective map of  $\mathbf{N}$  onto  $A$ . Suppose that  $A$  is infinite. If  $A$  is countable, there exists a bijective map  $g$  of  $\mathbf{N}$  onto  $A$ , so  $f = g^{-1}$  is an injective map of  $A$  onto  $\mathbf{N}$ . If there exists an injective map  $f$  of  $A$  into  $\mathbf{N}$ , then by the last proposition there exists a bijective map  $\phi$  of  $f(A)$  onto  $\mathbf{N}$ , and then  $\phi \circ f$  is a bijective map of  $A$  onto  $\mathbf{N}$ , so  $A$  is countable. If



$g : \mathbf{N} \rightarrow A$  is surjective, define  $f : A \rightarrow \mathbf{N}$  by taking  $f(a)$  to be the smallest element of  $\{n \in \mathbf{N} : g(n) = a\}$ . It is easy to see that  $f$  is injective, so  $A$  is countable. ■

**1.4 Proposition.** *If  $A$  and  $B$  are countable, then so is  $A \times B$ .*

**Proof.** It suffices to prove this proposition for the special case  $A = B = \mathbf{N}$ . By Proposition 1.3, it suffices to produce an injective map  $f : \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$ . Let  $f$  be defined by  $f(m, n) = n + 2^{n+m}$ . Then  $f$  is injective; for if  $f(m, n) = f(j, k)$  with  $n \geq k$ , then we have

$$0 \leq n - k = 2^{k+j} - 2^{n+m} < n,$$

and since  $n < 2^n < 2^{n+m}$ , we get from the above that

$$2^{n+m} \leq 2^{k+j} < n + 2^{n+m} < 2^{n+m+1}.$$

It follows that  $n + m \leq k + j < n + m + 1$ , so  $n + m = k + j$ , and hence  $n - k = 2^{k+j} - 2^{n+m} = 0$ . Thus  $n = k$ , and  $m = j$ . ■

The more usual way to establish the last proposition is to list the elements of  $\mathbf{N} \times \mathbf{N}$  by listing in order the elements of the finite sets  $\{(j, k) : j + k = n\}$ , thus

$$(1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1), (1, 4), (2, 3), (3, 2), (4, 1), \dots,$$

which gives a bijective map of  $\mathbf{N}^2$  to  $\mathbf{N}$  directly.

**1.5 Corollary.** *The set  $\mathbf{Q}$  of all rational numbers is countable.*

**Proof.** The set  $\mathbf{Q}$  is countable by Proposition 1.4 and Corollary 1.2, and it follows from Proposition 1.3 that  $\mathbf{Q}$  is countable. ■

**1.6 Proposition.** *If  $A_n$  is a countable set for each  $n \in \mathbf{N}$ , then  $A = \bigcup_{n=1}^{\infty} A_n$  is countable.*

**Proof.** For each  $n \in \mathbf{N}$  there exists a surjective map  $\phi_n : \mathbf{N} \rightarrow A_n$ . Define  $f : \mathbf{N}^2 \rightarrow A$  by  $f(n, m) = \phi_n(m)$ . Then  $f$  is surjective, and since  $\mathbf{N}^2$  is countable by Proposition 1.4, it follows that  $A$  is countable by Proposition 1.3. ■

**1.7 Theorem.** *For any set  $A$ , there is no surjective mapping of  $A$  onto  $\mathcal{P}(A)$ . In particular, if  $A$  is a countably infinite set, then  $\mathcal{P}(A)$  is uncountable.*

**Proof.** Suppose that  $\phi : A \rightarrow \mathcal{P}(A)$ . Let  $B = \{a \in A : a \notin \phi(a)\}$ . If  $B = \phi(a)$  for some  $a \in A$ , then  $a \in B$  is impossible, since this would say  $a \in \phi(a)$ , so  $a \notin B$ . But also  $a \notin B$  is impossible, since this would say  $a \notin \phi(a)$ , so  $a \in B$ . The only conclusion possible is that  $B \neq \phi(a)$  for every  $a \in A$ , so  $\phi$  is not surjective. ■

**1.8 Corollary.** *If  $X$  is the set of all mappings of  $\mathbf{N}$  into  $\{0, 1\}$ , i.e., all sequences of zeros and ones, then  $X$  is uncountable.*

**Proof.** For each  $A \subset \mathbf{N}$ , let  $1_A : \mathbf{N} \rightarrow \{0, 1\}$  be defined by  $1_A(n) = 1$  if  $n \in A$ , and  $1_A(n) = 0$  if  $n \notin A$ . It is easy to check that the map  $A \mapsto 1_A$  is a bijective map of  $\mathcal{P}(\mathbf{N})$  onto  $X$ , so it follows from the last theorem that  $X$  is uncountable. ■

## 1.4 Incommensurability

The natural numbers are also called counting numbers; but from the earliest historic times, they were used for measuring as well, for instance, for measuring the length of a line segment. This application of course depends on choosing a unit of length, and then marking off on the given line segment a sequence of subsegments of this unit length, these subsegments to have only endpoints in common, while the segment to be measured is to be the union of the subsegments. If this is impossible, i.e., if the last subsegment marked off extends beyond the endpoint of the original segment, then hopefully a smaller choice of unit can rectify the situation. However, it was discovered in classical Greece (ca. 400 B.C.) that there is no choice of unit which makes the side and diagonal of a square simultaneously have integral lengths: these quantities are *incommensurable*. The classical proof cannot be improved:

Let  $s$  be the side, and let  $d$  be the diagonal of a square. Then  $d^2 = 2s^2$  by the Pythagorean theorem. If  $s$  and  $d$  are integers, then we may suppose them to have no common factor (in particular, we may suppose them to be not both even). But  $d^2 = 2s^2$  tells us that  $d^2$  is even, and it follows that  $d$  is even (it is easy to check that the square of an odd number is odd), so  $d = 2m$  for some integer  $m$ . But then  $2s^2 = d^2 = 4m^2$ , whence  $s^2 = 2m^2$ , so  $s^2$  is even, and hence  $s$  is even. This contradiction shows that the hypothesis that  $d$  and  $s$  are integers is untenable. ■

This proof is remarkable in that the geometric hypothesis is reduced at once to an algebraic one ( $d^2 = 2s^2$ ), and the rest of the argument is strictly algebraic, or number-theoretic if you will: it is shown that the equation  $x^2 = 2$  has no solution  $x$  in the rational numbers. Some readers may prefer an argument whose methods are as geometric as possible. Here is such a geometric proof:

Suppose that the side and diagonal of the square  $ABCD$  are commensurable, so that for some choice of unit length the side  $AB$  has length  $s$  and the diagonal  $AC$  has length  $d$ , where  $s$  and  $d$  are integers. We shall produce another square, with strictly

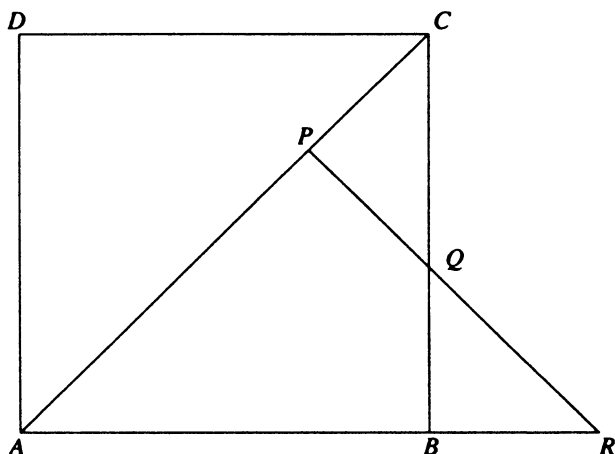


Figure 1.1. Geometric proof of incommensurability.

smaller sides, whose side and diagonal also are integral. On the diagonal  $AC$  of the square  $ABCD$ , mark off the point  $P$  which is at distance  $s$  from  $A$ . Draw the line through  $P$  perpendicular to  $AC$ , let  $Q$  be the intersection of this line with the side  $BC$ , and let  $R$  be the intersection of this line with the line through  $A$  and  $B$ . (See Figure 1.) We observe that the isosceles right triangle  $\triangle APR$  is congruent to  $\triangle ABC$ , since each has sides of length  $s$ . Hence the hypotenuse  $AR$  of  $\triangle ARP$  has length  $d$ , and hence  $BR = PC = d - s$ . (We note that  $d - s < s$ , i.e.,  $d < 2s$ , since the line segment  $AC$  joining the points  $A$  and  $C$  is shorter than the path made up of  $AB$  followed by  $BC$ .) But  $BQ = BR$ , so  $CQ$  has length  $s - (d - s) = 2s - d$ , an integer. Thus the isosceles right triangle  $\triangle CPQ$  has a hypotenuse of integral length, as well as sides which are of integral length  $d - s$ , strictly smaller than those of  $\triangle ABC$ . Completing it to a square, we have verified the claim made above. A repetition of this construction, fewer than  $s$  times, leads to a contradiction, since  $s \leq 1$  is obviously impossible. ■

The geometric proof above leads us to a different algebraic proof, perhaps the shortest possible. We can argue as follows: if  $s$  and  $d$  are positive integers with  $d^2 = 2s^2$ , and  $s$  is the smallest possible, we consider  $m = d - s$  and  $n = 2d - s$ . Since  $s < d < 2s$  is clear, we have  $0 < m < s$ , and a quick computation gives

$$n^2 = (2s - d)^2 = 4s^2 - 4sd + d^2 = 2d^2 - 4sd + 2s^2 = 2(d - s)^2 = 2m^2,$$

contradicting the minimality of  $s$ . This proof is indeed efficient, but rather uncivil.

The discovery of incommensurable quantities was a severe blow to the Pythagorean program of understanding nature by means of numbers. (The slogan of the Pythagoreans was "All is number.") The Greeks developed a sophisticated theory of ratios, presumably the work of Eudoxus, to work around the problem that certain quantities, even certain lengths, could not be reduced to numbers, as then understood. This theory anticipates the development of the real number system by Dedekind and Cantor in the nineteenth century.

## 1.5 Ordered Fields

In this section, we describe the formal properties that the set of rational numbers possesses, along with one useful property that it lacks.

**1.9 Definition.** *A field is a set  $K$ , containing at least two elements, together with two mappings of  $K \times K \rightarrow K$  called addition and multiplication, and written  $(a, b) \mapsto a + b$  and  $(a, b) \mapsto ab$ , respectively, with the following properties:*

1. for all  $a, b, c \in K$ ,  $(a + b) + c = a + (b + c)$ ;
2. for all  $a, b \in K$ ,  $a + b = b + a$ ;
3. for all  $a, b \in K$ , there exists  $c \in K$  such that  $a + c = b$ ;
4. for all  $a, b, c \in K$ ,  $(ab)c = a(bc)$ ;
5. for all  $a, b \in K$ ,  $ab = ba$ ;
6. for all  $a, b \in K$  such that  $a + b \neq b$ , there exists  $c \in K$  such that  $ac = b$ ; and
7. for all  $a, b, c \in K$ ,  $a(b + c) = ab + ac$ .

The reader can observe that the first three properties refer only to addition, the next three only to multiplication, while the last connects the two operations. The first and fourth are known as the associative laws, the second and fifth as the commutative laws, and the last as the distributive law. We now deduce a few properties from these axioms.

**1.10 Proposition.** *In any field  $K$ , there exist distinguished elements 0 and 1 with the properties:  $a + 0 = a$  for all  $a \in K$ ,  $a1 = a$  for all  $a \in K$ . If  $a, b \in K$  and  $a + b = a$ , then  $b = 0$ ; if  $a, b \in K$  and  $ab = a$ , then either  $a = 0$  or  $b = 1$ . Lastly,  $a0 = 0$  for all  $a \in K$ , and  $1 \neq 0$ .*

**Proof.** Let  $x \in K$  ( $K$  is not empty by assumption). By property 3, there exists  $0 \in K$  such that  $x + 0 = x$ . Now if  $a$  is any element of  $K$ , there exists  $b \in K$  with  $b + x = a$ ; but then

$$a + 0 = (b + x) + 0 = b + (x + 0) = b + x = a,$$

as claimed. Now suppose that  $a + b = a$ ; there exists  $c$  such that  $a + c = 0$ , and it follows that

$$b = b + (a + c) = (b + a) + c = a + c = 0.$$

Similarly, choose  $y \in K$  with  $y \neq 0$  ( $K$  has at least two elements by assumption.) For any  $a \in K$ , we have  $a + y \neq a$ , so by property 6 there exists an element  $1$  of  $K$  such that  $y1 = y$ . Now for any  $a \in K$ , there exists by property 6 some  $b \in K$  such that  $by = a$ ; it follows that

$$a1 = (by)1 = b(y1) = by = a,$$

as desired. Now suppose that  $ab = a$  and  $a \neq 0$ . Then there is some  $c \in K$  with  $ac = 1$ , and we have

$$1 = ac = (ab)c = (ba)c = b(ac) = b1 = b,$$

as desired. Finally, if  $a \in K$ , we have

$$a + a0 = a1 + a0 = a(1 + 0) = a1 = a,$$

so  $a0 = 0$ . It follows that if  $1 = 0$ , then  $a = a1 = 0$  for all  $a \in K$ , which contradicts the assumption that  $K$  has at least two elements. ■

**1.11 Corollary.** *For each  $a, b \in K$ , there is a unique  $c \in K$  such that  $a + c = b$ ; we denote this element  $c$  by  $b - a$ , or when  $b = 0$ , simply by  $-a$ . Clearly,  $-(-a) = a$  for all  $a \in K$ . Similarly, if  $a \in K$ ,  $a \neq 0$ , there is a unique  $b \in K$  with  $ab = 1$ ; we denote this element  $b$  as  $a^{-1}$ , or  $1/a$ .*

We will not go on and list and prove all the usual commonplaces of algebra, such as that  $-a = (-1)a$  for each  $a \in K$ , or the general associative laws which enable us to write  $a_1 + a_2 + \cdots + a_n$  and  $a_1 a_2 \cdots a_n$  without any ambiguity. If the reader has never engaged in such exercises, this might be a good time to do so. Observe that the symbols  $0$  and  $1$  are badly chosen; we should write  $0_K$  and  $1_K$  to avoid confusion with the integers (or rational numbers)  $0$  and  $1$ . By paying attention to context, we will hopefully always be clear what is meant. In fact, the fields we are most interested in can be regarded as containing the rational numbers, so there is no danger.

Let  $K$  be a field, and  $a \in K$ . We define  $na$  for each natural number  $n$  inductively:  $1a = a$ , and  $na = a + (n - 1)a$  for  $n > 1$ . It may happen that  $pa = 0$  for some  $p \in \mathbb{N}$  and  $a \in K$  with  $a \neq 0$  (this clearly occurs if and

only if  $p1 = 0$ , where 1 here is the element of  $K$ , not the natural number). If  $p$  is the smallest natural number for which this occurs, we say that  $K$  has characteristic  $p$ . If  $n1 \neq 0$  for every  $n \in \mathbf{N}$ , we say that  $K$  has characteristic 0. Of course, our basic example  $\mathbf{Q}$  is a field of characteristic 0.

Every field, according to Proposition 1.10, has the elements 0 and 1, and the smallest field consists of exactly these two elements, with the addition rule  $1 + 1 = 0$  and all the other addition and multiplication rules forced by that proposition. But among fields with characteristic 0, the smallest field is the field  $\mathbf{Q}$  of rational numbers. Indeed, if  $K$  has characteristic 0, the map  $n \mapsto n1$  is an injective mapping of  $\mathbf{N}$  into  $K$ . This injection extends in the obvious way to the field  $\mathbf{Q}$  of rationals, and this injection is an *isomorphism*, i.e., it carries the addition and multiplication of  $\mathbf{Q}$  to the addition and multiplication in the field  $K$ . We identify the image of this map with  $\mathbf{Q}$  itself, i.e., regard  $\mathbf{Q}$  as a subset of  $K$ . It is evidently the smallest subfield of  $K$ .

**1.12 Definition.** An ordered field is a field  $K$ , together with a total order relation  $<$  on  $K$ , satisfying the following conditions:

- (a) if  $a < b$ , then  $a + c < b + c$  for every  $c \in K$ ; and
- (b) if  $a < b$  and  $0 < c$ , then  $ac < bc$ .

We write  $a > b$  to mean  $b < a$ , and  $a \leq b$  to mean: either  $a < b$  or  $a = b$ . Similarly,  $a \geq b$  means that either  $a > b$  or  $a = b$ .

Taking  $c = -a$  in (a) of the definition, we see that  $a < b$  implies  $b - a > 0$ , and similarly  $b - a > 0$  implies  $a < b$ .

**1.13 Proposition.** Let  $K$  be an ordered field. Then

- (a) for each  $a \in K$ ,  $a > 0$  if and only if  $-a < 0$ ;
- (b) if  $a < b$  and  $c < 0$ , then  $ac > bc$ ;
- (c) for each  $a \in K$ , either  $a = 0$  or  $a^2 > 0$ ; and
- (d) for each  $a \in K$ , either  $a = 0$  or  $na \neq 0$  for every natural number  $n$ .

**Proof.** Suppose  $a > 0$ . Take  $c = -a$  in condition (a) of Definition 1.12 to deduce that  $-a < 0$ . Since  $a = -(-a)$ , this gives also that  $-a < 0$  implies  $a > 0$ . Thus (a) is proven. Since  $c < 0$  implies  $-c > 0$  by (a), we have  $a < b$  implies  $a(-c) < b(-c)$  by (a) of Definition 1.12, which gives  $-ac < -bc$  or  $ac > bc$  by another use of (a). If  $a \neq 0$ , then either  $a > 0$ , when  $a^2 > 0$  by property (b) of Definition 1.12, or  $-a > 0$ , when  $a^2 = (-a)^2 > 0$ . In particular,  $1 > 0$ , since  $1 = 1^2$ . It follows that  $1 + 1 > 1 > 0$  and by induction that  $n1 > 0$  for every natural number  $n$ ; in particular,  $n1 \neq 0$ . Thus  $na$  can be interpreted as the product of the positive element  $n1$  and  $a$ , so  $na > 0$  whenever  $a > 0$ , and  $na < 0$  whenever  $a < 0$ . ■

Of course, part (d) of this proposition can be restated as: every ordered field has characteristic zero.

**1.14 Proposition.** *Let  $K$  be a field, and let  $P \subset K$  have the properties:*

1. *if  $a, b \in P$ , then  $a + b \in P$ ;*
2. *if  $a, b \in P$ , then  $ab \in P$ ; and*
3. *for any  $a \in K$ , exactly one of the alternatives  $a \in P$ ,  $-a \in P$ ,  $a = 0$  holds.*

*If we define the relation  $<$  by the rule:  $a < b$  if and only if  $b - a \in P$ , then  $K$  with  $<$  is an ordered field.*

We leave the proof of this as an exercise. We note that the converse of this proposition is also trivially true: if  $K$  is an ordered field, and we put  $P = \{a \in K : a > 0\}$ , then  $P$  has the properties of the proposition, and  $b < a$  if and only if  $a - b > 0$ , as we have seen.

**1.15 Example.** It is evident that  $\mathbf{Q}$  is an ordered field, with the usual meaning of  $<$ . Here is a less obvious ordered field: let  $\mathbf{Q}(X)$  be the field of all rational functions in the indeterminate  $X$ . More precisely, consider the collection of all expressions of the form  $p(X)/q(X)$ , where  $p$  and  $q$  are polynomials in the indeterminate  $X$  with rational coefficients, and  $q \neq 0$ . Declare two such expressions  $p_1(X)/q_1(X)$  and  $p_2(X)/q_2(X)$  to be equivalent if  $p_1(X)q_2(X) = p_2(X)q_1(X)$ , and let  $\mathbf{Q}(X)$  be the set of equivalence classes. The addition and multiplication in  $\mathbf{Q}(X)$  are the familiar ones. To define an order, we decree that a polynomial  $p(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$  is positive if  $a_k > 0$ , where  $k$  is the smallest integer with  $a_k \neq 0$ . Thus, if  $a_0 > 0$ , then every polynomial of the form  $a_0 + a_1X + \cdots + a_nX^n$  is positive, and in particular, the constant polynomial  $a_0$  is bigger than any polynomial  $b_1X + \cdots + b_nX^n$  without a constant term. Next we define  $r(X) = p(X)/q(X)$  to be positive if either both  $p(X)$  and  $q(X)$  are positive, or both are negative. It is easy to see that this does not depend on the particular choice of  $p(X)$  and  $q(X)$  used to represent  $r(X)$ , and that the set  $P$  of such positive elements has the properties of the last proposition. Of course, we could have made such a construction starting with any ordered field  $K$  instead of  $\mathbf{Q}$ .

**1.16 Definition.** *Let  $S$  be a subset of the partially ordered set  $X$ . We say that  $M \in X$  is an upper bound for  $S$  if  $x \leq M$  for all  $x \in S$ . Similarly, we say that  $m \in X$  is a lower bound for  $S$  if  $m \leq x$  for all  $x \in S$ . We say that  $M$  is a maximal element of  $S$  if  $M \in S$  and there exists no  $x \in S$  with  $M < x$ . Similarly,  $m$  is a minimal element of  $S$  means that  $m \in S$  and for no  $x \in S$  do we have  $x < m$ .*

We note that when  $X$  is totally ordered, a maximal element  $M$  of  $S$  is an upper bound for  $S$ , and a minimal element of  $S$  is a lower bound. There can be at most one maximal element of  $S$  in this case; for if also  $M' \in S$  and  $M'$  is an upper bound for  $S$ , we would have  $M \leq M'$  and  $M' \leq M$ , so  $M = M'$ . Similarly, there exists at most one minimal element of  $S$ . Thus,  $M$  is a maximal element of  $S$  means  $M$  is the greatest element of  $S$ , and  $m$  is a minimal element of  $S$  means  $m$  is the least element of  $S$ . It is easy to see that if  $X$  is totally ordered, any finite subset  $F$  of  $X$  has a greatest and a least element, denoted, respectively, by  $\max\{x : x \in F\}$  and  $\min\{x : x \in F\}$ .

Here are two simple examples, with  $X = \mathbf{Q}$ . If  $S = \{1/n : n \in \mathbf{N}\}$ , then every  $q \in \mathbf{Q}$  with  $q \geq 1$  is an upper bound for  $S$ , and 1 is a maximal element of  $S$ . Every  $q \in \mathbf{Q}$  with  $q \leq 0$  is a lower bound for  $S$ , while for any  $q > 0$  there exists  $n \in \mathbf{N}$  with  $1/n < q$ , so  $q$  is not a lower bound for  $S$ . Thus the set of lower bounds for  $S$  is  $\{q : q \leq 0\}$ . The greatest lower bound for  $S$  is 0, but  $S$  admits no minimal element. Here is a more complicated example.

**1.17 Example.** Let  $S = \{q \in \mathbf{Q} : q^2 < 2\}$ . If  $M^2 \geq 2$  and  $M > 0$ , then  $M$  is an upper bound for  $S$ ; for if  $q \in \mathbf{Q}$  and  $q > M$ , then  $q^2 > qM > M^2 \geq 2$ , so  $q \notin S$ . Conversely, if  $M$  is an upper bound for  $S$ , then  $M > 0$  and  $M^2 \geq 2$ . For  $1 \in S$ , so  $M \geq 1 > 0$ . If  $M^2 < 2$ , then for each  $n \in \mathbf{N}$  we have

$$\left(M + \frac{1}{n}\right)^2 = M^2 + \frac{1}{n}\left(2M + \frac{1}{n}\right) \leq M^2 + \frac{1}{n}(2M + 1),$$

and we can choose  $n$  so that  $(1/n)(2M+1) < 2 - M^2$ , which gives  $M + 1/n \in S$ . Since  $M < M + 1/n$ , this contradicts the supposition that  $M$  is an upper bound for  $S$ . Thus  $M$  is an upper bound for  $S$  if and only if  $M > 0$  and  $M^2 \geq 2$ . Now if  $M^2 > 2$ , then  $(M - 1/n)^2 > 2$  for some  $n \in \mathbf{N}$ , by a calculation like the one just made. Thus a least upper bound for  $S$  would be a number  $M > 0$  with the property that  $M^2 = 2$ . As we saw in the last section, there exists no such  $M \in \mathbf{Q}$ . Thus  $S$  has upper bounds, but no least upper bound.

If  $K$  is an ordered field, it is easy to see that the set of upper bounds of a set  $S \subset K$  is either empty or infinite; indeed, if  $M$  is an upper bound, so is  $M'$  for any  $M' \geq M$ . We note that  $m$  is a lower bound for  $S$  (or minimal element of  $S$ ) if and only if  $-m$  is an upper bound for (resp., maximal element of) the set  $-S = \{-x : x \in S\}$ . This observation makes it easy to deduce generalities about lower bounds or minimal elements from the corresponding generalities about upper bounds or maximal elements.

**1.18 Definition.** We say that  $K$  is a complete ordered field if  $K$  is an ordered field with the property: given any subset  $S$  of  $K$ , if the set  $\mathcal{M}_S$  of upper bounds of the set  $S$  is nonempty, then  $\mathcal{M}_S$  possesses a least element.



In other words, the ordered field  $K$  is complete if and only if every subset  $S$  of  $K$  which has an upper bound must have a least upper bound. It is an immediate consequence that any subset  $S$  of a complete ordered field which has a lower bound must have a greatest lower bound. Example 1.17 shows that the ordered field  $\mathbf{Q}$  is not complete. We shall use the word *supremum* as a synonym for least upper bound, and *infimum* as a synonym for greatest lower bound. The notations

$$\sup S \text{ or } \sup_{x \in S} x, \quad \inf S \text{ or } \inf_{x \in S} x$$

will denote the supremum, infimum respectively, of the set  $S$ . If  $S$  has no upper bound, we write  $\sup S = +\infty$ , and if  $S$  has no lower bound, we write  $\inf S = -\infty$ . It is sometimes convenient to also take  $\sup \emptyset = -\infty$  and  $\inf \emptyset = +\infty$ . (After all, every  $x$  is an upper bound and a lower bound for  $\emptyset$ .)

It turns out that there exists a complete ordered field (not an obvious fact), and (essentially) only one. We sketch this result in a later section of this chapter, and for now go on to study the properties of such a field, assuming it does exist.

We introduce the following notation, to be fixed for the rest of the book: let  $\mathbf{R}$  be a complete ordered field. We call  $\mathbf{R}$  the field of real numbers. We define the set  $\bar{\mathbf{R}}$  of *extended real numbers* to be  $\mathbf{R}$  with two more elements adjoined, denoted  $+\infty$  and  $-\infty$ , with the order relation of  $\mathbf{R}$  extended by the rule  $-\infty < x < +\infty$  for all  $x \in \mathbf{R}$ . We will not at this time consider any algebraic operations among elements of  $\bar{\mathbf{R}}$ ; we emphasize that  $\bar{\mathbf{R}}$  is a totally ordered set, but not a field.

**1.19 Theorem.** *Let  $\epsilon \in \mathbf{R}$ ,  $\epsilon > 0$ . For any  $M \in \mathbf{R}$ , there exists  $n \in \mathbf{N}$  such that  $n\epsilon > M$ .*

**Proof.** Let  $S = \{n\epsilon : n \in \mathbf{N}\}$ . The theorem asserts that  $S$  is unbounded. If on the contrary  $S$  has an upper bound, then by the definition of complete ordered field, it has a least upper bound  $R$ . Then we have  $n\epsilon \leq R$  for every  $n \in \mathbf{N}$ , since  $R$  is an upper bound for  $S$ , but, since  $R$  is the smallest upper bound for  $S$ , there exists  $m \in \mathbf{N}$  with  $m\epsilon > R - \epsilon$ ; this implies  $(m+1)\epsilon > R$ , a contradiction. ■

**1.20 Corollary.** *Let  $x \in \mathbf{R}$ . There exists a unique  $n \in \mathbf{Z}$  such that  $n \leq x < n+1$ ; this integer  $n$  is called the largest integer in  $x$ , and denoted by  $[x]$ .*

**Proof.** Suppose  $x > 0$ . By Theorem 1.19,  $\{k \in \mathbf{N} : k > x\}$  is not empty, and hence there is a smallest integer  $n$  with  $n > x$ ; evidently,  $[x] = n - 1$ . If  $x < 0$ , let  $m = \min\{k \in \mathbf{N} : k \geq -x\}$ , which exists by the same reasoning; it is easy to see that  $-m = [x]$ . ■

The property of  $\mathbf{R}$  described in Theorem 1.19 is known as the Archimedean property. The property also holds for the field  $\mathbf{Q}$  (and is trivial), but it does not hold in every ordered field. For instance, it does not hold in the field  $\mathbf{Q}(X)$  discussed in Example 1.15 (or in  $\mathbf{R}(X)$  for that matter), since, for instance,  $X > 0$  but  $nX < 1$  for every  $n$ .

We next show that the rational numbers come arbitrarily close to any real number. Again, we use only the Archimedean property of  $\mathbf{R}$ .

**1.21 Definition.** A subset  $E$  of  $\mathbf{R}$  is said to be dense in  $\mathbf{R}$  if for any  $x \in \mathbf{R}$ , and any  $\epsilon > 0$ , there exists  $t \in E$  with  $x - \epsilon < t < x + \epsilon$ .

**1.22 Proposition.** The rational numbers  $\mathbf{Q}$  form a dense subset of  $\mathbf{R}$ .

**Proof.** Let  $x \in \mathbf{R}$  and  $\epsilon > 0$ . There exists  $n \in \mathbf{N}$  such that  $n\epsilon > 1$ , i.e., such that  $1/n < \epsilon$ . Let  $m = [nx]$ , so  $m \in \mathbf{N}$  and  $m \leq nx < m + 1$ . Put  $q = m/n$ ; then  $q \leq x < q + 1/n$ , so  $x - \epsilon < q < x + \epsilon$ . ■

The elements of  $\mathbf{R} \setminus \mathbf{Q}$  are called *irrational numbers*. Here is a pretty theorem of Dirichlet.

**1.23 Theorem.** If  $\alpha$  is an irrational number, then

$$S = \{n\alpha + m : m \in \mathbf{Z}, n \in \mathbf{N}\}$$

is a dense subset of  $\mathbf{R}$ .

**Proof.** Let us begin with the observation that if  $\xi \in S$  and  $k$  is a positive integer, then  $k\xi + m \in S$  for any integer  $m$ . To show that  $S$  is dense in  $\mathbf{R}$ , it suffices to show that for any positive integer  $N$ , there exists  $\xi \in S$  such that  $|\xi| < 1/N$ . For instance, if  $0 < \xi < 1/N$ , then given any  $t \in \mathbf{R}$ , we first choose  $m \in \mathbf{Z}$  so that  $t + m > 0$ , then let  $k$  be the smallest positive integer such that  $k\xi > t + m$ ; we have  $\eta = k\xi - m \in S$ , and  $t < \eta < t + 1/N$ . If  $-1/N < \xi < 0$ , we can choose  $m$  so  $t + m < 0$ , and proceed in a similar manner. Now if  $N$  is a positive integer, let  $\xi_n = n\alpha - [n\alpha]$ , for  $n = 1, 2, \dots, N + 1$ . (Here, as usual,  $[x]$  denotes the greatest integer not greater than  $x$ .) We observe that: (i)  $0 < \xi_n < 1$ , and (ii) if  $n < k$ , then  $\xi_k - \xi_n \in S$  (in particular,  $\xi_n \neq \xi_k$ ). Consider the  $N$  intervals  $((i-1)/N, i/N)$  ( $i = 1, \dots, N$ ); in view of (i), and the fact that each  $\xi_n$  is irrational, each  $\xi_n$  belongs to one of these intervals. But there are  $N + 1$  of these  $\xi_n$ , so there must exist an interval  $((i-1)/N, i/N)$  which contains  $\xi_n$  and  $\xi_k$  for some  $n < k$ . But then  $\xi = \xi_k - \xi_n \in S$  by (ii), and  $|\xi| < 1/N$ . ■

## 1.6 Functions on $\mathbf{R}$

The algebraic operations on  $\mathbf{R}$  can be used to define many functions from  $\mathbf{R}$  (or subsets of  $\mathbf{R}$ ) to  $\mathbf{R}$ . Thus, for each  $c \in \mathbf{R}$  we have the constant

function  $x \mapsto c$ , and the functions  $x \mapsto c + x$  and  $x \mapsto cx$ ; combining these two ideas, we construct  $x \mapsto c + dx$ , for any given  $c, d \in \mathbf{R}$ . Iterating these ideas, we obtain the polynomial functions  $p$  on  $\mathbf{R}$ , where  $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ . Such a polynomial, we recall, is said to have degree  $n$  if  $a_n \neq 0$  above; thus constant functions are polynomials of degree 0, with one exception: the zero polynomial (with all coefficients  $a_j = 0$ ) is said to have degree  $-\infty$ . (This enables the rule that the degree of a product is the sum of the degrees of the factors to be true without exceptions.) Similarly, if  $f$  and  $g$  are functions from  $X$  to  $\mathbf{R}$ ,  $g \neq 0$ , we define  $f/g$  by the formula  $(f/g)(x) = f(x)/g(x)$ ; the domain of  $f/g$  is, of course, not all of  $X$  in general, but  $\{x \in X : g(x) \neq 0\}$ . If  $p$  and  $q$  are polynomials,  $q \neq 0$ , the function  $p/q$  is called a rational function. When  $q \neq 0$ , the zero set of  $q$  is finite, as we see from the division algorithm for polynomials: if the degree of  $q$  is  $n > 0$ , then for any  $a \in \mathbf{R}$  we can write  $q(x) = (x - a)q_1(x) + r(x)$ , where the degree of  $r(x)$  is less than 1, i.e.,  $r(x)$  is a constant. If  $q(a) = 0$ , it follows that  $r(a) = 0$ , i.e., that  $r = 0$ , or  $q(x) = (x - a)q_1(x)$ . Here the degree of  $q_1$  is evidently  $n - 1$ . Since  $q(b) = 0$  if and only if  $b = a$  or  $q_1(b) = 0$ , we obtain inductively that  $q$  has at most  $n$  zeros.

Another useful function is  $|x|$ , defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0; \\ -x & \text{if } x < 0. \end{cases}$$

This function has, as is easily seen, the following properties: for all  $x, y \in \mathbf{R}$ ,

$$|xy| = |x||y|, \quad |x + y| \leq |x| + |y|, \quad |x| > 0 \text{ unless } x = 0.$$

We shall use these properties frequently throughout the rest of this book.

Related is the *sign* function:  $\operatorname{sgn} x = x/|x|$  for  $x \neq 0$ , and  $\operatorname{sgn} 0 = 0$ . We note that  $x = |x|\operatorname{sgn} x$  for all  $x \in \mathbf{R}$ . We will later want to use also the functions  $x \mapsto x^+$  and  $x \mapsto x^-$ , defined by  $x^+ = \max\{x, 0\}$  and  $x^- = \max\{-x, 0\}$ . We note that for any  $x$ , we have  $x^+ \geq 0$ ,  $x^- \geq 0$ ,  $x = x^+ - x^-$  and  $|x| = x^+ + x^-$ . We denote by  $\mathbf{R}_+$  the set of nonnegative real numbers, so  $\mathbf{R}_+ = \{x^+ : x \in \mathbf{R}\}$ .

The power functions  $x \mapsto x^n$  are defined on all  $\mathbf{R}$  when  $n \in \mathbf{N}$ , and they are injective when restricted to  $\mathbf{R}_+$ . Indeed, if  $0 \leq x < y$ , we have  $x^n < y^n$ . It follows that when  $n$  is odd,  $x \mapsto x^n$  is injective on  $\mathbf{R}$ , since  $(-x)^n = -x^n$  for  $n$  odd. These simple remarks are valid in any ordered field. The next theorem, which is a generalized complement to Example 1.17, uses the completeness of  $\mathbf{R}$  in an essential way. We need the following lemma:

**1.24 Lemma.** *Let  $n \in \mathbf{N}$ , and  $x, y \in \mathbf{R}$ ,  $x \geq 0$  and  $y > 0$ . If  $x^n < y$ , there exists  $t > x$  with  $t^n < y$ , and if  $x^n > y$ , there exists  $0 < t < x$  with  $t^n > y$ .*

**Proof.** We recall the algebraic identity

$$b^n - a^n = (b - a)(b^{n-1} + b^{n-2}a + \cdots + a^{n-1})$$

valid in any field. If  $0 \leq a < b$ , this gives the inequality

$$b^n < a^n + (b - a)nb^{n-1}.$$

If  $x^n < y$ , taking  $a = x$  and  $b = x + h$  in this inequality (with  $h > 0$ ) we get  $(x+h)^n < x^n + hn(x+h)^{n-1}$ , so for  $0 < h < 1$  with  $h < (y - x^n)/n(x+1)^{n-1}$ , we have  $(x+h)^n < y$ . Similarly, if  $0 < h < x$ , taking  $b = x$  and  $a = x - h$  in the inequality gives  $(x-h)^n > x^n - hn(x-h)^{n-1}$ , so if  $x^n > y$  and we take  $0 < h < x$  with  $h < (x^n - y)/nx^{n-1}$ , we get  $(x-h)^n > y$ . ■

**1.25 Theorem.** Let  $n \in \mathbf{N}$ . Then the map  $x \mapsto x^n$  of  $\mathbf{R}_+$  to  $\mathbf{R}_+$  is bijective. In other words, for each  $y \in \mathbf{R}$ ,  $y \geq 0$ , there exists a unique  $x \in \mathbf{R}$ , with  $x \geq 0$  and  $x^n = y$ .

**Proof.** We have already observed that the map is injective, so we have only to prove that it is surjective, i.e., that for every  $y \geq 0$ , there exists  $x \geq 0$  such that  $x^n = y$ . The result is obvious when  $y = 0$ , so we assume now that  $y > 0$ . Let  $S = \{t \geq 0 : t^n < y\}$ . Since  $0 \in S$ ,  $S \neq \emptyset$ . Furthermore,  $S$  is bounded above. In fact, if  $t > 1 + y$ , we have  $t^n > (1+y)^n > 1 + y > y$ , so  $t \notin S$ ; thus  $t \leq 1 + y$  for all  $t \in S$ . Hence there exists a least upper bound  $x$  for  $S$ . If  $x^n < y$ , there exists, by the last lemma,  $t > x$  with  $t^n < y$ ; but then  $t \in S$ , contradicting that  $x$  is an upper bound for  $S$ . On the other hand, if  $x^n > y$ , there exists  $u$  with  $0 < u < x$  and  $u^n > y$ . Now if  $t > u$ , it follows that  $t^n > y$ , so we conclude that  $t \leq u$  for all  $t \in S$ , i.e., that  $u$  is an upper bound for  $S$ . But  $u < x$ , so this is impossible. We conclude that  $x^n = y$ . ■

We denote the unique  $x \geq 0$  with  $x^n = y$  by  $y^{1/n}$ , or  $\sqrt[n]{y}$ . When  $n$  is odd, and  $y < 0$ , there exists  $t > 0$  with  $t^n = -y$ , so  $(-t)^n = (-1)^n t^n = y$ ; thus, when  $n$  is odd, there exists for every  $y \in \mathbf{R}$  a unique  $x \in \mathbf{R}$  with  $x^n = y$ . We denote this number  $x$  by  $y^{1/n}$ . We can now define (for  $x \geq 0$ )  $x^q$  for any rational  $q = m/n$  by  $x^{m/n}$  as  $(x^m)^{1/n}$ ; we should show that this does not depend on the particular choice of the representation  $q = m/n$ , that  $(x^m)^{1/n} = (x^{1/n})^m$ , that  $x^{q+r} = x^q x^r$ , and that  $x^{qr} = (x^q)^r$  for any rational  $q$  and  $r$ , etc.

## 1.7 Intervals in $\mathbf{R}$

**1.26 Definition.** Let  $K$  be an ordered field. We say  $J$  is an interval in  $K$  if  $J$  is a subset of  $K$  with the property: if  $a < b < c$  with  $a, c \in J$ , then  $b \in J$ .

Note that this definition counts both  $K$  and  $\emptyset$  as intervals. Note also that each singleton set  $\{a\}$  is an interval. Given any  $a, b \in K$  with  $a < b$ , we define  $(a, b) = \{x \in K : a < x < b\}$ . (The danger of confusion of the interval  $(a, b)$  with the ordered pair  $(a, b)$  is usually minimal. Some writers consider it serious enough to use the notation  $]a, b[$  instead of  $(a, b)$ , but we'll risk it.) It is immediate that  $(a, b)$  is an interval; we call it the open interval with endpoints  $a$  and  $b$ . Similarly, we define the closed interval with endpoints  $a$  and  $b$  to be  $[a, b] = \{x \in K : a \leq x \leq b\}$ , and two kinds of semiclosed intervals:  $[a, b) = \{x \in K : a \leq x < b\}$  and  $(a, b] = \{x \in K : a < x \leq b\}$ . This does not exhaust the possibilities for intervals; we also note that  $\{x \in K : x > a\}$  is an interval, which we denote by  $(a, +\infty)$ , as is  $\{x \in K : x \geq a\}$ , denoted by  $[a, +\infty)$ . The intervals  $(-\infty, b)$  and  $(-\infty, b]$  are defined in an analogous manner. We sometimes write  $(-\infty, +\infty)$  instead of  $K$ . We call  $(-\infty, b]$  and  $[a, +\infty)$  closed intervals, and call  $(-\infty, b)$  and  $(a, +\infty)$  open intervals.

In the case of a complete ordered field, it is easy to classify all the intervals, as follows. Let  $J$  be a nonempty interval in  $\mathbf{R}$ . Let  $a = \inf J$ , and let  $b = \sup J$ . We note  $-\infty \leq a \leq b \leq +\infty$ , but  $a = +\infty$  and  $b = -\infty$  are excluded since  $J$  is nonempty. If  $a < x < b$ , then (from the definition of  $\inf$  and  $\sup$ ) there exist  $c, d \in J$  with  $a < c < x < d < b$ , and it follows that  $x \in J$ . Thus  $J \supset (a, b)$ . We have  $a \leq x \leq b$  for all  $x \in J$ . If  $b < +\infty$ , there are two possibilities: either  $b \in J$  or  $b \notin J$ . Similarly, if  $a > -\infty$ , there are the two possibilities. Thus we see that every nonempty interval  $J$  in  $\mathbf{R}$  must have one of the forms described above:  $(a, b)$  or  $(a, b]$  or  $[a, b)$  or  $[a, b]$  or  $(-\infty, b)$  or  $(-\infty, b]$  or  $(a, +\infty)$  or  $[a, +\infty)$  or  $(-\infty, +\infty)$ . The first four of these are called bounded intervals, the next five are called unbounded.

**1.27 Theorem.** *Suppose that for each  $n \in \mathbf{N}$ ,  $J_n$  is a nonempty closed bounded interval in  $\mathbf{R}$ , with  $J_{n+1} \subset J_n$  for each  $n$ . Then  $\bigcap_{n=1}^{\infty} J_n \neq \emptyset$ .*

**Proof.** By hypothesis, each  $J_n$  has the form  $J_n = [a_n, b_n]$  for some  $a_n, b_n \in \mathbf{R}$ ,  $a_n \leq b_n$ . In fact, we have  $a_n \leq b_m$  for any  $m, n \in \mathbf{N}$ . For if  $k = \max\{m, n\}$ , we have  $J_k \subset J_n$ , which implies  $a_n \leq a_k$ , and  $J_k \subset J_m$ , which implies  $b_k \leq b_m$ ; thus we have  $a_n \leq a_k \leq b_k \leq b_m$ . Let  $c = \sup\{a_n : n \in \mathbf{N}\}$ . Then  $a_n \leq c$  for every  $n$ , and  $c \leq b_m$  for every  $m$ , so  $a_n \leq c \leq b_n$  for every  $n$ , i.e.,  $c \in \bigcap_{n=1}^{\infty} J_n$ . ■

The hypothesis that each  $J_n$  is closed cannot be dispensed with in this theorem, nor that each  $J_n$  is bounded, nor that the ordered field involved is complete. See the exercises at the end of this chapter.

**1.28 Theorem.** *The interval  $(0, 1)$  in  $\mathbf{R}$  is not countable.*

**Proof.** Let  $E$  be a countable subset of  $(0, 1)$ , say  $E = \{x_1, x_2, \dots\}$ . We make the following practically trivial remark: given a nonempty open interval  $(a, b)$  in  $\mathbf{R}$ , and  $x \in \mathbf{R}$ , there exist  $c < d$  such that  $[c, d] \subset (a, b)$  and  $x \notin$

$[c, d]$ . By this remark, we can choose  $a_1 < b_1$  such that  $x_1 \notin [a_1, b_1] \subset (0, 1)$ . Having chosen  $a_1, b_1, a_2, b_2, \dots, a_n, b_n$  such that  $a_k < b_k$  and  $x_k \notin [a_k, b_k]$  for  $k = 1, 2, \dots, n$ , and such that  $[a_{k+1}, b_{k+1}] \subset (a_k, b_k)$  for  $k = 1, \dots, n-1$ , the remark enables us to choose  $a_{n+1} < b_{n+1}$  such that  $[a_{n+1}, b_{n+1}] \subset (a_n, b_n)$  and  $x_{n+1} \notin [a_{n+1}, b_{n+1}]$ . Thus we have inductively defined for each  $n \in \mathbf{N}$  a closed interval  $J_n = [a_n, b_n]$ , such that  $J_{n+1} \subset J_n$  and  $x_n \notin J_n$  for every  $n \in \mathbf{N}$ . According to Theorem 1.27, there exists  $x \in \bigcap_{n=1}^{\infty} J_n$ . Since  $x_m \notin \bigcap_{n=1}^{\infty} J_n$  for every  $m \in \mathbf{N}$ , we conclude  $x \notin E$ . Thus  $E \neq (0, 1)$ . ■

## 1.8 Algebraic and Transcendental Numbers

**1.29 Definition.** A real number  $x$  is said to be algebraic if there exists a positive integer  $n$ , and integers  $a_0, a_1, \dots, a_n$ ,  $a_n \neq 0$ , such that

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0. \quad (1.1)$$

We say that  $x$  is algebraic of degree  $n$  if  $n$  is the smallest positive integer for which  $x$  satisfies an equation of the form (1.1). We say that  $x$  is transcendental if it is not algebraic.

We were motivated to expand from  $\mathbf{Q}$  to  $\mathbf{R}$  in order to be able to solve the equation  $x^2 = 2$ . We found that in  $\mathbf{R}$  we could solve any equation  $x^n = a$ , but that still leaves open the possibility that every real number is algebraic. There are at least two ways to see that this is not so.

**1.30 Proposition.** The set of transcendental numbers is uncountable.

**Proof.** Let  $A_N$  be the set of all numbers  $x$  which satisfy an equation of the form (1.1), with  $n + \sum_{j=0}^n |a_j| \leq N$ . Clearly, each  $A_N$  is finite, since each such equation has at most  $n$  solutions, and there are a finite number of such equations. But  $\bigcup_{N=2}^{\infty} A_N$  is the set of all algebraic numbers. Thus the set of algebraic numbers is the union of a sequence of finite sets, and hence is countable. Since  $\mathbf{R}$  is uncountable, the set of transcendental numbers is uncountable. ■

It is perhaps a little disappointing that this existence proof for transcendental numbers fails to exhibit a single one. Here is another approach. The following theorem of Liouville says that algebraic numbers which are not rational cannot be approximated too closely by rational numbers.

**1.31 Theorem.** Let  $x$  be an algebraic number of degree not more than  $n$ . Then there exists a constant  $C$  such that for any integers  $p, q$  ( $q > 0$ ) with  $x \neq p/q$ , we have

$$\left| x - \frac{p}{q} \right| > \frac{C}{q^n}.$$

**Proof.** Let  $f(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_0$ , where  $a_j$  is an integer for  $j = 0, 1, \dots, n$ , such that  $f(x) = 0$ . Then  $f(t) = (t - x)g(t)$ , where  $g$  is a polynomial of degree less than  $n$  (with real coefficients). Since  $g$  has at most  $n - 1$  zeros, there exists  $\delta > 0$  such that  $0 < |t - x| \leq \delta$  implies  $g(t) \neq 0$ , and hence also  $f(t) \neq 0$ . It is easy to see that there exists  $M$  such that  $|g(t)| \leq M$  for all  $t$  with  $|t - x| \leq \delta$ . For instance, if  $g(t) = \sum_{j=0}^{n-1} b_j t^j$ , choose  $N$  large enough so that  $[x - \delta, x + \delta] \subset [-N, N]$ , and take  $M = \sum_{j=0}^{n-1} |b_j| N^j$ . Now suppose  $|x - p/q| \leq \delta$ ,  $x \neq p/q$ . Then  $g(p/q) \neq 0$ , so we have

$$\frac{p}{q} - x = \frac{f(p/q)}{g(p/q)} = \frac{a_n p^n + a_{n-1} p^{n-1} q + \cdots + a_0 q^n}{q^n g(p/q)}.$$

Now the numerator of this last fraction is an integer, and it is not 0, since  $f(t) \neq 0$  for  $0 < |t - x| \leq \delta$ . Hence the numerator has absolute value at least 1, and we have

$$\left| x - \frac{p}{q} \right| \geq \frac{1}{M q^n}.$$

But if  $|x - p/q| > \delta$ , then of course  $|x - p/q| > \delta/q^n$ , since  $q \geq 1$ . Thus taking any  $C$  smaller than both  $\delta$  and  $1/M$  gives us the theorem. ■

**1.32 Example.** Let  $x_1 = \frac{1}{2}$ , and inductively define

$$x_n = x_{n-1} + \frac{1}{2^{n!}}.$$

Then  $x_{n+m} - x_n = 2^{-(n+1)!} + \cdots + 2^{-(n+m)!} < 2 \cdot 2^{-(n+1)!}$  for every  $n, m$ . Let  $x = \sup\{x_1, x_2, \dots\}$ . We have then  $x - x_n \leq 2 \cdot 2^{-(n+1)!}$ . But  $x_n = p_n 2^{-n!}$  for some integer  $p_n$ . Thus we have (taking  $q_n = 2^{n!}$ ) inequalities  $0 < x - p_n/q_n < 2 \cdot 2^{-n-1}/q_n$  for every  $n$ . According to the theorem, this is impossible if  $x$  is algebraic, so  $x$  is transcendental.

## 1.9 Existence of $\mathbf{R}$

In this section, we outline the construction first given by Dedekind. Dedekind presented his real numbers as “cuts” of the line, i.e., as pairs of sets of rationals, one set lying entirely to the left of the other, the union being the set of all rationals. Nowadays, we dispense with one element of the pair, since the left side of the cut carries all the information anyway. The following sketch omits many, perhaps most of the details, which are rather tedious.

**1.33 Theorem.** *There exists a complete ordered field  $\mathbf{R}$ . If  $R_1$  and  $R_2$  are complete ordered fields, there exists an isomorphism of ordered fields between them, i.e., there exists a bijective mapping  $\psi : R_1 \rightarrow R_2$  which*

*preserves the structure: for any  $x, y \in R_1$ , we have  $\psi(x + y) = \psi(x) + \psi(y)$  and  $\psi(xy) = \psi(x)\psi(y)$ ; for any  $x, y \in R_1$  with  $x < y$ , we have  $\psi(x) < \psi(y)$ .*

**Proof.** We define the set  $\mathbf{R}$  to consist of all subsets  $\alpha$  of the rational numbers  $\mathbf{Q}$ , having the properties:

- (a) if  $q \in \alpha$  and  $r < q$ , then  $r \in \alpha$ ;
- (b)  $\alpha$  has no greatest element; and
- (c)  $\emptyset \neq \alpha \neq \mathbf{Q}$ .

For each  $q \in \mathbf{Q}$ , the set  $\bar{q} = \{r \in \mathbf{Q} : r < q\}$  evidently belongs to  $\mathbf{R}$ , and the map  $q \mapsto \bar{q}$  is clearly injective. We observe that each  $\alpha \in \mathbf{R}$  is bounded above, as a subset of  $\mathbf{Q}$ . We define  $\alpha < \beta$  to mean that  $\alpha$  is a proper subset of  $\beta$ . We define the sum  $\alpha + \beta$  by  $\alpha + \beta = \{a + b : a \in \alpha, b \in \beta\}$ . It is not hard to show that  $\alpha + \beta \in \mathbf{R}$ . It is clear that  $\alpha + \beta = \beta + \alpha$ , and that  $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ . We next construct  $-\alpha$  for any given  $\alpha \in \mathbf{R}$ , as follows: let  $\hat{\alpha}$  be the set of all upper bounds for  $\alpha$  in  $\mathbf{Q}$ , with the exception of the least upper bound, if that exists ( $\alpha$  has a least upper bound if and only if  $\alpha = \bar{q}$  for some  $q \in \mathbf{Q}$ ). Define  $-\alpha = \{-q : q \in \hat{\alpha}\}$ . It is easy to verify that  $-\alpha \in \mathbf{R}$  for each  $\alpha \in \mathbf{R}$ , and that given any  $\alpha \in \mathbf{R}$ ,  $\beta \in \mathbf{R}$ , the element  $\gamma = \beta + (-\alpha)$  satisfies the equation  $\alpha + \gamma = \beta$ . We have verified the first three properties of a field (see Definition 1.9). We next define multiplication in  $\mathbf{R}$ ; to do this, we first consider the case  $\alpha > \bar{0}$ ,  $\beta > \bar{0}$ . For such  $\alpha, \beta$  we put  $\alpha\beta = \{ab : a \in \alpha, a \geq 0, b \in \beta, b \geq 0\} \cup \bar{0}$ . If  $\alpha > \bar{0}$  and  $\beta < \bar{0}$  we define  $\alpha\beta = -\alpha(-\beta)$ , and if  $\alpha < \bar{0}$  and  $\beta < \bar{0}$ , we put  $\alpha\beta = (-\alpha)(-\beta)$ . We define  $\bar{0}\alpha = \bar{0}$  for all  $\alpha$ . Again, it is clear that the operation is commutative and associative (properties 4 and 5 of Definition 1.9). If  $\alpha > \bar{0}$ , we define  $1/\alpha = \{1/q : q \in \hat{\alpha}\} \cup \{q \in \mathbf{Q} : q \leq 0\}$ . If  $\alpha < \bar{0}$ , we define  $1/\alpha = -1/(-\alpha)$ . It is not hard to verify that  $\alpha(1/\alpha) = \bar{1}$ , and more generally, that setting  $\gamma = (1/\alpha)\beta$  gives a solution to the equation  $\alpha\gamma = \beta$ , whenever  $\alpha \neq \bar{0}$ . Thus property 6 of Definition 1.9 is satisfied. It remains to check the distributive law to see that  $\mathbf{R}$  is a field. The path to this is to first prove it for positive elements, then consider the several other cases. Next, we have to check that  $\mathbf{R}$  is an ordered field (see Definition 1.12). Since  $\alpha < \beta$  is easily seen to be equivalent to  $\beta - \alpha > \bar{0}$ , this follows from the easy observation that  $\alpha + \beta > \bar{0}$  and  $\alpha\beta > \bar{0}$  whenever  $\alpha > \bar{0}$  and  $\beta > \bar{0}$ . Finally, we have to prove that  $\mathbf{R}$  is a complete ordered field. Suppose that  $A$  is any subset of  $\mathbf{R}$ . We define  $\sup A$  to be  $\bigcup_{\alpha \in A} \alpha$ . If  $A$  is bounded above, say  $\alpha \leq \gamma$  for all  $\alpha \in A$ , there exists  $M \in \mathbf{Q}$  such that  $\alpha \leq \bar{M}$  for all  $\alpha \in A$  (take any  $M \in \bar{\gamma}$ ). This means that  $a < M$  for every  $a \in \alpha$ , every  $\alpha \in A$ , so  $\bigcup_{\alpha \in A} \alpha \subset \bar{M}$ , i.e.,  $\sup A \in \mathbf{R}$ . It is clear that  $\sup A$  is the least upper bound of  $A$ . The existence of  $\mathbf{R}$  is thus established.

Note that if we leave out condition (c) in the definition of  $\mathbf{R}$ , and put  $+\infty = \mathbf{Q}$  and  $-\infty = \emptyset$ , we arrive at  $\bar{\mathbf{R}}$ , the extended reals.



Finally, we give a sketch of the proof of the uniqueness of  $\mathbf{R}$ . If  $R_1$  and  $R_2$  are complete ordered fields, we want to show that there exists an order-preserving isomorphism  $\psi$  of  $R_1$  onto  $R_2$ . Let  $Q_j$  be the smallest subfield of  $R_j$  ( $j = 1, 2$ ). Each  $Q_j$  is isomorphic to the field of rational numbers  $\mathbf{Q}$ , so there exists an isomorphism  $\phi$  of  $Q_1$  onto  $Q_2$ . This isomorphism is unique, and order-preserving. We now define a map  $\psi$  of  $R_1$  into  $R_2$  as follows: for  $x \in R_1$ , let  $\psi(x) = \sup\{\phi(q) : q \in Q_1, q < x\}$ . We note this makes sense: for any  $x \in R_1$ , there exists  $M \in Q_1$  with  $M > x$ , and hence  $M > q$  for any  $q < x$ ; then  $\phi(q) < \phi(M)$  for every  $q \in Q_1$  with  $q < x$ , since  $\phi$  is order-preserving, so  $\{\phi(q) : q \in Q_1, q < x\}$  has an upper bound in  $R_2$ , and hence a least upper bound. If  $x, y \in R_1$  with  $x < y$ , there exists  $q, r \in Q_1$  with  $x < q < r < y$ ; it follows easily that  $\psi(x) \leq \phi(q) < \phi(r) \leq \psi(y)$ , so  $\psi$  is order-preserving, and in particular is injective. It is easy to see that  $\psi$  is surjective: given  $y \in R_2$ , let  $x = \sup\{\phi^{-1}(q_2) : q_2 \in Q_2, q_2 < y\}$ , and check that  $\psi(x) = y$ . It remains to verify that  $\psi$  is a field homomorphism, i.e., that  $\psi(x + y) = \psi(x) + \psi(y)$  and  $\psi(xy) = \psi(x)\psi(y)$  for every  $x, y \in R_1$ . We leave this to the reader.

## 1.10 Exercises

1. List all the subsets of  $\mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset)))$ .
2. Show that a set  $X$  is infinite if and only if there exists a bijective map of  $X$  to a proper subset of  $X$ .
3. Criticize the following proof by induction of the proposition, "Happy families are all alike." Consider a set consisting of one happy family. Obviously all its elements are the same. Suppose it has been shown that for a set of  $n$  happy families, say  $\{f_1, \dots, f_n\}$ , we have  $f_1 = f_2 = \dots = f_n$ . Consider a set  $\{f_1, f_2, \dots, f_{n+1}\}$  of  $n + 1$  happy families. Then  $\{f_1, f_2, \dots, f_n\}$  is a set of  $n$  happy families, so  $f_1 = f_2 = \dots = f_n$ . Similarly,  $\{f_2, f_3, \dots, f_{n+1}\}$  is a set of  $n$  happy families, so  $f_2 = f_3 = \dots = f_{n+1}$ . Thus  $f_{n+1} = f_1$  also, and the set of  $n + 1$  happy families are all alike. By the principle of induction, we see that for any finite set of happy families, they are all alike. Since the set of all happy families is finite, we conclude: happy families are all alike.
4. (The results of this exercise will be used in later chapters.) The binomial coefficients  $\binom{n}{k}$  (pronounced "n choose k") are defined for nonnegative integers  $n$  and  $k$  by the formulas

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad k = 0, 1, \dots, n$$

where  $0! = 1$  and  $n! = n(n-1)!$  for  $n = 1, 2, \dots$ . We put  $\binom{n}{k} = 0$  for  $k > n$ .

(a) Show that

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$$

for  $k = 1, 2, \dots, n$ .

(b) Show that

$$(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k$$

for any nonnegative integer  $n$ , and deduce

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

for any  $a, b$ .

(c) Find

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots$$

and

$$\binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \dots$$

5. Prove by induction that for any positive integers  $a, b, n$ ,

$$\binom{a}{0} \binom{b}{n} + \binom{a}{1} \binom{b}{n-1} + \dots + \binom{a}{n} \binom{b}{0} = \binom{a+b}{n}$$

6. Prove, using induction or otherwise, that

$$\sum_{k=1}^n (2k-1) = n^2, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

7. Show that for each positive integer  $n$ ,

$$2(\sqrt{n}-1) < 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots + \frac{1}{\sqrt{n}} < 2\sqrt{n}.$$

HINT: The identity  $\sqrt{k+1} - \sqrt{k} = (\sqrt{k+1} + \sqrt{k})^{-1}$  might be helpful.

8. Which is larger,  $99^{50} + 100^{50}$  or  $101^{50}$ ? Try to answer without using a calculator.

9. Show that  $\sqrt{10}$  and  $\sqrt{15}$  are irrational. Better yet, show that if  $n \in \mathbb{N}$ , then  $\sqrt{n}$  is either an integer or irrational.

10. Show that if a field  $K$  has characteristic  $p > 0$ , then  $p$  is a prime.

11. Let  $K = \{q + r\sqrt{2} : q, r \in \mathbf{Q}\}$ . Show that  $K$ , with the operations it inherits from  $\mathbf{R}$ , is an ordered field with the Archimedean property.

12. The complex field  $\mathbf{C}$  is defined to be the set  $\mathbf{R} \times \mathbf{R}$  of all ordered pairs of real numbers, together with the operations of addition and multiplication given by the rules

$$(a, b) + (c, d) = (a + c, b + d)$$

and

$$(a, b)(c, d) = (ac - bd, ad + bc).$$

(a) Show that  $\mathbf{C}$  is a field, having a subfield isomorphic to  $\mathbf{R}$ . (We will denote this subfield also by  $\mathbf{R}$ , sowing a crop of confusion that is unlikely to ever be reaped.)

(b) Show that  $\mathbf{C}$  cannot be made into an ordered field.

13. Let  $K$  be an ordered field, in which the Archimedean property is satisfied. Show that if every bounded increasing sequence in  $K$  has a least upper bound in  $K$ , then  $K$  is a complete ordered field.

14. Find a bijective map from  $(0, 1)$  to  $\mathbf{R}$ .

15. Find a bijective map from  $(0, 1)$  to  $[0, 1]$ .

16. Deduce from Theorem 1.23 the slightly stronger result: if  $\alpha \in \mathbf{R}$  is irrational, and  $M$  is any integer, then  $\{n\alpha + m : n > M, m \in \mathbf{Z}\}$  is dense in  $\mathbf{R}$ .

17. Show that for any  $\alpha \in \mathbf{R}$ , there exist infinitely many rational numbers  $m/n$  with  $|\alpha - m/n| < 1/n^2$ . This is a stronger, quantitative, version of the proposition that the rational numbers are dense in  $\mathbf{R}$ .

18. Give an example of each of the following:

(a) Nonempty intervals  $J_n$  in  $\mathbf{R}$  ( $n \in \mathbf{N}$ ) which are bounded, with  $J_{n+1} \subset J_n$  for each  $n$ , and  $\bigcap_{n=1}^{\infty} J_n = \emptyset$ .

(b) Nonempty intervals  $J_n$  in  $\mathbf{R}$  ( $n \in \mathbf{N}$ ) which are closed, with  $J_{n+1} \subset J_n$  for each  $n$ , and  $\bigcap_{n=1}^{\infty} J_n = \emptyset$ .

(c) Nonempty intervals  $J_n = [a_n, b_n]$  in  $\mathbf{Q}$  ( $n \in \mathbf{N}$ ) such that  $J_{n+1} \subset J_n$  for every  $n$ , and  $\bigcap_{n=1}^{\infty} J_n = \emptyset$ .

## 1.11 Notes

1.1 For a closer look at set theory, with an introduction to cardinal and ordinal numbers, I highly recommend Halmos [2]. An important point of view among mathematicians (in fact, the dominant one) believes that

the foundations of mathematics lie in set theory. From this point of view, it is natural to want to construct the set  $\mathbf{N}$  of natural numbers, with its order relation and algebraic structure, from a reasonable set of axioms for set theory. In this book, the natural numbers are taken for granted.

- 1.2 According to Whittaker and Watson [15], even the concept of negative number was rejected by conservative mathematicians at surprisingly late times. They cite authors of works on algebra, trigonometry, etc., in the latter half of the eighteenth century in which the use of negative numbers was disallowed, although Descartes had used them unrestrictedly more than a hundred years before. The notations  $\mathbf{N}$  (for number),  $\mathbf{Q}$  (for quotient), and  $\mathbf{R}$  (for real) seem obvious to English speakers, but  $\mathbf{Z}$  may seem obscure. It derives from the German *Zahl*, meaning number. Its use has become universal.
- 1.3 The ideas of countable and uncountable are due to Cantor, as indeed is the whole idea of set. Theorem 1.4, for instance, is due to Cantor. The problems of infinity of course presented themselves much earlier; Galileo, for instance, discussed the apparent paradox that there are as many even numbers as whole numbers, at the same time that there are only half as many. Cantor, incidentally, was not led to the creation of set theory by the desire for greater abstraction, but through his research on the convergence of trigonometric series.
- 1.4 The discovery of incommensurability, or rather the fact that it had gone so long undiscovered, had a great effect on Plato, for one. I quote from Toeplitz [14]:

Plato puts considerable emphasis on the fundamental nature of this discovery. In the *Laws*, at the point where he assigns that mathematical discovery a place in higher school instruction, he mentions that he first learned of it when he was a comparatively old man and that he had felt ashamed, for himself and for all Greeks, of this ignorance which "befits more the level of swine than of men." (Toeplitz. *Calculus: A Genetic Approach*. Copyright University of Chicago Press, Chicago, 1963)

- 1.5 The concept of field (*Körper* in German, hence the conventional  $K$  or  $k$ ) seems to be due to Dedekind. Theorem 1.10 was probably first proved by Eudoxus, who dealt of course with the prevailing Greek idea of quantity, not quite (though close to) the modern idea of real number. Archimedes employed it, realizing its fundamental importance, and explicitly credited Eudoxus with the theorem. Today, a reference to the property of Eudoxus, as it should be called, would only earn a blank stare.

- 1.6 One imagines that the name absolute value, and the symbol  $|x|$ , go back to ancient times. In fact, they were introduced by Karl Weierstrass in 1859.
- 1.7 Theorem 1.27 is of fundamental significance; it could be used as the basic “completeness” property of the real numbers, instead of the “order completeness” that we have chosen. The uncountability of the real numbers was first proved by Cantor, who gave two different proofs at different times. We have followed his first proof; his other proof uses a different method (the so-called *diagonal argument*).
- 1.8 Theorem 1.31, as already mentioned, is due to Liouville (1851), who first exhibited a transcendental number (the one given, or a close relative), and predated Proposition 1.30, which is due to Cantor. Cantor’s argument, which does not look constructive, can be seen to “exhibit” a transcendental number; for the listing of the algebraic numbers as in the proof of Proposition 1.30 can be used to provide a nested sequence of closed bounded intervals (with rational endpoints) whose intersection contains no algebraic number. The supremum of the left endpoints of these intervals is then a transcendental number which has been “exhibited” in the same sense as the one in Liouville’s example. It was shown by Hermite in 1873 that  $e$  is transcendental, and by Lindemann in 1882 that  $\pi$  is transcendental. (The reader has heard of  $e$  and  $\pi$  elsewhere; we will introduce  $e$  in the next chapter, and  $\pi$  later.) For a proof of Hermite’s theorem, see Herstein [5].
- 1.9 Dedekind developed his construction of real numbers in 1858, though it was not published until 1872, the same year that Cantor published on the same subject. Dedekind declared a real number to be a division (a “Dedekind cut”) of the rational numbers  $\mathbf{Q}$  into two nonempty sets,  $L$  and  $R$  say, with the property that  $q < r$  for every  $q \in L$  and  $r \in R$ ; for convenience, we can also assume that  $L$  has no greatest element. While it must have been obvious from the start that in working with Dedekind cuts it sufficed to consider the left half of the cut, this artifice seems to have been first suggested by Bertrand Russell, better known to the world at large as a philosopher than as a mathematician. Hardy was aware of this, but preferred (in his classic calculus text [3]) to carry on with a cut being defined as two intervals of rationals.

# 2

## Sequences and Series

### 2.1 Sequences

In the first chapter, we defined a sequence in  $X$  to be a mapping from  $\mathbf{N}$  to  $X$ . Let us broaden this definition slightly, and allow the mapping to have a domain of the form  $\{n \in \mathbf{Z} : m \leq n \leq p\}$ , or  $\{n \in \mathbf{Z} : n \geq m\}$ , for some  $m \in \mathbf{Z}$  (usually, but not always,  $m = 0$  or  $m = 1$ ). The most common notation is to write  $n \mapsto x_n$  instead of  $n \mapsto x(n)$ . If the domain of the sequence is the finite set  $\{m, m+1, \dots, p\}$ , we write the sequence as  $(x_n)_{n=m}^p$ , and speak of a finite sequence (though we emphasize that the sequence should be distinguished from the set  $\{x_n : m \leq n \leq p\}$ ). If the domain of the sequence is a set of the form  $\{m, m+1, m+2, \dots\} = \{n \in \mathbf{Z} : n \geq m\}$ , we write it as  $(x_n)_{n=m}^\infty$ , and speak of an infinite sequence. Note that the corresponding set of values  $\{x_n : n \geq m\}$  may be finite. When the domain of the sequence is understood from the context, or is not relevant to the discussion, we write simply  $(x_n)$ . In this chapter, we shall be concerned with infinite sequences in  $\mathbf{R}$ .

**2.1 Definition.** A sequence  $(x_n)_{n=m}^\infty$  in  $\mathbf{R}$  is said to converge to the limit  $x \in \mathbf{R}$ , and we write  $x_n \rightarrow x$  as  $n \rightarrow \infty$ , or  $\lim_{n \rightarrow \infty} x_n = x$ , or simply  $\lim x_n = x$ , if for every  $\epsilon > 0$  there exists  $n_0 \in \mathbf{N}$  such that  $x - \epsilon < x_n < x + \epsilon$  for every  $n \geq n_0$ . A sequence which does not converge to any limit in  $\mathbf{R}$  is said to diverge. We say  $x_n \rightarrow +\infty$ , or  $\lim x_n = +\infty$ , if for every  $M \in \mathbf{R}$  there exists  $n_0 \in \mathbf{N}$  such that  $x_n > M$  for every  $n \geq n_0$ , and we say that  $x_n \rightarrow -\infty$ , or  $\lim x_n = -\infty$ , if for every  $M \in \mathbf{R}$  there exists  $n_0 \in \mathbf{N}$  such that  $x_n < M$  for every  $n \geq n_0$ .

Note that if  $x_n \rightarrow +\infty$ , then  $(x_n)$  diverges. A simple example of a divergent sequence is given by putting  $x_n = (-1)^n$ . We begin with a proposition establishing a few basic rules for dealing with convergent sequences.

**2.2 Proposition.** *Suppose that  $(c_n)$  and  $(d_n)$  are convergent sequences in  $\mathbf{R}$ , say  $\lim_{n \rightarrow \infty} c_n = C$  and  $\lim_{n \rightarrow \infty} d_n = D$ . Let  $a \in \mathbf{R}$ . Then the sequences  $(ac_n)$ ,  $(c_n + d_n)$ , and  $(c_n d_n)$  are all convergent, and in fact*

$$\lim_{n \rightarrow \infty} (ac_n) = aC, \quad \lim_{n \rightarrow \infty} (c_n + d_n) = C + D, \quad \lim_{n \rightarrow \infty} (c_n d_n) = CD.$$

If  $D \neq 0$ , then for some  $k \geq m$ ,  $d_n \neq 0$  for all  $n \geq k$ , and  $(1/d_n)_{n=k}^{\infty}$  converges; in fact  $\lim_{n \rightarrow \infty} 1/d_n = 1/D$ .

**Proof.** Let  $\epsilon > 0$ . If  $a \neq 0$ , there exists  $n_0$  such that  $|c_n - C| < \epsilon/|a|$  for all  $n \geq n_0$ , and then  $|ac_n - aC| = |a||c_n - C| < \epsilon$  for all  $n \geq n_0$ . If  $a = 0$ , this is trivial. Thus  $\lim(ac_n) = aC$ .

There exist  $n_1$  and  $n_2$  such that  $|c_n - C| < \epsilon/2$  for all  $n \geq n_1$  and  $|d_n - D| < \epsilon/2$  for all  $n \geq n_2$ . Let  $n_0 = \max\{n_1, n_2\}$ , and we have

$$\begin{aligned} |(c_n + d_n) - (C + D)| &= |(c_n - C) + (d_n - D)| \\ &\leq |c_n - C| + |d_n - D| < \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

for all  $n \geq n_0$ . Thus  $\lim(c_n + d_n) = C + D$ .

Choose  $M > \max\{|C|, |D|\}$ . Choose  $n_1$  so that  $|c_n - C| < \epsilon/(2M)$  for all  $n \geq n_1$ , and  $n_2$  so that  $|d_n - D| < \epsilon/(2M)$  and  $|d_n - D| < M - |D|$  for all  $n \geq n_2$ . Then  $|d_n| = |d_n - D + D| \leq |d_n - D| + |D| < M$  for all  $n \geq n_2$ . For  $n \geq n_0 = \max\{n_1, n_2\}$ , we have

$$\begin{aligned} |c_n d_n - CD| &= |(c_n - C)d_n + C(d_n - D)| \\ &\leq |c_n - C||d_n| + |C||d_n - D| \\ &\leq M(|c_n - C| + |d_n - D|) \\ &< \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

Thus  $\lim(c_n d_n) = CD$ .

Finally, choose  $k$  so that  $|d_n - D| < \frac{1}{2}|D|$  for  $n \geq k$ ; it follows that  $|d_n| > \frac{1}{2}|D|$  for all  $n \geq k$ . Next choose  $n_0 > k$  so that  $|d_n - D| < |D|^2\epsilon/2$  for all  $n \geq n_0$ . We have then for  $n \geq n_0$

$$\left| \frac{1}{d_n} - \frac{1}{D} \right| = \left| \frac{D - d_n}{d_n D} \right| \leq \frac{2}{|D|^2} |d_n - D| < \epsilon.$$

Thus  $\lim(1/d_n) = 1/D$ . ■

This proposition will be used quite frequently in the future, usually without explicit citation. Another, even simpler, fact is the following:

**2.3 Proposition.** If  $(x_n)_{n=m}^{\infty}$  is a convergent sequence in  $\mathbf{R}$ , and for some  $k \in \mathbf{N}$  we have  $x_n \geq 0$  for all  $n \geq k$ , then  $\lim x_n \geq 0$ . Similarly, if  $a \leq x_n \leq b$  for some  $a, b \in \mathbf{R}$  and all  $n \geq k$ , then  $a \leq \lim x_n \leq b$ .

**Proof.** If  $x_n \geq 0$  for all  $n \geq k$ , and  $L < 0$ , taking  $\epsilon = |L|$  in the definition of limit shows that  $L$  cannot be the limit of  $(x_n)$ . The second statement follows from the first by considering the sequences  $(x_n - a)$  and  $(b - x_n)$  and applying the last proposition. ■

**2.4 Example.** Let

$$s_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

for each  $n \in \mathbf{N}$ . Suppose that  $(s_n)$  converges, say  $\lim s_n = s$ . It is clear from the definition of convergence that if  $s_n \rightarrow s$ , then also  $s_{2n} \rightarrow s$ , but we observe that

$$s_{2n} - s_n = \frac{1}{n+1} + \cdots + \frac{1}{2n} > \frac{1}{2n} + \cdots + \frac{1}{2n} = \frac{1}{2},$$

so we would have  $0 = \lim(s_{2n} - s_n) \geq 1/2$ , using the last two propositions. This contradiction shows that  $(s_n)$  does not converge.

The next proposition presents a few examples of convergent sequences. We will use two simple inequalities, which can be proved directly by induction on  $n$ , or (for  $t \geq 0$ ) seen immediately from the binomial theorem.

**2.5 Lemma.** For every real  $t > -1$ , and every  $n \in \mathbf{N}$ , we have

$$(1+t)^n \geq 1+nt, \quad (1+t)^n \geq 1+nt+\frac{1}{2}n(n-1)t^2;$$

when  $t \neq 0$ , the inequalities are strict for  $n > 1$ ,  $n > 2$ , respectively.

The proof is left as an exercise.

**2.6 Proposition.** If  $a > 0$ , then  $na \rightarrow +\infty$  as  $n \rightarrow \infty$ , and  $a/n \rightarrow 0$  as  $n \rightarrow \infty$ ; if  $a > 1$ , then  $a^n \rightarrow +\infty$ , and if  $0 < b < 1$ , then  $b^n \rightarrow 0$ , as  $n \rightarrow \infty$ . If  $a > 0$ , then  $\lim_{n \rightarrow \infty} a^{1/n} = 1$ . Finally,  $\lim_{n \rightarrow \infty} n^{1/n} = 1$ .

**Proof.** That  $na \rightarrow +\infty$  as  $n \rightarrow \infty$  is simply a rephrasing of the Archimedean property of  $\mathbf{R}$ , i.e., of Theorem 1.19.

If  $\epsilon > 0$ , then Theorem 1.19 asserts the existence of  $n_0$  such that  $n_0\epsilon > a$ , and hence  $n\epsilon > a$  for all  $n \geq n_0$ , which gives  $0 < a/n < \epsilon$  for all  $n \geq n_0$ , so  $\lim_{n \rightarrow \infty} (a/n) = 0$ .

If  $a > 1$ , let  $\delta = a - 1$ , so  $\delta > 0$ ; then  $a^n = (1+\delta)^n > 1+n\delta$ , so  $a^n \rightarrow +\infty$  since  $n\delta \rightarrow +\infty$ .

If  $0 < b < 1$ , let  $a = 1/b$ , so  $a > 1$ . If  $\epsilon > 0$ , then there exists  $n_0$  such that  $a^n > 1/\epsilon$  for  $n \geq n_0$ , which is equivalent to  $0 < b^n < \epsilon$  for  $n \geq n_0$ , so  $\lim_{n \rightarrow \infty} b^n = 0$ .



For  $n > 1$ , let  $\delta_n = n^{1/n} - 1$ ; clearly,  $\delta_n > 0$ . Then

$$n = (1 + \delta_n)^n > 1 + n\delta_n + \frac{1}{2}n(n-1)\delta_n^2 > \frac{1}{2}n(n-1)\delta_n^2,$$

from which we conclude that  $0 < \delta_n^2 < 2/(n-1)$ . Given  $\epsilon > 0$ , choose  $n_0 \in \mathbf{N}$  with  $n_0 > 1 + 2/\epsilon^2$ . Then  $1 < n^{1/n} < 1 + [2/(n-1)]^{1/2} < 1 + \epsilon$  for every  $n \geq n_0$ . Thus  $\lim_{n \rightarrow \infty} n^{1/n} = 1$ .

If  $a \geq 1$ , then  $1 \leq a^{1/n} \leq n^{1/n}$  for every  $n > a$ , so  $\lim a^{1/n} = 1$ . Finally, if  $0 < a < 1$ , let  $b = 1/a$ , so  $\lim a^{1/n} = 1/\lim b^{1/n} = 1$ . ■

**2.7 Example.** Fix the integer  $b \geq 2$ , and  $x \in \mathbf{R}$ ,  $0 \leq x < 1$ . Define  $d_1 = [bx]$ , so  $d_1$  is an integer with  $0 \leq d_1 < b$ , and put  $x_1 = d_1/b$ , so  $0 \leq x - x_1 < 1/b$ ; having obtained such integers  $d_1, \dots, d_k$ , and numbers  $x_1, \dots, x_k$ , so that  $0 \leq x - x_k < b^{-k}$ , we proceed to set  $d_{k+1} = [b^{k+1}(x - x_k)]$  and  $x_{k+1} = x_k + d_{k+1}b^{-k-1}$ , so that inductively we obtain a sequence  $(d_n)_{n=1}^\infty$  such that  $0 \leq x - x_n < b^{-n}$ , where each  $d_n$  is an element of  $\{0, 1, \dots, b-1\}$  and  $x_n = d_1b^{-1} + d_2b^{-2} + \dots + d_nb^{-n}$ . If  $x = x_n$  for some  $n$ , then  $d_m = 0$  for all  $m > n$ . We write symbolically

$$x = 0.d_1d_2\dots,$$

where each  $d_j$  is an integer,  $0 \leq d_j < b$ , to mean that the sequence  $(x_k)$ , where  $x_k = d_1b^{-1} + \dots + d_kb^{-k}$ , converges to  $x$ . This is called a representation, or expansion, of  $x$  in the base  $b$ . When  $b = 10$  this is called the decimal expansion, when  $b = 2$  the binary expansion, when  $b = 16$  the hexadecimal expansion. We have seen that any  $x \in [0, 1)$  admits a representation in any base  $b$ . We observe that if  $x$  is a rational that can be expressed in the form  $p/b^n$  for some integers  $p$  and  $n$ , there exist two representations of  $x$  in the base  $b$ ; one has the form  $0.d_1d_2\dots d_n000\dots$ , and the other looks like  $0.d_1d_2\dots(d_n-1)(b-1)(b-1)(b-1)\dots$ . The above procedure produces the “terminating” expansion, when there is a choice, e.g., with  $b = 10$  the expansion of  $x = \frac{1}{2}$  is  $0.5000\dots$  rather than  $0.4999\dots$ . If  $x$  does not have the form  $p/b^n$ , the expansion is unique. We leave this fact as an exercise.

**2.8 Example.** Theorem 1.25 demonstrates that every positive real number has a square root, but does not show how to find it. Indeed, what does it mean to “find”  $\sqrt{2}$ , for instance? We know it is not a rational number. We have seen that each real number has a decimal expansion, so one answer to the question would be to display the decimal expansion of  $\sqrt{2}$ , perhaps by finding an explicit formula for the integer  $d_n$  in the  $n$ th place after the decimal point, or perhaps by finding an inductive procedure for determining  $d_n$ . Now we can interpret the truncated finite decimal expansions as being a sequence of rational numbers which converge to the number represented by the infinite decimal expansion. Any other sequence of rational numbers converging to  $\sqrt{2}$  would be just as good in principle, and possibly better in

practice, in that the  $n$ th term in the sequence might be much closer to  $\sqrt{2}$  than the decimal fraction  $1.d_1d_2\dots d_n$ . Here is one such sequence, which represents the oldest known method for computing square roots.

Given  $a > 0$ , we define a sequence  $(x_n)$  inductively, as follows. Choose  $x_0 > 0$  and let  $x_{n+1} = \frac{1}{2}(x_n + a/x_n)$  for  $n \geq 0$ . It is clear from the definition of convergence that if  $x_n \rightarrow x$ , then also  $x_{n+1} \rightarrow x$ , so that  $x = \frac{1}{2}(x + a/x)$ , or  $2x^2 = x^2 + a$ , that is  $x^2 = a$ . It is also clear that  $x_n > 0$  for all  $n$ , so  $x \geq 0$ : thus  $x = \sqrt{a}$ . To show that in fact  $(x_n)$  does converge, we calculate as follows:

$$\begin{aligned} x_{n+1} - \sqrt{a} &= \frac{1}{2} \left( x_n + \frac{a}{x_n} \right) - \sqrt{a} \\ &= \frac{x_n^2 + a - 2x_n\sqrt{a}}{2x_n} = \frac{(x_n - \sqrt{a})^2}{2x_n}. \end{aligned}$$

In particular, we see that  $x_{n+1} \geq \sqrt{a}$  for all  $n$ , and  $x_n^2 \geq a$  for every  $n \geq 1$ . Similarly, we find that

$$x_{n+1} + \sqrt{a} = \frac{(x_n + \sqrt{a})^2}{2x_n},$$

so

$$\frac{x_{n+1} - \sqrt{a}}{x_{n+1} + \sqrt{a}} = \frac{(x_n - \sqrt{a})^2}{(x_n + \sqrt{a})^2}$$

By induction, it follows that for every  $n$

$$\frac{x_n - \sqrt{a}}{x_n + \sqrt{a}} = \left( \frac{x_0 - \sqrt{a}}{x_0 + \sqrt{a}} \right)^{2^n}. \quad (2.1)$$

Now  $|x_0 - \sqrt{a}| < x_0 + \sqrt{a}$ , since  $x_0 > 0$  and  $\sqrt{a} > 0$ . Thus equation (2.1) implies that  $x_n \rightarrow \sqrt{a}$  as  $n \rightarrow \infty$ , and indeed, that the convergence is quite rapid. We note incidentally that

$$\begin{aligned} x_n - x_{n+1} &= x_n - \frac{1}{2}(x_n + a/x_n) = \frac{1}{2}(x_n - a/x_n) \\ &= \frac{1}{2}(x_n^2 - a)/x_n \geq 0, \end{aligned}$$

so that  $x_{n+1} \leq x_n$  for all  $n \geq 1$ . In particular,  $x_n + \sqrt{a} \leq x_1 + \sqrt{a} \leq 2x_1$  for every  $n \geq 1$ . From this, and equation (2.1), we can write down the estimate

$$0 \leq x_n - \sqrt{a} \leq 2x_1 \left( \frac{x_0 - \sqrt{a}}{x_0 + \sqrt{a}} \right)^{2^n}. \quad (2.2)$$

Thus this method seems to be a highly efficient method for calculating square roots to any given degree of accuracy. For instance, with  $a = 10$ , we might take a rough first guess  $x_0 = 3$ . We see easily that  $3 < \sqrt{10} < 3.5$ , so

$|x_0 - \sqrt{a}| < 1/2$ , and  $x_0 + \sqrt{a} > 6$ . Thus the estimate (2.2) gives us (after we calculate  $2x_1 = 19/3 < 7$ )

$$0 < x_n - \sqrt{a} < 7 \cdot 12^{-2^n}$$

Thus  $x_4$  approximates  $\sqrt{10}$  to an accuracy better than  $10^{-15}$ , which should be good enough for household use.

**2.9 Definition.** Let  $(a_n)$  be a sequence in  $\mathbf{R}$ . We say that  $(a_n)$  is an increasing sequence if  $a_n \leq a_{n+1}$  for all  $n$ ; we call it strictly increasing if  $a_n < a_{n+1}$  for all  $n$ . Similarly,  $(a_n)$  is called decreasing (or strictly decreasing) if  $a_n \geq a_{n+1}$  (resp.,  $a_n > a_{n+1}$ ) for all  $n$ . A sequence is called monotone if it is either increasing or decreasing.

**2.10 Proposition.** If  $(a_n)_{n=m}^{\infty}$  is an increasing sequence in  $\mathbf{R}$ , then either  $(a_n)$  is convergent, or  $a_n \rightarrow +\infty$  as  $n \rightarrow \infty$ .

**Proof.** If  $(a_n)$  is not bounded above, for every  $M$  there exists  $k$  such that  $a_k > M$ ; since  $(a_n)$  is increasing, we have  $a_n \geq a_k > M$  for every  $n \geq k$ . Thus  $a_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Suppose  $(a_n)$  is bounded above; then there exists a least upper bound  $M$  for the set  $\{a_n : n \geq m\}$ . Given any  $\epsilon > 0$ ,  $M - \epsilon$  is not an upper bound for  $\{a_n : n \geq m\}$ . Thus there exists  $k$  with  $a_k > M - \epsilon$ ; since  $(a_n)$  is increasing, it follows that  $M - \epsilon < a_n \leq M$  for every  $n \geq k$ . Thus,  $a_n \rightarrow M$  as  $n \rightarrow \infty$ . ■

**2.11 Corollary.** If  $(a_n)$  is a decreasing sequence in  $\mathbf{R}$ , then either  $(a_n)$  is convergent, or  $(a_n) \rightarrow -\infty$  as  $n \rightarrow \infty$ .

Thus every bounded monotone sequence converges.

**2.12 Example.** Let  $c_n = (1 + 1/n)^n$  for each  $n \in \mathbf{N}$ . Then  $(c_n)$  is an increasing sequence. One way to see this is to recall the identity

$$b^r - a^r = (b - a)(b^{r-1} + b^{r-2}a + \cdots + ba^{r-2} + a^{r-1})$$

valid for any  $a, b \in \mathbf{R}$  (or any field) and any positive integer  $r$ , which we used to advantage in the first chapter. It gives the inequality

$$b^r < a^r + r(b - a)b^{r-1} \quad (2.3)$$

whenever  $0 < a < b$ . Taking  $a = 1 + 1/(n + 1)$ ,  $b = 1 + 1/n$ , and using  $r = n + 1$  in the inequality (2.3), we get

$$\left(1 + \frac{1}{n}\right)c_n = b^{n+1} < a^{n+1} + \frac{(n+1)b^n}{n(n+1)} = c_{n+1} + \frac{1}{n}c_n,$$

from which  $c_n < c_{n+1}$  follows at once. Furthermore, the sequence  $(c_n)$  is bounded; in fact, taking  $a = 1$  and  $b = 1 + 1/(2n)$  with  $r = n$  in the inequality (2.3), we get

$$\left(1 + \frac{1}{2n}\right)^n < 1 + n \frac{1}{2n} \left(1 + \frac{1}{2n}\right)^{n-1} < 1 + \frac{1}{2} \left(1 + \frac{1}{2n}\right)^n,$$

from which we get

$$\left(1 + \frac{1}{2n}\right)^n < 2,$$

so  $c_{2n} < 4$ . Since  $(c_n)$  is an increasing sequence, it follows that  $c_n < 4$  for all  $n$ . Hence, by Proposition 2.10 it follows that  $(c_n)$  converges. We denote the limit by  $e$ . We have not seen the last of this number.

The notion of *subsequence* is fairly natural, but let us give a formal definition.

**2.13 Definition.** A sequence  $(b_n)_{n=p}^{\infty}$  is said to be a subsequence of the sequence  $(a_n)_{n=m}^{\infty}$  if there exists a strictly increasing sequence  $(n_k)_{k=p}^{\infty}$  in  $\mathbf{Z}$ , such that  $b_k = a_{n_k}$  for every  $k \geq p$ .

Let  $(a_n)_{n=k}^{\infty}$  be a bounded sequence. For each  $m \geq k$ , let us set  $b_m = \sup_{n \geq m} a_n = \sup\{a_m, a_{m+1}, a_{m+2}, \dots\}$ . Clearly,  $(b_m)_{m=k}^{\infty}$  is a decreasing sequence; each  $b_m \in \mathbf{R}$  since  $(a_n)$  is bounded above, and the sequence  $(b_m)$  is bounded below, since the sequence  $(a_n)$  is bounded below. By the Corollary above, the sequence  $(b_n)$  is convergent; we call this limit the *limes superior*, or *upper limit*, of the sequence  $(a_n)$ . The *limes inferior* or *lower limit* is defined in an analogous way. If  $(a_n)$  is not bounded above, we put its upper limit to be  $+\infty$ , and if it is not bounded below, its lower limit is said to be  $-\infty$ . We summarize:

**2.14 Definition.** If  $(a_n)_{n=k}^{\infty}$  is any sequence in  $\mathbf{R}$ , we define

$$\begin{aligned} \limsup_{n \rightarrow \infty} a_n &= \inf_{m \geq k} \sup_{n \geq m} a_n = \lim_{m \rightarrow \infty} \sup_{n \geq m} a_n \\ \liminf_{n \rightarrow \infty} a_n &= \sup_{m \geq k} \inf_{n \geq m} a_n = \lim_{m \rightarrow \infty} \inf_{n \geq m} a_n \end{aligned}$$

One often sees the symbol  $\overline{\lim}$  used for  $\limsup$ , and  $\underline{\lim}$  for  $\liminf$ . We can describe the upper and lower limits also in the following way:

**2.15 Proposition.** Let  $(a_n)$  be a sequence in  $\mathbf{R}$ . The following are equivalent:

- (a)  $\limsup a_n = A$ ; and
- (b) for every  $A' > A$ ,  $a_n < A'$  for all but finitely many  $n$ ; for every  $A'' < A$ ,  $a_n > A''$  for infinitely many  $n$ .

**Proof.** Suppose  $A = \limsup a_n$ . Then for any  $A' > A$ , there exists  $m$  such that  $\sup_{n \geq m} a_n < A'$ , so, in particular,  $a_n < A'$  for all  $n \geq m$ . But since  $\sup_{n \geq m} a_n \geq A$  for every  $m$ , it follows that if  $A'' < A$ , then for every  $m$  there exists  $n \geq m$  with  $a_n > A''$ . Thus (a) implies (b). Now suppose (b) holds. Then for every  $A' > A$ , there exists  $m$  such that  $a_n < A'$  for all  $n \geq m$ , and hence  $\sup_{n \geq m} a_n \leq A'$ ; it follows that  $\limsup a_n \leq A'$  for every  $A' > A$ , and hence that  $\limsup a_n \leq A$ . On the other hand, for any  $A'' < A$ , (b) assures us that for every  $m$  there exists  $n \geq m$  with  $a_n > A''$ . It follows that for every  $m$ ,  $\sup_{n \geq m} a_n > A''$ , and hence that  $\limsup a_n \geq A''$ . Since this holds for every  $A'' < A$ , it follows that  $\limsup a_n \geq A$ . Thus  $\limsup a_n = A$ . We have proved (b) implies (a). ■

We leave it to the reader to formulate and prove the corresponding characterization of the lower limit.

Some of the basic properties of the upper and lower limits are summarized in the next proposition.

**2.16 Proposition.** *If  $(a_n)$  and  $(b_n)$  are sequences in  $\mathbf{R}$ , then:*

- (a)  $\limsup(-a_n) = -\liminf a_n$ ;
- (b)  $\limsup ca_n = c \limsup a_n$  for any  $c > 0$ ;
- (c)  $\limsup(a_n + b_n) \leq \limsup a_n + \limsup b_n$ ;
- (d)  $\liminf a_n \leq \limsup a_n$ , with equality if and only if  $(a_n)$  is convergent, in which case  $\limsup a_n = \lim a_n$ ; and
- (e) if  $(b_n)$  is a subsequence of  $(a_n)$ , then

$$\liminf a_n \leq \liminf b_n \leq \limsup b_n \leq \limsup a_n.$$

The proof of this proposition is left as an exercise.

**2.17 Theorem.** *Every bounded sequence in  $\mathbf{R}$  has a convergent subsequence.*

**Proof.** Let  $(a_n)$  be a bounded sequence in  $\mathbf{R}$ . Let  $A = \limsup a_n$ . We shall construct a subsequence of  $(a_n)$  which converges to  $A$ . By Proposition 2.15 there exists  $n_1$  such that  $a_{n_1} > A - 1$ . Having obtained  $n_1 < n_2 < \dots < n_k$  such that  $a_{n_j} > A - 1/j$  for  $j = 1, 2, \dots, k$ , we can find (by Proposition 2.15)  $n_{k+1} > n_k$  such that  $a_{n_{k+1}} > A - 1/(k+1)$ , thus defining the subsequence  $(a_{n_k})$  inductively. We have  $A \leq \liminf a_{n_k} \leq \limsup a_{n_k} \leq \limsup a_n = A$ , so  $\lim a_{n_k} = A$ . ■

This very important theorem is known as the Bolzano–Weierstrass theorem. Of course, we could equally well have constructed a subsequence converging to  $B = \liminf a_n$ . According to Proposition 2.16(e), any convergent subsequence would have a limit between  $A$  and  $B$ .

**2.18 Definition.** Let  $(a_n)$  be a sequence in  $\mathbf{R}$ . We say that  $(a_n)$  is a Cauchy sequence if for every  $\epsilon > 0$  there exists  $n_0$  such that  $|a_n - a_m| < \epsilon$  for every  $n, m \geq n_0$ .

The next result is known as the Cauchy criterion for convergence.

**2.19 Theorem.** A sequence in  $\mathbf{R}$  is convergent if and only if it is a Cauchy sequence.

**Proof.** Suppose  $(a_n)$  converges to  $A \in \mathbf{R}$ . Then for every  $\epsilon > 0$ , there exists  $n_0$  such that  $|a_n - A| < \epsilon/2$  for every  $n \geq n_0$ . Then for every  $n, m \geq n_0$  we have  $|a_n - a_m| = |a_n - A + A - a_m| \leq |a_n - A| + |a_m - A| < \epsilon$ , so  $(a_n)$  is a Cauchy sequence. Less trivial is the converse. If  $(a_n)$  is a Cauchy sequence, then for each  $\epsilon > 0$  there exists  $n_0$  such that  $|a_n - a_m| < \epsilon$  for every  $n \geq n_0, m \geq n_0$ . Then  $a_{n_0} - \epsilon < a_n < a_{n_0} + \epsilon$  whenever  $n \geq n_0$ , so

$$a_{n_0} - \epsilon \leq \inf_{n \geq n_0} a_n \leq \liminf a_n \leq \limsup a_n \leq \sup_{n \geq n_0} a_n \leq a_{n_0} + \epsilon,$$

which gives  $\limsup a_n - \liminf a_n \leq 2\epsilon$ . Since  $\epsilon > 0$  was arbitrary, it follows that  $\limsup a_n = \liminf a_n$ , so  $(a_n)$  is convergent. ■

## 2.2 Continued Fractions

This section is not needed for subsequent developments, and may be omitted in a first reading.

Let  $a_0$  be an integer, and let  $a_n \in \mathbf{N}$  for each  $n \in \mathbf{N}$ . The expression

$$a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_n}}}$$

is called a *continued fraction* of order  $n$ ; we shall denote it more concisely as  $[a_0; a_1, \dots, a_n]$ . Such an expression makes sense more generally if  $a_j$  are any real numbers, as long as  $a_j \neq 0$  for  $j > 0$ . We call  $[a_0; a_1, a_2, \dots]$  an infinite continued fraction; it denotes the sequence of continued fractions of order  $n$  described above. We obtain an explicit representation of the finite continued fractions as quotients of integers as follows. Define sequences  $(p_n)$  and  $(q_n)$  by putting  $p_{-1} = 1, q_{-1} = 0, p_0 = a_0, q_0 = 1$ , and define inductively for  $n \geq 1$

$$p_n = a_n p_{n-1} + p_{n-2}, \tag{2.4}$$

$$q_n = a_n q_{n-1} + q_{n-2}, \tag{2.5}$$

so that each  $p_n$  and  $q_n$  is a positive integer when each  $a_n$  is a positive integer. We observe that

$$\frac{p_0}{q_0} = a_0, \quad \frac{p_1}{q_1} = \frac{a_1 a_0 + 1}{a_1 + 0} = a_0 + \frac{1}{a_1},$$

and indeed we have in general:

**2.20 Lemma.** For every  $n \geq 0$ , and any positive reals  $a_1, a_2, \dots, a_n$ ,

$$\frac{p_n}{q_n} = [a_0; a_1, a_2, \dots, a_n].$$

**Proof.** We just observed the formula holds for  $n = 0$  and  $n = 1$ . Assume it holds for a particular  $n$ . Then

$$\begin{aligned} [a_0; a_1, \dots, a_n, a_{n+1}] &= [a_0; a_1, \dots, a_{n-1}, a_n + 1/a_{n+1}] \\ &= \frac{(a_n + 1/a_{n+1})p_{n-1} + p_{n-2}}{(a_n + 1/a_{n+1})q_{n-1} + q_{n-2}} \\ &= \frac{p_n + p_{n-1}/a_{n+1}}{q_n + q_{n-1}/a_{n+1}} = \frac{a_{n+1}p_n + p_{n-1}}{a_{n+1}q_n + q_{n-1}} \\ &= \frac{p_{n+1}}{q_{n+1}}, \end{aligned}$$

completing the induction. ■

**2.21 Lemma.** For each  $n \geq 0$ , we have  $p_{n-1}q_n - q_{n-1}p_n = (-1)^n$ .

**Proof.** Using the defining equations, we have

$$\begin{aligned} p_{n-1}q_n - q_{n-1}p_n &= p_{n-1}(a_n q_{n-1} + q_{n-2}) - q_{n-1}(a_n p_{n-1} + p_{n-2}) \\ &= -(p_{n-2}q_{n-1} - q_{n-2}p_{n-1}) \\ &\vdots \\ &= (-1)^n(p_{-1}q_0 - q_{-1}p_0) = (-1)^n. \end{aligned} \quad \blacksquare$$

**2.22 Corollary.** If  $a_n \in \mathbb{N}$  for each  $n$ , then the positive integers  $p_n$  and  $q_n$  have no common divisor greater than 1.

**2.23 Corollary.** For each  $n \geq 1$ ,

$$\left| \frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} \right| = \frac{1}{q_n q_{n-1}}.$$

**2.24 Lemma.** If  $a_n \in \mathbb{N}$  for each  $n$ , then  $q_n > (\sqrt{2})^{n-1}$  for each  $n > 0$ .

**Proof.** Since  $q_n = a_n q_{n-1} + q_{n-2}$  and  $a_n \geq 1$ , it follows that  $q_n > q_{n-1}$  for all  $n$ , and hence that  $q_n > 2q_{n-2}$ . Iterating this estimate, we arrive at  $q_{2n} > 2^n q_0 = 2^n$ , and  $q_{2n+1} > 2^n q_1 \geq 2^n$ , and the lemma follows. ■

**2.25 Theorem.** For any sequence  $(a_n)_{n=1}^{\infty}$  in  $\mathbf{N}$ , and any  $a_0 \in \mathbf{Z}$ , the infinite continued fraction  $[a_0; a_1, a_2, \dots]$  converges.

**Proof.** By Corollary 2.23 and Lemma 2.24, we see that

$$\left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| = \frac{1}{q_{n+1}q_n} < 2^{-(n-1)},$$

which implies that  $(p_n/q_n)$  is a Cauchy sequence, hence convergent. ■

**2.26 Theorem.** For every irrational real number  $x$ , there exists a unique infinite continued fraction which converges to  $x$ .

**Proof.** Let  $a_0 = [x]$ , the greatest integer not greater than  $x$ , so  $0 < x - a_0 < 1$ . Define  $r_1 = (x - a_0)^{-1}$ , so  $r_1 > 1$ , and let  $a_1 = [r_1]$ , so  $a_1$  is a positive integer. Having defined  $r_n$  and  $a_n = [r_n]$ , we define  $r_{n+1} = (r_n - a_n)^{-1}$  and put  $a_{n+1} = [r_{n+1}]$ . Since  $x$  is irrational, we see that every  $r_n$  is irrational, so  $0 < r_n - a_n < 1$ , and  $r_{n+1}$  is well-defined, with  $r_{n+1} > 1$ . Now

$$x = [a_0; a_1, a_2, \dots, a_{n-1}, r_n] \quad (2.6)$$

for every  $n \geq 1$ . Indeed,  $x = a_0 + 1/r_1 = [a_0; r_1]$ , and if the formula (2.6) holds for some  $n$ , then using  $r_n = a_n + 1/r_{n+1}$ , we get equation (2.6) with  $n$  replaced by  $n + 1$ , so that (2.6) holds for all  $n$ . Now from equation (2.6) and Lemma 2.20 we have

$$x = \frac{r_n p_{n-1} + p_{n-2}}{r_n q_{n-1} + q_{n-2}}$$

for  $n \geq 2$ . Correspondingly, we have the formula

$$\frac{p_n}{q_n} = \frac{a_n p_{n-1} + p_{n-2}}{a_n q_{n-1} + q_{n-2}},$$

so we have, after simplifying,

$$x - \frac{p_n}{q_n} = \frac{(r_n - a_n)(p_{n-1}q_{n-2} - q_{n-1}p_{n-2})}{(r_n q_{n-1} + q_{n-2})(a_n q_{n-1} + q_{n-2})}$$

so that

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2} < \frac{1}{2^{n-1}},$$

the last inequality coming from Lemma 2.24. Thus,  $p_n/q_n \rightarrow x$ , and the existence is proved. To show uniqueness, suppose that we had two continued fraction representations

$$x = [a_0; a_1, a_2, \dots] = [a'_0; a'_1, a'_2, \dots].$$



Clearly,  $a_0 = a'_0 = [x]$ . Suppose that we have established  $a_j = a'_j$  for  $j = 0, 1, \dots, n$ . Then (with an obvious notation)  $p_j = p'_j$  and  $q_j = q'_j$  for  $0 \leq j \leq n$  as well. Putting  $r_k = [a_k; a_{k+1}, a_{k+2}, \dots]$ , with the analogous definition of  $r'_k$ , we have

$$x = [a_0; a_1, a_2, \dots, a_n, r_{n+1}] = [a'_0; a'_1, a'_2, \dots, a'_n, r'_{n+1}],$$

so that we have

$$x = \frac{p_n r_{n+1} + p_{n-1}}{q_n r_{n+1} + q_{n-1}} = \frac{p'_n r'_{n+1} + p'_{n-1}}{q'_n r'_{n+1} + q'_{n-1}} = \frac{p_n r'_{n+1} + p_{n-1}}{q_n r'_{n+1} + q_{n-1}},$$

from which we can solve to get  $r_{n+1} = r'_{n+1}$ , and hence  $a_{n+1} = [r_{n+1}] = [r'_{n+1}] = a'_{n+1}$ . Thus  $a_n = a'_n$  for all  $n$ . ■

It is not hard to see that if  $x$  was rational, the process described above would terminate after finitely many steps, and  $x$  would be represented by a finite continued fraction.

Note that Theorems 2.25 and 2.26 establish a one-to-one correspondence between the set of irrational numbers in  $(0, 1)$  and the set  $\mathbf{N}^{\mathbf{N}}$  of mappings of  $\mathbf{N}$  into  $\mathbf{N}$ . But we also have seen (the binary expansion) that there is an injective mapping of the irrational numbers in  $(0, 1)$  into the set  $2^{\mathbf{N}}$  of mappings of  $\mathbf{N}$  into  $\{0, 1\}$ . Thus there is an injective mapping of  $\mathbf{N}^{\mathbf{N}}$  into  $2^{\mathbf{N}}$ .

## 2.3 Infinite Series

Given a sequence  $(a_n)_{n=m}^{\infty}$  in  $\mathbf{R}$ , consider the associated sequence  $(s_n)_{n=m}^{\infty}$  defined by

$$s_n = a_m + a_{m+1} + \dots + a_n = \sum_{k=m}^n a_k;$$

the number  $s_n$  is called the  $n$ th partial sum of the infinite series  $\sum_{n=m}^{\infty} a_n$ , and we say that the series converges if the sequence  $(s_n)$  converges. If  $\lim_{n \rightarrow \infty} s_n = s$ , we write  $\sum_{n=m}^{\infty} a_n = s$ . (Actually, this is an abuse of language, since, properly,  $\sum_{n=m}^{\infty} a_n$  denotes the sequence  $(s_n)_{n=m}^{\infty}$ , rather than its limit.) We also write  $\sum_{n=m}^{\infty} a_n = \pm\infty$  if  $\lim_{n \rightarrow \infty} s_n = \pm\infty$ . Thus the notion of infinite series is totally equivalent to that of infinite sequence; given any sequence  $(s_n)_{n=m}^{\infty}$ , we can realize it as the sequence of partial sums of the series  $\sum_{n=m}^{\infty} a_n$ , where  $a_m = s_m$  and  $a_n = s_n - s_{n-1}$  for  $n > m$ . We shall often omit the limits of summation, i.e., write  $\sum a_n$  instead of  $\sum_{n=m}^{\infty} a_n$ , when  $m$  is either understood from the context or irrelevant.

**2.27 Proposition.** *If  $\sum a_n$  converges, then  $\lim a_n = 0$ .*

**Proof.** If  $s_n$  denotes the  $n$ th partial sum of the series  $\sum a_n$ , then the convergence of  $(s_n)$  to the limit  $s$  implies the convergence of  $(s_{n-1})$  to the same limit, which implies  $a_n = s_n - s_{n-1} \rightarrow 0$  as  $n \rightarrow \infty$ . ■

The necessary condition  $a_n \rightarrow 0$  is not sufficient for the series  $\sum a_n$  to converge; Example 2.4 shows that the series  $\sum_{n=1}^{\infty} (1/n)$  diverges.

**2.28 Proposition.** *If the series  $\sum a_n$  and  $\sum b_n$  converge, then so does the series  $\sum (a_n + b_n)$ ; for any  $c \in \mathbf{R}$  the series  $\sum ca_n$  converges, and we have*

$$\sum (a_n + b_n) = \sum a_n + \sum b_n, \quad \sum ca_n = c \sum a_n.$$

**Proof.** This follows immediately from the two corresponding facts about sequences. ■

**2.29 Proposition.** *Suppose that  $a_n \geq 0$  for each  $n \in \mathbf{N}$ . Then either  $\sum_{n=1}^{\infty} a_n = +\infty$  or the series converges.*

**Proof.** If each  $a_n \geq 0$ , then  $(s_n)$  is an increasing sequence; thus this proposition simply echoes Proposition 2.10. ■

In view of the last proposition, we can indicate the convergence of the series  $\sum a_n$  when every  $a_n \geq 0$  by writing  $\sum a_n < \infty$ .

**2.30 Corollary.** *If  $0 \leq b_n \leq a_n$  for every  $n$ , and if  $\sum_{n=1}^{\infty} a_n$  converges, then  $\sum_{n=1}^{\infty} b_n$  converges.*

**Proof.** If  $s_n = \sum_{k=1}^n a_k$  and  $t_n = \sum_{k=1}^n b_k$ , then evidently  $t_n \leq s_n$  for every  $n$ , so  $(t_n)$  is bounded if  $(s_n)$  is bounded. ■

This last result is often used in the equivalent form: if  $\sum_{n=1}^{\infty} b_n$  diverges, then so does  $\sum_{n=1}^{\infty} a_n$ .

**2.31 Corollary.** *If  $\sum |a_n|$  converges, then  $\sum a_n$  converges.*

**Proof.** Recall the notation  $x^+ = \max\{x, 0\}$  and  $x^- = \max\{-x, 0\}$  for any  $x \in \mathbf{R}$ . Since  $0 \leq a_n^+ \leq |a_n|$  and  $0 \leq a_n^- \leq |a_n|$ , we conclude from Corollary 2.30 that  $\sum a_n^+$  and  $\sum a_n^-$  converge. But  $a_n = a_n^+ - a_n^-$ , so it follows from Proposition 2.28 that  $\sum a_n$  converges. ■

A series  $\sum a_n$  with the property that  $\sum |a_n| < +\infty$  is said to be *absolutely convergent*. The last corollary then states simply that an absolutely convergent series is convergent. A series which is convergent but not absolutely convergent is called *conditionally convergent*. The next corollary is a trivial generalization of Corollary 2.30. We will refer to it as the *comparison test*.

**2.32 Corollary.** If there exist  $C > 0$  and  $n_0$  such that  $|b_n| \leq Ca_n$  for all  $n \geq n_0$ , and  $\sum a_n < \infty$ , then  $\sum b_n$  converges (absolutely).

**2.33 Example.** The series  $\sum_{n=1}^{\infty} 1/n$  (known as the harmonic series) was seen to diverge in Example 2.4. It follows that  $\sum_{n=1}^{\infty} n^{-s}$  diverges for  $s \leq 1$ .

**2.34 Example.** Consider the series  $\sum_{n=1}^{\infty} 1/n(n+1)$ ; since  $1/n(n+1) = 1/n - 1/(n+1)$ , we see that

$$s_n = \frac{1}{1} - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots + \frac{1}{n} - \frac{1}{n+1} = 1 - \frac{1}{n+1}$$

(we say that the series *telescopes*), so that the series converges to 1. It follows from the comparison test that the series  $\sum_{n=1}^{\infty} 1/n^2$  converges, since  $1/n^2 \leq 2/n(n+1)$ , and that its sum is less than 2. It then follows from the comparison test that  $\sum_{n=1}^{\infty} n^{-s}$  converges for all  $s \geq 2$ .

The next example gives one of the most useful series with which to compare other series.

**2.35 Example.** Consider the series  $\sum_{n=0}^{\infty} x^n$ , known as the *geometric series*. We have

$$s_n = \sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x}$$

when  $x \neq 1$ ; it follows that  $s_n \rightarrow 1/(1-x)$  when  $|x| < 1$ ; the series diverges for  $|x| \geq 1$  by Proposition 2.27. Thus  $\sum_{n=0}^{\infty} x^n = 1/(1-x)$  for  $|x| < 1$ , and the series diverges for all other  $x$ .

**2.36 Theorem.** Let  $(a_n)$  be a sequence, and let  $\lambda = \limsup |a_n|^{1/n}$ . If  $\lambda < 1$ , then  $\sum_{n=1}^{\infty} a_n$  converges absolutely; if  $\lambda > 1$ , then  $\sum_{n=1}^{\infty} a_n$  diverges.

**Proof.** If  $\lambda < 1$ , choose  $x$  with  $\lambda < x < 1$ . According to Proposition 2.15, there exists  $n_0$  such that  $|a_n|^{1/n} < x$  for all  $n \geq n_0$ . Then  $|a_n| < x^n$  for all  $n \geq n_0$ , and  $\sum x^n$  converges as we saw in the example above, so the series  $\sum a_n$  converges absolutely by the comparison test. Suppose now that  $\lambda > 1$ . Then, again by Proposition 2.15, there exist infinitely many  $n$  such that  $|a_n|^{1/n} > 1$ , and hence  $|a_n| > 1$ . Hence  $\sum a_n$  diverges, by Proposition 2.27. ■

Note that Theorem 2.36 asserts nothing if  $\limsup |a_n|^{1/n} = 1$ . In fact, this holds if  $a_n = 1/n$  (Proposition 2.6), and hence if  $a_n = 1/n^2$ . As we have just seen, the first of these gives a divergent series, the second a convergent series. The assertion of Theorem 2.36 is known as the *root test*; the following similar theorem is called the *ratio test*.

**2.37 Theorem.** Suppose that  $a_n > 0$  for all  $n$ . Let

$$\lambda = \liminf \frac{a_{n+1}}{a_n}, \quad \mu = \limsup \frac{a_{n+1}}{a_n}.$$

If  $\mu < 1$ , then  $\sum a_n$  converges, and if  $\lambda > 1$ , then  $\sum a_n$  diverges.

**Proof.** If  $\mu < 1$ , choose  $x$  with  $\mu < x < 1$ . There exists  $m$  such that  $|a_{n+1}/a_n - \mu| < x - \mu$  for all  $n \geq m$ , and hence  $a_{n+1}/a_n < x$  for all  $n \geq m$ . It follows that  $a_{n+1} < xa_n$  for all  $n \geq m$ , and inductively that  $a_{m+k} < x^k a_m$  for every  $k > 0$ . Thus  $a_n < Cx^n$  for all  $n \geq m$ , where  $C = x^{-m}a_m$ , and hence  $\sum a_n$  converges by the comparison test. If  $\lambda > 1$ , then there exists  $m$  such that  $a_{n+1}/a_n > 1$  for all  $n \geq m$ , i.e.,  $(a_n)_{n=m}^\infty$  is an increasing sequence, so  $\lim a_n = 0$  is impossible. (In fact,  $a_n \rightarrow +\infty$ , as a closer look will tell you.) ■

**2.38 Example.** The series  $\sum_{n=0}^\infty x^n/n!$  converges absolutely for every  $x \in \mathbf{R}$ ; for putting  $a_n = |x|^n/n!$ , we have  $a_{n+1}/a_n = |x|/(n+1) \rightarrow 0$  as  $n \rightarrow \infty$ . Thus the series converges absolutely for every  $x \in \mathbf{R}$  by the ratio test. Consider the series

$$\sum_{n=0}^\infty \frac{(nx)^n}{n!}.$$

Putting  $a_n = n^n|x|^n/n!$ , we have

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^{n+1}}{(n+1)!} \frac{n!}{n^n} |x| = \left(\frac{n+1}{n}\right)^n |x| \rightarrow e|x|,$$

as we saw in Example 2.12. Thus the given series converges absolutely for  $|x| < 1/e$  and diverges for  $|x| > 1/e$ .

The next example shows a new way of establishing divergence of a series.

**2.39 Example.** Here is another proof that the harmonic series  $\sum_{n=1}^\infty (1/n)$  diverges. We note that  $s_1 = 1$ ,  $s_2 - s_1 = \frac{1}{2}$ ,  $s_4 - s_2 = \frac{1}{3} + \frac{1}{4} > \frac{1}{2}$ ,  $s_8 - s_4 = \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{2}$ , and, in general,

$$s_{2^k} - s_{2^{k-1}} = \frac{1}{2^{k-1}+1} + \frac{1}{2^{k-1}+2} + \cdots + \frac{1}{2^k} > 2^{k-1} \frac{1}{2^k} = \frac{1}{2}$$

so

$$s_{2^n} = s_1 + \sum_{k=1}^n (s_{2^k} - s_{2^{k-1}}) > 1 + \frac{n}{2}.$$

Thus the series diverges to  $+\infty$ . This argument, based on the same idea as our earlier proof of this fact in Example 2.4, can be generalized to check many other series for convergence, as the next theorem shows.

**2.40 Theorem.** If  $(a_n)_{n=1}^{\infty}$  is a decreasing sequence, with  $a_n > 0$  for each  $n$ , then  $\sum a_n$  converges if and only if  $\sum 2^n a_{2^n}$  converges.

**Proof.** Since  $a_n > 0$ , the partial sums  $s_n$  of  $\sum_{n=1}^{\infty} a_n$  form an increasing sequence, as do the partial sums  $t_n$  of  $\sum_{n=0}^{\infty} 2^n a_{2^n}$ . Now

$$s_{2^k} - s_{2^{k-1}} = a_{2^{k-1}+1} + \cdots + a_{2^k}$$

gives

$$2^{k-1} a_{2^k} \leq s_{2^k} - s_{2^{k-1}} \leq 2^{k-1} a_{2^{k-1}},$$

since  $a_j \geq a_{j+1}$  for every  $j$ . Adding, we find

$$\frac{1}{2} \sum_{k=1}^n 2^k a_{2^k} \leq s_{2^n} - s_1 \leq \sum_{k=1}^n 2^{k-1} a_{2^{k-1}},$$

so  $\frac{1}{2} t_n \leq s_{2^n} \leq t_{n-1} + a_1$ . Thus  $(s_n)$  is bounded if and only if  $(t_n)$  is bounded. ■

**2.41 Example.** If  $s > 0$ , the sequence  $(n^{-s})$  is decreasing, so the last theorem (known, by the way, as *Cauchy's condensation test*) applies. Since  $2^n (2^n)^{-s} = 2^{n-n s} = (2^{1-s})^n$ , we see that  $\sum n^{-s}$  converges if and only if  $2^{1-s} < 1$ , i.e., if and only if  $s > 1$ . This extends our earlier findings for  $s \leq 1$  and  $s \geq 2$ .

The next example shows yet another idea in establishing the convergence of series.

**2.42 Example.** Consider the series  $\sum_{n=0}^{\infty} (-1)^n a_n$ , where  $a_n > a_{n+1}$  for all  $n$  and  $\lim a_n = 0$ . Then for each  $n$ ,  $s_{2n} - s_{2n-2} = -a_{2n-1} + a_{2n} < 0$ , so  $(s_{2n})$  is a decreasing sequence, while  $s_{2n+1} - s_{2n-1} = a_{2n} - a_{2n+1} > 0$ , so  $(s_{2n-1})$  is an increasing sequence. Since  $s_{2n} - s_{2n-1} = a_{2n} > 0$ , we have for all  $m < n$  (choosing  $k = \max\{m, n\}$ ) that

$$s_{2m-1} \leq s_{2k-1} < s_{2k} \leq s_{2n};$$

in particular, the monotone sequences  $(s_{2n})$  and  $(s_{2n-1})$  are bounded, hence convergent, say to  $s'$  and  $s''$ , respectively, and  $s' \geq s''$ . But  $s_{2n} - s_{2n-1} = a_{2n} \rightarrow 0$  as  $n \rightarrow \infty$ , so we conclude that  $s' = s''$ , and the series is convergent. Taking  $a_n = 1/(n+1)$ , we see that the series  $\sum_{n=0}^{\infty} (-1)^n/(n+1) = \sum_{n=1}^{\infty} (-1)^{n+1}/n$  is convergent; as we have seen, it is not absolutely convergent, so it is an example of a conditionally convergent series.

The next theorem can be regarded as a generalization of this example.

**2.43 Theorem.** Let  $(a_n)_{n=1}^{\infty}$  and  $(b_n)_{n=1}^{\infty}$  be two sequences in  $\mathbf{R}$ , let  $s_n = \sum_{k=1}^n b_k$  for each  $n \in \mathbf{N}$ , and suppose the following conditions hold:

- (a)  $\sum_{n=1}^{\infty} |a_n - a_{n+1}| < \infty$ ;  
 (b)  $(s_n)$  is a bounded sequence; and  
 (c)  $a_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Then  $\sum_{n=1}^{\infty} a_n b_n$  converges.

**Proof.** For any  $n > 1$  and any  $p$ , we have

$$\begin{aligned}
 \sum_{k=n}^{n+p} a_k b_k &= \sum_{k=n}^{n+p} a_k (s_k - s_{k-1}) \\
 &= \sum_{k=n}^{n+p} a_k s_k - \sum_{k=n}^{n+p} a_k s_{k-1} \\
 &= \sum_{k=n}^{n+p} a_k s_k - \sum_{k=n-1}^{n+p-1} a_{k+1} s_k \\
 &= \sum_{k=n}^{n+p-1} (a_k - a_{k+1}) s_k + a_{n+p} s_{n+p} - a_n s_{n-1}, \quad (2.7)
 \end{aligned}$$

a formula which is called *summation by parts*, in analogy with a similar formula for integrals. Now by hypothesis there exists  $M$  such that  $|s_n| < M$  for all  $n$ , and given  $\epsilon > 0$  there exists  $n_0$  such that  $|a_k| < \epsilon$  for all  $k \geq n_0$ , and such that  $\sum_{k=n_0}^{\infty} |a_k - a_{k+1}| < \epsilon$ . From formula (2.7) we have for any  $n > n_0$  and any  $p$ ,

$$\begin{aligned}
 \left| \sum_{k=n}^{n+p} a_k b_k \right| &\leq \sum_{k=n}^{n+p-1} |a_k - a_{k+1}| |s_k| + |a_{n+p}| |s_{n+p}| + |a_n| |s_{n-1}| \\
 &\leq M \sum_{k=n}^{\infty} |a_k - a_{k+1}| + 2M \sup_{j \geq n} |a_j| \\
 &\leq M\epsilon + 2M\epsilon = 3M\epsilon,
 \end{aligned}$$

which shows that the partial sums of the series  $\sum a_k b_k$  form a Cauchy sequence, and thus the series converges. ■

**2.44 Corollary.** If  $a_n > a_{n+1}$  for every  $n \in \mathbf{N}$  and  $\lim a_n = 0$ , and if  $(\sum_{k=1}^n b_k)$  is a bounded sequence, then  $\sum_{n=1}^{\infty} a_n b_n$  converges.

**Proof.** We have

$$\sum_{k=1}^n |a_k - a_{k+1}| = \sum_{k=1}^n (a_k - a_{k+1}) = a_1 - a_{n+1} < a_1$$

for all  $n$ , so all the hypotheses of Theorem 2.43 are satisfied. ■

**2.45 Corollary.** If  $(a_n)$  is a decreasing sequence, such that  $\lim a_n = 0$ , then  $\sum (-1)^n a_n$  converges.

**Proof.** If  $b_n = (-1)^n$ , the hypotheses of the last corollary are satisfied. ■

## 2.4 Rearrangements of Series

**2.46 Definition.** A sequence  $(b_n)_{n=1}^{\infty}$  is called a rearrangement of the sequence  $(a_n)_{n=1}^{\infty}$  if there exists a bijective mapping  $\phi : \mathbf{N} \rightarrow \mathbf{N}$  such that  $b_n = a_{\phi(n)}$  for every  $n$ .

**2.47 Theorem.** If  $a_n \geq 0$  for every  $n$ , then  $\sum a_n = \sum b_n$  for every rearrangement  $(b_n)$  of  $(a_n)$ .

**Proof.** It suffices to show that  $\sum b_n \leq \sum a_n$ , since  $(a_n)$  is a rearrangement of  $(b_n)$  whenever  $(b_n)$  is a rearrangement of  $(a_n)$ . Now if  $b_k = a_{\phi(k)}$  for each  $k$ , we have

$$\sum_{k=1}^n b_k = \sum_{k=1}^n a_{\phi(k)} \leq \sum_{n=1}^N a_n \leq \sum_{n=1}^{\infty} a_n,$$

where we took  $N = \max\{\phi(1), \dots, \phi(n)\}$ . Since this holds for every  $n$ , we conclude  $\sum b_k \leq \sum a_n$ . ■

**2.48 Corollary.** If the series  $\sum a_n$  is absolutely convergent, then so is  $\sum b_n$  for any rearrangement  $(b_n)$  of  $(a_n)$ , and  $\sum b_n = \sum a_n$ .

**Proof.** It is obvious that  $(|b_n|)$  is a rearrangement of  $(|a_n|)$  whenever  $(b_n)$  is a rearrangement of  $(a_n)$ , so  $\sum |b_n| = \sum |a_n| < +\infty$ . Similarly,  $\sum b_n^+ = \sum a_n^+$ , and  $\sum b_n^- = \sum a_n^-$ , so

$$\sum b_n = \sum (b_n^+ - b_n^-) = \sum b_n^+ - \sum b_n^- = \sum a_n^+ - \sum a_n^- = \sum a_n,$$

as we claimed. ■

A convergent series which remains convergent for any rearrangement is called *unconditionally convergent*. The last corollary asserts that an absolutely convergent series is unconditionally convergent. The converse is also true.

**2.49 Theorem.** Let  $(a_n)$  be a sequence in  $\mathbf{R}$  which is convergent but not absolutely convergent. For any extended real numbers  $\alpha$  and  $\beta$ , such that  $\alpha \leq \beta$ , there exists a rearrangement  $(b_n)$  of  $(a_n)$  such that, with  $s_n = \sum_{k=1}^n b_k$ ,

$$\liminf s_n = \alpha, \quad \limsup s_n = \beta.$$

**Proof.** Since  $\sum |a_n| = +\infty$ , while  $\sum a_n$  converges, we see that  $\sum a_n^+ = \sum a_n^- = +\infty$ . Let  $\{n : a_n > 0\} = \{n_1, n_2, \dots\}$ , and let  $\{n : a_n \leq 0\} = \{m_1, m_2, \dots\}$ , where  $n_1 < n_2 < \dots$  and  $m_1 < m_2 < \dots$ . Then

$$\sum_{k=1}^{\infty} a_{n_k} = \sum a_n^+ = +\infty, \quad (2.8)$$

$$\sum_{k=1}^{\infty} a_{m_k} = -\sum a_n^- = -\infty. \quad (2.9)$$

Assume first that  $-\infty < \alpha \leq \beta < +\infty$ . The idea of the construction is to list first just enough of the positive terms of the given sequence to get a sum larger than  $\beta$ , then just enough of the negative terms to get a sum less than  $\alpha$ , and continue in this manner. Let's do this in a more formal manner. We define inductively two sequences of positive integers,  $(k_j)_{j=1}^{\infty}$  and  $(l_j)_{j=1}^{\infty}$ , as follows. Define  $k_1$  to be the smallest positive integer  $k$  such that  $\sum_{j=1}^k a_{n_j} > \beta$ ; such  $k$  exist by virtue of (2.8), so  $k_1$  is well-defined. Let  $l_1$  be the smallest positive integer  $l$  such that

$$\sum_{j=1}^{k_1} a_{n_j} + \sum_{j=1}^l a_{m_j} < \alpha.$$

Such  $l$  exist by (2.9), so  $l_1$  is well-defined. Define  $\phi(j) = n_j$  for  $j = 1, \dots, k_1$ , and  $\phi(k_1 + j) = m_j$  for  $j = 1, \dots, l_1$ . Let  $N_1 = k_1 + l_1$ . Having defined  $k_1, \dots, k_j$  and  $l_1, \dots, l_j$ , and  $\phi(i)$  for  $i = 1, \dots, N_j$ , where  $N'_j = \sum_{i=1}^j k_i$ ,  $N''_j = \sum_{i=1}^j l_i$ , and  $N_j = N'_j + N''_j$ , we let  $k_{j+1}$  be the smallest positive integer  $k$  such that  $\sum_{i=1}^{N_j} a_{\phi(i)} + \sum_{i=N'_j+1}^{N'_j+k} a_{n_i} > \beta$ . Such  $k$  exist because of (2.8), so  $k_{j+1}$  is well-defined. We put  $\phi(N_j + i) = n_{N'_j+i}$  for  $j = 1, \dots, n_{k_{j+1}}$ . Similarly, we take  $l_{j+1}$  to be the smallest positive integer  $l$  such that  $\sum_{i=1}^{N_j+k_{j+1}} a_{\phi(i)} + \sum_{i=N''_j+1}^{N''_j+l} a_{m_i} < \alpha$ . In view of (2.9),  $l_{j+1}$  is well-defined, and we put  $\phi(N_j + k_{j+1} + i) = m_{N''_j+i}$  for  $j = 1, \dots, m_{l_{j+1}}$ . In this way, we define a bijective mapping  $\phi$  of  $\mathbf{N}$  on itself. Putting  $b_i = a_{\phi(i)}$  and  $s_n = \sum_{i=1}^n b_i$ , we have by the construction that  $s_{N_j} < \alpha$  and  $s_{N_j+k_{j+1}} > \beta$  for every  $j$ . Furthermore, we see that  $\alpha - |b_n| \leq s_n \leq \beta + |b_n|$  for all  $n$ . Since  $\lim b_n = 0$  (since the convergence of  $\sum a_n$  implies  $a_n \rightarrow 0$  as  $n \rightarrow \infty$ ), we conclude that  $s_n > \beta$  for infinitely many  $n$ , while for any  $\epsilon > 0$  we have  $s_n < \beta + \epsilon$  for all sufficiently large  $n$ , so  $\limsup s_n = \beta$ , and similarly we get  $\liminf s_n = \alpha$ .

If  $-\infty < \alpha \leq \beta = +\infty$ , we modify the above construction to get  $\sum_{i=1}^{N_j} b_i < \alpha$  and  $\sum_{i=1}^{N_j+k_j} b_i > j$  for each  $j$ , to get  $\limsup s_n = +\infty$ . A similar modification of the argument takes care of the other cases where  $\alpha$  or  $\beta$  equals  $\pm\infty$ . ■



## 2.5 Unordered Series

Let  $A$  be a set (perhaps uncountable), and let  $\mathcal{F}(A)$  denote the collection of all finite subsets of  $A$  (a notation to be used only in this section). If  $x : \alpha \mapsto x_\alpha$  is a function from  $A$  to  $\mathbf{R}$ , then  $\sum_{\alpha \in F} x_\alpha$  has an obvious well-defined meaning for each  $F \in \mathcal{F}(A)$ , in view of the commutative and associative laws for addition in  $\mathbf{R}$ .

**2.50 Definition.** Let  $x : \alpha \rightarrow x_\alpha$  be a real-valued function on  $A$ . We say that the unordered series  $\sum_{\alpha \in A} x_\alpha$  converges to the sum  $s$ , and write (with an abuse of language)  $\sum_{\alpha \in A} x_\alpha = s$ , if for every  $\epsilon > 0$  there exists  $F_0 \in \mathcal{F}(A)$  such that

$$\left| \sum_{\alpha \in F} x_\alpha - s \right| < \epsilon$$

for every  $F \in \mathcal{F}(A)$  with  $F \supset F_0$ . Similarly, we say that  $\sum_{\alpha \in A} x_\alpha$  diverges to  $+\infty$ , and write  $\sum_{\alpha \in A} x_\alpha = +\infty$ , if for every  $M \in \mathbf{R}$  there exists  $F_0 \in \mathcal{F}(A)$  such that  $\sum_{\alpha \in F} x_\alpha > M$  for every  $F \in \mathcal{F}$  with  $F \supset F_0$ .

We remark that in the case  $A = \mathbf{N}$ , the convergence of  $\sum_{n \in \mathbf{N}} x_n$  implies the convergence of  $\sum_{n=1}^{\infty} x_n$ ; the converse, as we will see shortly, is false.

The next two propositions are quite straightforward, and their proofs are omitted.

**2.51 Proposition.** Let  $x : A \rightarrow \mathbf{R}$ ,  $y : A \rightarrow \mathbf{R}$ , and  $c \in \mathbf{R}$ . If  $\sum_{\alpha \in A} x_\alpha$  and  $\sum_{\alpha \in A} y_\alpha$  converge, then so do  $\sum_{\alpha \in A} (x_\alpha + y_\alpha)$  and  $\sum_{\alpha \in A} (cx_\alpha)$ , and we have

$$\sum_{\alpha \in A} (x_\alpha + y_\alpha) = \sum_{\alpha \in A} x_\alpha + \sum_{\alpha \in A} y_\alpha, \quad \sum_{\alpha \in A} (cx_\alpha) = c \sum_{\alpha \in A} x_\alpha.$$

**2.52 Proposition.** Let  $x : A \rightarrow \mathbf{R}$ , with  $x_\alpha \geq 0$  for all  $\alpha \in A$ . Then  $\sum_{\alpha \in A} x_\alpha$  either converges or diverges to  $+\infty$ , and

$$\sum_{\alpha \in A} x_\alpha = \sup \left\{ \sum_{\alpha \in F} x_\alpha : F \in \mathcal{F}(A) \right\}.$$

**2.53 Theorem.** If  $x : A \rightarrow \mathbf{R}$ , then  $\sum_{\alpha \in A} x_\alpha$  converges if and only if  $\sum_{\alpha \in A} |x_\alpha|$  converges.

**Proof.** If  $\sum_{\alpha \in A} |x_\alpha|$  converges, then  $\sum_{\alpha \in A} x_\alpha^+$  converges by Proposition 2.52, since  $x^+ \leq |x|$  for all  $x \in \mathbf{R}$ ; it now follows from Proposition 2.51 that  $\sum_{\alpha \in A} x_\alpha$  converges, since  $x = 2x^+ - |x|$  for all  $x \in \mathbf{R}$ .

Now suppose  $\sum_{\alpha \in A} x_\alpha$  converges to  $s$ . Then there exists  $F_0 \in \mathcal{F}(A)$  such that  $|\sum_{\alpha \in F} x_\alpha - s| < 1$  for every  $F \in \mathcal{F}(A)$  with  $F \supset F_0$ , and

hence  $|\sum_{\alpha \in F} x_\alpha| < 1 + |s|$  for every  $F \in \mathcal{F}(A)$  with  $F \supset F_0$ . Now for any  $F \in \mathcal{F}(A)$ , we have

$$\sum_{\alpha \in F} x_\alpha = \sum_{\alpha \in F \cup F_0} x_\alpha - \sum_{\alpha \in F_0 \setminus F} x_\alpha \leq 1 + |s| + \sum_{\alpha \in F_0} x_\alpha^-,$$

since  $-x \leq x^-$  for all  $x \in \mathbf{R}$ . Let  $P = \{\alpha \in A : x_\alpha \geq 0\}$ . For any  $F \in \mathcal{F}(A)$  we have then

$$\sum_{\alpha \in F} x_\alpha^+ = \sum_{\alpha \in F \cap P} x_\alpha \leq 1 + |s| + \sum_{\alpha \in F_0} x_\alpha^-,$$

so  $\sum_{\alpha \in A} x_\alpha^+$  converges by Proposition 2.52, and hence  $\sum_{\alpha \in A} |x_\alpha|$  converges by Proposition 2.51, since  $|x| = 2x^+ - x$  for all  $x \in \mathbf{R}$ . ■

**2.54 Corollary.** *If  $\sum_{\alpha \in A} x_\alpha$  converges, then  $\{\alpha \in A : x_\alpha \neq 0\}$  is countable.*

**Proof.** Let  $S = \{\alpha \in A : x_\alpha \neq 0\}$ . If  $\sum_{\alpha \in A} x_\alpha$  converges, it follows that  $\sum_{\alpha \in A} |x_\alpha| = M < \infty$ , as we have just seen. Let  $S_n = \{\alpha \in A : |x_\alpha| > 1/n\}$ . Then  $S_n$  is finite, in fact  $\#S_n < nM$ . But  $S = \bigcup_{n=1}^{\infty} S_n$ , so  $S$  is countable. ■

**2.55 Theorem.** *Let  $(A_n)_{n=1}^{\infty}$  be a partition of  $A$ , i.e.,  $A = \bigcup_{n=1}^{\infty} A_n$  and  $A_j \cap A_k = \emptyset$  for every  $j \neq k$ . Let  $x : A \rightarrow \mathbf{R}$ . Then  $\sum_{\alpha \in A} x_\alpha$  converges if and only if*

$$\sum_{n=1}^{\infty} \sum_{\alpha \in A_n} |x_\alpha| < \infty, \quad (2.10)$$

and in this case,  $\sum_{\alpha \in A_n} x_\alpha$  converges for every  $n$  and

$$\sum_{\alpha \in A} x_\alpha = \sum_{n=1}^{\infty} \left( \sum_{\alpha \in A_n} x_\alpha \right). \quad (2.11)$$

**Proof.** Suppose (2.10) holds. If  $F \in \mathcal{F}(A)$ , then there exists  $N$  such that  $F \subset \bigcup_{k=1}^N A_k$ , so

$$\sum_{\alpha \in F} |x_\alpha| = \sum_{n=1}^N \sum_{\alpha \in F \cap A_n} |x_\alpha| \leq \sum_{n=1}^N \sum_{\alpha \in A_n} |x_\alpha| \leq \sum_{n=1}^{\infty} \sum_{\alpha \in A_n} |x_\alpha|,$$

so  $\sum_{\alpha \in A} |x_\alpha| \leq \sum_{n=1}^{\infty} \sum_{\alpha \in A_n} |x_\alpha| < \infty$ .

Now if  $\sum_{\alpha \in A} x_\alpha$  converges, so  $\sum_{\alpha \in A} |x_\alpha| < \infty$ , then  $\sum_{\alpha \in A_n} |x_\alpha| < \infty$  for every  $n$ . Given  $\epsilon > 0$ , there exists for each  $n$  a finite subset  $F_n$  of  $A_n$

such that  $\sum_{\alpha \in F_n} |x_\alpha| > \sum_{\alpha \in A_n} |x_\alpha| - \epsilon 2^{-n}$ . Fix  $N$ , and let  $F = \bigcup_{n=1}^N F_n$ . Then

$$\begin{aligned} \sum_{n=1}^N \sum_{\alpha \in A_n} |x_\alpha| &< \sum_{n=1}^N \left( \sum_{\alpha \in F_n} |x_\alpha| + \frac{\epsilon}{2^n} \right) \\ &= \sum_{\alpha \in F} |x_\alpha| + \sum_{n=1}^N \frac{\epsilon}{2^n} \leq \sum_{\alpha \in A} |x_\alpha| + \epsilon. \end{aligned}$$

Since this holds for every  $N$ ,

$$\sum_{n=1}^{\infty} \sum_{\alpha \in A_n} |x_\alpha| \leq \sum_{\alpha \in A} |x_\alpha| + \epsilon,$$

and since  $\epsilon > 0$  was arbitrary, we have equation (2.11), with  $x_\alpha$  replaced by  $|x_\alpha|$ , so (2.11) holds whenever  $x_\alpha \geq 0$  for all  $\alpha$ . In particular, it holds with  $x_\alpha$  replaced by  $x_\alpha^+$  or by  $x_\alpha^-$ . Using as usual  $x_\alpha = x_\alpha^+ - x_\alpha^-$  and Proposition 2.51, we get the validity of (2.11) in general. ■

**2.56 Corollary.** *If  $x : \mathbf{N} \rightarrow \mathbf{R}$ , then  $\sum_{n \in \mathbf{N}} x_n$  converges if and only if  $\sum_{n=1}^{\infty} |x_n|$  converges, and in this case,  $\sum_{n \in \mathbf{N}} x_n = \sum_{n=1}^{\infty} x_n$ .*

**2.57 Corollary.** *If  $x : A \rightarrow \mathbf{R}$ , and  $(A_n)_{n=1}^{\infty}$  and  $(B_n)_{n=1}^{\infty}$  are partitions of  $A$ , then  $\sum_{n=1}^{\infty} \sum_{\alpha \in A_n} |x_\alpha|$  converges if and only if  $\sum_{n=1}^{\infty} \sum_{\alpha \in B_n} |x_\alpha|$  converges, and in this case*

$$\sum_{n=1}^{\infty} \sum_{\alpha \in A_n} x_\alpha = \sum_{n=1}^{\infty} \sum_{\alpha \in B_n} x_\alpha = \sum_{\alpha \in A} x_\alpha.$$

A special case of this occurs quite frequently.

**2.58 Theorem.** *Let  $x_{nm} \in \mathbf{R}$  for all nonnegative integers  $n$  and  $m$ . Then*

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} x_{nm} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} x_{nm} = \sum_{k=0}^{\infty} \sum_{n+m=k} x_{nm} = \lim_{N \rightarrow \infty} \sum_{n=0}^N \sum_{m=0}^N x_{nm},$$

*provided any one of the four iterated series is convergent when  $x_{nm}$  is replaced by  $|x_{nm}|$ .*

One standard application of the last theorem is to the multiplication of infinite series.

**2.59 Definition.** *Let  $(a_n)_{n=0}^{\infty}$  and  $(b_n)_{n=0}^{\infty}$  be sequences in  $\mathbf{R}$ . The Cauchy product of the series  $\sum a_n$  and  $\sum b_n$  is defined to be the series  $\sum c_n$ , where  $c_n = \sum_{k=0}^n a_k b_{n-k}$  for  $n = 0, 1, \dots$ .*

**2.60 Theorem.** If the series  $\sum_{n=0}^{\infty} a_n$  and  $\sum_{n=0}^{\infty} b_n$  are absolutely convergent, then so is their Cauchy product  $\sum_{n=0}^{\infty} c_n$ , and

$$\sum_{n=0}^{\infty} c_n = \left( \sum_{n=0}^{\infty} a_n \right) \left( \sum_{n=0}^{\infty} b_n \right).$$

**Proof.** Let  $x_{nm} = a_n b_m$ . Then

$$\sum_{n=0}^{\infty} \left( \sum_{m=0}^{\infty} |x_{nm}| \right) = \sum_{n=0}^{\infty} \left( \sum_{m=0}^{\infty} |a_n| |b_m| \right) = \left( \sum_{n=0}^{\infty} |a_n| \right) \left( \sum_{n=0}^{\infty} |b_n| \right) < \infty,$$

so Theorem 2.58 tells us that

$$\left( \sum_{n=0}^{\infty} a_n \right) \left( \sum_{m=0}^{\infty} b_m \right) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} x_{nm} = \sum_{k=0}^{\infty} \sum_{n+m=k} a_n b_m = \sum_{k=0}^{\infty} c_k,$$

as was claimed. ■

In fact, the Cauchy product converges if only  $\sum a_n$  converges absolutely, and then it converges to the product of the two series. But the Cauchy product of two convergent series need not be convergent (see the exercises below.)

## 2.6 Exercises

1. Prove Lemma 2.5.
2. Show that the following sequences  $(x_n)$  converge, and find their limits:
  - (a)  $x_n = \frac{1}{n^2} + \frac{2}{n^2} + \cdots + \frac{n}{n^2}$ ;
  - (b)  $x_n = \sqrt{n}(\sqrt{n+1} - \sqrt{n})$ ;
  - (c)  $x_0 = 0$ ,  $x_1 = 1$ , and  $x_n = \frac{1}{2}(x_{n-1} + x_{n-2})$  for  $n \geq 2$ .
3. Let  $x_0 = 1$  and let  $x_{n+1} = 1 + 1/x_n$  for all  $n \geq 0$ . Show that  $(x_n)$  converges, and find its limit.
4. Suppose  $a > 0$ . Let  $x_1 = \sqrt{a}$ , and define  $x_{n+1} = \sqrt{a + x_n}$  for  $n \geq 1$ . Show that  $x_n < 1 + \sqrt{a}$  for all  $n$ , and that  $(x_n)$  is an increasing sequence. Then show that  $(x_n)$  converges, and find its limit.
5. Let  $J = \{x : 1 \leq x \leq \frac{3}{2}\}$ . Define  $f(x) = -\frac{1}{2}x^2 + x + 1$ , for  $x \in J$ .
  - (a) Show that  $f$  is strictly decreasing, i.e., that if  $x < y$ , then  $f(x) > f(y)$ .
  - (b) Show that if  $x \in J$ , then  $f(x) \in J$ .

- (c) Show that  $|f(x) - f(y)| \leq \frac{1}{2}|x - y|$  for all  $x, y \in J$ .
- (d) Let  $x_0 = 1$ , and  $x_{n+1} = f(x_n)$  for  $n \geq 0$ . Show that  $\lim x_n = \sqrt{2}$ .
- (e) Show that  $(x_{2n})$  increases, and  $(x_{2n-1})$  decreases, to  $\sqrt{2}$ .
6. Suppose  $0 < a < b$ . Define  $a_0 = a$ ,  $b_0 = b$ , and

$$a_{n+1} = \sqrt{a_n b_n}, \quad b_{n+1} = \frac{1}{2}(a_n + b_n)$$

for every  $n \in \mathbf{N}$ . Show that  $(a_n)$  is an increasing sequence, that  $(b_n)$  is a decreasing sequence, and that both converge to the same limit.

7. Prove Proposition 2.16. Give an example where we have strict inequality in (c) of this proposition.

8. Show that  $(n!)^{1/n} \rightarrow \infty$  as  $n \rightarrow \infty$ .

9. Use the Bolzano-Weierstrass theorem (Theorem 2.17) to give another proof of (the nontrivial half of) Theorem 2.19.

10. Let  $(a_n)_{n=1}^{\infty}$  be a sequence in  $\mathbf{R}$ .

- (a) Show that if  $\lim a_n = A$  exists, then  $\lim(1/n)(a_1 + a_2 + \cdots + a_n) = A$ , but that the converse is false.
- (b) Show that if  $\sum_{n=1}^{\infty} a_n$  converges, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n k a_k = 0.$$

11. Does the series  $\sum_{n=1}^{\infty} \frac{1}{n \sqrt[n]{n}}$  converge?

12. Show that

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{2n+1}{n(n+1)}$$

converges, and find its sum.

13. Let  $(a_n)$  be a sequence in  $\mathbf{R}$ , with  $a_n \rightarrow A \neq 0$  as  $n \rightarrow \infty$ . Suppose also that  $a_n \neq 0$  for every  $n$ . Show that the two series

$$\sum_{n=1}^{\infty} |a_{n+1} - a_n|, \quad \sum_{n=1}^{\infty} \left| \frac{1}{a_{n+1}} - \frac{1}{a_n} \right|$$

either both converge or both diverge.

14. Show that if  $a_n > 0$  for every  $n$ , then

$$\liminf \frac{a_{n+1}}{a_n} \leq \liminf a_n^{1/n} \leq \limsup a_n^{1/n} \leq \limsup \frac{a_{n+1}}{a_n}.$$

Hence the root test is always decisive when the ratio test is. Of course, there are many cases where the ratio test is easier to apply.

15. Show that if  $(a_n)$  is a sequence in  $\mathbf{R}$ , and  $a_n r^n \rightarrow 0$  for some  $r \neq 0$ , then  $\sum_{n=0}^{\infty} a_n x^n / n!$  converges absolutely for every  $x \in \mathbf{R}$ .

16. Show that the series

$$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^s}$$

diverges for  $s = 1$ , but converges for every  $s > 1$ . [The logarithm function has not yet been defined, but for the purpose of this problem take for granted the fact that there is such a function on  $(0, +\infty)$ , with the properties  $\log x > 0$  for all  $x > 1$ , and  $\log(xy) = \log x + \log y$  for all  $x, y \in (0, +\infty)$ .] HINT: Use Proposition 2.40.

17. Show that if  $\sum |a_n - a_{n+1}| < \infty$ , then  $(a_n)$  converges, but not conversely.

18. Let  $(a_n)_{n=1}^{\infty}$  and  $(b_n)_{n=1}^{\infty}$  be sequences in  $\mathbf{R}$ . Show that if  $\sum b_n$  converges and  $\sum |a_n - a_{n+1}| < \infty$ , then  $\sum a_n b_n$  converges.

19. Let  $(p_n)$  be an increasing sequence of positive real numbers, with  $p_n \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $(a_n)$  a sequence of real numbers.

(a) Show that if  $\sum a_n$  converges, then

$$\frac{1}{p_n} \sum_{k=1}^n p_k a_k \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (2.12)$$

(b) Conversely, if  $\sum a_n$  does not converge, show that there exists an increasing sequence of positive numbers  $(p_n)$  with  $\lim p_n = \infty$  such that equation (2.12) does not hold.

20. Show that if the sequence  $(a_n)$  converges, then so does any rearrangement of  $(a_n)$ , and to the same limit.

21. Let  $(a_n)$  be a sequence of positive real numbers. Show that if  $\sum a_n$  converges, then  $\liminf_{n \rightarrow \infty} n a_n = 0$ ; show that if also  $(a_n)$  is decreasing, then  $\lim_{n \rightarrow \infty} n a_n = 0$ .

22. Let  $a_n = \frac{(-1)^n}{\sqrt{n+1}}$  for  $n = 0, 1, \dots$ . Show that  $\sum_{n=0}^{\infty} a_n$  converges, but that the Cauchy product of  $\sum a_n$  with itself diverges.

## 2.7 Notes

- 2.1 The inequality  $(1+t)^n > 1+nt$  (for  $t > 0$ ,  $n > 1$ ) was first published by Jakob Bernoulli. The most common conception of a real number is probably that of an infinite decimal expansion, which Example 2.7 shows to be justified. However, taking this as the definition of real number, while possible, would present several awkward points. For instance, defining the sum and product of infinite decimals is not trivial. The earliest use of the method of calculating square roots described in Example 2.8 that we know of is found on clay tablets dating to the Old Babylonian period (1800-1600 B.C.). It was also known to Chinese and Arab mathematicians centuries before Newton, yet it is still sometimes referred to as "Newton's method." The symbol  $e$  was introduced by Euler (as was also the symbol  $\pi$ ). The Cauchy criterion is sometimes also attributed to Bolzano.
- 2.2 For more about continued fractions, see, for instance, the little book of Khinchin [7], or Hardy and Wright [4].
- 2.3 It is amusing that while we have established  $\sum_{n=1}^{\infty} 1/n = +\infty$ , the proof of Theorem 2.40 gives the estimate  $s_{2^n} < n+1$ , so that to get a partial sum that exceeds 100 we would need more than  $2^{100}$  terms of the series. Adding a million terms a second, this would take more than 10 quadrillion years.

Example 2.42 (the convergence of alternating series) was found by Leibniz. Its generalization, Theorem 2.43, or at least Corollary 2.44, is due to Dirichlet. The related Exercise 18 is due to Abel, for the case of a bounded monotone sequence  $(a_n)$ .

- 2.4 The striking Theorem 2.49 (or at least the special case  $\alpha = \beta$ ) is due to Riemann.
- 2.5 It was shown by Mertens that if  $\sum a_n$  is convergent,  $\sum b_n$  is absolutely convergent, and  $\sum c_n$  is the Cauchy product of  $\sum a_n$  and  $\sum b_n$ , then  $\sum c_n$  is convergent. It was shown by Abel that if all three series are convergent, then  $\sum c_n = (\sum a_n)(\sum b_n)$ . The sense of convergence described in this section is a special case of the following. A *directed set* is a partially ordered set with the property that given any  $x$ ,  $y$ , there exists  $z$  such that  $x \leq z$  and  $y \leq z$ . A totally ordered set is directed, as is the set  $\mathcal{F}$  of all finite subsets of a set  $A$ , with the partial ordering of set inclusion. Let  $I$  be a directed set, and let  $x: I \rightarrow \mathbf{R}$ . We say that  $(x_\alpha)_{\alpha \in I}$  is a *generalized sequence* in  $\mathbf{R}$ , or a *net* in  $\mathbf{R}$ . We say that  $x_\alpha \rightarrow L$  if for every  $\epsilon > 0$  there exists  $\alpha_0 \in I$  such that  $|x_\alpha - L| < \epsilon$  for every  $\alpha \in I$  with  $\alpha \geq \alpha_0$ . This notion will appear again in the definition of the Riemann integral later on.

- 2.6 Exercise 3, meant to be done by a direct approach, can also be treated by the methods of Section 2, and then gives an evaluation of the limit of  $(f_{n+1}/f_n)$ , where  $(f_n)$  is the sequence of Fibonacci numbers. Exercise 5 is an example of finding a fixed point of a mapping by the method of iteration. A generalization will be discussed later in the context of metric spaces. The sequences in Exercise 6 were introduced by Gauss, who called their common limit the *arithmetic-geometric mean* of the numbers  $a$  and  $b$ . Exercise 19 is a theorem of Kronecker.



# 3

## Continuous Functions on Intervals

In this chapter, we begin the study of continuous functions with the special case of real-valued functions defined on an interval in  $\mathbf{R}$ . The concepts we develop here will be reexamined in a more general setting in later chapters.

### 3.1 Limits and Continuity

**3.1 Definition.** Let  $A \subset \mathbf{R}$ , and let  $f : A \rightarrow \mathbf{R}$ . We say that  $f$  has the right-hand limit  $L$  ( $L \in \mathbf{R}$ ) at  $c$ , and write  $\lim_{x \rightarrow c+} f(x) = L$ , if there exists  $b > c$  such that  $(c, b) \subset A$ , and for any  $\epsilon > 0$  there exists  $\delta > 0$  such that  $|f(x) - L| < \epsilon$  for every  $x \in A$  satisfying  $c < x < c + \delta$ .

We say that  $f$  has the left-hand limit  $L$  at  $c$ , and write  $\lim_{x \rightarrow c-} f(x) = L$ , if there exists  $a < c$  such that  $(a, c) \subset A$ , and for any  $\epsilon > 0$  there exists  $\delta > 0$  such that  $|f(x) - L| < \epsilon$  for every  $x \in A$  satisfying  $c - \delta < x < c$ .

If  $\lim_{x \rightarrow c+} f(x)$  and  $\lim_{x \rightarrow c-} f(x)$  both exist and are equal, we call their common value the limit of  $f$  at  $c$ , and denote it by  $\lim_{x \rightarrow c} f(x)$ .

We also want to consider the extended real numbers  $\pm\infty$  in the roles of  $c$  or  $L$  above.

**3.2 Definition.** Let  $A \subset \mathbf{R}$ , and let  $f : A \rightarrow \mathbf{R}$ . We say  $\lim_{x \rightarrow c+} f(x) = +\infty$  if for every  $M \in \mathbf{R}$  there exists  $\delta > 0$  such that  $(c, c + \delta) \subset A$ , and  $f(x) > M$  for every  $x \in (c, c + \delta)$ .

We say that  $\lim_{x \rightarrow +\infty} f(x) = L$  if for every  $\epsilon > 0$  there exists  $x_0$  such that  $(x_0, +\infty) \subset A$  and  $|f(x) - L| < \epsilon$  for all  $x > x_0$ .

The definitions of  $\lim_{x \rightarrow c-} f(x) = +\infty$ ,  $\lim_{x \rightarrow -\infty} f(x) = L$ , etc., etc., follow the same pattern, and will be left to the reader.

We will often write  $f(a+)$  for  $\lim_{x \rightarrow a+} f(x)$ , and  $f(b-)$  for  $\lim_{x \rightarrow b-} f(x)$ .

There is a convenient terminology which saves us from writing too many  $\epsilon$ 's and  $\delta$ 's.

**3.3 Definition.** Let  $a \in \mathbf{R}$  and  $U \subset \mathbf{R}$ . We say that  $U$  is a neighborhood of  $a$  if there exists  $\epsilon > 0$  such that  $(a - \epsilon, a + \epsilon) \subset U$ . If  $a \in A \subset \mathbf{R}$ , we say that  $V$  is a neighborhood of  $a$  relative to  $A$  if there exists a neighborhood  $U$  of  $a$  such that  $V = A \cap U$ .

We make three extremely simple remarks: every open interval  $(a, b)$  with  $a < b$  is a neighborhood of  $c$  for every  $c \in (a, b)$ ; if  $U$  is a neighborhood of  $a$  and  $U \subset V$ , then  $V$  is a neighborhood of  $a$ ; if  $U$  and  $V$  are neighborhoods of  $a$ , then so is  $U \cap V$ .

We can rephrase the definition of limit as follows: let  $I$  be a neighborhood of  $c$ , and  $f : I \setminus \{c\} \rightarrow \mathbf{R}$ . We say  $\lim_{x \rightarrow c} f(x) = L$  if for every neighborhood  $V$  of  $L$  there exists a neighborhood  $U$  of  $c$  such that  $f(U \setminus \{c\}) \subset V$ .

**3.4 Example.** Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be defined by

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

Then for all  $t$  neither  $f(t+)$  nor  $f(t-)$  exists. For the intervals  $(t, t + \delta)$  and  $(t - \delta, t)$  contain both rational and irrational numbers, so with  $\epsilon = 1/2$  and any  $L$ , the inequality  $|f(s) - L| < \epsilon$  for all  $s \in (t, t + \delta)$  (or for all  $s \in (t - \delta, t)$ ) is impossible for any  $\delta > 0$ .

**3.5 Example.** Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be defined by  $f(t) = 1/n$  if  $t = m/n$ , where  $m$  and  $n$  are integers without common factor, ( $n > 0$ ), and  $f(t) = 0$  if  $t$  is irrational. Then  $\lim_{s \rightarrow t} f(s) = 0$  for every  $t \in \mathbf{R}$ . For given  $\epsilon > 0$ , there exist only finitely many numbers  $m/n$  with  $n \leq 1/\epsilon$  in the interval  $(t - 1, t + 1)$ , so there exists  $\delta > 0$  such that  $(t - \delta, t)$  and  $(t, t + \delta)$  contain no such numbers: thus  $f(s) < \epsilon$  for all  $s$  with  $0 < |t - s| < \delta$ .

**3.6 Proposition.** Let  $I$  be an interval with endpoints  $a$  and  $b$  ( $a < b$ ), and let  $f$  be an increasing function on  $I$ . Then for every  $t \in (a, b)$  the one-sided limits  $f(t+)$  and  $f(t-)$  exist, and  $f(t-) \leq f(t) \leq f(t+)$ . Furthermore,  $f(a+)$  and  $f(b-)$  exist (in the extended sense), with  $f(a) \leq f(a+)$  if  $a \in I$  and  $f(b-) \leq f(b)$  if  $b \in I$ .

**Proof.** If  $t \in (a, b)$ , let  $A = \sup\{f(s) : s \in I, s < t\}$  and let  $B = \inf\{f(s) : s \in I, s > t\}$ . Clearly  $A \leq f(t) \leq B$  since  $f$  is increasing. For any  $\epsilon > 0$ , there exists  $s_0 \in I$  with  $s_0 < t$  such that  $f(s_0) > A - \epsilon$ . Since  $f$  is increasing, it follows that  $A - \epsilon < f(s) \leq A$  for every  $s$  with  $s_0 < s < t$ , so  $A = f(t-)$ .

Similarly,  $B = f(t+)$ . If  $t = a$ , let  $B = \inf\{f(s) : s \in I, s > a\}$  (possibly  $B = -\infty$ ). The same argument as before shows that  $B = f(a+)$ . Similarly, if  $t = b$ , the same argument shows that with  $A = \sup\{f(s) : s \in I, s < b\}$  (possibly  $A = +\infty$ ), we have  $A = f(b-)$ . ■

**3.7 Definition.** Let  $I$  be an interval in  $\mathbf{R}$ ,  $c \in I$ , and  $f : I \rightarrow \mathbf{R}$ . We say that  $f$  is continuous at  $c$  if for every  $\epsilon > 0$  there exists  $\delta > 0$  such that  $|f(x) - f(c)| < \epsilon$  for every  $x \in I$  with  $|x - c| < \delta$ . We say that  $f$  is continuous on the interval  $I$  if it is continuous at every point of  $I$ .

Thus the function in Example 3.4 is not continuous at any point of  $\mathbf{R}$ , while the function in Example 3.5 is continuous at every irrational  $x$ , but discontinuous at every rational  $q$ .

There are several equivalent ways to formulate the notion of continuity.

**3.8 Proposition.** Let  $I$  be an interval in  $\mathbf{R}$ ,  $c \in I$ , and  $f : I \rightarrow \mathbf{R}$ . The following are equivalent:

- (a)  $f$  is continuous at  $c$ .
- (b)  $\lim_{x \rightarrow c} f(x) = f(c)$ . If  $c$  is the left endpoint of  $I$ , this is to be read as  $f(c+) = f(c)$ , and if  $c$  is the right endpoint of  $I$ , this is to be read as  $f(c-) = f(c)$ .
- (c) For every neighborhood  $V$  of  $f(c)$  there exists a neighborhood  $U$  of  $c$ , relative to  $I$ , such that  $f(U) \subset V$ .
- (d) For every neighborhood  $V$  of  $f(c)$ ,  $f^{-1}(V)$  is a neighborhood of  $c$ , relative to  $I$ .
- (e) For every sequence  $(x_n)$  in  $I$  with  $\lim x_n = c$ , the sequence  $(f(x_n))$  converges to  $f(c)$ .

**Proof.** The equivalence of (a), (b), (c), and (d) is a matter of looking at the definitions of limit and neighborhood. We show the equivalence of (a) and (e).

Suppose  $f$  is continuous at  $c$  and  $(x_n)$  is a sequence in  $I$  with  $\lim x_n = c$ . Let  $\epsilon > 0$ . There exists  $\delta > 0$  such that  $|f(x) - f(c)| < \epsilon$  for every  $x \in I$  such that  $|x - c| < \delta$ . There exists  $n_0$  such that  $|x_n - c| < \delta$  for every  $n \geq n_0$ . It follows that  $|f(x_n) - f(c)| < \epsilon$  for all  $n \geq n_0$ . Thus  $\lim f(x_n) = f(c)$  for every sequence  $(x_n)$  in  $I$  converging to  $c$ .

Now suppose that  $f$  is not continuous at  $c$ . Then there exists  $\epsilon > 0$  such that for every  $\delta > 0$  there exists  $x \in I$  with  $|x - c| < \delta$  but  $|f(x) - f(c)| \geq \epsilon$ . In particular, for every  $n$  there exists  $x_n \in I$  with  $|x_n - c| < 1/n$  and  $|f(x_n) - f(c)| \geq \epsilon$ . Then  $(x_n)$  converges to  $c$  but  $(f(x_n))$  does not converge to  $f(c)$ . ■

If  $f$  and  $g$  are real-valued functions defined on a set  $X$ , we can combine them in various ways to obtain other functions on  $X$ : we define their sum  $f+g$  by the rule  $(f+g)(x) = f(x)+g(x)$ , their product by the rule  $(fg)(x) = f(x)g(x)$ , and their maximum by  $\max\{f, g\}(x) = \max\{f(x), g(x)\}$ . The function  $1/f$  is defined by the rule  $(1/f)(x) = 1/f(x)$ ; its domain is  $\{x \in X : f(x) \neq 0\}$ . Similarly, we can define  $|f|$ ,  $f^+$ ,  $f^-$  by composing  $f$  with the appropriate map of  $\mathbf{R}$  to itself.

**3.9 Proposition.** *Let  $f$  and  $g$  be real-valued functions on an interval  $I$ , and  $c \in I$ . If  $f$  and  $g$  are continuous at  $c \in I$ , then  $f+g$  and  $fg$  are continuous at  $c$ . If also  $f(c) \neq 0$ , then  $1/f$  is defined in a relative neighborhood of  $c$ , and is continuous at  $c$ .*

**Proof.** This is an immediate consequence of Proposition 3.8(e) and the corresponding properties for sequences (Proposition 2.1). ■

**3.10 Example.** It is obvious that constant functions are continuous everywhere, and the identity function  $x \mapsto x$  is continuous everywhere. Using the last proposition, we conclude that every polynomial function is continuous everywhere, and that every rational function is continuous on its domain. It is also trivial that the maps  $x \mapsto x^+$ ,  $x \mapsto x^-$ , and  $x \mapsto |x|$  are all continuous. The greatest integer function  $x \mapsto [x]$  is continuous at each  $c \notin \mathbf{Z}$ , discontinuous at each  $c \in \mathbf{Z}$ .

**3.11 Proposition.** *Let  $I$  be an interval in  $\mathbf{R}$ ,  $f : I \rightarrow \mathbf{R}$ , and  $c \in I$ . Let  $J$  be an interval in  $\mathbf{R}$  with  $f(I) \subset J$ , and  $g : J \rightarrow \mathbf{R}$ . If  $f$  is continuous at  $c$  and  $g$  is continuous at  $f(c)$ , then the composition  $g \circ f$  is continuous at  $c$ .*

**Proof.** Let  $V$  be a neighborhood of  $g(f(c))$ . Then, since  $g$  is continuous at  $f(c)$ ,  $g^{-1}(V)$  is a neighborhood of  $f(c)$ , relative to  $J$ , i.e.,  $g^{-1}(V) = U \cap J$ , where  $U$  is a neighborhood of  $f(c)$  in  $\mathbf{R}$ . Since  $f$  is continuous at  $c$ ,  $f^{-1}(U)$  is a neighborhood of  $c$  relative to  $I$ . Since  $f(I) \subset J$ ,  $f^{-1}(g^{-1}(V)) = f^{-1}(U) \cap f^{-1}(J) = f^{-1}(U)$ . Hence  $(g \circ f)^{-1}(V) = f^{-1}(g^{-1}(V)) = f^{-1}(U)$  is a neighborhood of  $c$  relative to  $I$ . Thus  $g \circ f$  is continuous at  $c$ . ■

**3.12 Corollary.** *If  $f$  is continuous at  $c$ , then so are  $f^+$ ,  $f^-$ ,  $|f|$ , and  $p(f)$ , whenever  $p$  is a polynomial.*

**3.13 Proposition.** *If  $f$  is an increasing function on an interval  $I$ , then  $D = \{d \in I : f \text{ is not continuous at } d\}$  is countable.*

**Proof.** For any  $d \in I$ ,  $d$  not an endpoint of  $I$ , we know (Proposition 3.6) that  $f(d-)$  and  $f(d+)$  exist, with  $f(d-) \leq f(d) \leq f(d+)$ , so  $d \in D$  if and only if  $f(d-) < f(d+)$ . Similarly, if  $a$  is the left endpoint of  $I$  and  $a \in I$ , then  $a \in D$  if and only if  $f(a) < f(a+)$ , and if  $I$  contains its right endpoint  $b$ , then  $b \in D$  if and only if  $f(b-) < f(b)$ . For each

$d \in D$ , the interval  $(f(d-), f(d+))$  contains a rational number  $q_d$  (make the obvious modification if  $d$  is an endpoint of  $I$ ). If  $d, d' \in D$  with  $d < d'$ , then  $f(d+) \leq f(d'-)$ , so  $q_d < q_{d'}$ . Thus the map  $d \mapsto q_d$  of  $D$  into  $\mathbf{Q}$  is injective; since  $\mathbf{Q}$  is countable, it follows that  $D$  is countable. ■

It is left as one of the exercises at the end of this chapter to show that for any countable subset  $S$  of  $\mathbf{R}$  there exists an increasing function on  $\mathbf{R}$  whose discontinuity set is precisely  $S$ .

## 3.2 Two Fundamental Theorems

The next two theorems are of use in many situations. They will be generalized in a later chapter.

**3.14 Theorem.** *A continuous real-valued function on a closed bounded interval attains maximum and minimum values.*

**Proof.** Suppose  $f : J \rightarrow \mathbf{R}$  is continuous, where  $J = [a, b]$  ( $a < b$ ). Let  $M = \sup\{f(x) : x \in J\}$ . If  $M = +\infty$ , there exists for each  $n \in \mathbf{N}$  some  $x_n \in J$  such that  $f(x_n) > n$ . According to the Bolzano–Weierstrass theorem (Theorem 2.17) there exists a subsequence  $(x_{n_k})$  of  $(x_n)$  which converges to some  $c$ ; since  $a \leq x_n \leq b$  for every  $n$ , we have  $a \leq c \leq b$ . According to Proposition 3.8(e), we have  $f(c) = \lim f(x_{n_k})$ , but this is impossible since  $f(x_{n_k}) > n_k \rightarrow +\infty$ . Thus  $M < +\infty$ . Now choose, for each  $n \in \mathbf{N}$ ,  $x_n \in J$  such that  $f(x_n) > M - 1/n$ . Since also  $f(x_n) \leq M$ , we have  $f(x_n) \rightarrow M$  as  $n \rightarrow \infty$ . The sequence  $(x_n)$  has a convergent subsequence  $(y_n)$ . Then  $(f(y_n))$  is a subsequence of  $(f(x_n))$ , so  $f(y_n) \rightarrow M$ ; but if  $y_n \rightarrow c$ , Proposition 3.8 assures us that  $f(y_n) \rightarrow f(c)$ . Thus  $f(c) = M$ . The proof that  $f$  attains a minimum value is similar, or can be deduced from what we have proved by considering the function  $-f$ . ■

**3.15 Theorem.** *If  $f$  is continuous on the interval  $[a, b]$ , and  $f(a) < y < f(b)$ , or  $f(a) > y > f(b)$ , there exists  $x$ , with  $a < x < b$ , such that  $f(x) = y$ .*

**Proof.** We may assume that  $f(a) < y < f(b)$ . Let  $E = \{t \in [a, b] : f(t) < y\}$ , so  $E$  is a nonempty ( $a \in E$ ) subset of  $[a, b]$ . Let  $x = \sup E$ , so  $x \in [a, b]$ . For each  $n$  there exists  $x_n \in E$  such that  $x - 1/n < x_n \leq x$ . Thus  $f(x_n) < y$  for every  $n$ . Since  $x_n \rightarrow x$ , we have (by Proposition 3.8(e))  $\lim f(x_n) = f(x)$ , so  $f(x) \leq y$ . But  $f(b) > y$  implies (since  $f$  is continuous at  $b$ ) that there exists  $\delta > 0$  such that  $f(t) > y$  for all  $t$  with  $b - \delta < t \leq b$ . Thus  $x < b$ . Hence there exist  $t_n \in J$  with  $x < t_n$  and  $\lim t_n = x$ . Since  $t_n > x$ , we have  $t_n \notin E$ , i.e.,  $f(t_n) \geq y$ , so  $f(x) = \lim f(t_n) \geq y$ . Thus  $f(x) = y$ . ■

This result is known as the *intermediate value theorem*. Another way to express this fact: if  $J$  is an interval, and  $f : J \rightarrow \mathbf{R}$  is continuous, then  $f(J)$  is an interval. The last two theorems together yield: if  $J$  is a closed bounded interval, then  $f(J)$  is a closed bounded interval.

**3.16 Theorem.** *Let  $I$  be a nonempty interval in  $\mathbf{R}$ , and let  $f : I \rightarrow \mathbf{R}$  be continuous. If  $f$  is injective, then  $J = f(I)$  is an interval, and  $g = f^{-1}$  is a continuous function on  $J$ .*

**Proof.** We have already observed that  $J = f(I)$  is an interval, just because  $I$  is an interval and  $f$  is continuous. The condition that  $f$  is injective is equivalent to the condition that  $f$  is either strictly increasing or strictly decreasing on  $I$ . Indeed, if  $f$  is neither strictly increasing nor strictly decreasing, there would exist  $a, b, c \in I$  with  $a < b < c$  such that either  $f(a) \leq f(b) \geq f(c)$  or  $f(a) \geq f(b) \leq f(c)$ . If equality holds in any of these inequalities,  $f$  is not injective. Suppose  $f(a) < f(c)$ . If  $f(b) < f(a)$ , Theorem 3.15 tells us there exists  $t \in (b, c)$  such that  $f(t) = f(a)$ , so  $f$  is not injective. If  $f(b) > f(c)$ , Theorem 3.15 tells us there exists  $t \in (a, b)$  such that  $f(t) = f(c)$ , so again  $f$  is not injective. The case where  $f(a) > f(c)$  is dealt with in an entirely analogous manner. Thus  $f$  injective implies  $f$  is either strictly increasing or strictly decreasing. The converse is obvious.

The interesting part is the continuity of the inverse function  $g$ , whose existence comes from the definition of injective. Let  $y \in J$ , so  $y = f(x)$  for some  $x \in I$ . Suppose  $f$  is strictly increasing (the case of  $f$  strictly decreasing is treated in a similar fashion, or can be deduced from the strictly increasing case by considering the function  $-f$ ). Let  $V$  be a neighborhood of  $x$  relative to  $I$ ; we must show that  $g^{-1}(V) = f(V)$  is a neighborhood of  $y$  relative to  $J$ . If  $x$  is not an endpoint of  $I$ , there exist  $a, b \in V$  with  $a < x < b$ . Then  $f(a) < y < f(b)$  since  $f$  is strictly increasing, and for any  $z$  with  $f(a) < z < f(b)$  there exists (by Theorem 3.15) some  $t$ ,  $a < t < b$ , with  $f(t) = z$ . Thus  $g^{-1}(V) \supset (f(a), f(b))$ , so  $g^{-1}(V)$  is a neighborhood of  $y$  for any neighborhood  $V$  of  $g(y)$ . If  $x$  is an endpoint of  $I$ , say the left endpoint, then  $y$  is necessarily the left endpoint of  $J$ , since  $f$  is strictly increasing, and there exists  $b > x$  such that  $[x, b) \subset V$ . Then  $[y, f(b)) \subset J$  is a neighborhood of  $y$  relative to  $J$ , and as before  $g^{-1}(V) \supset [y, f(b))$  is a neighborhood of  $y$  relative to  $J$ . Thus  $g$  is continuous at  $y$  for any  $y \in J$ . ■

As an application of this result, consider the function  $f : x \mapsto x^n$ , which is strictly increasing on  $\mathbf{R}$  if  $n$  is odd, and strictly increasing on  $\mathbf{R}^+ = [0, \infty)$  when  $n$  is even. The theorem not only tells us that  $f$  maps onto  $\mathbf{R}$  when  $n$  is odd,  $\mathbf{R}^+$  when  $n$  is even, i.e., gives a new proof of Theorem 1.13, but also gives the bonus that the inverse function  $x \mapsto x^{1/n}$  is continuous.

### 3.3 Uniform Continuity

**3.17 Definition.** Let  $I$  be an interval in  $\mathbf{R}$ , and let  $f : I \rightarrow \mathbf{R}$ . We say that  $f$  is uniformly continuous on  $I$  if for every  $\epsilon > 0$  there exists  $\delta > 0$  such that  $|f(s) - f(t)| < \epsilon$  for all  $s, t \in I$  such that  $|s - t| < \delta$ .

Thus, if  $f$  is uniformly continuous on  $I$ , then  $f$  is continuous on  $I$ , but the converse need not hold. Consider the function  $f$  on  $(0, 1)$  defined by  $f(x) = 1/x$ . As we saw above,  $f$  is continuous at each point of  $(0, 1)$ . But for any  $\delta > 0$  we can find  $s$  and  $t$  in  $(0, 1)$  such that  $|s - t| < \delta$  and  $|f(s) - f(t)| \geq 1$ . For instance, take any  $s < \delta$  and  $t = s/(1 + s)$ . Clearly,  $0 < t < s$ , so  $|s - t| = s - t < s < \delta$ , and  $1/t - 1/s = 1$ . Thus there exists no  $\delta$  fulfilling the job description when  $\epsilon = 1$ , so  $f$  is not uniformly continuous. Similarly, it is easy to see that the function  $t \mapsto t^2$  is not uniformly continuous on  $[0, \infty)$ . However, no such example exists when  $I$  is a closed bounded interval.

**3.18 Theorem.** If  $I$  is a closed bounded interval, and  $f : I \rightarrow \mathbf{R}$  is continuous on  $I$ , then  $f$  is uniformly continuous on  $I$ .

**Proof.** If  $f$  is not uniformly continuous on  $I$ , there exists  $\epsilon > 0$  such that for every  $\delta > 0$  there exist  $s$  and  $t$  in  $I$  with  $|s - t| < \delta$  but  $|f(s) - f(t)| \geq \epsilon$ . In particular, for each  $n \in \mathbf{N}$  there exist  $s_n$  and  $t_n$  in  $I$  such that  $|s_n - t_n| < 1/n$  and  $|f(s_n) - f(t_n)| \geq \epsilon$ . According to the Bolzano–Weierstrass theorem (Theorem 2.17) there exists a convergent subsequence  $(s_{n_k})$  of  $(s_n)$ . Then  $s = \lim s_{n_k} \in I$ , since  $I$  is a closed interval. Since  $|t_{n_k} - s_{n_k}| < 1/n_k < 1/k$ , we see that  $\lim t_{n_k} = s$  also. According to Proposition 3.8,  $f(s) = \lim f(s_{n_k}) = \lim f(t_{n_k})$ , but this contradicts  $|f(s_{n_k}) - f(t_{n_k})| \geq \epsilon$ . Thus the hypothesis that  $f$  is not uniformly continuous is untenable. ■

This theorem will prove itself invaluable in the future. At this point, we content ourselves with applying it to show that every continuous function on a bounded closed interval can be approximated in a natural sense by a function whose graph consists of a finite number of line segments.

**3.19 Definition.** Let  $f$  be a real-valued function on an interval  $I$  with endpoints  $a$  and  $b$ . We say that  $f$  is linear if there exist  $c, d \in \mathbf{R}$  such that  $f(t) = c + td$  for all  $t \in I$ . We say that  $f$  is piecewise linear if there exists a finite sequence  $(x_k)_{k=0}^n$  with  $a = x_0 < x_1 < \cdots < x_n = b$  such that the restriction of  $f$  to  $(x_{k-1}, x_k)$  is linear, for  $k = 1, 2, \dots, n$ .

We observe that if  $f$  is continuous and piecewise linear, then the restriction of  $f$  to each  $[x_{k-1}, x_k]$  is linear.

**3.20 Theorem.** If  $f$  is a continuous real-valued function on the closed bounded interval  $I$ , then for any  $\epsilon > 0$  there exists a piecewise linear continuous function  $g$  such that  $|f(t) - g(t)| < \epsilon$  for all  $t \in I$ .

**Proof.** According to Theorem 3.18, there exists  $\delta > 0$  such that  $|f(s) - f(t)| < \epsilon$  for all  $s, t \in I$  with  $|s - t| < \delta$ . Suppose  $I = [a, b]$ . Choose  $n > (b - a)/\delta$ , and set  $x_k = a + (k/n)(b - a)$ , for  $k = 0, 1, \dots, n$ , so  $a = x_0 < x_1 < \dots < x_n = b$ , with  $x_k - x_{k-1} = (b - a)/n < \delta$  for each  $k$ ,  $1 \leq k \leq n$ . Define  $g$  to be the piecewise linear function such that  $g(x_k) = f(x_k)$  for each  $k$ ,  $0 \leq k \leq n$ . In other words, put

$$\begin{aligned} g(t) &= f(x_{k-1}) + \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}(t - x_{k-1}) \\ &= f(x_{k-1}) \frac{x_k - t}{x_k - x_{k-1}} + f(x_k) \frac{t - x_{k-1}}{x_k - x_{k-1}} \end{aligned}$$

for  $x_{k-1} \leq t \leq x_k$ ,  $1 \leq k \leq n$ . Now for any  $t \in I$ , there is some  $k$  such that  $t \in [x_{k-1}, x_k]$ . We observe that the value of  $g(t)$  lies between the values at  $x_{k-1}$  and  $x_k$  of the function  $f$ , and therefore, by Theorem 3.15,  $g(t) = f(s)$  for some  $s \in [x_{k-1}, x_k]$ . Since  $|s - t| < \delta$ , we obtain  $|f(t) - g(t)| = |f(t) - f(s)| < \epsilon$ . ■

## 3.4 Sequences of Functions

If  $(f_n)$  is a sequence of functions, there are many ways that we might understand the statement that  $(f_n)$  converges to a function  $f$ . We briefly consider two of these ways in this section.

**3.21 Definition.** Let  $(f_n)$  be a sequence of real-valued functions on a set  $X$ . We say that  $f_n \rightarrow f$  pointwise, or that  $f$  is the pointwise limit of the sequence  $(f_n)$ , if  $\lim f_n(x) = f(x)$  for every  $x \in X$ .

In other words,  $f_n \rightarrow f$  pointwise if and only if for every  $x \in X$  and every  $\epsilon > 0$  there exists  $n_0$  such that  $|f_n(x) - f(x)| < \epsilon$  for every  $n \geq n_0$ .

**3.22 Definition.** We say that  $f_n \rightarrow f$  uniformly on  $X$ , or that  $f$  is the uniform limit on  $X$  of the sequence  $(f_n)$ , if for every  $\epsilon > 0$  there exists an integer  $n_0$  such that  $|f(x) - f_n(x)| < \epsilon$  for every  $x \in X$  whenever  $n \geq n_0$ .

Comparing this to the definition of pointwise convergence, we see that the difference is that  $n_0$  can be chosen to work simultaneously for all  $x$ , rather than choosing  $n_0$  for each  $x$  separately.

It is clear that if  $f_n \rightarrow f$  uniformly, then  $f_n \rightarrow f$  pointwise, but the converse need not hold. For instance, let  $f_n(x) = x^n$  for  $x \in [0, 1]$ . Then  $f_n(x) \rightarrow 0$  if  $0 \leq x < 1$ , but  $f_n(1) = 1$  for all  $n$ . Thus  $f_n \rightarrow f$  pointwise, where  $f$  is defined by  $f(x) = 0$  for  $0 \leq x < 1$ ,  $f(1) = 1$ . We note that  $f$  is not continuous at 1. The convergence is not uniform, since if we take any  $0 < \epsilon < 1$ , we can find for any  $n_0$  some  $x \in [0, 1)$  such that  $x^{n_0} > \epsilon$ , i.e.,  $f_{n_0}(x) - f(x) > \epsilon$ .



Theorem 3.20 is equivalent to the statement that for any continuous real-valued function on a closed bounded interval  $I$ , there exists a sequence  $(g_n)$  of continuous piecewise linear functions which converges uniformly on  $I$  to  $f$ .

**3.23 Proposition.** *Let  $(f_n)$  be a sequence of real-valued functions on a set  $X$ . Then  $(f_n)$  is uniformly convergent to some  $f$  if and only if it is uniformly Cauchy, i.e., if and only if for every  $\epsilon > 0$  there exists  $n_0$  such that  $|f_n(x) - f_m(x)| < \epsilon$  for every  $n, m \geq n_0$  and every  $x \in X$ .*

**Proof.** If  $(f_n)$  converges uniformly to  $f$ , then for any  $\epsilon > 0$  there exists  $n_0$  such that  $|f(x) - f_n(x)| < \epsilon/2$  for every  $n \geq n_0$ , every  $x \in X$ , and it follows that

$$\begin{aligned} |f_m(x) - f_n(x)| &= |f_m(x) - f(x) + f(x) - f_n(x)| \\ &\leq |f_m(x) - f(x)| + |f(x) - f_n(x)| < \epsilon \end{aligned}$$

for all  $x \in X$  whenever  $n, m \geq n_0$ , so  $(f_n)$  is uniformly Cauchy.

Suppose now that  $(f_n)$  is uniformly Cauchy. Then  $(f_n(x))$  is a Cauchy sequence in  $\mathbf{R}$  for each  $x \in X$ , hence convergent for each  $x \in X$ . Let  $f(x) = \lim f_n(x)$  for each  $x \in X$ . Given  $\epsilon > 0$ , choose  $n_0$  so that  $|f_n(x) - f_m(x)| < \epsilon/2$  for every  $x \in X$ , whenever  $n, m \geq n_0$ . Then, for every  $n \geq n_0$  and every  $x \in X$ , we have

$$|f_n(x) - f(x)| \leq |f_n(x) - f_m(x)| + |f_m(x) - f(x)| < \epsilon/2 + |f_m(x) - f(x)|$$

for every  $m \geq n_0$ . Since  $\lim f_m(x) = f(x)$ , we can choose  $m$  so that  $|f_m(x) - f(x)| < \epsilon/2$ , and it follows that  $|f_n(x) - f(x)| < \epsilon$  for every  $n \geq n_0$  and every  $x \in X$ , i.e., that  $(f_n)$  converges uniformly on  $X$  to  $f$ . ■

**3.24 Theorem.** *Let  $(f_n)$  be a sequence of continuous real-valued functions on the interval  $I \subset \mathbf{R}$ . If  $f_n \rightarrow f$  uniformly on  $I$ , then  $f$  is continuous on  $I$ .*

**Proof.** Let  $x \in I$  and  $\epsilon > 0$ . Choose  $n$  so that  $|f_n(t) - f(t)| < \epsilon/3$  for every  $t \in I$ , and then choose  $\delta > 0$  so that  $|f_n(y) - f_n(x)| < \epsilon/3$  for all  $y \in I$  with  $|y - x| < \delta$ . Then for all  $y \in I$  with  $|x - y| < \delta$  we have

$$\begin{aligned} |f(x) - f(y)| &= |f(x) - f_n(x) + f_n(x) - f_n(y) + f_n(y) - f(y)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon, \end{aligned}$$

so  $f$  is continuous at  $x$ . ■

We have seen an example where the lack of uniform convergence was revealed by the discontinuity of the limit function. However, it is possible for a sequence  $(f_n)$  of continuous functions to converge pointwise to a continuous function  $f$ , and yet the convergence not be uniform. For instance,

take  $f_n(x) = 1/(nx)$  for  $x \in (0, 1)$ . For an example where the functions are defined on a closed bounded interval, define

$$f_n(x) = \begin{cases} nx & \text{if } 0 \leq x \leq 1/n, \\ 2 - nx & \text{if } 1/n \leq x \leq 2/n, \\ 0 & \text{if } 2/n \leq x \leq 1. \end{cases}$$

Then each  $f_n$  is continuous on  $[0, 1]$ , the sequence  $(f_n)$  converges to 0 pointwise on  $[0, 1]$ , but  $\max\{f_n(x) : 0 \leq x \leq 1\} = 1$  for every  $n$ , so the convergence is not uniform.

There is one special case where pointwise convergence to a continuous function does imply uniform convergence.

**3.25 Theorem.** *Let  $(f_n)$  be a sequence of continuous real-valued functions on a closed bounded interval  $I$ , and suppose  $(f_n)$  is a monotone sequence, i.e., that either  $f_n(x) \leq f_{n+1}(x)$  for every  $n$  and every  $x \in I$ , or that  $f_n(x) \geq f_{n+1}(x)$  for every  $n$  and every  $x \in I$ . If  $(f_n)$  converges pointwise to a continuous function  $f$ , then  $(f_n)$  converges to  $f$  uniformly on  $I$ .*

**Proof.** If  $(f_n)$  is increasing, let  $g_n = f - f_n$ , or if  $(f_n)$  is decreasing, let  $g_n = f_n - f$ . Then  $(g_n)$  is a decreasing sequence of continuous functions, with  $\lim g_n(x) = 0$  for every  $x \in I$ . We show that  $(g_n)$  converges to 0 uniformly on  $I$ , and this is equivalent to the uniform convergence on  $I$  of  $(f_n)$  to  $f$ . Let  $M_n = \sup\{g_n(x) : x \in I\}$ . Evidently  $M_{n+1} \leq M_n$  for all  $n$ . The uniform convergence of  $(g_n)$  to 0 is equivalent to  $\lim M_n = 0$ . By Theorem 3.14 there exists  $x_n \in I$  such that  $g_n(x_n) = M_n$ . According to the Bolzano Weierstrass theorem (Theorem 2.17) there exists a subsequence  $(x_{n_k})$  of  $(x_n)$  which converges to some  $x^* \in I$ . Since  $\lim g_n(x^*) = 0$ , there exists  $m$  such that  $g_n(x^*) < \epsilon$  for all  $n \geq m$ . Since  $g_m$  is continuous, there is a neighborhood  $U$  of  $x^*$  such that  $g_m(x) < \epsilon$  for all  $x \in U$ . Choose  $k$  such that  $x_{n_k} \in U$ , and  $n_k > m$ . Then  $g_{n_k}(x_{n_k}) \leq g_m(x_{n_k}) < \epsilon$ , i.e.,  $M_{n_k} < \epsilon$ , and it follows that  $M_n < \epsilon$  for every  $n \geq n_k$ . Thus,  $(g_n)$  converges to 0 uniformly on  $I$ . ■

Infinite series of functions are a special case of infinite sequences of functions. We say that the infinite series of functions converges uniformly on the set  $X$  if the associated sequence of partial sums converges uniformly on  $X$ . The following is probably the most often used criterion for uniform convergence; it is known as the Weierstrass  $M$ -test.

**3.26 Theorem.** *Let  $(f_n)$  be a sequence of real-valued functions on a set  $X$ . If there exists a sequence of constants  $(M_n)$  such that  $|f_n(x)| \leq M_n$  for every  $x \in X$ , and  $\sum M_n < +\infty$ , then  $\sum f_n$  converges uniformly on  $X$ .*

**Proof.** The series  $\sum f_n(x)$  converges (absolutely) for every  $x \in X$  by the comparison test (Corollary 2.32). Let  $s_n(x) = \sum_{k=1}^n f_k(x)$ , and let  $s(x) =$

$\sum_{k=1}^{\infty} f_k(x)$ . For any  $\epsilon > 0$ , there exists  $m$  such that  $\sum_{k=m+1}^{\infty} M_k < \epsilon$ ; then we have, for all  $n \geq m$  and every  $x \in X$ ,

$$|s(x) - s_n(x)| = \left| \sum_{k=n+1}^{\infty} f_k(x) \right| \leq \sum_{k=n+1}^{\infty} |f_k(x)| \leq \sum_{k=m+1}^{\infty} M_k < \epsilon,$$

so that  $\sum f_n$  converges to  $s$  uniformly on  $X$ . ■

## 3.5 The Exponential Function

The series  $\sum_{n=0}^{\infty} x^n/n!$  converges absolutely for every  $x \in \mathbf{R}$ , as we see by applying the ratio test (Theorem 2.37). Let us denote its sum by  $E(x)$ . We note that the convergence is uniform on any bounded interval; in fact, for all  $x \in [-L, L]$ , we have

$$\frac{|x^n|}{n!} \leq \frac{L^n}{n!} = M_n,$$

and the series  $\sum M_n$  converges, so the series for  $E(x)$  converges uniformly by the Weierstrass  $M$ -test (Theorem 3.26). It follows (Theorem 3.24) that  $E$  is continuous on any bounded interval in  $\mathbf{R}$ , and thus continuous on  $\mathbf{R}$ . We also note that  $E$  is strictly increasing on  $[0, +\infty)$ . The essential property of the function  $E$  is expressed in the following proposition.

**3.27 Proposition.** *For every  $x, y \in \mathbf{R}$ ,  $E(x+y) = E(x)E(y)$ .*

**Proof.** We calculate

$$\begin{aligned} E(x+y) &= \sum_{n=0}^{\infty} \frac{(x+y)^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \\ &= \sum_{n=0}^{\infty} \sum_{k+j=n} \frac{x^k}{k!} \frac{y^j}{j!} = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{x^k}{k!} \frac{y^j}{j!} \\ &= E(x)E(y), \end{aligned}$$

where we used the binomial formula (Exercise 3 in Chapter 1), and Theorem 2.58. ■

**3.28 Corollary.** *For all  $x \in \mathbf{R}$ ,  $E(x)E(-x) = E(0) = 1$ ; also,  $E$  is a strictly increasing positive function on  $\mathbf{R}$ , with*

$$\lim_{x \rightarrow +\infty} E(x) = +\infty, \quad \lim_{x \rightarrow -\infty} E(x) = 0.$$

**Proof.** The equation  $E(0) = 1$  is immediate from the definition of  $E$ ; then  $E(0) = E(x)E(-x)$  follows from the proposition, and obviously implies  $E(x) > 0$ , since we already observed that  $E(x) \geq 1$  for  $x \geq 0$ . If  $x < y < 0$ , then  $E(-x) > E(-y)$  since  $-x > -y$  and  $E$  is increasing on  $[0, \infty)$ , so  $E(x) = 1/E(-x) < 1/E(-y) = E(y)$ . The other cases of  $x < y$  are immediate, so  $E$  is increasing on  $\mathbf{R}$ . The definition of  $E$  gives the inequality  $E(x) > 1 + x$  for all  $x > 0$ , which shows that  $E(x) \rightarrow +\infty$  as  $x \rightarrow +\infty$ , and the relation  $E(-x) = 1/E(x)$  then shows that  $E(x) \rightarrow 0$  as  $x \rightarrow -\infty$ . ■

**3.29 Proposition.**  $E(1) = e$ .

**Proof.** Recall that  $e$  was defined as  $\lim_{n \rightarrow \infty} (1 + 1/n)^n$  (Example 2.12). Using the binomial formula, we have

$$\left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \left(\frac{1}{n}\right)^k = \sum_{k=0}^n \frac{1}{k!} \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k},$$

from which we see at once that  $(1 + 1/n)^n < \sum_{k=0}^n (1/k!) < E(1)$ , which implies that  $e \leq E(1)$ . Fix the positive integer  $m$ . Then for any  $n > m$ , we have

$$\begin{aligned} \left(1 + \frac{1}{n}\right)^n &= \sum_{k=0}^n \frac{1}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \\ &\geq \sum_{k=0}^m \frac{1}{k!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right); \end{aligned}$$

letting  $n \rightarrow \infty$  we get  $e \geq \sum_{k=0}^m (1/k!)$ , for every  $m$ , and then letting  $m \rightarrow \infty$  we get  $e \geq E(1)$ . ■

In view of the last two propositions, we denote  $E(x)$  by  $e^x$ , or sometimes by  $\exp x$ . Since  $E$  is a strictly increasing map of  $\mathbf{R}$  onto  $(0, +\infty)$ , there exists an inverse function defined on  $(0, +\infty)$  which we denote by  $\log$ ;  $\log$  is continuous by Theorem 3.16, it is strictly increasing, with  $\log 1 = 0$ ,  $\lim_{x \rightarrow 0+} \log x = -\infty$ , and  $\lim_{x \rightarrow +\infty} \log x = +\infty$ , and satisfies  $\log(xy) = \log x + \log y$  for all  $x, y \in (0, +\infty)$ .

We also define  $b^x$  for any  $b > 0$ ,  $x \in \mathbf{R}$ , by  $b^x = e^{x \log b}$ . It is easy to see that this agrees with our previous understanding of  $b^x$  when  $x$  is an integer, or rational number. It is clear that  $x \mapsto b^x$  is strictly increasing when  $b > 1$ , strictly decreasing when  $b < 1$ . We further define (for  $b > 0$ ,  $b \neq 1$ ),  $\log_b x = \log x / \log b$ , so that  $b^y = x$  if and only if  $y = \log_b x$ .

## 3.6 Trigonometric Functions

For any  $R > 0$ , if  $a_n = R^{2n}/(2n)!$  we have  $a_{n+1}/a_n = R^2/(2n+2)(2n+1) \rightarrow 0$  as  $n \rightarrow \infty$ , so the series  $\sum_{n=0}^{\infty} a_n$  converges by the ratio test

(Theorem 2.37).

**3.30 Definition.** *The cosine and sine functions are defined by*

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

In view of the remark just made, these series converge absolutely for every  $x \in \mathbf{R}$ , so the functions are well-defined. The Weierstrass  $M$ -test shows that both series converge uniformly on every interval  $[-R, R]$ , so the cosine and sine functions are continuous on  $\mathbf{R}$ . We calculate, using Theorem 2.60,

$$\begin{aligned} \cos^2 x &= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{(-1)^k x^{2k}}{(2k)!} \frac{(-1)^{n-k} x^{2n-2k}}{(2n-2k)!} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} \sum_{k=0}^n \binom{2n}{2k} \end{aligned}$$

and, similarly,

$$\begin{aligned} \sin^2 x &= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{(-1)^k x^{2k+1}}{(2k+1)!} \frac{(-1)^{n-k} x^{2n-2k+1}}{(2n-2k+1)!} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+2}}{(2n+2)!} \sum_{k=0}^n \binom{2n+2}{2k+1} \\ &= - \sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} \sum_{k=0}^{n-1} \binom{2n}{2k+1} \end{aligned}$$

Now

$$2^{2n} = (1+1)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} \quad (n \geq 0)$$

and

$$0 = (1-1)^{2n} = \sum_{k=0}^{2n} (-1)^k \binom{2n}{k} \quad (n \geq 1),$$

so adding and subtracting these equations we see that (for  $n \geq 1$ )

$$2^{2n} = 2 \sum_{k=0}^n \binom{2n}{2k} = 2 \sum_{k=0}^{n-1} \binom{2n}{2k+1}.$$

Thus we have

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

so adding and subtracting we get the equations

$$\cos^2 x + \sin^2 x = 1, \quad (3.1)$$

$$\cos^2 x - \sin^2 x = \cos 2x. \quad (3.2)$$

The first of these formulas shows that  $|\cos x| \leq 1$  and  $|\sin x| \leq 1$  for all  $x \in \mathbf{R}$ .

Now for  $x > 0$ ,  $x^{n+2}/(n+2)! < x^n/n!$  if and only if  $x^2 < (n+1)(n+2)$ ; hence  $x^{n+2}/(n+2)! < x^n/n!$  for all  $n \geq 1$  if  $0 < x < \sqrt{6}$ . Looking back to Example 2.42 (the discussion of alternating series), we see that we have the inequalities

$$\frac{x^2}{2!} - \frac{x^4}{4!} < 1 - \cos x < \frac{x^2}{2!}, \quad x - \frac{x^3}{3!} < \sin x < x \quad (3.3)$$

for all  $x$  with  $0 < x < \sqrt{6}$ . In particular, we have  $\cos 0 = 1$  and  $\cos 2 < 1 - 2^2/2 + 2^4/4! = 1 - 2 + 2/3 < 0$ . By the intermediate value theorem, there exists  $t \in (0, 2)$  such that  $\cos t = 0$ .

**3.31 Definition.** We define  $\pi = 2 \inf\{t > 0 : \cos t = 0\}$ .

Since the cosine is continuous and  $\cos 0 = 1$ , we see that  $\pi > 0$ , and the estimates above give  $\pi < 4$ . We will find better estimates later. The inequalities above give  $\sin x > 0$  for  $0 < x < \sqrt{6}$ , so  $\sin(\pi/2) > 0$ . Since  $\cos(\pi/2) = 0$  and  $\cos^2 x + \sin^2 x = 1$  for all  $x$ , we conclude  $\sin(\pi/2) = +1$ . From the equation  $\cos^2 x - \sin^2 x = \cos 2x$  proved above, we get  $\cos \pi = \cos^2(\pi/2) - \sin^2(\pi/2) = -1$ , which in turn shows  $\sin \pi = 0$ . Using the formula again, we find  $\cos 2\pi = 1$ ,  $\sin 2\pi = 0$ .

From the defining equation for the sine function, we have

$$\begin{aligned} \sin(x+y) &= \sum_{n=0}^{\infty} (-1)^n \frac{(x+y)^{2n+1}}{(2n+1)!} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \sum_{k=0}^{2n+1} \binom{2n+1}{k} x^k y^{2n+1-k} \\ &= \sum_{n=0}^{\infty} (-1)^n \sum_{k=0}^{2n+1} \frac{x^k}{k!} \frac{y^{2n+1-k}}{(2n+1-k)!} \\ &= \sum_{n=0}^{\infty} \left( \sum_{k=0}^n \frac{(-1)^k x^{2k}}{(2k)!} \frac{(-1)^{n-k} y^{2n-2k+1}}{(2n-2k+1)!} \right. \\ &\quad \left. + \sum_{k=0}^n \frac{(-1)^k x^{2k+1}}{(2k+1)!} \frac{(-1)^{n-k} y^{2n-2k}}{(2n-2k)!} \right) \\ &= \cos x \sin y + \sin x \cos y. \end{aligned}$$

Similarly, we find

$$\begin{aligned}
 \cos(x+y) &= \sum_{n=0}^{\infty} (-1)^n \frac{(x+y)^{2n}}{(2n)!} \\
 &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \sum_{k=0}^{2n} \binom{2n}{k} x^k y^{2n-k} \\
 &= \sum_{n=0}^{\infty} (-1)^n \sum_{k=0}^{2n} \frac{x^k}{k!} \frac{y^{2n-k}}{(2n-k)!} \\
 &= \sum_{n=0}^{\infty} \left( \sum_{k=0}^n \frac{(-1)^k x^{2k}}{(2k)!} \frac{(-1)^{n-k} y^{2n-2k}}{(2n-2k)!} \right. \\
 &\quad \left. + \sum_{k=0}^{n-1} \frac{(-1)^k x^{2k+1}}{(2k+1)!} \frac{(-1)^{n-k} y^{2n-2k-1}}{(2n-2k-1)!} \right) \\
 &= \cos x \cos y - \sin x \sin y.
 \end{aligned}$$

From these formulas, we find that

$$\begin{aligned}
 \sin(x+\pi) &= \sin x \cos \pi + \cos x \sin \pi = -\sin x, \\
 \cos(x+\pi) &= \cos x \cos \pi - \sin x \sin \pi = -\cos x,
 \end{aligned}$$

and therefore

$$\sin(x+2\pi) = \sin x, \quad \cos(x+2\pi) = \cos x$$

for all  $x \in \mathbf{R}$ .

### 3.7 Exercises

1. Let  $f: \mathbf{R} \rightarrow \mathbf{R}$ . If  $\lim_{h \rightarrow 0} [f(x+h) - f(x-h)] = 0$  for every  $x \in \mathbf{R}$ , does it follow that  $f$  is continuous? If  $\lim_{h \rightarrow 0} [f(x+h) + f(x-h) - 2f(x)] = 0$  for every  $x \in \mathbf{R}$ , does it follow that  $f$  is continuous? Give proofs or counterexamples.
2. Show that if  $f: \mathbf{R} \rightarrow \mathbf{R}$  satisfies the equation  $f(x+y) = f(x) + f(y)$  for every  $x, y \in \mathbf{R}$ , and if  $f$  is continuous at (at least) one point, then there exists  $c$  such that  $f(x) = cx$  for all  $x \in \mathbf{R}$ . HINT: Show that  $f$  is continuous everywhere; let  $c = f(1)$  and show that  $f(q) = cq$  for every rational  $q$ ; deduce that  $f(x) = cx$  for every real  $x$ .
3. Prove that a polynomial (with real coefficients) of odd degree has at least one real root.

4. Let  $I$  be a bounded closed interval, and let  $f : I \rightarrow \mathbf{R}$  be continuous. Show that there exists  $x \in I$  such that  $f(x) = x$ . Are the hypotheses about  $I$  both necessary?

5. Let  $I$  be an interval in  $\mathbf{R}$ . For each  $t \in I$ , let  $\mathcal{N}_t$  denote the set of all neighborhoods of  $t$ , relative to  $I$ . If  $f : I \rightarrow \mathbf{R}$ , define, for each  $t \in I$ ,

$$\limsup f(t) = \inf_{U \in \mathcal{N}_t} \sup\{f(s) : s \in U\},$$

$$\liminf f(t) = \sup_{U \in \mathcal{N}_t} \inf\{f(s) : s \in U\}.$$

Thus  $-\infty \leq \liminf f(t) \leq \limsup f(t) \leq +\infty$ . Prove the following:

- (a) if  $c > 0$ , then  $\limsup(cf)(t) = c \limsup f(t)$  and  $\limsup(-f)(t) = -\liminf f(t)$ .
- (b)  $\limsup(f+g)(t) \leq \limsup f(t) + \limsup g(t)$ , and equality need not hold.
- (c)  $\lim_{x \rightarrow t} f(x) = L$  if and only if

$$\limsup f(t) = \liminf f(t) = L.$$

6. The function  $f : I \rightarrow \mathbf{R}$  is called *upper semicontinuous* at  $t \in I$  if  $f(t) \geq \limsup f(t)$ , and upper semicontinuous in  $I$  if it is upper semicontinuous at each point of  $I$ . We abbreviate upper semicontinuous by u.s.c.

- (a) Show that  $f$  is u.s.c. at  $t \in I$  if and only if for every  $\epsilon > 0$  there exists a neighborhood  $U$  of  $t$ , relative to  $I$ , such that  $f(s) < f(t) + \epsilon$  for every  $s \in U$ .
- (b) Show that if  $J$  is a closed interval, then the function  $f$  defined by  $f(t) = 1$  if  $t \in J$ ,  $f(t) = 0$  if  $t \notin J$ , is u.s.c. on  $\mathbf{R}$ .
- (c) Show that if  $I$  is a closed and bounded interval, and  $f$  is u.s.c. on  $I$ , then there exists  $c \in I$  such that  $f(t) \leq f(c)$  for every  $t \in I$ .
- (d) Define the notion of lower semicontinuous function. Deduce the analogues for lower semicontinuous functions of (a), (b), and (c).

7. A real-valued function  $f$  on an interval  $I$  is called *convex* if it satisfies the following condition: for every  $a, b \in I$  and every  $t$ ,  $0 < t < 1$ ,

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b).$$

Let  $f$  be convex on  $I$ .

- (a) Show that if  $I$  is open, then  $f$  is continuous.
- (b) Show that if  $g$  is convex and increasing on an interval containing  $f(I)$ , then  $g \circ f$  is convex on  $I$ .



8. Let  $S$  be a countable infinite subset of  $\mathbf{R}$ , let  $n \mapsto q_n$  be a bijective map of  $\mathbf{N}$  onto  $S$ , and define

$$f(x) = \sum_{\{n: q_n \leq x\}} 2^{-n}$$

for each  $x \in \mathbf{R}$ . Show that  $f$  is an increasing function, that  $f$  is continuous at each  $x \notin S$ , and that  $f(q_n+) - f(q_n-) = 2^{-n}$ . Taking  $S$  to be the set  $\mathbf{Q}$  of all rationals, we get an example of an increasing function which is discontinuous at each rational, continuous at each irrational.

9. Show that if  $I$  is a bounded interval, and  $f: I \rightarrow \mathbf{R}$  is uniformly continuous, then  $f$  is bounded.

10. Show that a bounded continuous function on a bounded interval need not be uniformly continuous.

11. Let  $I$  be an interval in  $\mathbf{R}$ , and suppose that  $f: I \rightarrow \mathbf{R}$  is uniformly continuous. Show that  $(f(x_n))$  is a Cauchy sequence whenever  $(x_n)$  is a Cauchy sequence in  $I$ .

12. Let  $f$  be uniformly continuous on the open interval  $(a, b)$ . Show that  $\lim_{t \rightarrow a+} f(t)$  and  $\lim_{t \rightarrow b-} f(t)$  exist, and hence that  $f$  can be extended to a continuous function on  $[a, b]$ .

13. Show that any piecewise linear continuous function  $f$  can be expressed in the form

$$f(x) = a + bx + \sum_{k=0}^n c_k |x - x_k|$$

for some constants  $a, b, x_0, x_1, \dots, x_n, c_0, c_1, \dots, c_n$ .

14. Show that if  $(f_n)$  and  $(g_n)$  are uniformly convergent sequences of real-valued functions on  $X$ , then  $(f_n + g_n)$  is also uniformly convergent on  $X$ , but that  $(f_n g_n)$  need not be uniformly convergent.

15. For which real  $x$  is the series

$$\sum_{n=1}^{\infty} \frac{x^n}{1 + x^{2n}}$$

convergent? For which intervals  $I$  is the series uniformly convergent on  $I$ ?

16. Show that the series

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

is uniformly convergent on every bounded interval in  $\mathbf{R}$ , but is not absolutely convergent for any  $x$ .

17. Let  $I$  be an interval, and let  $f_n : I \rightarrow \mathbf{R}$  for each  $n \in \mathbf{N}$ . Suppose that  $x \in I$  and every  $f_n$  is continuous at  $x$ . Suppose that  $(f_n)$  converges uniformly on  $I$  to  $f$ . Show that if  $x_n \in I$  and  $x_n \rightarrow x$  as  $n \rightarrow \infty$ , then  $f_n(x_n) \rightarrow f(x)$  as  $n \rightarrow \infty$ .

18. Show that  $\lim_{x \rightarrow +\infty} x^n e^{-x} = 0$  for any positive integer  $n$ , and show that  $\lim_{x \rightarrow 0+} x^p \log x = 0$  for any  $p > 0$ .

19. Suppose that  $F : \mathbf{R} \rightarrow \mathbf{R}$  is continuous, and has the property that  $F(x+y) = F(x)F(y)$  for all  $x$  and  $y$  in  $\mathbf{R}$ . Show that either  $F(x) = 0$  for all  $x \in \mathbf{R}$  or there exists  $b > 0$  such that  $F(x) = b^x$  for all  $x \in \mathbf{R}$ .

20. Use the results of the last section to show that  $\sin(\pi/2 - x) = \cos x$  and  $\cos(\pi/2 - x) = \sin x$  for all real  $x$ ; show that  $\sin(\pi/4) = \cos(\pi/4) = \sqrt{2}/2$ , and  $\sin(\pi/6) = \cos(\pi/3) = 1/2$ .

## 3.8 Notes

3.1 Cauchy is often credited with the introduction of rigorous ideas into analysis, but he frequently referred to infinitely small quantities, which is not considered good form in standard analysis today. Here is a quote from one of Cauchy's books (1821):

The function  $f(x)$  will remain continuous with respect to  $x$  between the given limits, if between these limits an infinitely small increase of the variable always produces an infinitely small increase of the function itself.

This is a far cry from the definition we use, which was first published by Heine in 1872, and probably derives from the lectures of Weierstrass. The state of rigor in the 1820s, and the growing desire for it, is illustrated by Abel's writing in 1826, "There is in mathematics hardly a single infinite series of which the sum is determined in a rigorous way." Something like the modern idea of function was first given by Dirichlet in 1837. It was Dirichlet who first considered the function in Example 3.4. Many of the basic ideas in analysis that later became common currency were first developed by Bolzano before 1820, but his work did not become widely known until long after his death.

3.2 The intermediate value theorem was published by Cauchy; perhaps in previous generations it was too obvious a geometric fact to require proof. Bolzano had published a pamphlet devoted to this theorem, but it did not receive wide attention at the time. Darboux observed that a function need not be continuous to enjoy the intermediate value property: an example would be the function defined by  $f(0) = 0$  and  $f(x) = \sin(1/x)$  for  $x \neq 0$ .

- 3.3 The notion of uniform continuity possibly originates with Weierstrass. Theorem 3.18 is due to Heine.
- 3.4 The notion of uniform convergence, and the fundamental Theorem 3.24, were found independently in the 1840s by Stokes, Seidel, and Weierstrass. It was Weierstrass who introduced the modern terminology *uniform* (“gleichmässig”) in his lectures of that period. Cauchy had erred in concluding too much from pointwise convergence, but also came to the concept of uniform convergence in a paper of 1853. Theorem 3.25 is due to Dini.
- 3.5 The symbol  $e$  was introduced by Euler, who also found the series expansion for the exponential function. Historically, the logarithm was developed before the exponential.
- 3.6 Euler also found the beautiful relation between the exponential and trigonometric functions that becomes available when one expands one's horizons from the real to the complex numbers, expressed by the formula  $e^{iz} = \cos z + i \sin z$ , valid for any complex number  $z$ . Taking  $z = x$  real, this equation can serve as a definition of the trigonometric functions  $\cos x$  and  $\sin x$ , as, respectively, the real and imaginary parts of  $e^{iz}$ . The addition formulas for the trigonometric functions then follow at once from the addition formula for the exponential:  $e^{z+w} = e^z e^w$  for any  $z, w \in \mathbb{C}$ . This is proved word-for-word as is Proposition 3.27, and the result is a much easier proof of the results of Section 3.6.

# 4

## Differentiation

In this chapter we develop the basic theory of derivatives of real-valued functions of one variable. The geometric motivation of the derivative of  $f$  at  $c$  is that it is the limit of the slope of lines joining  $(c, f(c))$  to nearby points on the graph of  $f$ , as these points approach  $c$ , and therefore represents the slope of a “line passing through two consecutive points” of the graph of  $f$ , i.e., the tangent line. We can enjoy such language without succumbing to it.

### 4.1 Derivatives

**4.1 Definition.** Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval in  $\mathbf{R}$ , and let  $x \in I$ . We say that  $f$  is differentiable at  $x$  if

$$\lim_{y \rightarrow x} \frac{f(y) - f(x)}{y - x} = f'(x)$$

exists. (If  $x$  is an endpoint of  $I$ , the appropriate one-sided limit is intended.) The limit  $f'(x)$  is called the derivative of  $f$  at  $x$ .

It may happen that a one-sided limit

$$\lim_{y \rightarrow x+} \frac{f(y) - f(x)}{y - x}, \quad \text{or} \quad \lim_{y \rightarrow x-} \frac{f(y) - f(x)}{y - x}$$

exists; these will be called the right-hand derivative (or left-hand derivative) of  $f$  at  $x$ . Evidently, the derivative of  $f$  at an interior point  $x$  of  $I$  exists if and only if both one-sided derivatives exist and are equal.

It is clear that differentiability is a local property, in the following sense: if  $f$  and  $g$  are defined in neighborhoods  $U$  and  $V$  of the point  $x$ , and if for some neighborhood  $W$  of  $x$ , with  $W \subset U \cap V$  we have  $f(y) = g(y)$  for every  $y \in W$ , then  $f$  is differentiable at  $x$  if and only if  $g$  is, and  $f'(x) = g'(x)$ .

The function  $\phi(y) = (f(y) - f(x))/(y - x)$  is defined in  $I \setminus \{x\}$ , and the definition says precisely that  $f$  is differentiable at  $x$  if and only if  $\phi$  can be defined at  $x$ , i.e., extended to all of  $I$ , in such a way that the extended function is continuous at  $x$ . We make this comment formally as a proposition.

**4.2 Proposition.** *Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval in  $\mathbf{R}$ , and let  $x \in I$ . Then  $f$  is differentiable at  $x$  if and only if there exists  $\phi : I \rightarrow \mathbf{R}$ , with  $\phi$  continuous at  $x$ , such that  $f(y) = f(x) + (y - x)\phi(y)$  for all  $y \in I$ . In this case,  $f'(x) = \phi(x)$ .*

**4.3 Corollary.** *If  $f$  is differentiable at  $x$ , then  $f$  is continuous at  $x$ .*

Of course, a continuous function need not be differentiable. The simplest example is perhaps the function  $|x|$ , which is not differentiable at 0. In fact, a continuous function need not have a derivative at any point.

**4.4 Example.** Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be defined by  $f(t) = |t|$  for  $-\frac{1}{2} \leq t \leq \frac{1}{2}$ , and let  $f(t + n) = f(t)$  for every  $n \in \mathbf{Z}$ . Thus the graph of  $f$  consists of a sequence of line segments. For each nonnegative integer  $n$ , define  $f_n$  by  $f_n(t) = 4^{-n}f(4^n t)$ . Then the graph of  $f_n$  consists of a sequence of line segments, having slope  $\pm 1$ . We note that  $f_n(t + 4^{-n}) = f_n(t)$  for all  $t \in \mathbf{R}$  and every  $n \geq 0$ . Evidently,  $f_n$  is continuous on  $\mathbf{R}$  for every  $n$ . Since  $|f_n(t)| \leq \frac{1}{2}4^{-n}$ , we see that the series  $\sum_{n=0}^{\infty} f_n(t)$  converges uniformly on  $\mathbf{R}$  (the Weierstrass  $M$ -test, Theorem 3.26) and hence  $g = \sum_{n=0}^{\infty} f_n$  is continuous on  $\mathbf{R}$ , by Theorem 3.24. We shall show that  $g$  is nowhere differentiable. Fix  $t \in \mathbf{R}$ . For each  $n$ , choose  $h_n = \pm 4^{-n-1}$ , with the sign chosen so that  $4^n t$  and  $4^n(t + h_n)$  lie in the same interval  $[\frac{k}{2}, \frac{k+1}{2}]$ . Then we have  $f_n(t + h_n) - f_n(t) = \pm h_n$ , and in fact  $f_m(t + h_n) - f_m(t) = \pm h_n$  for  $m \leq n$ ; but  $f_m(t + h_n) = f_m(t)$  for every  $m > n$ . Hence we have

$$\frac{g(t + h_n) - g(t)}{h_n} = \sum_{m=0}^n \frac{f_m(t + h_n) - f_m(t)}{h_n} = \sum_{m=0}^n \epsilon_m,$$

where  $\epsilon_m = \pm 1$  for  $m = 0, \dots, n$ . Thus the difference quotient  $(g(t + h_n) - g(t))/h_n$  is an odd integer when  $n$  is even, and an even integer when  $n$  is odd; since  $h_n \rightarrow 0$  as  $n \rightarrow \infty$ , the derivative  $g'(t)$  cannot exist.

It is often convenient to rephrase the definition of the derivative as

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

which obviously has the same meaning as the equation in our definition. This definition in turn leads to the following formulation:

**4.5 Proposition.** *Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an open interval, and let  $x \in I$ . Then  $f$  is differentiable at  $x$  if and only if there exists a linear function  $L : \mathbf{R} \rightarrow \mathbf{R}$  such that the function  $r$ , defined in a neighborhood of 0 by the equation  $f(x+h) = f(x) + L(h) + r(h)$ , satisfies  $\lim_{h \rightarrow 0} r(h)/h = 0$ .*

The word linear in this proposition is to be understood in the sense of linear algebra, i.e.,  $L(ah + bk) = aL(h) + bL(k)$  for all  $a, b \in \mathbf{R}$ , all  $h, k \in \mathbf{R}$ . It is trivial that any such function has the form  $L(h) = \lambda h$  for some constant  $\lambda$ . For  $L$  meeting the requirements of the proposition, of course the associated  $\lambda = f'(x)$ . The function  $L$  is called the *differential* of  $f$  at  $x$ , and is usually denoted by  $df$ , if the point  $x$  is understood, or  $df_x$  if there are various points at which we consider the differentiability of  $f$ . This way of looking at derivatives seems overly complicated now, but is the route to understanding derivatives of functions on  $\mathbf{R}^n$  later. For now, we content ourselves with the observation that the existence of the derivative at  $x$  is equivalent to a “close fit” near  $x$  between the graph of  $f$  and the straight line graph of the function  $f(x) + L$ .

## 4.2 Derivatives of Some Elementary Functions

**4.6 Proposition.** *Let  $n \in \mathbf{Z}$ , and define  $f$  by  $f(x) = x^n$  for all  $x \in \mathbf{R}$ , or all  $x \neq 0$  if  $n < 0$ . Then  $f$  is differentiable in  $\mathbf{R}$  (or in  $\mathbf{R} \setminus \{0\}$  when  $n < 0$ ), and  $f'(x) = nx^{n-1}$  for all  $x$  in the domain of  $f$ .*

**Proof.** The cases  $n = 0$  and  $n = 1$  are especially obvious. If  $f(x) = C$  for all  $x \in \mathbf{R}$ , then  $f(y) - f(x) = 0$  for any  $y, x \in \mathbf{R}$ , so  $f'(x) = 0$  for every  $x$ . If  $f(x) = x$  for all  $x$ , then  $f(y) - f(x) = y - x$ , so  $f'(x) = 1$  for all  $x$ . In general, if  $f(x) = x^n$  for some  $n \in \mathbf{N}$ , then for fixed  $x$  we have

$$f(y) - f(x) = y^n - x^n = (y - x)(y^{n-1} + y^{n-2}x + \cdots + x^{n-1}) = (y - x)\phi(y)$$

so  $f$  is differentiable at  $x$  and  $f'(x) = \phi(x) = nx^{n-1}$ .

If  $f(x) = x^{-n}$  for some  $n \in \mathbf{N}$ , then for any  $x \neq 0$ ,  $y \neq 0$ , we have

$$f(y) - f(x) = \frac{x^n - y^n}{x^n y^n} = (y - x)\phi(y),$$

where

$$\phi(y) = -\frac{y^{n-1} + y^{n-2}x + \cdots + x^{n-1}}{x^n y^n},$$

so  $f'(x) = \phi(x) = -nx^{n-1}$ . Thus the formula  $f'(x) = nx^{n-1}$  holds whenever  $f(x) = x^n$  for some  $n \in \mathbf{Z}$ . ■

We constructed the exponential and trigonometric functions in the last chapter, using infinite series. The next two propositions use the functional equations which they satisfy to compute their derivatives.

**4.7 Proposition.** *If  $E(x) = e^x$  for  $x \in \mathbf{R}$ , then  $E$  is differentiable in  $\mathbf{R}$ , and  $E' = E$ .*

**Proof.** We have the formula  $E(x+h) = E(x)E(h)$  for all  $x, h \in \mathbf{R}$ , so

$$\frac{E(x+h) - E(x)}{h} = E(x) \frac{E(h) - 1}{h},$$

from which we deduce that  $E'(x)$  exists for any  $x \in \mathbf{R}$  if and only if  $E'(0)$  exists, and then  $E'(x) = E(x)E'(0)$ . Now the formula

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

gives the inequalities

$$1 + x < E(x) < 1 + x + x^2 \sum_{n=1}^{\infty} \frac{1}{2^n} = 1 + x + x^2$$

for any  $x$  with  $0 < x < 1$ , since  $n! \geq 2^{n-1}$  for all  $n \geq 2$ . Hence

$$1 < \frac{E(h) - E(0)}{h} < 1 + h$$

for all  $h$ ,  $0 < h < 1$ , so

$$\lim_{h \rightarrow 0+} \frac{E(h) - E(0)}{h} = 1.$$

But for  $h < 0$ , we have

$$\frac{E(h) - E(0)}{h} = \frac{1/E(-h) - 1}{h} = \frac{1}{E(|h|)} \frac{E(|h|) - 1}{|h|},$$

so

$$\lim_{h \rightarrow 0-} \frac{E(h) - 1}{h} = \lim_{k \rightarrow 0+} \frac{E(k) - 1}{E(k)k} = 1.$$

Thus  $E'(0) = 1$ , so  $E'(x) = E(x)$  for all  $x$ . ■

**4.8 Proposition.** *The functions  $\sin$  and  $\cos$  are differentiable in  $\mathbf{R}$ , and  $\sin' = \cos$ ,  $\cos' = -\sin$ .*

**Proof.** We recall the inequalities (3.3) of the last chapter:

$$\frac{x^2}{2!} - \frac{x^4}{4!} < 1 - \cos x < \frac{x^2}{2!}, \quad x - \frac{x^3}{3!} < \sin x < x,$$

valid for  $0 < x < \sqrt{6}$ . It follows that

$$\frac{1}{2} - \frac{x^2}{4!} < \frac{1 - \cos x}{x^2} < \frac{1}{2}, \quad 1 - \frac{x^2}{6} < \frac{\sin x}{x} < 1$$

for all  $x$  with  $0 < x < \sqrt{6}$ , and hence for all  $x$  with  $|x| < \sqrt{6}$ , and hence that

$$\lim_{h \rightarrow 0} \frac{1 - \cos h}{h} = 0, \quad \lim_{h \rightarrow 0} \frac{\sin h}{h} = 1. \quad (4.1)$$

Now for any  $x, h \in \mathbf{R}$  we have

$$\frac{\sin(x+h) - \sin x}{h} = \sin x \frac{\cos h - 1}{h} + \cos x \frac{\sin h}{h}$$

and

$$\frac{\cos(x+h) - \cos x}{h} = \cos x \frac{\cos h - 1}{h} - \sin x \frac{\sin h}{h},$$

so the equations (4.1) give  $\sin'(x) = \cos x$ , and  $\cos'(x) = -\sin x$ . ■

## 4.3 Convex Functions

**4.9 Definition.** Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval. We say that  $f$  is convex if for every  $a, b \in I$  and every  $t$  with  $0 < t < 1$ ,

$$f(tb + (1-t)a) \leq tf(b) + (1-t)f(a).$$

For  $a = b$ , the definition says nothing. If  $a < b$ , then  $a < tb + (1-t)a < b$ ; conversely, for each  $x \in (a, b)$ , we can write  $x = tb + (1-t)a$  by taking  $t = (x-a)/(b-a)$ , so  $1-t = (b-x)/(b-a)$ . Thus the following is a simple restatement of the definition:

**4.10 Proposition.** Suppose that  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval. Then  $f$  is convex if and only if for every  $a, b \in I$  and  $a < x < b$  we have

$$f(x) \leq \frac{x-a}{b-a}f(b) + \frac{b-x}{b-a}f(a). \quad (4.2)$$

The geometric meaning of convexity is that the graph of  $f$ , between  $(a, f(a))$  and  $(b, f(b))$ , lies below the line segment joining these two points, the so-called “secant line” (for any  $a < b$  in the interval  $I$ ). This implies that the graph of  $f$  lies *above* this secant line, outside the interval  $(a, b)$ .



**4.11 Proposition.** *If  $f$  is a convex function on an interval  $I$ , and  $a < b < c \in I$ , then*

$$f(x) \geq \frac{x-a}{b-a}f(b) + \frac{b-x}{b-a}f(a) \quad \text{for } b < x < c$$

and

$$f(x) \geq \frac{c-x}{c-b}f(b) + \frac{x-b}{c-b}f(c) \quad \text{for } a < x < b.$$

**Proof.** If  $b < x < c$ , a change of notation in the inequality (4.2) gives

$$f(b) \leq \frac{b-a}{x-a}f(x) + \frac{x-b}{x-a}f(a),$$

which is equivalent to

$$f(x) \geq \frac{x-a}{b-a}f(b) + \frac{b-x}{b-a}f(a).$$

The other inequality follows similarly, when  $a < x < b$ . ■

**4.12 Corollary.** *If  $f$  is a convex function on an open interval  $I$ , then  $f$  is continuous on  $I$ .*

**Proof.** If  $b \in I$ , there exist  $a$  and  $c$  in  $I$  with  $a < b < c$ . Then the last two propositions give the inequalities

$$\frac{x-a}{b-a}f(b) + \frac{b-x}{b-a}f(a) \leq f(x) \leq \frac{x-b}{c-b}f(c) + \frac{c-x}{c-b}f(b)$$

for  $b < x < c$ , from which it follows that  $f(b+) = f(b)$ . Similarly,

$$\frac{c-x}{c-b}f(b) \leq f(x) \leq \frac{x-a}{b-a}f(b) + \frac{b-x}{b-a}f(a)$$

for  $a < x < b$ , which gives  $f(b-) = f(b)$ . Thus  $\lim_{x \rightarrow b} f(x) = f(b)$ , so  $f$  is continuous at  $b$ . ■

The hypothesis that  $I$  is open is necessary; for instance, the function  $f$  on  $[0, 1]$  defined by  $f(0) = 1$ ,  $f(t) = 0$  for  $0 < t \leq 1$  is convex. We next observe inequalities among the difference quotients of a convex function.

**4.13 Proposition.** *If  $f$  is convex on  $I$ , then*

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \leq \frac{f(b) - f(x)}{b - x} \quad (4.3)$$

whenever  $a, b \in I$  and  $a < x < b$ .

**Proof.** Simply subtract  $f(a)$  from both sides of the inequality (4.2) to get the left-hand inequality in (4.3), and subtract  $f(b)$  from both sides of (4.2) to get the right-hand inequality. ■

The geometric meaning of the inequalities (4.3) is that the slope of the secant line joining  $(a, f(a))$  and  $(b, f(b))$  for a convex function is increased if  $b$  is increased, or if  $a$  is increased.

**4.14 Theorem.** Let  $f$  be a convex function on the open interval  $I$ . Then the right- and left-hand derivatives of  $f$  at  $c$  exist for all  $c \in I$ ; denoting them by  $m_r(c)$  and  $m_l(c)$ , respectively, we have

$$m_r(a) \leq m_l(c) \leq m_r(c) \leq m_l(b)$$

whenever  $a < c < b$ .

**Proof.** Let  $c \in I$ . For any  $t, u \in I$  with  $t < c < u$ , we have by (4.3) the inequality

$$\frac{f(c) - f(t)}{c - t} \leq \frac{f(u) - f(c)}{u - c},$$

so that

$$\sup_{t < c} \frac{f(c) - f(t)}{c - t} \leq \inf_{u > c} \frac{f(u) - f(c)}{u - c}, \quad (4.4)$$

and in particular, both sides of this inequality are finite. Suppose that  $s, v \in I$ , and  $s < t < c < u < v$ . Then from the left-hand inequality in (4.3) we obtain

$$\frac{f(u) - f(c)}{u - c} \leq \frac{f(v) - f(c)}{v - c},$$

so the function  $u \mapsto (f(u) - f(c))/(u - c)$  is increasing on  $I \cap (c, +\infty)$ , and hence

$$m_r(c) = \lim_{u \rightarrow c+} \frac{f(u) - f(c)}{u - c} = \inf_{u > c} \frac{f(u) - f(c)}{u - c}$$

exists. Similarly, the right-hand inequality in (4.3) gives

$$\frac{f(c) - f(s)}{c - s} \leq \frac{f(c) - f(t)}{c - t},$$

so the function  $t \mapsto (f(c) - f(t))/(c - t)$  is increasing on  $I \cap (-\infty, c)$ , and therefore

$$m_l(c) = \lim_{t \rightarrow c-} \frac{f(c) - f(t)}{c - t} = \sup_{t < c} \frac{f(c) - f(t)}{c - t}$$

exists. We have seen in (4.4) that  $m_l(c) \leq m_r(c)$ , and that both are real numbers, so they are the left- and right-hand derivatives of  $f$  at  $c$ . Finally, suppose  $a < c < b$ . Choose  $t \in (a, c)$ . Then

$$m_r(a) \leq \frac{f(t) - f(a)}{t - a} \leq \frac{f(c) - f(t)}{c - t} \leq m_l(c),$$

and choosing  $u \in (c, b)$  we get

$$m_r(c) \leq \frac{f(u) - f(c)}{u - c} \leq \frac{f(b) - f(u)}{b - u} \leq m_l(b),$$

so everything is proven. ■

**4.15 Corollary.** *Let  $f$  be convex on the open interval  $I$ . Then there exists a countable subset  $S$  of  $I$  such that  $f'(x)$  exists for all  $x \in I \setminus S$ .*

**Proof.** Let  $S = \{x \in I : f'(x) \text{ does not exist}\}$ . For each  $x \in S$ , we have  $m_l(x) < m_r(x)$ , so we can find a rational number  $q = q(x)$  such that  $m_l(x) < q < m_r(x)$ . If  $x, y \in S$  and  $x < y$ , the inequality  $m_r(x) \leq m_l(y)$  from the last theorem shows that  $q(x) < q(y)$ . Thus the map  $x \mapsto q(x)$  is an injective map of  $S \rightarrow \mathbf{Q}$ , so  $S$  is countable. ■

**4.16 Corollary.** *Let  $f$  be convex on the open interval  $I$ , and let  $c \in I$ . For any  $m$  such that  $m_l(c) \leq m \leq m_r(c)$ , we have  $f(x) \geq f(c) + m(x - c)$  for all  $x \in I$ .*

**Proof.** If  $x > c$ , then  $f(x) - f(c) \geq (x - c)m_r(c) \geq m(x - c)$ . If  $x < c$ , then  $f(c) - f(x) \leq (c - x)m_l(c) \leq m(c - x)$ , and the desired inequality follows if we multiply by  $-1$ . For  $x = c$ , there is nothing to prove. ■

This corollary says that at each point  $(c, f(c))$  of the graph of  $f$  there exists a straight line passing through that point, and lying below the graph of  $f$ . Such a line is called a *support line* of  $f$ . If  $f'(c)$  exists, then there is only one support line, and it is the line tangent to the graph of  $f$ , i.e., the line through  $(c, f(c))$  with slope  $f'(c)$ .

## 4.4 The Differential Calculus

In this section, we obtain the short and simple set of rules for differentiating any function made up from simpler functions by the algebraic operation of addition, multiplication, and division, and the functional operation of composition. This set of rules was originally termed the “calculus of derivatives” or “differential calculus”; the term has now come to stand for the whole theory of derivatives.

**4.17 Proposition.** *Let  $f$  and  $g$  be real-valued functions on the interval  $I$ , and suppose that  $f$  and  $g$  are differentiable at a point  $x \in I$ . Then  $f + g$  and  $fg$  are differentiable at  $x$ , and  $(f + g)'(x) = f'(x) + g'(x)$ ,  $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ .*

**Proof.** By Proposition 4.2, we can write  $f(y) - f(x) = (y - x)\phi(y)$  and  $g(y) - g(x) = (y - x)\psi(y)$ , where  $\phi$  and  $\psi$  are continuous at  $x$ . Then

$$(f + g)(y) - (f + g)(x) = f(y) + g(y) - f(x) - g(x) = (y - x)(\phi(y) + \psi(y))$$

and

$$\begin{aligned}(fg)(y) - (fg)(x) &= f(y)g(y) - f(x)g(y) + f(x)g(y) - f(x)g(x) \\ &= (y - x)(\phi(y)g(y) + f(x)\psi(y)),\end{aligned}$$

and since the sum and product of functions continuous at  $x$  are again continuous at  $x$ , Proposition 4.2 gives  $(f + g)'(x) = \phi(x) + \psi(x) = f'(x) + g'(x)$ , and  $(fg)'(x) = \phi(x)g(x) + f(x)\psi(x) = f'(x)g(x) + f(x)g'(x)$ . ■

We note that the formula  $f'(x) = nx^{n-1}$  when  $f(x) = x^n$ , established above by direct calculation, can also be proved for positive  $n$  by induction using the last proposition, and the trivial result for  $n = 1$ . It now follows that for any polynomial function, i.e.,  $f$  defined by  $f(x) = \sum_{k=0}^n a_k x^k$ , we have  $f'(x) = \sum_{k=1}^n k a_k x^{k-1}$ .

The next result is known as the *chain rule*.

**4.18 Proposition.** Let  $I$  and  $J$  be intervals in  $\mathbf{R}$ , and  $f : I \rightarrow \mathbf{R}$ ,  $f(I) \subset J$ , and  $g : J \rightarrow \mathbf{R}$ . If  $f$  is differentiable at  $x$ , and  $g$  is differentiable at  $f(x)$ , then  $g \circ f$  is differentiable at  $x$ , and  $(g \circ f)'(x) = g'(f(x))f'(x)$ .

**Proof.** Let  $u = f(x)$ . There exists  $\phi : I \rightarrow \mathbf{R}$ , continuous at  $x$ , such that  $f(y) - f(x) = (y - x)\phi(y)$  for all  $y \in I$ , and there exists  $\psi : J \rightarrow \mathbf{R}$ , such that  $\psi$  is continuous at  $u$ , and  $g(v) - g(u) = (v - u)\psi(v)$  for all  $v \in J$ . Then, putting  $h = g \circ f$ ,

$$\begin{aligned}h(y) - h(x) &= g(f(y)) - g(f(x)) = (f(y) - f(x))\psi(f(y)) \\ &= (y - x)\phi(y)\psi(f(y)),\end{aligned}$$

and since the composition of continuous functions and the product of continuous functions are continuous, it follows from Proposition 4.1 that  $h$  is differentiable at  $x$ , with  $h'(x) = \phi(x)\psi(f(x)) = g'(f(x))f'(x)$ . ■

The classical notation, introduced by Leibniz, used the language of *variables*, a concept we have avoided in favor of numbers and functions. This notation reads: if the variable  $y$  is a function of the variable  $x$ , i.e.,  $y = f(x)$ , we denote  $f'(x)$  by  $dy/dx$ . This notation has one great advantage: if also  $z = g(y)$ , so that  $z$  is indirectly a function of  $x$ , then

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx},$$

a wonderful mnemonic for the chain rule. We can also make sense of this notation using the notion of differentials introduced in the first section of this chapter. To do this, we use  $x$  to denote the identity function on  $\mathbf{R}$ , i.e.,  $x(t) = t$  for all  $t \in \mathbf{R}$ , rather than having this letter stand for a real number, as we have been doing. It is easy to see that the differential of this function is the identity map of  $\mathbf{R}$  again, i.e.,  $dx(h) = h$  for all  $h \in \mathbf{R}$ . Let  $y$  be a function, differentiable at  $c$ . Recalling that  $dy_c(h) = y'(c)h$  for all  $h \in \mathbf{R}$ , we see that in fact  $y' = dy/dx$ , in the sense that  $y'(c) = dy_c(h)/dx_c(h)$ , for any  $h \neq 0$ .

**4.19 Example.** Let  $f$  and  $g$  be real-valued functions on an interval  $I$ , and suppose that, for all  $x \in I$ ,  $f(x) = [g(x)]^n$  for some  $n \in \mathbf{Z}$ . Then if  $g$  is differentiable at  $c$ , so is  $f$ , and  $f'(c) = n[g(c)]^{n-1}g'(c)$ . In particular, with  $n = -1$  we find that  $(1/g)'(c) = -g'(c)/g(c)^2$ . Combined with the rule for differentiating products, we find the formula  $(f/g)'(c) = [g(c)f'(c) - f(c)g'(c)]/g(c)^2$ .

Similarly, if  $f(x) = e^{g(x)}$  for all  $x \in I$ , then  $f'(c) = f(c)g'(c)$ .

**4.20 Definition.** Let  $I$  be an interval, and  $f : I \rightarrow \mathbf{R}$ . We say that  $f$  has a local maximum (or local minimum) at  $c \in I$  if there exists a neighborhood  $U$  of  $c$ , relative to  $I$ , such that  $f(c) \geq f(x)$  for every  $x \in U$  (respectively,  $f(c) \leq f(x)$  for every  $x \in U$ ).

**4.21 Proposition.** Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval. If  $f$  has a local maximum or minimum at  $c \in I$ , where  $c$  is not an endpoint of  $I$ , and if  $f$  is differentiable at  $c$ , then  $f'(c) = 0$ .

**Proof.** We may assume that  $f$  has a local maximum at  $c$ . Then there exists  $\delta > 0$  such that  $f(t) - f(c) \leq 0$  for every  $t$  with  $|t - c| < \delta$ . It follows that

$$\frac{f(t) - f(c)}{t - c} \leq 0$$

for every  $t$  with  $c < t < c + \delta$ , so  $f'(c) \leq 0$ . But it also follows that

$$\frac{f(t) - f(c)}{t - c} \geq 0$$

for every  $t$  with  $c - \delta < t < c$ , so  $f'(c) \geq 0$ . Thus  $f'(c) = 0$ . ■

The next result is known as the mean value theorem of the differential calculus, and is fundamental to the theory of derivatives.

**4.22 Theorem.** Let  $f : [a, b] \rightarrow \mathbf{R}$  be continuous on  $[a, b]$ , and differentiable in  $(a, b)$ . Then there exists  $\xi \in (a, b)$  such that  $f(b) - f(a) = f'(\xi)(b - a)$ .

**Proof.** Define the function  $g$  on  $[a, b]$  by

$$g(t) = f(t) - f(a) - \frac{f(b) - f(a)}{b - a}(t - a).$$

Then  $g$  is continuous on  $[a, b]$  and differentiable in  $(a, b)$ , and  $g(a) = g(b) = 0$ . We want to show that  $g'(\xi) = 0$  for some  $\xi \in (a, b)$ . If  $g(t) = 0$  for all  $t \in (a, b)$ , we may choose any  $\xi \in (a, b)$ . Suppose  $g(t) > 0$  for some  $t \in (a, b)$ . By Theorem 3.14, there exists  $\xi \in [a, b]$  such that  $g$  has a maximum value at  $\xi$ , and we must have  $\xi \in (a, b)$ . According to Theorem 4.21, we have  $g'(\xi) = 0$ . The argument if  $g(t) < 0$  for some  $t$  is similar. ■

**4.23 Corollary.** Let  $f : [a, b] \rightarrow \mathbf{R}$  be continuous on  $[a, b]$ , and differentiable in  $(a, b)$ . If  $f'(t) \geq 0$  (respectively,  $f'(t) > 0$ ) for all  $t \in (a, b)$ , then  $f$  is increasing (respectively, strictly increasing) on  $[a, b]$ . If  $f'(t) \leq 0$  (respectively,  $f'(t) < 0$ ) for all  $t \in (a, b)$ , then  $f$  is decreasing (respectively, strictly decreasing) on  $[a, b]$ . If  $f'(t) = 0$  for all  $t \in (a, b)$ , then  $f$  is constant on  $[a, b]$ .

**Proof.** For any  $x, y$  with  $a \leq x < y \leq b$ , we can apply the last theorem to get  $f(y) - f(x) = f'(\xi)(y - x)$  for some  $\xi$  with  $x < \xi < y$ . ■

We remark that if we only know that  $f'(x) > 0$ , it follows that there exists  $\delta > 0$  such that  $f(y) > f(x)$  for all  $y$  with  $x < y < x + \delta$ , and  $f(y) < f(x)$  if  $x - \delta < y < x$  (proof?), but it does not follow that  $f$  is increasing on any interval containing  $x$  (see the Exercises.)

It does not follow from the existence of  $f'(x)$  in an interval that the function  $f'$  is continuous in that interval.

**4.24 Example.** Let  $g : \mathbf{R} \rightarrow \mathbf{R}$  be defined by  $g(0) = 0$ , and  $g(x) = x \sin(1/x)$  for  $x \neq 0$ . Let  $f(x) = xg(x)$ . It is easy to see that  $g$  is continuous at 0, as well as differentiable at any  $x \neq 0$ , so  $f$  is differentiable everywhere, and  $f'(0) = g(0) = 0$ . Since  $f'(x) = 2x \sin(1/x) - \cos(1/x)$  for  $x \neq 0$ , we see that  $\lim_{x \rightarrow 0} f'(x)$  does not exist, so  $f'$  is not continuous at 0.

However, if  $f'$  exists throughout an interval, it does have one property associated with continuous functions, namely, the intermediate value property.

**4.25 Theorem.** Let  $f : [a, b] \rightarrow \mathbf{R}$  be differentiable at each point of  $[a, b]$ . If  $f'(a) < y < f'(b)$ , or if  $f'(a) > y > f'(b)$ , there exists  $x \in (a, b)$  such that  $f'(x) = y$ .

**Proof.** We consider only the case  $f'(a) < y < f'(b)$ ; the other case follows the same pattern, or can be deduced from this one. By the definition of the derivative, there exists  $h$ ,  $0 < h < b - a$ , such that

$$\frac{f(a+h) - f(a)}{h} < y \quad \text{and} \quad \frac{f(b) - f(b-h)}{h} > y.$$

Define  $g : [a, b - h] \rightarrow \mathbf{R}$  by  $g(t) = (1/h)(f(t + h) - f(t))$ . Since  $f$  is differentiable on  $[a, b]$ , it is continuous there, and it follows that  $g$  is continuous on  $[a, b - h]$ . Since  $g(a) < y$  and  $g(b - h) > y$ , it follows from Theorem 3.15 that there exists  $c \in (a, b - h)$  such that  $g(c) = y$ . Applying the mean value theorem (Theorem 4.22) we have

$$y = \frac{f(c + h) - f(c)}{h} = f'(x)$$

for some  $x \in (c, c + h) \subset (a, b)$ . ■

We next establish an analogue of an earlier result for continuous functions.

**4.26 Theorem.** *Let  $f$  be differentiable in the open interval  $I$ , and suppose  $f' > 0$  in  $I$ . Then  $f$  maps  $I$  bijectively onto an open interval  $J$ , and the inverse map  $g : J \rightarrow I$  is differentiable in  $J$ , with  $g'(u) = [f'(g(u))]^{-1}$  for all  $u \in J$ .*

**Proof.** According to Corollary 4.23,  $f$  is strictly increasing on  $I$ , and  $f$  is continuous since it is differentiable, so (by Theorem 3.16)  $f$  maps  $I$  bijectively onto an interval  $J$ , and the inverse map  $g : J \rightarrow I$  is continuous. We want to show that  $g$  is differentiable in  $J$ . Let  $u \in J$ , so  $u = f(x)$  for some  $x = g(u) \in I$ . Since  $f$  is differentiable at  $x$ , there exists  $\phi : I \rightarrow \mathbf{R}$  such that  $\phi$  is continuous at  $x$ , and  $f(y) - f(x) = (y - x)\phi(y)$  for all  $y \in I$ . We note that  $\phi(x) = f'(x) > 0$  and  $\phi(y) > 0$  for all  $y \neq x$  since  $f$  is strictly increasing. Then for any  $v \in J$ , taking  $y = g(v)$ , we have  $g(v) - g(u) = (v - u)\psi(v)$ , where  $\psi(v) = 1/\phi(y) = 1/\phi(g(v))$ . As we have seen,  $\psi$  is well-defined on  $J$ , and  $\psi$  is continuous at  $u$  since  $g$  is continuous, and  $\phi$  is continuous at  $x = g(u)$ . Thus  $g$  is differentiable at  $u$ , and  $g'(u) = \psi(u) = 1/\phi(x) = 1/f'(x)$ . ■

Of course, the same conclusions can be drawn if we assume that  $f' < 0$  on  $I$ .

**4.27 Example.** If we take  $f(x) = e^x$ , then  $f' = f > 0$  so Theorem 4.26 applies, yielding the fact that the inverse function  $g$  is differentiable everywhere on its domain  $(0, +\infty)$ , with  $g'(x) = 1/f'(g(x)) = 1/f(g(x)) = 1/x$ . Thus  $\log' x = 1/x$ .

Fix  $p > 0$ , and let  $f(x) = x^p = e^{p \log x}$ . Applying the chain rule, Proposition 4.18, we find that  $f'(x) = (p/x)e^{p \log x} = px^{p-1}$ , thus extending the formula from the case we earlier established ( $p \in \mathbf{Z}$ ).

**4.28 Example.** Let

$$f(x) = \tan x = \frac{\sin x}{\cos x}, \quad -\frac{\pi}{2} < x < \frac{\pi}{2};$$

then using Proposition 4.17 we find

$$f'(x) = \frac{\cos x \cos x - \sin x(-\sin x)}{\cos^2 x} = \frac{1}{\cos^2 x} > 0$$

for all  $x$  in the domain of  $f$ ; clearly the range of  $f$  is the entire line  $\mathbf{R}$ , and the inverse function  $g$  satisfies

$$g'(x) = \frac{1}{f'(g(x))} = \cos^2(g(x));$$

using the identity  $1/\cos^2 u = 1 + \tan^2 u$  that follows from  $\sin^2 u + \cos^2 u = 1$ , we find  $g'(x) = (1 + x^2)^{-1}$  for all  $x \in \mathbf{R}$ . The function  $g$  is called the arctangent function, and we usually denote  $g(x)$  by  $\arctan x$ , sometimes by  $\tan^{-1} x$ .

## 4.5 L'Hospital's Rule

We begin with a generalization of the mean value theorem (Theorem 4.22).

**4.29 Theorem.** *Let  $f$  and  $g$  be continuous real-valued functions on the interval  $[a, b]$ , differentiable in  $(a, b)$ . Then there exists  $c \in (a, b)$  such that  $f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)]$ .*

**Proof.** Define  $\phi : [a, b] \rightarrow \mathbf{R}$  by

$$\phi(t) = [f(t) - f(a)][g(b) - g(a)] + [g(b) - g(t)][f(b) - f(a)],$$

so  $\phi$  is continuous on  $[a, b]$ , differentiable in  $(a, b)$ , and we see that  $\phi(a) = [g(b) - g(a)][f(b) - f(a)] = \phi(b)$ . It follows from Theorem 4.22 that  $\phi'(c) = 0$  for some  $c \in (a, b)$ , and this gives the desired relation. ■

The geometric interpretation of this theorem is that if  $t \mapsto (f(t), g(t))$  ( $a \leq t \leq b$ ) is a differentiable curve in the plane  $\mathbf{R}^2$ , there exists at least one point on the curve where the tangent line is parallel to the line through  $(a, f(a))$  and  $(b, f(b))$ . We will discuss curves, surfaces, etc., and their tangents in a later chapter.

**4.30 Theorem.** *Let  $-\infty \leq a < b \leq +\infty$ , and let  $f$  and  $g$  be differentiable in  $(a, b)$ , with  $g'(t) \neq 0$  for all  $t \in (a, b)$ . Suppose that*

$$\lim_{x \rightarrow a+} \frac{f'(x)}{g'(x)} = L, \quad (4.5)$$

where  $-\infty \leq L \leq +\infty$ , and that either

$$f(a+) = g(a+) = 0 \quad (4.6)$$



or that

$$\lim_{x \rightarrow a+} |g(x)| = +\infty. \quad (4.7)$$

Then

$$\lim_{x \rightarrow a+} \frac{f(x)}{g(x)} = L.$$

**Proof.** We begin by observing that if  $a < s < t < b$ , we have  $g(t) - g(s) = g'(v)(t - s)$  for some  $v \in (s, t)$ , and hence  $g(s) \neq g(t)$ . Therefore we can deduce from Theorem 4.29 that there exists  $u \in (s, t)$ , such that

$$\frac{f(t) - f(s)}{g(t) - g(s)} = \frac{f'(u)}{g'(u)}. \quad (4.8)$$

Let us first consider the case when  $L$  is a real number. Let  $\epsilon > 0$ , and choose  $\eta$  with  $0 < \eta < \epsilon$ . There exists  $c \in (a, b)$  such that

$$\frac{f'(u)}{g'(u)} \in (L - \eta, L + \eta)$$

for every  $u \in (a, c)$ , and hence from equation (4.8)

$$L - \eta < \frac{f(t) - f(s)}{g(t) - g(s)} < L + \eta \quad (4.9)$$

for every  $s, t$  with  $a < s < t \leq c$ .

Suppose that (4.6) holds. Taking the limit as  $s \rightarrow a+$ , we have  $L - \epsilon < L - \eta \leq f(t)/g(t) \leq L + \eta < L + \epsilon$  for every  $t$ ,  $a < t < c$ . Thus  $\lim_{t \rightarrow a+} f(t)/g(t) = L$ .

Now suppose that (4.7) holds. We may suppose that  $g(s) \rightarrow +\infty$  as  $s \rightarrow a+$ , and that  $g(c) > 0$ . Take  $t = c$  in the inequalities (4.9), and multiply by  $(g(s) - g(c))/g(s)$ ; provided this multiplier is positive, we get

$$(L - \eta) \left(1 - \frac{g(c)}{g(s)}\right) < \frac{f(s) - f(c)}{g(s)} < (L + \eta) \left(1 - \frac{g(c)}{g(s)}\right).$$

For any  $\delta > 0$ , we can find  $d$  with  $a < d < c$  such that  $|f(c)|/g(s) < \delta$  and  $g(c)/g(s) < \delta$  for all  $s$  with  $a < s < d$ . Then we have

$$(L - \eta)(1 - \delta) - \delta < \frac{f(s)}{g(s)} < (L + \eta)(1 - \delta) + \delta$$

for all  $s \in (a, d)$ , and if  $\delta$  is chosen small enough, this implies that  $L - \epsilon < f(s)/g(s) < L + \epsilon$  for all  $s \in (a, d)$ . Thus  $f(s)/g(s) \rightarrow L$  as  $s \rightarrow a+$ .

Next we consider the case  $L = +\infty$ . Let  $M > 0$ . There exists  $c \in (a, b)$  such that  $f'(u)/g'(u) > 2M$  for all  $u \in (a, c)$ , and it follows from equation (4.8) that

$$\frac{f(t) - f(s)}{g(t) - g(s)} > 2M, \quad (4.10)$$

whenever  $a < s < t < c$ .

If the hypothesis (4.6) holds, by letting  $s \rightarrow a+$  in (4.10) we obtain  $f(t)/g(t) \geq 2M > M$  for all  $t$ ,  $a < t \leq c$ , and since  $M$  was arbitrary, we have  $f(t)/g(t) \rightarrow +\infty$  as  $t \rightarrow a+$ .

Suppose now that the hypothesis (4.7) holds. We may suppose that  $g(s) \rightarrow +\infty$  as  $s \rightarrow a+$ , and that  $g(t) > 0$  for all  $t \in (a, c)$ . Choose  $d \in (a, c)$  such that  $|f(c)|/g(s) < 1/2$  and  $g(c)/g(s) < 1/2$  for every  $s$  with  $a < s < d$ . We fix  $t = c$  in the inequality (4.10) and multiply by  $1 - g(c)/g(s)$  to get

$$\frac{f(s) - f(c)}{g(s)} > 2M \left(1 - \frac{g(c)}{g(s)}\right),$$

so

$$\frac{f(s)}{g(s)} > M - \frac{1}{2},$$

whenever  $a < s < d$ . Since  $M$  was arbitrary, we conclude that  $f(s)/g(s) \rightarrow +\infty$  as  $s \rightarrow a+$ .

The proof for the case  $L = -\infty$  is left to the reader. ■

We stated Theorem 4.30 for limits from the right; of course, the corresponding theorem for limits from the left is true. It can be deduced by a simple change of variable, and does not need a separate statement.

## 4.6 Higher Order Derivatives

If  $f'(t)$  exists for every  $t$  in an interval  $I$ , we can raise the question of the existence of the derivative of  $f'$  at points of  $I$ .

**4.31 Definition.** Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval. If  $f'(t)$  exists for all  $t$  in some neighborhood (relative to  $I$ ) of  $c \in I$ , and if the derivative of  $f'$  at  $c$  exists, we denote it by  $f''(c)$ . Inductively, we define  $f^{(0)} = f$ ,  $f^{(1)} = f'$ , and for each positive integer  $n$  we define  $f^{(n)}(t)$ , the  $n$ th order derivative of  $f$  at  $t$ , to be the derivative at  $t$  (if it exists) of the function  $f^{(n-1)}$ .

Thus the existence of  $f^{(n)}$  in an interval  $I$  implies the existence, and continuity, of  $f^{(k)}$  in  $I$  for  $k = 0, 1, \dots, n-1$ . One more notation:

**4.32 Definition.** Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval. We say that  $f$  is of class  $C^n$  in  $I$ , and write  $f \in C^n(I)$ , if  $f^{(n)}$  is defined on all of  $I$ , and is continuous. We say  $f$  is of class  $C^\infty$  if it is of class  $C^n$  for every  $n$ .

The main theorem involving higher order derivatives is the following, known as Taylor's theorem. It can be thought of as an extension of the mean value theorem, or as a statement concerning the local approximation

of an  $n$ -times differentiable function by polynomials of degree  $\leq n$ . From the formula for differentiating the monomial  $x^n$ , we quickly see that the polynomial

$$P(x) = \sum_{k=0}^n a_k \frac{(x-c)^k}{k!}$$

has the property that  $P^{(k)}(c) = a_k$  for  $0 \leq k \leq n$ , and  $P^{(k)}(x) = 0$  for all  $k > n$ , and all  $x$ . The polynomial  $P_n$  in the next theorem (called the Taylor polynomial for  $f$  at  $c$ ) is thus the unique polynomial of order  $\leq n$  whose  $k$ th order derivatives at  $c$  agree with those of  $f$  at  $c$ , for each  $k \leq n$ . The theorem gives a formula for the error made in approximating  $f$  by  $P_n$ .

**4.33 Theorem.** *Let  $I$  be an interval, and suppose that  $f : I \rightarrow \mathbf{R}$  satisfies the condition that  $f^{(n+1)}(t)$  exists at each  $t \in I$ . Fix  $c \in I$ , and define*

$$P_n(t) = \sum_{k=0}^n \frac{f^{(k)}(c)}{k!} (t-c)^k.$$

*Then for each  $x \in I$ , there exists  $\xi$  between  $c$  and  $x$  such that*

$$f(x) = P_n(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-c)^{n+1}. \quad (4.11)$$

**Proof.** Let

$$g(t) = f(t) - P_n(t) - C(t-c)^{n+1},$$

where  $C$  is chosen so that  $g(x) = 0$ . We observe that  $g$  has derivatives up to order  $n+1$  on  $I$ , that  $g(c) = g'(c) = \dots = g^{(n)}(c) = 0$ , and that  $g^{(n+1)}(t) = f^{(n+1)}(t) - C(n+1)!$ . We also note that the conclusion of the theorem can be stated as  $C = f^{(n+1)}(\xi)/(n+1)!$  for some  $\xi$  between  $c$  and  $x$ , and hence is equivalent to the claim that  $g^{(n+1)}(\xi) = 0$  for some  $\xi$  between  $c$  and  $x$ . Now since  $g(c) = g(x) = 0$ , there exists by the mean value theorem some  $\xi_1$  between  $c$  and  $x$  such that  $g'(\xi_1) = 0$ . Having found  $\xi_k$  between  $c$  and  $x$  such that  $g^{(k)}(\xi_k) = 0$ , where  $0 \leq k \leq n$ , we can find by the mean value theorem  $\xi_{k+1}$  between  $c$  and  $\xi_k$  such that  $g^{(k+1)}(\xi_{k+1}) = 0$ . Thus with  $\xi = \xi_{n+1}$ , we have  $\xi$  between  $c$  and  $x$  with  $g^{(n+1)}(\xi) = 0$ . ■

**4.34 Example.** Suppose  $f$  is a function on  $\mathbf{R}$  satisfying the equation  $f' = kf$ , for some constant  $k$ . Then we find inductively that  $f^{(n)}(t) = k^n f(t)$  for all  $t$ , and every  $n$ . Since  $f$  is continuous, by virtue of being differentiable, we see that  $f$  is of class  $C^\infty$ . By Taylor's theorem, we have for every positive integer  $n$  and every  $x \in \mathbf{R}$ ,

$$f(x) = f(0) \sum_{j=0}^n \frac{k^j x^j}{j!} + \frac{k^{n+1} f(\xi)}{(n+1)!} x^{n+1}$$

for some  $\xi$  between 0 and  $x$ . Let  $R > 0$ . Since  $f$  is continuous on  $[-R, R]$ , it is bounded there, say  $|f(t)| \leq M$  for  $-R \leq t \leq R$ . Then

$$\left| f(x) - f(0) \sum_{j=0}^n \frac{k^j x^j}{j!} \right| \leq M \frac{(|kx|)^{n+1}}{(n+1)!} \leq M \frac{(|k|R)^{n+1}}{(n+1)!}$$

for all  $x \in [-R, R]$ . Thus

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

the series converging uniformly on every bounded interval. In other words,  $f(x) = f(0)e^{kx}$  for all  $x \in \mathbf{R}$ . (An exercise asks you to prove this in a more elementary way.)

## 4.7 Analytic Functions

**4.35 Proposition.** Let  $(a_n)_{n=0}^{\infty}$  be a sequence in  $\mathbf{R}$ . Define  $R$  by the equation  $1/R = \limsup |a_n|^{1/n}$ , where we put  $1/0 = +\infty$  and  $1/\infty = 0$ . Then for each  $c \in \mathbf{R}$ , the series

$$\sum_{n=0}^{\infty} a_n (x - c)^n$$

converges absolutely for every  $x \in \mathbf{R}$  such that  $|x - c| < R$ , and diverges for every  $x$  with  $|x - c| > R$ .

**Proof.** This follows at once from the root test (Theorem 2.36). ■

We call a series of this form a *power series*, and the extended real  $R$  of the proposition is called the *radius of convergence* of the power series. The next result is that a convergent power series can be differentiated term-by-term.

**4.36 Theorem.** Let  $\sum_{n=0}^{\infty} a_n (x - c)^n$  be a power series with positive radius of convergence  $R$ , and let  $I = (c - R, c + R)$ . If  $f$  is defined by  $f(x) = \sum_{n=0}^{\infty} a_n (x - c)^n$  for  $x \in I$ , then  $f$  is differentiable in  $I$ , and

$$f'(x) = \sum_{n=1}^{\infty} n a_n (x - c)^{n-1}$$

for all  $x \in I$ .

**Proof.** It clearly suffices to consider the case  $c = 0$ ,  $I = (-R, R)$ . Let  $x \in I$ . Choose  $r$  with  $|x| < r < R$ , and let  $\delta = r - |x|$ . For any  $y \in I$ , we have

$$f(y) - f(x) = \sum_{n=0}^{\infty} a_n y^n - \sum_{n=0}^{\infty} a_n x^n = \sum_{n=1}^{\infty} a_n (y^n - x^n),$$

since convergent series can be added term-by-term. Define  $\phi_1(y) = 1$  for all  $y$ , and for  $n > 1$  let  $\phi_n(y) = \sum_{k=0}^{n-1} x^k y^{n-1-k}$ , so  $y^n - x^n = (y-x)\phi_n(y)$  for all  $y$ , and every  $n \geq 1$ . We observe that if  $|y-x| < \delta$ , then  $|y| < r$  and  $|x| < r$ , so  $|\phi_n(y)| < nr^{n-1}$ . Since  $r < R$  and  $n^{1/n} \rightarrow 1$  as  $n \rightarrow \infty$ , the series  $\sum_{n=1}^{\infty} na_n r^{n-1}$  converges. It follows by the Weierstrass  $M$ -test (Theorem 3.26) that the series  $\sum_{n=1}^{\infty} a_n \phi_n$  converges uniformly on  $J = \{y : |y-x| < \delta\}$ , and hence its sum  $\phi$  is continuous on  $J$ . Since we have seen that  $f(y) - f(x) = (y-x)\phi(y)$ , it follows that  $f$  is differentiable at  $x$ , and  $f'(x) = \phi(x) = \sum_{n=1}^{\infty} na_n x^{n-1}$ . ■

**4.37 Corollary.** *If the power series  $\sum_{n=0}^{\infty} a_n(x-c)^n$  has a positive radius of convergence  $R$ , then its sum  $f$  is a function of class  $C^\infty$  on  $(c-R, c+R)$ , and  $a_k = f^{(k)}(c)/k!$  for every nonnegative integer  $k$ .*

**Proof.** The last theorem shows that  $f'$  can be expressed as a power series with radius of convergence  $R$ , and hence itself has a derivative of the same kind. Clearly,  $f'(0) = a_1$ . By induction, we prove that

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)a_n(x-c)^{n-k}$$

for every  $k$  and every  $x \in (c-R, c+R)$ , and hence  $f^{(k)}(c) = k!a_k$ . ■

**4.38 Definition.** *Let  $I$  be an open interval in  $\mathbf{R}$ , and let  $f : I \rightarrow \mathbf{R}$ . We say that  $f$  is analytic in  $I$  if for each  $c \in I$  there exists a sequence  $(a_n)_{n=0}^{\infty}$  in  $\mathbf{R}$ , and  $\delta > 0$ , such that  $f(x) = \sum a_n(x-c)^n$  for all  $x$  with  $|x-c| < \delta$ .*

The last corollary implies that if  $f$  is analytic on  $I$ , then  $f \in C^\infty(I)$ . The converse to this statement is false.

**4.39 Example.** Define  $f : \mathbf{R} \rightarrow \mathbf{R}$  by

$$f(x) = \begin{cases} 0 & \text{for all } x \leq 0, \\ e^{-1/x} & \text{for } x > 0. \end{cases}$$

Clearly,  $f(0+) = 0$ , so  $f$  is continuous at 0, and clearly  $f$  is of class  $C^\infty$  when restricted to  $(-\infty, 0)$  or to  $(0, +\infty)$ . We will show that  $f \in C^\infty(\mathbf{R})$ ; it suffices to show that  $f^{(n)}(0) = 0$  for all  $n$ , since this implies that  $f^{(n-1)}$  is continuous at 0, and hence everywhere. We begin by calculating

$$\lim_{h \rightarrow 0+} \frac{f(h)}{h} = \lim_{x \rightarrow \infty} x e^{-x} = 0,$$

so  $f'(0)$  exists and equals 0. Next we observe that  $f'(x) = x^{-2}e^{-1/x}$  for  $x > 0$ , and, by an easy induction argument, that  $f^{(n)}(x) = p_n(1/x)f(x)$  for  $x > 0$ , where  $p_n$  is a polynomial. Hence

$$\lim_{h \rightarrow 0+} \frac{f^{(n)}(h)}{h} = \lim_{h \rightarrow 0+} (1/h)p_n(1/h)e^{-1/h} = \lim_{x \rightarrow +\infty} x p_n(x)e^{-x}$$

for every  $n$ . To show that this last limit is 0 it will suffice to show that  $\lim_{x \rightarrow +\infty} x^k e^{-x} = 0$  for every nonnegative  $k$ . But this follows at once from the inequality  $e^x > x^{k+1}/(k+1)!$  for  $x > 0$ . (It would be overkill to drag in L'Hospital's rule at this point.) Thus  $f^{(n)}(0) = 0$  for every  $n$ . But this shows that  $f$  cannot be analytic, for if  $f$  had a power series representation  $f(x) = \sum a_n x^n$  in a neighborhood  $(-\delta, \delta)$  of 0, we would have  $a_n = f^{(n)}(0)/n! = 0$  for every  $n$ , which would give  $f(x) = 0$  for all  $x \in (-\delta, \delta)$ , contradicting the fact that  $f(x) > 0$  for  $x > 0$ .

Our last result in this section says that a function given by a power series converging in an interval is analytic in that interval.

**4.40 Proposition.** Suppose  $f(x) = \sum_{n=0}^{\infty} a_n x^n$ , where the series converges for  $-R < x < R$ . Then for each  $c \in (-R, R)$ , there is a sequence  $(b_n)$  such that  $f(x) = \sum_{k=0}^{\infty} b_k (x-c)^k$ , converging for  $|x-c| < R-|c|$ .

**Proof.** Since

$$x^n = (x-c+c)^n = \sum_{k=0}^n \binom{n}{k} (x-c)^k c^{n-k},$$

we have formally

$$\begin{aligned} \sum_{n=0}^{\infty} a_n x^n &= \sum_{n=0}^{\infty} \sum_{k=0}^n a_n \binom{n}{k} c^{n-k} (x-c)^k \\ &= \sum_{0 \leq k \leq n} a_n \binom{n}{k} c^{n-k} (x-c)^k \\ &= \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} a_n \binom{n}{k} c^{n-k} (x-c)^k. \end{aligned}$$

Thus, if we put

$$b_k = \sum_{n=k}^{\infty} a_n \binom{n}{k} c^{n-k},$$

we have  $f(x) = \sum_{k=0}^{\infty} b_k (x-c)^k$ , provided that the interchange of order of summation that we have carried out can be justified. Now if we let

$$B_k = \sum_{n=k}^{\infty} |a_n| \binom{n}{k} |c|^{n-k},$$

we have  $|b_k| \leq B_k$ , and

$$\sum_{k=0}^{\infty} B_k t^k = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} |a_n| \binom{n}{k} |c|^{n-k} t^k$$

$$\begin{aligned}
&= \sum_{n=0}^{\infty} a_n \sum_{k=0}^n \binom{n}{k} |c|^{n-k} t^k \\
&= \sum_{n=0}^{\infty} |a_n| (|c| + t)^n,
\end{aligned}$$

which converges for  $t > 0$ ,  $|c| + t < R$ . Theorem 2.58, which justified the equation above, now justifies the interchange of order of summation which gives the desired equation

$$\sum_{k=0}^{\infty} b_k (x - c)^k = \sum_{n=0}^{\infty} a_n x^n,$$

whenever  $|x - c| + |c| < R$ . ■

We observe in passing that the converse to this theorem is false; a function analytic in an interval  $(-R, R)$  need not have a power series expansion converging throughout that interval. The function defined by  $f(x) = 1/(1 + x^2)$  is perhaps the simplest example; it is analytic on all of  $\mathbf{R}$ , but its power series expansion at 0 converges only in the interval  $(-1, 1)$ .

## 4.8 Exercises

1. Let  $f(0) = 0$ , and  $f(x) = x/(1 + e^{1/x})$  for  $x \neq 0$ . Find the right- and left-hand derivatives of  $f$  at 0.
2. Use Corollary 4.23 to show that if  $f$  is a differentiable function on an interval  $I$  satisfying the equation  $f' = kf$ , for some constant  $k$ , then there exists a constant  $C$  such that  $f(x) = Ce^{kx}$  for all  $x \in I$ . HINT: Consider the function  $g$  defined by  $g(x) = e^{-kx} f(x)$ .
3. Show that the formulas for  $\sin'$  and  $\cos'$ , together with the equations  $\sin 0 = 0$  and  $\cos 0 = 1$ , imply that  $\sin^2 x + \cos^2 x = 1$  for all  $x \in \mathbf{R}$ .
4. Let  $f(x) = (x^2 - 1)^n$ , and let  $g = f^{(n)}$ . Show that the polynomial  $g$  has  $n$  distinct real roots, all in the interval  $[-1, 1]$ .
5. Suppose that  $f$  is differentiable on  $(a, +\infty)$ . Show that if  $f'(x) \rightarrow L$  as  $x \rightarrow +\infty$ , where  $-\infty \leq L \leq +\infty$ , then  $f(x)/x \rightarrow L$  as  $x \rightarrow +\infty$ . Deduce that if  $f(x) \rightarrow M$  as  $x \rightarrow +\infty$ , where  $M$  is real, and  $f'(x) \rightarrow L$  as  $x \rightarrow +\infty$ , where  $-\infty \leq L \leq +\infty$ , then  $L = 0$ .
6. If  $P(x) = \sum_{k=0}^n a_k x^k$ , and

$$a_0 + \frac{a_1}{2} + \cdots + \frac{a_n}{n+1} = 0,$$

show that there exists  $x$  with  $0 < x < 1$  and  $P(x) = 0$ .

7. Let  $f$  be differentiable on an interval  $I$ , and suppose that  $f'$  is an increasing function on  $I$ . Show that  $f$  is convex. Deduce that if  $f'' \geq 0$  on  $I$ , then  $f$  is convex.

8. Let  $p > 1$ , and put  $q = p/(p-1)$ , so  $1/p + 1/q = 1$ . Show that for any  $x > 0$ ,  $y > 0$ , we have

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q},$$

and find the case where equality holds.

9. Let  $a > 1$ . Show that for all  $x \geq 0$ ,

$$\frac{1}{1+x} - \frac{1}{1+ax} \leq \frac{\sqrt{a}-1}{\sqrt{a}+1},$$

with equality only for  $x = 1/\sqrt{a}$ .

10. Let  $P(x) = x^n + \sum_{k=0}^{n-1} a_k x^k$ . Find

$$\lim_{x \rightarrow +\infty} ([P(x)]^{1/n} - x).$$

11. Let  $S$  and  $C$  be real-valued functions on  $\mathbf{R}$ , such that for all  $x, y$

$$S(x+y) = S(x)C(y) + C(x)S(y),$$

$$C(x+y) = C(x)C(y) - S(x)S(y),$$

$$S^2(x) + C^2(x) = 1,$$

and such that  $\lim_{x \rightarrow 0} S(x)/x = 1$ .

(a) Show that  $C(0) = 1$  and  $S(0) = 0$ , and that for all  $x$ ,  $S(-x) = -S(x)$  and  $C(-x) = C(x)$ .

(b) Show that for all  $x, y$

$$S(x) - S(y) = 2C\left(\frac{x+y}{2}\right)S\left(\frac{x-y}{2}\right),$$

and deduce that  $S$  is differentiable on  $\mathbf{R}$ , and  $S' = C$ .

(c) Find a corresponding formula for  $C(x) - C(y)$ , and use it to show that  $C$  is differentiable on  $\mathbf{R}$ , with  $C' = S$ .

(d) Show that  $S(x) = \sin x$  and  $C(x) = \cos x$ .

12. Suppose that  $f$  is defined in an open interval containing  $x$ , and that  $f'''(x)$  exists. Show that

$$f''(x) = \lim_{h \rightarrow 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}.$$

Give an example where this limit exists, but  $f''(x)$  does not. HINT: Use L'Hospital's rule.



13. Show that if  $f^{(n)} = 0$  in an interval  $I$ , and  $f(a) = f'(a) = \cdots = f^{(n-1)}(a) = 0$  for some  $a \in I$ , then  $f = 0$ .

14. Suppose that  $k \in \mathbf{R}$ ,  $k > 0$ , and that  $f$  is a real-valued function on an interval  $I$  such that  $f''(x)$  exists for all  $x \in I$ , and  $f''(x) + k^2 f(x) = 0$  for all  $x \in I$ . Show that there exist constants  $A$  and  $B$  such that  $f(x) = A \cos kx + B \sin kx$  for all  $x \in I$ .

15. Let  $f \in C^n(I)$ , where  $I$  is an interval, and let  $c \in I$ . Suppose that there exists a polynomial  $P$  of degree  $\leq n$  such that

$$|f(x) - P(x)| \leq C|x - c|^{n+1}$$

for some constant  $C$ . Show that  $P$  is the Taylor polynomial for  $f$  at  $c$ , i.e.,

$$P(x) = \sum_{k=0}^n \frac{f^{(k)}(c)}{k!} (x - c)^k.$$

16. Let  $f$  be a function of class  $C^2$  on  $(0, +\infty)$ , and let  $M_j = \sup |f^{(j)}(t)|$  for  $j = 0, 1, 2$ . Show that  $M_1^2 \leq 4M_0M_2$ . HINT: Use the second-order Taylor expansion about the point  $x$  to show that

$$|f'(x)| \leq hM_2 + \frac{M_0}{h}$$

for every  $h > 0$ , and then choose  $h$  to minimize the right-hand side.

17. Let  $p \in \mathbf{R}$ , and let  $f(x) = (1 + x)^p$  for  $-1 < x < +\infty$ . Show that the Taylor series  $\sum_{n=0}^{\infty} f^{(n)}(0)x^n/n!$  converges to  $f(x)$  in the interval  $-1 < x < 1$ .

18. Show that if  $f$  is analytic in an interval  $I$ , and not identically zero, then the zeros of  $f$  are *isolated*; that is, if  $f(c) = 0$  for some  $c \in I$ , there exists  $\delta > 0$  such that  $f(t) \neq 0$  for all  $t \in I$  satisfying  $0 < |t - c| < \delta$ .

## 4.9 Notes

4.1 The notation  $f'(x)$  was introduced by Lagrange; the related  $\dot{y} = \dot{f}(x)$  was Newton's notation, and (as mentioned above) the  $dy/dx$  notation is due to Leibniz. The notation  $Df$ , or  $D_x f(x)$ , was introduced by Arbogast in 1800, and has some real advantages. It is interesting that the idea behind the integral goes back to ancient times (Archimedes, for instance, calculated several nontrivial integrals), but the derivative did not really arise before the seventeenth century. Example 4.4 is due to van der Waerden (1930); the first continuous but nowhere differentiable function was exhibited by Weierstrass in 1861 (but not published before 1874). Bolzano is said to have found such an example as early as 1830.

- 4.2 The usual way in which the sine and cosine functions are introduced in calculus courses is through the geometric notions of angle and arc length; one establishes the crucial inequality  $\sin x/x < 1$  for  $x > 0$  by comparing the area of a sector of a circle with that of a triangle which it contains. The ideas of angle and arc length are still beyond our scope, so we had to work with infinite series definitions of sine and cosine.
- 4.3 Convex functions appear in very many contexts in mathematics. The notion of convex function is related to that of convex set. A subset  $K$  of the plane is called convex if with any two points which it contains it also contains the entire line segment joining them. A function  $f$  is convex if and only if the set  $\{(x, y) : y \geq f(x)\}$  of all points in the plane lying above the graph of  $f$  is a convex set.
- 4.4 The mean value theorem is due to Lagrange. The special case where  $f(a) = f(b) = 0$  is called Rolle's theorem; Rolle proved that between any two roots of a polynomial  $P(x)$  there lies a root of  $P'(x)$ . In fact, Rolle was strongly critical of the developing theory of calculus in his day, and it is ironic that his name is associated with such a crucial step in setting the calculus on a firm basis. The proof we gave of the mean value theorem is due to O. Bonnet, and was first published in 1868. Bonnet also proved mean value theorems for integrals, as well as derivatives. Theorem 4.25 is due to Darboux. Theorem 4.26 generalizes to differentiable mappings from  $\mathbf{R}^n$  to  $\mathbf{R}^n$ , as we shall see, and the proof becomes noticeably more difficult.
- 4.5 The general form of the mean value theorem, Theorem 4.29, is due to Cauchy. Theorem 4.30 is known as L'Hospital's rule. It first appeared in L'Hospital's book, published in 1696, the first textbook of the calculus. L'Hospital, whose name is often spelled L'Hôpital or l'Hôpital, was a French marquis who studied mathematics with Johann Bernoulli, and apparently acknowledged his debt to Bernoulli for all the ideas of which he wrote an exposition. This book achieved a wide readership; the next major calculus textbook was that of Maria Agnesi (1748), best remembered today for the plane curve known as the "witch of Agnesi." Not long after, Euler wrote the books which remained the standard for a long time.
- 4.6 Theorem 4.33 was first stated and proved by Lagrange in 1797. Taylor had derived the formula for the power series expansion of a function in 1715, by considering an expression in terms of finite differences and passing to the limit.
- 4.7 The results of this section remain valid for complex functions, that is, for functions  $\sum_{n=0}^{\infty} a_n(z - c)^n$  where  $a_n$ ,  $c$ , and  $z$  are all complex numbers. In fact, this is the natural setting for the study of power

series, and many striking results exist; but that is the subject of a different course. One can also take power series where the powers are of, for example, matrices. Thus, if  $A$  is an  $n \times n$  matrix, it is useful to form the matrix

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!},$$

which solves the system of differential equations  $\mathbf{x}'(t) = A\mathbf{x}$  with initial condition  $\mathbf{x}(0) = \mathbf{x}_0$  by  $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$ . Here is another example of a "power series." If  $f$  is a function analytic on  $\mathbf{R}$ , then we can write

$$f(x+h) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)h^n}{n!}.$$

Let us use the notation  $Df$  for  $f'$ , so  $D^n f = f^{(n)}$ ;  $D$  is a function whose domain and range is the set of functions analytic on  $\mathbf{R}$ . Another such function is the "translation operator"  $T_h$ , defined by  $(T_h f)(x) = f(x+h)$ . The Taylor expansion above then reads

$$(T_h f)(x) = \sum_{n=0}^{\infty} \frac{h^n D^n f(x)}{n!},$$

or on the operator level,  $T_h = e^{hD}$ . This discussion is merely formal here, but in fact good sense can be made of the idea that translation is the exponential of differentiation.

# 5

## The Riemann Integral

In this chapter we give an exposition of the definite integral of a real-valued function defined on a closed bounded interval. We assume familiarity with this concept from a previous study of calculus, but want to develop the theory in a more precise way than is typical for calculus courses, and also take a closer look at what kind of functions can be integrated. The integral to be defined and studied here is now widely known as the Riemann integral; in a later chapter we will study the more general Lebesgue integral.

### 5.1 Riemann Sums

**5.1 Definition.** Let  $[a, b]$  be a bounded closed interval in  $\mathbf{R}$ . A finite sequence  $(x_k)_{k=0}^n$  is called a partition of  $[a, b]$  if  $a = x_0 < x_1 < \cdots < x_n = b$ .

It is clear that the set of all partitions of  $[a, b]$  is in a one-one correspondence with the set of all finite subsets of  $(a, b)$ , and this correspondence induces a partial ordering among partitions.

**5.2 Definition.** Let  $\pi = (x_k)_{k=0}^n$  and  $\pi' = (y_j)_{j=0}^m$  be two partitions of the bounded closed interval  $[a, b]$ . We say that  $\pi'$  is a refinement of  $\pi$ , and write  $\pi \leq \pi'$ , if  $\{x_0, x_1, \dots, x_n\} \subset \{y_0, y_1, \dots, y_m\}$ .

**5.3 Definition.** If  $\pi = (x_k)_{k=0}^n$  is a partition of  $[a, b]$ , a selection associated to  $\pi$  is a finite sequence  $(\xi_k)_{k=1}^n$  such that  $x_{k-1} \leq \xi_k \leq x_k$  for  $k = 1, \dots, n$ .

If  $f$  is a real-valued function whose domain contains  $[a, b]$ ,  $\pi = (x_k)_{k=1}^n$  is a partition of  $[a, b]$ , and  $\sigma = (\xi_k)_{k=1}^n$  is a selection associated to  $\pi$ , we put

$$S(f, \pi, \sigma) = \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}),$$

and we call  $S(f, \pi, \sigma)$  a Riemann sum for the function  $f$ , associated to the partition  $\pi$  and selection  $\sigma$ .

The (Riemann) integral of  $f$  over  $[a, b]$  is the limit of Riemann sums, in the following sense:

**5.4 Definition.** Let  $f$  be a real-valued function whose domain contains  $[a, b]$ . We say that  $f$  is Riemann integrable over  $[a, b]$  if there exists a real number  $I$  with the following property: for any  $\epsilon > 0$ , there exists a partition  $\pi_0$  of  $[a, b]$  such that for every partition  $\pi \geq \pi_0$ , i.e., every  $\pi$  which is a refinement of  $\pi_0$ , and every selection  $\sigma$  associated to  $\pi$ , we have  $|S(f, \pi, \sigma) - I| < \epsilon$ . The number  $I$  is called the integral of  $f$  over  $[a, b]$ , and denoted by  $\int_a^b f$ , or  $\int_a^b f(x) dx$ .

Throughout this chapter, we will use the word integrable to mean Riemann integrable, and will often say  $\int_a^b f$  exists to mean  $f$  is Riemann integrable over  $[a, b]$ .

Since the number  $I$ , when it exists, is determined solely by the function  $f$  and the interval  $[a, b]$ , the first of the two notations,  $\int_a^b f$ , is clearly logically preferable to the second,  $\int_a^b f(x) dx$ . Here the letter  $x$  serves a purely ceremonial purpose, and may be replaced by any other letter; thus,  $\int_a^b f(t) dt$  or  $\int_a^b f(\omega) d\omega$  have the same meaning as  $\int_a^b f(x) dx$ . Nevertheless, this second notation has some practical advantages (besides being the notation most of us grew up with). For instance, if we were to insist on the first notation for integrals,  $\int_0^1 (x^2 + xy + 1) dx$  would require a long-winded circumlocution such as "for each  $y$ , define the function  $f_y$  by  $f_y(x) = x^2 + xy + 1$  for each  $x \in [0, 1]$ , and consider  $\int_0^1 f_y$ ."

**5.5 Proposition.** In order for  $f$  to be integrable over  $[a, b]$ , it is necessary that  $f$  be bounded on  $[a, b]$ .

**Proof.** If  $f$  is integrable, there exists a partition  $\pi = (x_k)_{k=1}^n$  of  $[a, b]$  such that  $|S(f, \pi, \sigma) - \int_a^b f| < 1$  for every selection  $\sigma$  associated to  $\pi$ . But if  $f$  were not bounded above on  $[a, b]$ , it would have no upper bound on some interval  $[x_{k-1}, x_k]$ ; varying the choice  $\xi_k$  while leaving  $\xi_j$  fixed for  $j \neq k$ , we would obtain selections  $\sigma$  with  $S(f, \pi, \sigma)$  arbitrarily large, a contradiction. A similar argument shows that  $f$  is bounded below. ■

**5.6 Example.** Not every bounded function is integrable. Consider the function of Example 3.4 in Chapter 3:

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

For any partition  $\pi = (x_k)_{k=0}^n$  of  $[0, 1]$ , we note that each interval  $[x_{k-1}, x_k]$  contains both rational and irrational numbers; thus for every partition  $\pi$  there exist selections  $\sigma_1$  and  $\sigma_2$  associated to  $\pi$  such that  $S(f, \pi, \sigma_1) = 1$  and  $S(f, \pi, \sigma_2) = 0$ . Thus  $\int_0^1 f$  does not exist. We recall that this function is discontinuous at every point.

**5.7 Definition.** Let  $f$  be a real-valued function whose domain contains  $[a, b]$ , let  $\pi = (x_k)_{k=0}^n$  be a partition of  $[a, b]$ . We define the upper sum of  $f$  for the partition  $\pi$  to be

$$\overline{S}(f, \pi) = \sup\{S(f, \pi, \sigma) : \sigma \text{ associated to } \pi\},$$

and the lower sum of  $f$  for the partition  $\pi$  to be

$$\underline{S}(f, \pi) = \inf\{S(f, \pi, \sigma) : \sigma \text{ associated to } \pi\}.$$

It is clear that  $-\infty \leq \underline{S}(f, \pi) \leq \overline{S}(f, \pi) \leq +\infty$ , and that when  $f$  is bounded, say  $m \leq f(t) \leq M$  for all  $t \in [a, b]$ , we have

$$m(b-a) \leq \underline{S}(f, \pi) \leq \overline{S}(f, \pi) \leq M(b-a).$$

We can refine this estimate as follows. Given the partition  $\pi = (x_k)_{k=0}^n$  of  $[a, b]$  and the function  $f$  on  $[a, b]$ , let  $m_k = \inf\{f(t) : x_{k-1} \leq t \leq x_k\}$  and  $M_k = \sup\{f(t) : x_{k-1} \leq t \leq x_k\}$ . It is obvious that we have

$$\overline{S}(f, \pi) = \sum_{k=1}^n M_k(x_k - x_{k-1}), \quad \underline{S}(f, \pi) = \sum_{k=1}^n m_k(x_k - x_{k-1}).$$

We next show that every lower sum is less than or equal to any upper sum.

**5.8 Lemma.** Let  $\pi$  and  $\pi'$  be partitions of  $[a, b]$ , and suppose that  $\pi'$  is a refinement of  $\pi$ . For any function  $f$  defined on  $[a, b]$ , we have

$$\underline{S}(f, \pi) \leq \underline{S}(f, \pi') \leq \overline{S}(f, \pi') \leq \overline{S}(f, \pi).$$

**Proof.** Suppose that  $\pi = (x_k)_{k=0}^n$  and  $\pi' = (y_j)_{j=0}^m$ . Let  $M_k = \sup\{f(t) : x_{k-1} \leq t \leq x_k\}$  for each  $k$ ,  $1 \leq k \leq n$ . For each  $j$ ,  $1 \leq j \leq m$ , there exists a unique  $k = k(j)$  such that  $x_{k-1} < y_j \leq x_k$ . Let  $F_k = \{j : k(j) = k\}$  for each  $k$ ,  $1 \leq k \leq n$ . We note that  $[x_{k-1}, x_k] = \bigcup_{j \in F_k} [y_{j-1}, y_j]$ , and hence

that  $\sum_{j \in F_k} (y_j - y_{j-1}) = x_k - x_{k-1}$ . Now if  $\sigma = (\xi_j)_{j=1}^m$  is any selection associated to the partition  $\pi'$ , we have

$$\begin{aligned} S(f, \pi', \sigma) &= \sum_{j=1}^m f(\xi_j)(y_j - y_{j-1}) = \sum_{k=1}^n \sum_{j \in F_k} f(\xi_j)(y_j - y_{j-1}) \\ &\leq \sum_{k=1}^n \sum_{j \in F_k} M_k(y_j - y_{j-1}) = \sum_{k=1}^n M_k \sum_{j \in F_k} (y_j - y_{j-1}) \\ &= \sum_{k=1}^n M_k(x_k - x_{k-1}) = \bar{S}(f, \pi), \end{aligned}$$

and it follows that  $\bar{S}(f, \pi') \leq \bar{S}(f, \pi)$ . The proof that  $\underline{S}(f, \pi') \geq \underline{S}(f, \pi)$  is entirely similar. ■

**5.9 Proposition.** *Let  $\pi_1$  and  $\pi_2$  be any partitions of  $[a, b]$ . If  $f$  is a real-valued function on  $[a, b]$ , then*

$$\underline{S}(f, \pi_1) \leq \bar{S}(f, \pi_2).$$

**Proof.** There exists a partition  $\pi_3$  which is at the same time a refinement of  $\pi_1$  and a refinement of  $\pi_2$ ; for instance, take the union of the finite sets associated to  $\pi_1$  and  $\pi_2$ , and list in increasing order. Then

$$\underline{S}(f, \pi_1) \leq \underline{S}(f, \pi_3) \leq \bar{S}(f, \pi_3) \leq \bar{S}(f, \pi_2)$$

by Lemma 5.8. ■

**5.10 Theorem.** *If  $f$  is a real-valued function on  $[a, b]$ , then  $\int_a^b f$  exists if and only if for every  $\epsilon > 0$  there exists a partition  $\pi$  of  $[a, b]$  such that  $\bar{S}(f, \pi) - \underline{S}(f, \pi) < \epsilon$ . If  $(\pi_n)$  is a sequence of partitions of  $[a, b]$  such that  $\bar{S}(f, \pi_n) - \underline{S}(f, \pi_n) \rightarrow 0$  as  $n \rightarrow \infty$ , and  $\sigma_n$  is a selection associated to  $\pi_n$  for each  $n$ , then  $\int_a^b f = \lim_{n \rightarrow \infty} S(f, \pi_n, \sigma_n)$ .*

**Proof.** The necessity of the condition is immediate from the definition of the integral. We prove the sufficiency. Let  $I = \inf_{\pi} \bar{S}(f, \pi)$ , where the infimum is taken over all partitions  $\pi$  of  $[a, b]$ . By the last proposition, we have

$$\underline{S}(f, \pi) \leq I \leq \bar{S}(f, \pi)$$

for every partition  $\pi$  of  $[a, b]$ . Given  $\epsilon > 0$ , there exists by hypothesis a partition  $\pi_0$  of  $[a, b]$  such that  $\bar{S}(f, \pi_0) - \underline{S}(f, \pi_0) < \epsilon$ , and by Lemma 5.8, this inequality remains true when  $\pi_0$  is replaced by any refinement  $\pi$  of  $\pi_0$ . Now for any refinement  $\pi$  of  $\pi_0$  and any selection  $\sigma$  associated to  $\pi$ , we have

$$\underline{S}(f, \pi) \leq S(f, \pi, \sigma) \leq \bar{S}(f, \pi)$$

as well as

$$\underline{S}(f, \pi) \leq I \leq \overline{S}(f, \pi),$$

so  $|I - S(f, \pi, \sigma)| \leq \overline{S}(f, \pi) - \underline{S}(f, \pi) < \epsilon$ . Thus  $f$  is Riemann integrable over  $[a, b]$ , and  $\int_a^b f = I$ . The second assertion also follows immediately. ■

## 5.2 Existence Results

**5.11 Definition.** If  $\pi = (x_k)_{k=0}^n$  is a partition of  $[a, b]$ , we let

$$\mu(\pi) = \max\{x_k - x_{k-1} : k = 1, \dots, n\};$$

$\mu(\pi)$  is called the mesh of the partition  $\pi$ .

**5.12 Theorem.** If  $f$  is a monotone function on  $[a, b]$ , then  $\int_a^b f$  exists. Furthermore, if  $(\pi_n)_{n=1}^\infty$  is any sequence of partitions of  $[a, b]$ , with  $\mu(\pi_n) \rightarrow 0$  as  $n \rightarrow \infty$ , and if  $(\sigma_n)$  is any sequence of associated selections, then  $S(f, \pi_n, \sigma_n) \rightarrow \int_a^b f(t) dt$ .

**Proof.** We assume  $f$  is increasing; an obvious modification of the argument works for the case when  $f$  is decreasing. Let  $\epsilon > 0$ . Choose a partition  $\pi = (x_k)_{k=0}^n$  such that  $[f(b) - f(a)]\mu(\pi) < \epsilon$ . [For instance, one can choose a positive integer  $n$  such that  $n > [f(b) - f(a) + 1](b - a)/\epsilon$ , and define the partition  $\pi = (x_k)_{k=0}^n$  by  $x_k = a + (k/n)(b - a)$ , so  $x_k - x_{k-1} = (b - a)/n$  for each  $k$ ,  $1 \leq k \leq n$ .] Then  $m_k = \inf\{f(t) : x_{k-1} \leq t \leq x_k\} = f(x_{k-1})$  and  $M_k = \sup\{f(t) : x_{k-1} \leq t \leq x_k\} = f(x_k)$ , so

$$\begin{aligned} \overline{S}(f, \pi) - \underline{S}(f, \pi) &= \sum_{k=1}^n [f(x_k) - f(x_{k-1})](x_k - x_{k-1}) \\ &\leq \mu(\pi) \sum_{k=1}^n [f(x_k) - f(x_{k-1})] \\ &= \mu(\pi)[f(b) - f(a)] < \epsilon. \end{aligned}$$

The theorem now follows from Theorem 5.10. ■

The next example shows how to compute the value of  $\int_a^b f$  in a special case; it shows that it can be convenient not to use the equally spaced partition suggested in the proof above.

**5.13 Example.** Let  $f(t) = t^p$ , for  $t \geq 0$ , where  $p \neq -1$ . Let us compute  $\int_1^b f(t) dt$  for any  $b > 1$ . Fix a positive integer  $n$ , let  $\delta = b^{1/n}$ , and let



$x_k = \delta^k$ , for  $k = 0, 1, \dots, n$ . Thus  $\pi = (x_k)_{k=0}^n$  is a partition of  $[1, b]$ . We have

$$\begin{aligned}\bar{S}(f, \pi) &= \sum_{k=1}^n f(x_k)(x_k - x_{k-1}) \\ &= \sum_{k=1}^n \delta^{kp}(\delta^k - \delta^{k-1}) \\ &= \frac{\delta - 1}{\delta} \sum_{k=1}^n \delta^{(p+1)k} \\ &= \frac{\delta - 1}{\delta} \frac{\delta^{n(p+1)} - 1}{\delta^{p+1} - 1} \delta^{p+1} \quad (\text{here we use } p \neq -1) \\ &= \frac{b^{p+1} - 1}{b^p + \delta^{p-1} + \dots + \delta + 1} \delta^p.\end{aligned}$$

Letting  $n \rightarrow \infty$ , we have  $\delta \rightarrow 1$ , so that these upper sums approach  $(b^{p+1} - 1)/(p + 1)$ .

Since we have seen that  $\int_1^b f(t) dt$  exists, and is given by the limit of any sequence of Riemann sums  $S(f, \pi_n, \sigma_n)$  such that  $\mu(\pi_n) \rightarrow 0$ , it follows that  $\int_1^b f(t) dt = (b^{p+1} - 1)/(p + 1)$  for any real  $p \neq -1$ .

**5.14 Theorem.** *If  $f$  is a continuous real-valued function on  $[a, b]$ , then  $\int_a^b f$  exists. Furthermore,  $\int_a^b f(t) dt = \lim_{n \rightarrow \infty} S(f, \pi_n, \sigma_n)$  for any sequence  $(\pi_n)$  of partitions of  $[a, b]$  such that  $\mu(\pi_n) \rightarrow 0$ , and any sequence of associated selections  $(\sigma_n)$ .*

**Proof.** Let  $\epsilon > 0$ . According to Theorem 3.18,  $f$  is uniformly continuous on  $[a, b]$ , so there exists  $\delta > 0$  such that  $|f(x) - f(y)| < \epsilon/(b - a)$  whenever  $|x - y| < \delta$ . Let  $\pi = (x_k)_{k=0}^n$  be any partition of  $[a, b]$  such that  $\mu(\pi) < \delta$ . Since  $|s - t| < \delta$  whenever  $x_{k-1} \leq s, t \leq x_k$ , we have  $|f(s) - f(t)| < \epsilon/(b - a)$  for any  $s, t \in [x_{k-1}, x_k]$ , and hence  $M_k - m_k \leq \epsilon/(b - a)$ , where  $M_k$  and  $m_k$  have their customary meanings. Thus

$$\begin{aligned}\bar{S}(f, \pi) - \underline{S}(f, \pi) &= \sum_{k=1}^n (M_k - m_k)(x_k - x_{k-1}) \\ &\leq \frac{\epsilon}{b - a} \sum_{k=1}^n (x_k - x_{k-1}) = \epsilon,\end{aligned}$$

so  $\int_a^b f$  exists by Theorem 5.10. ■

We recall that our example of a nonintegrable function was not continuous at any point. It turns out that a bounded function is integrable if and only if its set of discontinuities is sufficiently small, in a sense to be

made precise in a later chapter, when we study measure theory. For now, we content ourselves with the following result in this direction:

**5.15 Theorem.** *Let  $f$  be a bounded function on  $[a, b]$ , let  $D$  be the set of discontinuities of  $f$ , i.e., the set of all  $x \in [a, b]$  such that  $f$  is not continuous at  $x$ , and suppose that for every  $\epsilon > 0$  there exists a finite collection of disjoint intervals  $\{(a_j, b_j) : 1 \leq j \leq m\}$  such that  $D \subset \bigcup_{j=1}^m (a_j, b_j)$  and  $\sum_{j=1}^m (b_j - a_j) < \epsilon$ . Then  $f$  is integrable over  $[a, b]$ .*

**Proof.** The set  $C = [a, b] \setminus \bigcup_{j=1}^m (a_j, b_j)$  is easily seen to be the union of a finite number of closed intervals, on each of which  $f$  is continuous. It follows from Theorem 3.18 that there exists  $\delta > 0$  such that  $|f(s) - f(t)| < \epsilon$  whenever  $s$  and  $t$  belong to a closed interval contained in  $C$  and  $|s - t| < \delta$ . Choose a partition  $\pi = (x_k)_{k=0}^n$  of  $[a, b]$  with the following property: for each  $k$ , either  $[x_{k-1}, x_k] \subset C$  and  $x_k - x_{k-1} < \delta$ , or  $[x_{k-1}, x_k] \subset [a_j, b_j]$  for some  $j$ ,  $1 \leq j \leq m$ . Write  $k \in G$  if the first alternative holds, and  $k \in B$  if the second holds. As usual, we let  $M_k = \sup\{f(t) : x_{k-1} \leq t \leq x_k\}$  and  $m_k = \inf\{f(t) : x_{k-1} \leq t \leq x_k\}$ ; also  $M = \max_k M_k$  and  $m = \min_k m_k$ . Then we have  $M_k - m_k < \epsilon$  for  $k \in G$ , and

$$\sum_{k \in B} (x_k - x_{k-1}) \leq \sum_{j=1}^m (b_j - a_j) < \epsilon.$$

Hence

$$\begin{aligned} \bar{S}(f, \pi) - \underline{S}(f, \pi) &= \sum_{k=1}^n (M_k - m_k)(x_k - x_{k-1}) \\ &= \sum_{k \in G} (M_k - m_k)(x_k - x_{k-1}) \\ &\quad + \sum_{k \in B} (M_k - m_k)(x_k - x_{k-1}) \\ &\leq \sum_{k \in G} \epsilon (x_k - x_{k-1}) + \sum_{k \in B} (M - m)(x_k - x_{k-1}) \\ &< (b - a + M - m)\epsilon. \end{aligned}$$

Theorem 5.10 now shows that  $\int_a^b f$  exists. ■

**5.16 Corollary.** *If  $f$  is continuous at all but a finite set of points in  $[a, b]$ , then  $\int_a^b f$  exists.*

Thus, for instance,  $\int_0^b \sin(1/x) dx$  exists. Here we define the integrand to have the value 0 at 0, or any other value.

It is clear that if  $f$  and  $g$  are functions on  $[a, b]$  such that  $\{t \in [a, b] : f(t) \neq g(t)\}$  is finite, then  $\int_a^b f = \int_a^b g$  (if either integral exists, then so

does the other, and they are equal). The proof of the last theorem shows that if  $f$  and  $g$  differ only on a set  $D$  with the property that for any  $\epsilon > 0$  there exist open intervals  $(a_j, b_j)$ ,  $1 \leq j \leq m$ , with  $D \subset \bigcup_{j=1}^m (a_j, b_j)$  and  $\sum(b_j - a_j) < \epsilon$ , then  $\int_a^b f = \int_a^b g$ .

There exists an increasing function  $f$  which is discontinuous at every rational, as we saw in the exercises in Chapter 3. It is not hard to see that the rationals in  $[0, 1]$  cannot be contained in a finite union of intervals of total length less than one, so the last theorem does not apply, though Theorem 5.12 tells us that  $\int_0^1 f$  does exist. Thus the condition of the last theorem is sufficient, but not necessary.

**5.17 Theorem.** *If  $f : [a, b] \rightarrow \mathbf{R}$ , then  $\int_a^b f$  exists if and only if for every  $\epsilon > 0$ , there exist continuous functions  $g$  and  $h$  on  $[a, b]$  such that  $g(t) \leq f(t) \leq h(t)$  for all  $t \in [a, b]$ , and  $\int_a^b h - \int_a^b g < \epsilon$ .*

**Proof.** If there exist such  $g$  and  $h$ , choose a partition  $\pi$  of  $[a, b]$  such that  $\bar{S}(h, \pi) < \int_a^b h + \epsilon$  and  $\underline{S}(g, \pi) > \int_a^b g - \epsilon$ . Then  $\bar{S}(h, \pi) - \underline{S}(g, \pi) < 2\epsilon$ , and since

$$\underline{S}(g, \pi) \leq \underline{S}(f, \pi) \leq \bar{S}(f, \pi) \leq \bar{S}(h, \pi)$$

we get  $\bar{S}(f, \pi) - \underline{S}(f, \pi) < 2\epsilon$ , so  $f$  is Riemann integrable by Proposition 5.10.

Now suppose  $f$  is Riemann integrable. Let  $\pi = (x_k)_{k=0}^n$  be a partition of  $[a, b]$  such that  $\bar{S}(f, \pi) < \int_a^b f + \epsilon/2$ . As usual, let  $M_k = \sup\{f(t) : x_{k-1} \leq t \leq x_k\}$  for  $k = 1, \dots, n$ . We will construct a piecewise linear continuous function  $h$  with  $h(t) \geq M_k$  on each  $[x_{k-1}, x_k]$ .

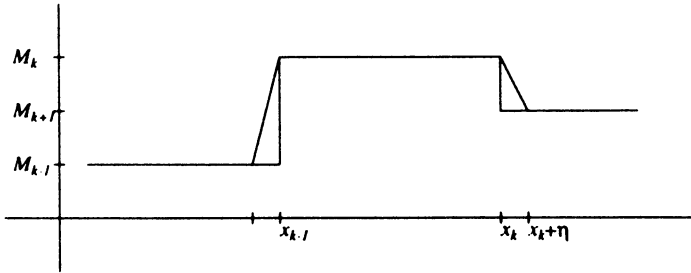


Figure 5.1. The construction in Theorem 5.17.

Let  $\xi_k = (x_{k-1} + x_k)/2$  be the midpoint of  $[x_{k-1}, x_k]$ , and  $0 < \eta < \mu(\pi)/2$ . We define the function  $h_\eta$  as follows: if  $a \leq t \leq \xi_1$ , set  $h_\eta(t) = M_1$ , and if  $\xi_n \leq t \leq b$ , set  $h_\eta(t) = M_n$ . The definition for  $h_\eta$  on each interval  $[x_k, x_{k+1}]$  will depend on which of  $M_k$  and  $M_{k+1}$  is larger. If  $M_k \leq M_{k+1}$ , let

$$h_\eta(t) = \begin{cases} M_k, & \text{for } \xi_k \leq t \leq x_k - \eta; \\ M_{k+1}, & \text{for } x_k \leq t \leq \xi_{k+1}; \\ (M_k(x_k - t) + M_{k+1}(\eta - x_k + t))/\eta & \text{for } x_k - \eta \leq t \leq x_k. \end{cases}$$

If  $M_k > M_{k+1}$ , we modify this definition in the obvious way: let

$$h_\eta(t) = \begin{cases} M_k, & \text{for } \xi_k \leq t \leq x_k; \\ M_{k+1}, & \text{for } x_k + \eta \leq t \leq \xi_{k+1}; \\ (M_{k+1}(t - x_k) + M_k(\eta - t + x_k))/\eta & \text{for } x_k + \eta \leq t \leq x_{k+1}. \end{cases}$$

Then  $h_\eta$  is continuous on  $[a, b]$ ,  $h_\eta(t) \geq f(t)$  for all  $t \in [a, b]$ , and

$$\int_a^b h_\eta(t) dt \leq \bar{S}(f, \pi) + \sum_{k=1}^n \eta |M_{k+1} - M_k|,$$

so taking  $\eta$  sufficiently small, we get  $\int_a^b h_\eta(t) dt < \int_a^b f(t) dt + \epsilon$ . Applying this argument to the function  $-f$ , we obtain a continuous function  $k$  with  $k \geq -f$  and  $\int_a^b k < -\int_a^b f + \epsilon$ , and taking  $g = -k$  we have a continuous  $g$  with  $g \leq f$  and  $\int_a^b g > \int_a^b f - \epsilon$ . ■

The method of dividing the elements of a partition into a “good set” and a “bad set” which we used in proving Theorem 5.15 can be used in other situations.

**5.18 Theorem.** Let  $f$  be integrable over  $[a, b]$ , with  $m \leq f(t) \leq M$  for all  $t \in [a, b]$ . If  $\phi$  is continuous on  $[m, M]$ , then  $g = \phi \circ f$  is integrable over  $[a, b]$ .

**Proof.** Let  $\epsilon > 0$ . Since  $\phi$  is continuous on  $[m, M]$ , there exists (Theorem 3.18)  $\delta > 0$  such that  $|\phi(s) - \phi(t)| < \epsilon$  for all  $s, t \in [m, M]$  with  $|s - t| < \delta$ . Since  $f$  is integrable, there exists a partition  $\pi = (x_k)_{k=0}^n$  of  $[a, b]$  such that  $\bar{S}(f, \pi) - \underline{S}(f, \pi) < \epsilon\delta$ . Let  $M_k = \sup\{f(t) : x_{k-1} \leq t \leq x_k\}$ ,  $M'_k = \sup\{g(t) : x_{k-1} \leq t \leq x_k\}$ , and let  $m_k, m'_k$  be the corresponding infima. Let  $G = \{k : M_k - m_k < \delta\}$  and  $B = \{k : M_k - m_k \geq \delta\}$ . Let  $C = \max_{m \leq t \leq M} \phi(t) - \min_{m \leq t \leq M} \phi(t)$ . Then  $|g(s) - g(t)| < \epsilon$  for all  $s, t \in [x_{k-1}, x_k]$  if  $k \in G$ , so  $M'_k - m'_k < \epsilon$  for all  $k \in G$ , and  $M'_k - m'_k \leq C$  for every  $k$ . Now

$$\epsilon\delta > \sum_{k=1}^n (M_k - m_k)(x_k - x_{k-1}) \geq \sum_{k \in B} \delta(x_k - x_{k-1});$$

it follows that  $\sum_{k \in B} (x_k - x_{k-1}) < \epsilon$ . Hence

$$\begin{aligned} \bar{S}(g, \pi) - \underline{S}(g, \pi) &= \sum_{k \in G} (M'_k - m'_k)(x_k - x_{k-1}) \\ &\quad + \sum_{k \in B} (M'_k - m'_k)(x_k - x_{k-1}) \\ &\leq \epsilon \sum_{k \in G} (x_k - x_{k-1}) + C \sum_{k \in B} (x_k - x_{k-1}) \\ &\leq \epsilon(b - a) + C\epsilon = (C + b - a)\epsilon. \end{aligned}$$

Since  $\epsilon > 0$  was arbitrary, it now follows from Theorem 5.10 that  $g$  is Riemann integrable. ■

We list a few of the most useful special cases of this theorem.

**5.19 Corollary.** *If  $f$  is integrable over  $[a, b]$ , then  $f^+$ ,  $f^-$ ,  $|f|$ , and  $f^2$  are also integrable over  $[a, b]$ .*

## 5.3 Properties of the Integral

**5.20 Proposition.** *If  $f$  and  $g$  are integrable over  $[a, b]$ , and  $A$  and  $B$  are real numbers, then  $Af + Bg$  is integrable over  $[a, b]$ , and*

$$\int_a^b (Af + Bg) = A \int_a^b f + B \int_a^b g.$$

**Proof.** If  $\pi$  is any partition of  $[a, b]$ , and  $\sigma$  any selection associated to  $\pi$ , then  $S(Af + Bg, \pi, \sigma) = AS(f, \pi, \sigma) + BS(f, \pi, \sigma)$ , and the proposition follows. ■

**5.21 Proposition.** *If  $f$  is integrable over  $[a, b]$ , and  $m \leq f(t) \leq M$  for all  $t \in [a, b]$ , then  $m(b-a) \leq \int_a^b f \leq M(b-a)$ . In particular, if  $|f(t)| \leq M$  for all  $t \in [a, b]$ , then  $|\int_a^b f(t) dt| \leq M(b-a)$ .*

**Proof.** We already observed that  $m(b-a) \leq S(f, \pi, \sigma) \leq M(b-a)$  for every partition  $\pi$  and associated selection  $\sigma$ , which implies the proposition. ■

**5.22 Corollary.** *If  $f$  and  $g$  are integrable over  $[a, b]$ , and  $f(t) \leq g(t)$  for all  $t \in [a, b]$ , then  $\int_a^b f(t) dt \leq \int_a^b g(t) dt$ . If  $f$  is integrable over  $[a, b]$ , then*

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt.$$

**Proof.** Since  $g(t) - f(t) \geq 0$ , Proposition 5.21 shows that

$$0 \leq \int_a^b (g(t) - f(t)) dt = \int_a^b g(t) dt - \int_a^b f(t) dt,$$

which gives the first statement. Now  $\pm f(t) \leq |f(t)|$ , so the first statement of the theorem implies the second statement. ■

Another consequence of the basic estimate in Corollary 5.22 is the following, known as the mean value theorem for integrals:

**5.23 Theorem.** Let  $p$  be a nonnegative function integrable over  $[a, b]$ . If  $f$  is continuous on  $[a, b]$ , then there exists  $\xi \in [a, b]$  such that

$$\int_a^b f(t)p(t) dt = f(\xi) \int_a^b p(t) dt.$$

**Proof.** Let  $m = \min\{f(t) : a \leq t \leq b\}$  and  $M = \max\{f(t) : a \leq t \leq b\}$ . (The existence of  $M$  and  $m$  comes from Theorem 3.14.) Since  $p(t) \geq 0$  for all  $t \in [a, b]$ , we have  $mp(t) \leq f(t)p(t) \leq Mp(t)$  for all  $t \in [a, b]$ , so

$$m \int_a^b p(t) dt \leq \int_a^b f(t)p(t) dt \leq M \int_a^b p(t) dt$$

by Corollary 5.22. In particular, if  $\int_a^b p(t) dt = 0$ , then  $\int_a^b f(t)p(t) dt = 0$  also, and we can choose any  $\xi$  for the desired equation. When  $\int_a^b p(t) dt > 0$ , the number

$$y = \frac{\int_a^b f(t)p(t) dt}{\int_a^b p(t) dt}$$

satisfies  $m \leq y \leq M$ , and hence by the intermediate value theorem (Theorem 3.15) there exists  $\xi$  with  $f(\xi) = y$ . ■

**5.24 Theorem.** If  $f$  and  $g$  are integrable over  $[a, b]$ , then so is their product  $fg$ . Furthermore,

$$\left(\int_a^b fg\right)^2 \leq \left(\int_a^b f^2\right)\left(\int_a^b g^2\right). \quad (5.1)$$

**Proof.** By Proposition 5.20,  $f + g$  and  $f - g$  are integrable over  $[a, b]$ , and hence by Theorem 5.18 so are  $(f + g)^2$  and  $(f - g)^2$ . Since  $fg = (1/4)[(f + g)^2 - (f - g)^2]$ , another application of Theorem 5.20 tells us that  $fg$  is integrable over  $[a, b]$ .

Similarly,  $tf \pm g/t$  is integrable for any positive number  $t$ , and hence so is  $(tf \pm g/t)^2$ . From Propositions 5.21 and 5.20 we see that

$$0 \leq \int_a^b \left(tf \pm \frac{g}{t}\right)^2 = t^2 \int_a^b f^2 + \frac{1}{t^2} \int_a^b g^2 \pm 2 \int_a^b fg,$$

from which we have

$$\left|\int_a^b fg\right| \leq \frac{1}{2} \left(t^2 \int_a^b f^2 + \frac{1}{t^2} \int_a^b g^2\right)$$

for every  $t > 0$ . If  $\int_a^b f^2 = 0$ , it follows (letting  $t \rightarrow \infty$ ) that  $\int_a^b fg = 0$ , so the inequality holds. When  $\int_a^b f^2 > 0$ , we can choose  $t$  to make the two

terms equal, i.e., such that  $t^4 = \left(\int_a^b f^2\right)^{-1} \int_a^b g^2$ . Then

$$\left(\int_a^b fg\right)^2 \leq \left(t^2 \int_a^b f^2\right)^2 = \left(t^2 \int_a^b f^2\right) \left(\frac{1}{t^2} \int_a^b g^2\right) = \int_a^b f^2 \int_a^b g^2,$$

establishing the inequality. ■

The inequality (5.1) is known as the Schwarz, or the Bunyakovsky-Schwarz, inequality. It is of great utility in a variety of situations in analysis.

**5.25 Theorem.** Let  $a < c < b$ , and let  $f$  be a function defined on  $[a, b]$ . The integral  $\int_a^b f$  exists if and only if  $\int_a^c f$  and  $\int_c^b f$  both exist, and in this case  $\int_a^b f = \int_a^c f + \int_c^b f$ .

**Proof.** Let  $\pi = (x_k)_{k=0}^N$  be a partition of  $[a, b]$  which includes the point  $c$ , say  $c = x_n$ , for some  $0 < n < N$ . Then  $(x_k)_{k=0}^n$  is a partition of  $[a, c]$ , and  $(x_{n+k})_{k=0}^{N-n}$  is a partition of  $[c, b]$ . As usual, we put  $M_k = \sup\{f(t) : x_{k-1} \leq t \leq x_k\}$  and  $m_k = \inf\{f(t) : x_{k-1} \leq t \leq x_k\}$ . Let  $f_1$  and  $f_2$  denote the restrictions of  $f$  to  $[a, c]$  and  $[c, b]$ , respectively. Then

$$\begin{aligned} \overline{S}(f, \pi) - \underline{S}(f, \pi) &= \sum_{k=1}^n (M_k - m_k)(x_k - x_{k-1}) \\ &\quad + \sum_{j=1}^{N-n} (M_{n+j} - m_{n+j})(x_{n+j} - x_{n+j-1}) \\ &= \overline{S}(f_1, \pi) - \underline{S}(f_1, \pi) + \overline{S}(f_2, \pi) - \underline{S}(f_2, \pi). \end{aligned}$$

Now if  $\int_a^b f$  exists, and  $\epsilon > 0$ , there exists a partition  $\pi$  of  $[a, b]$  such that  $\overline{S}(f, \pi) - \underline{S}(f, \pi) < \epsilon$ ; there is no loss of generality in assuming that  $c$  is an element of  $\pi$ , in view of Lemma 5.8. The above shows that  $\overline{S}(f_1, \pi) - \underline{S}(f_1, \pi) < \epsilon$  and  $\overline{S}(f_2, \pi) - \underline{S}(f_2, \pi) < \epsilon$ . Thus  $\int_a^c f$  and  $\int_c^b f$  exist. Conversely, if these two integrals exist, then we combine suitable partitions of  $[a, c]$  and  $[c, b]$  to get a partition of  $[a, b]$ , and use the equation above to show  $\int_a^b f$  exists, and equals  $\int_a^c f + \int_c^b f$ . ■

**5.26 Definition.** Let  $f$  be integrable over the interval  $[a, b]$ , where  $a < b$ . We define  $\int_b^a f = -\int_a^b f$ , and we define  $\int_a^a f = 0$ .

With this convention, we can now expand the scope of the last theorem as follows:

**5.27 Corollary.** For any  $a, b, c$ ,

$$\int_a^b f = \int_a^c f + \int_c^b f,$$

in the sense that if any two of these integrals exist, then so does the third, and the equation holds.

**Proof.** We may reformulate the assertion of the theorem in a more symmetric manner as

$$\int_a^b f + \int_b^c f + \int_c^a f = 0,$$

in view of the definition above. In this form, we know it to be true when  $a < b < c$ , and the left-hand side is unchanged by a cyclic permutation of  $\{a, b, c\}$ , so it is true for the cases  $b < c < a$  and  $c < a < b$ . We check that the interchange of  $a$  and  $b$  changes the expression into its negative, hence again leaves it zero. This remains true for either of the other two interchanges, since they can be achieved by a cyclic permutation followed by the interchange of  $a$  and  $b$ . Thus all six permutations of  $\{a, b, c\}$  leave the left-hand side equal to 0. Finally, if  $a = b$  or  $b = c$  or  $a = c$ , the result is obvious. ■

The properties of the integral  $\int_a^b f$  we have obtained for the case  $a < b$  remain true in general, with the exception of the inequalities obtained in Proposition 5.21 and its corollary, which must be reversed if  $a > b$ . We still have the basic inequality: if  $|f(t)| \leq M$  for all  $t$  in an interval  $I$ , then  $|\int_a^b f(t) dt| \leq M|b - a|$  for all  $a, b \in I$ .

## 5.4 Fundamental Theorems of Calculus

Each of the next two theorems is often referred to as the fundamental theorem of calculus.

**5.28 Definition.** Let  $f : I \rightarrow \mathbf{R}$ . We say that  $F : I \rightarrow \mathbf{R}$  is a primitive of  $f$  if  $F'(x) = f(x)$  for every  $x \in I$ .

A primitive of  $f$  is also called an *antiderivative* of  $f$ , or an *indefinite integral* of  $f$ .

**5.29 Theorem.** Let  $f$  be integrable over the interval  $[a, b]$ . If  $F$  is a primitive of  $f$ , then  $\int_a^b f(t) dt = F(b) - F(a)$ .



**Proof.** Since  $F'(t) = f(t)$  for all  $t \in [a, b]$ , it follows that for any partition  $\pi = (x_k)_{k=0}^n$  of  $[a, b]$ , we have  $F(x_k) - F(x_{k-1}) = f(\xi_k)(x_k - x_{k-1})$  for some  $\xi_k \in (x_{k-1}, x_k)$ , by the mean value theorem (Theorem 4.22). Then  $\sigma = (\xi_k)$  is a selection associated to  $\pi$ , and we have

$$\begin{aligned} S(f, \pi, \sigma) &= \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}) \\ &= \sum_{k=1}^n (F(x_k) - F(x_{k-1})) = F(b) - F(a). \end{aligned}$$

If  $f$  is integrable over  $[a, b]$ , it follows that  $\int_a^b f(t) dt = F(b) - F(a)$ . ■

**5.30 Theorem.** Let  $f : I \rightarrow \mathbf{R}$ , where  $I$  is an interval, and suppose that  $f$  is integrable over any closed bounded interval contained in  $I$ . Let  $a \in I$ . If  $F$  is defined by

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

for each  $x \in I$ , then  $F$  is continuous on  $I$ . If  $f$  is continuous at  $x \in I$ , then  $F'(x) = f(x)$ .

**Proof.** Let  $x \in I$ . Let  $J$  be a closed interval contained in  $I$  which is a neighborhood of  $x$  relative to  $I$ . Let  $M = \sup\{|f(t)| : t \in J\}$ . Then for any  $h$  with  $x + h \in J$  we have

$$F(x + h) - F(x) = \int_a^{x+h} f(t) dt - \int_a^x f(t) dt = \int_x^{x+h} f(t) dt,$$

in view of Corollary 5.27, and hence  $|F(x + h) - F(x)| \leq M|h|$ . This shows that  $F$  is continuous at  $x$ . Now suppose  $f$  is continuous at  $x$ . Let  $\epsilon > 0$ . There exists  $\delta > 0$  such that  $|f(t) - f(x)| < \epsilon$  for all  $t \in I$  with  $|t - x| < \delta$ . Then for  $h$  with  $x + h \in I$  and  $|h| < \delta$ ,

$$\begin{aligned} \left| \frac{F(x + h) - F(x)}{h} - f(x) \right| &= \left| \frac{1}{h} \int_x^{x+h} f(t) dt - f(x) \right| \\ &= \left| \frac{1}{h} \int_x^{x+h} [f(t) - f(x)] dt \right| \leq \frac{|h|}{|h|} \epsilon = \epsilon. \end{aligned}$$

Thus  $F'(x) = f(x)$ . ■

**5.31 Corollary.** If  $f$  is continuous on the interval  $I$ , and  $a \in I$ , then  $F$ , defined by  $F(x) = \int_a^x f(t) dt$ , is a primitive of  $f$  on  $I$ .

We can deduce Theorem 5.29 from Theorem 5.30, if we assume that  $f$  is continuous, as follows.

Define  $G(x) = \int_a^x f(t) dt$  for each  $x \in I$ . By Theorem 5.30, we have  $G'(x) = f(x) = F'(x)$  for every  $x \in I$ , so  $(F - G)'(x) = 0$  for every  $x \in I$ , so  $F - G$  is a constant function on  $I$ , i.e.,  $F(x) - G(x) = F(a) - G(a) = F(a)$  for every  $x \in I$ . In particular,  $\int_a^b f(t) dt = G(b) = F(b) - F(a)$ . ■

The next result, known as *integration by parts*, can be very useful in dealing with integrals.

**5.32 Theorem.** Let  $f$  and  $g$  be integrable over  $[a, b]$ . If  $F$  and  $G$  are primitives of  $f$  and  $g$ , respectively, then

$$\int_a^b F(t)g(t) dt = F(b)G(b) - F(a)G(a) - \int_a^b f(t)G(t) dt.$$

**Proof.** Since  $(FG)' = F'G + FG' = fG + Fg$ , Theorem 5.29 tells us that

$$\int_a^b [f(t)G(t) + F(t)g(t)] dt = F(b)G(b) - F(a)G(a),$$

and the result follows. ■

Another important tool in dealing with integrals is the change of variables.

**5.33 Theorem.** Let  $\phi$  be of class  $C^1$  on the interval  $[\alpha, \beta]$ , with  $a = \phi(\alpha)$  and  $b = \phi(\beta)$ . If  $f$  is continuous on  $\phi([\alpha, \beta])$  and  $g = f \circ \phi$ , then

$$\int_a^b f(t) dt = \int_\alpha^\beta g(u)\phi'(u) du.$$

**Proof.** Note that  $g$  is continuous on  $[\alpha, \beta]$ . Let  $F(x) = \int_a^x f(t) dt$  for  $x \in [a, b]$ , so  $F$  is differentiable on  $[a, b]$ , and  $F'(x) = f(x)$  for all  $x \in [a, b]$ . Then  $G = F \circ \phi$  is differentiable in  $[\alpha, \beta]$ , and  $G'(u) = F'(\phi(u))\phi'(u) = g(u)\phi'(u)$  for all  $u \in [\alpha, \beta]$ . Hence

$$\int_\alpha^\beta g(u)\phi'(u) du = G(\beta) - G(\alpha) = F(b) - F(a) = \int_a^b f(t) dt,$$

as desired. ■

We observe that in this theorem, it is not necessary to assume that  $\phi([\alpha, \beta]) \subset [a, b]$ .

## 5.5 Integrating Sequences and Series

**5.34 Theorem.** Suppose that  $f_n$  is integrable over  $[a, b]$  for each  $n$ , and that  $f_n \rightarrow f$  uniformly on  $[a, b]$ . If  $f$  is integrable over  $[a, b]$ , then

$$\lim_{n \rightarrow \infty} \int_a^b f_n(t) dt = \int_a^b f(t) dt.$$

**Proof.** Given  $\epsilon > 0$  there exists  $n_0$  such that  $|f_n(t) - f(t)| < \epsilon/(b-a)$  for all  $n \geq n_0$ ; then by Proposition 5.21 we have

$$\left| \int_a^b f_n(t) dt - \int_a^b f(t) dt \right| = \left| \int_a^b (f_n(t) - f(t)) dt \right| \leq \epsilon$$

for every  $n \geq n_0$ . ■

**5.35 Corollary.** If  $f_n$  is integrable over  $[a, b]$  for each  $n$ , and  $f = \sum_{n=1}^{\infty} f_n$ , where the series converges uniformly on  $[a, b]$ , then, assuming  $f$  is integrable over  $[a, b]$ ,

$$\int_a^b f = \sum_{n=1}^{\infty} \int_a^b f_n.$$

In fact, the assumption that  $f$  is integrable in the last theorem and corollary is unnecessary (see the exercises at the end of this chapter), but since this is rarely an issue in applications, we do not press the point.

**5.36 Example.** Let  $f_n(x) = x/n(x+n)$  for  $0 \leq x \leq 1$ ,  $n \in \mathbf{N}$ . Then  $0 \leq f_n(x) \leq 1/n(n+1)$  for every  $x \in [0, 1]$ , and as we saw in Chapter 2,  $\sum_{n=1}^{\infty} 1/n(n+1) = 1$ , so by the Weierstrass  $M$ -test, the series  $\sum_{n=1}^{\infty} f_n$  converges uniformly on  $[0, 1]$  to some continuous function  $f$ , where  $0 \leq f(x) \leq 1$  for all  $x \in [0, 1]$ . By Corollary 5.35, we have

$$\sum_{n=1}^{\infty} \int_0^1 f_n(x) dx = \int_0^1 f(x) dx = \gamma,$$

where evidently  $0 < \gamma < 1$ . But

$$\int_0^1 \frac{x}{n(x+n)} dx = \int_0^1 \left( \frac{1}{n} - \frac{1}{x+n} \right) dx = \frac{1}{n} - \log \frac{n+1}{n},$$

so

$$\sum_{n=1}^N \int_0^1 f_n(x) dx = \sum_{n=1}^N \frac{1}{n} - \log(N+1) \rightarrow \gamma$$

as  $N \rightarrow \infty$ . Since  $\log(N+1) - \log N \rightarrow 0$  as  $N \rightarrow \infty$ , we can also say

$$\sum_{n=1}^N \frac{1}{n} - \log N \rightarrow \gamma$$

as  $N \rightarrow \infty$ . The number  $\gamma$  is known as Euler's constant.

Using the fundamental theorem of calculus, we can obtain results about differentiation from results concerning integration. Here is an example, formulated for series; it is simple to deduce the companion theorem for sequences.

**5.37 Theorem.** *Let  $(f_n)$  be a sequence of functions of class  $C^1$  on  $[a, b]$ , such that  $\sum_{n=1}^{\infty} f_n(c)$  converges for some  $c \in [a, b]$ , and suppose that  $\sum f'_n$  converges uniformly on  $[a, b]$ , to some function  $g$ . Then  $\sum_{n=1}^{\infty} f_n$  converges to a function  $f$  of class  $C^1$  on  $[a, b]$ , and  $f' = g$ .*

**Proof.** Since each  $f'_n$  is continuous, we know that  $g$  is continuous by Theorem 3.24. By Corollary 5.35, and using Theorem 5.29, we have

$$\int_c^x g(t) dt = \sum_{n=1}^{\infty} \int_c^x f'_n(t) dt = \sum_{n=1}^{\infty} [f_n(x) - f_n(c)]$$

and since  $\sum f_n(c)$  converges, it follows that  $\sum f_n(x)$  converges for each  $x \in [a, b]$ , say to  $f(x)$ . But we then have

$$f(x) - f(c) = \int_c^x g(t) dt$$

for each  $x \in [a, b]$ , and hence by Theorem 5.30 we obtain that  $f'(x) = g(x)$  for every  $x \in [a, b]$ . ■

## 5.6 Improper Integrals

In this section we consider certain integrals of the form  $\int_a^{\infty} f(x) dx$  or  $\int_a^b f(x) dx$  where  $f$  is not bounded on  $[a, b]$ ; these are no longer defined as Riemann integrals, and involve an extra limiting process.

**5.38 Definition.** *Let  $f : [a, +\infty) \rightarrow \mathbf{R}$  be integrable over  $[a, b]$  for every  $b > a$ . We say that the improper integral  $\int_a^{\infty} f(t) dt$  converges if*

$$L = \lim_{b \rightarrow +\infty} \int_a^b f(t) dt$$

*exists, and we write  $\int_a^{\infty} f(t) dt = L$  in this case.*

**5.39 Definition.** *Let  $f : (a, b) \rightarrow \mathbf{R}$  be integrable over  $[c, b]$  for every  $c \in (a, b)$ . If  $L = \lim_{c \rightarrow a+} \int_c^b f(t) dt$  exists, we say that  $\int_a^b f(t) dt$  converges, and equals  $L$ .*

**5.40 Example.** The integral  $\int_0^1 x^p dx$  exists for  $p \geq 0$ , since  $x \mapsto x^p$  is continuous on  $[0, 1]$ , but fails to exist for  $p < 0$ , since the function is unbounded. However, for  $p \neq -1$ ,

$$\lim_{\delta \rightarrow 0} \int_{\delta}^1 x^p dx = \lim_{\delta \rightarrow 0} \frac{1 - \delta^{p+1}}{p+1} = \begin{cases} 1/(p+1) & \text{if } p > -1, \\ +\infty & \text{if } p < -1, \end{cases}$$

so the improper integral  $\int_0^1 x^{-p} dx$  converges for  $p < 1$ , and diverges for  $p > 1$ . It is easy to check that it also diverges for  $p = 1$ . Similarly, we calculate that  $\int_1^{\infty} x^{-p} dx$  converges for  $p > 1$  and diverges for  $p \leq 1$ .

In this section we will only discuss improper integrals of the first kind. There are analogous results for integrals of the second kind. By a change of variable, each integral of the second kind can be transformed into one of the first kind. There are of course complicated integrals of mixed type, i.e., which involve both sorts of difficulties. An important example is the gamma function integral

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

When  $x \geq 1$ , this is simply an integral of the first kind defined above, but when  $x < 0$ , the integrand is unbounded near 0, so  $\int_0^R t^{x-1} e^{-t} dt$  is itself an improper integral, of the second kind. Such situations are generally easy to deal with by writing the integral in question as the sum of two integrals each of a pure type, in this case as

$$\Gamma(x) = \int_0^1 t^{x-1} e^{-t} dt + \int_1^{\infty} t^{x-1} e^{-t} dt.$$

The theory of improper integrals of the first kind has many analogies to the theory of infinite series. The next proposition follows at once from the definition. The analogue for integrals of the second kind is obvious.

**5.41 Proposition.** *If  $f$  and  $g$  are functions on  $[a, \infty)$  which are integrable over every  $[a, b]$ , and the integrals  $\int_a^{\infty} f(t) dt$  and  $\int_a^{\infty} g(t) dt$  converge, then for any  $A, B \in \mathbf{R}$  the integral  $\int_a^{\infty} (Af(t) + Bg(t)) dt$  converges, with value  $A \int_a^{\infty} f(t) dt + B \int_a^{\infty} g(t) dt$ .*

**5.42 Proposition.** *Let  $f$  be integrable over  $[a, b]$  for every  $b > a$ , and suppose that  $f(t) \geq 0$  for all  $t \geq a$ . Then either  $\int_a^{\infty} f(t) dt$  converges, or  $\int_a^x f(t) dt \rightarrow +\infty$  as  $x \rightarrow +\infty$ , in which case we write  $\int_a^{\infty} f(t) dt = +\infty$ .*

**Proof.** The function  $F$  defined on  $[a, +\infty)$  by  $F(x) = \int_a^x f(t) dt$  is increasing, since  $f(t) \geq 0$  for all  $t$ , so  $\lim_{x \rightarrow \infty} F(x) = \sup_{t \geq a} F(t)$  exists, either as a real number or  $+\infty$ . ■

**5.43 Corollary.** Let  $f$  and  $g$  be functions on  $[a, \infty)$  which are integrable over every interval  $[a, b]$ , and suppose that  $|g(t)| \leq f(t)$  for all  $t \geq a$  and that  $\int_a^\infty f(t) dt$  converges. Then  $\int_a^\infty g(t) dt$  converges.

**Proof.** We have  $0 \leq g^+(t) \leq |g(t)| \leq f(t)$  for all  $t$ , so  $\int_a^R g^+(t) dt \leq \int_a^R |g(t)| dt \leq \int_a^\infty f(t) dt$  for all  $R$ , so  $\int_a^\infty g^+(t) dt$  and  $\int_a^\infty |g(t)| dt$  converge. Since  $g = 2g^+ - |g|$ , it follows that  $\int_a^\infty g(t) dt$  converges. ■

Again, the analogue for integrals of the second kind is obvious.

For example, since  $t^{x-1}e^{-t} \leq t^{x-1}$  for  $t > 0$ , this corollary and the example above shows that  $\int_0^1 t^{x-1}e^{-t} dt$  converges if and only if  $x > 0$ . Since  $t^{x-1}e^{-t} < Ce^{-t/2}$  for some constant  $C$  and all  $t \geq 1$ , we see that  $\int_1^\infty t^{x-1}e^{-t} dt$  converges for all  $x$ . Thus the integral defining  $\Gamma(x)$  converges for all  $x > 0$ .

We say that the improper integral  $\int_a^\infty f(t) dt$  is absolutely convergent if  $\int_a^\infty |f(t)| dt$  converges. The last corollary, a comparison test for improper integrals, also contains the assertion that an absolutely convergent improper integral is convergent. It is easy to construct an example of an improper integral which is conditionally convergent, i.e., convergent but not absolutely convergent. For instance, one could define  $f(x) = (-1)^n/n$  for  $[x] = n$ . Here is a more natural example, an improper integral with many applications.

**5.44 Example.** Let  $f(x) = \sin x/x$  for  $x > 0$ ,  $f(0) = 1$ . Then  $f$  is continuous on  $[0, \infty)$ , so  $\int_0^\infty f(x) dx$  converges if and only if  $\int_1^\infty f(x) dx$  converges. Now integrating by parts (Theorem 5.32) we have

$$\int_1^R \frac{\sin x}{x} dx = \cos 1 - \frac{\cos R}{R} - \int_1^R \frac{\cos x}{x^2} dx.$$

Since  $|\cos x/x^2| \leq 1/x^2$  and  $\int_1^\infty (1/x^2) dx$  converges, we see that the integral  $\int_1^\infty (\sin x/x) dx$  converges, and hence  $\int_0^\infty f(x) dx$  converges. On the other hand, the integral  $\int_0^\infty |f(x)| dx$  diverges, since

$$\begin{aligned} \int_\pi^{(n+1)\pi} \frac{|\sin x|}{x} dx &= \sum_{k=1}^n \int_{k\pi}^{(k+1)\pi} \frac{|\sin x|}{x} dx \\ &\geq \sum_{k=1}^n \frac{1}{(k+1)\pi} \int_{k\pi}^{(k+1)\pi} |\sin x| dx \\ &= \sum_{k=1}^n \frac{1}{(k+1)\pi} \int_0^\pi \sin x dx \\ &= \sum_{k=1}^n \frac{2}{(k+1)\pi} \end{aligned}$$

which diverges to  $+\infty$  as  $n \rightarrow \infty$ .

The method used in this example can be generalized to prove an analog of Theorem 2.43.

**5.45 Theorem.** Let  $f$  be continuous on  $[a, \infty)$  and let  $g$  be a function of class  $C^1$  on  $[a, \infty)$ . Let  $F(x) = \int_a^x f(t) dt$  for  $x \geq a$ . Suppose that the following conditions hold:

- (a)  $\int_a^\infty |g'(t)| dt < \infty$ ;
- (b)  $F$  is a bounded function; and
- (c)  $g(x) \rightarrow 0$  as  $x \rightarrow \infty$ .

Then  $\int_a^\infty f(x)g(x) dx$  converges.

**Proof.** Integrating by parts, we find

$$\int_a^R f(x)g(x) dx = F(R)g(R) - F(a)g(a) - \int_a^R F(x)g'(x) dx;$$

since  $F$  is bounded and  $\int_a^\infty |g'(x)| dx$  converges, the integral on the right converges by Corollary 5.43, and since  $F$  is bounded and  $g(R) \rightarrow 0$  as  $R \rightarrow \infty$ , the first term on the right approaches 0 as  $R \rightarrow \infty$ . Thus  $\lim_{R \rightarrow \infty} \int_a^R f(x)g(x) dx$  exists. ■

**5.46 Corollary.** Let  $f$  be continuous on  $[a, \infty)$  and let  $g$  be a function of class  $C^1$  on  $[a, \infty)$ . Let  $F(x) = \int_a^x f(t) dt$  for  $x \geq a$ . If  $g$  is decreasing and  $g(x) \rightarrow 0$  as  $x \rightarrow \infty$ , and if  $F$  is bounded, then  $\int_a^\infty f(x)g(x) dx$  converges.

**Proof.** Since  $g'(x) \leq 0$ , we have

$$\int_a^R |g'(x)| dx = - \int_a^R g'(x) dx = g(a) - g(R) \rightarrow g(a)$$

as  $R \rightarrow \infty$ , so  $\int_a^\infty |g'(x)| dx$  converges, and the theorem applies. ■

**5.47 Corollary.** If  $g$  is of class  $C^1$  on  $[a, \infty)$  and  $g(x)$  decreases to 0 as  $x \rightarrow \infty$ , then  $\int_a^\infty g(x) \sin x dx$  and  $\int_a^\infty g(x) \cos x dx$  converge.

This corollary of course includes the last example. Here is another application of it.

**5.48 Example.** We show that  $\int_0^\infty \cos(x^2) dx$  converges. Using the change of variable  $u = x^2$  (see Theorem 5.33) we find

$$\int_1^R \cos(x^2) dx = \int_1^{R^2} \frac{\cos u}{2\sqrt{u}} du.$$

According to the last corollary, the integral on the right approaches a limit as  $R \rightarrow \infty$ , so  $\int_1^\infty \cos(x^2) dx$  converges; this implies the convergence of  $\int_0^\infty \cos(x^2) dx$ .

If  $\sum a_n$  is convergent, then  $a_n \rightarrow 0$  as  $n \rightarrow \infty$ . Here the analogy between series and improper integrals breaks down, as the last example shows. It is not hard to show that the integral in this example is only conditionally convergent. Here is an example of an absolutely convergent integral with an unbounded integrand.

**5.49 Example.** Define  $\phi: \mathbf{R} \rightarrow \mathbf{R}$  by  $\phi(t) = (1 - |t|)^+$ , and let

$$f(t) = \sum_{n=1}^{\infty} \sqrt{n} \phi(2n^2(t - n)).$$

Note that for any given  $t$  the series contains at most one nonzero term. (The reader should sketch  $\phi$  and  $f$ .) For every  $n$ ,

$$\int_{n-1/2}^{n+1/2} f(t) dt = \sqrt{n} \frac{1}{2n^2} \int_{-1}^1 \phi(t) dt = \frac{\sqrt{n}}{2n^2},$$

so  $\int_0^{\infty} f(t) dt = (1/2) \sum_{n=1}^{\infty} n^{-3/2} < +\infty$ . But  $f(t)$  does not approach 0 as  $t \rightarrow \infty$ , in fact,  $\sup_{t \geq R} f(t) = +\infty$  for every  $R$ .

## 5.7 Exercises

1. Using only the definition of the integral, show that (assuming the integrals exist)

$$\int_{-a}^a f(x^2) dx = 2 \int_0^a f(x^2) dx, \quad \int_{-a}^a x f(x^2) dx = 0.$$

2. Show that if  $f$  is continuous on  $[0, 1]$ , then

$$\begin{aligned} \int_0^{\pi/2} f(\cos x) dx &= \int_0^{\pi/2} f(\sin x) dx = \frac{1}{2} \int_0^{\pi} f(\sin x) dx \\ \int_0^{n\pi} f(\cos^2 x) dx &= n \int_0^{\pi} f(\cos^2 x) dx. \end{aligned}$$

3. Show that if  $f$  is the function of Example 3.5, then  $f$  is integrable over  $[0, 1]$ , and  $\int_0^1 f(t) dt = 0$ .

4. Find the limit, as  $n \rightarrow \infty$ , of

$$\sum_{k=1}^n \frac{k^3}{n^4}.$$



5. Show that

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{n}{n^2 + k^2} = \frac{\pi}{4}.$$

6. Show that if  $f$  is integrable over  $[a, b]$ , and  $x_{k,n} = a + k(b-a)/n$  for  $k = 1, 2, \dots, n$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(x_{k,n}) = \frac{1}{b-a} \int_a^b f(x) dx.$$

7. Find all functions  $f$  on  $[0, 1]$  such that  $f$  is continuous on  $[0, 1]$ , and

$$\int_0^x f(t) dt = \int_x^1 f(t) dt$$

for every  $x \in (0, 1)$ .

8. Show that  $\int_0^\pi f(\sin x) \cos x dx = 0$  for any function  $f$  continuous on  $[0, 1]$ .

9. Show that if  $0 < L < R < +\infty$ , then

$$\left| \int_L^R \frac{\sin x}{x} dx \right| < \frac{2}{L}.$$

10. Let  $f$  be of class  $C^1$  on  $[a, b]$ , with  $f(a) = f(b) = 0$ . Show that

$$\int_a^b x f(x) f'(x) dx = -\frac{1}{2} \int_a^b [f(x)]^2 dx.$$

Deduce that if also  $\int_a^b [f(x)]^2 dx = 1$ , then

$$\int_a^b [f'(x)]^2 dx \cdot \int_a^b [x f(x)]^2 dx > \frac{1}{4}.$$

11. Let  $b_{m,n} = \int_0^1 x^m (1-x)^n dx$ , where  $m$  and  $n$  are nonnegative integers. Show that

$$b_{m,n} = \frac{m! n!}{(m+n+1)!}.$$

HINT: Integrate by parts.

12. Let  $f$  be continuous in an interval  $I$  containing 0, and define

$$f_1(x) = \int_0^x f(t) dt, \quad f_2(x) = \int_0^x f_1(t) dt,$$

and, in general,  $f_n(x) = \int_0^x f_{n-1}(t) dt$  for  $n \geq 2$ . Show that

$$f_{n+1}(x) = \int_0^x \frac{(x-t)^n}{n!} f(t) dt$$

for every  $n \geq 0$ . HINT: Integrate by parts.

13. Use the last exercise to prove the following form of Taylor's theorem: if  $f$  is a function of class  $C^{n+1}$  on an interval  $I$  containing 0, then for all  $x \in I$  we have

$$f(x) = \sum_{k=1}^n \frac{f^{(k)}(0)}{k!} x^k + \int_0^x f^{(n+1)}(t) \frac{(x-t)^{n+1}}{(n+1)!} dt.$$

14. If  $f$  is continuous on  $[0, \infty)$  and  $f(x) \rightarrow L$  as  $x \rightarrow \infty$ , show that for  $a > 0$ ,

$$\lim_{n \rightarrow \infty} \int_0^a f(nx) dx = aL.$$

15. Suppose that  $f_n$  is integrable over  $[a, b]$  for every  $n \in \mathbf{N}$ , and that  $(f_n)$  converges uniformly on  $[a, b]$  to some function  $f$ . Show that  $f$  is integrable over  $[a, b]$ .

16. Let  $\phi_n$  be nonnegative functions integrable over  $[-1, 1]$ , satisfying the conditions:

- (a)  $\int_{-1}^1 \phi_n(t) dt = 1$  for every  $n$ ; and
- (b) for every  $\delta > 0$ ,  $\phi_n \rightarrow 0$  uniformly on  $[-1, -\delta] \cup [\delta, 1]$ .

Show that for every  $f$  which is integrable over  $[-1, 1]$  and continuous at 0, we have  $\int_{-1}^1 f(t)\phi_n(t) dt \rightarrow f(0)$  as  $n \rightarrow \infty$ .

17. Let  $f$  be a positive, decreasing continuous function on  $[0, +\infty)$ . Show that for each nonnegative integer  $m$ ,

$$\sum_{n=m+1}^{\infty} f(n) \leq \int_m^{\infty} f(t) dt \leq \sum_{n=m}^{\infty} f(n),$$

and deduce that the improper integral  $\int_0^{\infty} f(t) dt$  converges if and only if the infinite series  $\sum_{n=0}^{\infty} f(n)$  converges.

18. Show that

$$\int_0^{\infty} \frac{\sin t}{t^p} dt$$

converges if and only if  $0 < p < 2$ . Show that for  $0 < p < 2$ , the integrals

$$I_n = \int_{(n-1)\pi}^{n\pi} \frac{\sin t}{t^p} dt$$

have the properties  $I_n = (-1)^n |I_n|$  and  $|I_{n+1}| \leq |I_n|$  for all  $n \in \mathbf{N}$ .

19. Show that

$$\int_x^{\infty} e^{-t^2} dt < \frac{1}{2x} e^{-x^2}$$

for all  $x > 0$ .

20. Discuss the convergence of  $\int_0^\infty x \sin(e^x) dx$ .
21. Show that Theorem 5.34 fails for improper integrals, by exhibiting a sequence  $(f_n)$  of continuous functions on  $[0, \infty)$ , each vanishing outside a bounded interval, which converges uniformly to 0 on  $[0, \infty)$ , such that  $\int_0^\infty f_n = 1$  for each  $n$ .

## 5.8 Notes

- 5.1 The integral sign was introduced by Leibniz (1675); it is a stylized  $S$ , denoting something that originates from summation, but was used for the indefinite integral only. The word integral was first used (1690) for such “sums” by Jakob Bernoulli. The notation with upper and lower limits of integration was introduced by Fourier in 1822, and rapidly became widely adopted; Gauss never used this notation. Cauchy first defined the definite integral as a limit in some sense of Riemann sums, but only for continuous functions, and always sums of the form  $\sum f(x_{j-1})(x_j - x_{j-1})$ . The formulation we give of the definition of integral is essentially due to Riemann, in the middle of the nineteenth century.
- 5.2 The computation of Example 5.2 is due to Fermat, one of the great number of ingenious calculations made in computing areas in the era before the fundamental theorem of calculus became widely known. The origins of integration really go back to Archimedes, or perhaps earlier. Jordan (nineteenth century) defined the *outer content* of a subset  $A$  of  $\mathbf{R}$  to be the infimum of all sums  $\sum_{k=1}^n (b_k - a_k)$  where  $A \subset \bigcup_{k=1}^n (a_k, b_k)$ . We can thus rephrase Theorem 5.15 as follows: if  $f : [a, b] \rightarrow \mathbf{R}$  is bounded, and continuous, except at a set of points having outer content zero, then  $f$  is integrable. The notion of content has been superseded in our century by the notion of measure, which we will discuss in a later chapter.
- 5.3 The inequality (5.1) was first published by Bunyakovsky in 1859 (in French, in a St. Petersburg publication). It was rediscovered in 1885 by Schwarz. The inequality generalizes a similar inequality for finite sums, published by Cauchy in 1821. Both results are subsumed in the corresponding inequality for integrals defined by a measure, which will be given in Chapter 10. The mean value theorem for integrals, Theorem 5.23, was published by Dirichlet in 1837.
- 5.4 The fundamental theorem of calculus, along with a number of other rules for dealing with derivatives and integrals, was independently found by Newton and Leibniz; a famous priority war took place during

the early eighteenth century. Actually, Isaac Barrow had previously published Theorem 5.30.

- 5.5 Cauchy gave as a theorem the permissibility of integrating a point-wise convergent series of functions term-by-term. The problem of justifying the term-by-term integration of a series of functions was first recognized in the early nineteenth century, in connection with Fourier series. Theorem 5.34 and its corollary for series is only the first of such results; we will find more effective theorems in a later chapter.

# 6

## Topology

In this chapter, we extend the notions of neighborhoods, convergent sequences, and continuous functions, which we have studied in the setting of the real line, to more general situations. We are interested especially in the setting of  $\mathbf{R}^d$ , the Euclidean space of dimension  $d$ , but it turns out that the means by which we formulate and analyze the notions of continuity and convergence, and related ideas, carry over to much more general settings, with little or no adaptation; as a result, we shall introduce some fairly abstract notions right from the beginning.

### 6.1 Topological Spaces

**6.1 Definition.** *Let  $X$  be a set. A topology on  $X$  is a collection  $\mathcal{T}$  of subsets of  $X$ , called open sets, having the following properties:*

- (a) *If  $A$  is any set, and  $U_\alpha \in \mathcal{T}$  for every  $\alpha \in A$ , then  $\bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}$ ;*
- (b) *If  $U_\alpha \in \mathcal{T}$  for each  $\alpha$  in some finite set  $F$ , then  $\bigcap_{\alpha \in F} U_\alpha \in \mathcal{T}$ ; and*
- (c)  *$X \in \mathcal{T}$  and  $\emptyset \in \mathcal{T}$ .*

*A topological space is a pair  $(X, \mathcal{T})$ , where  $\mathcal{T}$  is a topology on  $X$ .*

We will usually write simply “ $X$  is a topological space” when no confusion is likely; sometimes there is more than one topology to be considered for  $X$ , and we have to be more careful.

**6.2 Example.** On any set  $X$ , we can introduce two simple-minded topologies. The *trivial topology* on  $X$  consists of  $\{\emptyset, X\}$ . The *discrete topology* on  $X$  is  $\mathcal{P}(X)$ , i.e., every subset of  $X$  is declared to be open. Another artificial topology: let a subset  $U$  of  $X$  be called open if and only if  $U^C = X \setminus U$  is finite.

**6.3 Example.** We define a topology on  $\mathbf{R}$ , called the usual topology, by declaring those subsets of  $\mathbf{R}$  to be open which can be expressed as unions of open intervals. Thus,  $U \subset \mathbf{R}$  is open if and only if for each  $p \in U$  there exists an open interval  $(a, b)$  with  $p \in (a, b) \subset U$ . It is easy to check that the collection of all such sets is a topology; we can obtain  $\emptyset$  open by recognizing it as the union of an empty collection of intervals (or if this is annoying, declare  $\emptyset$  to be open by special dispensation.) We define the usual topology on  $E$ , for any subset  $E$  of  $\mathbf{R}$ , by defining a subset  $A$  of  $E$  to be open if and only if  $A = U \cap E$  for some open subset  $U$  of  $\mathbf{R}$ .

**6.4 Example.** We can define a topology  $\mathcal{T}_u$  on  $\mathbf{R}$  by declaring the empty set,  $\mathbf{R}$ , and the unbounded open intervals  $(-\infty, a)$  ( $a \in \mathbf{R}$ ) to be the open sets. Similarly, we could define a topology  $\mathcal{T}_l$  on  $\mathbf{R}$  to be the empty set and  $\mathbf{R}$ , together with the collection of all  $(a, +\infty)$  ( $a \in \mathbf{R}$ ).

Since topologies on a set  $X$  are subsets of  $\mathcal{P}(X)$ , there is a partial order relation on the set of topologies on  $X$ : if  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are topologies on  $X$ , we say that  $\mathcal{T}_1$  is *weaker than*  $\mathcal{T}_2$ , or that  $\mathcal{T}_2$  is *stronger than*  $\mathcal{T}_1$ , if  $\mathcal{T}_1 \subset \mathcal{T}_2$ . The words *coarser* and *finer* are often used as synonyms for weaker and stronger, respectively. We note that the trivial topology is weaker than every topology, and the discrete topology is stronger than every topology. Of the topologies on  $\mathbf{R}$  mentioned above, we see that the usual topology is stronger than both the topology  $\mathcal{T}_u$  and the topology  $\mathcal{T}_l$ ; these two topologies are not comparable. If  $\mathcal{T}$  is a topology on  $\mathbf{R}$  which is stronger than both  $\mathcal{T}_u$  and  $\mathcal{T}_l$ , then  $\mathcal{T}$  is stronger than the usual topology.

**6.5 Definition.** Let  $X$  be a topological space, and  $x \in X$ . We say that  $N$  is a neighborhood of  $x$  if there exists an open set  $G$  such that  $x \in G \subset N$ . We say that  $x$  is an interior point of  $N$  if  $N$  is a neighborhood of  $x$ .

Thus, a subset  $U$  of  $X$  is open if and only if  $U$  is a neighborhood of  $x$  for every  $x \in U$ ; for if this is true, there exists for each  $x \in U$  an open set  $G_x$  with  $x \in G_x \subset U$ , so  $U = \bigcup_{x \in U} G_x$  is open. We note that this definition of neighborhood agrees with that of Chapter 3 for the case of  $\mathbf{R}$  with its usual topology.

Along with the concept of open set, there is the companion concept of closed set.

**6.6 Definition.** Let  $X$  be a topological space. A subset  $F$  of  $X$  is called closed if its complement  $F^C = X \setminus F$  is open.

It is clear that the class of closed sets has the properties:

- (a) if  $F_\alpha$  is a closed set for each  $\alpha \in A$ , then  $\bigcap_{\alpha \in A} F_\alpha$  is closed;
- (b) if  $F_1, \dots, F_n$  are closed, then  $\bigcup_{k=1}^n F_k$  is closed; and
- (c)  $X$  and  $\emptyset$  are closed.

Conversely, any collection of sets having the properties (a), (b), and (c) can be used to define a topology (the set of their complements). It is often convenient to describe a situation in terms of closed sets rather than open ones.

**6.7 Definition.** If  $E$  is a subset of a topological space  $X$ , the closure of  $E$  is the intersection of all the closed subsets of  $X$  which contain  $E$ . We denote the closure of  $E$  by  $\overline{E}$ , or sometimes by  $\text{cl } E$ .

In view of (c) above, it is clear that  $\overline{E}$  is closed, and since it is contained in every closed set which contains  $E$ , it can be simply described as the smallest closed set which contains  $E$ . Here is another characterization of  $\overline{E}$ .

**6.8 Proposition.** If  $E$  is a subset of a topological space  $X$ , and  $x \in X$ , then  $x \in \overline{E}$  if and only if  $U \cap E \neq \emptyset$  for every open neighborhood  $U$  of  $x$  (and hence for every neighborhood  $U$  of  $x$ ).

**Proof.** If there is an open neighborhood  $U$  of  $x$  such that  $U \cap E = \emptyset$ , then  $X \setminus U$  is a closed set which contains  $E$ , and hence contains  $\overline{E}$ , so  $x \notin \overline{E}$ . If  $x \notin \overline{E}$ , then  $X \setminus \overline{E}$  is an open neighborhood of  $x$  which does not meet  $E$ . ■

**6.9 Definition.** Let  $X$  be a topological space, and let  $E \subset X$ . We say that  $x \in X$  is a limit point of  $E$  if  $E \cap U \setminus \{x\} \neq \emptyset$  for every neighborhood  $U$  of  $x$ . A point of  $E$  which is not a limit point of  $E$  is called an isolated point of  $E$ .

Thus the last proposition says that the closure of any set  $E$  consists of the points of  $E$ , and the limit points of  $E$ . We note that a limit point of  $E$  may or may not belong to  $E$ . In the special case  $E = X$ , we see that a point  $x \in X$  is an isolated point of  $X$  if and only if  $\{x\}$  is open.

**6.10 Definition.** If  $E$  is a subset of a topological space  $X$ , the interior of  $E$ , denoted by  $E^\circ$  or by  $\text{int } E$ , is the union of all open sets which are contained in  $E$ .

It is easy to see that  $E^\circ$  is open, and is, in fact, the largest open set contained in  $E$ . We also note that  $E^\circ$  consists precisely of all interior points of  $E$ , as previously defined. It is not hard to show that the complement in  $X$  of the interior of  $E$  is exactly the closure of the complement of  $E$ . The set  $\overline{E} \setminus E^\circ$  is referred to as the *boundary* of  $E$ ; it is easy to see that the boundary of  $E$  is empty if and only if  $E$  is both open and closed.

**6.11 Definition.** Let  $X$  be a topological space, and let  $D \subset X$ . We say that  $D$  is dense in  $X$  if  $\overline{D} = X$ . More generally, we say that  $D$  is dense in  $E$ , for some  $E \subset X$ , if  $\overline{D} \supset E$ .

Thus  $D$  is dense in  $E$  if and only if for every  $x \in E$  and every neighborhood  $U$  of  $x$ ,  $D \cap U \neq \emptyset$ .

For instance, the set of rational numbers is dense in  $\mathbf{R}$ , as is the set of irrational numbers. A topological space in which there exists a countable dense set is called *separable*.

## 6.2 Continuous Mappings

The purpose of topological spaces is to have a setting for the notion of continuous mappings. The following definition expands the notion of continuity given in Chapter 3 (see Proposition 3.8).

**6.12 Definition.** Let  $X$  and  $Y$  be topological spaces, and let  $f : X \rightarrow Y$ . We say that  $f$  is continuous at  $x \in X$  if  $f^{-1}(N)$  is a neighborhood of  $x$  for every neighborhood  $N$  of  $f(x)$ . We say that  $f$  is continuous on  $A \subset X$  if  $f$  is continuous at each  $x \in A$ , and we say simply that  $f$  is continuous if  $f$  is continuous at each  $x \in X$ .

We can rephrase this definition without using the word neighborhood:  $f$  is continuous at  $x$  if for any open  $V \subset Y$  with  $f(x) \in V$ , there exists an open subset  $U$  of  $X$ , with  $x \in U$ , such that  $f(y) \in V$  for every  $y \in U$ .

**6.13 Example.** If  $f : X \rightarrow Y$  is a constant map, i.e., there exists  $y_0 \in Y$  such that  $f(x) = y_0$  for all  $x \in X$ , then  $f$  is continuous. If  $Y$  has the trivial topology, then every map of  $X$  into  $Y$  is continuous. If  $X$  has the discrete topology, then every  $f : X \rightarrow Y$  is continuous.

If  $X$  is a topological space, then the identity map of  $X$  is continuous. More generally, if  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are topologies on  $X$ , then the identity map  $i : (X, \mathcal{T}_1) \rightarrow (X, \mathcal{T}_2)$  is continuous if and only if  $\mathcal{T}_2$  is weaker than  $\mathcal{T}_1$ .

**6.14 Proposition.** Let  $X$  and  $Y$  be topological spaces. If  $f : X \rightarrow Y$ , then  $f$  is continuous if and only if  $f^{-1}(V)$  is open (in  $X$ ) whenever  $V$  is open (in  $Y$ ), if and only if  $f^{-1}(F)$  is closed whenever  $F$  is closed.

**Proof.** Suppose  $f$  is continuous, and  $V$  is an open subset of  $Y$ . Let  $U = f^{-1}(V)$ . Then  $V$  is a neighborhood of  $f(x)$  for each  $x \in U$ , so  $U$  is a neighborhood of  $x$  for each  $x \in U$ , by Definition 6.12. Thus  $U$  is an open subset of  $X$ .

Suppose next that  $f^{-1}(V)$  is open whenever  $V$  is open. Then for each  $x \in X$ , and any neighborhood  $N$  of  $f(x)$ , there exists an open  $V$  such that  $f(x) \in V \subset N$ ; if  $U = f^{-1}(V)$ , we have  $x \in U$ ,  $U$  is open, and



$U \subset f^{-1}(N)$ , so  $f^{-1}(N)$  is a neighborhood of  $x$ . Thus  $f$  is continuous at  $x$  for each  $x \in X$ , i.e.,  $f$  is continuous.

Finally, we remark that  $f^{-1}(A^C) = [f^{-1}(A)]^C$  for any set  $A$  and any map  $f$ , so the condition that  $f^{-1}(G)$  is open for every open  $G$  is equivalent to the condition that  $f^{-1}(F)$  is closed for every closed  $F$ . ■

**6.15 Proposition.** *Let  $X, Y, Z$  be topological spaces,  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$ . If  $f$  is continuous at  $x$ , and  $g$  is continuous at  $f(x)$ , then the composition  $g \circ f$  is continuous at  $x$ .*

**Proof.** Let  $W$  be a neighborhood of  $g \circ f(x)$ ; then  $V = g^{-1}(W)$  is a neighborhood of  $f(x)$  in  $Y$ , since  $g$  is continuous at  $f(x)$ , and therefore  $(g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W)) = f^{-1}(V)$  is a neighborhood of  $x$ , since  $f$  is continuous. ■

**6.16 Definition.** *Let  $X$  and  $Y$  be topological spaces, and let  $f : X \rightarrow Y$ . We say that  $f$  is a homeomorphism of  $X$  onto  $Y$  if  $f$  is bijective, and  $f$  and  $f^{-1}$  are continuous.*

The identity map is a homeomorphism of any topological space onto itself. If  $f : X \rightarrow Y$  is a homeomorphism, then  $f^{-1} : Y \rightarrow X$  is also a homeomorphism. We say that the spaces  $X$  and  $Y$  are *homeomorphic* if there exists a homeomorphism from  $X$  to  $Y$ . Using Proposition 6.15, it is easy to see that this defines an equivalence relation on any set of topological spaces.

**6.17 Example.** The interval  $(-1, 1)$  is homeomorphic to  $\mathbf{R}$ ; for instance, the map  $x \mapsto \tan \pi x/2$  is a continuous bijective map of  $(-1, 1)$  onto  $\mathbf{R}$ , whose inverse is (necessarily) continuous. Another homeomorphism of  $(-1, 1)$  onto  $\mathbf{R}$  is given by taking  $f(x) = x/(1 - x^2)$ .

An interval  $[a, b]$  is not homeomorphic to any open interval  $(c, d)$  in  $\mathbf{R}$ . For if  $f : [a, b] \rightarrow (c, d)$  is continuous and injective, then it is strictly monotone, as we saw in the proof of Theorem 3.16. If  $f$  is strictly increasing, then  $f(t) > f(a) > c$  for all  $t \in [a, b]$ , so  $f$  is not surjective, and if  $f$  is decreasing, then  $f(t) < f(a) < d$  for all  $t$ , so  $f$  is not surjective. Similarly,  $[a, b]$  is not homeomorphic to any interval  $(c, d)$  or  $[c, d]$ , etc.

## 6.3 Metric Spaces

**6.18 Definition.** *Let  $X$  be a nonempty set. A metric on  $X$ , or distance function on  $X$ , is a map  $\rho : X \times X \rightarrow \mathbf{R}$  with the following properties:*

- (a)  $\rho(x, y) \geq 0$  for every  $x, y \in X$ , with equality holding if and only if  $x = y$ ;

- (b)  $\rho(x, y) = \rho(y, x)$  for every  $x, y \in X$ ; and  
 (c)  $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$  for every  $x, y, z \in X$ .

A metric space is a pair  $(X, \rho)$ , where  $\rho$  is a metric on  $X$ .

We will usually abuse language by writing “the metric space  $X$ ” instead of the correct “the metric space  $(X, \rho)$ ” when no confusion is likely. In those situations where more than one metric is considered on the same set  $X$ , we must of course use the proper language.

**6.19 Definition.** Let  $(X, \rho)$  be a metric space. For each  $a \in X$  and  $r > 0$ , we define the open ball with center  $a$  and radius  $r$  to be the set

$$B(a, r) = \{x \in X : \rho(x, a) < r\}.$$

**6.20 Definition.** Let  $X$  be a set, and  $\rho$  a metric on  $X$ . The topology on  $X$  induced by  $\rho$  is defined as follows:  $G \subset X$  is called open if for every  $a \in G$  there exists  $\delta > 0$  such that  $B(a, \delta) \subset G$ .

We note that an open ball is indeed an open set by this definition: for if  $x \in B(a, r)$ , and we take  $\delta = r - \rho(x, a)$ , so  $\delta > 0$ , then for all  $y \in B(x, \delta)$  we have  $\rho(y, a) \leq \rho(y, x) + \rho(x, a) < \delta + r - \delta = r$ , so  $B(x, \delta) \subset B(a, r)$ . It is not hard to check that the class of sets called open by Definition 6.20 is indeed a topology on  $X$ .

On any set  $X$ , we can define a metric  $\rho$  by putting  $\rho(x, y) = 1$  for any  $x, y \in X$  with  $x \neq y$ , and  $\rho(x, x) = 0$  for all  $x \in X$ . This is called the discrete metric, and the topology it induces is evidently the discrete topology, since  $B(x, 1) = \{x\}$  for every  $x \in X$ .

The standard metric on  $\mathbf{R}$  is given by  $\rho(x, y) = |x - y|$ ; it is trivial to verify that this is a metric, and that it induces the usual topology on  $\mathbf{R}$ .

The most important examples of metric spaces are Euclidean space and subsets of Euclidean space.

If  $X$  is any set, and  $d$  is a positive integer, we denote by  $X^d$  the  $d$ -fold Cartesian product of  $X$  with itself:  $X^d = X \times X \times \cdots \times X$ , which is also described as the set of all finite sequences  $(x_1, \dots, x_d)$ , with  $x_j \in X$  for  $j = 1, \dots, d$ . The space  $\mathbf{R}^d$  is known as  $d$ -dimensional Euclidean space. If  $\mathbf{x} = (x_1, \dots, x_d)$  and  $\mathbf{y} = (y_1, \dots, y_d)$ , then the sum of  $\mathbf{x}$  and  $\mathbf{y}$ , and the product of a scalar  $c \in \mathbf{R}$  with  $\mathbf{x}$ , are defined by

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, \dots, x_d + y_d), \quad c\mathbf{x} = (cx_1, cx_2, \dots, cx_d)$$

and  $\mathbf{R}^d$  with these operations is a vector space. It is also endowed with a standard inner product, given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^d x_j y_j,$$

where  $\mathbf{x} = (x_1, \dots, x_d)$ ,  $\mathbf{y} = (y_1, \dots, y_d)$ , and  $c \in \mathbf{R}$ . It is easy to see that for all  $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$  and  $c \in \mathbf{R}$ ,

$$\begin{aligned}\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle &= \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle, \\ \langle c\mathbf{x}, \mathbf{y} \rangle &= c\langle \mathbf{x}, \mathbf{y} \rangle, \\ \langle \mathbf{x}, \mathbf{y} \rangle &= \langle \mathbf{y}, \mathbf{x} \rangle,\end{aligned}$$

and that  $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$  for all  $\mathbf{x} \in \mathbf{R}^d$ , with equality only when  $\mathbf{x} = \mathbf{0}$ , where  $\mathbf{0} = (0, \dots, 0)$ . The inner product is also often denoted by  $\mathbf{x} \cdot \mathbf{y}$ , and sometimes called the *dot product*. We assume the reader to be familiar with these notions. We define the *length* of  $\mathbf{x}$ , or *norm* of  $\mathbf{x}$ , to be  $|\mathbf{x}| = \left(\sum_{j=1}^d x_j^2\right)^{1/2} = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$ . It is evident that  $|\mathbf{x}| \geq 0$  for all  $\mathbf{x}$ , with equality only for  $\mathbf{x} = \mathbf{0}$ . Of course, when  $d = 1$  the length of  $x$  is the absolute value of  $x$ , so there is no conflict of notation. The next proposition is called the *Cauchy inequality*.

**6.21 Proposition.** *For any  $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$ , we have*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq |\mathbf{x}||\mathbf{y}|.$$

**Proof.** We have

$$\begin{aligned}0 &\leq \sum_{i,j=1}^d (x_i y_j - x_j y_i)^2 = \sum_{i,j=1}^d (x_i^2 y_j^2 + x_j^2 y_i^2 - 2x_i x_j y_i y_j) \\ &= \sum_{i=1}^d \sum_{j=1}^d x_i^2 y_j^2 + \sum_{i=1}^d \sum_{j=1}^d x_j^2 y_i^2 - 2 \sum_{i=1}^d \sum_{j=1}^d x_i y_i x_j y_j \\ &= \sum_{i=1}^d \langle \mathbf{y}, \mathbf{y} \rangle x_i^2 + \sum_{i=1}^d \langle \mathbf{x}, \mathbf{x} \rangle y_i^2 - 2 \sum_{i=1}^d \langle \mathbf{x}, \mathbf{y} \rangle x_i y_i \\ &= 2\langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle - 2\langle \mathbf{x}, \mathbf{y} \rangle^2,\end{aligned}$$

from which the desired inequality follows. ■

**6.22 Corollary.** *For any  $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$ , we have  $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$ .*

**Proof.** Using the Cauchy inequality, we get

$$\begin{aligned}|\mathbf{x} + \mathbf{y}|^2 &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle \\ &\leq |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2|\mathbf{x}||\mathbf{y}| = (|\mathbf{x}| + |\mathbf{y}|)^2,\end{aligned}$$

and the desired inequality follows by taking square roots. ■

**6.23 Definition.** *The standard metric on  $\mathbf{R}^d$  is given by  $\rho(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$ .*

It is an easy consequence of the last corollary that the standard metric on  $\mathbf{R}^d$  is indeed a metric on  $\mathbf{R}^d$ . Having taken the conceptual leap in regarding  $\mathbf{R}^d$  as a "space" for  $d$  different from 1, 2, or 3, we might as well go farther, and consider spaces of infinite dimension. The most natural example is perhaps the following.

**6.24 Example.** Let  $\mathcal{H}$  be the set of all sequences  $(x_n)_{n=1}^{\infty}$  in  $\mathbf{R}$  with the property that  $\sum_{n=1}^{\infty} x_n^2$  converges. It follows from the easy inequality  $(a+b)^2 \leq 2(a^2 + b^2)$  ( $a, b \in \mathbf{R}$ ) that  $(x_n + y_n) \in \mathcal{H}$  whenever  $(x_n) \in \mathcal{H}$  and  $(y_n) \in \mathcal{H}$ , and thus  $\mathcal{H}$  is a vector space over  $\mathbf{R}$ . If  $\mathbf{x} = (x_n)$  and  $\mathbf{y} = (y_n)$  are elements of  $\mathcal{H}$ , then, since  $|x_n y_n| \leq \frac{1}{2}(x_n^2 + y_n^2)$  for every  $n$ , the series  $\sum_{n=1}^{\infty} x_n y_n$  converges (absolutely); we denote its sum by  $\langle \mathbf{x}, \mathbf{y} \rangle$ , and call it the inner product of  $\mathbf{x}$  and  $\mathbf{y}$ . We define  $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ . Since  $|\sum_{k=1}^n x_k y_k|^2 \leq (\sum_{k=1}^n x_k^2)(\sum_{k=1}^n y_k^2) \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2$  for every  $n$ , we have the inequality  $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$  for every  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ . Just as in the finite-dimensional case, it follows that we have the triangle inequality  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$  for every  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ . We define the distance between  $\mathbf{x}$  and  $\mathbf{y}$  to be  $\|\mathbf{x} - \mathbf{y}\|$ . Our intuition about distance will not often lead us astray in thinking about  $\mathcal{H}$ . In addition, we say that  $\mathbf{x}$  and  $\mathbf{y}$  are orthogonal if  $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ , and our experience in the plane or 3-space will prove a reliable guide in thinking about orthogonality in  $\mathcal{H}$ . The space  $\mathcal{H}$  is known as Hilbert space, and is often denoted by  $l^2$  (pronounced "little ell two").

Two different metrics on a set  $X$  may induce the same topology. For instance, if  $\rho$  is a metric on  $X$ , and we put  $\rho'(x, y) = \min\{1, \rho(x, y)\}$ , it is easy to check that  $\rho'$  is also a metric on  $X$ . Since  $B_\rho(x, \delta) = B_{\rho'}(x, \delta)$  for any  $x \in X$  and  $\delta \in (0, 1)$ , where the meaning of the notation should be obvious, it follows that the topologies induced by  $\rho$  and  $\rho'$  are identical.

More generally, if there exists a constant  $C$  such that  $\rho' \leq C\rho$ , then the topology induced by the metric  $\rho$  is stronger than the topology induced by  $\rho'$ . For  $B_\rho(x, \delta/C) \subset B_{\rho'}(x, \delta)$  for any  $x \in X$  and  $\delta > 0$ , so that if  $x$  is an interior point of  $G$  with respect to  $\rho'$ , then  $x$  is an interior point of  $G$  with respect to  $\rho$ ; it follows that if  $G$  is open for the topology defined by  $\rho'$ , then it is open for the topology defined by  $\rho$ . The inequality  $\rho'(x, y) \leq C\rho(x, y)$  need only hold for all  $x, y \in X$  such that  $\rho(x, y) < \epsilon$  (for some fixed  $\epsilon > 0$ ) for this argument to work.

It is easy to see that if  $(X, \rho)$  and  $(Y, \rho')$  are metric spaces, and  $f$  maps  $X$  into  $Y$ , then  $f$  is continuous at  $x \in X$  if and only if for every  $\epsilon > 0$  there exists  $\delta > 0$  such that  $\rho'(f(x), f(y)) < \epsilon$  for all  $y \in X$  with  $\rho(x, y) < \delta$ . This was our original definition of continuity for real-valued functions of a real variable. We can also generalize the notion of uniform continuity to the context of metric spaces in the obvious way.

**6.25 Definition.** Let  $(X, \rho)$  and  $(Y, \rho')$  be metric spaces, and let  $f : X \rightarrow Y$ . We say that  $f$  is uniformly continuous if for every  $\epsilon > 0$  there exists

$\delta > 0$  such that  $\rho'(f(x), f(y)) < \epsilon$  for all  $x, y \in X$  with  $\rho(x, y) < \delta$ .

## 6.4 Constructing Topological Spaces

In this section, we give some standard ways to construct topologies, or to form new topological spaces from old ones.

**6.26 Definition.** Let  $(X, \mathcal{T})$  be a topological space, and let  $Y \subset X$ . The relative topology of  $Y$  is the collection  $\mathcal{T}_Y = \{U \cap Y : U \in \mathcal{T}\}$ . We call  $(Y, \mathcal{T}_Y)$  a subspace of  $(X, \mathcal{T})$ .

It is trivial to verify that the relative topology is indeed a topology on  $Y$ . An element of this topology is described as being relatively open, or open in  $Y$ . Thus every subset of  $X$  is open in itself. If  $Y$  is an open subset of  $X$ , then (and only then) the relatively open subsets of  $Y$  are the open subsets of  $Y$ , i.e., those subsets of  $Y$  which are open (in  $X$ ). If  $Y$  is a closed subset of  $X$ , then (and only then) the relatively closed subsets of  $Y$  are just the closed subsets of  $Y$ , i.e., the subsets of  $Y$  which are closed (as subsets of  $X$ ). If  $\rho$  is a metric on  $X$ , then the restriction of  $\rho$  to  $Y \times Y$  is obviously a metric on  $Y$ ; the topology it induces on  $Y$  is precisely the relative topology. As usual, when there seems no danger of confusion, we refer to  $Y$  as a subspace of  $X$ , rather than the accurate  $(Y, \mathcal{T}_Y)$  as a subspace of  $(X, \mathcal{T})$ .

For example, we observe that  $\{q \in \mathbf{Q} : |q| < \sqrt{2}\}$  is neither open nor closed in  $\mathbf{R}$ , but is both open and closed in  $\mathbf{Q}$ .

**6.27 Definition.** Let  $\mathcal{T}$  be a topology on a set  $X$ . A base for  $\mathcal{T}$  is a collection  $\mathcal{B} \subset \mathcal{T}$  with the property that for every  $U \in \mathcal{T}$ , we have  $U = \bigcup \{G \in \mathcal{B} : G \subset U\}$ .

For example, the open balls form a base for the topology induced by a metric on  $X$ . In fact, the open balls of radius less than  $\epsilon$  form a base for any  $\epsilon > 0$ ; also, the open balls of radius  $1/n$  ( $n \in \mathbf{N}$ ) also form a base. For  $\mathbf{R}$ , the collection  $\{(q - 1/n, q + 1/n) : q \in \mathbf{Q}, n \in \mathbf{N}\}$  is a base for the usual topology. It is often useful to know that a topological space has a countable base; such spaces are said to *satisfy the second axiom of countability*.

**6.28 Proposition.** Let  $X$  be a set. A family  $\mathcal{B}$  of subsets of  $X$  is a base for some topology on  $X$  if and only if the following two conditions hold:

- (a)  $\bigcup_{U \in \mathcal{B}} U = X$ ; and
- (b) for every finite subset  $\{U_1, \dots, U_n\}$  of  $\mathcal{B}$ , and every  $x \in \bigcap_{k=1}^n U_k$ , there exists  $V \in \mathcal{B}$  such that  $x \in V \subset \bigcap_{k=1}^n U_k$ .

**Proof.** If  $\mathcal{B}$  satisfies the two conditions, we define  $\mathcal{T}$  to be the collection of all  $U \subset X$  which are unions of sets in  $\mathcal{B}$ , along with the empty set. By the first condition,  $X \in \mathcal{T}$ , and it is obvious that  $\mathcal{T}$  is closed under arbitrary unions. Suppose that  $G_k \in \mathcal{T}$  for  $k = 1, \dots, n$ , and let  $G = \bigcap_{k=1}^n G_k$ . For each  $x \in G$ , there exists for each  $k$  some  $U_k \in \mathcal{B}$  such that  $x \in U_k \subset G_k$  (this is the definition of  $\mathcal{T}$ ), and the hypothesis then gives the existence of  $V \in \mathcal{T}$  such that  $x \in V \subset G$ . Thus  $G$  is the union of members of  $\mathcal{B}$ , so  $G \in \mathcal{T}$ . Thus  $\mathcal{T}$  is a topology on  $X$ . The proof of "only if" is left to the reader. ■

**6.29 Definition.** Let  $(X_j, \mathcal{T}_j)$  be topological spaces, for  $j = 1, 2, \dots, n$ . We define the product topology on  $\prod_{k=1}^n X_k = X_1 \times X_2 \times \dots \times X_n$  by taking the collection  $\{U_1 \times \dots \times U_n : U_j \in \mathcal{T}_j\}$  to be a base.

Since

$$(U_1 \times \dots \times U_n) \cap (V_1 \times \dots \times V_n) = U_1 \cap V_1 \times \dots \times U_n \cap V_n$$

we see that the sufficient condition of Proposition 6.28 is satisfied, so the definition makes sense: the collection named is indeed the base for a topology. It is possible to define other topologies on the product space  $X_1 \times \dots \times X_n$ , but in the future the product topology will always be understood. If each  $X_k$  is a metric space, with metric  $\rho_k$ , we can define the *product metric* on  $X_1 \times \dots \times X_n$  by

$$\rho((x_1, \dots, x_n), (y_1, \dots, y_n)) = \max_{1 \leq k \leq n} \rho_k(x_k, y_k).$$

It is easy to see that the product metric induces the product topology. It is also easy to see that the product topology on  $\mathbf{R}^n$ , derived from the usual topology on  $\mathbf{R}$ , coincides with the topology on  $\mathbf{R}^n$  derived from the standard metric.

If  $X = \prod_{k=1}^n X_k$ , the map  $\pi_k$  of  $X$  to  $X_k$  defined by  $\pi_k(x_1, \dots, x_n) = x_k$  is called the *k*th coordinate projection on  $X$ .

**6.30 Lemma.** If  $X_1, \dots, X_n$  are topological spaces, then for each  $k$ , the coordinate projection  $\pi_k : \prod_{j=1}^n X_j \rightarrow X_k$  is continuous.

**Proof.** If  $V$  is an open set in  $X_k$ , then

$$\pi_k^{-1}(V) = X_1 \times \dots \times V \times \dots \times X_n$$

which is open in  $X$ . ■

**6.31 Proposition.** Let  $X, Y_1, \dots, Y_n$  be topological spaces, and let  $Y = \prod_{k=1}^n Y_k$ . A map  $f : X \rightarrow Y$  is continuous if and only if  $\pi_k \circ f$  is continuous for each  $k$ ,  $1 \leq k \leq n$ .

**Proof.** If  $f$  is continuous, then  $\pi_k \circ f$  is continuous for each  $k$  by the last lemma, and Proposition 6.15. Now suppose that  $\pi_k \circ f$  is continuous for each  $k$ . Suppose  $V = \prod_{k=1}^n V_k$ , where  $V_k$  is open in  $Y_k$  for each  $k$ . Then

$$f^{-1}(V) = \{x \in X : \pi_k(f(x)) \in V_k, 1 \leq k \leq n\} = \bigcap_{k=1}^n (\pi_k \circ f)^{-1}(V_k),$$

and since  $\pi_k \circ f$  is continuous for each  $k$ , we know that  $(\pi_k \circ f)^{-1}(V_k)$  is open for each  $k$  (Proposition 6.14), and hence  $f^{-1}(V)$  is open. Since any open subset  $G$  of  $Y$  is the union of sets of this type, it follows that  $f^{-1}(G)$  is open in  $X$  for every open subset  $G$  of  $Y$ , so  $f$  is continuous by Proposition 6.14. ■

If  $\mathcal{T}$  and  $\mathcal{T}'$  are topologies on a set  $X$ , then  $\mathcal{T} \cap \mathcal{T}'$  is again a topology on  $X$ ; more generally, if  $\mathcal{T}_\alpha$  is a topology on  $X$  for each  $\alpha$  in some index set  $A$ , then  $\mathcal{T} = \bigcap_{\alpha \in A} \mathcal{T}_\alpha$  is again a topology on  $X$ , as is easily verified. Evidently,  $\mathcal{T}$  is the strongest topology on  $X$  which is weaker than every  $\mathcal{T}_\alpha$ . Given any collection  $\mathcal{S}$  of subsets of  $X$ , there is a unique weakest topology which contains  $\mathcal{S}$ : it is the intersection of all topologies on  $X$  which contain  $\mathcal{S}$  (this is a nonempty collection of topologies, since the discrete topology  $\mathcal{P}(X)$  is one such). We say that  $\mathcal{S}$  is a *subbase* for the topology  $\mathcal{T}$ , or that  $\mathcal{T}$  is the *topology generated by  $\mathcal{S}$*  if  $\mathcal{T}$  is the weakest topology containing  $\mathcal{S}$ . If  $X$  is the union of sets in  $\mathcal{S}$ , then the collection of all finite intersections of sets in  $\mathcal{S}$  forms a base for the topology generated by  $\mathcal{S}$ .

If  $X$  is a set,  $Y$  a topological space, and  $f : X \rightarrow Y$ , then there is a weakest topology on  $X$  which makes  $f$  continuous, namely,  $f^{-1}(\mathcal{T})$ , where  $\mathcal{T}$  is the topology on  $Y$ . More generally, if  $X$  is a set, and  $\{Y_\alpha : \alpha \in A\}$  is a collection of topological spaces, indexed by some set  $A$ , and if  $f_\alpha : X \rightarrow Y_\alpha$  for each  $\alpha \in A$ , there is a weakest topology on  $X$  for which each  $f_\alpha$  is continuous; this is the topology generated by  $\bigcup_{\alpha \in A} f_\alpha^{-1}(\mathcal{T}_\alpha)$ , where  $\mathcal{T}_\alpha$  is the topology that comes with  $Y_\alpha$ . In other words, we take as a base for a topology on  $X$  the collection of all sets of the form  $\bigcap_{\alpha \in F} f_\alpha^{-1}(U_\alpha)$ , where  $F$  runs through all finite subsets of  $A$ , and  $U_\alpha \in \mathcal{T}_\alpha$  for each  $\alpha \in F$ .

An important special case of this idea (which will not be used later in this book) is the construction of the product of an arbitrary collection of topological spaces. If  $X_\alpha$  is a nonempty set for each  $\alpha$  in some (nonempty) index set  $A$ , the Cartesian product  $\prod_{\alpha \in A} X_\alpha$  is defined to be the set of all maps  $x : \alpha \mapsto x_\alpha$  of  $A$  into  $\bigcup_{\alpha \in A} X_\alpha$  with the property that  $x_\alpha \in X_\alpha$  for each  $\alpha \in A$ . A frequent case is when  $X_\alpha = X$  for every  $\alpha$ , when  $\prod_{\alpha \in A} X_\alpha$  is also written  $X^A$ . When  $A = \{1, \dots, n\}$ , then  $X^A$  reduces to  $X^n$ . We see that  $\mathbf{R}^{\mathbf{N}}$  is the set of all maps of  $\mathbf{N}$  into  $\mathbf{R}$ , i.e., all infinite sequences  $(x_n)_{n=1}^\infty$  of real numbers, and if  $I = [0, 1]$ , then  $\mathbf{R}^I$  denotes the set of all real-valued functions on  $[0, 1]$ . For each  $\alpha \in A$ , the map  $\pi_\alpha : X \rightarrow X_\alpha$  defined by  $\pi_\alpha(x) = x_\alpha$  is called the  $\alpha$ th coordinate function.

If  $(X_\alpha, \mathcal{T}_\alpha)$  is a topological space for each  $\alpha \in A$ , we define the *product topology* on  $X = \prod_{\alpha \in A} X_\alpha$  to be the weakest topology on  $X$  for which each coordinate projection  $\pi_\alpha$  is continuous. For each fixed  $\alpha$ ,  $\pi_\alpha^{-1}(s_{\mathcal{T}_\alpha})$  is the weakest topology on  $X$  which makes  $\pi_\alpha$  continuous; so the product topology on  $X$  is the topology generated by the collection of all sets of the form  $\pi_\alpha^{-1}(U_\alpha)$ , where  $\alpha \in A$  and  $U_\alpha \in \mathcal{T}_\alpha$ . According to the remark made above, a base for this topology then consists of all finite intersections of such sets, i.e., of all sets having the form

$$\{x \in X : x_\alpha \in U_\alpha \text{ for each } \alpha \in F\},$$

where  $F$  is a finite subset of  $A$ , and  $U_\alpha$  is open in  $X_\alpha$  for each  $\alpha \in F$ . Of course, when  $A$  is finite, this reduces to the definition previously given. It is important to notice that if  $U_\alpha$  is open in  $X_\alpha$  for each  $\alpha \in A$ , then  $\prod_{\alpha \in A} U_\alpha$  is not necessarily open in  $X$ . The analogue of Proposition 6.31 is true: if  $f : Y \rightarrow \prod_{\alpha \in A} X_\alpha$ , where  $Y$  and  $X_\alpha$  are topological spaces for each  $\alpha \in A$ , then  $f$  is continuous if and only if  $\pi_\alpha \circ f$  is continuous for every  $\alpha \in A$ . The proof of Proposition 6.31 carries over unchanged to this situation.

Besides being continuous, the coordinate functions on a product space have another nice property.

**6.32 Definition.** Let  $X$  and  $Y$  be topological spaces, and let  $f : X \rightarrow Y$ . We say that  $f$  is an *open mapping* if  $f(U)$  is open for every open  $U$ . We say that  $f$  is a *closed mapping* if  $f(C)$  is closed for every closed  $C$ .

We remarked above that  $f^{-1}(G)$  is open for every  $G$  if and only if  $f^{-1}(F)$  is closed for every  $F$ , since  $f^{-1}(A^c) = [f^{-1}(A)]^c$  for any map  $f$ , any  $A$ . However, if  $f$  is an open map, it need not be true that  $f(F)$  is a closed set whenever  $F$  is closed.

**6.33 Proposition.** If  $X = \prod_{\alpha \in A} X_\alpha$  is the topological product of the topological spaces  $X_\alpha$  ( $\alpha \in A$ ), then each coordinate projection  $\pi_\alpha : X \rightarrow X_\alpha$  ( $\alpha \in A$ ) is an open map.

**Proof.** Since  $f(\bigcup_{\beta \in B} U_\beta) = \bigcup_{\beta \in B} f(U_\beta)$  for any function  $f$ , any collection of sets  $\{U_\beta : \beta \in B\}$ , it suffices to show that  $\pi_\alpha(U)$  is open for every  $U$  in a base for the topology of  $X$ . Thus it suffices to show that  $\pi_\alpha(U)$  is open when  $U = \{x \in X : x_\beta \in U_\beta, \beta \in F\}$  when  $F$  is a finite subset of  $A$ , and  $U_\beta$  is open in  $X_\beta$  for each  $\beta \in F$ . But if  $\alpha \notin F$ , then  $\pi_\alpha(U) = X_\alpha$ , and if  $\alpha \in F$ , then  $\pi_\alpha(U) = U_\alpha$ , so  $\pi_\alpha(U)$  is open in either case. ■

The coordinate functions are not, in general, closed maps, as we see in the example  $\mathbf{R}^2$ . The set  $F = \{(x, y) \in \mathbf{R}^2 : xy = 1\}$  is a closed subset of  $\mathbf{R}^2$  (it is the inverse image of  $\{1\}$  under the continuous map  $(x, y) \mapsto xy$  of  $\mathbf{R}^2 \rightarrow \mathbf{R}$ ), but  $\pi_1(F) = \mathbf{R} \setminus 0$ , which is not a closed subset of  $\mathbf{R}$ .



## 6.5 Sequences

**6.34 Definition.** Let  $X$  be a topological space, and let  $(x_n)$  be a sequence in  $X$ . We say that  $(x_n)$  converges to  $x$ , and write  $x_n \rightarrow x$  as  $n \rightarrow \infty$ , or that  $\lim_{n \rightarrow \infty} x_n = x$ , if  $x \in X$  and for every neighborhood  $U$  of  $x$  there exists  $n_0$  such that  $x_n \in U$  for every  $n \geq n_0$ .

In general, a sequence in a topological space can converge to more than one point. For instance, if  $X$  has the trivial topology, every sequence converges simultaneously to every point in  $X$ . To avoid this kind of outrage, we can restrict ourselves to better behaved topological spaces.

**6.35 Definition.** A topological space is called a Hausdorff space if for every pair of points  $x, y \in X$  with  $x \neq y$ , there exist disjoint open sets  $U$  and  $V$  with  $x \in U$  and  $y \in V$ .

In particular, if  $X$  is a Hausdorff space, each singleton set  $\{x\}$  is a closed set. A metric space is a Hausdorff space. The proof of the next proposition is very easy, and will be omitted.

**6.36 Proposition.** If  $X$  is a Hausdorff space,  $(x_n)$  a sequence in  $X$ , and if  $x_n \rightarrow x$  and  $x_n \rightarrow y$  as  $n \rightarrow \infty$ , then  $x = y$ .

**6.37 Proposition.** If  $X$  is a metric space,  $x \in X$ , and  $E \subset X$ , then  $x \in \overline{E}$  if and only if there exists a sequence  $(x_n)$  in  $E$  which converges to  $x$ . In particular,  $E$  is closed if and only if it contains the limits of all convergent sequences in  $E$ .

**Proof.** If  $x_n \rightarrow x$  as  $n \rightarrow \infty$ , and  $U$  is any neighborhood of  $x$ , then  $x_n \in U$  for all sufficiently large  $n$ ; if  $(x_n)$  is a sequence in  $E$ , it follows that  $x \in \overline{E}$ . Conversely, if  $x \in \overline{E}$ , then for every  $n$ , there exists  $x_n \in E$  with  $\rho(x_n, x) < 1/n$ , so the sequence  $(x_n)$  in  $E$  converges to  $x$ . ■

**6.38 Proposition.** Let  $X$  be a metric space and  $Y$  a topological space. A map  $f$  of  $X$  into  $Y$  is continuous at  $x \in X$  if and only if for every sequence  $(x_n)$  in  $X$  such that  $\lim x_n = x$ , we have  $\lim f(x_n) = f(x)$ .

**Proof.** Suppose  $f$  is continuous at  $x$ , and  $x_n \rightarrow x$  as  $n \rightarrow \infty$ . Let  $N$  be any neighborhood of  $f(x)$ . Then  $f^{-1}(N)$  is a neighborhood of  $x$ , so there exists  $n_0$  such that  $x_n \in f^{-1}(N)$  for every  $n \geq n_0$ , i.e., such that  $f(x_n) \in N$  for every  $n \geq n_0$ . Thus  $\lim f(x_n) = f(x)$ . This part of the argument holds for any topological spaces  $X$  and  $Y$ .

Now suppose that for every sequence  $(x_n)$  in  $X$  with  $\lim x_n = x$  we have  $\lim f(x_n) = f(x)$ . Let  $N$  be any neighborhood of  $f(x)$ . If  $f^{-1}(N)$  is not a neighborhood of  $x$ , then for every  $n$ ,  $B(x, 1/n) \not\subset f^{-1}(N)$ . Thus there exists for each  $n$  some  $x_n \in X$  with  $\rho(x_n, x) < 1/n$  and  $f(x_n) \notin N$ . Then  $(x_n)$

converges to  $x$ , but  $(f(x_n))$  does not converge to  $f(x)$ . This contradiction shows that  $f^{-1}(N)$  is a neighborhood of  $x$  for every neighborhood  $N$  of  $f(x)$ , so  $f$  is continuous at  $x$ . ■

**6.39 Definition.** A sequence  $(x_n)$  in a metric space is called a Cauchy sequence if for every  $\epsilon > 0$  there exists  $n_0$  such that  $\rho(x_m, x_n) < \epsilon$  for every  $n \geq n_0$  and  $m \geq n_0$ .

Just as in the special case of the real line, every convergent sequence in a metric space is necessarily Cauchy; just as in the special case of the rational numbers, the converse need not be true.

**6.40 Definition.** A metric space  $X$  is said to be complete if every Cauchy sequence in  $X$  is convergent.

We have seen (Theorem 2.19) that  $\mathbf{R}$  is complete. It follows easily that  $\mathbf{R}^d$  is complete for any positive integer  $d$ . For if  $(\mathbf{x}_n)$  is a Cauchy sequence in  $\mathbf{R}^d$ , with  $\mathbf{x}_n = (x_{1,n}, \dots, x_{d,n})$ , then the inequality  $|x_{j,n} - x_{j,m}| \leq |\mathbf{x}_n - \mathbf{x}_m|$  implies that each sequence  $(x_{j,n})_{n=1}^\infty$  ( $1 \leq j \leq d$ ) is a Cauchy sequence in  $\mathbf{R}$ , hence convergent. If  $\mathbf{x} = (x_1, \dots, x_d)$ , where  $x_j = \lim x_{j,n}$ , then the inequality  $|\mathbf{x} - \mathbf{x}_n| \leq \sqrt{d} \max_j \{|x_j - x_{j,n}|\}$  shows that  $\mathbf{x}_n \rightarrow \mathbf{x}$  as  $n \rightarrow \infty$ .

**6.41 Proposition.** Let  $X$  be a complete metric space. A subspace  $E$  of  $X$  is complete if and only if  $E$  is a closed subset of  $X$ .

**Proof.** Suppose  $E$  is complete. According to Proposition 6.37, for each  $x \in \bar{E}$ , there exists a sequence  $(x_n)$  in  $E$  with  $x_n \rightarrow x$  as  $n \rightarrow \infty$ . Then  $(x_n)$  is a Cauchy sequence in  $E$ , so converges in  $E$ , so  $x \in E$ . Thus  $\bar{E} \subset E$ , so  $E$  is closed.

If  $E$  is closed, then every Cauchy sequence in  $E$  converges to a limit in  $X$  since  $X$  is complete, and this limit is in  $E$  by Proposition 6.37, so  $E$  is complete. ■

This chapter has so far consisted largely of definitions, straightforward examples, and propositions which largely amounted to studying the definitions. We now come to a theorem with some real content; it is not trivial even in the case  $X = \mathbf{R}$ .

**6.42 Theorem.** If  $X$  is a complete metric space, and  $G_n$  is a dense open subset of  $X$  for each  $n \in \mathbf{N}$ , then  $\bigcap_{n=1}^\infty G_n$  is dense.

**Proof.** We must show that for every  $p \in X$ , any open neighborhood  $V$  of  $p$  must contain points of  $\bigcap_{n=1}^\infty G_n$ . We shall do so by constructing for each  $n$  an open ball  $B_n = B(\mathbf{x}_n, \epsilon_n)$  with the properties that  $\bar{B}_1 \subset V \cap G_1$ ,  $\bar{B}_{n+1} \subset G_n \cap B_n$  for every  $n$ , and  $(\epsilon_n)$  decreases to 0. Then, in particular, we have  $B_{n+k} \subset B_n$  for all positive integers  $n$  and  $k$ , so  $\rho(x_n, x_{n+k}) \leq \epsilon_n$ ;

it follows that  $(x_n)$  is a Cauchy sequence, and hence, since  $X$  was assumed complete, that  $(x_n)$  converges to some  $x \in X$ . Now since  $x_k \in \overline{B}_n$  for every  $k \geq n$ , and  $\overline{B}_n$  is closed, it follows from Proposition 6.37 that  $x \in \overline{B}_n$  for every  $n$ , and thus  $x \in G_n \cap B_n$  for every  $n$ , so  $x \in V \cap \bigcap_{n=1}^{\infty} G_n$ , as desired.

To obtain the sequence with these properties, we begin by observing that since  $G_1$  is dense,  $V \cap G_1 \neq \emptyset$ , so there exists  $x_1 \in V \cap G_1$ . Since  $V \cap G_1$  is open, there exists  $\epsilon_1$  such that  $B(x_1, 2\epsilon_1) \subset V \cap G_1$ . We put  $B_1 = B(x_1, \epsilon_1)$ . Then  $\overline{B}_1 \subset \{y \in X : \rho(y, x_1) \leq \epsilon_1\} \subset B(x_1, 2\epsilon_1) \subset V \cap G_1$ . Suppose that  $B_1, \dots, B_n$  have been found, with  $B_j = B(x_j, \epsilon_j)$ , such that  $\overline{B}_1 \subset V \cap G_1$  and (if  $n > 1$ )  $\overline{B}_j \subset B_{j-1} \cap G_j$  for  $j = 2, \dots, n$ , and such that  $\epsilon_j \leq \frac{1}{2}\epsilon_{j-1}$ . Then since  $G_{n+1}$  is dense in  $X$ , there exists  $x_{n+1} \in G_{n+1} \cap B_n$ , and since  $G_{n+1}$  is open, there exists  $\epsilon_{n+1} > 0$  such that  $B(x_{n+1}, 2\epsilon_{n+1}) \subset G_{n+1} \cap B_n$ . Then  $\epsilon_{n+1} \leq \frac{1}{2}\epsilon_n$ , and putting  $B_{n+1} = B(x_{n+1}, \epsilon_{n+1})$  we have

$$\overline{B}_{n+1} \subset \{y : \rho(y, x_{n+1}) \leq \epsilon_{n+1}\} \subset B(x_{n+1}, 2\epsilon_{n+1}) \subset G_{n+1} \cap B_n,$$

so the construction can be carried on. ■

This theorem is known as the Baire category theorem. It has many applications in analysis, but for now we give only the following:

**6.43 Theorem.** *Let  $(f_k)$  be a sequence of continuous real-valued functions on the complete metric space  $X$ . If  $(f_k(x))$  is a bounded sequence for each  $x \in X$ , then there exists a nonempty open subset  $V$  of  $X$  in which the sequence  $(f_k)$  is uniformly bounded, i.e., there exists  $M$  with  $|f_k(x)| \leq M$  for every  $k$  and every  $x \in V$ .*

**Proof.** Let  $U_{k,n} = \{x \in X : |f_k(x)| > n\}$  for each pair of positive integers  $k, n$ . Since  $f_k$  is continuous,  $U_{k,n}$  is open for every  $k, n$ , and hence  $G_n = \bigcup_{k=1}^{\infty} U_{k,n}$  is open. Now for each  $x \in X$ , there exists by hypothesis some  $M_x$  such that  $|f_k(x)| \leq M_x$  for every  $k$ , so  $x \notin G_n$  for  $n \geq M_x$ . Thus  $\bigcap_{n=1}^{\infty} G_n = \emptyset$ . Then Theorem 6.42 tells us that some  $G_n$  is not dense in  $X$ , so there exists an open subset  $V$  of  $X$ , disjoint from  $G_n$ . But this means that  $|f_k(x)| \leq n$  for every  $k$ , and every  $x \in V$ . ■

Here is one more major result about complete metric spaces.

**6.44 Theorem.** *Let  $X$  be a complete metric space, and suppose that  $f : X \rightarrow X$  has the property that there exists  $\alpha < 1$  such that*

$$\rho(f(x), f(y)) \leq \alpha \rho(x, y)$$

*for every  $x, y \in X$ . Then there exists a unique point  $x \in X$  such that  $f(x) = x$ . If  $x_0 \in X$ , and  $x_{n+1} = f(x_n)$  for every  $n \geq 0$ , then  $x = \lim_{n \rightarrow \infty} x_n$ .*

**Proof.** We begin by observing that if  $f(x) = x$  and  $f(y) = y$ , then

$$\rho(x, y) = \rho(f(x), f(y)) \leq \alpha \rho(x, y),$$

which implies  $\rho(x, y) = 0$ , so  $y = x$ . Thus uniqueness of a fixed point for  $f$  is proved.

We turn now to the existence. Choose any  $x_0 \in X$ . Define the sequence  $(x_n)$  inductively by  $x_{n+1} = f(x_n)$ . Then

$$\rho(x_n, x_{n+1}) = \rho(f(x_{n-1}), f(x_n)) \leq \alpha \rho(x_{n-1}, x_n)$$

for all  $n \geq 1$ , and it follows by induction that  $\rho(x_n, x_{n+1}) \leq \alpha^n \rho(x_0, x_1)$ . Let  $C = \rho(x_0, x_1)$ . Then for any positive integers  $n$  and  $k$  we have from the triangle inequality (polygon inequality?) that

$$\rho(x_n, x_{n+k}) \leq \sum_{j=1}^k \rho(x_{n+j-1}, x_{n+j}) \leq \sum_{j=1}^k \alpha^{n+j-1} C \leq \frac{C\alpha^n}{1-\alpha}.$$

Thus  $(x_n)$  is a Cauchy sequence in  $X$ , and since  $X$  is complete, there exists  $x$  such that  $x_n \rightarrow x$  as  $n \rightarrow \infty$ . Since  $f$  is continuous, we have  $f(x) = \lim f(x_n) = \lim x_{n+1} = x$ , and the existence of a fixed point for  $f$  has been shown. ■

A map with the above property is called a contraction map, and Theorem 6.44 is usually referred to as the *contraction map principle*. Here is a simple example of how it can be used.

**6.45 Example.** Let  $I$  be a closed interval in  $\mathbf{R}$  (not necessarily bounded), and let  $f$  be a differentiable real-valued function on  $I$ , with  $f(I) \subset I$ . Suppose that  $|f'(t)| \leq \alpha$  for all  $t \in I$ , where  $\alpha < 1$ . Then the equation  $f(x) = x$  has a unique solution in  $I$ , given by  $x = \lim x_n$ , where  $x_0$  is any point of  $I$  and  $x_{n+1} = f(x_n)$  for every  $n \geq 0$ . For the mean value theorem tells us that  $|f(s) - f(t)| = |f'(\xi)(s - t)|$  for some  $\xi$  between  $s$  and  $t$ , and hence that  $|f(s) - f(t)| \leq \alpha|s - t|$  for all  $s, t \in I$ . Now we apply Theorem 6.44.

Here is a more interesting example, which involves an infinite-dimensional metric space.

**6.46 Theorem.** Let  $f : G \rightarrow \mathbf{R}$  be a bounded continuous real-valued function on an open subset  $G$  of  $\mathbf{R}^2$ , and suppose that  $f$  satisfies a Lipschitz condition with respect to the second variable, i.e., that there exists a constant  $M$  such that

$$|f(x, y_1) - f(x, y_2)| \leq M|y_1 - y_2|$$

whenever  $(x, y_1)$  and  $(x, y_2)$  are points of  $G$ . Then for any  $(x_0, y_0) \in G$ , the differential equation  $y' = f(x, y)$ , with the initial condition  $y(x_0) = y_0$ , has a unique solution in some interval  $[x_0 - \delta, x_0 + \delta]$ . In other words, there exists  $\delta > 0$ , and  $\phi : [x_0 - \delta, x_0 + \delta] \rightarrow \mathbf{R}$ , such that  $\phi(x_0) = y_0$  and  $\phi'(x) = f(x, \phi(x))$  for all  $x$  with  $|x - x_0| \leq \delta$ .

**Proof.** We aim to produce  $\delta > 0$  and  $\phi$  continuous on  $I = [x_0 - \delta, x_0 + \delta]$  such that

$$\phi(x) = y_0 + \int_{x_0}^x f(t, \phi(t)) dt$$

for all  $x \in I$ . By Theorem 5.30, we will have  $\phi'(x) = f(x, \phi(x))$ , and evidently  $\phi(x_0) = y_0$ .

Since  $f$  is bounded, there exists  $K$  such that  $|f(x, y)| \leq K$  for all  $(x, y) \in G$ . Now choose  $\delta > 0$  such that  $M\delta < 1$ , and such that

$$\{(x, y) : |x - x_0| \leq \delta, |y - y_0| \leq K\delta\} \subset G.$$

This is possible since  $G$  is open. Let  $I = [x_0 - \delta, x_0 + \delta]$ .

Let  $X$  be the space of all continuous functions  $g : I \rightarrow [y_0 - K\delta, y_0 + K\delta]$ , and define the metric  $\rho$  on  $X$  by  $\rho(g, h) = \max_{t \in I} |g(t) - h(t)|$ . (By Theorem 3.14,  $\rho$  is well-defined.) It is easy to see that  $\rho$  is a metric on  $X$ , and Proposition 3.23 and Theorem 3.24 show that  $X$  is complete. For  $g \in X$ , define the function  $Tg$  by

$$(Tg)(x) = y_0 + \int_{x_0}^x f(t, g(t)) dt$$

for  $x \in I$ . Then  $Tg$  is continuous (in fact, differentiable) on  $I$  (Theorem 5.30), and

$$|(Tg)(x) - y_0| = \left| \int_{x_0}^x f(t, g(t)) dt \right| \leq K|x - x_0| \leq K\delta,$$

so  $T$  maps  $X$  into  $X$ . Furthermore, we have for any  $g, h \in X$ , and any  $x \in I$ ,

$$|(Tg)(x) - (Th)(x)| \leq \left| \int_{x_0}^x [f(t, g(t)) - f(t, h(t))] dt \right| \leq M\delta\rho(g, h),$$

so  $\rho(Tg, Th) \leq M\delta\rho(g, h)$ . Since  $M\delta = \alpha < 1$ , the map  $T$  is a contraction map of  $X$  into itself, and Theorem 6.44 guarantees the existence of a unique  $\phi \in X$  such that  $T\phi = \phi$ . But if  $\phi$  is any function satisfying  $\phi'(x) = f(x, \phi(x))$  for all  $x \in I$ , then  $\phi'$  is continuous, so

$$\phi(x) - \phi(x_0) = \int_{x_0}^x \phi'(t) dt = \int_{x_0}^x f(t, \phi(t)) dt$$

for all  $x \in I$ . Hence any solution of our differential equation with initial condition satisfies

$$|\phi(x) - y_0| = \left| \int_{x_0}^x f(t, \phi(t)) dt \right| \leq K|x - x_0| \leq K\delta$$

for all  $x \in I$ , so  $\phi \in X$ . Thus the solution found is the unique solution on  $I$ . ■

## 6.6 Compactness

**6.47 Definition.** Let  $X$  be a topological space,  $E \subset X$ . An open cover of  $E$  is a collection  $\mathcal{U} = \{U_\alpha : \alpha \in A\}$  of open subsets of  $X$  such that  $E \subset \bigcup_{\alpha \in A} U_\alpha$ . If  $B \subset A$ , the collection  $\mathcal{V} = \{U_\alpha : \alpha \in B\}$  is called a subcover of  $\mathcal{U}$  if it is itself a cover.

**6.48 Definition.** Let  $X$  be a topological space. A subset  $K$  of  $X$  is called compact if every open cover of  $K$  has a finite subcover.

It is clear that any finite subset of  $X$  is compact; if  $X$  has the discrete topology, every compact set is finite. The next theorem, known as the Heine-Borel theorem, gives a more interesting example.

**6.49 Theorem.** A closed bounded interval in  $\mathbf{R}$  is compact.

**Proof.** Suppose  $I = [a, b]$ , where  $-\infty < a < b < +\infty$ , and suppose that  $\mathcal{U} = \{U_\alpha : \alpha \in A\}$  is an open cover of  $I$ . Let

$$E = \{x \in I : [a, x] \subset \bigcup_{j=1}^n U_{\alpha_j} \text{ for some } \{\alpha_1, \dots, \alpha_n\} \subset A\}.$$

We want to show that  $b \in E$ . Clearly,  $a \in E$  (there exists  $\alpha \in A$  such that  $a \in U_\alpha$ ). Let  $c = \sup E$ , so  $a \leq c \leq b$ . Now there exists  $\beta$  such that  $c \in U_\beta$ , and since  $U_\beta$  is open, there exists  $\epsilon > 0$  such that  $U_\beta \supset (c - \epsilon, c + \epsilon)$ . There exists  $x \in E$  with  $x > c - \epsilon$ , since  $c$  is the least upper bound of  $E$ , so there exist  $\alpha_1, \dots, \alpha_n$  such that  $[a, x] \subset \bigcup_{k=1}^n U_{\alpha_k}$ . Then we have  $[a, c + \epsilon/2] \subset U_\beta \cup \bigcup_{k=1}^n U_{\alpha_k}$ . If  $c < b$ , this contradicts the fact that  $c$  is an upper bound of  $E$ . If  $c = b$ , this shows that  $b \in E$ . ■

We will give another proof of this important result, and expand our repertory of compact spaces, later in this section. The next result is trivial, but points out a great difference in the nature of being compact, as opposed to, for instance, being open, or being closed.

**6.50 Proposition.** A subset of a topological space is compact if and only if it is compact in itself (with the relative topology).

The next result is reminiscent of Theorem 1.27.

**6.51 Proposition.** *If  $X$  is a compact space, and  $(K_n)$  a sequence of nonempty closed subsets of  $X$ , with  $K_{n+1} \subset K_n$  for all  $n$ , then  $\bigcap_{n=1}^{\infty} K_n$  is not empty.*

**Proof.** Let  $U_n = X \setminus K_n$ . Then  $\bigcup U_n = X \setminus \bigcap K_n$ , so if  $\bigcap K_n = \emptyset$ , then  $\{U_n : n \in \mathbf{N}\}$  is an open cover of  $X$ . Since  $X$  is compact, this implies that there exists  $m$  such that  $\bigcup_{n=1}^m U_n = X$ , which is equivalent to  $\bigcap_{n=1}^m K_n = \emptyset$ . Since  $\bigcap_{n=1}^m K_n = K_m \neq \emptyset$ , this is a contradiction. ■

**6.52 Proposition.** *If  $K$  is a compact subset of the topological space  $X$ , then every closed subset of  $K$  is compact.*

**Proof.** If  $F \subset K$  is closed, then  $F^C$  is open; if  $\mathcal{U}$  is an open cover of  $F$ , then  $\mathcal{V} = \mathcal{U} \cup \{F^C\}$  is an open cover of  $K$ , and a finite subcover of  $\mathcal{V}$  (of  $K$ ) produces a finite subcover of  $\mathcal{U}$  (of  $F$ ). ■

**6.53 Corollary.** *Every closed bounded subset of  $\mathbf{R}$  is compact.*

**6.54 Proposition.** *If  $X$  is a Hausdorff space, then every compact subset of  $X$  is closed.*

**Proof.** Let  $K$  be a compact subset of  $X$ . Since  $X$  is Hausdorff, for each  $x \in K^C$  and each  $y \in K$ , there exist open sets  $U_{xy}$  and  $V_{xy}$  such that  $x \in U_{xy}$ ,  $y \in V_{xy}$ , and  $U_{xy} \cap V_{xy} = \emptyset$ . Then for each  $x$ , the collection  $\{V_{xy} : y \in K\}$  is an open cover of  $K$ , and since  $K$  is compact, there exist  $y_1, \dots, y_n \in K$  such that  $K \subset \bigcup_{j=1}^n V_{xy_j}$ . Let  $U = \bigcap_{j=1}^n U_{xy_j}$ ; then  $U$  is open,  $U \cap K = \emptyset$ , and  $x \in U$ . Thus  $K^C$  is a neighborhood of each  $x \in K^C$ , so  $K^C$  is open, which means  $K$  is closed. ■

**6.55 Corollary.** *A subset  $K$  of  $\mathbf{R}$  is compact if and only if  $K$  is closed and bounded.*

**Proof.** We have already seen that if  $K$  is closed and bounded, then  $K$  is compact. If  $K$  is compact, the last proposition shows that  $K$  is closed. Let  $U_n = (-n, n)$  for each  $n \in \mathbf{N}$ ; then  $K \subset \bigcup_{n=1}^{\infty} U_n$ , so  $K \subset \bigcup_{n=1}^m U_n$  for some  $m$ , but this means  $K \subset U_m$  for some  $m$ ,  $K$  is bounded. ■

The next theorem leads to a generalization of Theorem 3.14.

**6.56 Theorem.** *If  $X$  is a compact space, and if  $f$  is a continuous mapping of  $X$  into a topological space  $Y$ , then  $f(X)$  is compact.*

**Proof.** If  $\mathcal{U}$  is an open cover of  $f(X)$ , then  $f^{-1}(\mathcal{U}) = \{f^{-1}(U) : U \in \mathcal{U}\}$  is an open cover of  $X$ . Since  $X$  is compact, there exist  $U_1, \dots, U_n \in \mathcal{U}$  such that  $X = \bigcup_{j=1}^n f^{-1}(U_j)$ , and it follows that  $f(X) \subset \bigcup_{j=1}^n U_j$ . ■

**6.57 Corollary.** *If  $X$  is a compact space, and if  $f : X \rightarrow \mathbf{R}$  is a continuous function, then  $f$  takes a maximum and a minimum value on  $X$ .*

**Proof.** By Theorem 6.56,  $f(X)$  is a compact subset of  $\mathbf{R}$ . By the last corollary,  $f(X)$  is closed and bounded. Then  $\sup f(X) \in f(X)$  and  $\inf f(X) \in f(X)$ , and this is the assertion of the theorem. ■

In general, the inverse of an injective continuous function need not be continuous. For instance, if  $X = [0, 2\pi)$  and  $Y = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$ , the map  $f$  defined by  $f(t) = (\cos t, \sin t)$  is a continuous bijective map of  $X$  to  $Y$ , but its inverse is not continuous. However, we have the following result:

**6.58 Corollary.** *If  $X$  is a compact space,  $Y$  a Hausdorff space, and  $f : X \rightarrow Y$  is continuous and bijective, then  $f^{-1}$  is continuous, i.e.,  $f$  is a homeomorphism.*

**Proof.** For each closed  $F \subset X$ ,  $F$  is compact by Proposition 6.52, so  $f(F)$  is compact by Theorem 6.56, and hence  $f(F)$  is closed by Proposition 6.54. This says that if  $g = f^{-1}$ , then  $g^{-1}(F)$  is closed for every closed  $F \subset X$ , and thus  $g$  is continuous by Proposition 6.14. ■

The next theorem generalizes Theorem 3.18.

**6.59 Theorem.** *If  $f$  is a continuous map of a compact metric space  $X$  into a metric space  $Y$ , then  $f$  is uniformly continuous.*

**Proof.** Let  $\epsilon > 0$ . For each  $x \in X$ , there exists  $\delta(x) > 0$  such that  $\rho(f(y) - f(x)) < \epsilon/2$  for all  $y \in X$  with  $\rho(x, y) < 2\delta(x)$ . Since  $\{B(x, \delta(x)) : x \in X\}$  is an open cover of  $X$ , and  $X$  is compact, there exist  $x_1, \dots, x_n$  such that  $X = \bigcup_{j=1}^n B_j$ , where we put  $B_j = B(x_j, \delta(x_j))$  for  $j = 1, \dots, n$ . Let  $\delta = \min\{\delta(x_j) : 1 \leq j \leq n\}$ . If  $x$  and  $y$  are points of  $X$  with  $\rho(x, y) < \delta$ , then there exists  $j$  such that  $x \in B_j$ , i.e.,  $\rho(x, x_j) < \delta(x_j)$ . It follows that  $\rho(y, x_j) \leq \rho(y, x) + \rho(x, x_j) < \delta(x_j) + \delta \leq 2\delta(x_j)$ , so  $\rho(f(x), f(x_j)) < \epsilon/2$  and  $\rho(f(y), f(x_j)) < \epsilon/2$ , and hence  $\rho(f(x), f(y)) < \epsilon$ . ■

**6.60 Definition.** *If  $E$  is a subset of a metric space  $X$ , we define the diameter of  $E$  to be  $\text{diam } E = \sup\{\rho(x, y) : x, y \in E\}$ . A subset  $E$  of a metric space  $X$  is said to be bounded if  $\text{diam } E < +\infty$ .*

It is easy to see that  $E$  is bounded if and only if  $E \subset B(a, r)$  for some  $a \in X$  and some  $r > 0$ .

**6.61 Definition.** *A subset  $E$  of a metric space  $X$  is said to be totally bounded if for every  $\epsilon > 0$ , there exists a finite subset  $\{x_1, \dots, x_n\}$  of  $X$  such that  $E \subset \bigcup_{k=1}^n B(x_k, \epsilon)$ .*



A set  $F$  such that  $\bigcup_{x \in F} B(x, \epsilon) \supset E$  is called an  $\epsilon$ -net for  $E$ . Thus,  $E$  is totally bounded if and only if for every  $\epsilon > 0$  there exists a finite  $\epsilon$ -net for  $E$ . In this definition, we get the same meaning if we demand that  $\{x_1, \dots, x_n\} \subset E$  (see the exercises). The next proposition records some basic facts about totally bounded sets; the easy proof is left to the reader.

**6.62 Proposition.** *Any bounded set in  $\mathbf{R}^n$  is totally bounded. If  $E$  is a totally bounded subset of a metric space  $X$ , then*

- (a)  $E$  is bounded;
- (b)  $\bar{E}$ , the closure of  $E$ , is totally bounded; and
- (c) any subset of  $E$  is totally bounded.

We remark that, in general, a bounded subset of a metric space need not be totally bounded. For instance, if  $X$  has the discrete metric, then  $X$  is bounded, in fact,  $\text{diam } X = 1$ , but if  $X$  is an infinite set, then no finite collection of balls of radius 1 or less can cover  $X$ . Perhaps a more interesting example is the closed unit ball in Hilbert space,  $B = \{x \in \mathcal{H} : \|x\| \leq 1\}$ . If  $e_n = (\delta_{nk})_{k=1}^\infty$ , where  $\delta_{nk} = 1$  if  $k = n$  and  $\delta_{nk} = 0$  for  $k \neq n$ , we note that  $e_n \in B$  for every  $n$ , and  $\|e_n - e_m\| = \sqrt{2}$  for all  $m \neq n$ . Thus any ball of radius not greater than  $\frac{1}{2}$  can contain at most one of the elements  $e_n$ , so  $B$  is not totally bounded.

**6.63 Theorem.** *A metric space  $X$  is compact if and only if it is complete and totally bounded.*

**Proof.** Suppose  $X$  is compact. For any  $\epsilon > 0$ ,  $X = \bigcup_{x \in X} B(x, \epsilon)$ , so there exist  $x_1, \dots, x_n$  such that  $X = \bigcup_{k=1}^n B(x_k, \epsilon)$ . Thus  $X$  is totally bounded. Now suppose  $\{x_n\}$  is a Cauchy sequence in  $X$ . Let  $A_n = \{x_k : k \geq n\}$ . Then  $\bar{A}_{n+1} \subset \bar{A}_n$  for every  $n$ , so (Proposition 6.51)  $\bigcap_{n=1}^\infty \bar{A}_n \neq \emptyset$  since  $X$  is compact. Let  $x \in \bigcap \bar{A}_n$ . If  $\epsilon > 0$ , there exists  $n_0$  such that  $d(x_m, x_n) < \epsilon/2$  for all  $m, n \geq n_0$ , and  $B(x, \epsilon/2) \cap A_{n_0} \neq \emptyset$  (since  $x \in \bar{A}_{n_0}$ ), so there exists  $m \geq n_0$  with  $d(x, x_m) < \epsilon/2$ , and it follows that  $d(x, x_n) < \epsilon$  for every  $n \geq n_0$ . Thus  $x_n \rightarrow x$  as  $n \rightarrow \infty$ . We have proved  $X$  is complete.

Next, suppose that  $X$  is complete and totally bounded. Let  $\{U_\alpha : \alpha \in A\}$  be an open cover of  $X$ . Since  $X$  can be expressed as the union of finitely many sets of diameter  $\leq 1$ , if there exists no finite subcollection of  $\{U_\alpha : \alpha \in A\}$  which covers  $X$ , there exists a set  $F_1 \subset X$  with  $\text{diam } F_1 \leq 1$ , such that no finite subcollection of  $\{U_\alpha : \alpha \in A\}$  covers  $F_1$ . We construct inductively a sequence  $\{F_n\}$  of subsets of  $X$  with the properties: (i)  $\text{diam } F_n \leq 1/n$ ; (ii)  $F_{n+1} \subset F_n$  for every  $n$ ; and (iii) no finite subcollection of  $\{U_\alpha : \alpha \in A\}$  covers  $F_n$ . Such  $F_n$  exist since each  $F_n$  is itself totally bounded, hence can be decomposed into a finite number of sets of diameter  $\leq 1/(n+1)$ , one of which can be chosen as  $F_{n+1}$ . Now choose  $x_n \in F_n$  for each  $n$ . If  $m \leq n$  and  $m \leq p$ , then  $x_n, x_p \in F_m$ , so  $d(x_p, x_n) \leq 1/m$ : thus  $\{x_n\}$  is a Cauchy

sequence, so there exists  $x \in X$  such that  $x_n \rightarrow x$  as  $n \rightarrow \infty$ . Now there exists  $\beta$  such that  $x \in U_\beta$ , and since  $U_\beta$  is open, there exists  $\epsilon > 0$  such that  $B(x, \epsilon) \subset U_\beta$ . For some  $m$ , we have  $x_n \in B(x, \epsilon/2)$  for all  $n \geq m$ ; choosing  $n \geq m$  such that also  $n > 2/\epsilon$ , we have  $F_n \subset B(x, \epsilon) \subset U_\beta$ , contradicting that  $F_n$  is not covered by any finite subcollection of  $\{U_\alpha : \alpha \in A\}$ . Thus the assumption that  $\{U_\alpha : \alpha \in A\}$  admits no finite subcover of  $X$  is untenable, and we have proved  $X$  is compact. ■

**6.64 Corollary.** *A subset of  $\mathbf{R}^d$  is compact if and only if it is closed and bounded.*

**Proof.** By Proposition 6.41, a subset  $K$  of  $\mathbf{R}^d$  is complete if and only if it is closed, and as we remarked above, it is totally bounded if and only if it is bounded. ■

Some other terminology: a subset  $E$  of a topological space is called *relatively compact*, or *precompact*, if its closure is compact.

**6.65 Corollary.** *A subset of a complete metric space is relatively compact if and only if it is totally bounded.*

**6.66 Definition.** *A space  $X$  is said to be sequentially compact, or to possess the Bolzano–Weierstrass property, if every sequence in  $X$  has a convergent subsequence.*

**6.67 Theorem.** *A metric space  $X$  is compact if and only if it is sequentially compact.*

**Proof.** Suppose  $X$  is sequentially compact. Then  $X$  is complete, since if a Cauchy sequence has a convergent subsequence, it is easily seen to be itself convergent. To show that  $X$  is totally bounded, let  $\epsilon > 0$ , and suppose there exists no finite  $\epsilon$ -net. Choose  $x_1 \in X$ . If  $B(x_1, \epsilon) = X$  we have constructed a finite  $\epsilon$ -net, so there exists  $x_2 \in X$  such that  $d(x_2, x_1) \geq \epsilon$ . If  $B(x_1, \epsilon) \cup B(x_2, \epsilon) = X$ , we have constructed a finite  $\epsilon$ -net, so there exists  $x_3 \in X$  such that  $d(x_3, x_j) \geq \epsilon$  for  $j = 1, 2$ . In this way, we may construct an infinite sequence  $(x_n)$  such that  $d(x_n, x_j) \geq \epsilon$  whenever  $n \neq j$ . Clearly, such a sequence admits no convergent subsequence. Thus the assumption that there exists no finite  $\epsilon$ -net is untenable, and we have shown that  $X$  is totally bounded. We have shown that  $X$  is complete and totally bounded, so  $X$  is compact.

Now suppose  $X$  is compact. Let  $(x_n)$  be a sequence in  $X$ . Let  $A_n = \{x_n, x_{n+1}, \dots\}$  for each  $n$ . Then  $A_{n+1} \subset A_n$  for every  $n$ , so  $\overline{A_{n+1}} \subset \overline{A_n}$  for every  $n$ ; thus the sets  $(\overline{A_n})$  form a nested sequence of nonempty closed sets. Since  $X$  is compact,  $\bigcap_{n=1}^{\infty} \overline{A_n} \neq \emptyset$ , by Proposition 6.51. If  $x \in \bigcap_{n=1}^{\infty} \overline{A_n}$ , and  $\epsilon > 0$ , then  $B(x, \epsilon) \cap A_n \neq \emptyset$  for every  $n$ . Thus there exists  $n_1 \geq 1$  such that  $d(x_{n_1}, x) < 1$ , there exists  $n_2 > n_1$  such that  $d(x_{n_2}, x) < 1/2$ ,

and proceeding inductively we can find  $n_1 < n_2 < \cdots < n_k < \cdots$  such that  $d(x_{n_k}, x) < 1/k$  for every  $k$ . Thus we have constructed a convergent subsequence of  $(x_n)$ . ■

**6.68 Definition.** If  $X$  and  $Y$  are topological spaces,  $C(X, Y)$  will denote the set of all continuous maps from  $X$  to  $Y$ . We abbreviate  $C(X, \mathbf{R})$  by  $C(X)$ .

**6.69 Proposition.** Let  $X$  be a compact topological space, and let  $Y$  be a metric space. For  $f, g \in C(X, Y)$ , define  $d(f, g) = \sup\{\rho(f(x), g(x)) : x \in X\}$ . Then  $d$  is a metric on  $C(X, Y)$ , and the metric space  $(C(X, Y), d)$  is complete if  $Y$  is complete.

**Proof.** The verification that  $d$  is a metric is routine. Suppose that  $Y$  is complete, and suppose that  $(f_n)$  is a Cauchy sequence in  $C(X, Y)$ . Then for each  $x \in X$ ,  $\rho(f_n(x), f_m(x)) \leq d(f_n, f_m)$ , so  $(f_n(x))$  is a Cauchy sequence in  $Y$ , so there exists some point  $f(x) \in Y$  such that  $f_n(x) \rightarrow f(x)$  as  $n \rightarrow \infty$ . We must check that  $f$  is continuous, and that  $d(f_n, f) \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $x \in X$ , and  $\epsilon > 0$ . There exists  $n$  such that  $\rho(f(x), f_n(x)) < \epsilon/3$ , and a neighborhood  $U$  of  $x$  such that  $\rho(f_n(x), f_n(y)) < \epsilon/3$  for every  $y \in U$ . It follows that for every  $y \in U$ ,

$$\rho(f(x), f(y)) \leq \rho(f(x), f_n(x)) + \rho(f_n(x), f_n(y)) + \rho(f_n(y), f(y)) < \epsilon.$$

Thus  $f \in C(X, Y)$ . Now given  $\epsilon > 0$ , choose  $n_0$  such that  $d(f_n, f_m) < \epsilon$  for all  $n, m \geq n_0$ . Then for every  $x \in X$ , we have

$$\rho(f_n(x), f(x)) = \lim_{m \rightarrow \infty} \rho(f_n(x), f_m(x)) \leq \epsilon$$

for every  $n \geq n_0$ ; this says  $d(f_n, f) \leq \epsilon$ . ■

The metric on  $C(X, Y)$  defined in the last proposition is known as the *uniform metric* on  $C(X, Y)$ , and the associated topology as the *uniform topology*. We note that  $d(f_n, f) \rightarrow 0$  if and only if  $(f_n)$  converges uniformly to  $f$ .

**6.70 Definition.** Let  $X$  be a topological space,  $Y$  a metric space, and  $\mathcal{F}$  a subset of  $C(X, Y)$ . We say that  $\mathcal{F}$  is equicontinuous at  $x \in X$  if for every  $\epsilon > 0$  there exists a neighborhood  $U$  of  $x$  such that  $\rho(f(x), f(y)) < \epsilon$  for every  $y \in U$  and every  $f \in \mathcal{F}$ . We say that  $\mathcal{F}$  is equicontinuous if it is equicontinuous at each point of  $X$ .

**6.71 Theorem.** Let  $X$  be a compact space, and let  $Y$  be a compact metric space. A subset  $\mathcal{F}$  of  $C(X, Y)$  is totally bounded if and only if it is equicontinuous.

**Proof.** Suppose  $\mathcal{F}$  is equicontinuous. Let  $\epsilon > 0$ . For each  $x \in X$ , there exists an open neighborhood  $U_x$  such that  $\rho(f(x), f(y)) < \epsilon/3$  for all  $y \in U_x$  and all  $f \in \mathcal{F}$ . Since  $X$  is compact, there exist  $x_1, \dots, x_n$ , such that  $X = \bigcup_{j=1}^n U_{x_j}$ . Let

$$Z = \{(f(x_1), \dots, f(x_n)) : f \in \mathcal{F}\} \subset Y^n.$$

We give  $Y_n$  the product metric introduced earlier:

$$\tilde{\rho}((y_1, \dots, y_n), (y'_1, \dots, y'_n)) = \max\{\rho(y_k, y'_k) : 1 \leq k \leq n\}$$

and observe that  $Y^n$  is totally bounded; indeed, if  $\{y_1, \dots, y_N\}$  is an  $\epsilon$ -net for  $Y$ , then  $\{(y_{j_1}, \dots, y_{j_n}) : 1 \leq j_k \leq N, 1 \leq k \leq n\}$  is evidently a finite  $\epsilon$ -net for  $Y^n$ . It follows that  $Z$ , as a subset of  $Y^n$ , is totally bounded. Thus there exist  $f_1, \dots, f_m \in \mathcal{F}$  such that for every  $f \in \mathcal{F}$  there exists some  $k$ ,  $1 \leq k \leq m$ , such that  $\rho(f(x_j), f_k(x_j)) < \epsilon/3$  for every  $j$ ,  $1 \leq j \leq n$ . Now for any  $x \in X$ , there exists  $j$  such that  $x \in U_j$ , and we then have

$$\begin{aligned} \rho(f(x), f_k(x)) &\leq \rho(f(x), f(x_j)) + \rho(f(x_j), f_k(x_j)) + \rho(f_k(x_j), f_k(x)) \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

Thus  $\{f_1, \dots, f_m\}$  is a finite  $\epsilon$ -net for  $\mathcal{F}$ ; we have shown that  $\mathcal{F}$  is totally bounded.

Now suppose that  $\mathcal{F}$  is totally bounded, and let  $\epsilon > 0$ . Then there exist  $f_1, \dots, f_n \in \mathcal{F}$  such that for any  $f \in \mathcal{F}$  there exists  $j$ ,  $1 \leq j \leq n$ , with  $d(f, f_j) < \epsilon/3$ . If  $x \in X$ , for each  $j$  there exists a neighborhood  $U_j$  of  $x$  such that  $\rho(f_j(x), f_j(y)) < \epsilon/3$  for all  $y \in U_j$ . Let  $U = \bigcap_{j=1}^n U_j$ ; then  $U$  is a neighborhood of  $x$ , and for any  $f \in \mathcal{F}$ , choosing  $j$  so that  $d(f_j, f) < \epsilon/3$ , we have

$$\rho(f(x), f(y)) \leq \rho(f(x), f_j(x)) + \rho(f_j(x), f_j(y)) + \rho(f_j(y), f(y)) < \epsilon.$$

Thus  $\mathcal{F}$  is equicontinuous at  $x$  for each  $x \in X$ . ■

**6.72 Corollary.** Let  $\mathcal{F}$  be an equicontinuous subset of  $C(X, Y)$ , where  $X$  is compact and  $Y$  is a compact metric space. Then every sequence in  $\mathcal{F}$  has a uniformly convergent subsequence.

**6.73 Corollary.** If  $\mathcal{F}$  is a pointwise bounded and equicontinuous family of continuous real-valued functions on the compact space  $X$ , then every sequence of functions in  $\mathcal{F}$  has a uniformly convergent subsequence.

**Proof.** We need only observe that  $\mathcal{F}$  is uniformly bounded, i.e., that there exists  $M \in \mathbb{R}$  such that  $|f(x)| \leq M$  for every  $x \in X$  and  $f \in \mathcal{F}$ . Then we can regard  $\mathcal{F}$  as a subset of  $C(X, [-M, M])$ , so Corollary 6.73 applies. ■

## 6.7 Connectedness

**6.74 Definition.** A topological space  $X$  is said to be disconnected if there exist disjoint nonempty open sets  $U$  and  $V$  with  $X = U \cup V$ . The space  $X$  is said to be connected if it is not disconnected. A subset  $A$  of a topological space  $X$  is said to be connected if  $A$ , with its relative topology inherited from  $X$ , is connected.

The following proposition, whose proof is left to the reader, gives various alternative ways to express the content of this definition.

**6.75 Proposition.** The topological space  $X$  is connected if and only if it is not the union of two nonempty disjoint closed sets, or equivalently, if and only if there exists no subset of  $X$  which is simultaneously open and closed, other than  $X$  and  $\emptyset$ . The subset  $A$  of  $X$  is connected if and only if for every pair  $U, V$  of open subsets of  $X$  such that  $A \subset U \cup V$  and  $U \cap V \cap A = \emptyset$ , we have either  $A \subset U$  or  $A \subset V$ .

It is obvious that if  $X$  has the trivial topology, every subset of  $X$  is connected, while if  $X$  has the discrete topology, the only connected subsets of  $X$  are the singletons. Let us examine the situation in a more interesting special case.

**6.76 Theorem.** A subset  $A$  of  $\mathbf{R}$  is connected if and only if it is an interval.

**Proof.** Suppose that  $A$  is not an interval, so there exist  $a, b \in A$ , and  $c \notin A$ , with  $a < c < b$ . Let  $U = (-\infty, c)$  and  $V = (c, +\infty)$ . Then  $U$  and  $V$  are open subsets of  $\mathbf{R}$ , with  $A \subset U \cup V$  and  $U \cap V = \emptyset$ . Since  $a \in U \cap A$  and  $b \in V \cap A$ , we see from Proposition 6.75 that  $A$  is disconnected.

Now suppose that  $A$  is an interval, and that there exist open subsets  $U$  and  $V$  of  $\mathbf{R}$  such that  $A \subset U \cup V$ ,  $U \cap V \cap A = \emptyset$ , and neither  $U \cap A$  nor  $V \cap A$  is empty. Let  $a \in U \cap A$  and  $b \in V \cap A$ . We may assume  $a < b$ . Since  $A$  is an interval, for each  $t$  with  $a \leq t \leq b$ , we have  $t \in A$ , and hence either  $t \in U$  or  $t \in V$ . Let  $E = \{t \in [a, b] : t \in U\}$ , and let  $c = \sup E$ . Then  $a \leq c \leq b$ , so  $c \in A$ . If  $c \in V$ , then  $c > a$ , and  $(c - \epsilon, c + \epsilon) \subset V$  for some  $\epsilon > 0$ , since  $V$  is open. But since  $c$  is the least upper bound of  $E$ , there exists  $t \in (c - \epsilon, c]$  with  $t \in U$ . This contradiction shows that  $c \notin V$ . If  $c \in U$ , then  $c < b$ , and  $(c - \epsilon, c + \epsilon) \subset U$  for some  $\epsilon > 0$ , since  $U$  is open. But then there exists  $t > c$  with  $t \in [a, b] \cap U$ , contradicting the fact that  $c$  is an upper bound of  $E$ . Thus our assumption that both  $A \cap U$  and  $A \cap V$  are nonempty is untenable. Thus  $A$  is connected. ■

We have seen that the image of an interval under a continuous real-valued function is again an interval (Theorem 3.15). Here is a generalization of that theorem.

**6.77 Theorem.** Let  $X$  and  $Y$  be topological spaces, and let  $f : X \rightarrow Y$  be continuous. If  $X$  is connected, then  $f(X)$  is connected.

**Proof.** If  $f(X)$  is disconnected, there exist open subsets  $U$  and  $V$  of  $Y$  such that  $f(X) \subset U \cup V$ ,  $U \cap V \cap f(X) = \emptyset$ , and  $U \cap f(X) \neq \emptyset$ ,  $V \cap f(X) \neq \emptyset$ . But then  $f^{-1}(U)$  and  $f^{-1}(V)$  are open since  $f$  is continuous,  $f^{-1}(U) \cup f^{-1}(V) = X$  since  $U \cup V \supset f(X)$ ,  $f^{-1}(U) \cap f^{-1}(V) = \emptyset$  since  $f(X) \cap U \cap V = \emptyset$ , and neither  $f^{-1}(U)$  nor  $f^{-1}(V)$  is empty. Thus  $X$  is disconnected whenever  $f(X)$  is disconnected. ■

If  $E$  and  $F$  are connected sets with a point in common, then  $E \cup F$  is connected. More generally, we have the following:

**6.78 Proposition.** Let  $X$  be a topological space, and suppose that  $E_\alpha$  is a connected subset of  $X$  for every  $\alpha \in A$ . If  $\bigcap_{\alpha \in A} E_\alpha \neq \emptyset$ , then  $\bigcup_{\alpha \in A} E_\alpha$  is connected.

**Proof.** Let  $E = \bigcup_{\alpha \in A} E_\alpha$ , and suppose  $E \subset U \cup V$ , where  $U$  and  $V$  are open subsets of  $X$  with  $E \cap U \cap V = \emptyset$ . By hypothesis, there exists  $c \in \bigcap_{\alpha \in A} E_\alpha$ ; we may assume  $c \in U$ . Now for every  $\alpha \in A$ , we have  $E_\alpha \subset U \cup V$  and  $E_\alpha \cap U \cap V = \emptyset$ ; since  $E_\alpha$  is connected, either  $E_\alpha \cap U = \emptyset$  or  $E_\alpha \cap V = \emptyset$ . Since  $c \in E_\alpha \cap U$ , we conclude  $E_\alpha \subset U$ . Thus  $E \subset U$ . ■

Another simple but useful fact about connected sets:

**6.79 Proposition.** If  $E$  is a connected subset of  $X$ , then  $\overline{E}$  is connected.

**Proof.** If  $\overline{E}$  is not connected, there exist nonempty sets  $F$  and  $C$  closed in  $\overline{E}$  with  $\overline{E} = F \cup C$  and  $F \cap C = \emptyset$ . Since  $\overline{E}$  is closed,  $F$  and  $C$  are closed (in  $X$ ). Then  $E \subset F \cup C$ , and  $E \cap F \cap C = \emptyset$ . Since  $E$  is connected, either  $E \subset F$  or  $E \subset C$ . But then  $\overline{E} \subset F$ , or  $\overline{E} \subset C$ , since  $\overline{E}$  is the smallest closed set containing  $E$ . Thus  $\overline{E}$  is connected. ■

**6.80 Definition.** Let  $X$  be a topological space. For each  $x \in X$ , let  $C_x$  be the union of all the connected subsets of  $X$  which contain  $x$ . Each  $C_x$  is called a component (or connected component) of  $X$ , or the component in  $X$  of the point  $x$ .

**6.81 Proposition.** Let  $X$  be a topological space, and for each  $x \in X$  let  $C_x$  be the connected component of  $x$ . Then:

- (a) for each  $x \in X$ ,  $C_x$  is connected and closed; and
- (b) for any  $x, y \in X$ , either  $C_x = C_y$  or  $C_x \cap C_y = \emptyset$ .

**Proof.** By Proposition 6.78,  $C_x$  is connected, and by Proposition 6.79,  $\overline{C}_x$  is connected; hence, by the definition of  $C_x$ ,  $\overline{C}_x \subset C_x$ , so  $\overline{C}_x = C_x$ , and  $C_x$  is closed. If  $C_x \cap C_y \neq \emptyset$ , then  $C_x \cup C_y$  is connected by Proposition 6.78, so  $C_x \cup C_y \subset C_x$  by the definition of  $C_x$ , so  $C_y \subset C_x$ , and symmetrically  $C_x \subset C_y$ , so  $C_x = C_y$ . ■

If  $X$  is a discrete space, every subset of  $X$  is both open and closed, so  $X$  (if it contains more than one point) is as far from being connected as possible. In particular, the connected component of each  $x \in X$  reduces to  $\{x\}$ . But this property is shared by many nondiscrete spaces.

**6.82 Definition.** The topological space  $X$  is called totally disconnected if the connected component of each  $x \in X$  is  $\{x\}$ ; equivalently, if every connected subset of  $X$  is a singleton set. A subset  $E$  of a topological space is called totally disconnected if it is totally disconnected with the relative topology.

For example, the space  $\mathbf{Q}$  of rationals, with its usual metric topology, is totally disconnected, as is its complement in  $\mathbf{R}$ , the set of irrationals. Of course, a discrete subset of a topological space is totally disconnected. The next example shows that a closed set which is as far from discrete as is possible can still be totally disconnected. The set constructed here is known as the Cantor set.

**6.83 Example.** There exists a closed subset  $K$  of  $[0, 1]$  such that  $K$  is totally disconnected, but every point of  $K$  is a limit point of  $K$ . We construct  $K$  as follows. For any closed bounded interval  $I = [a, b]$ , we define  $I' = [a, (2a+b)/3]$  and  $I'' = [(a+2b)/3, b]$ ; thus  $I'$  and  $I''$  are the two closed intervals remaining after we remove the open middle third of  $[a, b]$ ; each has length  $(b-a)/3$ . Now we define a sequence  $(K_n)_{n=0}^{\infty}$  of closed subsets of  $[0, 1]$  inductively. Let  $K_0 = [0, 1]$ . Suppose we have obtained  $K_0, \dots, K_n$  such that for  $0 \leq j \leq n$ ,  $K_j = \bigcup_{k=1}^{2^j} I_{j,k}$ , where each  $I_{j,k}$  is a closed interval of length  $3^{-j}$ , and  $I_{j,k} \cap I_{j,l} = \emptyset$  for  $k \neq l$ . Then define  $I_{n+1,2k-1} = I'_{n,k}$  and  $I_{n+1,2k} = I''_{n,k}$ , and put  $K_{n+1} = \bigcup_{k=1}^{2^{n+1}} I_{n+1,k}$ . In other words, we obtain  $K_{n+1}$  by removing the open middle third of each of the  $2^n$  closed intervals that make up  $K_n$ . We define  $K = \bigcap_{n=0}^{\infty} K_n$ . Since the intersection of closed sets is closed,  $K$  is closed. Since  $K_n$  contains no interval of length greater than  $3^{-n}$ ,  $K$  contains no interval of positive length, and thus no connected set consisting of more than one point. This also shows that  $K$  has empty interior. Finally, if  $x \in K$ , then for each  $n$ ,  $x \in I_{n,k}$  for some  $k$ ; let  $x_n$  be the right endpoint of  $I_{n,k}$ , unless  $x$  is the right endpoint of  $I_{n,k}$ , in which case let  $x_n$  be the left endpoint. Then  $x_n \in K$  for every  $n$ , and  $0 < |x_n - x| \leq 3^{-n}$ , so  $x$  is a limit point of  $K$ .

We observe finally that  $K$  is uncountable. We give an argument which can be applied to any closed subset of a complete metric space which consists

entirely of limit points. Suppose not, so  $K = \{x_j : j \in \mathbb{N}\}$ . We construct a sequence of open intervals  $(J_n)_{n=1}^\infty$  with the properties  $\bar{J}_{n+1} \subset J_n$  for every  $n$ ,  $J_n \cap K \neq \emptyset$ ,  $\text{diam } J_n \rightarrow 0$ , and  $x_n \notin \bar{J}_{n+1}$ . Let  $J_1$  be any open interval containing  $x_1$ . Having found  $J_n$  with  $J_n \cap K \neq \emptyset$ , there exist at least two points in  $J_n \cap K$ , since each point of  $K$  is a limit point of  $K$ , and hence there exists an open interval  $J_{n+1}$  such that  $\bar{J}_{n+1} \subset J_n$ ,  $J_{n+1} \cap K \neq \emptyset$ ,  $\text{diam } J_{n+1} < \frac{1}{2} \text{diam } J_n$ , and  $x_n \notin \bar{J}_{n+1}$ . Choose  $y_n \in J_n \cap K$ ; then  $|y_n - y_{n+k}| \leq \text{diam } J_n$  since  $y_{n+k} \in J_{n+k} \subset J_n$ , so  $(y_n)$  is a Cauchy sequence, hence converges to some  $y$ . Since  $y_n \in K$  for every  $n$ , and  $K$  is closed, it follows that  $y \in K$ . But since  $x_n \notin \bar{J}_m$  for every  $m > n$ ,  $y \neq x_n$  for every  $n$ . This contradiction shows that  $K$  is not countable. (An exercise describes an alternate proof.)

## 6.8 Exercises

1. Let  $X$  be a metric space, and  $a \in X$ ,  $r > 0$ . Show that  $\overline{B(a, r)} \subset \{x \in X : \rho(x, a) \leq r\}$ , and give an example to show that the inclusion can be proper.
2. Let  $X$  be a topological space. Show that  $X \setminus E^\circ = \overline{X \setminus E}$ , for any  $E \subset X$ .
3. Show that if the topological space  $X$  has a countable base, then every base contains a countable base.
4. A Hausdorff space  $X$  is called a *door space* if every subset of  $X$  is either open or closed. Show that if  $X$  is a door space, then  $X$  has at most one limit point. Show that if  $x \in X$  is not a limit point, then  $\{x\}$  is open.
5. Let  $X$  be a topological space. For each  $E \subset X$ , the set of limit points of  $E$  is called the *derived set* of  $E$ , and denoted  $E'$ . Show that for any  $E \subset X$ ,  $E'$  is closed, and show that  $E' = (\bar{E})'$ .
6. Let  $X$  be a topological space and  $E \subset X$ . We say that  $x \in X$  is a *condensation point* of  $E$  if every neighborhood of  $x$  contains uncountably many points of  $E$ . Suppose that  $X$  has a countable base, and that  $E$  is an uncountable subset of  $X$ , and let  $C$  be the set of all condensation points of  $E$ . Show that  $C$  is closed, that every point of  $C$  is a limit point of  $C$ , and that  $E \setminus C$  is countable. HINT: Let  $\{U_n\}$  be a countable base for  $X$ , let  $A = \{n : U_n \cap E \text{ is countable}\}$ , and let  $G = \bigcup_{n \in A} U_n$ . Show that  $C = X \setminus G$ .
7. Let  $f$  be a continuous real-valued function on a topological space  $X$ . Show that the zero set of  $f$ ,  $Z(f) = \{x \in X : f(x) = 0\}$ , is closed.
8. Let  $X$  be a topological space,  $Y$  a Hausdorff space, and let  $f : X \rightarrow Y$  and  $g : X \rightarrow Y$  be continuous. Show that  $\{x \in X : f(x) = g(x)\}$  is closed. Hence if  $f(x) = g(x)$  for all  $x$  in a dense subset of  $X$ , then  $f = g$ .



9. Let  $X$  be a metric space. If  $x \in X$  and  $E \subset X$ , define the distance from  $x$  to  $E$  as  $\rho(x, E) = \inf\{\rho(x, y) : y \in E\}$ .

- Show that the function  $x \mapsto \rho(x, E)$  is uniformly continuous on  $X$ , for each  $E \subset X$ , and that  $\overline{E} = \{x : \rho(x, E) = 0\}$ .
- Deduce from (a) that if  $F$  is a closed subset of  $X$ , there exist open sets  $G_n$  ( $n \in \mathbb{N}$ ) such that  $F = \bigcap_{n=1}^{\infty} G_n$ .
- If  $E$  and  $F$  are disjoint closed subsets of  $X$ , show that  $f$ , defined by

$$f(x) = \frac{\rho(x, E)}{\rho(x, E) + \rho(x, F)},$$

is a continuous real-valued function on  $X$ , with  $0 \leq f \leq 1$ , and  $E = f^{-1}(\{0\})$ ,  $F = f^{-1}(\{1\})$ . Deduce that there exist disjoint open sets  $U$  and  $V$  with  $E \subset U$  and  $F \subset V$ .

10. Let  $X$  be a topological space, and let  $f : X \rightarrow \mathbb{R}$ . Show that the set  $C$  of all points  $x \in X$  such that  $f$  is continuous at  $x$  is the intersection of a countable family of open subsets of  $X$ .

11. Let  $X$  and  $Y$  be topological spaces, and let  $f : X \rightarrow Y$ . Show that  $f$  is continuous if and only if  $f(\overline{E}) \subset \overline{f(E)}$  for every  $E \subset X$ , if and only if  $f^{-1}(\overline{E}) \subset \overline{f^{-1}(E)}$  for every  $E \subset Y$ .

12. Let  $X$  be a topological space, and let  $f : X \rightarrow [-\infty, +\infty)$ . We say that  $f$  is *upper semicontinuous*, abbreviated u.s.c., if  $f$  is continuous when  $\mathbb{R} \cup \{-\infty\}$  is given the topology  $\mathcal{T}_u$  of Example 6.4. Similarly, a mapping  $f$  of  $X$  into  $\mathbb{R} \cup \{+\infty\}$  is called *lower semicontinuous*, or l.s.c., if it is continuous when the target space has the topology  $\mathcal{T}_l$ .

- Show that  $f : X \rightarrow \mathbb{R}$  is continuous ( $\mathbb{R}$  having its usual topology) if and only if  $f$  is both u.s.c. and l.s.c.
- Show that if  $f$  and  $g$  are u.s.c., then so are  $f + g$  and  $\lambda f$ , for any real  $\lambda > 0$ .
- Show that if  $\mathcal{F}$  is any collection of u.s.c. functions on  $X$ , and  $g$  is defined by  $g(x) = \inf\{f(x) : f \in \mathcal{F}\}$ , then  $g$  is u.s.c.
- Show that  $f$  is l.s.c. if and only if  $-f$  is u.s.c., and deduce the analogues of (b) and (c) for l.s.c. functions.

13. A collection  $\mathcal{F}$  of subsets of a set  $X$  is said to have the *finite intersection property* if  $\bigcap_{j=1}^n F_j \neq \emptyset$  for any finite subcollection  $\{F_1, F_2, \dots, F_n\}$  of  $\mathcal{F}$ . Show that a topological space  $X$  is compact if and only if for every collection  $\{F_\alpha : \alpha \in A\}$  of closed subsets of  $X$  which has the finite intersection property, we have  $\bigcap_{\alpha \in A} F_\alpha \neq \emptyset$ .

14. Show that if  $E$  is a totally bounded subset of a metric space  $X$ , then for every  $\epsilon > 0$  there exists a finite subset  $\{x_1, \dots, x_n\}$  of  $E$  such that  $E \subset \bigcup_{k=1}^n B(x_k, \epsilon)$ .

15. Let  $(X, \mathcal{T})$  be a compact Hausdorff space. Let  $\mathcal{T}'$  be a topology on  $X$  which is strictly stronger than  $\mathcal{T}$ , and let  $\mathcal{T}''$  be a topology on  $X$  which is strictly weaker than  $\mathcal{T}$ . Show that  $(X, \mathcal{T}')$  is Hausdorff but not compact, while  $(X, \mathcal{T}'')$  is compact, but not Hausdorff.

16. Let  $X$  be a compact Hausdorff space, and let  $f : X \rightarrow \mathbf{R}$ . The *graph* of  $f$  is the set  $G(f) = \{(x, f(x)) : x \in X\} \subset X \times \mathbf{R}$ . Show that  $f$  is continuous if and only if  $G(f)$  is compact.

17. Let  $X$  be a compact metric space, and  $f : X \rightarrow X$  an isometry, i.e.,  $\rho(f(x), f(y)) = \rho(x, y)$  for every  $x, y \in X$ . Show that  $f$  is bijective. HINT: Let  $Y = f(X)$ , and suppose  $X \setminus Y \neq \emptyset$ . Let  $x_0 \in X \setminus Y$ , and let  $\delta = \rho(x_0, Y)$ . Define the sequence  $(x_n)$  inductively by  $x_{n+1} = f(x_n)$  for every  $n \geq 0$ . Show that  $\rho(x_n, x_m) \geq \delta$  for all  $m < n$ .

18. Show that an upper semicontinuous function on a compact space assumes a maximum value.

19. Prove Proposition 6.75.

20. A topological space  $X$  is called *pathwise connected* if for every  $x, y \in X$  there exists a continuous map  $\gamma : [0, 1] \rightarrow X$  with  $\gamma(0) = x$  and  $\gamma(1) = y$ . Show that every pathwise connected space is connected. Show that the converse is false, by considering the following subspace of  $\mathbf{R}^2$ :

$$X = \{(t, \sin(1/t)) : t \neq 0\} \cup \{(0, t) : -1 \leq t \leq 1\}.$$

21. A topological space  $X$  is called *locally connected* if for each  $x \in X$ , any neighborhood of  $x$  contains an open connected neighborhood of  $x$ . (Equivalently,  $X$  has a base of connected sets.) Show that if  $X$  is locally connected, and  $G$  is an open subset of  $X$ , then every component of  $G$  is open.

22. Show that every open subset of  $\mathbf{R}$  is the union of a disjoint sequence of open intervals.

23. A subset  $C$  of  $\mathbf{R}^d$  is called *convex* if it has the following property: for every  $\mathbf{x}$  and  $\mathbf{y}$  in  $C$ , and every real  $t$  with  $0 \leq t \leq 1$ , we have  $t\mathbf{x} + (1-t)\mathbf{y} \in C$ . Show that the intersection of an arbitrary collection of convex sets is convex. Show that a convex subset of  $\mathbf{R}^d$  is connected.

24. Let  $X = \{(x, y) \in \mathbf{R}^2 : \text{either } x \text{ or } y \text{ is irrational}\}$ . Show that  $X$  is connected.

25. Show that the Cantor set  $K$  can be described as the set of all  $x \in [0, 1]$  such that  $x = \sum_{n=1}^{\infty} a_n 3^{-n}$ , where  $a_n \in \{0, 2\}$  for every  $n$ . Deduce that  $K$  is uncountable.

## 6.9 Notes

- 6.1 The word topology in the title of this chapter is perhaps misleading, in that it does not refer to the subject of study of topologists. Rather it is the framework in which analysts study the notion of convergence and continuity. The points of the topological spaces or metric spaces that we are concerned with are likely to be functions. The subject is called general topology, or point set topology. The concepts of neighborhood and limit point were introduced by Cantor in his paper of 1872. The basic role of open sets in the study of continuity and convergence in general settings emerged only gradually.
- 6.2 Hausdorff in 1914, in the first book on general topology, defined continuity in terms of the concept of open set.
- 6.3 Metric spaces were introduced by Fréchet in 1916.
- 6.4 The topological product of infinitely many topological spaces was introduced by Tychonoff in 1930; he proved the eponymous theorem that the product of any collection of compact spaces is again compact.
- 6.5 For metric spaces, all the topological notions can be defined in terms of sequences, but this is decidedly not the case for general topological spaces. A concept called *nets* or *generalized sequences*, however, does suffice, and has the advantage of enabling us to use the intuition we have developed for sequences. See the classic book of Kelley [6] for more on this. The important role that completeness plays was first put into evidence by Cauchy. Theorem 6.42 was proved by Baire for  $\mathbf{R}^n$  in his 1899 thesis; it had previously been proved for  $\mathbf{R}$  by Osgood, who proved Theorem 6.43.
- 6.6 Bolzano stated the result that every infinite subset of a bounded interval in  $\mathbf{R}$  has a limit point, but apparently never wrote down a proof. Weierstrass independently found this result, in the form that every bounded sequence in  $\mathbf{R}$  has a convergent subsequence. The word "compact" meant "sequentially compact" until almost the middle of this century; what we now call "compact" (the Heine–Borel property) was introduced as "bicomcompact" by Alexandroff and Urysohn in 1924, and eventually took over as the dominant notion in the period when attention was being paid more and more to topological spaces,

rather than metric spaces. Many of the results in this section were obtained by Alexandroff in the 1920s. In topological spaces, compactness and sequential compactness are distinct notions; neither implies the other. Hausdorff introduced the notion of total boundedness, and proved Theorem 6.63. Total boundedness is surely a more intuitive idea than sequential compactness, which in turn is easier to grasp than compactness; however, compactness has proved its utility, and become a central notion. Theorem 6.71 (actually, Corollary 6.73) is due to Arzelà and Ascoli, who found it independently in the 1880s.

- 6.7 The notion of connectedness is perhaps the most intuitive of all the basic ideas of topology, certainly far more intuitive than compactness. Yet there are surprising facts to be discovered about it, even at an elementary level. For instance, Sierpiński constructed a subset of  $\mathbf{R}^2$  which is connected, but becomes totally disconnected when one point is removed.

# Function Spaces

In this chapter, we give a few applications of the results obtained in the preceding chapters, especially the last chapter. The common ground is that we are considering spaces whose points are functions, and functions on such spaces.

## 7.1 The Weierstrass Polynomial Approximation Theorem

Continuous real-valued functions on an interval can be quite nasty, for instance, nowhere differentiable, but polynomial functions are as pleasant as possible: their values are easily computable, they can be easily differentiated and integrated, they vanish at only a finite number of points, etc. Therefore, it is a very pleasant fact that any continuous real-valued function on a closed bounded interval in  $\mathbf{R}$  can be uniformly approximated by polynomials. The following theorem is known as the Weierstrass polynomial approximation theorem.

**7.1 Theorem.** *The polynomial functions are dense in  $C([a, b])$ , with its uniform topology.*

**Proof.** The assertion of the theorem is that for any continuous real-valued function  $f$  on the closed bounded interval  $[a, b]$ , and any  $\epsilon > 0$ , there exists a polynomial function  $p$  such that  $|p(x) - f(x)| < \epsilon$  for every  $x \in [a, b]$ ; equivalently, there exists a sequence  $(p_n)$  of polynomials which converges

uniformly to  $f$  on  $[a, b]$ . We begin by proving a special case. We denote the identity function by  $x$ ; thus  $x(t) = t$  for  $t \in \mathbf{R}$ .

**7.2 Lemma.** *There exists a sequence  $(q_n)$  of polynomials which converges uniformly to  $|x|$  on  $[-1, 1]$ .*

**Proof.** Let  $q_0 = 1$ , and inductively define  $q_{n+1} = \frac{1}{2}[x^2 + 2q_n - q_n^2]$  for  $n \geq 0$ . Clearly, every  $q_n$  is a polynomial. We note that if  $|x| \leq q_n \leq 1$ , then

$$q_n - q_{n+1} = \frac{1}{2}(2q_n - x^2 - 2q_n + q_n^2) = \frac{1}{2}(q_n^2 - x^2) \geq 0,$$

and

$$q_{n+1} - |x| = \frac{1}{2}(x^2 - 2|x| + 2q_n - q_n^2) = \frac{1}{2}[(1 - |x|)^2 - (1 - q_n)^2],$$

so if  $|x| \leq q_n \leq 1$ , we have  $|x| \leq q_{n+1} \leq q_n \leq 1$ . Since  $q_0 = 1$ , it follows by induction that  $|x| \leq q_{n+1} \leq q_n$  for all  $n$ . Hence  $(q_n(t))$  converges for every  $t \in [-1, 1]$ ; if  $q = \lim q_n$ , we have  $q = \frac{1}{2}[x^2 + 2q - q^2]$ , and thus  $q^2 = x^2$ . Since  $q \geq 0$ , we have  $q = |x|$ . Since the continuous functions  $q_n$  decrease to the continuous function  $|x|$ , we know by Dini's theorem (Theorem 3.25) that the convergence is uniform. ■

**7.3 Lemma.** *For any  $c \in \mathbf{R}$ , there exists a sequence  $(p_n)$  of polynomials which converges to  $|x - c|$  uniformly on every compact subset of  $\mathbf{R}$ .*

**Proof.** By Lemma 7.2, we can find for each  $n \in \mathbf{N}$  a polynomial  $Q_n$  such that  $|Q_n(t) - |t|| < 1/n^2$  for all  $t \in [-1, 1]$ . Let  $p_n(t) = nQ_n((t - c)/n)$ ; then  $|p_n(t) - |t - c|| < 1/n$  for all  $n > |t| + |c|$ . ■

Returning to the proof of Theorem 7.1, let  $A$  be the set of all functions  $f : \mathbf{R} \rightarrow \mathbf{R}$  with the property that for any  $\epsilon > 0$  there exists a polynomial  $p$  such that  $|f(t) - p(t)| < \epsilon$  for all  $t \in [a, b]$ . It is quite easy to see that if  $f, g \in A$  and  $a, b \in \mathbf{R}$ , then  $af + bg \in A$  (in other words,  $A$  is a vector space over  $\mathbf{R}$ ). Since  $u^+ = \frac{1}{2}(u + |u|)$  for any  $u \in \mathbf{R}$ , it follows from Lemma 7.3 that the function  $(x - c)^+$  belongs to  $A$  for any  $c \in \mathbf{R}$ , and hence that any function of the form

$$g = A + \sum_{j=1}^n m_j (x - c_j)^+ \quad (7.1)$$

belongs to  $A$ . But any piecewise linear continuous function on  $[a, b]$  can be expressed in the form (7.1). Indeed, if  $g$  is piecewise linear on  $[a, b]$ , there exists a partition  $(c_j)_{j=0}^n$  of  $[a, b]$ , and  $m_j \in \mathbf{R}$  for  $j = 1, 2, \dots, n$ , such that  $g(t) = f(c_{j-1}) + m_j(t - c_{j-1})$  for all  $t \in [c_{j-1}, c_j]$ ,  $1 \leq j \leq n$ . But then we have

$$g(t) = f(a) + \sum_{j=1}^n m_j (t - c_{j-1})^+$$

for all  $t \in [a, b]$ , i.e.,  $g$  has the form (7.1) with  $A = f(a)$ . Now given any  $f \in C([a, b])$ , there exists a piecewise linear continuous function  $g$  such that  $|g(t) - f(t)| < \epsilon/2$  for all  $t \in [a, b]$  (Theorem 3.20), and, by what we have just proved, a polynomial  $p$  such that  $|p(t) - g(t)| < \epsilon/2$  for all  $t \in [a, b]$ ; then we have  $|f(t) - p(t)| < \epsilon$  for all  $t \in [a, b]$ . ■

We now turn to the space of all real continuous functions on a compact space, and establish some sufficient conditions for a subset of this space of functions to be dense. For the proof of the next theorem, it is convenient to introduce some new notation.

**7.4 Definition.** If  $a, b \in \mathbf{R}$ , we put  $a \vee b = \max\{a, b\}$  and  $a \wedge b = \min\{a, b\}$ . If  $f, g : X \rightarrow \mathbf{R}$ , we define  $f \vee g : X \rightarrow \mathbf{R}$  by  $(f \vee g)(x) = f(x) \vee g(x)$ ,  $x \in X$ , and  $f \wedge g : X \rightarrow \mathbf{R}$  by  $(f \wedge g)(x) = f(x) \wedge g(x)$ ,  $x \in X$ .

**7.5 Theorem.** Let  $X$  be a compact topological space, and let  $\mathcal{L}$  be a subset of  $C(X)$  having the following properties:

- (a) if  $f, g \in \mathcal{L}$ , and  $a, b \in \mathbf{R}$ , then  $af + bg \in \mathcal{L}$ ;
- (b) if  $f, g \in \mathcal{L}$ , then  $f \vee g \in \mathcal{L}$  and  $f \wedge g \in \mathcal{L}$ ;
- (c) for any  $x, y \in X$ , with  $x \neq y$ , there exists  $f \in \mathcal{L}$  with  $f(x) \neq f(y)$ ; and
- (d) each constant function belongs to  $\mathcal{L}$ .

Then  $\mathcal{L}$  is dense in  $C(X)$ .

**Proof.** We remark that condition (a) says that  $\mathcal{L}$  is a vector space, with the usual operations on functions; condition (b) is often described with the words “ $\mathcal{L}$  is a lattice”; condition (c) is described by “ $\mathcal{L}$  separates points.” Thus the theorem says that any vector lattice contained in  $C(X)$  which separates the points of  $X$  and contains the constants, is necessarily dense in  $C(X)$ .

We observe that conditions (a), (c), and (d) imply the following: if  $x, y \in X$  with  $x \neq y$ , and if  $a, b \in \mathbf{R}$ , then there exists  $f \in \mathcal{L}$  such that  $f(x) = a$  and  $f(y) = b$ . For by (c) there exists  $g \in \mathcal{L}$  with  $g(x) = \alpha$  and  $g(y) = \beta$ , where  $\alpha \neq \beta$ . By (a) and (d), for any  $s, t \in \mathbf{R}$  we have  $sg + t \in \mathcal{L}$ . It is trivial to choose  $s$  and  $t$  such that  $f = sg + t$  satisfies  $f(x) = a$  and  $f(y) = b$ .

Now to the proof of the theorem. Let  $f \in C(X)$  and  $\epsilon > 0$ . We must produce  $g \in \mathcal{L}$  with  $\|f - g\| < \epsilon$ , i.e., with  $f(x) - \epsilon < g(x) < f(x) + \epsilon$  for every  $x \in X$ . We proceed as follows. For each  $x, y \in X$ , there exists  $g_{xy} \in \mathcal{L}$  with  $g_{xy}(x) = f(x)$  and  $g_{xy}(y) = f(y)$ . Since  $f$  and  $g_{xy}$  are continuous, there exists an open neighborhood  $U_{xy}$  of  $y$  such that  $g_{xy}(z) < f(z) + \epsilon$  for all  $z \in U_{xy}$ . Since  $X$  is compact, there exist  $y_1, y_2, \dots, y_n$  such

that  $X = \bigcup_{j=1}^n U_{xy_j}$ . We put  $g_x = g_{xy_1} \wedge \cdots \wedge g_{xy_n}$ . Then  $g_x \in \mathcal{L}$  by hypothesis (b),  $g_x(x) = f(x)$ , and  $g_x(z) < f(z) + \epsilon$  for every  $z \in X$ . Since  $f$  and  $g_x$  are continuous, there exists an open neighborhood  $V_x$  of  $x$  such that  $g_x(z) > f(z) - \epsilon$  for every  $z \in V_x$ . Since  $X$  is compact, there exist  $x_1, \dots, x_m$  such that  $X = \bigcup_{j=1}^m V_{x_j}$ . Define  $g = g_{x_1} \vee \cdots \vee g_{x_m}$ . Then  $g \in \mathcal{L}$ ,  $g(z) < f(z) + \epsilon$  for every  $z \in X$ , and  $g(z) > f(z) - \epsilon$  for every  $z \in X$ . ■

This theorem is due to Stone. From it we can deduce another approximation theorem, which contains the Weierstrass approximation theorem as a special case. It is known as the Stone–Weierstrass theorem.

**7.6 Theorem.** *Let  $X$  be a compact topological space, and suppose that  $A \subset C(X)$  has the following properties:*

- (a) if  $f, g \in A$ , and  $a, b \in \mathbf{R}$ , then  $af + bg \in A$ ;
- (b) if  $f, g \in A$ , then  $fg \in A$ ;
- (c) for any  $x, y \in X$ , with  $x \neq y$ , there exists  $f \in A$  with  $f(x) \neq f(y)$ ; and
- (d) each constant function belongs to  $A$ .

*Then  $A$  is dense in  $C(X)$ .*

**Proof.** It is clear that the closure  $\bar{A}$  of  $A$  in  $C(X)$  satisfies the same conditions (a)–(d), so we can assume that  $A$  is uniformly closed. We notice that conditions (a), (c), and (d) are taken unchanged from Theorem 7.5, so it suffices to show that  $f \vee g \in A$  and  $f \wedge g \in A$  whenever  $f, g \in A$ . (Of course, one of these suffices, since  $a \vee b = -(-a) \wedge (-b)$ .) We make the following remark: since

$$a \vee b = \frac{a+b}{2} + \frac{|a-b|}{2}, \quad a \wedge b = \frac{a+b}{2} - \frac{|a-b|}{2},$$

and  $|a| = a \vee (-a)$ , conditions (a) and (b) of Theorem 7.5 are equivalent to (a) and the condition: if  $f \in \mathcal{L}$ , then  $|f| \in \mathcal{L}$ . Now conditions (a) and (b) of our theorem imply that if  $g \in A$  and  $p$  is a polynomial with real coefficients, then  $p(g) \in A$ . According to Lemma 7.3, we can find polynomials  $P_n$  such that  $P_n(x)$  converges uniformly to  $|x|$  on any interval  $[-M, M]$  in  $\mathbf{R}$ . It follows that  $P_n(g)$  converges uniformly on  $X$  to  $|g|$ . Thus  $A$  satisfies all the hypotheses of Theorem 7.5, and is closed, so  $A = C(X)$ . ■

**7.7 Corollary.** *Let  $K$  be a closed and bounded subset of  $\mathbf{R}^n$ . Then the set  $P(K)$  of polynomials  $p(x_1, \dots, x_n)$  in the coordinate functions  $x_1, \dots, x_n$  is dense in  $C(K)$ .*

**Proof.** We need only observe that  $A = P(K)$  satisfies the hypotheses of the last theorem. ■



## 7.2 Lengths of Paths

Let  $(X, \rho)$  be a metric space. By a *path* in  $X$ , we will mean a continuous map  $f : I \rightarrow X$ , where  $I$  is a closed bounded interval in  $\mathbf{R}$ . We want to define the length of a path in  $X$ , and it is natural to do it as follows.

If  $f : [a, b] \rightarrow X$  is a path in  $X$ , and  $\pi = (t_j)_{j=0}^n$  is a partition of  $[a, b]$  (see Definition 5.1), we define

$$L(f, \pi) = \sum_{j=1}^n \rho(f(t_j), f(t_{j-1}))$$

and define the *length of  $f$*  by

$$L(f) = \sup\{L(f, \pi) : \pi \text{ a partition of } [a, b]\}.$$

(We think of  $L(f)$  as the supremum of the lengths of polygons inscribed in  $f$ .) We say that  $f$  is *rectifiable* if  $L(f) < +\infty$ . It is clear that  $L(f) = 0$  if and only if  $f$  is a constant path.

Let  $\mathcal{P}$  be the set of all rectifiable paths in  $X$  with domain  $[0, 1]$ . We define

$$d(f, g) = \sup\{\rho(f(t), g(t)) : 0 \leq t \leq 1\}$$

for  $f, g \in \mathcal{P}$ . As we saw in the last chapter, this is a metric on  $\mathcal{P}$ . Convergence in this metric is just uniform convergence of functions.

The map  $L : \mathcal{P} \rightarrow \mathbf{R}$  is not continuous. For instance, let us define  $g : \mathbf{R} \rightarrow \mathbf{R}$  by  $g(t) = 1 - |1 - 2t|$  for  $0 \leq t \leq 1$ , and set  $g(k+t) = g(t)$  for  $k \leq t \leq k+1$  ( $k \in \mathbf{Z}$ ). Define the path  $f_n$  in  $\mathbf{R}^2$  by  $f_n(t) = (t, g(nt)/n)$  for  $0 \leq t \leq 1$ . (The reader might make a quick sketch.) Then  $L(f_n) = \sqrt{5}$  for every  $n$ , as is easily checked, but  $(f_n)$  converges uniformly to the path  $f$ , given by  $f(t) = (t, 0)$ ,  $0 \leq t \leq 1$ , and  $L(f) = 1$ .

However, the map  $L$  does have the property, called *lower semicontinuity*, described in the next theorem.

**7.8 Theorem.** *If  $f_k \in \mathcal{P}$  for each  $k \in \mathbf{N}$ , and  $(f_k)$  converges uniformly to  $f$ , then  $L(f) \leq \liminf L(f_k)$ .*

**Proof.** Let  $\lambda < L(f)$ . There exists a partition  $\pi = (t_j)_{j=0}^n$  of  $[0, 1]$  such that  $\lambda < L(f, \pi) = \sum_{j=1}^n \rho(f(t_{j-1}), f(t_j))$ . For any  $\epsilon > 0$ , if  $g \in \mathcal{P}$  and  $d(f, g) < \epsilon/n$ , we have

$$\rho(f(t_j), f(t_{j-1})) \leq \rho(f(t_j), g(t_j)) + \rho(g(t_j), g(t_{j-1})) + \rho(g(t_{j-1}), f(t_{j-1}))$$

and hence

$$\begin{aligned} L(g, \pi) &= \sum_{j=1}^n \rho(g(t_j), g(t_{j-1})) \\ &\geq \sum_{j=1}^n \rho(f(t_{j-1}), f(t_j)) - 2\epsilon = L(f, \pi) - 2\epsilon > \lambda - 2\epsilon. \end{aligned}$$

Thus  $L(g) \geq L(g, \pi) > \lambda - 2\epsilon$ . Now if  $(f_k)$  converges to  $f$  uniformly, there exists  $k_0$  such that  $d(f, f_k) < \epsilon/n$  for all  $k \geq k_0$ , and hence  $L(f_k) > \lambda - 2\epsilon$  for all  $k \geq k_0$ . Thus  $\liminf L(f_k) \geq \lambda - 2\epsilon$ . Since  $\epsilon > 0$  was arbitrary, we have  $\liminf L(f_k) \geq \lambda$ . Since  $\lambda$  was an arbitrary number smaller than  $L(f)$ , we conclude  $\liminf L(f_k) \geq L(f)$ . ■

The next theorem says that in a compact metric space there always exists a shortest path between any two points. The proof will use the following lemma:

**7.9 Lemma.** *For any  $g \in \mathcal{P}$ , there exists  $g^* \in \mathcal{P}$  with the properties:*

- (a)  $g^*(0) = g(0)$  and  $g^*(1) = g(1)$ ;
- (b)  $L(g^*) = L(g)$ ; and
- (c) for all  $t, t' \in [0, 1]$ ,  $\rho(g^*(t), g^*(t')) \leq L(g)|t - t'|$ .

**Proof.** Let  $g_t$  be the restriction of  $g$  to the interval  $[0, t]$ , and let  $\lambda(t) = L(g_t)$ . It is clear that  $\lambda$  is an increasing function on  $[0, 1]$ , with  $\lambda(0) = 0$  and  $\lambda(1) = L(g)$ . Furthermore,  $\lambda$  is continuous. For given  $\epsilon > 0$ , we can choose a partition  $\pi = (t_j)_{j=0}^n$  of  $[0, 1]$  such that  $L(g, \pi) > L(g) - \epsilon$ . Since  $g$  is uniformly continuous, there exists  $\delta > 0$  such that  $\rho(g(t), g(t')) < \epsilon$  whenever  $|t - t'| < \delta$ , and we can also assume that  $t_j - t_{j-1} < \delta$  for  $1 \leq j \leq n$ . Suppose  $0 \leq t < t' \leq 1$  and  $t' - t < \eta = \min\{t_j - t_{j-1}\}$ . I claim that  $\lambda(t') - \lambda(t) \leq 3\epsilon$ . Suppose not. Then the restriction  $h$  of  $g$  to  $[t, t']$  has length  $L(h) > 3\epsilon$ . Hence there is a partition  $\pi'$  of  $[t, t']$  such that  $L(h, \pi') > 3\epsilon$ . Now since  $t' - t < t_j - t_{j-1}$  for every  $j$ ,  $1 \leq j \leq n$ ,  $t$  and  $t'$  belong to either the same or, at worst, two adjacent intervals  $[t_{j-1}, t_j]$ . Let  $\pi''$  be the partition obtained by throwing together the points of  $\pi$  and  $\pi'$ . Then

$$\begin{aligned} L(g, \pi'') &\geq L(g, \pi) - \rho(g(t_{j-1}), g(t)) - \rho(g(t'), g(t_{j+1})) + L(h, \pi') \\ &> L(g) - \epsilon - \epsilon - \epsilon + 3\epsilon = L(g), \end{aligned}$$

which is impossible. Thus  $\lambda$  is continuous on  $[0, 1]$ , and hence is a surjective map of  $[0, 1]$  onto  $[0, L(g)]$ .

We next define a map  $\phi : [0, L(g)] \rightarrow [0, 1]$  by  $\phi(s) = \inf\{t \in [0, 1] : \lambda(t) = s\}$ . If  $\lambda$  is strictly increasing, then of course  $\phi = \lambda^{-1}$ , but  $\lambda$  need not be strictly increasing ( $g$  may “stop to rest” in some subinterval of  $[0, 1]$ ). In any case,  $\phi$  is well-defined since  $\lambda$  is surjective. We define  $\tilde{g}(s) = g(\phi(s))$ . Then  $\tilde{g} : [0, L(g)] \rightarrow X$  is continuous (even though  $\phi$  need not be), and in fact we have for  $0 \leq s, s' \leq L(g)$ ,

$$\rho(\tilde{g}(s), \tilde{g}(s')) = \rho(g(t), g(t')),$$

where  $\lambda(t) = s$  and  $\lambda(t') = s'$ . It follows that  $\rho(\tilde{g}(s), \tilde{g}(s')) \leq |s' - s|$  for all  $s, s' \in [0, L(g)]$ . It is clear that  $\tilde{g}(0) = g(0)$  and  $\tilde{g}(L(g)) = g(1)$ . (We call

$\tilde{g}$  the reparametrization of  $g$  by arc length.) Now let  $g^*(t) = \tilde{g}(L(g)t)$  for  $t \in [0, 1]$ . It is easy to see that  $g^*(0) = g(0)$  and  $g^*(1) = g(1)$ , that  $L(g^*) = L(\tilde{g}) = L(g)$ , and that  $\rho(g^*(t), g^*(t')) \leq L(g)|t - t'|$  for all  $t, t' \in [0, 1]$ . ■

**7.10 Theorem.** Let  $X$  be a compact metric space, and let  $p, q \in X$ . Let  $\mathcal{S}$  be the set of all  $g \in \mathcal{P}$  with  $g(0) = p$  and  $g(1) = q$ . If  $\mathcal{S} \neq \emptyset$ , then there exists  $f \in \mathcal{S}$  such that  $L(f) \leq L(g)$  for every  $g \in \mathcal{S}$ .

**Proof.** Let  $D = \inf\{L(g) : g \in \mathcal{S}\}$ , and let  $(g_n)$  be a sequence in  $\mathcal{S}$  with  $L(g_n) \rightarrow D$  as  $n \rightarrow \infty$ . Using the last lemma, we obtain a sequence  $(g_n^*)$  in  $\mathcal{S}$  with  $L(g_n^*) = L(g_n)$  which also satisfies the condition

$$\rho(g_n^*(t), g_n^*(t')) \leq L(g_n)|t - t'|$$

for all  $t, t' \in [0, 1]$ . But this condition implies that  $(g_n^*)$  is an equicontinuous sequence of maps of  $[0, 1]$  into the compact metric space  $X$ , and hence by the Arzelà-Ascoli theorem, Theorem 6.71, there exists a uniformly convergent subsequence  $(g_{n_k}^*)$ . If  $f$  is the limit of this subsequence, we have by Theorem 7.8 that  $L(f) \leq \liminf L(g_{n_k}) = D$ , and the theorem is proved. ■

## 7.3 Fourier Series

In this section and the next, it will be convenient to use complex numbers, which we have mentioned in this book so far only in passing. Recall that the complex number field  $\mathbf{C}$  is the set  $\mathbf{R}^2$ , endowed with its natural addition, and the multiplication law  $(a, b)(c, d) = (ac - bd, bc + ad)$ , and that  $\mathbf{C}$  is a field, with the subfield  $\{(a, 0) : a \in \mathbf{R}\}$  being identified with the real field  $\mathbf{R}$ . We then write 1 for  $(1, 0)$ , 0 for  $(0, 0)$ , and set  $i = (0, 1)$ , so that  $i^2 = -1$ . Each complex number  $z$  can be expressed uniquely in the form  $z = x + iy$ , with  $x, y \in \mathbf{R}$ , and we set  $\bar{z} = x - iy$ , and call  $\bar{z}$  the *complex conjugate*, or simply *conjugate*, of  $z$ . We observe that  $z + \bar{z} = 2x$ ,  $z - \bar{z} = 2iy$ . We call  $x$  the real part of  $z$ , and denote it by  $\Re z$ , and  $y$  the imaginary part of  $z$ , denoted by  $\Im z$ . The norm  $|z| = \sqrt{x^2 + y^2}$  of a complex number  $z$  is called its *absolute value* or *modulus*. We observe that  $|z|^2 = x^2 + y^2 = z\bar{z}$ . The inequality  $|z + w| \leq |z| + |w|$  is a familiar fact about  $\mathbf{R}^2$ . We note that  $\overline{zw} = (\bar{z})(\bar{w})$ , and consequently  $|zw| = |z||w|$ . Any complex number  $z$  can be written in the form  $z = r\lambda$ , where  $r \geq 0$  and  $|\lambda| = 1$ ; if  $z \neq 0$ , then  $r = |z|$  and  $\lambda = z/|z|$  are uniquely determined, while if  $z = 0$ , then  $r = 0$  and  $\lambda$  can be any number of absolute value 1.

We will deal in this section with complex-valued functions of a real variable, not functions of a complex variable, the subject matter of an entirely different course.

If  $f : X \rightarrow \mathbf{C}$ , then  $u = \Re f$  and  $v = \Im f$  map  $X$  into  $\mathbf{R}$ , and  $f = u + iv$ . When  $X$  is an interval in  $\mathbf{R}$ , we can discuss the derivative of  $f$  at a point,

or the integral of  $f$  over the interval; our previous definitions make sense in the context of complex-valued functions. It is very easy to see that when  $f : I \rightarrow \mathbf{C}$ ,  $I$  an interval in  $\mathbf{R}$ , with  $u = \Re f$  and  $v = \Im f$ , then  $f'(t)$  exists if and only if  $u'(t)$  and  $v'(t)$  exist, and  $f'(t) = u'(t) + iv'(t)$  in this case. Similarly,  $\int_a^b f$  exists if and only if  $\int_a^b u$  and  $\int_a^b v$  exist, and  $\int_a^b f = \int_a^b u + i \int_a^b v$  in this case. It is obvious that if  $f$  and  $g$  are complex-valued functions on  $[a, b]$  which are Riemann integrable, then  $f + g$  is again such a function, and  $\int_a^b (f + g) = \int_a^b f + \int_a^b g$ . It is very easy to see also that for any constant  $c \in \mathbf{C}$ ,  $\int_a^b (cf) = c \int_a^b f$ . It is not as trivial to establish the useful inequality

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt.$$

To see this, choose  $\lambda \in \mathbf{C}$  with  $|\lambda| = 1$  and  $\lambda \int_a^b f \geq 0$ . Then

$$\begin{aligned} \left| \int_a^b f(t) dt \right| &= \lambda \int_a^b f(t) dt = \int_a^b \lambda f(t) dt \\ &= \Re \int_a^b \lambda f(t) dt = \int_a^b \Re(\lambda f(t)) dt \\ &\leq \int_a^b |\lambda f(t)| dt = \int_a^b |f(t)| dt. \end{aligned}$$

We say that a function  $f : \mathbf{R} \rightarrow Y$  is *periodic*, with period  $T$ , or simply  $T$ -periodic, if  $f(t+T) = f(t)$  for all  $t$ . In this definition,  $Y$  may be any set. Obviously, a constant function is periodic, with every  $T$  as period. If  $Y$  is a metric space, and  $f : \mathbf{R} \rightarrow Y$  is periodic and continuous, then either  $f$  is constant, or  $f$  has a smallest positive period, and every period of  $f$  is an integer multiple of this one (we leave the proof as an exercise.) It is obvious that sums and products of complex-valued  $T$ -periodic functions are again  $T$ -periodic, as are pointwise limits of such functions. We note that if  $f$  is  $T$ -periodic, then the function  $t \mapsto f(tT)$  is periodic with period 1.

The reader can easily verify that for any  $\alpha \in \mathbf{R}$ ,  $\int_\alpha^{\alpha+T} g(t) dt = \int_0^T g(t) dt$  whenever  $g$  is  $T$ -periodic and integrable over some interval of length  $T$ .

Examples of functions with the period  $2\pi$  are the sine and cosine functions, and hence any function of the form

$$f(t) = A_0 + \sum_{n=1}^N (A_n \cos 2\pi nt + B_n \sin 2\pi nt), \quad (7.2)$$

is periodic, with period 1; we will refer to such functions as *trigonometric polynomials*, of degree  $\leq N$ .

We will make use of the complex exponential function, which we define by the rule  $e^{it} = \cos t + i \sin t$ . We note that  $|e^{it}| = 1$  for all  $t \in \mathbf{R}$ , and

that if  $\lambda \in \mathbf{C}$  and  $|\lambda| = 1$ , then there exist  $t \in \mathbf{R}$  such that  $\lambda = e^{it}$ ;  $t$ , of course, is not uniquely determined, only up to addition by an integer multiple of  $2\pi$ . Thus any  $z \in \mathbf{C}$  can be expressed in the form  $z = re^{it}$  with  $r \geq 0$  and  $t \in \mathbf{R}$ ; any such  $t$  is called an *argument* of  $z$ . Clearly, we have  $\overline{e^{it}} = e^{-it}$ . The addition laws for the trigonometric functions which we obtained in Chapter 3 quickly lead to the identity  $e^{i(a+t)} = e^{ia}e^{it}$ , which is sufficient to justify the notation. We note that  $\cos t = \frac{1}{2}(e^{it} + e^{-it})$ , and  $\sin t = \frac{1}{2i}(e^{it} - e^{-it})$ . Hence the trigonometric polynomial described in (7.2) can also be expressed in the form

$$f(t) = \sum_{n=-N}^N c_n e^{2\pi i n t}, \quad (7.3)$$

with  $c_0 = A_0$ ,  $c_n = \frac{1}{2}(A_n - iB_n)$  and  $c_{-n} = \frac{1}{2}(A_n + iB_n)$  for  $n > 0$ . Conversely, any function of the form (7.3) can also be expressed in the form (7.2) by taking  $A_0 = c_0$  and  $A_n = c_n + c_{-n}$ ,  $B_n = i(c_n - c_{-n})$  for  $n > 0$ . We note that in this correspondence, the coefficients  $A_n$  and  $B_n$  are real if and only if  $c_n = \overline{c_{-n}}$ .

**7.11 Proposition.** *Let  $f$  be a trigonometric polynomial,*

$$f(t) = \sum_{n=-N}^N c_n e^{2\pi i n t} = \frac{a_0}{2} + \sum_{n=1}^N (a_n \cos 2\pi n t + b_n \sin 2\pi n t).$$

Then, for any  $\alpha \in \mathbf{R}$ ,

$$c_n = \int_{\alpha}^{\alpha+1} f(t) e^{-2\pi i n t} dt, \quad -N \leq n \leq N, \quad (7.4)$$

and, consequently,

$$a_n = 2 \int_{\alpha}^{\alpha+1} f(t) \cos 2\pi n t dt, \quad b_n = 2 \int_{\alpha}^{\alpha+1} f(t) \sin 2\pi n t dt. \quad (7.5)$$

**Proof.** It suffices to observe that

$$\int_{\alpha}^{\alpha+1} e^{ikt} dt = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k \neq 0 \end{cases}$$

to verify equation (7.4). We deduce Equation (7.5) by using the relations  $a_n = c_n + c_{-n}$ ,  $b_n = i(c_n - c_{-n})$ . ■

We will fix the following notation for the rest of this chapter:

**7.12 Definition.** Let  $\mathcal{R}$  denote the set of all 1-periodic complex-valued functions on  $\mathbf{R}$  which are Riemann integrable over bounded intervals, and let  $\mathcal{C}$  denote the set of all continuous 1-periodic complex-valued functions on  $\mathbf{R}$ .

**7.13 Definition.** For each  $n \in \mathbb{Z}$ , let  $e_n : \mathbb{R} \rightarrow \mathbb{C}$  be defined by  $e_n(t) = e^{2\pi i n t}$ .

For each  $f \in \mathcal{X}$ , define  $\hat{f} : \mathbb{Z} \rightarrow \mathbb{C}$  by

$$\hat{f}_n = \int_0^1 f e_{-n} = \int_0^1 f(t) e^{-2\pi i n t} dt$$

and call  $\hat{f}_n$  the  $n$ th Fourier coefficient of  $f$ . The series  $\sum_{n=-\infty}^{\infty} \hat{f}_n e^{2\pi i n x}$  is called the Fourier series of  $f$ ; it is to be regarded as the sequence of partial sums  $(s_n)$ , where

$$s_n(x) = \sum_{k=-n}^n \hat{f}_k e^{2\pi i k x}$$

for  $n = 0, 1, 2, \dots$

**7.14 Proposition.** Let  $f, g \in \mathcal{X}$ . Then  $(\widehat{f+g})_n = \hat{f}_n + \hat{g}_n$ ; if  $c \in \mathbb{C}$ ,  $(\widehat{cf})_n = c\hat{f}_n$ ; if  $f_a$  is defined by  $f_a(t) = f(t-a)$ , then  $(\hat{f}_a)_n = e^{-2\pi i n a} \hat{f}_n$ .

**Proof.** We leave the first two statements to the reader. For the last statement, we can compute

$$\begin{aligned} (\hat{f}_a)_n &= \int_0^1 f(t-a) e^{-2\pi i n t} dt = \int_a^{a+1} f(t-a) e^{-2\pi i n t} dt \\ &= \int_0^1 f(u) e^{-2\pi i n (u+a)} du = e^{-2\pi i n a} \hat{f}_n, \end{aligned}$$

as claimed. ■

From Proposition 7.11 we see that if  $f$  is a trigonometric polynomial, then the Fourier series of  $f$  converges to  $f$  at each point; in fact, we have  $s_n = f$  for every  $n \geq N$ , if  $N$  is the degree of  $f$ . In general, a Fourier series, even the Fourier series of a continuous function, need not converge at every point. If the Fourier series of a function  $f$  does converge at a point  $t$ , there is no guarantee that it converges to  $f(t)$ . In fact, if we change the definition of  $f$  at one point, the Fourier coefficients remain unchanged. However, we will show that for a reasonably nice function  $f$ , the Fourier series of  $f$  does converge uniformly to  $f$ . First, we establish that a continuous function is determined by its Fourier series.

**7.15 Theorem.** Let  $f \in \mathcal{C}$ . If  $\hat{f}_n = 0$  for every  $n \in \mathbb{Z}$ , then  $f = 0$ .

**Proof.** Suppose that  $f \in \mathcal{C}$ ,  $f \neq 0$ , with  $\hat{f}_n = 0$  for every  $n$ ; using Proposition 7.14, we can assume that  $f$  is real-valued, and  $f(0) > 0$ . Then there exist  $\delta > 0$  and  $\epsilon > 0$  such that  $f(t) \geq \epsilon$  for all  $t \in (-\delta, \delta)$ . Since  $\hat{f}_n = 0$  for all  $n$ , we see that  $\int_{-1/2}^{1/2} f(t)g(t)dt = 0$  for every trigonometric

polynomial  $g$ . We construct a sequence of real trigonometric polynomials  $(g_n)$  with the properties: (a)  $g_n(t) \geq 1$  for all  $t \in (-\delta, \delta)$ ; (b)  $|g_n(t)| \leq 1$  for  $\delta \leq |t| \leq \frac{1}{2}$ ; and (c)  $g_n \rightarrow +\infty$  uniformly on  $[-\delta', \delta']$  for any  $\delta' < \delta$ . With such  $g_n$ , we readily find that  $\int_{-1/2}^{1/2} f(t)g_n(t) dt \rightarrow +\infty$  as  $n \rightarrow \infty$ , a contradiction. It suffices to set  $g_n(t) = [g(t)]^n$ , where  $g(t) = 1 + \cos 2\pi t - \cos 2\pi\delta$ . Since  $\cos 2\pi t < \cos 2\pi\delta$  for  $\delta < |t| \leq \frac{1}{2}$ , we see that  $-\cos \delta \leq g(t) \leq 1$  for  $\delta \leq |t| \leq \frac{1}{2}$ , while  $g(t) \geq 1 + \cos \delta' - \cos \delta > 1$  for  $|t| \leq \delta' < \delta$ . ■

**7.16 Corollary.** *If  $f, g \in \mathcal{C}$  and  $\hat{f}_n = \hat{g}_n$  for all  $n$ , then  $f = g$ .*

**Proof.** If  $h = f - g$ , then  $h \in \mathcal{C}$  and  $\hat{h}_n = 0$  for every  $n$ , so  $h = 0$  by the last theorem. ■

**7.17 Corollary.** *Let  $f \in \mathcal{C}$ . If the Fourier series of  $f$  converges uniformly, then it converges to  $f$ .*

**Proof.** If  $s_n \rightarrow g$  uniformly, then  $s_n e_{-k}$  converges uniformly to  $g e_{-k}$ , so  $(s_n)_k = \int_0^1 s_n e^{-ik} \rightarrow \int_0^1 g e^{-ik} = \hat{g}_k$ . But by Proposition 7.11,  $(s_n)_k = \hat{f}_k$  for all  $n \geq k$ , so we conclude that  $\hat{g}_k = \hat{f}_k$  for all  $k$ , and hence by the last corollary, that  $g = f$ . ■

**7.18 Definition.** *If  $f, g \in \mathcal{R}$ , we define their inner product as*

$$\langle f, g \rangle = \int_0^1 f(t) \overline{g(t)} dt$$

*and the 2-norm of  $f$  by  $\|f\|_2 = \langle f, f \rangle^{1/2}$ .*

Since  $\bar{e}_n = e_{-n}$  for every  $n$ , we can now write  $\hat{f}_n = \langle f, e_n \rangle$  for each  $f \in \mathcal{R}$  and  $n \in \mathbb{Z}$ .

We summarize the properties of this complex inner product:

**7.19 Proposition.** *For all  $f \in \mathcal{R}$ ,  $\|f\|_2 \geq 0$ , and  $\|cf\|_2 = |c|\|f\|_2$  for any  $c \in \mathbb{C}$ . If  $f \in \mathcal{C}$ , then  $\|f\|_2 = 0$  only if  $f = 0$ . For any  $f, g \in \mathcal{R}$ ,  $\langle f, g \rangle = \overline{\langle g, f \rangle}$ . For fixed  $g$ , the map  $f \mapsto \langle f, g \rangle$  is linear, i.e.,*

$$\langle c_1 f_1 + c_2 f_2, g \rangle = c_1 \langle f_1, g \rangle + c_2 \langle f_2, g \rangle$$

*for any  $f_1, f_2, g \in \mathcal{R}$  and any  $c_1, c_2 \in \mathbb{C}$ . The Schwarz–Bunyakovsky inequality holds:  $|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$  for all  $f, g \in \mathcal{R}$ ; if  $f, g \in \mathcal{C}$ , then equality holds if and only if  $f$  and  $g$  are linearly dependent.*

**Proof.** We omit the easy proofs of the first few statements. To verify the inequality, we observe that for any  $\lambda \in \mathbb{C}$ ,

$$0 \leq \|f - \lambda g\|_2^2 = \langle f - \lambda g, f - \lambda g \rangle$$

$$\begin{aligned}
&= \langle f, f - \lambda g \rangle - \lambda \langle g, f - \lambda g \rangle \\
&= \overline{\langle f - \lambda g, f \rangle} - \lambda \overline{\langle f - \lambda g, g \rangle} \\
&= \langle f, f \rangle - \overline{\lambda \langle g, f \rangle} - \lambda \overline{\langle f, g \rangle} + |\lambda|^2 \langle g, g \rangle.
\end{aligned}$$

Let  $\lambda = t\langle f, g \rangle$ , with  $t$  real. Then we have

$$2t|\langle f, g \rangle|^2 \leq \|f\|_2^2 + t^2|\langle f, g \rangle|^2\|g\|_2^2.$$

If  $\langle g, g \rangle = 0$ , then we have

$$2t|\langle f, g \rangle|^2 \leq \|f\|_2^2$$

for every real  $t$ , which gives  $\langle f, g \rangle = 0$ . If  $\|g\|_2 \neq 0$ , we can choose  $t = 1/\|g\|_2^2$ , obtaining

$$\frac{2|\langle f, g \rangle|^2}{\|g\|_2^2} \leq \|f\|_2^2 + \frac{|\langle f, g \rangle|^2}{\|g\|_2^2}$$

which again gives  $|\langle f, g \rangle|^2 \leq \|f\|_2^2\|g\|_2^2$ .

We note that equality holds if and only if either  $\|g\|_2 = 0$  or  $\|f - \lambda g\|_2 = 0$ , which for continuous  $f$  and  $g$  implies that  $f = \lambda g$ . ■

**7.20 Corollary.** For any  $f, g \in \mathcal{R}$ , we have  $\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$ .

**Proof.** We have

$$\begin{aligned}
\|f + g\|_2^2 &= \langle f + g, f + g \rangle = \|f\|_2^2 + \langle f, g \rangle + \langle g, f \rangle + \|g\|_2^2 \\
&\leq \|f\|_2^2 + 2\|f\|_2\|g\|_2 + \|g\|_2^2 = (\|f\|_2 + \|g\|_2)^2,
\end{aligned}$$

and taking square roots gives the desired result. ■

It follows at once that  $\rho_2$  defined by  $\rho_2(f, g) = \|f - g\|_2$  is a metric on  $\mathcal{C}$ , and a pseudometric on  $\mathcal{R}$ . (That is, on  $\mathcal{R}$  it has all the properties of a metric except that  $\rho_2(f, g) = 0$  does not imply that  $f = g$ .) Since  $\|f\|_2 \leq \|f\|$  for all  $f \in \mathcal{R}$ , uniform convergence implies convergence with respect to the metric  $\rho_2$ . Convergence in the metric  $\rho_2$  is also called *mean-square convergence*. Because the inner product for functions shares the formal properties of the familiar inner product of vectors in  $\mathbf{R}^2$  or  $\mathbf{R}^3$ , our intuition for low-dimensional spaces can serve as a reliable guide in thinking about infinite-dimensional spaces such as  $\mathcal{R}$  or  $\mathcal{C}$ .

**7.21 Definition.** If  $f, g \in \mathcal{R}$ , we say that  $f$  and  $g$  are orthogonal, and write  $f \perp g$ , if  $\langle f, g \rangle = 0$ . A subset  $\mathcal{S}$  of  $\mathcal{R}$  is called orthogonal if  $f, g \in \mathcal{S}$  and  $f \neq g$  implies  $f \perp g$ ;  $\mathcal{S}$  is called orthonormal if  $\mathcal{S}$  is orthogonal and also  $\|f\|_2 = 1$  for every  $f \in \mathcal{S}$ .



For example, the set  $\{e_n : n \in \mathbf{Z}\}$  is an orthonormal set, as we saw in the proof of Proposition 7.11.

Another interesting orthonormal set: let  $r : \mathbf{R} \rightarrow \mathbf{C}$  be defined by

$$r(t) = \begin{cases} 1 & \text{if } 0 \leq t < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} \leq t < 1, \end{cases}$$

and  $r(t+n) = r(t)$  for all  $n \in \mathbf{Z}$ . Define  $r_n$  ( $n \in \mathbf{N}$ ) by

$$r_n(t) = r(2^{n-1}t), \quad t \in \mathbf{R}.$$

It is easy to see that  $\{r_n : n \in \mathbf{N}\}$  is an orthonormal set in  $\mathcal{R}$ ; in fact, it has the stronger property: if  $g : \mathbf{R}^n \rightarrow \mathbf{R}$  is an arbitrary function, then  $\langle g(r_1, \dots, r_n), r_k \rangle = 0$  for every  $k > n$ . These functions are called Rademacher functions.

We next observe that the Pythagorean theorem holds in  $\mathcal{R}$ .

**7.22 Proposition.** *If  $f, g \in \mathcal{R}$  and  $f \perp g$ , then  $\|f + g\|_2^2 = \|f\|_2^2 + \|g\|_2^2$ .*

**Proof.** Look at the proof of the last corollary. ■

Of course, this generalizes to any number of summands.

**7.23 Corollary.** *If  $f_1, \dots, f_n \in \mathcal{R}$ , with  $f_j \perp f_k$  whenever  $j \neq k$ , then*

$$\left\| \sum_{k=1}^n f_k \right\|_2^2 = \sum_{k=1}^n \|f_k\|_2^2.$$

**Proof.** Since  $f_n \perp \sum_{k=1}^{n-1} f_k$ , this follows by induction from the preceding proposition. ■

**7.24 Proposition.** *Let  $F$  be a finite set, and suppose  $\{\phi_\alpha : \alpha \in F\}$  is an orthonormal set in  $\mathcal{R}$ . If  $f \in \mathcal{R}$ , and  $c_\alpha = \langle f, \phi_\alpha \rangle$  for each  $\alpha \in F$ , then  $g = \sum_{\alpha \in F} c_\alpha \phi_\alpha$  has the following properties:*

- (a)  $f - g \perp \phi_\alpha$  for each  $\alpha \in F$ ; and
- (b) if  $h = \sum_{\alpha \in F} d_\alpha \phi_\alpha$ , where  $d_\alpha \in \mathbf{C}$  ( $\alpha \in F$ ), then

$$\|f - h\|_2^2 = \|f - g\|_2^2 + \sum_{\alpha \in F} |d_\alpha - c_\alpha|^2;$$

in particular,  $\|f - h\|_2 > \|f - g\|_2$  unless  $h = g$ .

**Proof.** We have, for each  $\alpha \in F$ ,

$$\begin{aligned} \langle f - g, \phi_\alpha \rangle &= \langle f, \phi_\alpha \rangle - \langle g, \phi_\alpha \rangle \\ &= c_\alpha - \sum_{\beta \in F} \langle c_\beta \phi_\beta, \phi_\alpha \rangle \\ &= c_\alpha - c_\alpha = 0, \end{aligned}$$

which establishes (a). Now it follows that  $f - g \perp g - h$ , so by Proposition 7.22 it follows that

$$\|f - h\|_2^2 = \|f - g\|_2^2 + \|g - h\|_2^2 = \|f - g\|_2^2 + \sum_{\alpha \in F} |d_\alpha - c_\alpha|^2,$$

where we used Corollary 7.23 to evaluate  $\|g - h\|_2$ . ■

The next Corollary is known as *Bessel's inequality*.

**7.25 Corollary.** Let  $\{\phi_\alpha : \alpha \in A\}$  be an orthonormal set in  $\mathcal{R}$ . For any  $f \in \mathcal{R}$ ,  $\sum_{\alpha \in A} |\langle f, \phi_\alpha \rangle|^2 \leq \|f\|_2^2$ .

**Proof.** Taking  $d_\alpha = 0$  in the last proposition shows that  $\sum_{\alpha \in F} |\langle f, \phi_\alpha \rangle|^2 \leq \|f\|_2^2$  for every finite subset  $F$  of  $A$ , which is what this corollary states. ■

Taking the orthonormal set in the last corollary to be the trigonometric system  $\{e_n : n \in \mathbb{Z}\}$ , we obtain the classical Bessel inequality:

**7.26 Corollary.** If  $f \in \mathcal{R}$ , then  $\sum_{n \in \mathbb{Z}} |\hat{f}_n|^2 \leq \|f\|_2^2$ .

The next corollary, which is an immediate consequence of the last, is known as the *Riemann-Lebesgue lemma*.

**7.27 Corollary.** If  $f \in \mathcal{R}$ , then  $\hat{f}_n \rightarrow 0$  as  $|n| \rightarrow \infty$ .

**7.28 Definition.** We say that  $f : [a, b] \rightarrow \mathbb{C}$  is piecewise smooth if there exists a partition  $(x_k)_{k=0}^n$  of  $[a, b]$  such that

- (a)  $f'(t)$  exists for all  $t \in (x_{k-1}, x_k)$  for every  $k$ ,  $1 \leq k \leq n$ ; and
- (b) for each  $k$ , the restriction of  $f'$  to  $(x_{k-1}, x_k)$  has a continuous extension to  $[x_{k-1}, x_k]$ .

For instance, a piecewise linear function is piecewise smooth.

**7.29 Theorem.** Let  $f \in \mathcal{C}$ , and suppose that the restriction of  $f$  to  $[0, 1]$  is piecewise smooth. Then the Fourier series of  $f$  converges uniformly on  $\mathbb{R}$  to  $f$ .

**Proof.** Using integration by parts (Theorem 5.32) over each of the intervals  $[x_{k-1}, x_k]$ , and adding, we obtain for  $n \neq 0$ ,

$$\int_0^1 f(t) e^{-2\pi i n t} dt = \frac{1}{2\pi i n} \int_0^1 f'(t) e^{-2\pi i n t} dt,$$

so  $\hat{f}_n = \hat{f}'_n/2\pi in$ . (Since  $f'$  is bounded and continuous except at finitely many points,  $f' \in \mathcal{A}$ , and  $\hat{f}'$  is well-defined.) Hence

$$\begin{aligned}\sum_{n \in \mathbf{Z}} |\hat{f}_n| &= |\hat{f}_0| + \sum_{n \neq 0} \frac{|\hat{f}'_n|}{2\pi|n|} \\ &\leq |f(0)| + \frac{1}{2\pi} \left( \sum_{n \neq 0} \frac{1}{n^2} \right)^{1/2} \left( \sum_{n \neq 0} |\hat{f}'_n|^2 \right)^{1/2}\end{aligned}$$

by the Cauchy-Schwarz inequality. Since  $\sum_{n=1}^{\infty} n^{-2} < \infty$ , and from Corollary 7.26 we know  $\sum |\hat{f}'_n|^2 < \infty$ , we obtain  $\sum |\hat{f}_n| < \infty$ , so the Fourier series  $\sum_{n \in \mathbf{Z}} \hat{f}_n e_n$  converges uniformly on  $\mathbf{R}$  by the Weierstrass  $M$ -test. By Corollary 7.17, the sum of the series is  $f$ . ■

**7.30 Theorem.** *For every  $f \in \mathcal{C}$ , there exists a sequence of trigonometric polynomials which converges uniformly to  $f$ .*

**Proof.** We can assume that  $f$  is real-valued. Let  $\epsilon > 0$ . By Theorem 3.20, there exists  $g$ , a continuous piecewise linear real function on  $[0, 1]$ , such that  $|f(x) - g(x)| < \epsilon/2$  for all  $x \in [0, 1]$ . Clearly,  $g$  can be chosen so that  $g(0) = g(1)$ , so the 1-periodic extension of  $g$ , which we still denote by  $g$ , belongs to  $\mathcal{C}$  and satisfies  $\|f - g\| < \epsilon/2$ . By Theorem 7.29, there exists a trigonometric polynomial  $h$  with  $\|g - h\| < \epsilon/2$ , and thus  $\|f - h\| < \epsilon$ . The theorem follows. ■

**7.31 Theorem.** *For every  $f \in \mathcal{A}$ , there exists a sequence of trigonometric polynomials which converges to  $f$  in the mean square metric.*

**Proof.** We may assume that  $f$  is real-valued. Let  $\epsilon > 0$ . Let  $M = \sup |f(t)|$ . By Theorem 5.17, there exists a continuous function  $g$  on  $[0, 1]$  such that  $\int_0^1 |f(t) - g(t)| dt < \epsilon^2/(2M)$ ; clearly, we can choose  $g$  with  $g(0) = g(1)$ , so that the 1-periodic extension of  $g$  to  $\mathbf{R}$  is continuous, and also we can take  $g$  with  $\|g\| \leq M$  (replace  $g$  by  $\max\{-M, \min\{g, M\}\}$  if necessary). Then

$$\begin{aligned}\int_0^1 |f(t) - g(t)|^2 dt &= \int_0^1 |f(t) - g(t)| |f(t) - g(t)| dt \\ &\leq 2M \int_0^1 |f(t) - g(t)| dt < \epsilon^2,\end{aligned}$$

so  $\|f - g\|_2 < \epsilon$ . By Theorem 7.30, there exists a trigonometric polynomial  $h$  with  $\|g - h\| < \epsilon$ , and hence  $\|g - h\|_2 < \epsilon$ , and thus we have  $\|f - h\|_2 < 2\epsilon$ . The theorem follows. ■

We can now sharpen Bessel's inequality (Corollary 7.26) to an equality, which is known as *Parseval's relation*.

**7.32 Theorem.** For every  $f \in \mathcal{R}$ , the Fourier series of  $f$  converges to  $f$  in the mean square, and

$$\|f\|_2^2 = \sum_{-\infty}^{\infty} |\hat{f}_n|^2. \quad (7.6)$$

**Proof.** Let  $\epsilon > 0$ , and let  $s_n$  be the  $n$ th partial sum of the Fourier series of  $f$ , i.e.,  $s_n = \sum_{k=-n}^n \hat{f}_k e_k$ . By Theorem 7.31 there exists a trigonometric polynomial  $g$  with  $\|f - g\|_2 < \epsilon$ . By Proposition 7.24,  $\|f - s_n\|_2 \leq \|f - g\|_2$  for every  $n \geq m$ , where  $m$  is the degree of  $g$ . That proposition also tells us that  $\|f - s_n\|_2^2 = \|f\|_2^2 - \sum_{k=-n}^n |\hat{f}_k|^2$ , so that Parseval's relation follows. ■

## 7.4 Weyl's Theorem

Recall that for any real number  $x$ ,  $[x]$  denotes the greatest integer in  $x$ , i.e.,  $[x] \in \mathbf{N}$  and  $[x] \leq x < [x] + 1$ . We define  $\langle x \rangle$ , the *fractional part* of  $x$ , by  $\langle x \rangle = x - [x]$ . If  $\xi$  is an irrational real number, then  $\{\langle n\xi \rangle : n \in \mathbf{N}\}$  is dense in  $[0, 1]$ , by Dirichlet's theorem (Theorem 1.23). We devote this section to proving the following stronger result, known as Weyl's theorem:

**7.33 Theorem.** If  $\xi \in \mathbf{R}$  is irrational, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \# \{k \in \mathbf{N} : 1 \leq k \leq n, \langle k\xi \rangle \in [a, b]\} = b - a$$

for any  $0 \leq a < b < 1$ .

**Proof.** Define, for each  $n \in \mathbf{N}$ , the map  $L_n : \mathcal{R} \rightarrow \mathbf{C}$  by

$$L_n(f) = \frac{1}{n} \sum_{k=1}^n f(k\xi).$$

If  $\psi$  is the periodic extension of the indicator function of the interval  $[a, b]$ , i.e.,  $\psi$  is defined by

$$\psi(t) = \begin{cases} 1 & \text{if } \langle t \rangle \in [a, b], \\ 0 & \text{if } \langle t \rangle \notin [a, b], \end{cases}$$

then

$$L_n(\psi) = \frac{1}{n} \# \{k \in \mathbf{N} : 1 \leq k \leq n, \langle k\xi \rangle \in [a, b]\},$$

and  $\int_0^1 \psi(t) dt = b - a$ . We shall prove that for any  $f \in \mathcal{R}$ ,  $L_n(f) \rightarrow \int_0^1 f(t) dt$  as  $n \rightarrow \infty$ , thus proving a generalization of the theorem.

We begin by observing two obvious properties of the functions  $L_n$ . These are:

(a) for any  $f, g \in \mathcal{R}$  and any  $c, d \in \mathbf{C}$ ,  $L_n(cf + dg) = cL_n(f) + dL_n(g)$ ; and

(b) if  $f, g \in \mathcal{R}$  and  $f \leq g$ , then  $L_n(f) \leq L_n(g)$ .

We first prove a special case of the generalized theorem: if  $f$  is a trigonometric polynomial of period 1, i.e.,  $f(t) = \sum_{k=-N}^N c_k e^{2\pi i k t}$ , then  $L_n(f) \rightarrow \int_0^1 f(t) dt$  as  $n \rightarrow \infty$ .

In view of (a) above, it suffices to prove this when  $f(t) = e_k(t) = e^{2\pi i k t}$ . We note that  $\int_0^1 e_k(t) dt = 0$  if  $k \neq 0$  and  $\int_0^1 e_0(t) dt = 1$ . Now if  $k \neq 0$ , then  $e^{2\pi i k \xi} \neq 1$ , since  $\xi$  is irrational, so

$$L_n(e_k) = \frac{1}{n} \sum_{m=1}^n e^{2\pi i m k \xi} = \frac{e^{2\pi i k \xi}}{n} \frac{1 - e^{2\pi i n k \xi}}{1 - e^{2\pi i k \xi}}.$$

Thus, for  $k \neq 0$ ,

$$|L_n(e_k)| \leq \frac{2}{n} \frac{1}{|1 - e^{2\pi i k \xi}|}$$

which approaches 0 as  $n \rightarrow \infty$ . When  $k = 0$ ,  $L_n(e_0) = 1 = \int_0^1 e_0(t) dt$ . Our special case is proven.

Now let  $f \in \mathcal{R}$ . We suppose, without loss of generality, that  $f$  is real-valued. By Theorem 5.17, there exist continuous functions  $g$  and  $h$  on  $[0, 1]$  with  $g < f < h$  and  $\int_0^1 h(t) dt - \int_0^1 g(t) dt < \epsilon$ . It is easy to see that we can choose such  $g$  and  $h$  such that  $g(0) = g(1)$ ,  $h(0) = h(1)$ , so that they can be extended to be continuous 1-periodic functions on  $\mathbf{R}$ . Now by Theorem 7.30 there exist trigonometric polynomials  $p$  and  $q$  with  $\|g - \epsilon - p\| < \epsilon$  and  $\|h + \epsilon - q\| < \epsilon$ . Then we have  $p < f < q$  and  $\int_0^1 q(t) dt - \int_0^1 f(t) dt < 2\epsilon$ ,  $\int_0^1 f(t) dt - \int_0^1 p(t) dt < 2\epsilon$ . Since  $L_n(p) \rightarrow \int_0^1 p(t) dt$  and  $L_n(q) \rightarrow \int_0^1 q(t) dt$ , from  $L_n(p) \leq L_n(f) \leq L_n(q)$  we get

$$\begin{aligned} \int_0^1 f(t) dt - 2\epsilon &< \int_0^1 p(t) dt \\ &\leq \liminf L_n(f) \leq \limsup L_n(f) \\ &\leq \int_0^1 q(t) dt < \int_0^1 f(t) dt + 2\epsilon, \end{aligned}$$

and since  $\epsilon > 0$  is arbitrary, we see that  $L_n(f) \rightarrow \int_0^1 f(t) dt$  as  $n \rightarrow \infty$ . ■

## 7.5 Exercises

1. Show that if  $f$  is a function of class  $C^1$  on  $[a, b]$ , and  $\epsilon > 0$ , there exists a polynomial  $p$  such that

$$\sup\{|f(t) - p(t)| + |f'(t) - p'(t)| : t \in [a, b]\} < \epsilon.$$

2. Let  $\Gamma = \{\lambda \in \mathbf{C} : |\lambda| = 1\}$ . To each  $f \in C(\Gamma)$ , we can associate the function  $f \circ e_1$ , i.e.,  $t \mapsto f(e^{2\pi it})$ , in  $\mathcal{C}$ . Note that  $f \circ e_1$  is a trigonometric polynomial if (and only if)  $f$  is (the restriction to  $\Gamma$  of) an ordinary polynomial in the coordinate functions  $x$  and  $y$  of  $\mathbf{R}^2$ . Use this idea and the Stone–Weierstrass theorem to give another proof of Theorem 7.30.

3. Deduce Theorem 7.1 from Theorem 7.30.

4. If  $A$  is an algebra of complex-valued functions on a compact space  $X$ , which separates points and contains the constant functions,  $A$  is not necessarily dense in  $C(X, \mathbf{C})$  (compare Theorem 7.6). For example, let  $X = \{z \in \mathbf{C} : |z| = 1\}$ , and let  $A$  be the set of all polynomial functions, i.e., functions of the form  $p(z) = \sum_{k=0}^n c_k z^k$ . Show that  $|p(z) - \bar{z}| \geq 1$  for all  $z \in X$  and every polynomial  $p$ . HINT: Observe that  $\int_0^{2\pi} (1 - e^{it} p(e^{it})) dt = 1$  for any polynomial  $p$ .

5. Show that the Taylor series for  $\sqrt{1-x}$  converges uniformly for  $0 \leq x \leq 1$ , and deduce another proof of Lemma 7.2.

6. Suppose that  $f : [0, 1] \rightarrow \mathbf{R}$  is continuous, and that  $\int_0^1 f(t) t^n dt = 0$  for every positive integer  $n$ . Show that  $f = 0$ .

7. Let  $f : [a, b] \rightarrow X$  be a path in the metric space  $X$ , and let  $\phi$  be a continuous strictly increasing function on  $[c, d]$ , with  $\phi(c) = a$  and  $\phi(d) = b$ . If  $g = f \circ \phi$ , show that  $L(g) = L(f)$ .

8. Let  $Y$  be a metric space. Show that if  $f : \mathbf{R} \rightarrow Y$  is periodic and continuous, then either  $f$  is constant or there exists a smallest positive period of  $f$ .

9. Show that if  $f : \mathbf{R} \rightarrow \mathbf{C}$  has periods  $T_1$  and  $T_2$ , then  $mT_1 + nT_2$  is a period of  $f$  for every pair of integers  $m, n$ . Deduce that if  $f$  is continuous, then either  $f$  is constant or  $T_1/T_2$  is rational.

10. Let  $A$  be the collection of all finite subsets of  $\mathbf{N}$ . We define a collection of functions,  $\{W_\alpha : \alpha \in A\}$ , called the Walsh functions, as follows. We put  $W_\emptyset = 1$ . If  $\alpha$  is a nonempty finite subset of  $\mathbf{N}$ , say  $\alpha = \{k_1, \dots, k_n\}$  ( $k_1 < k_2 < \dots < k_n$ ) we define  $W_\alpha = r_{k_1} r_{k_2} \cdots r_{k_n}$ , where  $(r_k)$  is the sequence of Rademacher functions.

(a) Show that  $\{W_\alpha : \alpha \in A\}$  is an orthonormal set in  $\mathcal{R}$ .

(b) Show that if  $n$  is a positive integer, and  $f \in \mathcal{R}$  is constant on each interval  $[(k-1)2^{-n}, k2^{-n}]$  ( $k \in \mathbf{Z}$ ), then  $f$  is a linear combination of a finite number of Walsh functions.

(c) Show that if  $f \in \mathcal{R}$ , and  $\langle f, W_\alpha \rangle = 0$  for every  $\alpha \in A$ , then  $\|f\|_2 = 0$ .

11. Calculate the Fourier coefficients of the function  $f \in \mathcal{R}$  defined by  $f(t) = t$  for  $0 \leq t < 1$ . Use Parseval's theorem to deduce the value of  $\sum_{n=1}^{\infty} n^{-2}$ .

12. Calculate the Fourier coefficients of the function  $g \in \mathcal{C}$  given by  $g(t) = |t|$  for  $-\frac{1}{2} \leq t \leq \frac{1}{2}$ . Use your result to check your answer in the last exercise for  $\sum_{n=1}^{\infty} n^{-2}$ , and also to evaluate  $\sum_{n=1}^{\infty} n^{-4}$ .

13. Let  $f \in \mathcal{R}$ , and  $s_n(x) = \sum_{k=-n}^n \hat{f}_k e^{2\pi i k x}$ . Show that

$$s_n(x) = \int_{-1/2}^{1/2} f(x-t) D_n(t) dt,$$

where

$$D_n(t) = \frac{\sin(2n+1)\pi t}{\sin \pi t}.$$

14. Show that if  $f \in \mathbf{R}$ , and  $f(t) = 0$  for all  $t \in (a, b)$ , then the Fourier series of  $f$  converges to 0 uniformly in any closed interval  $[c, d] \subset (a, b)$ . HINT: Use the last exercise.

15. Show that if  $f \in \mathcal{R}$  has a continuous derivative in an interval  $(a, b)$ , then the Fourier series of  $f$  converges to  $f$  uniformly in any closed interval  $[c, d] \subset (a, b)$ . HINT: Use the last exercise, and Theorem 7.29.

16. Let  $\sigma_n(x) = (1/(n+1)) \sum_{k=0}^n s_k(x)$ , where  $s_k$  is the  $k$ th partial sum of the Fourier series of  $f$ . Show that

$$\sigma_n(x) = \int_{-1/2}^{1/2} f(x-t) K_n(t) dt,$$

where

$$K_n(t) = \frac{1}{n+1} \frac{\sin^2(2n+1)\pi t}{\sin^2 \pi t}.$$

17. With the notation of the last exercise, show that if  $f \in \mathcal{C}$ , then  $(\sigma_n)$  converges uniformly to  $f$ . This provides another proof of Theorem 7.30.

18. Show that if  $f, g \in \mathcal{R}$ , then

$$\langle f, g \rangle = \sum_{n \in \mathbf{Z}} \hat{f}_n \overline{\hat{g}_n}.$$

## 7.6 Notes

7.1 Weierstrass published Theorem 7.1 in 1886; he may well have known it for some time. The proof given is that of Lebesgue (1898). There

are many other proofs; especially worth citing is that of Serge Bernstein (1912), whose proof draws on ideas from elementary probability theory. The Weierstrass theorem was generalized in many ways over the years before Stone obtained Theorem 7.6 in 1937. The proofs of Theorems 7.5 and 7.6 given here were published by Stone in 1947. Other proofs, and generalizations of these theorems, have since been given. One of the attractive early generalizations of the Weierstrass theorem was found by Müntz in 1914, and improved upon by Szász a year or two later. It says that if  $0 = p_0 < p_1 < p_2 < \dots$  is a sequence of real numbers, then the set of all linear combinations  $\sum_{k=0}^n c_k x^{p_k}$  is dense in  $C([0, 1])$  if and only if  $\sum_{k=1}^{\infty} p_k^{-1} = +\infty$ .

- 7.2 In a later chapter, we will consider the length of a smooth (continuously differentiable) curve, as a special case ( $k = 1$ ) of the volume of a  $k$ -dimensional manifold.
- 7.3 The subject of trigonometric series, and especially Fourier series, is a vast one, and we have given only the smallest taste of this subject. It has played an important role in mathematical history ever since Fourier asserted, in the early years of the nineteenth century, that “any” function could be expanded in a Fourier series. Among the significant developments spurred by Fourier’s ideas (which were not received with enthusiasm by the mathematical establishment of the day) was the modern concept of function, which probably began with Dirichlet, and the whole theory of sets, initiated by Cantor to deal with problems about trigonometric series. A good place to learn more is Körner’s book [8]. Fourier theory continues to be a central area in mathematical research today.
- 7.4 The theorem which is now called Weyl’s theorem was discovered independently around the year 1909 by Bohl, Sierpiński, and Weyl. It is a special case of the ergodic theorem.



# Differentiable Maps

If  $f$  is a real-valued function on an interval  $(a, b)$ , we recall that  $f$  is differentiable at the point  $c \in (a, b)$  if

$$\lim_{t \rightarrow c} \frac{f(t) - f(c)}{t - c}$$

exists; if it does, we denote its value by  $f'(c)$ , and call it the derivative of  $f$  at  $c$ . In the last chapter, we remarked that this definition makes sense for a complex-valued function on  $(a, b)$ . If  $\mathbf{f} : (a, b) \rightarrow \mathbf{R}^n$  is a vector-valued function on  $(a, b)$ , the same definition still makes perfect sense. If  $\mathbf{f} = (f_1, \dots, f_n)$ , we see that  $\mathbf{f}'$  exists if and only if  $f'_j$  exists for each  $j$ ,  $1 \leq j \leq n$ , and that in this case,  $\mathbf{f}' = (f'_1, \dots, f'_n)$ . If we try to extend the definition in another direction, however, we run into trouble. If  $f$  is a real-valued function defined in some neighborhood of the point  $\mathbf{c} \in \mathbf{R}^n$ , the definition above makes no sense for  $n > 1$ , since we can't divide by vectors. We are led to the right idea by focusing our attention not on the number  $f'(\mathbf{c})$ , which as we know represents the slope of the tangent line to the graph of  $f$ , but on the tangent line itself, which is the graph of a linear function.

Linear maps (synonyms: mappings, transformations) from  $\mathbf{R}^n$  to  $\mathbf{R}^m$  can be regarded as an especially simple subset of the set of all mappings from  $\mathbf{R}^n$  to  $\mathbf{R}^m$ ; in this chapter, we begin by studying linear maps (i.e., reviewing linear algebra). We then find that a fairly wide class of maps can be approximated locally by linear maps, and thereby become accessible to analysis.

## 8.1 Linear Algebra

In this section, we review some of the basic notions concerning linear mappings of vector spaces. With the possible exception of the last part of this section, all of the material should be quite familiar.

Let us recall some basic notions and results from linear algebra. Let  $V$  and  $W$  be vector spaces over the real field  $\mathbf{R}$ , and  $L : V \rightarrow W$ . We say that  $L$  is *linear* if  $L(sv + tw) = sLv + tLw$  for all  $s, t \in \mathbf{R}$  and  $\mathbf{v}, \mathbf{w} \in V$ . (It is customary with linear maps to write  $L\mathbf{v}$  instead of  $L(\mathbf{v})$ , and to write  $LM$  for the composition  $L \circ M$  of  $L$  and  $M$ .) Let  $L : V \rightarrow W$  be a linear map, and suppose that  $\mathbf{v}_1, \dots, \mathbf{v}_n$  is a basis for  $V$ . Thus each element  $\mathbf{v}$  of  $V$  has a unique expression as a linear combination of  $\mathbf{v}_1, \dots, \mathbf{v}_n$ ,

$$\mathbf{v} = \sum_{j=1}^n v^j \mathbf{v}_j,$$

where we have used superscripts rather than subscripts to keep track of the components of  $\mathbf{v}$ . It follows that  $L\mathbf{v} = \sum v^j L\mathbf{v}_j$ . If also  $\mathbf{w}_1, \dots, \mathbf{w}_m$  is a basis for  $W$ , we can write each vector  $L\mathbf{v}_j$  as a linear combination of  $\mathbf{w}_1, \dots, \mathbf{w}_m$ ,

$$L\mathbf{v}_j = \sum_{i=1}^m a_j^i \mathbf{w}_i \quad (j = 1, \dots, n)$$

and thus find that

$$L\mathbf{v} = \sum_{j=1}^n v^j L\mathbf{v}_j = \sum_{j=1}^n v^j \sum_{i=1}^m a_j^i \mathbf{w}_i = \sum_{i=1}^m \left( \sum_{j=1}^n a_j^i v^j \right) \mathbf{w}_i. \quad (8.1)$$

It is customary to visualize the set  $A = \{a_j^i : i = 1, \dots, m; j = 1, \dots, n\}$  of  $mn$  real numbers as a rectangular array, with  $m$  rows and  $n$  columns, the number  $a_j^i$  occupying the position in the  $i$ th row (from the top) and  $j$ th column (from the left). Such an array is called an  $m \times n$  *matrix*, and in this case, we speak of the matrix of  $L$ , with respect to the bases  $\mathbf{v}_1, \dots, \mathbf{v}_n$  and  $\mathbf{w}_1, \dots, \mathbf{w}_m$ . Thus,

$$A = [a_j^i] = \begin{bmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \cdots & a_n^m \end{bmatrix}.$$

We can write  $A = [L]$  to indicate that  $A$  is the matrix associated to the linear transformation  $L$ ; this notation implies a prior understanding of which bases  $\mathbf{v}_1, \dots, \mathbf{v}_n$  and  $\mathbf{w}_1, \dots, \mathbf{w}_m$  for  $V$  and  $W$  are being used. When  $V = \mathbf{R}^n$  and  $W = \mathbf{R}^m$ , there are the so-called "natural" bases: let  $\mathbf{e}_j$  denote the  $n$ -tuple (or  $m$ -tuple) having 1 in the  $j$ th position and 0's elsewhere.

When  $L$  is a linear mapping from  $\mathbf{R}^n$  to  $\mathbf{R}^m$ ,  $[L]$  will denote the matrix of  $L$  with respect to the natural bases, unless some other bases are specifically indicated. Each choice of basis for  $V$  gives rise to an isomorphism of  $V$  with the space  $\mathbf{R}^n$ ; our superscript-subscript notation for matrices implies that the  $n$ -tuple  $[\mathbf{v}] = (v^1, \dots, v^n) \in \mathbf{R}^n$ , corresponding to the vector  $\mathbf{v} = \sum v^j \mathbf{v}_j \in V$ , is to be thought of as a *column vector*, i.e., a matrix with  $n$  rows and 1 column.

Suppose that  $U$  is a third vector space, with basis  $\mathbf{u}_1, \dots, \mathbf{u}_p$ , and that  $M: U \rightarrow V$  is another linear map, with the corresponding matrix  $[M] = B = [b_j^i]$ . Then the composition  $LM$  is a linear map of  $U$  into  $W$ , and we find

$$\begin{aligned}(LM)\mathbf{u}_j &= L(M\mathbf{u}_j) = L\left(\sum_{k=1}^n b_j^k \mathbf{v}_k\right) \\ &= \sum_{k=1}^n b_j^k \sum_{i=1}^m a_k^i \mathbf{w}_i \\ &= \sum_{i=1}^m \left(\sum_{k=1}^n a_k^i b_j^k\right) \mathbf{w}_i,\end{aligned}$$

which gives rise to the notion of *matrix product*: if  $A = [a_j^i]$  is an  $m \times n$  matrix, and  $B = [b_j^i]$  is an  $n \times p$  matrix, their product  $AB$  is defined to be the  $m \times p$  matrix  $C = [c_j^i]$ , where  $c_j^i = \sum_{k=1}^n a_k^i b_j^k$ . This "multiplication law" for matrices gives the desirable formula  $[LM] = [L][M]$  when  $L$  and  $M$  are linear mappings whose composition is defined, and  $[L]$ ,  $[M]$ ,  $[LM]$  are the corresponding matrices with respect to some choice of basis for each of the three spaces involved. Since the associative law  $L(MN) = (LM)N$  is an immediate consequence of the definition of composition, the associative law for matrix multiplication follows.

We note that the multiplication law applies to the product of an  $m \times n$  matrix and an  $n \times 1$  vector, and that  $L\mathbf{v}$  is represented by the column vector  $[L\mathbf{v}] = [L][\mathbf{v}]$ . In particular, when  $V = \mathbf{R}^n$  and  $W = \mathbf{R}^m$ , and we use the natural bases, we cannot distinguish between  $\mathbf{v}$  and  $[\mathbf{v}]$ , nor between  $L\mathbf{v} = L(\mathbf{v})$  and the matrix product  $[L][\mathbf{v}]$ .

We note that if  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  is linear, then the matrix of  $f$  is a  $1 \times n$  matrix, i.e., a row vector:  $[f] = [f_1, \dots, f_n]$ , where  $f_j = f(\mathbf{e}_j)$  for  $j = 1, \dots, n$ . In this case,  $f(\mathbf{v}) = [f][\mathbf{v}]$  is the product of the  $1 \times n$  matrix  $[f]$  with the  $n \times 1$  matrix  $[\mathbf{v}]$ . If we associate to  $[f]$  the column vector

$$\mathbf{f} = [f]^t = \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{bmatrix}$$

we see that  $f(\mathbf{v}) = \mathbf{f} \cdot \mathbf{v}$  for all  $\mathbf{v} \in \mathbf{R}^n$ , where  $\mathbf{f} \cdot \mathbf{v}$  denotes the usual inner

product in  $\mathbf{R}^n$ . In general, if  $A = [a_j^i]$  is an  $m \times n$  matrix, we denote by  $A^t$  the  $n \times m$  matrix  $[b_j^i]$ , where  $b_j^i = a_i^j$ , for  $1 \leq j \leq n$ ,  $1 \leq i \leq m$ ;  $A^t$  is called the *transpose* of  $A$ .

**8.1 Definition.** Let  $L: V \rightarrow W$  be a linear transformation, where  $V$  and  $W$  are vector spaces of dimension  $n$  and  $m$ , respectively. The sets

$$\mathcal{R} = \mathcal{R}_L = \{\mathbf{w} \in W : \mathbf{w} = L\mathbf{v} \text{ for some } \mathbf{v} \in V\}$$

and

$$\mathcal{N} = \mathcal{N}_L = \{\mathbf{v} \in V : L\mathbf{v} = \mathbf{0}\}$$

are called the *image* and *kernel* of  $L$ , respectively.

The image and kernel of  $L$  are also referred to as the *range* and *null space* of  $L$ , respectively. It is easy to see that  $\mathcal{R}$  and  $\mathcal{N}$  are subspaces of  $W$  and  $V$ , respectively. The fundamental theorem of linear algebra is perhaps the following:

**8.2 Theorem.** With the notations just introduced,

$$\dim \mathcal{R} + \dim \mathcal{N} = \dim V.$$

**Proof.** Let  $\mathbf{v}_1, \dots, \mathbf{v}_k$  be a basis for  $\mathcal{N}$ ; we may extend it to a basis for  $V$ , i.e., there exist  $\mathbf{v}_{k+1}, \dots, \mathbf{v}_n$  such that  $\mathbf{v}_1, \dots, \mathbf{v}_n$  is a basis for  $V$ . Let  $\mathbf{w}_j = L\mathbf{v}_j$  for  $j = k+1, \dots, n$ . I claim that  $\mathbf{w}_{k+1}, \dots, \mathbf{w}_n$  form a basis of  $\mathcal{R}$ . First of all, they span  $\mathcal{R}$ ; if  $\mathbf{w} \in \mathcal{R}$ , then  $\mathbf{w} = L\mathbf{v}$  for some  $\mathbf{v} \in V$ , and we may write  $\mathbf{v} = \sum_{j=1}^n v^j \mathbf{v}_j$  for some scalars  $v^1, \dots, v^n$ . Then, since  $L\mathbf{v}_j = \mathbf{0}$  for  $j = 1, \dots, k$ , we have

$$\mathbf{w} = L\mathbf{v} = \sum_{j=1}^n v^j L\mathbf{v}_j = \sum_{j=k+1}^n v^j L\mathbf{v}_j = \sum_{j=k+1}^n v^j \mathbf{w}_j.$$

Thus,  $\mathbf{w}_{k+1}, \dots, \mathbf{w}_n$  span  $\mathcal{R}$ . Also, these vectors form a linearly independent set; for if  $c^{k+1}, \dots, c^n$  are scalars such that  $\sum_{j=k+1}^n c^j \mathbf{w}_j = \mathbf{0}$ , then we have

$$L\left(\sum_{j=k+1}^n c^j \mathbf{v}_j\right) = \sum_{j=k+1}^n c^j \mathbf{w}_j = \mathbf{0},$$

so  $\sum_{j=k+1}^n c^j \mathbf{v}_j \in \mathcal{N}$ . But then there exist scalars  $a^1, \dots, a^k$  such that  $\sum_{j=k+1}^n c^j \mathbf{v}_j = \sum_{j=1}^k a^j \mathbf{v}_j$ . Since  $\mathbf{v}_1, \dots, \mathbf{v}_n$  are linearly independent, this implies that every  $c^j$  (as well as every  $a^j$ ) is 0. ■

The dimension of the kernel of  $L$  is called the *nullity* of  $L$ , and the dimension of the image of  $L$  is called the *rank* of  $L$ . The mapping  $L$  is called *nonsingular* if it has nullity 0; this is equivalent to  $L$  being injective (one-one). The mapping  $L$  is surjective (onto) if and only if it has rank  $m$ . We deduce immediately from the last theorem:

**8.3 Corollary.** Suppose that  $L : V \rightarrow W$  is a linear mapping, where  $\dim V = \dim W = n$ . Then  $L$  is injective if and only if  $L$  is surjective; nonsingularity is equivalent to invertibility.

Let  $\mathcal{L}(V, W)$  denote the set of all linear transformations from the vector space  $V$  to the vector space  $W$ ; we write  $\mathcal{L}(V)$  for  $\mathcal{L}(V, V)$ . The assignment of a matrix to each linear transformation enables us to regard each element of  $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$  as a point of Euclidean space  $\mathbf{R}^{nm}$ , and thus we can speak of open sets in  $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ , of continuous functions of linear transformations, etc. The usual metric on  $\mathbf{R}^{nm}$  can be described in matrix terms as follows: the distance from  $S$  to  $T$  is  $\|S - T\|_{\text{tr}}$ , where the “trace norm” of the linear transformation  $T$  is defined by

$$\|T\|_{\text{tr}}^2 = \sum_{i,j} (a_j^i)^2 = \text{tr } A^t A,$$

where  $A = [a_j^i]$  is the  $m \times n$  matrix  $[T]$ ,  $A^t$  denotes the transpose of  $A$ , and  $\text{tr } B$  denotes the *trace* of the square matrix  $B$ :  $\text{tr } B = \sum_j b_j^j$ .

A perhaps more natural way to define the distance between linear transformations is by using the so-called “operator norm” defined by

$$\|T\| = \sup\{|T\mathbf{v}| : \mathbf{v} \in \mathbf{R}^n, |\mathbf{v}| \leq 1\}. \quad (8.2)$$

It is not hard to verify that this definition is equivalent to

$$\|T\| = \inf\{C \geq 0 : |T\mathbf{v}| \leq C|\mathbf{v}| \text{ for all } \mathbf{v} \in \mathbf{R}^n\}. \quad (8.3)$$

The finiteness of  $\|T\|$  is easy to see, and it is obvious that  $\|T\| = 0$  if and only if  $T = 0$ . In fact, we have the following comparison:

**8.4 Lemma.** If  $T \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ , then

$$\|T\| \leq \|T\|_{\text{tr}} \leq \sqrt{n}\|T\|.$$

**Proof.** Let  $A = [a_k^j]$  be the matrix of  $T$ . Let  $\mathbf{v} \in \mathbf{R}^n$ , with  $|\mathbf{v}| = 1$ . If  $\mathbf{v} = \sum_k v^k \mathbf{e}_k$ , so  $\sum_k (v^k)^2 = 1$ , then

$$\begin{aligned} |T\mathbf{v}|^2 &= \left| \sum_{j=1}^m \sum_{k=1}^n a_k^j v^k \mathbf{e}_j \right|^2 = \sum_{j=1}^m \left( \sum_{k=1}^n a_k^j v^k \right)^2 \\ &\leq \sum_{j=1}^m \left( \sum_{k=1}^n (a_k^j)^2 \right) \left( \sum_{k=1}^n (v^k)^2 \right) \quad (\text{Cauchy inequality}) \\ &= \sum_{j,k} (a_k^j)^2 = \|T\|_{\text{tr}}^2, \end{aligned}$$

so  $\|T\| \leq \|T\|_{\text{tr}}$ . On the other hand, since  $\sum_{j=1}^m (a_k^j)^2 = |T\mathbf{e}_k|^2 \leq \|T\|^2$  for every  $k$ ,  $1 \leq k \leq n$ , we have

$$\|T\|_{\text{tr}}^2 = \sum_{k=1}^n \sum_{j=1}^m (a_k^j)^2 \leq n\|T\|^2,$$

proving the other inequality. ■

We leave it to the reader to show that the sup and inf in (8.2) and (8.3) are actually max and min, i.e., that they are attained. The next proposition lists the basic properties of the operator norm. We note that these properties are shared by the trace norm. We leave the proof to the reader.

**8.5 Proposition.** *For every  $S, T \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ ,  $R \in \mathcal{L}(\mathbf{R}^m, \mathbf{R}^l)$ ,  $c \in \mathbf{R}$ , we have:*

- (a)  $\|S + T\| \leq \|S\| + \|T\|$ ;
- (b)  $\|cT\| = |c| \|T\|$ ; and
- (c)  $\|RS\| \leq \|R\| \|S\|$ .

**8.6 Proposition.** *The space  $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ , with the distance function  $\rho$  defined by  $\rho(S, T) = \|S - T\|$ , is a complete metric space.*

**Proof.** Property (a) of Proposition 8.5 shows that the distance function satisfies the triangle inequality; the other properties of a metric are trivially satisfied. Lemma 8.4 shows that a sequence in  $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$  is Cauchy, or is convergent, in the metric defined by  $\|\cdot\|$  if and only if it is Cauchy (resp., convergent) in the metric defined by  $\|\cdot\|_{\text{tr}}$ . Since  $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$  is certainly complete in the metric defined by the trace norm (the usual Euclidean metric of  $\mathbf{R}^{nm}$ ), it follows that it is complete in the metric defined by the operator norm. ■

The next proposition is interesting, though not essential for our application of linear algebra to the study of more general mappings. As usual,  $I$  will denote the identity mapping of  $\mathbf{R}^n$ .

**8.7 Proposition.** *If  $L \in \mathcal{L}(\mathbf{R}^n)$ , and  $\|I - L\| < 1$ , then  $L$  is invertible, and  $\|I - L^{-1}\| \leq \|I - L\| (1 - \|I - L\|)^{-1}$ .*

**Proof.** Let  $T = I - L$ , so  $\|T\| = t < 1$ . Let  $S_m = \sum_{k=0}^m T^k$ ; then

$$\begin{aligned} \|S_m - S_{m+p}\| &= \left\| \sum_{k=m+1}^{m+p} T^k \right\| \leq \sum_{k=m+1}^{m+p} \|T^k\| \\ &\leq \sum_{k=m+1}^{\infty} \|T\|^k = \frac{t^{m+1}}{1-t} \end{aligned}$$

using properties (a) and (c) of Proposition 8.5. Thus,  $\{S_m\}$  is a Cauchy sequence in  $\mathcal{L}(\mathbf{R}^n)$ ; it follows from Proposition 8.6 that  $S_m \rightarrow S$  for some  $S \in \mathcal{L}(\mathbf{R}^n)$ . In particular, taking  $m = 0$  above, we find

$$\|I - S_p\| \leq \frac{t}{1-t}$$

for every  $p$ , and hence  $\|I - S\| \leq t/(1-t)$ . Now  $LS_m = (I - T)S_m = I - T^{m+1}$ , so

$$\|I - LS\| = \lim_{m \rightarrow \infty} \|I - LS_m\| = \lim_{m \rightarrow \infty} \|T^{m+1}\| = 0,$$

i.e.,  $LS = I$ . Since  $LS_m = S_m L$  for every  $m$ , we have also  $SL = I$ . ■

**8.8 Corollary.** *Let  $S, T \in \mathcal{L}(\mathbf{R}^n)$ . If  $T$  is invertible, and  $\|S - T\| < \|T^{-1}\|^{-1}$ , then  $S$  is invertible, and*

$$\|S^{-1} - T^{-1}\| \leq \|S - T\| \frac{C^2}{1 - C\|S - T\|},$$

where  $C = \|T^{-1}\|$ .

**Proof.** Since

$$\|I - ST^{-1}\| = \|(T - S)T^{-1}\| \leq \|T - S\| \|T^{-1}\| < 1,$$

it follows from Proposition 8.7 that  $ST^{-1}$  is invertible. If  $U = (ST^{-1})^{-1}$ , we have  $S(T^{-1}U) = (ST^{-1})U = I$ , so  $T^{-1}U$  is a right inverse for  $S$ ; also, the equation  $U(ST^{-1}) = I$  gives  $US = T$  which leads to  $T^{-1}US = I$ , so  $T^{-1}U$  is also a left inverse for  $S$ . Now  $\|S^{-1} - T^{-1}\| = \|T^{-1}U - T^{-1}\| \leq \|T^{-1}\| \|I - U\|$ ; from Theorem 8.7 we have the estimate  $\|I - U\| \leq t/(1-t)$ , where  $t = \|I - ST^{-1}\|$ ; since  $t = \|T^{-1}(T - S)\| \leq \|T^{-1}\| \|T - S\|$ , we easily deduce the estimate of the corollary for  $\|S^{-1} - T^{-1}\|$ . ■

This corollary says that the set of invertible elements is an open subset  $U$  of  $\mathcal{L}(\mathbf{R}^n)$ , and that the map  $T \mapsto T^{-1}$  is a continuous map of  $U$  onto itself. In fact, the proof shows that  $T^{-1}$  can be expressed as a power series in  $T$ ; on the matrix level, this says that each entry in  $[T^{-1}]$  can be expressed as a power series in the  $n^2$  entries of  $[T]$ . Another way to see these facts is to use the theory of determinants; a matrix  $A$  is invertible if and only if  $\det A \neq 0$ , and  $\det A$ , being a polynomial in the coefficients of  $A$ , is certainly a continuous function on  $\mathcal{L}(\mathbf{R}^n)$ , so  $\{A : \det A \neq 0\}$  is open. Furthermore, the coefficients of  $A^{-1}$  can be expressed (using Cramer's rule) as rational functions of the coefficients of  $A$ , so the map  $A \mapsto A^{-1}$  is a very nice kind of continuous function.

## 8.2 Differentials

The following definition of differentiability seems to have been first used by Maurice Fréchet, in the early years of this century, and then in the context of mappings between infinite-dimensional vector spaces. It is the universally accepted definition today.

**8.9 Definition.** Let  $U \subset \mathbf{R}^n$ , and  $f : U \rightarrow \mathbf{R}^m$ . We say that  $f$  is differentiable at the point  $\mathbf{p} \in U$  if  $\mathbf{p}$  is an interior point of  $U$ , and there exists a linear mapping  $L : \mathbf{R}^n \rightarrow \mathbf{R}^m$  such that

$$\lim_{\mathbf{h} \rightarrow 0} \frac{1}{|\mathbf{h}|} (f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) - L\mathbf{h}) = 0. \quad (8.4)$$

**Remark.** The linear map  $L$  of the above definition is unique, when it exists. For if  $M$  is another linear map such that (8.4) holds, then

$$\lim_{\mathbf{h} \rightarrow 0} \frac{1}{|\mathbf{h}|} (L\mathbf{h} - M\mathbf{h}) = 0;$$

but then for any fixed  $\mathbf{k} \neq 0$ , and  $t > 0$ , we have

$$\frac{1}{|t\mathbf{k}|} (L(t\mathbf{k}) - M(t\mathbf{k})) = \frac{1}{|\mathbf{k}|} (L\mathbf{k} - M\mathbf{k}),$$

so letting  $t \rightarrow 0$  we conclude that  $L\mathbf{k} = M\mathbf{k}$ ; since  $\mathbf{k}$  was an arbitrary nonzero element of  $\mathbf{R}^n$ , we conclude that  $L = M$ .

We say that  $T : \mathbf{R}_n \rightarrow \mathbf{R}^m$  is an *affine map* if it has the form  $\mathbf{x} \mapsto \mathbf{c} + L\mathbf{x}$ , for some  $\mathbf{c} \in \mathbf{R}^m$  and linear map  $L : \mathbf{R}^n \rightarrow \mathbf{R}^m$ . Thus Definition 8.9 says roughly that a function is differentiable at  $\mathbf{p}$  if it can be approximated near  $\mathbf{p}$  by an affine function.

**8.10 Definition.** If  $f$  is differentiable at  $\mathbf{p}$ , we denote the linear map  $L$  of Definition 8.9 by  $df_{\mathbf{p}}$ , and its matrix with respect to the standard bases of  $\mathbf{R}^n$  and  $\mathbf{R}^m$  by  $f'(\mathbf{p})$ . We call  $df_{\mathbf{p}}$  the differential of  $f$  at  $\mathbf{p}$ .

Notation and terminology vary widely in this subject. Some authors refer to  $df_{\mathbf{p}}$  as the *derivative* of  $f$  at  $\mathbf{p}$ , or the Fréchet derivative;  $df$  is sometimes denoted by  $Df$ , or by  $f'$ , the symbol we have reserved for its matrix. The matrix  $f'(\mathbf{p})$  is often called the *Jacobian matrix* of  $f$  at  $\mathbf{p}$ ; its determinant (when  $n = m$ ) is the *Jacobian determinant*, or simply the Jacobian, of  $f$  at  $\mathbf{p}$ , and denoted by  $J_f(\mathbf{p})$ .

It is easy to see that if  $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$  is itself a linear map, then  $f$  is differentiable everywhere, and  $df_{\mathbf{p}} = f$  at every point  $\mathbf{p}$ . It is also easy to see that when  $m = n = 1$ , the definitions we have given agree with the familiar ones. Suppose  $n = 1$ , so  $f$  is a vector-valued function of one



variable:  $\mathbf{f}(t) = (f_1(t), \dots, f_m(t))$ . Then we see that  $\mathbf{f}'(t)$  is an  $m \times 1$  matrix, or column vector; in terms of coordinates, we see that

$$\mathbf{f}'(t) = \begin{bmatrix} f'_1(t) \\ f'_2(t) \\ \vdots \\ f'_m(t) \end{bmatrix},$$

so that our definition reduces to the previously familiar one. The next result is also extremely simple, and we omit the proof.

**8.11 Proposition.** *If  $\mathbf{f}$  and  $\mathbf{g}$  are differentiable at  $\mathbf{p}$ , then so is  $\mathbf{f} + \mathbf{g}$ , and  $d(\mathbf{f} + \mathbf{g})_{\mathbf{p}} = d\mathbf{f}_{\mathbf{p}} + d\mathbf{g}_{\mathbf{p}}$ ; if  $\mathbf{f}$  is differentiable at  $\mathbf{p}$  and  $c$  is a constant, then  $c\mathbf{f}$  is differentiable at  $\mathbf{p}$ , and  $d(c\mathbf{f})_{\mathbf{p}} = cd\mathbf{f}_{\mathbf{p}}$ .*

**8.12 Proposition.** *Let  $\mathbf{f} : U \rightarrow \mathbf{R}^m$ , where  $U$  is an open set in  $\mathbf{R}^n$ . If  $\mathbf{f} = (f_1, \dots, f_m)$ , then  $\mathbf{f}$  is differentiable at  $\mathbf{p}$  if and only if each  $f_j$  ( $j = 1, \dots, m$ ) is differentiable at  $\mathbf{p}$ , and in this case we have*

$$d\mathbf{f}(\mathbf{h}) = (df_1(\mathbf{h}), \dots, df_m(\mathbf{h}))$$

(where both sides are to be evaluated at  $\mathbf{p}$ ), or in terms of matrices,

$$\mathbf{f}'(\mathbf{p}) = \begin{bmatrix} f'_1(\mathbf{p}) \\ f'_2(\mathbf{p}) \\ \vdots \\ f'_m(\mathbf{p}) \end{bmatrix}$$

(here, each  $f'_j(\mathbf{p})$  is a  $1 \times n$  matrix, i.e., a row vector of length  $n$ ).

This proposition is easy to prove, using the inequalities

$$\max_{1 \leq j \leq m} |a_j| \leq |\mathbf{a}| \leq \sum_{j=1}^m |a_j|,$$

which make statements about limits of vectors equivalent to statements about limits of their coordinates. We next observe that differentiability at a point implies continuity at that point.

**8.13 Proposition.** *If  $\mathbf{f}$  is differentiable at  $\mathbf{p}$ , and  $C > \|d\mathbf{f}_{\mathbf{p}}\|$ , then there exists  $\delta > 0$  such that  $|\mathbf{f}(\mathbf{p} + \mathbf{h}) - \mathbf{f}(\mathbf{p})| \leq C|\mathbf{h}|$  for all  $|\mathbf{h}| \leq \delta$ . In particular,  $\mathbf{f}$  is continuous at  $\mathbf{p}$ .*

**Proof.** Let  $\epsilon = C - \|d\mathbf{f}_{\mathbf{p}}\|$ , so  $\epsilon > 0$ ; by Definition 8.9 there exists  $\delta > 0$  such that

$$|\mathbf{f}(\mathbf{p} + \mathbf{h}) - \mathbf{f}(\mathbf{p}) - d\mathbf{f}_{\mathbf{p}}\mathbf{h}| \leq \epsilon|\mathbf{h}|$$

whenever  $|\mathbf{h}| < \delta$ . It follows that

$$|\mathbf{f}(\mathbf{p} + \mathbf{h}) - \mathbf{f}(\mathbf{p})| \leq \|d\mathbf{f}_{\mathbf{p}}\mathbf{h}\| + \epsilon|\mathbf{h}| \leq (\|d\mathbf{f}_{\mathbf{p}}\| + \epsilon)|\mathbf{h}| = C|\mathbf{h}|,$$

as claimed. ■

**8.14 Proposition.** *Let  $U$  be an open set in  $\mathbf{R}^n$ . Suppose  $f : U \rightarrow \mathbf{R}$  has a local maximum or minimum at  $\mathbf{p} \in U$ , and that  $f$  is differentiable at  $\mathbf{p}$ . Then  $df_{\mathbf{p}} = 0$ .*

**Proof.** Suppose that  $f$  has a local maximum at  $\mathbf{p}$ , so that there exists  $\delta > 0$  such that  $f(\mathbf{q}) \leq f(\mathbf{p})$  for all  $\mathbf{q} \in U$  with  $|\mathbf{q} - \mathbf{p}| < \delta$ . Let  $L = df_{\mathbf{p}}$ ; then  $f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) = L\mathbf{h} + r(\mathbf{h})$ , where  $r(\mathbf{h})/|\mathbf{h}| \rightarrow 0$  as  $\mathbf{h} \rightarrow 0$ . For any  $\mathbf{v} \in \mathbf{R}^n$ , and  $t > 0$  sufficiently small, we find (taking  $\mathbf{h} = t\mathbf{v}$  above) that  $L(t\mathbf{v}) + r(t\mathbf{v}) \leq 0$ , or  $L\mathbf{v} \leq r(t\mathbf{v})/t$ , so letting  $t \rightarrow 0$  we have  $L\mathbf{v} \leq 0$ ; replacing  $\mathbf{v}$  by  $-\mathbf{v}$ , we also find that  $L\mathbf{v} \geq 0$ , so  $L\mathbf{v} = 0$ . Since  $\mathbf{v}$  was an arbitrary vector in  $\mathbf{R}^n$ , we have shown that  $L = 0$ , as desired. The case where  $f$  has a local minimum reduces to the case we considered by looking at the function  $-f$ . ■

The next result is known as the chain rule; when  $m = n = 1$ , it reduces to the chain rule of elementary calculus.

**8.15 Theorem.** *Let  $U \subset \mathbf{R}^n$ ,  $V \subset \mathbf{R}^m$ , and let  $\mathbf{f} : U \rightarrow \mathbf{R}^m$  and  $\mathbf{g} : V \rightarrow \mathbf{R}^k$ . Suppose that  $\mathbf{f}$  is differentiable at  $\mathbf{p} \in U$ , that  $\mathbf{q} = \mathbf{f}(\mathbf{p}) \in V$ , and that  $\mathbf{g}$  is differentiable at  $\mathbf{q}$ . Then  $\mathbf{g} \circ \mathbf{f}$  is differentiable at  $\mathbf{p}$ , and  $d(\mathbf{g} \circ \mathbf{f})_{\mathbf{p}} = d\mathbf{g}_{\mathbf{q}} \circ d\mathbf{f}_{\mathbf{p}}$ .*

**Proof.** Since  $\mathbf{f}$  is differentiable at  $\mathbf{p}$ , and  $\mathbf{g}$  is differentiable at  $\mathbf{q} = \mathbf{f}(\mathbf{p})$ , we know that  $\mathbf{p}$  is an interior point of  $U$  and  $\mathbf{q}$  is an interior point of  $V$ , so  $\mathbf{p}$  is an interior point of the domain  $U \cap \mathbf{f}^{-1}(V)$  of  $\mathbf{g} \circ \mathbf{f}$ . Let  $T = d\mathbf{f}_{\mathbf{p}}$  and  $S = d\mathbf{g}_{\mathbf{q}}$ ; we want to show that, setting

$$\mathbf{r}(\mathbf{h}) = (\mathbf{g} \circ \mathbf{f})(\mathbf{p} + \mathbf{h}) - (\mathbf{g} \circ \mathbf{f})(\mathbf{p}) - STh,$$

we have  $\mathbf{r}(\mathbf{h})/|\mathbf{h}| \rightarrow 0$  as  $\mathbf{h} \rightarrow 0$ . Let

$$\mathbf{r}_{\mathbf{f}}(\mathbf{h}) = \mathbf{f}(\mathbf{p} + \mathbf{h}) - \mathbf{f}(\mathbf{p}) - T\mathbf{h}$$

for all  $\mathbf{h}$  such that  $\mathbf{p} + \mathbf{h} \in U$ , and

$$\mathbf{r}_{\mathbf{g}}(\mathbf{k}) = \mathbf{g}(\mathbf{q} + \mathbf{k}) - \mathbf{g}(\mathbf{q}) - S\mathbf{k}$$

for all  $\mathbf{k}$  such that  $\mathbf{q} + \mathbf{k} \in V$ . Then by Definition 8.9,  $|\mathbf{r}_{\mathbf{f}}(\mathbf{h})|/|\mathbf{h}| \rightarrow 0$  as  $\mathbf{h} \rightarrow 0$ , and  $|\mathbf{r}_{\mathbf{g}}(\mathbf{k})|/|\mathbf{k}| \rightarrow 0$  as  $\mathbf{k} \rightarrow 0$ . Let  $\mathbf{k} = \mathbf{k}(\mathbf{h}) = \mathbf{f}(\mathbf{p} + \mathbf{h}) - \mathbf{f}(\mathbf{p})$ . Then  $|\mathbf{k}| \leq C|\mathbf{h}|$  for  $|\mathbf{h}|$  sufficiently small, if  $C > \|T\|$ , by Proposition 8.13; it follows that

$$\frac{\mathbf{r}_{\mathbf{g}}(\mathbf{k}(\mathbf{h}))}{|\mathbf{h}|} \rightarrow 0$$

as  $\mathbf{h} \rightarrow \mathbf{0}$ . Now

$$\begin{aligned} (\mathbf{g} \circ \mathbf{f})(\mathbf{p} + \mathbf{h}) - (\mathbf{g} \circ \mathbf{f})(\mathbf{p}) &= \mathbf{g}(\mathbf{q} + \mathbf{k}) - \mathbf{g}(\mathbf{q}) = S(\mathbf{k}) + \mathbf{r}_{\mathbf{g}}(\mathbf{k}) \\ &= S(T\mathbf{h} + \mathbf{r}_{\mathbf{f}}(\mathbf{h})) + \mathbf{r}_{\mathbf{g}}(\mathbf{k}), \end{aligned}$$

so that

$$\mathbf{r}(\mathbf{h}) = S(\mathbf{r}_{\mathbf{f}}(\mathbf{h})) + \mathbf{r}_{\mathbf{g}}(\mathbf{k})$$

and hence

$$\frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|} \leq \|S\| \frac{|\mathbf{r}_{\mathbf{f}}(\mathbf{h})|}{|\mathbf{h}|} + \frac{|\mathbf{r}_{\mathbf{g}}(\mathbf{k})|}{|\mathbf{h}|}$$

which  $\rightarrow 0$  as  $\mathbf{h} \rightarrow \mathbf{0}$ . ■

## 8.3 The Mean Value Theorem

We have seen in Chapter 4 that one of the most useful theoretical tools in dealing with derivatives of functions of one variable is the mean value theorem. Here is a version of the mean value theorem for real-valued functions of several variables.

**8.16 Definition.** If  $\mathbf{a}, \mathbf{b} \in \mathbf{R}^n$ , we put

$$[\mathbf{a}, \mathbf{b}] = \{(1-t)\mathbf{a} + t\mathbf{b} : 0 \leq t \leq 1\},$$

and refer to  $[\mathbf{a}, \mathbf{b}]$  as the line segment from  $\mathbf{a}$  to  $\mathbf{b}$ .

A subset  $E$  of  $\mathbf{R}^n$  is called convex if  $[\mathbf{a}, \mathbf{b}] \subset E$  whenever  $\mathbf{a} \in E$  and  $\mathbf{b} \in E$ .

When  $n = 1$ , this notation conflicts with the standard usage of  $[a, b]$  if  $a > b$ ; let us live with the risk.

**8.17 Theorem.** Let  $E \subset \mathbf{R}^n$ , and suppose that  $f : E \rightarrow \mathbf{R}$  is differentiable at each point of  $E$ . If  $[\mathbf{a}, \mathbf{b}] \subset E$ , then there is a point  $\mathbf{c} \in [\mathbf{a}, \mathbf{b}]$  such that  $f(\mathbf{b}) - f(\mathbf{a}) = df_{\mathbf{c}}(\mathbf{b} - \mathbf{a}) = f'(\mathbf{c})(\mathbf{b} - \mathbf{a})$ .

**Proof.** Let  $\phi(t) = t\mathbf{b} + (1-t)\mathbf{a}$ , so that  $[\mathbf{a}, \mathbf{b}] = \{\phi(t) : 0 \leq t \leq 1\}$ . We observe that  $\phi'(t) = \mathbf{b} - \mathbf{a}$ . Let  $g(t) = f(\phi(t))$ ; according to Theorem 8.15,  $g$  is differentiable at each point of  $[0, 1]$ , and  $dg_t = df_{\phi(t)} \circ d\phi_t$ , or  $g'(t) = f'(\phi(t))(\mathbf{b} - \mathbf{a})$ . Applying the mean value theorem, Theorem 4.22, we find there exists  $c$ ,  $0 < c < 1$ , such that  $g'(c) = g(1) - g(0)$ . This translates, with  $\mathbf{c} = \phi(c)$ , to the statement  $f(\mathbf{b}) - f(\mathbf{a}) = f'(\mathbf{c})(\mathbf{b} - \mathbf{a})$ . ■

**8.18 Corollary.** Let  $U$  be a convex open set in  $\mathbf{R}^n$ , and let  $f : U \rightarrow \mathbf{R}$ . Suppose that  $f$  is differentiable at each point of  $U$ , and that  $\|df_{\mathbf{p}}\| \leq M$  for all  $\mathbf{p} \in U$ . Then

$$|f(\mathbf{b}) - f(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|$$

for all  $\mathbf{a}, \mathbf{b} \in U$ .

There is no mean value theorem for vector-valued functions of several variables, but a vector-valued version of the estimate in the last corollary does hold, and is very useful.

**8.19 Corollary.** *Let  $U$  be a convex open set in  $\mathbf{R}^n$  and  $\mathbf{f} : U \rightarrow \mathbf{R}^m$ . If  $\mathbf{f}$  is differentiable at each point of  $U$ , and if  $\|\mathbf{df}_{\mathbf{p}}\| \leq M$  for every  $\mathbf{p} \in U$ , then we have  $|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|$  for all  $\mathbf{a}, \mathbf{b} \in U$ .*

**Proof.** Let  $\mathbf{u} \in \mathbf{R}^m$ ,  $|\mathbf{u}| = 1$ , and define  $\phi : \mathbf{R}^m \rightarrow \mathbf{R}$  by  $\phi(\mathbf{t}) = \mathbf{u} \cdot \mathbf{t}$ . We note that  $d\phi = \phi$  at each point, since  $\phi$  is linear, and  $\|d\phi\| = |\mathbf{u}| = 1$ . Now define  $g = \phi \circ \mathbf{f}$ . Then  $dg_{\mathbf{p}} = d\phi_{\mathbf{f}(\mathbf{p})} \circ \mathbf{df}_{\mathbf{p}}$ , by the chain rule, and so

$$\|dg_{\mathbf{p}}\| \leq \|d\phi\| \|\mathbf{df}_{\mathbf{p}}\| = \|\mathbf{df}_{\mathbf{p}}\| \leq M.$$

Applying the last corollary, we have

$$\mathbf{u} \cdot (\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})) = g(\mathbf{b}) - g(\mathbf{a}) \leq M|\mathbf{b} - \mathbf{a}|$$

for every unit vector  $\mathbf{u} \in \mathbf{R}^m$ . But we may choose such a unit vector  $\mathbf{u}$  so that  $\mathbf{u} \cdot (\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})) = |\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})|$ . ■

**8.20 Corollary.** *Let  $U$  be a connected open set in  $\mathbf{R}^n$ . If  $\mathbf{f} : U \rightarrow \mathbf{R}^m$  is differentiable at each point of  $U$ , and  $\mathbf{f}'(\mathbf{p}) = \mathbf{0}$  for every  $\mathbf{p} \in U$ , then  $\mathbf{f}$  is constant.*

**Proof.** Fix  $\mathbf{p}_0 \in U$ , and let  $V = \{\mathbf{p} \in U : \mathbf{f}(\mathbf{p}) = \mathbf{f}(\mathbf{p}_0)\}$ . Then  $V$  is nonempty, and  $V$  is closed in  $U$  since  $\mathbf{f}$  is continuous. Furthermore,  $V$  is open: for if  $\mathbf{p} \in V$ , we may choose an open ball  $B$  centered at  $\mathbf{p}$  which lies entirely in  $U$ ; the last corollary applies, with  $M = 0$ , to yield  $|\mathbf{f}(\mathbf{q}) - \mathbf{f}(\mathbf{p})| = 0$  for every  $\mathbf{q} \in B$ , so  $B \subset V$ . Thus  $V$  is open, as well as closed. Since  $U$  is connected, it follows that  $V = U$ . ■

## 8.4 Partial Derivatives

Let  $U$  be an open set in  $\mathbf{R}^n$ , and  $f : U \rightarrow \mathbf{R}$ . If  $\mathbf{p} = (p_1, \dots, p_n) \in U$ , then (for  $j = 1, \dots, n$ ) the function  $t \mapsto f(p_1, \dots, p_{j-1}, t, p_{j+1}, \dots, p_n)$  is defined in an interval  $(p_j - \delta, p_j + \delta)$  for some  $\delta > 0$ , so it makes sense to ask if it has a derivative at  $p_j$ ; if it does, we call that derivative the  $j$ th partial derivative of  $f$  at  $\mathbf{p}$ , or the partial derivative with respect to the  $j$ th coordinate function  $x_j$ , and denote it by  $D_j f(\mathbf{p})$ , or by  $(\partial f / \partial x_j)(\mathbf{p})$ . We may define the partial derivatives of a vector-valued function, i.e., a map  $\mathbf{f} : U \rightarrow \mathbf{R}^m$ , in the same way. In other words,

$$D_j \mathbf{f}(\mathbf{p}) = \lim_{h \rightarrow 0} \frac{\mathbf{f}(\mathbf{p} + h\mathbf{e}_j) - \mathbf{f}(\mathbf{p})}{h},$$

provided this limit exists.

**8.21 Proposition.** Let  $U$  be open in  $\mathbf{R}^n$ , and suppose that  $\mathbf{f} : U \rightarrow \mathbf{R}^m$  is differentiable at  $\mathbf{p} \in U$ . Then  $D_j \mathbf{f}(\mathbf{p})$  exists for each  $j$ ,  $1 \leq j \leq n$ , and  $d\mathbf{f}_{\mathbf{p}}(\mathbf{h}) = \sum_{j=1}^n h^j D_j \mathbf{f}(\mathbf{p})$ , for  $\mathbf{h} = (h^1, \dots, h^n)$ .

**Proof.** If  $L = d\mathbf{f}_{\mathbf{p}}$ , we are given that

$$\mathbf{f}(\mathbf{p} + h\mathbf{e}_j) - \mathbf{f}(\mathbf{p}) = L(h\mathbf{e}_j) + \mathbf{r}(h),$$

where  $\mathbf{r}(h)/h \rightarrow \mathbf{0}$  as  $h \rightarrow 0$ . Since  $L$  is linear, we have

$$\frac{\mathbf{f}(\mathbf{p} + h\mathbf{e}_j) - \mathbf{f}(\mathbf{p})}{h} = L(\mathbf{e}_j) + \frac{\mathbf{r}(h)}{h},$$

which shows that  $D_j \mathbf{f}(\mathbf{p})$  exists and equals  $L\mathbf{e}_j$ . It follows that

$$L\mathbf{h} = L\left(\sum_{j=1}^n h^j \mathbf{e}_j\right) = \sum_{j=1}^n h^j L\mathbf{e}_j = \sum_{j=1}^n h^j D_j \mathbf{f}(\mathbf{p}),$$

as claimed. ■

This proposition can also be formulated as a description of the matrix  $\mathbf{f}'(\mathbf{p})$  of the linear transformation  $d\mathbf{f}_{\mathbf{p}}$  associated to the map  $\mathbf{f} = (f_1, \dots, f_m)$ , namely,

$$\mathbf{f}'(\mathbf{p}) = [D_1 \mathbf{f}(\mathbf{p}) \quad D_2 \mathbf{f}(\mathbf{p}) \quad \cdots \quad D_n \mathbf{f}(\mathbf{p})],$$

where  $D_j \mathbf{f}$  is the column vector  $[D_j f_1, \dots, D_j f_m]^t$ , that is

$$(\mathbf{f}'(\mathbf{p}))_j^i = D_j f_i(\mathbf{p}) = \frac{\partial f_i}{\partial x_j},$$

where the left-hand side denotes the entry in the  $i$ th row and  $j$ th column of the matrix  $\mathbf{f}'(\mathbf{p})$ . Let us express the chain rule (Theorem 8.15) in terms of partial derivatives. In the situation of Theorem 8.15, if  $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ , we have  $\mathbf{h}'(\mathbf{p}) = \mathbf{g}'(\mathbf{f}(\mathbf{p})) \mathbf{f}'(\mathbf{p})$ , and thus

$$D_j h_i(\mathbf{p}) = \sum_{k=1}^m (D_k g_i)(\mathbf{f}(\mathbf{p}))(D_j f_k)(\mathbf{p}).$$

With the classical notation, it is customary to write  $y_k = f_k(x_1, \dots, x_n)$  and  $z_j = g_j(y_1, \dots, y_m)$ , and write the chain rule (suppressing the points at which functions are to be evaluated) in the form

$$\frac{\partial z_i}{\partial x_j} = \sum_{k=1}^m \frac{\partial z_i}{\partial y_k} \frac{\partial y_k}{\partial x_j}.$$

**8.22 Example.** The converse of Proposition 8.21 is false: the existence of the partial derivatives of  $f$  at a point, or even throughout an entire neighborhood of a point, does not imply that  $f$  is differentiable at the point. Indeed, it does not even imply that  $f$  is continuous at the point. Consider the example

$$f(s, t) = \begin{cases} \frac{2st}{s^2 + t^2} & \text{if } s^2 + t^2 > 0; \\ 0 & \text{if } s = t = 0. \end{cases}$$

It is easy to see that  $f$  is differentiable at each point other than the origin, and that  $D_1 f(0, 0) = D_2 f(0, 0) = 0$ , since  $f(s, 0) = 0 = f(0, t)$  for all real  $s$  and  $t$ . But  $f(t, t) = 1$  for all  $t \neq 0$ , so  $f$  is not even continuous at  $(0, 0)$ .

However, with a slightly stronger assumption, we can deduce differentiability.

**8.23 Theorem.** Let  $U$  be open in  $\mathbf{R}^n$ , and let  $f : U \rightarrow \mathbf{R}$ . Suppose that for each  $j$ ,  $1 \leq j \leq n$ ,  $D_j f$  exists in a neighborhood of  $\mathbf{p} \in U$ , and is continuous at  $\mathbf{p}$ . Then  $f$  is differentiable at  $\mathbf{p}$ .

**Proof.** Let  $r(\mathbf{h}) = f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) - \sum_{j=1}^n h^j D_j f(\mathbf{p})$ ; we must show that  $r(\mathbf{h})/|\mathbf{h}| \rightarrow 0$  as  $\mathbf{h} \rightarrow \mathbf{0}$ . Let  $\mathbf{k}_j = \sum_{i=1}^j h^i \mathbf{e}_i$  for  $j = 1, \dots, n$ , and let  $\mathbf{k}_0 = \mathbf{0}$ . Thus  $\mathbf{k}_1 = h^1 \mathbf{e}_1$  and  $\mathbf{k}_n = \mathbf{h}$ . By applying the mean value theorem in one variable  $n$  times, we find

$$\begin{aligned} r(\mathbf{h}) &= \sum_{j=1}^n [f(\mathbf{p} + \mathbf{k}_j) - f(\mathbf{p} + \mathbf{k}_{j-1}) - h^j D_j f(\mathbf{p})] \\ &= \sum_{j=1}^n [f(\mathbf{p} + \mathbf{k}_{j-1} + h^j \mathbf{e}_j) - f(\mathbf{p} + \mathbf{k}_{j-1}) - h^j D_j f(\mathbf{p})] \\ &= \sum_{j=1}^n h^j [D_j f(\mathbf{q}_j) - D_j f(\mathbf{p})], \end{aligned}$$

where  $\mathbf{q}_j = \mathbf{p} + \mathbf{k}_{j-1} + \theta^j h^j \mathbf{e}_j$  for some  $\theta^j \in (0, 1)$ . By hypothesis, we can, given any  $\epsilon > 0$ , find  $\delta > 0$  such that  $|D_j f(\mathbf{q}) - D_j f(\mathbf{p})| < \epsilon/n$  whenever  $|\mathbf{q} - \mathbf{p}| < \delta$ . Now if  $|\mathbf{h}| < \delta$ , it follows that  $|\mathbf{k}_j| < \delta$  for each  $j$ , and hence that  $|\mathbf{q}_j - \mathbf{p}| < \delta$  for each  $j$ . We conclude that

$$|r(\mathbf{h})| \leq \frac{\epsilon}{n} \sum_{j=1}^n |h^j| \leq \epsilon |\mathbf{h}|$$

whenever  $|\mathbf{h}| < \delta$ . ■

If  $f : \Omega \rightarrow \mathbf{R}$ , then  $D_j f(\mathbf{p})$  exists for  $\mathbf{p}$  in some (possibly empty) subset  $U$  of  $\Omega$ , so it is possible that  $D_k D_j f(\mathbf{p})$  exists at some  $\mathbf{p} \in U$ , for some

$k$ . This is a *second-order partial derivative*. Similarly, we can have partial derivatives of order  $r$  for any positive integer  $r$ ; such a partial derivative will have the form  $D_{j_1} D_{j_2} \cdots D_{j_r} f$ , where each  $j_i$  is an integer from  $\{1, \dots, n\}$ . It is customary to write  $D_j^2 f$  for  $D_j D_j f$ , etc. The notation

$$\frac{\partial^r f}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_r}}$$

is also commonly used. The following result makes it much easier to deal with higher order partial derivatives.

**8.24 Theorem.** Let  $\Omega$  be an open set in  $\mathbf{R}^n$ , let  $f : \Omega \rightarrow \mathbf{R}$ , and suppose that  $D_i f$ ,  $D_j f$  and  $D_i D_j f$  exist in a neighborhood of  $\mathbf{p} \in \Omega$ , and that  $D_i D_j f$  is continuous at  $\mathbf{p}$ . Then  $D_j D_i f(\mathbf{p})$  exists, and, in fact,  $D_j D_i f(\mathbf{p}) = D_i D_j f(\mathbf{p})$ .

**Proof.** We may assume without loss of generality that  $n = 2$ ,  $i = 1$ , and  $j = 2$ . Let  $\mathbf{p} = (a, b)$ , and choose  $\delta > 0$  so that  $(s, t) \in \Omega$  whenever  $|s - a| < \delta$  and  $|t - b| < \delta$ . Let

$$\Delta(h, k) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b). \quad (8.5)$$

Then

$$\lim_{h \rightarrow 0} \lim_{k \rightarrow 0} \frac{\Delta(h, k)}{hk} = \lim_{h \rightarrow 0} \frac{D_2 f(a + h, b) - D_2 f(a, b)}{h} = D_1 D_2 f(a, b),$$

and similarly

$$\lim_{k \rightarrow 0} \lim_{h \rightarrow 0} \frac{\Delta(h, k)}{hk} = D_2 D_1 f(a, b),$$

provided, of course, that this limit exists. Let  $g(t) = f(a + h, t) - f(a, t)$ . Then  $g$  is differentiable on the interval  $[b, b + k]$  (or  $[b + k, b]$  if  $k < 0$ ), and  $g'(t) = D_2 f(a + h, t) - D_2 f(a, t)$ . Now, by the mean value theorem (Theorem 4.22), we have

$$\Delta(h, k) = g(b + k) - g(b) = kg'(b + \mu k)$$

for some  $0 < \mu < 1$ . Another application of the mean value theorem gives

$$D_2 f(a + h, b + \mu k) - D_2 f(a, b + \mu k) = h D_1 D_2 f(a + \lambda h, b + \mu k)$$

for some  $0 < \lambda < 1$ . Combining these last two equations, we get

$$\frac{\Delta(h, k)}{hk} = D_1 D_2 f(a + \lambda h, b + \mu k).$$

Since  $D_1 D_2 f$  is continuous at  $(a, b)$ , we see that

$$\lim_{(h, k) \rightarrow (0, 0)} \frac{\Delta(h, k)}{hk} = D_1 D_2 f(a, b),$$

and this is more than we need to obtain the desired conclusion. ■

**8.25 Definition.** Let  $f : \Omega \rightarrow \mathbf{R}$ , where  $\Omega$  is an open set in  $\mathbf{R}^n$ . We say that  $f$  is of class  $C^r$  in  $\Omega$ , or  $f \in C^r(\Omega)$ , if all the partial derivatives of  $f$  of order  $r$  exist and are continuous throughout  $\Omega$ . (This implies that all partial derivatives of order less than or equal to  $r$  exist and are continuous.) We say that  $f$  is of class  $C^\infty$  if  $f$  is of class  $C^r$  for every positive integer  $r$ . We define  $C^0(\Omega)$  to be the set of all continuous functions on  $\Omega$ .

Thus

$$C^0(\Omega) \supset C^1(\Omega) \supset \cdots \supset C^r(\Omega) \supset C^{r+1}(\Omega)$$

and  $C^\infty(\Omega) = \bigcap_{r=0}^\infty C^r(\Omega)$ . Similarly, we say that a vector-valued function  $\mathbf{f}$  is of class  $C^r$  if  $\mathbf{f} = (f_1, \dots, f_m)$  where each  $f_j$  is of class  $C^r$ .

If  $f$  is of class  $C^r$ ,  $r > 1$ , then by a repeated application of Theorem 8.24 any mixed partial derivative of order  $r$  of  $f$  can be brought into a standard form:

$$D_{j_1} D_{j_2} \cdots D_{j_r} f = D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f,$$

where the nonnegative integers  $\alpha_i$  are obtained by  $\alpha_i = \#\{k : j_k = i\}$ . It is clear that  $\sum_{i=1}^n \alpha_i = r$  here. In the context of partial derivatives, an  $n$ -tuple  $\alpha = (\alpha_1, \dots, \alpha_n)$  of nonnegative integers is called a *multi-index*, and  $k = \sum \alpha_i$  is called the *order* of  $\alpha$ , and written  $|\alpha|$ . The corresponding partial derivative has multi-index notations:

$$D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f = D^\alpha f = \frac{\partial^\alpha f}{\partial x^\alpha}.$$

We note that for any multi-index  $\alpha$ ,  $|\alpha| = r$ , there are many ways to write  $D^\alpha$  as  $D_{j_1} \cdots D_{j_r}$ . How many ways? Well, we may select the  $\alpha_1$  positions where  $j_k = 1$  in

$$\binom{r}{\alpha_1} = \frac{r!}{\alpha_1! (r - \alpha_1)!}$$

ways; the  $\alpha_2$  positions where  $j_k = 2$  may be chosen from among the remaining  $r - \alpha_1$  available spots in

$$\binom{r - \alpha_1}{\alpha_2} = \frac{(r - \alpha_1)!}{\alpha_2! (r - \alpha_1 - \alpha_2)!}$$

ways, and so on. Combining the possibilities, we find that the number of distinct ways of expressing  $D^\alpha$  as a sequence of first-order partial derivatives is

$$\binom{r}{\alpha} = \frac{r!}{\alpha!} = \frac{r!}{\alpha_1! \alpha_2! \cdots \alpha_n!},$$

which is called a *multinomial coefficient*.

The multi-index notation makes it easier to deal with polynomials in  $n$  variables; if  $\mathbf{x} = (x_1, \dots, x_n)$ , we write  $\mathbf{x}^\alpha$  as shorthand for the monomial  $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ . Thus, the general polynomial in  $n$  variables, of degree  $\leq r$ , has the form  $\sum_{|\alpha| \leq r} c_\alpha \mathbf{x}^\alpha$ .

The following theorem, known as Taylor's theorem in  $n$  variables, may be regarded as a generalization of the mean value theorem.



**8.26 Theorem.** Let  $U$  be an open convex set in  $\mathbf{R}^n$ , and suppose that  $f: U \rightarrow \mathbf{R}$  is a function of class  $C^{r+1}$ ,  $r \geq 0$ . If  $\mathbf{p} \in U$  and  $\mathbf{q} = \mathbf{p} + \mathbf{h} \in U$ , we have the Taylor formula: there exists  $\theta$ ,  $0 < \theta < 1$ , such that

$$f(\mathbf{q}) = \sum_{|\alpha| \leq r} \frac{D^\alpha f(\mathbf{p})}{\alpha!} \mathbf{h}^\alpha + \sum_{|\alpha| = r+1} \frac{D^\alpha f(\theta \mathbf{h})}{\alpha!} \mathbf{h}^\alpha. \quad (8.6)$$

**Proof.** Fix  $\mathbf{p} + \mathbf{h} \in U$ , where  $\mathbf{h} = (h_1, \dots, h_n)$ , and define  $g(t) = f(\mathbf{p} + t\mathbf{h})$  for  $0 \leq t \leq 1$ . Then  $g$  is of class  $C^{r+1}$  on an interval containing  $[0, 1]$ , so by Taylor's theorem for functions of one variable (Theorem 4.33), we have

$$f(\mathbf{p} + \mathbf{h}) = g(1) = \sum_{k=1}^r \frac{g^{(k)}(0)}{k!} + \frac{g^{(r+1)}(\theta)}{(r+1)!} \quad (8.7)$$

for some  $\theta$ ,  $0 < \theta < 1$ . But according to the chain rule,

$$g'(t) = \sum_{i=1}^n D_i f(\mathbf{p} + t\mathbf{h}) h_i,$$

so

$$g''(t) = \sum_{j=1}^n D_j \left( \sum_{i=1}^n D_i f(\mathbf{p} + t\mathbf{h}) h_i \right) h_j = \sum_{i,j=1}^n D_j D_i f(\mathbf{p} + t\mathbf{h}) h_i h_j,$$

and, in general,

$$g^{(k)}(t) = \sum_{j_1, \dots, j_k} D_{j_1} D_{j_2} \cdots D_{j_k} f(\mathbf{p} + t\mathbf{h}) h_{j_1} h_{j_2} \cdots h_{j_k} \quad (8.8)$$

$$= \sum_{|\alpha|=k} \binom{k}{\alpha} D^\alpha f(\mathbf{p} + t\mathbf{h}) \mathbf{h}^\alpha. \quad (8.9)$$

Substituting (8.8) into equation (8.7), we obtain the desired formula. ■

The first term on the right in equation (8.6) is called the *Taylor polynomial of order  $r$  for  $f$  at  $\mathbf{p}$* ; it is the unique polynomial  $P_r(\mathbf{x})$  of degree  $\leq r$  with the property  $D^\alpha P_r(\mathbf{p}) = D^\alpha f(\mathbf{p})$  for every  $\alpha$  with  $|\alpha| \leq r$ . The proof is left as an exercise. Other versions of Taylor's theorem give different expressions for the remainder term  $f(\mathbf{q}) - P_r(\mathbf{q})$ .

## 8.5 Inverse and Implicit Functions

Let  $U$  be an open set in  $\mathbf{R}^n$ , and let  $\mathbf{f}: U \rightarrow \mathbf{R}^m$  be differentiable at  $\mathbf{p} \in U$ . If  $\mathbf{f}$  is injective, and maps  $U$  onto an open set  $V$  in  $\mathbf{R}^m$ , it does not follow that the inverse function  $\mathbf{g}: V \rightarrow U$  is differentiable at  $\mathbf{p}$ . [The

simplest example is with  $n = m = 1$ , and  $f(t) = t^3$ . It is easy to see that  $f$  is everywhere differentiable, and maps  $\mathbf{R}$  bijectively to itself, and that the inverse function  $g$  is continuous everywhere ( $g(t) = t^{1/3}$  for all  $t$ , of course), but  $g$  is not differentiable at 0.] Indeed, it is a simple consequence of the chain rule that if the inverse function  $\mathbf{g}$  to  $\mathbf{f}$  is differentiable at  $\mathbf{q} = \mathbf{f}(\mathbf{p})$ , then  $d\mathbf{f}_{\mathbf{p}}$  must be invertible; for by Theorem 8.15, we have  $d\mathbf{f}(\mathbf{p}) \circ d\mathbf{g}(\mathbf{q}) = I_m$ , and  $d\mathbf{g}(\mathbf{q}) \circ d\mathbf{f}(\mathbf{p}) = I_n$ , where  $I_m$  and  $I_n$  are the identity mappings of  $\mathbf{R}^m$  and  $\mathbf{R}^n$ , respectively.

This incidentally shows that necessarily  $m = n$  in this situation. This is hardly surprising, but the corresponding fact in the context of continuous mappings is rather deep; it turns out to be not at all easy to show that there is no bicontinuous map of an open set  $U$  in  $\mathbf{R}^n$  onto an open set  $V$  in  $\mathbf{R}^m$  when  $n \neq m$ . Differentiable functions are much easier to understand than the most general continuous functions.

The following theorem is called the inverse function theorem:

**8.27 Theorem.** *Let  $\Omega$  be an open set in  $\mathbf{R}^n$ , and let  $\mathbf{f} : \Omega \rightarrow \mathbf{R}^n$  be a mapping of class  $C^r$ ,  $r \geq 1$ . If  $\mathbf{p} \in \Omega$ , and  $d\mathbf{f}_{\mathbf{p}}$  is invertible, then there exists a neighborhood  $U$  of  $\mathbf{p}$  such that  $\mathbf{f}$  is one-one on  $U$ ,  $V = \mathbf{f}(U)$  is open, and the inverse  $\mathbf{g}$  of  $\mathbf{f}|_U$  is of class  $C^r$ .*

**Proof.** The proof is notationally much simpler when we make some simple normalizing assumptions, so we first prove:

**8.28 Lemma.** *Let  $\Omega$  be a neighborhood of  $\mathbf{0} \in \mathbf{R}^n$ , and suppose that  $\mathbf{f} : \Omega \rightarrow \mathbf{R}^n$  is differentiable in  $\Omega$ , with  $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ ,  $d\mathbf{f}_{\mathbf{0}} = I$ , and that  $d\mathbf{f}$  is continuous at  $\mathbf{0}$ . Then there exists a neighborhood  $U$  of  $\mathbf{0}$  such that:*

- (a) *for all  $\mathbf{s}_1, \mathbf{s}_2 \in U$ ,  $|\mathbf{f}(\mathbf{s}_1) - \mathbf{f}(\mathbf{s}_2)| \geq \frac{1}{2}|\mathbf{s}_1 - \mathbf{s}_2|$ ;*
- (b)  *$\mathbf{f}(U)$  contains a neighborhood  $V$  of  $\mathbf{0}$ ; and*
- (c) *there is a function  $\mathbf{g} : V \rightarrow U$  such that  $\mathbf{f}(\mathbf{g}(\mathbf{t})) = \mathbf{t}$  for all  $\mathbf{t} \in V$ , and such that  $\mathbf{g}$  is differentiable at  $\mathbf{0}$ , with  $d\mathbf{g}_{\mathbf{0}} = I$ .*

**Proof.** From the continuity hypothesis, we know there exists  $\delta > 0$  such that  $\|I - d\mathbf{f}_{\mathbf{s}}\| < \frac{1}{2}$  for all  $\mathbf{s}$  with  $|\mathbf{s}| \leq \delta$ . Let  $U = \{\mathbf{s} : |\mathbf{s}| < \delta\}$ . Define  $\mathbf{F}(\mathbf{s}) = \mathbf{s} - \mathbf{f}(\mathbf{s})$  for  $\mathbf{s} \in U$ . Then  $d\mathbf{F} = I - d\mathbf{f}$ , so  $\|d\mathbf{F}_{\mathbf{s}}\| \leq \frac{1}{2}$  for all  $\mathbf{s} \in U$ . Now for every  $\mathbf{s}_1, \mathbf{s}_2 \in U$  it follows from Corollary 8.19 that  $|\mathbf{F}(\mathbf{s}_1) - \mathbf{F}(\mathbf{s}_2)| \leq \frac{1}{2}|\mathbf{s}_1 - \mathbf{s}_2|$ , so

$$\begin{aligned} |\mathbf{f}(\mathbf{s}_1) - \mathbf{f}(\mathbf{s}_2)| &= |\mathbf{s}_1 - \mathbf{s}_2 - [\mathbf{F}(\mathbf{s}_1) - \mathbf{F}(\mathbf{s}_2)]| \\ &\geq |\mathbf{s}_1 - \mathbf{s}_2| - |\mathbf{F}(\mathbf{s}_1) - \mathbf{F}(\mathbf{s}_2)| \\ &\geq |\mathbf{s}_1 - \mathbf{s}_2| - \frac{1}{2}|\mathbf{s}_1 - \mathbf{s}_2| = \frac{1}{2}|\mathbf{s}_1 - \mathbf{s}_2|. \end{aligned}$$

We have established (a).

Choose  $\epsilon$  with  $0 < 4\epsilon < \delta$ , and let  $V = \{\mathbf{t} : |\mathbf{t}| < \epsilon\}$ . Let  $K = \{\mathbf{s} : |\mathbf{s}| \leq 4\epsilon\}$ , so  $K$  is a compact subset of  $U$ . Let  $\mathbf{t} \in V$ . Define the real-valued function  $\varphi$  on  $K$  by  $\varphi(\mathbf{s}) = |\mathbf{f}(\mathbf{s}) - \mathbf{t}|^2$ . Clearly,  $\varphi$  is continuous on the compact set  $K$ , so it assumes a minimum on  $K$ . Now if  $|\mathbf{s}| = 4\epsilon$ , then  $|\mathbf{f}(\mathbf{s})| \geq |\mathbf{s}|/2 = 2\epsilon$ , so  $\varphi(\mathbf{s}) \geq (2\epsilon - |\mathbf{t}|)^2 > \epsilon^2$ . On the other hand,  $\varphi(\mathbf{0}) = |\mathbf{t}|^2 < \epsilon^2$ . Hence  $\varphi$  assumes its minimum value at an interior point  $\mathbf{s}$  of  $K$ . Hence we have  $d\varphi(\mathbf{s}) = 0$ . Now an easy calculation shows that

$$D_j \varphi(\mathbf{s}) = 2 \sum_{i=1}^n (f_i(\mathbf{s}) - t_i) D_j f_i(\mathbf{s})$$

for  $j = 1, \dots, n$ , in other words, that  $d\mathbf{f}_{\mathbf{s}}(\mathbf{f}(\mathbf{s}) - \mathbf{t}) = \mathbf{0}$ . Since  $d\mathbf{f}_{\mathbf{s}}$  is non-singular, this implies that  $\mathbf{f}(\mathbf{s}) = \mathbf{t}$ . Thus  $\mathbf{f}(U) \supset V$ , and we have proved (b).

For each point  $\mathbf{t} \in V$ , we put  $\mathbf{g}(\mathbf{t})$  to be the point  $\mathbf{s} \in U$  for which  $\mathbf{f}(\mathbf{s}) = \mathbf{t}$ ; this point exists by (ii), and is unique in consequence of (i). The inequality of (i) shows that  $|\mathbf{g}(\mathbf{t}_1) - \mathbf{g}(\mathbf{t}_2)| \leq 2|\mathbf{t}_1 - \mathbf{t}_2|$ ; in particular, that  $\mathbf{g}$  is continuous in  $V$ . Now  $\mathbf{f}(\mathbf{h}) = \mathbf{h} + \mathbf{r}(\mathbf{h})$ , where  $\mathbf{r}(\mathbf{h})/|\mathbf{h}| \rightarrow \mathbf{0}$  as  $\mathbf{h} \rightarrow \mathbf{0}$ . Hence

$$|\mathbf{g}(\mathbf{h}) - \mathbf{h}| = |\mathbf{g}(\mathbf{f}(\mathbf{h}) - \mathbf{r}(\mathbf{h})) - \mathbf{g}(\mathbf{f}(\mathbf{h}))| \leq 2|\mathbf{r}(\mathbf{h})|,$$

which shows that  $\mathbf{g}$  is differentiable at  $\mathbf{0}$ , with  $d\mathbf{g}_0 = I$ , i.e., (c) holds. ■

Here is another proof of (b). Choose  $\epsilon$  so that  $0 < 2\epsilon < \delta$ , and put  $V = \{\mathbf{t} : |\mathbf{t}| < \epsilon\}$ ,  $K = \{\mathbf{s} : |\mathbf{s}| \leq 2\epsilon\}$ . Given  $\mathbf{t} \in V$ , define the function  $\mathbf{F}$  by  $\mathbf{F}(\mathbf{s}) = \mathbf{t} + \mathbf{s} - \mathbf{f}(\mathbf{s})$ ; then  $d\mathbf{F} = I - d\mathbf{f}$ , so  $\|d\mathbf{F}_{\mathbf{s}}\| < 1/2$  for all  $\mathbf{s} \in U$ . Hence, by Corollary 8.19, we have

$$|\mathbf{F}(\mathbf{s}_1) - \mathbf{F}(\mathbf{s}_2)| \leq \frac{1}{2}|\mathbf{s}_1 - \mathbf{s}_2| \quad (8.10)$$

for all  $\mathbf{s}_1, \mathbf{s}_2 \in U$ . In particular, if  $\mathbf{s} \in K$ , then  $|\mathbf{F}(\mathbf{s}) - \mathbf{F}(\mathbf{0})| \leq \frac{1}{2}|\mathbf{s}|$ , so  $|\mathbf{F}(\mathbf{s})| \leq \frac{1}{2}(|\mathbf{s}| + |\mathbf{t}|) < 2\epsilon$ ; thus,  $\mathbf{F}$  maps  $K$  into itself. Inequality (8.10) says that  $\mathbf{F}$  is a *contraction mapping* of  $K$  into itself; since  $K$  is complete (as a closed subset of  $\mathbf{R}^n$ ) the contraction mapping theorem (Theorem 6.44) tells us that  $\mathbf{F}$  has a (unique) fixed point  $\mathbf{s} \in K$ , but  $\mathbf{F}(\mathbf{s}) = \mathbf{s}$  means  $\mathbf{f}(\mathbf{s}) = \mathbf{t}$ . Thus  $\mathbf{f}(U) \supset V$ . The advantage of this proof over the one above is that it carries over to the case of differentiable mappings between infinite-dimensional complete normed linear spaces, whereas the proof above relies essentially on the mapping having domain a subset of  $\mathbf{R}^n$ , since it uses the compactness of closed bounded sets.

The proof of Theorem 8.27 is obtained from Lemma 8.28 by an affine change of variables. Let  $\mathbf{q} = \mathbf{f}(\mathbf{p})$ . Define maps  $\phi, \psi : \mathbf{R}^n \rightarrow \mathbf{R}^n$  by  $\phi(\mathbf{x}) =$

$\mathbf{x} - \mathbf{p}$ ,  $\psi(\mathbf{y}) = d\mathbf{f}_{\mathbf{p}}^{-1}(\mathbf{y} - \mathbf{q})$ . Thus  $\phi(\mathbf{p}) = \mathbf{0} = \psi(\mathbf{q})$ ,  $\phi'(\mathbf{p}) = I$ ,  $\psi'(\mathbf{q}) = \mathbf{f}'(\mathbf{p})^{-1}$ . Define  $\mathbf{F}$  on the open set  $\phi(\Omega)$  by  $\mathbf{F} = \psi \circ \mathbf{f} \circ \phi^{-1}$ . Then  $\mathbf{F}(\mathbf{0}) = \mathbf{0}$ ,  $\mathbf{F}'$  is continuous at  $\mathbf{0}$  and  $\mathbf{F}'(\mathbf{0}) = I$ . Applying Lemma 8.28, we find there exist open neighborhoods  $U_0$  and  $V_0$  of  $\mathbf{0}$ , and a map  $\mathbf{G} : V_0 \rightarrow U_0$ , such that  $\mathbf{F} \circ \mathbf{G} = I_{V_0}$ , and  $\mathbf{G}$  is differentiable at  $\mathbf{0}$ . Define  $V = \psi^{-1}(V_0)$  and  $U = \phi^{-1} \circ \mathbf{F}^{-1}(V_0) = \mathbf{f}^{-1}(V)$ . Then  $U$  and  $V$  are open neighborhoods of  $\mathbf{p}$  and  $\mathbf{q}$ , respectively. Define  $\mathbf{g} = \phi^{-1} \circ \mathbf{G} \circ \psi$ ; then  $\mathbf{g}$  is differentiable at  $\mathbf{q}$ , and  $\mathbf{g}(V) = U$ . Also,

$$\mathbf{f} \circ \mathbf{g} = (\psi^{-1} \circ \mathbf{F} \circ \phi)(\phi^{-1} \circ \mathbf{G} \circ \psi) = I_V.$$

Applying the argument with  $\mathbf{p}$  replaced by  $\mathbf{x} \in U$ , we conclude that  $\mathbf{g}$  is differentiable at every  $\mathbf{y} \in V$ . Hence by the chain rule, if  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ , then  $\mathbf{f}'(\mathbf{x})\mathbf{g}'(\mathbf{y}) = I$ , so  $\mathbf{g}' = (\mathbf{f}')^{-1}$ . Since the elements of  $A^{-1}$  are  $C^\infty$  functions of the elements of  $A$  for any matrix  $A$ , it follows that  $\mathbf{g}'$  is of class  $C^k$ , and thus  $\mathbf{g}$  is of class  $C^{k+1}$ , whenever  $\mathbf{f}'$  is of class  $C^k$ . Since  $\mathbf{f}$  is of class  $C^r$ , taking  $k = r - 1$  shows that  $\mathbf{g}$  is also of class  $C^r$ . ■

The following diagram may make it easier to follow the argument just given:

$$\begin{array}{ccc} \mathbf{p} \in U & \xrightarrow{\mathbf{f}} & \mathbf{q} \in V \\ \downarrow \phi & & \downarrow \psi \\ \mathbf{0} \in U_0 & \xrightarrow{\mathbf{F}} & \mathbf{0} \in V_0 \end{array}$$

We close this chapter with the very important implicit function theorem. The reader should first go through the proof of this theorem thinking of the special case  $n = m = 1$ . The notation in the statement of the theorem indicates that we are identifying  $\mathbf{R}^{n+m}$  with  $\mathbf{R}^n \times \mathbf{R}^m$ .

**8.29 Theorem.** Let  $\Omega$  be an open set in  $\mathbf{R}^{n+m}$ , and let  $\mathbf{f} : \Omega \rightarrow \mathbf{R}^m$  be a mapping of class  $C^r$ ,  $r \geq 1$ . Let  $\mathbf{p} = (\mathbf{a}, \mathbf{b}) \in \Omega$ , and let  $\Sigma$  be the "level surface" of  $\mathbf{f}$  through  $\mathbf{p}$ :

$$\Sigma = \{\mathbf{q} \in \Omega : \mathbf{f}(\mathbf{q}) = \mathbf{f}(\mathbf{p})\}.$$

Define  $S \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$  by  $S(\mathbf{h}) = d\mathbf{f}_{\mathbf{p}}(\mathbf{h}, \mathbf{0})$  and  $T \in \mathcal{L}(\mathbf{R}^m)$  by  $T(\mathbf{k}) = d\mathbf{f}_{\mathbf{p}}(\mathbf{0}, \mathbf{k})$ . Suppose that  $T$  is invertible. Then there exist a neighborhood  $U$  of  $\mathbf{p} \in \mathbf{R}^{n+m}$ , a neighborhood  $W$  of  $\mathbf{a}$  in  $\mathbf{R}^n$ , and a mapping  $\mathbf{g} : W \rightarrow \mathbf{R}^m$  which is of class  $C^r$ , such that

$$\Sigma \cap U = \{(\mathbf{s}, \mathbf{g}(\mathbf{s})) : \mathbf{s} \in W\}.$$

Furthermore,  $d\mathbf{g}_{\mathbf{a}} = -T^{-1}S$ .

**Proof.** We may assume  $\mathbf{f}(\mathbf{p}) = \mathbf{0}$ . Define the map  $\mathbf{F} : \Omega \rightarrow \mathbf{R}^{n+m}$  by  $\mathbf{F}(\mathbf{s}, \mathbf{t}) = (\mathbf{s}, \mathbf{f}(\mathbf{s}, \mathbf{t}))$ ; clearly,  $\mathbf{F}$  is of class  $C^r$  in  $\Omega$ , and we can easily verify that

$$d\mathbf{F}(\mathbf{h}, \mathbf{k}) = d\mathbf{F}(\mathbf{h}, \mathbf{0}) + d\mathbf{F}(\mathbf{0}, \mathbf{k}) = (\mathbf{h}, d\mathbf{f}(\mathbf{0}, \mathbf{k})),$$

and, in particular,  $d\mathbf{F}_{\mathbf{p}}(\mathbf{h}, \mathbf{k}) = (\mathbf{h}, T\mathbf{k})$ , so that our hypothesis guarantees that  $d\mathbf{F}_{\mathbf{p}}$  is nonsingular. We can thus apply the inverse function theorem: there exists a neighborhood  $U$  of  $\mathbf{p}$  (contained in  $\Omega$ ), and a neighborhood  $V$  of  $\mathbf{F}(\mathbf{0}) = (\mathbf{a}, \mathbf{0})$  in  $\mathbf{R}^{n+m}$ , such that  $\mathbf{F}$  maps  $U$  one-one onto  $V$ , and such that the inverse mapping  $\mathbf{G}$  of  $V$  onto  $U$  is of class  $C^r$ . We may assume

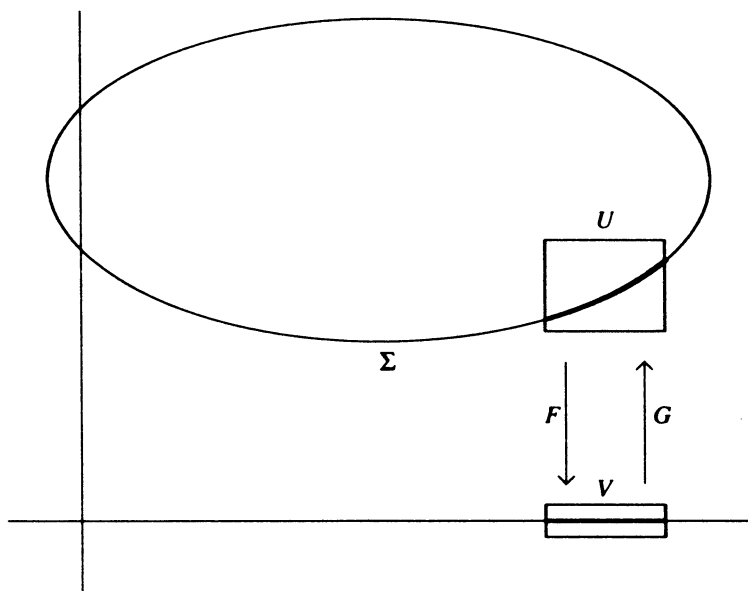


Figure 8.1. The construction in Theorem 8.29

(shrinking if necessary) that  $V$  has the form  $V = \{(\mathbf{s}, \mathbf{t}) : |\mathbf{s} - \mathbf{a}| < \epsilon, |\mathbf{t}| < \epsilon\}$  for some  $\epsilon > 0$ . Let  $W = \{\mathbf{s} \in \mathbf{R}^n : |\mathbf{s} - \mathbf{a}| < \epsilon\}$ . Since  $\mathbf{G}(\mathbf{s}, \mathbf{f}(\mathbf{s}, \mathbf{t})) = (\mathbf{s}, \mathbf{t})$  for all  $(\mathbf{s}, \mathbf{t}) \in U$ , it follows that  $\mathbf{G}$  has the form  $\mathbf{G}(\mathbf{u}, \mathbf{v}) = (\mathbf{u}, \phi(\mathbf{u}, \mathbf{v}))$  for all  $(\mathbf{u}, \mathbf{v}) \in V$ , where  $\phi : V \rightarrow \mathbf{R}^m$  is a mapping of class  $C^r$ . Define  $\mathbf{g} : W \rightarrow \mathbf{R}^m$  by  $\mathbf{g}(\mathbf{s}) = \phi(\mathbf{s}, \mathbf{0})$ . Then, for any  $(\mathbf{s}, \mathbf{t}) \in \Sigma \cap U$ , we have

$$(\mathbf{s}, \mathbf{t}) = \mathbf{G}(\mathbf{s}, \mathbf{0}) = (\mathbf{s}, \phi(\mathbf{s}, \mathbf{0})) = (\mathbf{s}, \mathbf{g}(\mathbf{s})),$$

i.e.,  $\mathbf{t} = \mathbf{g}(\mathbf{s})$ . Conversely, if  $\mathbf{s} \in W$ , then

$$\mathbf{f}(\mathbf{s}, \mathbf{g}(\mathbf{s})) = \mathbf{f}(\mathbf{s}, \phi(\mathbf{s}, \mathbf{0})) = \mathbf{0}.$$

It is a routine application of the chain rule to find an expression for the differential of the explicit function  $\mathbf{g}$ . The equation  $\mathbf{f}(\mathbf{s}, \mathbf{g}(\mathbf{s})) = \mathbf{0}$  which holds throughout a neighborhood  $W$  of  $\mathbf{a}$  can be read as  $\mathbf{f} \circ \mathbf{H} \equiv \mathbf{0}$ , where

$\mathbf{H}(\mathbf{s}) = (\mathbf{s}, \mathbf{g}(\mathbf{s})) = \mathbf{G}(\mathbf{s}, 0)$ . Then  $d\mathbf{f}_{\mathbf{p}} \circ d\mathbf{H}_{\mathbf{a}} = 0$ ; but it is easy to see that  $d\mathbf{H}_{\mathbf{a}}(\mathbf{h}) = (\mathbf{h}, d\mathbf{g}_{\mathbf{a}}(\mathbf{h}))$ , so we have

$$d\mathbf{f}_{\mathbf{p}} \circ d\mathbf{H}_{\mathbf{a}}(\mathbf{h}) = S\mathbf{h} + T(d\mathbf{g}_{\mathbf{a}}(\mathbf{h})) = 0,$$

or  $d\mathbf{g}_{\mathbf{a}}\mathbf{h} = -T^{-1}S\mathbf{h}$ . ■

The hypothesis that  $T$  is invertible is equivalent to the assertion that the last  $m$  columns of the matrix  $\mathbf{f}'(\mathbf{p})$  are linearly independent. If this matrix is only assumed to have rank  $m$ , then there exist  $m$  linearly independent columns, and a permutation transformation changes the situation to that hypothesized in the theorem. In terms of matrices, we have  $\mathbf{g}'(\mathbf{a}) = -B^{-1}A$ , where  $A$ , the matrix of  $S$ , is the  $m \times n$  matrix formed from the first  $n$  columns, and  $B$ , the matrix of  $T$ , is the  $m \times m$  matrix formed from the last  $m$  columns, of  $\mathbf{f}'(\mathbf{p})$ .

With the classical notation, this can be expressed as follows: let  $y_j = x_{n+j}$  for  $j = 1, \dots, m$ ; then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = -\left(\frac{\partial \mathbf{f}}{\partial \mathbf{y}}\right)^{-1} \frac{\partial \mathbf{f}}{\partial \mathbf{x}},$$

where

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial(y_1, \dots, y_m)}{\partial(x_1, \dots, x_n)} = \left[ \frac{\partial y_i}{\partial x_j} \right]$$

is the Jacobian matrix of  $\mathbf{y} = \mathbf{g}(\mathbf{x})$  and the right-hand side refers to the matrices

$$\left[ \frac{\partial f_i}{\partial x_j} \right], \quad \left[ \frac{\partial f_i}{\partial y_j} \right],$$

respectively.

## 8.6 Exercises

1. Prove Proposition 8.5.

2. The trace norm of a linear map is usually much easier to compute than the operator norm. Find  $\|L\|$  and  $\|L\|_{\text{tr}}$  when  $L: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  has the matrix

$$[L] = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

3. Let  $T \in \mathcal{L}(\mathbf{R}^n)$ . Show that the series

$$\sum_{k=0}^{\infty} \frac{T^k}{k!}$$

converges to an element of  $\mathcal{L}(\mathbf{R}^n)$ , which we denote by  $e^T$ . Show that if  $S, T \in \mathcal{L}(\mathbf{R}^n)$  and  $ST = TS$ , then  $e^{S+T} = e^S e^T$ . Deduce that  $e^T$  is invertible for every  $T \in \mathcal{L}(\mathbf{R}^n)$ .

4. Show that if  $T \in \mathcal{L}(\mathbf{R}^n)$  and  $\|I - T\| < 1$ , then there exists  $S \in \mathcal{L}(\mathbf{R}^n)$  such that  $e^S = T$ . Find an example of an invertible  $T \in \mathcal{L}(\mathbf{R}^n)$  such that there is no  $S$  with  $e^S = T$ .

5. Let  $\Omega$  be open in  $\mathbf{R}^n$ , and suppose that  $f$  and  $g$  are real-valued functions differentiable at  $\mathbf{p} \in \Omega$ . Show that  $fg$  is differentiable at  $\mathbf{p}$ , and that  $d(fg)_{\mathbf{p}} = f(\mathbf{p})dg_{\mathbf{p}} + g(\mathbf{p})df_{\mathbf{p}}$ .

6. Let  $\mathbf{f} : \mathbf{R} \rightarrow \mathbf{R}^n$  be differentiable, and suppose that  $|\mathbf{f}(t)| = 1$  for all real  $t$ . Show that  $\mathbf{f}'(t) \cdot \mathbf{f}(t) = 0$  for all  $t$ . What is the geometric meaning of this?

7. Let  $p, V, T$  be positive real variables, connected by the relation  $pV = kT$ , where  $k$  is a positive constant. Then each of  $p, V, T$  is an (implicitly defined) function of the other two variables. Show that

$$\frac{\partial p}{\partial V} \frac{\partial V}{\partial T} \frac{\partial T}{\partial p} = -1.$$

Show that this equation holds if we only assume the relation  $F(p, V, T) = 0$  for all  $p, V, T$  in the domain of  $F$ , for some function  $F$  of class  $C^1$  with  $D_j F(p, V, T) \neq 0$  for all  $p, V, T$  and  $j = 1, 2, 3$ .

[The aim of this problem, found in calculus texts, is to demonstrate that the Leibniz "fraction" notation for derivatives can lead to paradoxical-looking results when used for partial derivatives; it also intends to induce a preference in the reader for the notion of function over the notion of variable. The equation is the ideal gas law from freshman physics.]

8. If  $f : \Omega \rightarrow \mathbf{R}$ , where  $\Omega$  is open in  $\mathbf{R}^n$ , and if  $\mathbf{u}$  is a unit vector in  $\mathbf{R}^n$ , the *directional derivative* of  $f$  at  $\mathbf{p}$  in the direction  $\mathbf{u}$  is defined as

$$D_{\mathbf{u}}f(\mathbf{p}) = \lim_{h \rightarrow 0+} \frac{f(\mathbf{p} + h\mathbf{u}) - f(\mathbf{p})}{h}.$$

(a) Show that if  $f$  is differentiable at  $\mathbf{p}$ , then  $D_{\mathbf{u}}f(\mathbf{p})$  exists for every unit vector  $\mathbf{u}$ , and  $D_{\mathbf{u}}f(\mathbf{p}) = df_{\mathbf{p}}\mathbf{u}$ .

(b) By considering the function  $f : \mathbf{R}^2 \rightarrow \mathbf{R}$  defined by

$$f(s, t) = \frac{s^2 t}{s^4 + t^2}$$

for  $(s, t) \neq (0, 0)$ , and  $f(0, 0) = 0$ , show that  $f$  need not be differentiable at  $\mathbf{p}$  even if  $D_{\mathbf{u}}f(\mathbf{p})$  exists for every unit vector  $\mathbf{u}$ .

9. A function  $f : \mathbf{R}^n \setminus \{0\} \rightarrow \mathbf{R}$  is called *homogeneous of degree  $k$*  if  $f(t\mathbf{x}) = t^k f(\mathbf{x})$  for every  $\mathbf{x} \in \mathbf{R}^n$ ,  $\mathbf{x} \neq 0$ , and every  $t \in \mathbf{R}$ ,  $t > 0$ . Show that if  $f$  is homogeneous of degree  $k$  and differentiable, then  $\sum_{j=1}^n x^j D_j f(\mathbf{x}) = k f(\mathbf{x})$ .

10. Show that if  $f : \Omega \rightarrow \mathbf{R}$ , where  $\Omega$  is an open subset of  $\mathbf{R}^n$ , and if each partial derivative  $D_j f$  exists at every point of  $\Omega$  and is bounded, then  $f$  is continuous in  $\Omega$ . HINT: Imitate the proof of Theorem 8.23.

11. Let  $\mathbf{f} : \mathbf{R}^2 \setminus 0 \rightarrow \mathbf{R}^2$  be defined by

$$\mathbf{f}(s, t) = \left( \frac{s^2 - t^2}{s^2 + t^2}, \frac{st}{s^2 + t^2} \right).$$

Find the rank of  $d\mathbf{f}_{\mathbf{p}}$  for all  $\mathbf{p} \neq 0$ . Describe the image of  $\mathbf{f}$ .

12. Define  $f$  on  $\mathbf{R}^2$  by  $f(0, 0) = 0$ , and

$$f(s, t) = \frac{st(s^2 - t^2)}{s^2 + t^2}$$

for  $(s, t) \neq (0, 0)$ . Show that  $f$  is of class  $C^1$  in  $\mathbf{R}^2$ , and that the mixed partial derivatives  $D_1 D_2 f$  and  $D_2 D_1 f$  exist at every point of  $\mathbf{R}^2$ , but that  $D_1 D_2 f(0, 0) \neq D_2 D_1 f(0, 0)$ .

13. Let  $\alpha = (\alpha_1, \dots, \alpha_n)$  and  $\beta = (\beta_1, \dots, \beta_n)$  be multi-indexes, and let  $\mathbf{x} = (x_1, \dots, x_n)$  be the identity map of  $\mathbf{R}^n$  (i.e., let  $x_j$  be the  $j$ th coordinate function on  $\mathbf{R}^n$ , for  $j = 1, \dots, n$ ). Observe that  $dx_j = [0, \dots, 1, \dots, 0]$  (the row vector with all elements 0 except for 1 in the  $j$ th place), and that  $df = \sum_{j=1}^n D_j f dx_j$ . Show that

$$D^\alpha \mathbf{x}^\beta(0) = \begin{cases} \alpha! & \text{if } \alpha = \beta; \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

Deduce that if  $c_\alpha \in \mathbf{R}$  for each  $\alpha$  with  $|\alpha| \leq r$ , then

$$P(\mathbf{x}) = \sum_{|\alpha| \leq r} \frac{c_\alpha}{\alpha!} \mathbf{x}^\alpha$$

is the unique polynomial of degree  $\leq r$  satisfying  $D^\alpha P(0) = c_\alpha$  for each  $\alpha$  with  $|\alpha| \leq r$ .

14. Show that continuity of the differential is essential in the inverse function theorem by considering the function  $f : (-1, 1) \rightarrow \mathbf{R}$  defined by  $f(0) = 0$  and  $f(t) = t + 2t^2 \sin(1/t)$  for  $t \neq 0$ . Show that  $f$  is everywhere differentiable, and even that  $f'$  is bounded, that  $f'(0) = 1$ , but that  $f$  is not injective in any neighborhood of 0.

15. Show that if  $U$  is open in  $\mathbf{R}^n$ , and  $\mathbf{f} : U \rightarrow \mathbf{R}^n$  is continuously differentiable, with  $\mathbf{f}'(\mathbf{p})$  nonsingular for every  $\mathbf{p} \in U$ , then  $\mathbf{f}(U)$  is open.



16. Let  $\mathbf{f} : \mathbf{R}^2 \rightarrow \mathbf{R}^2$  be defined by

$$\mathbf{f}(x, y) = (e^x \cos y, e^x \sin y).$$

- (a) Show that  $\mathbf{f}$  is injective on the strip  $\{(x, y) : -\pi < y < \pi\}$ .  
 (b) If  $\mathbf{g}$  is the inverse function, find  $\mathbf{g}'(0, 1)$ .

17. Let  $\mathbf{f} : \mathbf{R}^2 \rightarrow \mathbf{R}^2$  be defined by  $\mathbf{f}(x, y) = (x^2 - y^2, 2xy)$ . Show that  $d\mathbf{f}_{\mathbf{p}}$  is invertible for all  $\mathbf{p} \neq \mathbf{0}$ . Find an explicit formula for the inverse function  $\mathbf{g}$  which the inverse function theorem says exists in a neighborhood  $V$  of  $(1, 0) = \mathbf{f}(1, 0)$ .

18. Define  $\mathbf{g} : \mathbf{R}^2 \rightarrow \mathbf{R}^2$  by  $\mathbf{g}(x, y) = (y \cos x, (x + y) \sin y)$ , and  $\mathbf{f} : \mathbf{R}^2 \rightarrow \mathbf{R}^3$  by  $\mathbf{f}(x, y) = (x^2 - y, 3x - 2y, 2xy + y^2)$ .

- (a) Show that  $\mathbf{g}$  maps a neighborhood of  $(0, \pi/2)$  bijectively to a neighborhood of  $(\pi/2, \pi/2)$ .  
 (b) If  $\mathbf{h} = \mathbf{f} \circ \mathbf{g}^{-1}$ , find the matrix  $\mathbf{h}'(\pi/2, \pi/2)$ .

19. The equations

$$\begin{aligned} uz - 2e^{vz} &= 0, \\ u - x^2 - y^2 &= 0, \\ v^2 - xy \log v - 1 &= 0, \end{aligned}$$

define  $z$  (implicitly) as a function of  $(u, v)$ , and  $(u, v)$  as a function of  $(x, y)$ , thus  $z$  as a function of  $(x, y)$ .

Describe the role of the inverse and implicit function theorems in the above statement, and compute

$$\frac{\partial z}{\partial x}(0, e).$$

(Note that when  $x = 0$  and  $y = e$ ,  $u = e^2$ ,  $v = 1$ , and  $z = 2$ .)

## 8.7 Notes

There are many good texts on linear algebra, and we have run through some familiar results largely to establish some notation and terminology; the last theorem in the first section, and its corollary, have an analytic flavor, and may well have not been mentioned in the usual linear algebra course. Determinants, mentioned in passing in this section, will play an important role later on. It is interesting that the theory of determinants was developed before that of matrices; it has its roots in the eighteenth century (and was studied in China independently of any contact with European mathematics). The Swiss mathematician Cramer found his famous

rule in 1750, but no general definition of determinants seems to exist before Cauchy (1815), and a definition immediately recognizable to modern readers first appeared with Jacobi (1841). Jacobi was above all an analyst, and all determinants to him were functional determinants (which Sylvester (1853) called Jacobians: the name has stuck.) Cayley, also in 1841, enclosed the array in two straight lines, and made major contributions to the theory; a few years later, he and Sylvester created matrices.

I know little about the developments that led to the inverse function theorem and the implicit function theorem, two quite important theorems that seem not to be attributed to any individual. Jacobi had shown that a set of  $n$  functions of  $n$  variables are functionally related if and only if their functional determinant (Jacobian) vanishes identically. The notion of rank of a matrix first appeared in the work of Sylvester some years later.

# 9

## Measures

In Chapter 5 we defined the Riemann integral of a real function  $f$  over a bounded interval  $[a, b]$  by

$$\int_a^b f(x) dx = \lim \sum_{j=1}^n f(\xi_j)(x_j - x_{j-1}),$$

where  $x_{j-1} \leq \xi_j \leq x_j$  for each  $j$ , and the limit is taken over increasingly fine partitions  $a = x_0 < x_1 < \cdots < x_n = b$  of the interval. We found that this limit existed whenever  $f$  was continuous on  $[a, b]$ , in fact, whenever  $f$  was bounded, with a set of discontinuities  $D$  which was “small,” in the sense that for any  $\epsilon > 0$ , there existed a finite collection of open intervals  $\{(a_k, b_k) : k = 1, \dots, r\}$  such that

$$D \subset \bigcup_{k=1}^r (a_k, b_k) \quad \text{and} \quad \sum_{k=1}^r (b_k - a_k) < \epsilon.$$

This is a fairly rich class of functions, including as it does not only every continuous function, but also some functions which have infinitely many, even uncountably many, discontinuities (recall that the Cantor set is small in the above sense.) However, the class of Riemann integrable functions does have at least one glaring weakness: it is not stable under pointwise convergence. That is, if  $f_n$  is Riemann integrable for each  $n$ , and if  $f_n(x) \rightarrow f(x)$  for every  $x$ ,  $a \leq x \leq b$ , it is entirely possible that  $f$  is not Riemann integrable. (For instance, take  $a = 0$  and  $b = 1$ , and set  $f_n(x) = 1$  if  $x = m/n!$  for some integer  $m$ , and  $f_n(x) = 0$  otherwise. Then each  $f_n$  is

Riemann integrable, and  $f_n$  converges pointwise to the function  $f$ , where  $f(x) = 1$  if  $x$  is rational, and  $f(x) = 0$  when  $x$  is irrational. We have seen that  $f$  is not Riemann integrable.) It would be agreeable to have a way of integrating functions which, while giving the same result when applied to continuous functions, has the property that the ordinary limit processes of analysis can be carried out freely. Also, it would be nice to have a less ad hoc treatment of the integration of unbounded functions, or functions on unbounded intervals, than we were able to give before. Finally, we want to develop integration in  $\mathbf{R}^n$  for  $n > 1$ , and the Lebesgue theory that we are about to develop makes it possible to do this with the same ease as for  $n = 1$ . In fact, the theory makes it possible to do integration in a very general setting. The definition of integral that we will introduce in the next chapter is based on the theory of measure, pioneered at the end of the last century by Émile Borel. This is a mathematical model applicable to the geometric notions of length, area, or volume, and to the physical notion of mass. It is also an appropriate mathematical model for the notion of probability, when “events” are interpreted as sets.

## 9.1 Additive Set Functions

It turns out to be impractical to try to assign a “length” to *every* subset of  $\mathbf{R}$ , for instance. We will have to content ourselves with having a reasonably large class of “measurable sets.” The next two definitions clearly single out the most essential property of measurement: that an object can be measured by breaking it up into smaller pieces, measuring those, and adding the results.

**9.1 Definition.** Let  $X$  be a set. A collection  $\mathcal{A}$  of subsets of  $X$  is called an algebra if:

- (a)  $\emptyset \in \mathcal{A}$ ;
- (b) if  $A \in \mathcal{A}$ , then  $A^C \in \mathcal{A}$ ; and
- (c) if  $A \in \mathcal{A}$  and  $B \in \mathcal{A}$ , then  $A \cup B \in \mathcal{A}$ .

Thus, an algebra of subsets of  $X$  is a nonempty collection closed under the operation of complementation, and closed under union. Since  $A \cap B = (A^C \cup B^C)^C$ , an algebra is also closed under the operation of intersection. Also, an obvious induction argument generalizes (c) above to the statement that  $\bigcup_{j=1}^n A_j \in \mathcal{A}$  whenever  $A_1, \dots, A_n \in \mathcal{A}$ . The two simplest examples of algebras are the *trivial algebra*  $\{\emptyset, X\}$ , and  $\mathcal{P}(X)$ , the collection of all subsets of  $X$ .

**9.2 Definition.** Let  $\mu$  be a function whose domain is an algebra  $\mathcal{A}$  of subsets of  $X$ . We say that  $\mu$  is finitely additive, or simply additive, if  $\mu(A \cup B) = \mu(A) + \mu(B)$ , whenever  $A, B \in \mathcal{A}$  and  $A \cap B = \emptyset$ .

We observe that if  $\mu$  is finitely additive on  $\mathcal{A}$ , an easy induction argument shows that

$$\mu\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k),$$

whenever  $A_1, \dots, A_n$  are pairwise disjoint elements of  $\mathcal{A}$ .

Note that we carelessly omitted to say what kind of values the set function  $\mu$  took. Well, all that is needed is that there be some meaning for addition; values in  $\mathbf{R}$  or  $\mathbf{R}^n$ , for instance, would make sense, or values which were linear transformations on some (fixed) vector space. However, the kind of values we will restrict ourselves to in the sequel are *nonnegative extended real* values; i.e.,  $\mu(A)$  is to be either a nonnegative real number, or the symbol  $\infty$ , with the addition rule  $t + \infty = \infty + t = \infty$  for any  $t$ . We note that if  $\mu$  is additive, then for any  $A, B \in \mathcal{A}$  with  $A \subset B$  we have  $B = A \cup (B \cap A^C)$ , so  $\mu(B) = \mu(A) + \mu(B \cap A^C) \geq \mu(A)$ . We express this property with the phrase " $\mu$  is monotone." Note also that if  $\mu$  is additive, then  $\mu(\emptyset) = 0$ , unless  $\mu(A) = \infty$  for every  $A$ .

**9.3 Example.** Let  $X$  be any nonempty set, and define  $\mu$  on  $\mathcal{P}(X)$  by  $\mu(A) = \#A$ , the number of points in  $A$ , if  $A$  is finite, and  $\mu(A) = \infty$  if  $A$  is infinite. We call this set function *counting measure* on  $X$ . It is easily seen to be additive.

**9.4 Example.** Let  $X$  be any nonempty set, fix  $x \in X$ , and define  $\delta_x : \mathcal{P}(X) \rightarrow \mathbf{R}$  by  $\delta_x(A) = 1$  if  $x \in A$ , and  $\delta_x(A) = 0$  if  $x \notin A$ . This set function is called the *unit point mass* at  $x$ . It is clearly additive.

**9.5 Example.** Let  $X = \mathbf{R}$ , and let  $\mathcal{J}$  be the collection of all semiclosed intervals  $(a, b]$ , where  $-\infty \leq a \leq b < \infty$ , together with the open intervals of form  $(a, \infty)$ . (The case  $a = b$  was included so that  $\emptyset \in \mathcal{J}$ .) We note that  $\mathcal{J}$  is closed under finite intersections, but not under complements or finite unions. Let  $\mathcal{A}$  be the class of all sets which can be expressed as finite unions of sets in  $\mathcal{J}$ . Then  $\mathcal{A}$  includes the complement of every element of  $\mathcal{J}$ , and hence is closed under complements as well as intersections. Thus  $\mathcal{A}$  is an algebra. Let  $\mu((a, b]) = b - a$ . Every  $A \in \mathcal{A}$  can be expressed as a *disjoint* union of intervals in  $\mathcal{J}$ :  $A = \bigcup_{j=1}^r I_j$ , with  $I_j \in \mathcal{J}$  and  $I_j \cap I_k = \emptyset$  for  $j \neq k$ . We can define  $\mu(A)$  to be  $\sum_{j=1}^r \mu(I_j)$ , provided we show that this sum depends only on  $A$ , and not on the particular decomposition into disjoint intervals. This can be done; we omit the details. Thus we can assign a "length" to each set in  $\mathcal{A}$ .

**9.6 Example.** We can generalize the last example to  $\mathbf{R}^n$ . Let  $X = \mathbf{R}^n$ , and let  $\mathcal{J}$  be the collection of all semiclosed “intervals” of the form

$$I = \{(x_1, \dots, x_n) : a_j < x_j \leq b_j, j = 1, \dots, n\},$$

where  $-\infty \leq a_j \leq b_j \leq +\infty$  for  $j = 1, \dots, n$  and we interpret  $x_j \leq b_j$  as  $x_j < \infty$  if  $b_j = \infty$ . (We may occasionally use the more compact notation  $I = \{\mathbf{x} : \mathbf{a} < \mathbf{x} \leq \mathbf{b}\}$ .) We observe that  $\mathcal{J}$  is closed under intersections, but not under complements or unions. Let  $\mathcal{A}$  be the class of all sets which can be expressed as finite unions of sets in  $\mathcal{J}$ . We can check that the complement of any  $I \in \mathcal{J}$  is an element of  $\mathcal{A}$ , and it follows that  $\mathcal{A}$  is an algebra. If  $I$  is an interval as above, define  $\mu(I) = \prod_{j=1}^n (b_j - a_j)$ . Thus  $\mu$  is the usual “volume” of a box in  $\mathbf{R}^n$ . Now every  $A \in \mathcal{A}$  can be expressed as a finite union of *disjoint* intervals  $I_k \in \mathcal{J}$ ; we want to define  $\mu(A) = \sum \mu(I_k)$ , but in order to do this, we must show that this formula is unambiguous. In other words, it must be shown that if

$$A = \bigcup_{j=1}^r I_j = \bigcup_{k=1}^s I'_k,$$

where  $I_1, \dots, I_r$  are pairwise disjoint intervals, as are  $I'_1, \dots, I'_s$ , then

$$\sum_{j=1}^r \mu(I_j) = \sum_{k=1}^s \mu(I'_k).$$

This can be done without much difficulty, but we omit the argument here.

**9.7 Example.** Again, let  $X = \mathbf{R}$ , and let  $\mathcal{A}$  be the algebra of Example 9.5. Let  $g$  be a nondecreasing real function on  $\mathbf{R}$ , and define the set function  $\mu_g$  by  $\mu_g((a, b]) = g(b) - g(a)$ , and  $\mu_g(A) = \sum_{j=1}^r \mu_g(I_j)$  if  $A$  is the disjoint union of  $I_1, \dots, I_r$ , with each  $I_j \in \mathcal{J}$ . One can show that  $\mu_g$  is well-defined (that  $\mu_g(A)$  is independent of the particular decomposition of  $A$  into intervals), and is finitely additive. Any finitely additive  $\mu$  on  $\mathcal{A}$  which is finite on bounded intervals is of the form  $\mu_g$  for some nondecreasing  $g$ . Simply define  $g(x) = \mu([0, x])$  for  $x > 0$ , and  $g(x) = -\mu([x, 0])$  for  $x \leq 0$ .

## 9.2 Countable Additivity

The last two definitions carry the absolutely minimal properties of our intuitive idea of measure. They are not sufficiently restrictive to enable us to deal effectively with limit processes, so we impose further conditions.

**9.8 Definition.** A collection  $\mathcal{A}$  of subsets of  $X$  is called a  $\sigma$ -algebra if the following hold:

- (a)  $\emptyset \in \mathcal{A}$ ;
- (b)  $A^C \in \mathcal{A}$  whenever  $A \in \mathcal{A}$ ; and
- (c) if  $A_k \in \mathcal{A}$  for  $k = 1, 2, \dots$ , then  $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$ .

In other words, a  $\sigma$ -algebra is an algebra which is closed under countable unions. The algebras  $\{\emptyset, X\}$  and  $\mathcal{P}(X)$  are  $\sigma$ -algebras, while the algebras of Examples 9.5 and 9.6 are not. By taking complements, one sees that a  $\sigma$ -algebra can also be described as an algebra closed under countable intersections.

It is not hard to see that the intersection of any nonempty collection of algebras (or  $\sigma$ -algebras) of subsets of  $X$  is again an algebra (or  $\sigma$ -algebra). In particular, given any class  $\mathcal{S}$  of subsets of  $X$ , we may take the intersection of all algebras (resp.,  $\sigma$ -algebras) of subsets of  $X$  which contain  $\mathcal{S}$  (this is a nonempty collection since it contains  $\mathcal{P}(X)$ ); this is the smallest algebra (resp.,  $\sigma$ -algebra) which contains  $\mathcal{S}$ , and we refer to it as the algebra (resp.,  $\sigma$ -algebra) *generated by*  $\mathcal{S}$ . In Examples 9.5 and 9.6 above, the algebra  $\mathcal{A}$  is exactly the algebra generated by the collection  $\mathcal{S}$ . We could describe the algebra generated by  $\mathcal{S}$  more constructively, as the set of all finite unions of finite intersections of sets in  $\mathcal{S}$  and their complements. If  $\mathcal{S}$  consists of the singletons (sets with exactly one point) in  $X$ , then the algebra generated by  $\mathcal{S}$  consists of the sets which are either finite or have finite complement, and the  $\sigma$ -algebra generated by  $\mathcal{S}$  consists of those sets which are either countable or have countable complement. In general, it is not so easy to describe the sets in the  $\sigma$ -algebra generated by a family  $\mathcal{S}$ .

**9.9 Definition.** Let  $X$  be a topological space. The  $\sigma$ -algebra generated by the closed subsets of  $X$  is called the *Borel algebra*, or the *class of Borel sets of*  $X$ .

Thus, every closed set is a Borel set, and every complement of a closed set, i.e., every open set. Every countable intersection of open sets is a Borel set, and every set which can be expressed as a countable union of such countable intersections, etc., etc.

The  $\sigma$ -algebra  $\mathcal{B}$  of subsets of  $\mathbf{R}$  generated by the intervals  $(-\infty, a]$  is the Borel algebra; for any interval  $(a, b]$  can be expressed as  $(-\infty, b] \cap (-\infty, a]^C$ , so  $(a, b] \in \mathcal{B}$ , and any open interval is the union of intervals  $(a, b_n]$ , so any open interval, and hence any open set, belongs to  $\mathcal{B}$ . Thus every Borel set belongs to  $\mathcal{B}$ ; on the other hand, since every  $(-\infty, a]$  is closed, every member of  $\mathcal{B}$  is a Borel set. If we restrict ourselves to intervals  $(-\infty, a]$  with  $a$  rational, we still get all Borel sets. Similarly, the  $\sigma$ -algebra generated by the semiclosed intervals  $\mathcal{S}$  of Example 9.6 coincides with the Borel sets of  $\mathbf{R}^n$ ; so does the  $\sigma$ -algebra generated by bounded open intervals with rational endpoints, for instance. It is hard to think of a subset of  $\mathbf{R}^n$  which is *not* a Borel set, but there are plenty of them.

**9.10 Definition.** Let  $\mu$  be a set function whose domain is a class  $\mathcal{A}$  of subsets of a set  $X$ , and whose values are nonnegative extended reals. We say that  $\mu$  is countably additive if

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$$

whenever  $(A_k)$  is a sequence of pairwise disjoint sets in  $\mathcal{A}$  whose union is also in  $\mathcal{A}$ . (If  $\mathcal{A}$  is a  $\sigma$ -algebra, the last clause can of course be omitted.)

A measure on  $X$  is a countably additive nonnegative set function whose domain is a  $\sigma$ -algebra of subsets of  $X$ .

The set functions of Examples 9.3 and 9.4 above (counting measure, the unit point mass) are easily seen to be measures. If  $\mu$  is defined on  $\mathcal{P}(X)$  by  $\mu(A) = 0$  if  $A$  is finite, and  $\mu(A) = \infty$  if  $A$  is infinite, we have an example of a finitely additive set function which is not countably additive (assuming  $X$  is infinite.) The set function  $\mu$  of Example 9.5 above can be shown to be countably additive; we don't call it a measure because its domain is not a  $\sigma$ -algebra.

A measure on a topological space  $X$  whose domain is the Borel algebra is called a *Borel measure*. (Be warned though, that some authors use Borel measure to refer to a measure defined on the Borel sets with the additional property that compact sets have finite measure.)

We next observe that the countable additivity property can also be expressed as the "continuity from below" of the set function  $\mu$ .

**9.11 Proposition.** Let  $\mu$  be a finitely additive set function, defined on the algebra  $\mathcal{A}$ . Then  $\mu$  is countably additive if and only if it has the following property: if  $A_n \in \mathcal{A}$  and  $A_n \subset A_{n+1}$  for each positive integer  $n$ , and if  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ , then  $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mu(A_n)$ .

**Proof.** Suppose  $\mu$  is countably additive. If we set  $B_1 = A_1$ , and  $B_n = A_n \setminus A_{n-1}$  for  $n > 1$ , we see that  $A_n = \bigcup_{k=1}^n B_k$ , and that the sets  $B_k$  are pairwise disjoint, so

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(B_k) = \lim_{n \rightarrow \infty} \mu(A_n).$$

Now suppose that  $\mu$  is finitely additive, and has the "continuity from below" property. If  $\{A_n\}$  is a sequence of pairwise disjoint sets in  $\mathcal{A}$ , and we put  $B_n = \bigcup_{k=1}^n A_k$ , then  $B_n \subset B_{n+1}$  for every  $n$ , so  $\mu(B_n) \rightarrow \mu(B)$ , where  $B = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ . But since  $\mu$  is finitely additive,  $\mu(B_n) = \sum_{k=1}^n \mu(A_k)$ , so  $\mu(B_n) \rightarrow \sum_{k=1}^{\infty} \mu(A_k)$ . Thus  $\mu$  is countably additive. ■

Here are some ways to get new measures from old ones.



**9.12 Proposition.** Let  $\mathcal{A}$  be a  $\sigma$ -algebra; if  $\mu, \nu$  are measures on  $\mathcal{A}$ ,  $t \in \mathbf{R}_+$ , and  $A \in \mathcal{A}$ , then the following are measures on  $\mathcal{A}$ :

- (a)  $\mu + \nu$ , defined by  $(\mu + \nu)(E) = \mu(E) + \nu(E)$ ;
- (b)  $t\mu$ , defined by  $(t\mu)(E) = t\mu(E)$ ; and
- (c)  $\mu_A$ , defined by  $\mu_A(E) = \mu(A \cap E)$ .

We leave the proofs to the reader. We can obviously generalize item (a) in this list to any finite sum of measures.

The collection of nonnegative set functions with domain  $\mathcal{A}$  have a natural order relation: if  $\mu$  and  $\nu$  are nonnegative set functions with the same domain  $\mathcal{A}$ , we say  $\mu \leq \nu$  if  $\mu(A) \leq \nu(A)$  for every  $A \in \mathcal{A}$ . Similarly, we can define the set function  $\mu \vee \nu$ , the maximum of  $\mu$  and  $\nu$ , by the natural  $(\mu \vee \nu)(A) = \max\{\mu(A), \nu(A)\}$ ; this idea extends to taking the supremum of an arbitrary collection of such set functions (all having the same domain  $\mathcal{A}$ , of course). In general, if  $\mu$  and  $\nu$  are measures, it does not follow that  $\mu \vee \nu$  is a measure. (The reader is invited to give a simple example where it is not even finitely additive.) However, we have the following result:

**9.13 Theorem.** Let  $\mathcal{A}$  be a  $\sigma$ -algebra of subsets of the set  $X$ , and let  $\mathcal{M}$  be a collection of measures with domain  $\mathcal{A}$ . Suppose that  $\mathcal{M}$  has the property: for any  $\mu_1, \mu_2 \in \mathcal{M}$  there exists  $\mu_3 \in \mathcal{M}$  such that  $\mu_1 \leq \mu_3$  and  $\mu_2 \leq \mu_3$ . If  $\nu$  is defined by  $\nu(E) = \sup\{\mu(E) : \mu \in \mathcal{M}\}$ , then  $\nu$  is a measure on  $\mathcal{A}$ .

**Proof.** Certainly,  $\nu$  is a well-defined, nonnegative set function with domain  $\mathcal{A}$ ; what we must prove is that  $\nu$  is countably additive. Let  $\{A_n\}$  be a disjoint sequence in  $\mathcal{A}$ . Then for each  $\mu \in \mathcal{M}$ , we have

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \leq \sum_{n=1}^{\infty} \nu(A_n),$$

whence

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \nu(A_n). \quad (9.1)$$

We have not yet used the key hypothesis. Fix the positive integer  $n$ , and for each  $k$ ,  $1 \leq k \leq n$ , choose  $c_k$  with  $c_k < \nu(A_k)$ . By the definition of  $\nu$ , there exists for each  $k$  some  $\mu_k \in \mathcal{M}$  with  $\mu_k(A_k) > c_k$ . By the hypothesis of the theorem (extended to  $n$  elements of  $\mathcal{M}$  by an obvious induction) there exists  $\mu \in \mathcal{M}$  such that  $\mu_k \leq \mu$  for  $k = 1, \dots, n$ . Then

$$\sum_{k=1}^n c_k < \sum_{k=1}^n \mu_k(A_k) \leq \sum_{k=1}^n \mu(A_k) = \mu\left(\bigcup_{k=1}^n A_k\right) \leq \nu\left(\bigcup_{k=1}^n A_k\right),$$

and since  $c_k < \nu(A_k)$  were arbitrary, we conclude that

$$\sum_{k=1}^n \nu(A_k) \leq \nu\left(\bigcup_{k=1}^n A_k\right) \leq \nu\left(\bigcup_{k=1}^{\infty} A_k\right)$$

for every  $n$ , and hence that

$$\sum_{k=1}^{\infty} \nu(A_k) \leq \nu\left(\bigcup_{k=1}^{\infty} A_k\right). \quad (9.2)$$

Combining inequalities (9.1)s and (9.2), we see that  $\nu$  is countably additive, and thus a measure. ■

**9.14 Corollary.** *Let  $\{\mu_n\}$  be a nondecreasing sequence of measures, and define  $\mu$  by  $\mu(A) = \sup \mu_n(A)$ . Then  $\mu$  is a measure.*

**9.15 Example.** Here is an application of Theorem 9.13. Let  $X$  be any set, and suppose  $p : X \rightarrow [0, \infty)$ . Define the set function  $\mu$  with domain  $\mathcal{P}(X)$  by  $\mu(A) = \sum_{x \in A} p(x)$ . Then  $\mu$  is a measure. For  $\mu = \sup_F \mu_F$ , where the sup is taken over all finite sets  $F$ , and  $\mu_F$ , defined by  $\mu_F = \sum_{x \in F} p(x)\delta_x$ , is a measure in view of Proposition 9.12. When  $p(x) \equiv 1$ , we recover the counting measure of Example 9.3 above. In the general case, we have theorems about series of positive terms, which we met in Chapter 2. For instance, if  $X$  is the disjoint union of countably many subsets  $X_n$ , we have

$$\sum_{x \in X} p(x) = \sum_{n=1}^{\infty} \sum_{x \in X_n} p(x);$$

as a special case, when  $X = \mathbf{N} \times \mathbf{N}$ , we obtain, whenever  $a_{mn} \geq 0$ , that

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{mn} = \sum_{k=2}^{\infty} \sum_{m+n=k} a_{mn},$$

in particular, the convergence of any of these three double series implies the convergence of the other two.

## 9.3 Outer Measures

It is often a nontrivial task to construct measures. For instance, we can define the length of any open subset of  $\mathbf{R}$  without any trouble, since each open set has a unique expression as a disjoint union of countably many open intervals, and from there we can easily find the only candidate for the length of a closed set; but there is no obvious way to express a Borel set in terms of open and closed sets, so that we can write a formula for

its length. There is no canonical way to express an open set in  $\mathbf{R}^2$  as the disjoint union of intervals, and thus no obvious way to define the area of an open set in  $\mathbf{R}^2$ . We begin by considering a class of set functions which lack (in general) the desired additivity property of measures.

**9.16 Definition.** Let  $X$  be a set. An outer measure on  $X$  is a nonnegative, extended-real valued function  $\mu^*$  whose domain consists of all subsets of  $X$ , and which satisfies:

- (a)  $\mu^*(\emptyset) = 0$ ;
- (b) if  $A \subset B$ , then  $\mu^*(A) \leq \mu^*(B)$ ; and
- (c) for any sequence  $(A_n)$  of subsets of  $X$ , we have

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n).$$

Property (b) is generally referred to as *monotonicity*, and property (c) as *countable subadditivity*. The next lemma describes a general method for constructing outer measures.

**9.17 Lemma.** Let  $\mathcal{C}$  be a collection of subsets of the set  $X$ , having the property that there is a countable subcollection of  $\mathcal{C}$  whose union is  $X$ . Let  $\lambda$  be a nonnegative real-valued function whose domain is  $\mathcal{C}$ . If  $\mu^*$  is defined by  $\mu^*(\emptyset) = 0$ , and

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \lambda(C_n) : C_n \in \mathcal{C}, \bigcup_{n=1}^{\infty} C_n \supset A \right\}. \quad (9.3)$$

for  $A \subset X$ ,  $A \neq \emptyset$ , then  $\mu^*$  is an outer measure on  $X$ .

**Proof.** The hypothesis guarantees that the set on the right-hand side of equation (9.3) is not empty; if we agree that the inf of the empty set is  $+\infty$ , we can dispense with it. Properties (a) and (b) of Definition 9.16 are obviously satisfied; the issue is countable subadditivity. Let  $A_n$  be a subset of  $X$ , for each positive integer  $n$ . Let  $\epsilon > 0$ . For each  $n$ , we may choose  $C_{n,k} \in \mathcal{C}$  ( $k = 1, 2, \dots$ ), such that

$$A_n \subset \bigcup_{k=1}^{\infty} C_{n,k} \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda(C_{n,k}) \leq \mu^*(A_n) + \frac{\epsilon}{2^n},$$

according to (9.3); then (since the countable family  $\{C_{n,k} : n, k = 1, 2, \dots\}$  covers  $\bigcup_{n=1}^{\infty} A_n$ ) it follows that

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n,k} \lambda(C_{n,k}) = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \lambda(C_{n,k})$$

$$\leq \sum_{n=1}^{\infty} (\mu^*(A_n) + \epsilon 2^{-n}) = \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon;$$

and since  $\epsilon > 0$  was arbitrary, the lemma is proved.  $\blacksquare$

**9.18 Example.** The single most important example of the construction given by Lemma 9.17 is *Lebesgue outer measure* on  $\mathbf{R}^n$ , which is obtained by taking  $\mathcal{C}$  to be the class of all bounded open intervals  $\{\mathbf{x} : \mathbf{a} < \mathbf{x} < \mathbf{b}\}$ , and

$$\lambda(\{\mathbf{a} < \mathbf{x} < \mathbf{b}\}) = \prod_{j=1}^n (b_j - a_j).$$

We denote Lebesgue outer measure by  $m^*$ . We would obtain the same outer measure by taking  $\mathcal{C}$  to be the set of all bounded intervals in the class  $\mathcal{J}$  of Example 9.6, and  $\lambda$  to be the  $\mu$  of that example.

Let us verify that  $m^*(I) = \lambda(I)$  for any interval  $I$ . First of all, it is clear that  $m^*(I) \leq \lambda(I)$ , by considering coverings by a single open interval. Let  $\alpha < \lambda(I)$ . Then we can find a closed bounded interval  $J \subset I$ , with  $\lambda(J) > \alpha$ . Now if  $\{C_k\}$  is any sequence of open intervals whose union contains  $I$ , then by the Heine-Borel theorem there exists  $n$  such that  $J \subset \bigcup_{k=1}^n C_k$ . Then  $\lambda(J) \leq \sum_{k=1}^n \lambda(C_k)$  (for finite unions, this is elementary), and hence  $\lambda(J) \leq \sum_{k=1}^{\infty} \lambda(C_k)$ , so  $\alpha < \lambda(J) \leq m^*(I)$ ; since  $\alpha$  was an arbitrary number with  $\alpha < \lambda(I)$ , we conclude  $\lambda(I) \leq m^*(I)$ , and thus  $m^*(I) = \lambda(I)$ .

**9.19 Example.** Let  $g$  be a nondecreasing real-valued function on  $\mathbf{R}$ , and define  $\lambda_g$  on the collection of bounded open intervals by  $\lambda_g((a, b)) = g(b-) - g(a+)$ . (Recall that  $g(a+)$  is defined as  $\lim_{h \rightarrow 0+} g(a+h)$ , the limit from the right at  $a$ ; similarly,  $g(b-)$  is the limit from the left of  $g$  at  $b$ .) The outer measure  $\mu_g^*$  obtained by application of Lemma 9.17 is known as the *Lebesgue-Stieltjes outer measure* associated to  $g$ . The reason for using  $g(b-) - g(a+)$  instead of just  $g(b) - g(a)$  will appear later.

**9.20 Example.** Another important example of an outer measure on  $\mathbf{R}^n$  is obtained as follows: fix a nonnegative real number  $p$ . For each  $\delta > 0$ , let  $\mathcal{C}_\delta$  be the collection of all subsets  $C$  of  $\mathbf{R}^n$  with  $\text{diam } C \leq \delta$ . Define  $\lambda_\delta(C) = (\text{diam } C)^p$ , and let  $\mathcal{H}_\delta^*$  be the resulting outer measure. It is obvious that  $\mathcal{H}_\delta^*(A) \leq \mathcal{H}_{\delta'}^*(A)$  if  $0 < \delta' < \delta$ . Let

$$\mathcal{H}^*(A) = \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^*(A) = \sup_{\delta > 0} \mathcal{H}_\delta^*(A);$$

from Theorem 9.13, or directly, we see that  $\mathcal{H}^*$  is again an outer measure; it is called *Hausdorff  $p$ -dimensional outer measure* on  $\mathbf{R}^n$ . The same idea makes sense in any separable metric space. With  $p = 0$ , we recover the “counting measure” of Example 9.3 above. With  $p = 1$  (the original version of this construction, due to Carathéodory), we get a definition of

length which can be shown to coincide with the usual idea for paths. It was Hausdorff's great idea not only to consider values of  $p$  different from 1, but to consider values of  $p$  which are not integers. It is not hard to see that if  $\mathcal{H}_p(E) = 0$ , then  $\mathcal{H}_q(E) = 0$  for every  $q > p$ . The *Hausdorff dimension* of  $E$  is defined to be  $\sup\{p \geq 0 : \mathcal{H}_p(E) > 0\}$ . It is not hard to verify that the Hausdorff dimension of a countable set is 0, that of a rectifiable curve is 1, that of an open set in  $\mathbf{R}^n$  is  $n$ . But this notion of dimension does not always yield an integer; for instance, the Hausdorff dimension of the Cantor set turns out to be  $\log 2 / \log 3$ .

## 9.4 Constructing Measures

We now show how to obtain measures from outer measures. The following definition may seem artificial, but it turns out to be a very useful technical tool.

**9.21 Definition.** Let  $X$  be a set, and let  $\mu^*$  be an outer measure on  $X$ . We say that a subset  $E$  of  $X$  is  $\mu^*$ -measurable if, for every  $A \subset X$ , we have

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^C). \quad (9.4)$$

In other words, a  $\mu^*$ -measurable set splits every set up into two pieces on which  $\mu^*$  is additive. Of course, we always have

$$\mu^*(A) \leq \mu^*(A \cap E) + \mu^*(A \cap E^C)$$

since  $\mu^*$  is subadditive; it is the opposite inequality which is at issue. Let us note that this is immediate when  $\mu^*(E) = 0$ , so every set of outer measure 0 is measurable.

**9.22 Theorem.** If  $\mu^*$  is an outer measure on  $X$ , and  $\mathcal{M}$  is the collection of all  $\mu^*$ -measurable subsets of  $X$ , then  $\mathcal{M}$  is a  $\sigma$ -algebra, and the restriction of  $\mu^*$  to  $\mathcal{M}$  is a measure on  $X$ .

**Proof.** We begin by showing that  $\mathcal{M}$  is an algebra. It is obvious that  $\emptyset \in \mathcal{M}$ , and that  $E^C \in \mathcal{M}$  whenever  $E \in \mathcal{M}$ . Suppose that  $E \in \mathcal{M}$  and  $F \in \mathcal{M}$ . Then for any  $A \subset X$ , we have

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^C),$$

and using  $A \cap E^C$  in the role of the test set  $A$  in (9.4), also

$$\mu^*(A \cap E^C) = \mu^*(A \cap E^C \cap F) + \mu^*(A \cap E^C \cap F^C);$$

substituting, we have

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^C \cap F) + \mu^*(A \cap E^C \cap F^C),$$

whence, using the subadditivity of  $\mu^*$  and DeMorgan's law,

$$\begin{aligned} \mu^*(A) &\geq \mu^*((A \cap E) \cup (A \cap E^C \cap F)) + \mu^*(A \cap E^C \cap F^C) \\ &= \mu^*(A \cap (E \cup F)) + \mu^*(A \cap (E \cup F)^C) \end{aligned}$$

which proves that  $E \cup F$  is  $\mu^*$ -measurable (as we saw above, the opposite inequality is automatic). Thus  $\mathcal{M}$  is an algebra.

We next establish the formula: for any  $E_1, \dots, E_n \in \mathcal{M}$ , with  $E_j \cap E_k = \emptyset$  when  $j \neq k$ , and any  $A \subset X$ ,

$$\mu^*(A \cap \bigcup_{j=1}^n E_j) = \sum_{j=1}^n \mu^*(A \cap E_j). \quad (9.5)$$

We proceed by induction. When  $n = 1$ , this is simply the definition of  $\mathcal{M}$ . Suppose that (9.5) is established for some integer  $n$ , and suppose that  $E_1, \dots, E_{n+1}$  are pairwise disjoint  $\mu^*$ -measurable sets. Let  $F = \bigcup_{j=1}^n E_j$ . Then  $F$  is  $\mu^*$ -measurable, since  $\mathcal{M}$  is an algebra. Applying the criterion (9.4) to the "test set"  $A' = A \cap (\bigcup_{k=1}^{n+1} E_k)$  and the measurable set  $F$ , and observing that  $A' \cap F^C = A \cap E_{n+1}$ , we have the desired formula (9.5) for  $n + 1$ , concluding the proof that (9.5) holds for all  $n$ . Taking  $A = X$ , we have, in particular, established that  $\mu^*$  is finitely additive on  $\mathcal{M}$ .

Now let  $\{E_n\}$  be a pairwise disjoint sequence of  $\mu^*$ -measurable sets, let  $F_n = \bigcup_{k=1}^n E_k$ , and let  $F = \bigcup_{n=1}^{\infty} E_n$ . Then, for any  $A \subset X$ , we have

$$\begin{aligned} \mu^*(A) &= \mu^*(A \cap F_n) + \mu^*(A \cap F_n^C) \\ &\geq \mu^*(A \cap F_n) + \mu^*(A \cap F^C) \quad \text{since } \mu^* \text{ is monotone} \\ &= \sum_{k=1}^n \mu^*(A \cap E_k) + \mu^*(A \cap F^C) \quad \text{by (9.5)} \end{aligned}$$

for every positive integer  $n$ , whence

$$\begin{aligned} \mu^*(A) &\geq \sum_{k=1}^{\infty} \mu^*(A \cap E_k) + \mu^*(A \cap F^C) \\ &\geq \mu^*(A \cap F) + \mu^*(A \cap F^C) \end{aligned}$$

proving that  $F \in \mathcal{M}$ . Now an algebra which is closed under countable disjoint unions must be, in fact, closed under arbitrary countable unions; thus we have established that  $\mathcal{M}$  is a  $\sigma$ -algebra, and (taking  $A = F$  above) that  $\mu^*$  is countably additive on  $\mathcal{M}$ . ■

**9.23 Definition.** Let  $m^*$  be Lebesgue outer measure on  $\mathbf{R}^n$  (see Example 9.18). An  $m^*$ -measurable set will be called Lebesgue measurable, and the  $\sigma$ -algebra of Lebesgue measurable sets will be denoted by  $\mathcal{M}$ . The restriction of  $m^*$  to  $\mathcal{M}$  will be called Lebesgue measure, and denoted by  $m$ .

We note that any  $E \subset \mathbf{R}^n$  with  $m^*(E) = 0$  is Lebesgue measurable, and  $m(E) = 0$ , but that we do not yet know that intervals are measurable. We recall that the Cantor set  $C \subset \mathbf{R}$  is a subset, for each positive integer  $n$ , of a set  $C_n$  which is the union of  $2^n$  disjoint closed intervals, each of length  $3^{-n}$ . Thus  $m^*(C_n) \leq (2/3)^n$ , and it follows that  $m^*(C) = 0$ . In particular,  $C$  is Lebesgue measurable.

**9.24 Theorem.** Suppose that the outer measure  $\mu^*$  was constructed from a set function  $\lambda$  and a "covering class"  $\mathcal{C}$  by the procedure of Lemma 9.17, but with the special circumstance that  $\mathcal{C}$  is an algebra, and  $\lambda$  countably additive on  $\mathcal{C}$ . Then each  $E \in \mathcal{C}$  is  $\mu^*$ -measurable, and  $\mu^*(E) = \lambda(E)$ .

**Proof.** Let  $E \in \mathcal{C}$ , and  $A \subset X$ . For any  $\epsilon > 0$ , there exist  $F_n \in \mathcal{C}$  such that

$$A \subset \bigcup_{n=1}^{\infty} F_n \quad \text{and} \quad \mu^*(A) \leq \sum_{n=1}^{\infty} \lambda(F_n) \leq \mu^*(A) + \epsilon.$$

Set  $F'_1 = F_1$ , and  $F'_n = F_n \setminus (\bigcup_{k=1}^{n-1} F_k)$  for  $n > 1$  (i.e., "disjointify" the sequence). Then since  $F'_n \subset F_n$  for each  $n$ , and  $\lambda$  is additive on the algebra  $\mathcal{C}$ , we have

$$\begin{aligned} \mu^*(A) + \epsilon &\geq \sum_{n=1}^{\infty} \lambda(F'_n) \\ &= \sum_{n=1}^{\infty} [\lambda(F'_n \cap E) + \lambda(F'_n \cap E^C)] \\ &\geq \mu^*(A \cap E) + \mu^*(A \cap E^C), \end{aligned}$$

which, since  $\epsilon > 0$  is arbitrary, shows that  $E$  is  $\mu^*$ -measurable. Now, obviously, we have  $\mu^*(E) \leq \lambda(E)$ ; taking  $A = E$  above, we see that

$$\mu^*(E) \geq \sum_{n=1}^{\infty} \lambda(E \cap F'_n) = \lambda(E),$$

using the countable additivity of  $\lambda$  on  $\mathcal{C}$ . ■

## 9.5 Metric Outer Measures

Theorem 9.24 gives us a way of extending a countably additive set function from an algebra to a  $\sigma$ -algebra containing it, but we will not exploit this important result in the sequel. We will instead use the next theorem, to avoid

the preliminary stage of constructing a countably additive set function in the process of constructing Borel measures. We want yet more terminology.

**9.25 Definition.** Let  $(X, \rho)$  be a metric space. If subsets  $A$  and  $B$  are subsets of  $X$ , we say that  $A$  and  $B$  are well separated if

$$\inf\{\rho(x, y) : x \in A, y \in B\} > 0.$$

**9.26 Definition.** We say that the outer measure  $\mu^*$  on the metric space  $X$  is a metric outer measure if it satisfies the condition

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$$

for any pair of sets  $A, B$  which are well separated.

**9.27 Theorem.** Let  $\mu^*$  be a metric outer measure on the metric space  $X$ . Then every Borel set is  $\mu^*$ -measurable.

**Proof.** We separate out the essential part of the argument as:

**9.28 Lemma.** Let  $\mu^*$  be a metric outer measure on the metric space  $X$ , and suppose that  $\{E_n\}$  is a sequence of sets in  $X$  with  $E_n \subset E_{n+1}$  for all  $n$ , with  $E = \bigcup_{n=1}^{\infty} E_n$ . If  $E_n$  and  $E \setminus E_{n+1}$  are well separated for every  $n$ , then  $\mu^*(E) = \lim \mu^*(E_n)$ .

**Proof of Lemma.** Since  $E_n \subset E_{n+1}$ , the sequence  $(\mu^*(E_n))$  is increasing, so  $L = \lim \mu^*(E_n)$  exists. Since  $\mu^*(E_n) \leq \mu^*(E)$  for every  $n$ , it is clear that  $L \leq \mu^*(E)$ . If  $L = \infty$ , the opposite inequality is free, so let us assume that  $L < \infty$ . Let  $A_1 = E_1$  and  $A_n = E_n \setminus E_{n-1}$  for  $n > 1$ . Then  $A_n$  and  $\bigcup_{k=n+2}^{\infty} A_k$  are well separated, so (by an obvious induction from the condition of Definition 9.26) we have, for any  $m$ ,

$$\sum_{n=1}^m \mu^*(A_{2n-1}) = \mu^*\left(\bigcup_{n=1}^m A_{2n-1}\right) \leq \mu^*(E_{2m-1}) \leq L$$

and

$$\sum_{n=1}^m \mu^*(A_{2n}) = \mu^*\left(\bigcup_{n=1}^m A_{2n}\right) \leq \mu^*(E_{2m}) \leq L$$

from which we see that the series  $\sum_{n=1}^{\infty} \mu^*(A_n)$  converges. Then from

$$\mu^*(E) = \mu^*\left(E_n \cup \bigcup_{k=n+1}^{\infty} A_k\right) \leq \mu^*(E_n) + \sum_{k=n+1}^{\infty} \mu^*(A_k)$$

we deduce that  $\mu^*(E) \leq \lim \mu^*(E_n)$ . ■



Returning to the proof of the theorem: it suffices to show that each closed set  $F$  in  $X$  is  $\mu^*$ -measurable. Let  $A$  be any subset of  $X$ , and let  $E = A \setminus F$  and  $E_n = \{x \in A : \rho(x, y) \geq 1/n \text{ for all } y \in F\}$ . Clearly,  $E_n \subset E_{n+1}$  for all  $n$ . Since  $F$  is closed, for any  $x \notin F$ , we have  $\rho(x, y) \geq \delta > 0$  for every  $y \in F$ , so  $\bigcup E_n = E$ . If  $x \in E_n$ , and  $y \in E \setminus E_{n+1}$ , there exists  $z \in F$  such that  $\rho(y, z) < 1/(n+1)$ , but  $\rho(x, z) \geq 1/n$ , so  $\rho(x, y) \geq \rho(x, z) - \rho(z, y) > 1/n - 1/(n+1)$ ; thus  $E_n$  and  $E \setminus E_{n+1}$  are well separated for every  $n$ . Applying the lemma, we have  $\mu^*(E) = \lim \mu^*(E_n)$ . But since  $F$  and  $E_n$  are well separated, we have  $\mu^*(A) \geq \mu^*(F \cup E_n) = \mu^*(F) + \mu^*(E_n)$  for every  $n$ , so passing to the limit we have  $\mu^*(A) \geq \mu^*(F) + \mu^*(E)$ . Thus  $F$  is  $\mu^*$ -measurable. ■

It is easy to see that  $m^*$  is a metric outer measure (Exercise: prove this!), and hence, by Theorem 9.27, we know that every Borel set is Lebesgue measurable. (We will see later that not every Lebesgue measurable set is a Borel set.) We saw earlier that  $m^*(I) = \lambda(I)$  (the length, or volume, of  $I$ ) for every interval  $I$ ; thus Lebesgue measure is an extension of the natural notion of volume, from the class of intervals to the  $\sigma$ -algebra of all Lebesgue measurable sets, a class which includes all Borel sets, and all subsets of sets of measure zero.

It is also not hard to see that  $p$ -dimensional Hausdorff outer measure is a metric outer measure (see the exercises at the end of this chapter), so every Borel set is measurable with respect to this outer measure; the measure obtained by restricting to the measurable sets is called, of course,  $p$ -dimensional Hausdorff measure.

## 9.6 Measurable Sets

In this section, we discuss the relation between Lebesgue measurable sets and Borel subsets of  $\mathbf{R}^n$ ; it turns out that every Lebesgue measurable set differs from some Borel set by a set of measure zero. We then show that not every subset of  $\mathbf{R}$  is Lebesgue measurable, and, in fact, give a stronger version of that statement.

**9.29 Theorem.** *If  $A$  is a Lebesgue measurable subset of  $\mathbf{R}^n$ , there exist Borel sets  $F$  and  $G$ , with  $F \subset A \subset G$  and  $m(G \setminus A) = m(A \setminus F) = 0$ . We may take  $F$  to be the union of a sequence of compact sets, and  $G$  to be the intersection of a sequence of open sets.*

**Proof.** We establish first:

**9.30 Lemma.** *For any bounded measurable set  $B$  in  $\mathbf{R}^n$ , and any  $\epsilon > 0$ , there exist compact  $K$  and open  $U$ , with  $K \subset B \subset U$ , such that  $m(U \setminus K) < \epsilon$ .*

**Proof.** From the definition of the outer measure  $m^*$ , we see that there exists an open  $U \supset B$  with  $m(U) < m(B) + \frac{1}{2}\epsilon$ . Let  $I$  be a bounded open interval containing  $\bar{B}$ ; there exists an open  $V$ , with  $I \supset V \supset I \setminus B$ , such that  $m(V) < m(I \setminus B) + \frac{1}{2}\epsilon$ . Since  $B$  is measurable, we know  $m(I) = m(B) + m(I \setminus B)$ , so

$$m(I \setminus V) = m(I) - m(V) \geq m(I) - m(I \setminus B) - \frac{\epsilon}{2} = m(B) - \frac{\epsilon}{2}.$$

Let  $K = I \setminus V$ ;  $K$  is closed in  $I$ , since  $V$  is open, and a subset of  $B$ , hence of  $\bar{B}$ , so  $K$  is in fact closed; since it is also bounded,  $K$  is compact. We have  $K \subset B \subset U$ , and  $m(U \setminus K) = m(U \setminus B) + m(B \setminus K) < \epsilon$ , as claimed. ■

Returning to the proof of the theorem, given any measurable set  $A$ , we may write  $A$  as the union of a disjoint sequence of bounded measurable sets  $A_n$ ; for each  $n$ , and each positive integer  $j$ , we can find, according to the lemma, compact  $K_{n,j}$  and open  $U_{n,j}$  such that  $K_{n,j} \subset A_n \subset U_{n,j}$  and  $m(U_{n,j} \setminus K_{n,j}) < 2^{-n}/j$ . Let  $F = \bigcup_{n,j} K_{n,j}$ . Then  $F \subset A$  and

$$m(A \setminus F) = \sum_{n=1}^{\infty} m(A_n \setminus \bigcup_{j=1}^{\infty} K_{n,j})$$

since the  $A_n$  are disjoint. Now, since  $m(A_n \setminus K_{n,j}) < 1/j$ , it follows that  $m(A_n \setminus \bigcup_j K_{n,j}) = 0$  for each  $n$ , and therefore  $m(A \setminus F) = 0$ . Let  $G_j = \bigcup_{n=1}^{\infty} U_{n,j}$ . Then  $G_j$  is open,  $G_j \supset A$ , and

$$m(G_j \setminus A) \leq \sum_{n=1}^{\infty} m(U_{n,j} \setminus A_n) < \frac{1}{j}.$$

Thus, putting  $G = \bigcap_n G_n$ , we have  $A \subset G$  and  $m(G \setminus A) = 0$ , as desired. ■

Thus, every measurable set is “essentially” a Borel set, in the sense that the set-theoretic difference is a null set for Lebesgue measure. It is a fact (which we have not yet shown, but will appear in the exercises in the next chapter) that there exist Lebesgue measurable sets which are not Borel sets. Another, less constructive way to prove this runs roughly as follows: it can be shown that the class of Borel sets in  $\mathbf{R}$  can be put in one-to-one correspondence with the real numbers, as can the Cantor set  $C$ ; since  $m(C) = 0$ , every subset of the Cantor set is measurable, and thus there are as many Lebesgue measurable sets as there are subsets of  $\mathbf{R}$ ; but there is no one-to-one correspondence between any set and the set of all its subsets (Theorem 1.7).

Next we show that not every subset of  $\mathbf{R}$  is Lebesgue measurable. The same statement applies to  $\mathbf{R}^n$  for any  $n$ , of course. We begin with a remark whose proof is completely obvious.

**9.31 Proposition.** *Lebesgue outer measure is translation invariant; i.e., for any subset  $A$  of  $\mathbf{R}^n$ , and any  $\mathbf{x} \in \mathbf{R}^n$ ,  $m^*(A + \mathbf{x}) = m^*(A)$ , where  $A + \mathbf{x}$  denotes  $\{\mathbf{a} + \mathbf{x} : \mathbf{a} \in A\}$ .*

**9.32 Proposition.** *There exists a subset of  $\mathbf{R}$  which is not Lebesgue measurable.*

**Proof.** The argument we give really takes place in the circle, rather than the line. Let us define an operation ("addition mod 1") on the interval  $[0, 1)$ , as follows: for  $0 \leq x < 1$ ,  $0 \leq y < 1$ , define

$$x \oplus y = \begin{cases} x + y, & \text{if } x + y < 1; \\ x + y - 1, & \text{if } x + y \geq 1. \end{cases}$$

It is a simple consequence of Proposition 9.31 that if  $A \subset [0, 1)$  and  $x \in [0, 1)$ , we have  $m^*(A \oplus x) = m^*(A)$ . Now define a relation on  $[0, 1)$  as follows: say that  $x \sim y$  if and only if  $x - y$  is a rational number. It is easy to see that  $\sim$  is an equivalence relation, and thus decomposes  $[0, 1)$  into equivalence classes  $\{E_\alpha\}_{\alpha \in J}$ ; thus,

$$[0, 1) = \bigcup_{\alpha \in J} E_\alpha, \quad \text{and} \quad E_\alpha \cap E_\beta = \emptyset \quad \text{for } \alpha \neq \beta.$$

Now let  $A$  be a set which contains exactly one point from each  $E_\alpha$  (the existence of such a set is the content of the "axiom of choice" of formal set theory). Let  $\{q_1, q_2, \dots\}$  be an enumeration of the rational numbers in  $[0, 1)$ . Then  $A$  has the following properties:

$$(A \oplus q_j) \cap (A \oplus q_k) = \emptyset \quad \text{whenever } j \neq k;$$

$$\bigcup_{j=1}^{\infty} A \oplus q_j = [0, 1).$$

Indeed, if  $x \oplus q_j = y \oplus q_k$ , then  $x - y = q_k - q_j$  (up to an integer), so  $x \sim y$ ; but if  $x$  and  $y$  are elements of  $A$ , this implies  $x = y$ , and hence  $j = k$ . Similarly, for any  $x \in [0, 1)$ , there exists  $a \in A$  such that  $x \sim a$ , i.e.,  $x - a$  is rational; then  $x = a \oplus q$ , where  $q = x - a$  or  $x - a + 1$  is a rational in  $[0, 1)$ . Now it is apparent that the set  $A$  cannot be Lebesgue measurable, for if it were, countable additivity would imply that

$$1 = m([0, 1)) = \sum_{j=1}^{\infty} m(A \oplus q_j) = \sum_{j=1}^{\infty} m(A),$$

which is, of course, impossible since the right-hand side is either 0 or  $\infty$ . ■

This example of an unmeasurable set is striking, but there is an even more striking example. We first prove a lemma which is interesting in its own right.

**9.33 Lemma.** *If  $A$  is a Lebesgue measurable subset of  $\mathbf{R}$ , with  $m(A) > 0$ , then  $A - A = \{x - y : x, y \in A\}$  contains a neighborhood of 0.*

**Proof.** We can assume  $m(A) < \infty$  without loss of generality. Then there exists a sequence of open intervals  $\{I_n\}$  such that

$$A \subset \bigcup_{n=1}^{\infty} I_n \quad \text{and} \quad \sum_{n=1}^{\infty} m(I_n) < \frac{4}{3}m(A).$$

Now, if  $m(A \cap I_n) \leq \frac{3}{4}m(I_n)$  for every  $n$ , then

$$m(A) \leq \sum_{n=1}^{\infty} m(I_n \cap A) \leq \frac{3}{4} \sum_{n=1}^{\infty} m(I_n) < m(A),$$

a contradiction. Thus there exists an interval  $I$  such that  $m(I \cap A) > \frac{3}{4}m(I)$ . Now if  $|x| < \frac{1}{2}m(I)$ , then  $x \in A - A$ . For if not, then  $(A+x) \cap A = \emptyset$ ; but then  $m((I \cap A) + x) \cup (I \cap A) = 2m(I \cap A) > \frac{3}{2}m(I)$ . Since  $m(I \cup (I+x)) < \frac{3}{2}m(I)$ , this is a contradiction. ■

**9.34 Theorem.** *There exists a subset  $A$  of  $\mathbf{R}$  with the property that  $A \cap B$  is unmeasurable for every Lebesgue measurable set  $B$  with  $m(B) > 0$ .*

**Proof.** Let  $\xi$  be irrational, and let  $G = \{n + m\xi : m, n \in \mathbf{Z}\}$ . We recall that  $G$  is dense in  $\mathbf{R}$  (Dirichlet's theorem, Theorem 1.23). We imitate the previous construction of an unmeasurable set, with the countable dense subgroup  $G$  of  $\mathbf{R}$  playing the role that  $\mathbf{Q}$  played. Thus, we observe that for any two real numbers  $x$  and  $y$ , either  $x+G = y+G$  or the sets  $x+G$  and  $y+G$  are disjoint; using the axiom of choice, there exists a set  $X$  which meets each coset  $x+G$  in exactly one point. Thus  $(x+G) \cap (y+G) = \emptyset$  if  $x, y \in X$  and  $x \neq y$ , and  $X+G = \mathbf{R}$ . We define  $H = \{2m + n\xi : m, n \in \mathbf{Z}\}$ , so  $H$  is a subgroup of  $G$ , and  $G \setminus H = H + 1$ . Let  $A = X + H$ . Then  $A^C = X + H + 1$ . We make the following observation:  $A - A$  and  $A^C - A^C$  are disjoint from  $H + 1$ . For instance, if  $A - A$  meets  $H + 1$ , there exist  $x_1, x_2 \in X$  and  $h_1, h_2, h_3 \in H$  such that  $x_1 + h_1 - (x_2 + h_2) = h_3 + 1$ ; this implies that  $x_1 - x_2 = h_3 + 1 + h_2 - h_1 \in G$ , so  $x_1 = x_2$ , and hence that  $h_1 - h_2 - h_3 = 1$ , which is impossible. Similarly, if  $A^C - A^C$  meets  $H + 1$ , there exist  $x_1, x_2 \in X$  and  $h_1, h_2, h_3 \in H$  such that  $x_1 + h_1 + 1 - (x_2 + h_2 + 2 + 1) = h_3 + 1$ , which again leads first to  $x_1 = x_2$ , and then to  $h_1 - h_2 = h_3 + 1$ , which is impossible. Since (as is easy to see)  $H + 1$  is dense in  $\mathbf{R}$ , we see that neither  $A - A$  nor  $A^C - A^C$  contains any interval. Thus any measurable subset of either  $A$  or  $A^C$  must have measure 0, according to Lemma 9.33. Hence if  $B$  is any measurable set with  $m(B) > 0$ , and  $A \cap B$  is measurable, then also  $B \setminus (A \cap B) = B \cap A^C$  is measurable, so  $0 < m(B) = m(B \cap A) + m(B \cap A^C)$  implies that either  $A \cap B$  or  $A^C \cap B$  has positive measure, contradicting what we have proven. Thus  $A \cap B$  is not measurable for any measurable  $B$  with  $m(B) > 0$ . ■

## 9.7 Exercises

1. If  $\mu$  is a finitely additive real-valued set function on the algebra  $\mathcal{A}$ , then  $\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F)$ . Can you write a corresponding formula for  $\mu(E \cup F \cup G)$ ? for arbitrary finite unions?

2. Show that if  $\mu$  is a measure with domain  $\mathcal{A}$ , and  $E_n \in \mathcal{A}$  for each positive integer  $n$ , then

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n).$$

(In other words, countable additivity implies countable subadditivity.)

3. Show that if  $\mu$  is a measure, with domain  $\mathcal{A}$ , and if  $E_n \in \mathcal{A}$  with  $E_{n+1} \subset E_n$  for every  $n$ , and if  $\mu(E_1) < \infty$ , then

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

Show that this conclusion may be false if we leave out the assumption  $\mu(E_1) < \infty$ .

4. For any sequence of subsets  $E_n$  of a set  $X$ , we define

$$\liminf E_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} E_m.$$

Let  $\mu$  be a measure, with domain  $\mathcal{A}$ . Show that for any sequence  $\{E_n\}$  of sets in  $\mathcal{A}$ , we have  $\mu(\liminf E_n) \leq \liminf \mu(E_n)$ .

5. For any sequence of subsets  $E_n$  of a set  $X$ , we define

$$\limsup E_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m,$$

i.e., as the set of points which belong to  $E_n$  for infinitely many  $n$ . Let  $\mu$  be a measure, with domain  $\mathcal{A}$ . Show that if  $\{E_n\}$  is a sequence of sets in  $\mathcal{A}$  such that  $\sum_{n=1}^{\infty} \mu(E_n) < \infty$ , then  $\mu(\limsup E_n) = 0$ .

6. A real number  $x$  will be called a *Liouville number* if, for each positive integer  $n > 2$ , there exist integers  $p_k$  and  $q_k$ , with  $q_k \rightarrow +\infty$ , such that

$$0 < \left| x - \frac{p_k}{q_k} \right| < \frac{1}{q_k^n}.$$

(We showed in Chapter 1 that Liouville numbers are transcendental.) Show that the set of Liouville numbers has Lebesgue measure 0.

7. For each subset  $E$  of  $I = [0, 1]$ , define  $m_*(E) = 1 - m^*(I \setminus E)$ ;  $m_*(E)$  is called the *inner measure* of  $E$ . Show that  $E \subset I$  is  $m^*$ -measurable if and only if  $m^*(E) = m_*(E)$ . HINT: Show that if  $m^*(E) = m_*(E)$ , then for any interval  $J \subset I$  we have  $m^*(J) = m^*(J \cap E) + m^*(J \cap E^C)$ .

8. Let  $\mu$  be a measure, with domain  $\mathcal{A}$ . Define, for  $E$  and  $F$  in  $\mathcal{A}$ ,

$$\rho(E, F) = \mu((E \cap F^C) \cup (F \cap E^C)).$$

Show that  $\rho$  is a *pseudometric* on  $\mathcal{A}$ , i.e., that  $\rho(E, F) = \rho(F, E)$  for all  $E, F \in \mathcal{A}$ , and that  $\rho(E, F) \leq \rho(E, G) + \rho(G, F)$ , for any  $E, F, G \in \mathcal{A}$ .

9. Let  $\mu$  be a measure on a set  $X$  with domain  $\mathcal{A}$ . Let  $\mathcal{N} = \{N \in \mathcal{A} : \mu(N) = 0\}$ , and let  $\overline{\mathcal{N}}$  be the collection of all subsets of elements of  $\mathcal{N}$ . Let  $\overline{\mathcal{A}} = \{E \cup N : E \in \mathcal{A}, N \in \overline{\mathcal{N}}\}$ . Show that  $\overline{\mathcal{A}}$  is a  $\sigma$ -algebra, and that  $\mu$  has a natural extension to a measure  $\bar{\mu}$  with domain  $\overline{\mathcal{A}}$ , such that if  $E \in \overline{\mathcal{A}}$  and  $\bar{\mu}(E) = 0$ , then every subset of  $E$  belongs to  $\overline{\mathcal{A}}$ .

10. Show that if  $X$  and  $Y$  are topological spaces, and  $f : X \rightarrow Y$  is a continuous mapping, then  $f^{-1}(B)$  is a Borel set in  $X$  whenever  $B$  is a Borel set in  $Y$ . HINT: Consider  $\{E \subset Y : f^{-1}(E) \in \mathcal{B}(X)\}$ , where  $\mathcal{B}(X)$  is the  $\sigma$ -algebra of Borel sets in  $X$ .

11. For each subset  $A$  of the positive integers  $\mathbb{N}$ , let  $s_n(A) = \#\{k \in A : k \leq n\}$ , and define

$$d(A) = \lim_{n \rightarrow \infty} \frac{s_n(A)}{n},$$

provided this limit exists. Let  $\mathcal{A}$  denote the collection of those  $A$  for which this limit exists. (We call  $d(A)$  the *density* of  $A$ .)

(a) Show that  $d$  is not countably additive on  $\mathcal{A}$ .

(b) Show that  $\mathcal{A}$  is not an algebra.

HINT: Let  $A$  be the set of even numbers, and define  $B$  by the rule:  $k \in B$  if either  $k$  is even, and  $2^{2n} \leq k \leq 2^{2n+1}$  for some  $n$ , or  $k$  is odd and  $2^{2n-1} \leq k \leq 2^{2n}$ . Which of  $A$ ,  $B$ ,  $A \cap B$  are in  $\mathcal{A}$ ?

12. Show that for any  $\epsilon > 0$  there exists a dense open subset  $G$  of  $\mathbb{R}$  such that  $m^*(G) < \epsilon$ .

13. Let  $\mu$  be a measure, defined on the  $\sigma$ -algebra of Borel sets of a metric space  $X$ , such that  $\mu(X) = 1$ , and  $\mu(\{x\}) = 0$  for every  $x \in X$ . Show that for any  $x \in X$  and any  $\epsilon > 0$ , there exists an open neighborhood  $U$  of  $x$  with  $\mu(U) < \epsilon$ . If  $X$  is separable, show that there exists a dense open subset  $G$  of  $X$  with  $\mu(G) < \epsilon$ .

14. We call a number  $x \in [0, 1]$  *satanic* if the decimal expansion of  $x$  contains somewhere the sequence 666; in other words, if  $x = \sum_1^\infty a_k 10^{-k}$ , where  $a_k \in \{0, 1, \dots, 9\}$  and there exists  $n$  such that  $a_n = a_{n+1} = a_{n+2} = 6$ . Show that “almost all” numbers in  $[0, 1]$  are satanic, i.e., that  $m([0, 1] \setminus S) = 0$ , where  $S$  is the set of satanic numbers.

15. Show that Lebesgue–Stieltjes outer measure is a metric outer measure.

16. Show that Hausdorff  $p$ -dimensional outer measure is a metric outer measure.

17. Let  $g$  be a nondecreasing function on  $\mathbf{R}$ , continuous from the right, and let  $\mu = \mu_g$  be the Lebesgue–Stieltjes measure associated with  $g$  (see Example 9.19 for the definition of Lebesgue–Stieltjes outer measure). Show that  $\mu((a, b]) = g(b) - g(a)$  for every  $a < b$ .

18. Show that there exists  $E \subset [0, 1]$  such that

$$m^*(E) = m^*([0, 1] \setminus E) = 1.$$

HINT: Use the construction of Theorem 9.34.

## 9.8 Notes

9.1 The word *field* is often used in this context instead of algebra, especially in probability theory; similarly, one often sees the term  $\sigma$ -field used instead of  $\sigma$ -algebra.

9.2 Émile Borel showed in 1895 that there is a measure on the  $\sigma$ -algebra generated by the open intervals in  $\mathbf{R}$  which agrees with the usual length of intervals. Radon first discussed the idea of measure in spaces more general than  $\mathbf{R}^n$ .

9.3 Lebesgue had introduced Lebesgue outer measure in 1902, but the general notion seems to have originated with Carathéodory in his 1918 book. Carathéodory introduced one-dimensional Hausdorff measure in 1914; the general idea was published by Hausdorff in 1919.

9.4 Definition 9.4 and Theorem 9.22 are due to Carathéodory. Theorem 9.24 was first proved by Fréchet in 1924; the proof given was found (independently) by Hahn and Kolmogorov, the latter in his groundbreaking book on probability (1933).

9.5 The concept of metric outer measure, and Theorem 9.27, are due to Carathéodory. Metric outer measures are also known as Carathéodory outer measures.

- 9.6 The existence of a nonmeasurable set was first demonstrated by Vitali in 1905. Theorem 9.34 was proved by Van Vleck (an American) in 1908.
- 9.7 The result of Exercise 5, known as the first Borel–Cantelli lemma, is frequently used in probability. Lebesgue defined inner measure for bounded subsets of  $\mathbf{R}$ , and defined a bounded set to be measurable if its inner and outer measures agreed.



# 10

## Integration

We now turn to the topic of integration. While our main interest is in Lebesgue integration in  $\mathbf{R}^n$ , we develop the general theory of integration with respect to an arbitrary measure—it is not harder to do, and there will be occasion to use the extra generality.

### 10.1 Measurable Functions

Until further notice, let  $X$  be a set,  $\mathcal{A}$  a  $\sigma$ -algebra of subsets of  $X$ , and  $\mu$  a measure with domain  $\mathcal{A}$ . Let  $\overline{\mathbf{R}}$  denote the set of extended real numbers. The algebraic operations on  $\mathbf{R}$  extend partially to  $\overline{\mathbf{R}} = [-\infty, +\infty]$ : we put

- (a)  $\pm\infty + t = \pm\infty$  if  $t \in \mathbf{R}$ ;
- (b)  $+\infty + (+\infty) = +\infty$ , and  $(-\infty) + (-\infty) = -\infty$ ;
- (c)  $t(\pm\infty) = \pm\infty$  if  $t > 0$ ,  $t(\pm\infty) = \mp\infty$  if  $t < 0$ ;
- (d)  $+\infty + (-\infty)$  is undefined; and
- (e)  $0 \cdot (\pm\infty) = 0$ .

While (a)–(d) are natural choices, reflecting our experience with limits, we take special note of (e), which is a convention not derived from any theorem about limits.

**10.1 Definition.** Let  $f : X \rightarrow \overline{\mathbf{R}}$ . We say that  $f$  is  $\mathcal{A}$ -measurable if:

- (a)  $f^{-1}(+\infty) \in \mathcal{A}$ ,  $f^{-1}(-\infty) \in \mathcal{A}$ ; and  
 (b)  $f^{-1}(U) \in \mathcal{A}$  for every open  $U \subset \mathbf{R}$ .

When there is no possibility of ambiguity, we will simply say *measurable*, instead of  $\mathcal{A}$ -measurable. The following lemma offers us a menu of convenient tests for measurability:

**10.2 Lemma.** For any  $f : X \rightarrow \overline{\mathbf{R}}$ , the following are equivalent:

- (a)  $f$  is  $\mathcal{A}$ -measurable;  
 (b)  $\{x : f(x) > t\} \in \mathcal{A}$  for every real  $t$ ;  
 (c)  $\{x : f(x) \leq t\} \in \mathcal{A}$  for every real  $t$ ;  
 (d)  $\{x : f(x) < t\} \in \mathcal{A}$  for every real  $t$ ;  
 (e)  $\{x : f(x) \geq t\} \in \mathcal{A}$  for every real  $t$ ; and  
 (f)  $f^{-1}(B) \in \mathcal{A}$  whenever  $B$  is a Borel set in  $\mathbf{R}$ , or when  $B = \{+\infty\}$  or  $B = \{-\infty\}$ .

**Proof.** Clearly, (a) implies (b), since

$$\{x : f(x) > t\} = f^{-1}(\{+\infty\}) \cup f^{-1}((t, +\infty)).$$

Clearly, (b) is equivalent to (c), since  $\{x : f(x) \leq t\} = \{x : f(x) > t\}^C$ . We see that (c) implies (d) by observing that

$$\{x : f(x) < t\} = \bigcup_{n=1}^{\infty} \{x : f(x) \leq t - 1/n\}.$$

Clearly, (d) is equivalent to (e) in the same way that (b) is equivalent to (c).

Suppose that (e) holds. Since  $f^{-1}(\{+\infty\}) = \bigcap_{n=1}^{\infty} \{x : f(x) \geq n\}$ , and  $f^{-1}(\{-\infty\})$  is the complement of  $\bigcup_{n=1}^{\infty} \{x : f(x) \geq -n\}$ , we see that the sets  $f^{-1}(\{+\infty\})$  and  $f^{-1}(\{-\infty\})$  belong to  $\mathcal{A}$ . Now consider the family  $\mathcal{G}$  of all subsets  $G$  of  $\mathbf{R}$  such that  $f^{-1}(G) \in \mathcal{A}$ . It is clear that  $\mathcal{G}$  is a  $\sigma$ -algebra. Since (e) holds,  $\mathcal{G}$  contains every interval  $[t, +\infty)$  in  $\mathbf{R}$ ; it follows that  $\mathcal{G}$  contains every Borel set (see the discussion after Definition 9.9), so (f) holds.

Obviously, (f) implies (a). ■

**10.3 Lemma.** If  $f$  is measurable, so are  $|f|$  and  $f^2$ . If  $f$  and  $g$  are measurable, so are  $f + g$  and  $fg$ . If  $f_n$  is measurable for  $n = 1, 2, \dots$ , then so are  $\sup f_n$ ,  $\inf f_n$ ,  $\limsup f_n$ ,  $\liminf f_n$ .

**Proof.** Use the last lemma, and observe

- (a)  $\{x : |f(x)| < t\} = \{x : f(x) \in (-t, t)\};$   
 (b)  $\{x : f^2(x) < t\} = \{x : f(x) \in (-\sqrt{t}, \sqrt{t})\}$  for  $t \geq 0$ ;  
 (c)  $\{x : (f + g)(x) < t\} = \bigcup_{q \in \mathbf{Q}} \{x : f(x) < t - q, g(x) < q\};$   
 (d)  $fg = \frac{1}{4}((f + g)^2 - (f - g)^2);$   
 (e)  $\{x : \sup_n f_n(x) > t\} = \bigcup_n \{x : f_n(x) > t\};$   
 (f)  $\{x : \inf_n f_n(x) < t\} = \bigcup_n \{x : f_n(x) < t\};$  and  
 (g)  $\limsup f_n = \inf_n \sup_{m \geq n} f_m$ , and  $\liminf f_n = \sup_n \inf_{m \geq n} f_m$ . ■

We also remark that if  $f$  is measurable, so are  $f^+$  and  $f^-$ . For instance,  $f^+ = \frac{1}{2}(f + |f|)$ , so Lemma 10.3 applies.

**10.4 Definition.** For each subset  $A$  of  $X$ , we define the characteristic function or indicator function of  $A$  by

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

It is obvious that  $1_A$  is  $\mathcal{A}$ -measurable if and only if  $A \in \mathcal{A}$ .

**10.5 Definition.** A real-valued function  $f$  on  $X$  is called simple if it assumes only finitely many distinct values.

It is easy to see that  $f$  is simple if and only if  $f$  can be expressed as a finite linear combination of indicator functions; if  $c_1, \dots, c_n$  are the distinct values of  $f$ , and  $A_j = \{x : f(x) = c_j\}$ , then  $f = \sum_1^n c_j 1_{A_j}$  is one such expression, which we call the *canonical representation* of  $f$ . Clearly,  $f$  is measurable if and only if each  $A_j \in \mathcal{A}$  in the canonical representation of  $f$ .

**10.6 Lemma.** If  $f$  is a nonnegative measurable function, then there exists a sequence  $(f_n)$  of nonnegative simple measurable functions, such that  $0 \leq f_n(x) \leq f_{n+1}(x)$  for all  $n \in \mathbf{N}$  and all  $x \in X$ , and such that  $\lim f_n(x) = f(x)$  for all  $x \in X$ .

**Proof.** It suffices to put

$$f_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{A_{n,k}} + n 1_{B_n},$$

where  $A_{n,k} = \{x : (k-1)2^{-n} < f(x) \leq k2^{-n}\}$  and  $B_n = \{x : f(x) > n\}$ . It is routine to verify that the sequence  $(f_n)$  has the desired properties. ■

**10.7 Corollary.** *If  $f$  is a measurable function, then there exists a sequence of simple measurable functions which converges pointwise to  $f$ ; when  $f$  is bounded, there is such a sequence where the convergence is uniform.*

**Proof.** We apply Lemma 10.6 to the functions  $f^+$  and  $f^-$ . Note that if the function  $f$  in Lemma 10.6 is bounded, then the construction yields simple  $f_n$  with  $f(x) - f_n(x) \leq 2^{-n}$  for all  $n \geq \sup f$ , so the convergence is uniform. ■

**10.8 Definition.** Suppose  $P(x)$  is a proposition, for each  $x \in X$ . We say that  $P(x)$  holds almost everywhere (abbreviated a.e.) or for almost all  $x$  (abbreviated a.a.  $x$ ) if the set  $F = \{x : P(x) \text{ is false}\} \in \mathcal{A}$  and  $\mu(F) = 0$ . If there is more than one measure that might be referred to, we write a.e. ( $\mu$ ) or  $\mu$ -a.a.  $x$ .

Here is an example of the usage.

**10.9 Theorem.** *If  $f$  is a Lebesgue measurable function on  $R^n$ , there exists a Borel measurable function  $g$  such that  $f = g$  a.e.*

**Proof.** The phrasing here implies that the measure  $\mu$  is taken here to be Lebesgue measure  $m$ . The theorem asserts that there is a Borel measurable  $g$  such that  $m\{x : f(x) \neq g(x)\} = 0$ . This is true if  $f = 1_A$ , where  $A$  is Lebesgue measurable, by Theorem 9.29, so it's true for any simple Lebesgue measurable  $f$ . By Corollary 10.7, combined with Lemma 10.3, it follows for any Lebesgue measurable  $f$ . ■

## 10.2 Integration

In this section we define the integral of certain functions, with respect to a measure  $\mu$ . We begin with nonnegative simple functions, extend to nonnegative measurable functions, and then to the class of summable (integrable) functions. This section also includes the three basic limit theorems of integration theory: the monotone convergence theorem, Fatou's lemma, and the dominated convergence theorem.

**10.10 Definition.** Let  $f$  be a nonnegative simple measurable function. If  $f = \sum_{k=1}^n c_k 1_{A_k}$  is the canonical representation of  $f$  as a finite linear combination of indicator functions, we define

$$\int f \, d\mu = \sum_{k=1}^n c_k \mu(A_k).$$

If  $A \in \mathcal{A}$ , we define

$$\int_A f \, d\mu = \int 1_A f \, d\mu.$$

We observe that  $0 \leq \int f d\mu \leq \infty$ , and that  $\int f d\mu = 0$  if and only if  $\mu(A_k) = 0$  for each  $k$  with  $c_k \neq 0$ , i.e., if and only if  $f = 0$  a.e., while  $\int f d\mu < \infty$  if and only if  $\mu(A_k) < \infty$  for each  $k$  such that  $c_k \neq 0$ . (Recall our convention that  $0 \cdot \infty = 0$ .)

**10.11 Lemma.** *If  $f$  and  $g$  are nonnegative simple measurable functions, then:*

- (a)  $\int f d\mu \leq \int g d\mu$  if  $f \leq g$ ;
- (b)  $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ ;
- (c)  $\int (tf) d\mu = t \int f d\mu$  for each  $t \geq 0$ ; and
- (d) the function  $E \mapsto \int_E f d\mu$  is a measure on  $\mathcal{A}$ .

**Proof.** Let

$$f = \sum_{j=1}^m a_j \mathbf{1}_{A_j}, \quad g = \sum_{k=1}^n b_k \mathbf{1}_{B_k}$$

be the canonical representations of  $f$  and  $g$ . We see that  $X = \bigcup_{j=1}^m A_j = \bigcup_{k=1}^n B_k$ , each of these unions being disjoint. Then

$$\int f d\mu = \sum_{j=1}^m a_j \mu(A_j) = \sum_{j=1}^m a_j \sum_{k=1}^n \mu(A_j \cap B_k) = \sum_{j,k} a_j \mu(A_j \cap B_k)$$

and, similarly,

$$\int g d\mu = \sum_{j,k} b_k \mu(A_j \cap B_k).$$

Suppose  $f \leq g$ . Then for each  $j, k$ , we see that  $a_j \leq b_k$  whenever  $A_j \cap B_k \neq \emptyset$ , so  $\int f d\mu \leq \int g d\mu$ , and (a) is proved.

Now let  $\{c_1, \dots, c_r\}$  be the distinct values assumed by  $f + g$ . Then

$$\begin{aligned} \int (f + g) &= \sum_{i=1}^r c_i \mu(\{f + g = c_i\}) \\ &= \sum_{i=1}^r c_i \mu\left(\bigcup_{a_j + b_k = c_i} A_j \cap B_k\right) \\ &= \sum_{i=1}^r c_i \sum_{a_j + b_k = c_i} \mu(A_j \cap B_k) \\ &= \sum_{j,k} (a_j + b_k) \mu(A_j \cap B_k) \\ &= \sum_{j=1}^m \sum_{k=1}^n a_j \mu(A_j \cap B_k) + \sum_{k=1}^n \sum_{j=1}^m b_k \mu(A_j \cap B_k) \end{aligned}$$

$$= \sum_{j=1}^m a_j \mu(A_j) + \sum_{k=1}^n b_k \mu(B_k) = \int f \, d\mu + \int g \, d\mu,$$

so (b) is proved. The truth of (c) is obvious, and (d) is an immediate consequence of Proposition 9.12. ■

We now extend the domain of integration to a larger class of functions.

**10.12 Definition.** If  $f$  is a nonnegative measurable function, we define

$$\int f \, d\mu = \sup \left\{ \int g \, d\mu : 0 \leq g \leq f, \, g \text{ simple measurable} \right\}.$$

As before, we define  $\int_A f \, d\mu$  to be  $\int f \mathbf{1}_A \, d\mu$ , for any  $A \in \mathcal{A}$ .

It is trivial that  $0 \leq \int f \, d\mu \leq +\infty$ , that  $\int f \, d\mu \leq \int g \, d\mu$  whenever  $f \leq g$ , and that  $\int f \, d\mu$  retains its previous meaning when  $f$  is simple.

**10.13 Proposition.** If  $f \geq 0$  is a measurable function, then  $\int f \, d\mu = 0$  if and only if  $f = 0$  a.e.

**Proof.** If  $f = 0$  a.e., then  $g = 0$  a.e. whenever  $0 \leq g \leq f$ , so  $\int f \, d\mu = 0$ . Now suppose that  $\int f \, d\mu = 0$ , and let  $E = \{x : f(x) > 0\}$ . Then  $E = \bigcup_{n=1}^{\infty} E_n$ , where  $E_n = \{x : f(x) \geq 1/n\}$ . Let  $g_n = (1/n)\mathbf{1}_{E_n}$ ; then  $g_n$  is a simple measurable function with  $0 \leq g_n \leq f$ , so  $(1/n)\mu(E_n) = \int g_n \, d\mu \leq \int f \, d\mu = 0$ . Thus  $\mu(E_n) = 0$  for each  $n \in \mathbf{N}$ , and hence  $\mu(E) = 0$ . ■

The next theorem generalizes Lemma 10.11(d).

**10.14 Theorem.** Let  $f \geq 0$  be a measurable function, and define the set function  $\nu$  on  $\mathcal{A}$  by  $\nu(A) = \int_A f \, d\mu$ . Then  $\nu$  is a measure on  $\mathcal{A}$ .

**Proof.** Let  $\mathcal{S}$  be the set of all simple measurable functions  $g$ , with  $0 \leq g \leq f$ . For each  $g \in \mathcal{S}$ , define the set function  $\nu_g$  by  $\nu_g(A) = \int_A g \, d\mu$ ; each  $\nu_g$  is a measure, by Lemma 10.11(d), and  $\nu = \sup\{\nu_g : g \in \mathcal{S}\}$ . We observe that if  $g \leq h$ , then  $\nu_g \leq \nu_h$ ; also, for any  $g, h \in \mathcal{S}$ , we have  $\max\{g, h\} \in \mathcal{S}$ . Thus, if  $\mathcal{M} = \{\nu_g : g \in \mathcal{S}\}$ , then  $\mathcal{M}$  is a collection of measures satisfying the hypotheses of Theorem 9.13, and it follows from that theorem that  $\nu$  is a measure. ■

Before extending the definition of the integral to a larger class of measurable functions, we obtain a key result about passing to the limit under the integral sign. It is known as the *monotone convergence theorem*, and is due to Beppo Levi.

**10.15 Theorem.** If  $(f_n)$  is a sequence of measurable functions, with  $0 \leq f_n \leq f_{n+1}$  for every  $n$ , then

$$\int \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

**Proof.** We note that  $f = \sup_n f_n$  is measurable by Lemma 10.3, so  $\int f d\mu$  is defined. Since  $f_n \leq f_{n+1}$ , we have  $\int f_n d\mu \leq \int f_{n+1} d\mu$ , so  $\lim \int f_n d\mu$  exists (in  $\overline{\mathbf{R}}$ ), and since  $f_n \leq f$  for every  $n$ ,  $\lim \int f_n d\mu \leq \int f d\mu$ . The reverse inequality is not as trivial.

Let  $g$  be a simple measurable function, with  $0 \leq g \leq f$ . Fix  $\epsilon$ ,  $0 < \epsilon < 1$ . Let  $A_n = \{x \in X : f_n(x) \geq (1 - \epsilon)g(x)\}$ , and observe that  $A_n \subset A_{n+1}$  for all  $n$ , since the sequence  $f_n$  is monotone increasing. Also, observe that  $\bigcup_{n=1}^{\infty} A_n = X$ , since  $\lim f_n = \sup f_n \geq g$ . Hence

$$\int f_n d\mu \geq \int_{A_n} f_n d\mu \geq (1 - \epsilon) \int_{A_n} g d\mu;$$

since  $E \mapsto \int_E g d\mu$  is a measure by Lemma 10.11, it follows from Proposition 9.11 that  $\lim \int f_n d\mu \geq (1 - \epsilon) \int g d\mu$  for every  $\epsilon > 0$ . It follows that  $\lim \int f_n d\mu \geq \int g d\mu$ ; since  $g$  was any nonnegative simple function with  $g \leq f$ , it follows that  $\lim \int f_n d\mu \geq \int f d\mu$ . ■

We note that this theorem gives an alternate proof of Theorem 10.14 above, based on the observation that  $1_A$  is the limit of the increasing sequence  $\{1_{A_n}\}$ , if  $A$  is the union of the increasing sequence  $\{A_n\}$ .

**10.16 Corollary.** If  $f \geq 0$  and  $g \geq 0$  are measurable functions, then

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu.$$

**Proof.** Choose, using Lemma 10.6, simple  $f_n$  which increase to  $f$ , and simple  $g_n$  which increase to  $g$ . Then  $f_n + g_n$  are simple, and increase to  $f + g$ ; since  $\int (f_n + g_n) d\mu = \int f_n d\mu + \int g_n d\mu$  by Lemma 10.11, this corollary now follows from Theorem 10.15. ■

We next give another very useful result relating pointwise limits and integration. The following theorem is known as *Fatou's Lemma*.

**10.17 Theorem.** If  $(f_n)$  is a sequence of nonnegative measurable functions, then

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu.$$

**Proof.** Let  $g_n = \inf_{m \geq n} f_m$ ; then each  $g_n$  is a nonnegative measurable function,  $g_n \leq g_{n+1}$  for each  $n$ , and  $\lim g_n = \liminf f_n$ , so

$$\int \liminf f_n d\mu = \int \lim g_n d\mu = \lim \int g_n d\mu$$

by the monotone convergence theorem (Theorem 10.15). But for any  $m \geq n$ , we have  $g_n \leq f_m$ , so  $\int g_n d\mu \leq \int f_m d\mu$ , and so

$$\int g_n d\mu \leq \inf_{m \geq n} \int f_m d\mu,$$

so  $\lim \int g_n d\mu \leq \liminf \int f_n d\mu$ . ■

**10.18 Definition.** Let  $f$  be a measurable function. We define

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu,$$

if at least one of the integrals on the right is finite. We say that  $f$  is integrable, or summable, and write  $f \in L(\mu)$ , if both  $\int f^+ d\mu < +\infty$  and  $\int f^- d\mu < +\infty$ , so that  $\int f d\mu$  is a real number.

**10.19 Proposition.** If  $f$  is a measurable function, then  $f$  is summable if and only if  $\int |f| d\mu < +\infty$ . If  $f$  is summable,  $g$  is measurable, and  $|g| \leq |f|$ , then  $g$  is summable.

**Proof.** Since  $|f| = f^+ + f^-$ , Corollary 10.16 shows that  $f \in L(\mu)$  if and only if  $|f| \in L(\mu)$ . It follows that if  $f \in L(\mu)$  and if  $g$  is a measurable function with  $|g| \leq |f|$ , then  $g \in L(\mu)$ . ■

**10.20 Proposition.** The summable functions form a vector space, and integration is a linear operation on  $L(\mu)$ ; i.e., if  $f, g \in L(\mu)$  and  $c \in R$ , then  $cf \in L(\mu)$  and  $f + g \in L(\mu)$ ; furthermore,  $\int (cf) d\mu = c \int f d\mu$  and  $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ .

**Proof.** The statements regarding  $cf$  are trivial. Since  $|f + g| \leq |f| + |g|$ ,  $f + g$  is summable whenever  $f$  and  $g$  are, by Proposition 10.19 and Corollary 10.16. Now  $f + g = (f + g)^+ - (f + g)^- = f^+ - f^- + g^+ - g^-$ , so

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+,$$

so by Corollary 10.16 it follows that

$$\int (f + g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f + g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu,$$

which leads to the desired  $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ . ■

The third and last of the major results about passing to the limit under the integral sign is known as Lebesgue's *dominated convergence theorem*.



**10.21 Theorem.** Suppose that  $(f_n)$  is a sequence of measurable functions, that  $f_n \rightarrow f$ , as  $n \rightarrow \infty$ , and that  $|f_n| \leq g$  for all  $n$ , where  $g$  is summable. Then  $f$  is summable, and  $\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu$ .

**Proof.** The summability of  $f$  follows from Proposition 10.19, since  $|f| \leq g$ , and  $g$  is summable. Since  $g \pm f_n \geq 0$  for every  $n$ , we can apply Fatou's lemma (Theorem 10.17) to get

$$\int g d\mu \pm \int f d\mu = \int (g \pm f) d\mu \leq \liminf \int (g \pm f_n) d\mu.$$

Now

$$\liminf \int (g + f_n) d\mu = \int g d\mu + \liminf \int f_n d\mu$$

and

$$\liminf \int (g - f_n) d\mu = \int g d\mu - \limsup \int f_n d\mu$$

as we saw in Chapter 2; utilizing linearity (Proposition 10.20), we have

$$\limsup \int f_n d\mu \leq \int f d\mu \leq \liminf \int f_n d\mu,$$

which gives the theorem. ■

Theorem 10.21, as well as Theorem 10.15, refer to pointwise convergence: the hypothesis is that  $f_n(x) \rightarrow f(x)$  for every  $x \in X$ , as  $n \rightarrow \infty$ . It is easy to see that the hypothesis may be weakened to *almost everywhere convergence*:  $f_n(x) \rightarrow f(x)$  for all  $x \in Y$ , where  $Y \in \mathcal{A}$  and  $\mu(X \setminus Y) = 0$ . For if we put  $g_n = 1_Y f_n$  and  $g = 1_Y f$ , then  $g_n \rightarrow g$  everywhere, and  $\int g_n d\mu = \int f_n d\mu$ , etc. We note that some sort of condition beyond pointwise convergence (such as monotonicity in Theorem 10.15, or the dominating function in Theorem 10.21) is necessary in order to "pass to the limit under the integral sign." For instance, with  $X = [0, 1]$  and  $\mu = m$  (Lebesgue measure), if  $f_n = n 1_{(0, 1/n)}$ , we see that  $f_n(x) \rightarrow 0$  for every  $x$ , but  $\int f_n dm = 1$ . This example shows incidentally that the inequality in Fatou's lemma may be strict, even when the  $\liminf$  is a limit.

## 10.3 Lebesgue and Riemann Integrals

In this section we fix a closed bounded interval  $[a, b]$  in  $\mathbf{R}$ , and examine the relation between the Riemann integral  $\int_a^b f(x) dx$  studied in Chapter 5 and the Lebesgue integral  $\int f dm = \int_{[a, b]} f dm$ .

**10.22 Proposition.** If  $f$  is continuous on  $[a, b]$ , then  $\int_a^b f(x) dx = \int f dm$ .

**Proof.** Let us call a function  $g : [a, b] \rightarrow \mathbf{R}$  a *step function* if there exists a partition  $(x_k)_{k=0}^n$  of  $[a, b]$ , and a sequence  $(c_k)_{k=1}^n$  such that  $g(t) = c_k$  for all  $t \in (x_{k-1}, x_k)$ ,  $1 \leq k \leq n$ . Step functions are a special case of simple functions, and it is very easy to see that if  $g$  is a step function, then the Riemann and Lebesgue integrals of  $g$  coincide. Now if  $f \in C([a, b])$ , there exists a sequence of step functions  $(g_n)$  which converges uniformly to  $f$  (see the proof of Theorem 3.20). It follows that  $\int_a^b f(x) dx = \lim \int_a^b g_n(x) dx$  (by Theorem 5.34), and that  $\int f dm = \lim \int g_n dm$  by the analogous theorem for Lebesgue integrals, or by the dominated convergence theorem. ■

**10.23 Theorem.** A bounded real-valued function on  $[a, b]$  is Riemann integrable if and only if it is continuous at almost every point of  $[a, b]$ ; in this case, its Riemann integral and Lebesgue integral are equal.

**Proof.** Let  $f$  be a bounded real function on  $[a, b]$ . Let  $\mathcal{L} = \{g \in C[a, b] : g \leq f\}$ , and let  $\mathcal{U} = \{h \in C[a, b] : f \leq h\}$ . The classes  $\mathcal{U}$  and  $\mathcal{L}$  are not empty, since  $f$  is bounded. We set

$$f_* = \sup\{g : g \in \mathcal{L}\} \quad \text{and} \quad f^* = \inf\{h : h \in \mathcal{U}\}.$$

We observe that for any real  $t$ ,  $\{x : f^*(x) < t\} = \bigcup_{h \in \mathcal{U}} \{x : h(x) < t\}$ , which is open in  $[a, b]$  since every  $h \in \mathcal{U}$  is continuous. Similarly, for any real  $t$ ,  $\{x : f_*(x) > t\}$  is open in  $[a, b]$ . In particular, we see that  $f^*$  and  $f_*$  are Borel measurable, with  $f_* \leq f \leq f^*$ . If  $f^*(x) = f_*(x)$ , then for any  $\epsilon > 0$  the set

$$\{y : |f(y) - f(x)| < \epsilon\} \supset \{y : f^*(y) < f(x) + \epsilon\} \cap \{y : f_*(y) > f(x) - \epsilon\}$$

is a neighborhood of  $x$ , so  $f$  is continuous at  $x$ . Conversely, if  $f$  is continuous at  $x$ , it is not hard to construct a function  $h \in \mathcal{U}$  with  $h(x) < f(x) + \epsilon$  for any given  $\epsilon > 0$ , so  $f^*(x) = f(x)$ ; similarly,  $f_*(x) = f(x)$ . Thus  $f$  is continuous at  $x$  if and only if  $f_*(x) = f^*(x)$ . Since  $f_* \leq f \leq f^*$ , it follows that if  $f^* = f_*$  a.e. then  $f = f^*$  a.e., and hence  $f$  is Lebesgue measurable, and  $\int f dm = \int f^* dm = \int f_* dm$ .

Suppose that  $f$  is Riemann integrable. We see from Theorem 5.17 that

$$\int_a^b f = \sup \left\{ \int_a^b g : g \in \mathcal{L} \right\} = \inf \left\{ \int_a^b h : h \in \mathcal{U} \right\},$$

so  $\int_a^b f \leq \int f_* dm \leq \int f^* dm \leq \int_a^b f$ . Hence  $\int f^* dm = \int f_* dm$ , and thus  $f^* = f_*$  a.e. by Proposition 10.13, so  $f$  is continuous almost everywhere.

We next observe that there exists a sequence  $(h_n)$  in  $\mathcal{U}$  such that  $h_{n+1} \leq h_n$  for every  $n$ , and  $\lim h_n(x) = f^*(x)$  for every  $x \in [a, b]$ . Indeed, we know there exists a countable subset  $\mathcal{D}$  of  $C([a, b])$  which is dense in  $C([a, b])$ , with its usual topology of uniform convergence. (We can take  $\mathcal{D}$  for instance to be the set of polynomials with rational coefficients.) Then

it is clear that  $f^* = \inf\{h : h \in \mathcal{D} \cap \mathcal{U}\}$ . If  $\mathcal{D} \cap \mathcal{U} = \{H_1, H_2, \dots\}$ , we can take  $h_n = \min\{H_1, H_2, \dots, H_n\}$ , and have the desired sequence. Similarly, there exists a sequence  $(g_n)$  in  $\mathcal{L}$  with  $g_n \leq g_{n+1}$  for all  $n$ , and  $\lim g_n = f_*$  pointwise. It follows from the monotone convergence theorem that  $\int f^* d\mu = \lim \int h_n d\mu$ , and  $\int f_* d\mu = \lim \int g_n d\mu$ . Hence if  $f^* = f_*$ , there exist continuous functions  $g \leq f$  and  $h \geq f$  such that  $\int h d\mu - \int g d\mu < \epsilon$ . By Theorem 5.17,  $f$  is Riemann integrable. ■

## 10.4 Inequalities for Integrals

In this section, we obtain a handful of inequalities which are very useful in many situations in analysis. The first is known as *Chebyshev's inequality*.

**10.24 Proposition.** *Let  $f$  be a nonnegative measurable function, and let  $t$  be a positive real number. Then*

$$\mu(\{x : f(x) \geq t\}) \leq \frac{1}{t} \int f d\mu.$$

**Proof.** If  $A = \{x : f(x) \geq t\}$ , then  $t1_A \leq f$ , so  $t\mu(A) \leq \int f d\mu$  from the definition of the integral. ■

We remark that we gave this argument earlier, in proving Proposition 10.13. The next result is called *Jensen's inequality*.

**10.25 Theorem.** *Let  $\mu$  be a measure, with  $\mu(X) = 1$ . If  $f$  is a summable function on  $X$ , taking values in an interval  $J$ , and  $\varphi$  is a convex function on  $J$ , then*

$$\varphi\left(\int f d\mu\right) \leq \int \varphi(f) d\mu.$$

**Proof.** Let  $c = \int f d\mu$ ; since  $\mu(X) = 1$ , we see that  $c \in J$ , so  $\varphi(c)$  is defined. Because  $\varphi$  is convex, there exists (see Corollary 4.16) a real number  $m$  with the property that

$$\varphi(y) \geq \varphi(c) + m(y - c) \quad \text{for all } y \in J;$$

in other words, such that the line with slope  $m$  through the point  $(c, \varphi(c))$  lies under the graph of  $\varphi$ . It follows that

$$\varphi(f(x)) \geq \varphi(c) + m(f(x) - c) \tag{10.1}$$

for all  $x \in X$ . This implies that  $\int [\varphi(f)]^- d\mu < \infty$ , so  $\int \varphi(f) d\mu$  is well-defined. Integrating the inequality (10.1), and using the fact that  $\int f d\mu = 1$ , we arrive at

$$\int \varphi(f) d\mu \geq \varphi(c) = \varphi\left(\int f d\mu\right)$$

as was to be proved. Note that the possibility that  $\int \varphi(f) d\mu = +\infty$  is not excluded. ■

**10.26 Corollary.** Let  $p_j > 0$ ,  $j = 1, \dots, n$ , with  $\sum p_j = 1$ . If  $\varphi$  is a convex function on the interval  $J$ , and  $x_j \in J$  for  $j = 1, \dots, n$ , then

$$\varphi\left(\sum_{j=1}^n x_j p_j\right) \leq \sum_{j=1}^n \varphi(x_j) p_j.$$

This corollary follows from Theorem 10.25 if we take  $X = \{1, \dots, n\}$ , and define  $\mu(\{j\}) = p_j$ .

**10.27 Corollary.** If  $\mu$  is a measure with  $\mu(X) = 1$ , and  $f$  is a nonnegative summable function on  $X$ , then

$$\exp\left(\int \log f d\mu\right) \leq \int f d\mu.$$

**Proof.** Apply Theorem 10.25 with  $\varphi(x) = e^x$ , and with  $\log f$  playing the role of  $f$ . Note that the case  $\int \log f d\mu = -\infty$  is not excluded; the inequality is of course then trivial (we take  $\exp(-\infty) = 0$  by convention). ■

Taking  $X = \{1, \dots, n\}$ , and  $\mu(\{j\}) = 1/n$  for each  $j$ , the last corollary reduces to the inequality

$$\left(\prod_{j=1}^n x_j\right)^{1/n} \leq \frac{1}{n} \sum_{j=1}^n x_j,$$

which is the classical *inequality of the geometric and arithmetic means*. The next inequality is sometimes referred to as *Liapounov's inequality*.

**10.28 Corollary.** Let  $\mu$  be a measure with  $\mu(X) = 1$  and  $1 \leq p < \infty$ . Then for any nonnegative measurable  $f$ , we have

$$\left(\int f d\mu\right)^p \leq \int f^p d\mu.$$

**Proof.** Take  $J = [0, \infty)$  and  $\varphi(x) = x^p$  in Theorem 10.25. ■

**10.29 Theorem.** Let  $p \in \mathbf{R}$ ,  $1 < p < \infty$ , and let  $q = p/(p-1)$ . If  $f$  and  $g$  are nonnegative measurable functions, then

$$\int fg d\mu \leq \left(\int f^p d\mu\right)^{1/p} \left(\int g^q d\mu\right)^{1/q}.$$

**Proof.** If  $\int g^q d\mu = \infty$ , then the inequality is trivial. If  $\int g^q d\mu = 0$ , then  $g = 0$  a.e., and again the result is trivial. The inequality is unaffected if we multiply  $g$  by a constant, so we may assume that  $\int g^q d\mu = 1$ . Define the measure  $\sigma$  by  $\sigma(E) = \int_E g^q d\mu$ ; then  $\sigma(X) = 1$ , and  $\int h d\sigma = \int h g^q d\mu$  for any measurable function  $h$ . In particular, we have

$$\begin{aligned} \int f g d\mu &= \int f g^{1-q} g^q d\mu = \int f g^{1-q} d\sigma \\ &\leq \left( \int f^p g^{(1-q)p} d\sigma \right)^{1/p} \quad (\text{by Corollary 10.28}) \\ &= \left( \int f^p d\mu \right)^{1/p} \quad (\text{since } p + q = pq), \end{aligned}$$

which was to be proved. ■

**10.30 Definition.** Let  $1 \leq p < \infty$ . We say that  $f \in L^p(\mu)$  if  $f$  is measurable and  $\int |f|^p d\mu < \infty$ ; we define the  $L^p$ -norm of  $f$  as

$$\|f\|_p = \left( \int |f|^p d\mu \right)^{1/p}.$$

We say that  $f \in L^\infty(\mu)$  if there exists  $C < +\infty$  such that  $|f| \leq C$  a.e. ( $\mu$ ); we define  $\|f\|_\infty$  to be the infimum of all such  $C$ .

The next result is known as *Hölder's inequality*. We note that when  $p = 2$ , then also  $q = 2$ , and the theorem becomes the Schwarz (Cauchy-Schwarz-Bunyakovsky) inequality. We note that taking  $g = 1$ , we can deduce Corollary 10.28 from the Hölder inequality.

**10.31 Corollary.** Let  $f \in L^p(\mu)$  and let  $g \in L^q(\mu)$ , where  $(1/p) + (1/q) = 1$ . Then  $fg$  is summable, and

$$\left| \int fg d\mu \right| \leq \|f\|_p \|g\|_q.$$

**Proof.** If  $1 < p < \infty$ , this follows at once from Theorem 10.29. If  $p = 1$ , then  $q = \infty$ , and the result is obvious. ■

**10.32 Theorem.** Let  $1 \leq p \leq +\infty$ . If  $f$  and  $g$  belong to  $L^p(\mu)$ , then so does  $f + g$ , and  $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ .

**Proof.** The cases  $p = 1$  and  $p = \infty$  are trivial, so assume  $1 < p < \infty$ . The integrability of  $|f + g|^p$  follows from the elementary inequality  $(a + b)^p \leq 2^p(a^p + b^p)$ , valid for any positive numbers  $a$  and  $b$ . We will make use of the

following calculation: if  $h \geq 0$  is any measurable function, and  $q = p/(p-1)$ , then

$$\|h^{p-1}\|_q = \left( \int h^{pq-q} d\mu \right)^{1/q} = \left( \int h^p d\mu \right)^{(p-1)/p} = \|h\|_p^{p-1}.$$

Using the Hölder inequality, we have

$$\begin{aligned} \|f + g\|_p^p &= \int |f + g|^{p-1} |f + g| d\mu \\ &\leq \int |f + g|^{p-1} (|f| + |g|) d\mu \\ &\leq \|f\|_p \| |f + g|^{p-1} \|_q + \|g\|_p \| |f + g|^{p-1} \|_q \\ &= (\|f\|_p + \|g\|_p) \|f + g\|_p^{p-1}, \end{aligned}$$

so dividing both sides of this equation by  $\|f + g\|_p^{p-1}$  yields the theorem. (If  $\|f + g\|_p = 0$ , there was nothing to prove.) ■

The result of the last theorem is known as *Minkowski's inequality*. Let us define the distance between two elements of  $L^p(\mu)$  by  $\rho(f, g) = \|f - g\|_p$ . Obviously,  $\rho(f, g) \geq 0$  for all  $f$  and  $g$  in  $L^p(\mu)$ , and  $\rho(f, f) = 0$ ; Minkowski's inequality easily implies that  $\rho(f, h) \leq \rho(f, g) + \rho(g, h)$  for any  $f, g, h \in L^p(\mu)$ . Thus  $\rho$  is a *pseudometric* on  $L^p(\mu)$ ; it fails to be a metric because  $\rho(f, g) = 0$  does not imply that  $f = g$ . It does, of course, imply that  $f = g$  a.e. ( $\mu$ ), and we usually slur the distinction between functions which are equal and those which are merely equal almost everywhere. To obtain a true metric space, one can consider the set of equivalence classes under the relation  $\equiv$ , defined by  $f \equiv g$  if and only if  $f = g$  a.e. ( $\mu$ ). We next show that this metric space is complete.

**10.33 Theorem.** Let  $f_n \in L^p$  for each  $n \in \mathbf{N}$ . If the series  $\sum \|f_n\|_p$  converges, then there exists  $F \in L^p$  such that the series  $\sum f_n$  converges to  $F$  in  $L^p$ , i.e., such that

$$\left\| F - \sum_{k=1}^n f_k \right\|_p \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

**Proof.** Let  $F_n = \sum_{k=1}^n f_k$  and  $G_n = \sum_{k=1}^n |f_k|$ , so  $|F_n| \leq G_n$  for every  $n$ . Obviously,  $\{G_n\}$  is an increasing sequence; denote its limit by  $G$ . Thus  $G$  is a measurable function,  $0 \leq G \leq \infty$ . By Minkowski's inequality we have

$$\left( \int G_n^p d\mu \right)^{1/p} \leq \sum_{k=1}^n \left( \int |f_k|^p d\mu \right)^{1/p} \leq \sum_{k=1}^{\infty} \|f_k\|_p = M < \infty$$

for every  $n$ ; by the monotone convergence theorem (Theorem 10.15) we see that  $\int G^p d\mu = \lim \int G_n^p d\mu \leq M^p < \infty$  (in particular, that  $G < \infty$  almost

everywhere). It follows that the series  $\sum f_n$  converges absolutely a.e. ( $\mu$ ), i.e., that  $F_n \rightarrow F$  a.e., where  $F$  is a measurable function with  $|F| \leq G$ , so  $F \in L^p$ . Also, we have

$$|F - F_n| = \left| \sum_{k=n+1}^{\infty} f_k \right| \leq \sum_{k=n+1}^{\infty} |f_k| \leq G,$$

so  $|F - F_n|^p \leq G^p$  for every  $n$ . Thus we can apply the dominated convergence theorem (Theorem 10.21) to conclude that  $\int |F - F_n|^p d\mu \rightarrow 0$  as  $n \rightarrow \infty$ . ■

**10.34 Corollary.** For each  $p$ ,  $1 \leq p < \infty$ ,  $L^p$  is a complete metric space; that is, if  $(g_n)$  is a Cauchy sequence in  $L^p$ , then there exists  $g \in L^p$  such that  $\|g - g_n\|_p \rightarrow 0$  as  $n \rightarrow \infty$ .

**Proof.** Choose an integer  $n_1$  such that  $\|g_n - g_m\|_p < 1/2$  for all  $n, m \geq n_1$ ; inductively, choose  $n_1 < n_2 < \dots$  such that  $\|g_{n_{k+1}} - g_{n_k}\|_p < 2^{-k}$  for every  $k$ . Let  $f_k = g_{n_k} - g_{n_{k-1}}$ , where  $g_{n_0} = 0$ , so  $\|f_k\|_p < 2^{-k}$  and  $\sum_{k=1}^n f_k = g_{n_n}$ . By Theorem 10.33 there exists  $g \in L^p$  such that  $\|g - g_{n_k}\|_p \rightarrow 0$  as  $k \rightarrow \infty$ . Since  $\{g_n\}$  is Cauchy, it follows also that  $\|g - g_n\|_p \rightarrow 0$  as  $n \rightarrow \infty$ . ■

Theorem 10.33 and Corollary 10.34 are both known as the *Riesz-Fischer theorem*.

## 10.5 Uniqueness Theorems

The construction of Lebesgue measure was natural and intuitive, given that the goal was to have countable additivity, while retaining the elementary idea of the volume of a rectangular parallelepiped (at least, one with sides parallel to the coordinate planes—we have yet to see that Lebesgue measure is invariant under rotations). It is worthwhile to see that in fact there is only one Borel measure on  $R^n$  which does this. We will prove some much more general results, since the effort is about the same. We begin by developing a powerful technical tool.

**10.35 Definition.** A class  $\mathcal{P}$  of subsets of a set  $X$  is called a  $\pi$ -system if it is closed under finite intersections, i.e., if  $A \cap B \in \mathcal{P}$  whenever  $A, B \in \mathcal{P}$ . A class  $\mathcal{L}$  of subsets of a set  $X$  is called a  $\lambda$ -system if it contains  $X$  and is closed under proper differences and increasing limits, i.e., if:

- (a)  $X \in \mathcal{L}$ ;
- (b) If  $A \subset B$ , with  $A$  and  $B$  in  $\mathcal{L}$ , then  $B \setminus A \in \mathcal{L}$ ; and
- (c) If  $A_n \in \mathcal{L}$  and  $A_n \subset A_{n+1}$  for every positive integer  $n$ , then  $\bigcup_1^{\infty} A_n \in \mathcal{L}$ .

Some examples of  $\pi$ -systems: all open intervals in  $R^n$ , or all bounded open intervals, or all bounded left-open and right-closed intervals with rational endpoints; or all intervals of the form  $\{x : -\infty < x_j \leq a_j\}$ , or just those with rational  $a_j$ . It is obvious that any  $\sigma$ -algebra is both a  $\pi$ -system and a  $\lambda$ -system, and the converse is nearly as obvious.

**10.36 Lemma.** *If  $\mathcal{A}$  is both a  $\pi$ -system and a  $\lambda$ -system, then  $\mathcal{A}$  is a  $\sigma$ -algebra.*

**Proof.** Using (a) and (b), we see that  $\mathcal{A}$  is closed under complementation, and hence under finite unions, so  $\mathcal{A}$  is an algebra. If  $(A_n)$  is a sequence in  $\mathcal{A}$ , then  $B_n = \bigcup_{k=1}^n A_k \in \mathcal{A}$  for each  $n$ , and  $B_n \subset B_{n+1}$ , so by (c) we have  $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$ . Thus  $\mathcal{A}$  is a  $\sigma$ -algebra. ■

The next result is the tool we want; it is known as *Dynkin's  $\pi$ - $\lambda$  theorem*.

**10.37 Theorem.** *If the  $\lambda$ -system  $\mathcal{L}$  contains the  $\pi$ -system  $\mathcal{P}$ , then it contains the  $\sigma$ -algebra generated by  $\mathcal{P}$ .*

**Proof.** Let  $\mathcal{A}$  be the intersection of all  $\lambda$ -systems containing  $\mathcal{P}$ . It is easy to see that  $\mathcal{A}$  is a  $\lambda$ -system, with  $\mathcal{P} \subset \mathcal{A} \subset \mathcal{L}$ . I claim that  $\mathcal{A}$  is a  $\pi$ -system. For each  $A \in \mathcal{A}$ , let  $\mathcal{G}_A = \{B \in \mathcal{A} : A \cap B \in \mathcal{A}\}$ ; it is easy to see that  $\mathcal{G}_A$  is a  $\lambda$ -system, for any  $A \in \mathcal{A}$ . It is also clear that if  $A \in \mathcal{P}$ , then  $\mathcal{G}_A \supset \mathcal{P}$ . Thus  $\mathcal{G}_A = \mathcal{A}$  whenever  $A \in \mathcal{P}$ , since  $\mathcal{A}$  is the minimal  $\lambda$ -system containing  $\mathcal{P}$ . But this says that  $A \cap B \in \mathcal{A}$  whenever  $A \in \mathcal{P}$  and  $B \in \mathcal{A}$ , i.e., that  $A \in \mathcal{G}_B$  whenever  $A \in \mathcal{P}$  and  $B \in \mathcal{A}$ . Thus  $\mathcal{P} \subset \mathcal{G}_B$  for all  $B \in \mathcal{A}$ , so again it follows that  $\mathcal{G}_B = \mathcal{A}$  for every  $B \in \mathcal{A}$ ; but this is just the assertion that  $\mathcal{A}$  is closed under intersections, i.e., is a  $\pi$ -system. It follows from Lemma 10.36 that  $\mathcal{A}$  is a  $\sigma$ -algebra. ■

**10.38 Theorem.** *Let  $\mathcal{P}$  be a  $\pi$ -system, and let  $\mathcal{A}$  be the  $\sigma$ -algebra generated by  $\mathcal{P}$ . If  $\mu$  and  $\nu$  are finite measures with domain  $\mathcal{A}$  such that  $\mu(X) = \nu(X)$  and  $\mu(A) = \nu(A)$  for every  $A \in \mathcal{P}$ , then  $\mu = \nu$ .*

**Proof.** Let  $\mathcal{L} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$ ; the claim is that  $\mathcal{L} = \mathcal{A}$ , and by Theorem 10.37, it suffices to show that  $\mathcal{L}$  is a  $\lambda$ -system. Now if  $A$  and  $B$  are in  $\mathcal{L}$ , and  $A \subset B$ , then  $\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$ , since  $\mu$  and  $\nu$  are finite. Thus  $B \setminus A \in \mathcal{L}$ . If  $A = \bigcup_{n=1}^{\infty} A_n$ , where  $A_n \in \mathcal{L}$  and  $A_n \subset A_{n+1}$  for every  $n$ , then

$$\mu(A) = \lim \mu(A_n) = \lim \nu(A_n) = \nu(A)$$

(using Proposition 9.11), so  $A \in \mathcal{L}$ . Since we were given  $X \in \mathcal{L}$ , it follows that  $\mathcal{L}$  is a  $\lambda$ -system. ■



The hypothesis in the last theorem that  $\mu$  and  $\nu$  are finite cannot be omitted. [Consider, for instance,  $\mu$  counting measure on  $\mathbf{R}$ , and  $\nu$  defined by  $\nu(A) = \sum_{q \in \mathbf{Q}} 1_A(q)$  (in other words,  $\nu$  is counting measure on the rationals  $\mathbf{Q}$ , regarded as a measure on  $\mathbf{R}$ ). Then  $\mu(A) = \nu(A) = \infty$  on all open intervals, a  $\pi$ -system which generates the Borel sets, but certainly  $\mu(A)$  does not equal  $\nu(A)$  for every Borel set.] However, it can be weakened enough to cover the cases we are most interested in.

**10.39 Corollary.** *Let  $\mathcal{P}$  be a  $\pi$ -system,  $\mathcal{A}$  the  $\sigma$ -algebra generated by  $\mathcal{P}$ , and  $\mu$  a measure with domain  $\mathcal{A}$ . Suppose that  $X = \bigcup_{j=1}^{\infty} K_j$ , where  $K_j \in \mathcal{P}$ ,  $K_j \subset K_{j+1}$ , and  $\mu(K_j) < +\infty$  for each  $j$ . If  $\nu$  is a measure with domain  $\mathcal{A}$  such that  $\nu(A) = \mu(A)$  for all  $A \in \mathcal{P}$ , then  $\mu = \nu$ .*

**Proof.** Define the measures  $\mu_j$  and  $\nu_j$  by  $\mu_j(A) = \mu(A \cap K_j)$ ,  $\nu_j(A) = \nu(A \cap K_j)$  for  $A \in \mathcal{A}$ . Then Theorem 10.38 gives  $\mu_j = \nu_j$  for each  $j$ , and Proposition 9.11 tells us that

$$\mu(A) = \lim_{j \rightarrow \infty} \mu_j(A) = \lim_{j \rightarrow \infty} \nu_j(A) = \nu(A)$$

for every  $A \in \mathcal{A}$ . ■

We single out the special case of Corollary 10.39 which interests us the most.

**10.40 Corollary.** *If  $\mu$  and  $\nu$  are Borel measures on  $\mathbf{R}^n$  such that  $\mu(I) = \nu(I) < \infty$  for every bounded interval  $I$ , or just for every bounded interval with rational sides, then  $\mu = \nu$ .*

We are now in a position to quickly prove that (the restriction to Borel sets of) Lebesgue measure is essentially the only Borel measure on  $\mathbf{R}^n$  which is translation invariant, and finite on bounded sets.

**10.41 Theorem.** *If  $\mu$  is a Borel measure on  $\mathbf{R}^n$  which is translation invariant and finite on bounded intervals, then there is a constant  $C$  such that  $\mu(A) = C m(A)$  for every Borel set  $A$ .*

**Proof.** Let  $I = [0, 1]^n = \{(x_1, \dots, x_n) : 0 \leq x_j < 1, 1 \leq j \leq n\}$  and let  $C = \mu(I)$ . Let  $k_1, \dots, k_n$  be positive integers, and let  $J = \{0 \leq x_j < 1/k_j, 1 \leq j \leq n\}$ . Then  $I$  is the union of  $(k_1 k_2 \cdots k_n)$  translates of  $J$ , which are pairwise disjoint, so  $\mu(I) = k_1 \cdots k_n \mu(J)$ , i.e.,  $\mu(J) = C m(J)$ . Now any interval  $[a, b]$  with rational sides, i.e., with  $b_j - a_j$  rational for each  $j$ ,  $1 \leq j \leq n$ , is the disjoint union of translates of such a  $J$ , so the measures  $\mu$  and  $Cm$  agree on all such intervals. By Corollary 10.40,  $\mu$  and  $Cm$  agree on all Borel sets. ■

## 10.6 Linear Transformations

Theorem 10.41 allows us to obtain a change-of-variables formula.

**10.42 Theorem.** *Let  $T : \mathbf{R}^n \rightarrow \mathbf{R}^n$  be a nonsingular linear transformation. Then  $m(T(A)) = |\det T| m(A)$  for every Borel set  $A$ .*

**Proof.** Let  $\mu_T(A) = m(T(A))$  for every Borel set  $A$ . (We note that  $T(A)$  is Borel whenever  $A$  is Borel, since  $T^{-1}$  is continuous, so this definition makes sense.) It is clear that  $\mu_T$  is a measure, and since

$$\mu_T(A + \mathbf{x}) = m(T(A + \mathbf{x})) = m(T(A) + T\mathbf{x}) = m(T(A)) = \mu_T(A),$$

we see that  $\mu_T$  is translation invariant. By Theorem 10.41, we conclude that there exists  $C_T$  such that  $\mu_T = C_T m$ . Now it is clear that if  $S$  and  $T$  are two such transformations, then  $C_{ST} = C_S C_T$ . If  $O$  is an orthogonal transformation, i.e., if  $O^t O = I$ , or equivalently,  $|O\mathbf{x}| = |\mathbf{x}|$  for all  $\mathbf{x} \in \mathbf{R}^n$ , then  $O(B) = B$ , where  $B = \{\mathbf{x} \in \mathbf{R}^n : |\mathbf{x}| \leq 1\}$ . It follows that  $\mu_O(B) = m(B)$ , and since  $0 < m(B) < \infty$ , we conclude  $C_O = 1$ . Recall that  $\det O = \pm 1$ , since  $1 = \det(O^t O) = \det O^t \det O = (\det O)^2$ . If  $D$  is a transformation represented by a diagonal matrix, with nonnegative entries  $d_j$ , and  $I = \{\mathbf{x} : 0 \leq x_j \leq 1, 1 \leq j \leq n\}$ , then  $D(I) = \{\mathbf{x} : 0 \leq x_j \leq d_j, 1 \leq j \leq n\}$ . It is then clear again that  $C_D = d_1 d_2 \cdots d_n = \det D$ . We now apply one of the more interesting theorems in linear algebra: any linear transformation  $T$  on  $\mathbf{R}^n$  can be expressed in the form  $T = O_1 D O_2$ , where  $O_1$  and  $O_2$  are orthogonal, and  $D$  is diagonal with nonnegative entries. We have  $C_T = C_{O_1} C_D C_{O_2} = \det D = |\det T|$ , as claimed. ■

Another proof of this theorem would express  $T$  as the product of elementary matrices, permutation matrices, and a diagonal matrix, corresponding to the usual way in which one computes determinants or solves systems of linear equations by elementary row operations. This is a more direct and elementary proof, but I wanted to call your attention to the factorization theorem used above.

**10.43 Corollary.** *If  $S$  is a subspace of  $\mathbf{R}^n$ , with  $\dim S < n$ , then  $m(S) = 0$ .*

**Proof.** We can find a nonsingular transformation which maps  $S$  into the hyperplane  $\{\mathbf{x} : x_1 = 0\}$ , which obviously has measure zero. The corollary now follows from Theorem 10.42. ■

**10.44 Corollary.** *If  $T$  is any linear transformation of  $\mathbf{R}^n$  into itself, and  $A$  is a Lebesgue measurable set in  $\mathbf{R}^n$ , then  $T(A)$  is Lebesgue measurable, and  $m(T(A)) = |\det T| m(A)$ .*

**Proof.** If  $T$  is singular, this follows from the last corollary. Assume  $T$  nonsingular; by Theorem 9.29 we can express  $A$  as the union of a Borel set and a set of measure zero; the set of measure zero is a subset of a Borel set of measure zero, and the corollary follows from Theorem 10.42. ■

We remark that if  $A$  is a Borel set in  $\mathbf{R}^n$ , and  $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$  is linear, it does not follow that  $T(A)$  is a Borel set.

Theorem 10.42 has an equivalent formulation in terms of integrals, rather than measures.

**10.45 Theorem.** *If  $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$  is an invertible linear transformation, and  $f$  is a measurable function on  $\mathbf{R}^n$ , with either  $f \geq 0$  or  $f$  summable, then*

$$\int f \, dm = |\det T| \int f \circ T \, dm.$$

**Proof.** If  $f = 1_A$ , where  $A$  is a measurable set in  $\mathbf{R}^n$ , then  $f \circ T = 1_{T^{-1}(A)}$ , so

$$\int f \, dm = m(A) = |\det T| m(T^{-1}(A)) = |\det T| \int f \circ T \, dm,$$

and the theorem is true for such  $f$ . Then it holds, by the linearity of integrals, for any simple measurable  $f$ , and then by the monotone convergence theorem, for any nonnegative measurable  $f$ , and finally, by consideration of  $f = f^+ - f^-$ , for every summable  $f$ . ■

## 10.7 Smooth Transformations

The next theorem deals with a more general change of variable result than Theorem 10.42 or Theorem 10.45.

**10.46 Theorem.** *Let  $U$  and  $V$  be open subsets of  $\mathbf{R}^n$ , and suppose  $\varphi$  is a continuously differentiable bijective map of  $U$  onto  $V$ , such that  $\varphi^{-1}$  is also continuously differentiable. Then, for any nonnegative measurable function  $f$  on  $V$ , we have*

$$\int_V f \, dm = \int_U (f \circ \varphi) |J_\varphi| \, dm, \quad (10.2)$$

where  $J_\varphi = \det \varphi'$  is the Jacobian determinant of  $\varphi$ , and, in particular,

$$m(\varphi(A)) = \int_A |J_\varphi| \, dm \quad (10.3)$$

for every measurable set  $A \subset U$ .

**Proof.** We observe that equation (10.3) is just the special case of equation (10.2) obtained by taking  $f = 1_{\varphi(A)}$ . We also observe that it suffices to consider Borel measurable functions, since every Lebesgue measurable function is almost everywhere equal to a Borel measurable function (Theorem 10.9).

We carry out the proof with a short sequence of lemmas. Let us introduce some terminology and notation. A *cube with center  $\xi$*  is a set of the form

$$Q = \{\mathbf{x} \in \mathbf{R}^n : \xi_j - h \leq x_j < \xi_j + h, 1 \leq j \leq n\}.$$

(This describes a “half-open” cube, which will be convenient for our purposes. Closed and open cubes are defined similarly.) If  $Q$  is a cube with center  $\xi$ , and  $\epsilon > 0$ , let  $Q^\epsilon$  denote the concentric cube with sides multiplied by  $1 + \epsilon$ ; thus, for the cube  $Q$  above,

$$Q' = \{\mathbf{x} \in \mathbf{R}^n : \xi_j - (1 + \epsilon)h \leq x_j < \xi_j + (1 + \epsilon)h, 1 \leq j \leq n\}.$$

It is clear that  $m(Q^\epsilon) = (1 + \epsilon)^n m(Q)$  for any cube  $Q$ .

**10.47 Lemma.** For each  $\xi \in U$ , let  $T_\xi$  be the affine map approximating  $\varphi$  near  $\xi$ , i.e.,  $T_\xi(\mathbf{x}) = \varphi(\xi) + \varphi'(\xi)(\mathbf{x} - \xi)$ . For any compact subset  $K$  of  $U$ , and any  $\epsilon > 0$ , there exists  $\delta = \delta(K, \epsilon) > 0$  such that, for any cube  $Q \subset U$  with center  $\xi$  in  $K$  and diameter  $< \delta$ , we have

$$\varphi(Q) \subset T_\xi(Q^\epsilon). \quad (10.4)$$

**Proof.** Let  $M = \sup_{\xi \in K} \|(\varphi')^{-1}(\xi)\|$ ; since  $K$  is compact and  $\varphi'$  is continuous, we know  $M < \infty$ . The definition of the derivative of a mapping tells us that for any  $\eta > 0$  there exists  $\delta = \delta(\xi, \eta) > 0$  such that  $|\varphi(\mathbf{x}) - T_\xi(\mathbf{x})| < \eta|\mathbf{x} - \xi|$  whenever  $0 < |\mathbf{x} - \xi| < \delta$ . Since  $K$  is compact, and  $\varphi'$  is continuous, we can choose  $\delta = \delta(\eta)$  so that this holds for every  $\xi \in K$ . It follows, since  $T_\xi^{-1}\mathbf{y} - T_\xi^{-1}\mathbf{z} = [\varphi'(\xi)]^{-1}(\mathbf{y} - \mathbf{z})$ , that

$$|T_\xi^{-1}\varphi(\mathbf{x}) - \mathbf{x}| = |T_\xi^{-1}[\varphi(\mathbf{x}) - T_\xi(\mathbf{x})]| \leq M\eta|\mathbf{x} - \xi|,$$

so that, if  $\eta > 0$  is chosen sufficiently small ( $\eta < \epsilon/(M\sqrt{n})$  will do), we have  $T_\xi^{-1}\varphi(\mathbf{x}) \in Q^\epsilon$ , or equivalently,  $\varphi(\mathbf{x}) \in T_\xi(Q^\epsilon)$ , whenever  $\mathbf{x} \in Q$  with  $Q$  a cube of diameter less than  $\delta$  and center  $\xi \in K$ . ■

**10.48 Lemma.** If  $B$  is a Borel subset of  $U$ , then

$$m(\varphi(B)) \leq \int_B |J_\varphi| dm. \quad (10.5)$$

**Proof.** Let  $C$  be a compact subset of  $\varphi(B)$ , and put  $K = \varphi^{-1}(C)$ , so  $K$  is a compact subset of  $U$ . Let  $\epsilon > 0$ . For each  $i \in \mathbf{N}$ , let  $K_i = \{\mathbf{x} : \rho(\mathbf{x}, K) \leq 1/i\}$ , where  $\rho(\mathbf{x}, K) = \inf\{|\mathbf{x} - \mathbf{y}| : \mathbf{y} \in K\}$ . Then  $K_i$  is compact, and for sufficiently large  $i$ , say  $i \geq i_0$ ,  $K_i \subset U$ . We note that  $K_{i+1} \subset K_i$  for all  $i$ , and  $\bigcap_i K_i = K$ . Since the function  $A \mapsto \int_A |J_\varphi| dm$  is a measure, and  $\int_{K_i} |J_\varphi| dm < \infty$  for  $i \geq i_0$ , it follows that  $\int_K |J_\varphi| dm = \lim \int_{K_i} |J_\varphi| dm$ .

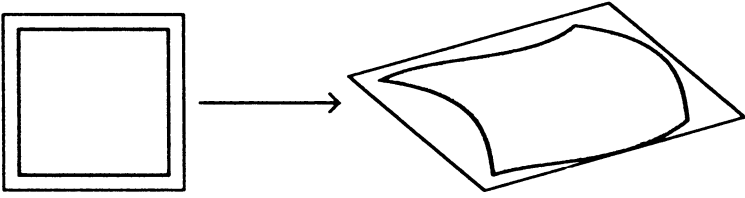


Figure 10.1. A picture for Lemma 10.47.

Hence there exists  $i \geq i_0$  such that  $\int_{K_i} |J_\varphi| dm < \int_K |J_\varphi| dm + \epsilon$ . Since  $K_i$  is compact, the continuous function  $J_\varphi$  is uniformly continuous on  $K_i$ , so there exists  $\delta > 0$  such that  $|J_\varphi(\mathbf{x}) - J_\varphi(\mathbf{y})| < \epsilon$  for all  $\mathbf{x}, \mathbf{y} \in K_i$  with  $|\mathbf{x} - \mathbf{y}| < \delta$ . We may assume that  $\delta \leq \delta(K_i, \epsilon)$  of Lemma 10.47; we may also assume that  $\delta < 1/i$ , so that every cube  $Q$  of diameter less than  $\delta$  which meets  $K$  is entirely contained in  $K_i$ . Since  $K$  is bounded, we can find a finite disjoint sequence of cubes  $Q_1, \dots, Q_N$  such that: (i)  $K \subset \bigcup_{j=1}^N Q_j$ ; (ii)  $Q_j \cap K \neq \emptyset$  for each  $j$ ; and (iii)  $\text{diam } Q_j < \delta$  for each  $j$ . Then we have  $C = \varphi(K) \subset \varphi(\bigcup_{j=1}^N Q_j)$ , a disjoint union, so (denoting the center of  $Q_j$  by  $\xi_j$ )

$$\begin{aligned}
 m(C) &\leq m\left(\varphi\left(\bigcup_{j=1}^N Q_j\right)\right) = \sum_{j=1}^N m(\varphi(Q_j)) \\
 &\leq \sum_{j=1}^N m(T_{\xi_j}(Q_j^\epsilon)) \quad (\text{by Lemma 10.47}) \\
 &= \sum_{j=1}^N |\det \varphi'(\xi_j)| m(Q_j^\epsilon) \quad (\text{by Theorem 10.42}) \\
 &= \sum_{j=1}^N (1 + \epsilon)^n |J_\varphi(\xi_j)| m(Q_j) \\
 &\leq (1 + \epsilon)^n \sum_{j=1}^N \int_{Q_j} (|J_\varphi| + \epsilon) dm \\
 &\leq (1 + \epsilon)^n \int_{K_i} (|J_\varphi| + \epsilon) dm.
 \end{aligned}$$

Since  $\int_{K_i} |J_\varphi| dm < \int_K |J_\varphi| dm + \epsilon$ , and  $\epsilon > 0$  was arbitrary, we conclude that

$$m(C) \leq \int_K |J_\varphi| dm \leq \int_B |J_\varphi| dm.$$

Since  $C$  was an arbitrary compact subset of  $\varphi(B)$ , it follows from Theorem 9.29 that  $m(\varphi(B)) \leq \int_B |J_\varphi| dm$ . ■

**10.49 Lemma.** *For every nonnegative Borel measurable function  $f$  on  $V$ , we have*

$$\int_V f \, dm \leq \int_U (f \circ \varphi) |J_\varphi| \, dm. \quad (10.6)$$

**Proof.** When  $f = \mathbf{1}_A$ , where  $A$  is a Borel subset of  $V$ , this is exactly the statement of the last lemma, with  $B = \varphi^{-1}(A)$ . Hence (10.6) holds whenever  $f$  is a nonnegative simple Borel measurable function, and taking monotone limits, it holds for every nonnegative Borel measurable  $f$ . ■

We conclude the proof of Theorem 10.46 by recasting the last lemma, with  $\varphi^{-1}$  now playing the role of  $\varphi$ , and  $(f \circ \varphi)|J_\varphi|$  in the role of  $f$ . For any differentiable maps  $\varphi$  and  $\psi$ , the chain rule tells us that  $(\varphi \circ \psi)' = (\varphi' \circ \psi)\psi'$ , and taking determinants, it follows that  $J_{\varphi \circ \psi} = (J_\varphi \circ \psi)J_\psi$ . Hence, for any nonnegative Borel measurable function  $f$  on  $V$ ,

$$\begin{aligned} \int_U (f \circ \varphi) |J_\varphi| \, dm &\leq \int_V (f \circ \varphi \circ \varphi^{-1}) (|J_\varphi| \circ \varphi^{-1}) |J_{\varphi^{-1}}| \, dm \\ &= \int_V f |J_{\varphi \circ \varphi^{-1}}| \, dm = \int_V f \, dm, \end{aligned}$$

so that we have, in fact, equality in (10.6), i.e., we have established (10.2), and the theorem is proved. ■

## 10.8 Multiple and Repeated Integrals

Finally, we turn to the very important topic of computing integrals over  $\mathbf{R}^n$  by repeated integrals over  $\mathbf{R}$ . We fix for now the positive integers  $k$  and  $l$ , and put  $n = k + l$ ; we identify  $\mathbf{R}^n$  with  $\mathbf{R}^k \times \mathbf{R}^l$ . If  $f$  is a function defined on  $\mathbf{R}^n$ , and  $y \in \mathbf{R}^l$ , we denote by  $f(\cdot, y)$  the function on  $\mathbf{R}^k$  such that  $x \mapsto f(x, y)$ ;  $f(x, \cdot)$  is defined in an analogous manner, for  $x \in \mathbf{R}^k$ . (We are dropping the convention of boldface letters for points of  $\mathbf{R}^n$  in this section.)

**10.50 Lemma.** *If  $f$  is a Borel measurable function on  $\mathbf{R}^n$ , then for every  $y \in \mathbf{R}^l$  the function  $f(\cdot, y)$  is Borel measurable on  $\mathbf{R}^k$ , and for every  $x \in \mathbf{R}^k$ , the function  $f(x, \cdot)$  is Borel measurable on  $\mathbf{R}^l$ .*

**Proof.** In view of Corollary 10.7, it suffices to prove the lemma when  $f$  is a simple Borel measurable function, and hence it suffices to prove the lemma when  $f = \mathbf{1}_A$ , where  $A$  is a Borel set in  $\mathbf{R}^n$ . Let  $\mathcal{G}$  be the collection of all Borel sets  $A$  in  $\mathbf{R}^n$  such that the lemma holds for  $\mathbf{1}_A$ . If  $A$  has the form  $A = B \times C$ , where  $B$  is a Borel set in  $\mathbf{R}^k$  and  $C$  is a Borel set in  $\mathbf{R}^l$ , then  $\mathbf{1}_A(x, y) = \mathbf{1}_B(x)\mathbf{1}_C(y)$  for any  $x \in \mathbf{R}^k$ ,  $y \in \mathbf{R}^l$ , and it is

obvious that  $A \in \mathcal{G}$ . If  $A_j \in \mathcal{G}$  and  $A_j \subset A_{j+1}$  for  $j = 1, 2, \dots$ , then for  $A = \bigcup_j A_j$ , we have  $\mathbf{1}_A = \lim \mathbf{1}_{A_j}$ ,  $\mathbf{1}_A(x, \cdot) = \lim \mathbf{1}_{A_j}(x, \cdot)$ , etc., and so clearly  $A \in \mathcal{G}$ . If  $A \in \mathcal{G}$ ,  $B \in \mathcal{G}$ , and  $A \subset B$ , then it follows that  $B \setminus A \in \mathcal{G}$ , since  $\mathbf{1}_{B \setminus A} = \mathbf{1}_B - \mathbf{1}_A$ . Obviously,  $\mathbf{R}^n \in \mathcal{G}$ , so  $\mathcal{G}$  is a  $\lambda$ -system, containing all intervals, and thus by Theorem 10.37,  $\mathcal{G}$  consists of all Borel sets, as was to be proved. ■

**10.51 Theorem.** *If  $f$  is a nonnegative Borel measurable function on  $\mathbf{R}^n$ , then the functions  $F$  and  $G$  defined by*

$$F(x) = \int_{\mathbf{R}^l} f(x, \cdot) dm, \quad G(y) = \int_{\mathbf{R}^k} f(\cdot, y) dm$$

*are Borel measurable, on  $\mathbf{R}^k$  and  $\mathbf{R}^l$ , respectively, and we have*

$$\int_{\mathbf{R}^n} f dm = \int_{\mathbf{R}^k} F dm = \int_{\mathbf{R}^l} G dm.$$

**Proof.** Using Lemma 10.6 and Theorem 10.15, it suffices to prove the theorem when  $f$  is simple, and hence it suffices to prove it for the special case  $f = \mathbf{1}_A$ , where  $A$  is a Borel set in  $\mathbf{R}^n$ . Let  $\mathcal{G}$  again denote the collection of all “good” sets, i.e., all Borel  $A$  such that the theorem holds for  $f = \mathbf{1}_A$ . If  $A = B \times C$ , where  $B$  and  $C$  are Borel sets in  $\mathbf{R}^k$  and  $\mathbf{R}^l$ , respectively, then  $f(x, \cdot) = \mathbf{1}_B(x) \mathbf{1}_C$ , so  $F = m(C) \mathbf{1}_B$ ; similarly,  $G = m(B) \mathbf{1}_C$ , so

$$\int_{\mathbf{R}^k} F dm = \int_{\mathbf{R}^l} G dm = m(B)m(C) = \int_{\mathbf{R}^n} f dm,$$

so  $A \in \mathcal{G}$ . Thus, in particular, every interval in  $\mathbf{R}^n$  belongs to the class  $\mathcal{G}$ . If  $A_j \in \mathcal{G}$  and  $A_j \subset A_{j+1}$  for  $j = 1, 2, \dots$ , then  $\bigcup_j A_j \in \mathcal{G}$ , as is easily seen from the monotone convergence theorem (Theorem 10.15). Let  $K_j = \{x \in \mathbf{R}^n : |x_i| \leq j, 1 \leq i \leq n\}$ . For each positive integer  $j$ , let  $\mathcal{G}_j$  denote the set of all Borel sets  $A$  such that  $A \cap K_j \in \mathcal{G}$ . If  $A \in \mathcal{G}_j$  for every  $j$ , then  $A \in \mathcal{G}$ , since  $A$  is the union of the increasing sequence  $(A \cap K_j)$ . But it is easy to see that each  $\mathcal{G}_j$  is a  $\lambda$ -system; for if  $A$  and  $B$  are in  $\mathcal{G}_j$ , with  $A \subset B$ , then  $\mathbf{1}_{B \setminus A} = \mathbf{1}_B - \mathbf{1}_A$ , and the integral of  $\mathbf{1}_A$  over any  $K_j$  is finite, as are the corresponding integrals over the projections of  $K_j$  onto  $\mathbf{R}^k$  and  $\mathbf{R}^l$ , so we can subtract (this subtraction was not available to us in  $\mathcal{G}$ ; that’s why we “localized”). Thus each  $\mathcal{G}_j$  is a  $\lambda$ -system, containing the  $\pi$ -system of all intervals, and hence  $\mathcal{G}_j$  is the algebra of all Borel sets. It follows that  $\mathcal{G}$  contains all Borel sets. ■

The more familiar way to write the conclusion of Theorem 10.51 is by using “dummy variables”:

$$\int_{\mathbf{R}^k} \int_{\mathbf{R}^l} f(x, y) dm(y) dm(x) = \int_{\mathbf{R}^l} \int_{\mathbf{R}^k} f(x, y) dm(x) dm(y),$$

both sides being equal to  $\int_{\mathbf{R}^n} f(x, y) \, d\mathbf{m}(x, y)$ . It is important to realize that the conclusion of Theorem 10.51 does not require that the integrals in question are finite. However, without the hypothesis that  $f$  is nonnegative, we do require the integrability of  $f$ .

**10.52 Theorem.** *Let  $f$  be a Borel measurable function on  $\mathbf{R}^n$  and suppose that  $f$  is summable, or, what is the same according to Theorem 10.51, that either*

$$\int_{\mathbf{R}^k} \int_{\mathbf{R}^l} |f(x, y)| \, d\mathbf{m}(y) \, d\mathbf{m}(x) < +\infty$$

or

$$\int_{\mathbf{R}^l} \int_{\mathbf{R}^k} |f(x, y)| \, d\mathbf{m}(x) \, d\mathbf{m}(y) < +\infty.$$

*Then  $f(x, \cdot)$  is integrable for almost all  $x \in \mathbf{R}^k$ ,  $f(\cdot, y)$  is integrable for almost all  $y \in \mathbf{R}^l$ , and*

$$\begin{aligned} \int_{\mathbf{R}^k} \int_{\mathbf{R}^l} f(x, y) \, d\mathbf{m}(y) \, d\mathbf{m}(x) &= \int_{\mathbf{R}^n} f \, d\mathbf{m} \\ &= \int_{\mathbf{R}^l} \int_{\mathbf{R}^k} f(x, y) \, d\mathbf{m}(x) \, d\mathbf{m}(y). \end{aligned}$$

**Proof.** It suffices to apply Theorem 10.51 to the functions  $f^+$  and  $f^-$ . ■

Theorems 10.51 and 10.52 are usually referred to as Fubini's theorem, but we will call them theorems of Fubini-Tonelli (see the notes at the end of this chapter.) These theorems were stated for Borel functions; naturally, we want to use these theorems when the given function  $f$  is only known to be Lebesgue measurable. The correct theorem then becomes somewhat clumsier to state, since it is no longer true that  $f(x, \cdot)$  is measurable on  $\mathbf{R}^l$  for every  $x \in \mathbf{R}^k$ , but only almost every  $x$ , etc. The validity of the interchange of order of integration in this case follows from Theorems 10.51 and 10.52 and the fact (Theorem 10.9) that any Lebesgue measurable  $f$  is equal almost everywhere to a Borel measurable function  $g$ . We omit the details.

The discussion above can be transferred to the following setting: let  $X$  and  $Y$  be sets, equipped with  $\sigma$ -algebras  $\mathcal{A}$  and  $\mathcal{B}$ , respectively, and let  $\mu$  and  $\nu$  be measures with domains  $\mathcal{A}$ ,  $\mathcal{B}$ . We let  $\mathcal{A} \times \mathcal{B}$  denote the  $\sigma$ -algebra of subsets of  $X \times Y$  generated by  $\{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$ . Assume  $\mu$  and  $\nu$  are  $\sigma$ -finite, i.e., that  $X = \bigcup_{j=1}^{\infty} X_j$ , with  $X_j \in \mathcal{A}$  and  $\mu(X_j) < \infty$  for each  $j$ , and a similar statement for  $Y$ . Then there exists a (unique) measure  $\lambda = \mu \times \nu$  on  $X \times Y$ , with domain  $\mathcal{A} \times \mathcal{B}$ , such that  $\lambda(A \times B) = \mu(A)\nu(B)$  for every  $A \in \mathcal{A}$ ,  $B \in \mathcal{B}$ , and

$$\int f \, d\lambda = \int \left( \int f(x, y) \, d\mu(x) \right) d\nu(y) = \int \left( \int f(x, y) \, d\nu(y) \right) d\mu(x)$$



for any  $\mathcal{A} \times \mathcal{B}$ -measurable  $f$  which is either nonnegative or  $\lambda$ -summable. The  $\sigma$ -finiteness condition is necessary here; if  $X = Y = [0, 1]$ ,  $\mu$  is Lebesgue-Borel measure and  $\nu$  is counting measure, then for  $f = 1_D$ , where  $D = \{(x, y) : x = y\}$ , we have  $\int f(x, y) d\nu(y) = 1$  for every  $x$ , so  $\int \int f(x, y) d\nu(y) d\mu(x) = 1$ , but  $\int f(x, y) d\mu(x) = 0$  for all  $y$ , so  $\int \int f(x, y) d\mu(x) d\nu(y) = 0$ .

The Fubini Tonelli theorem is one of the most frequently used tools in analysis. Here is one application, a generalization of the integration-by-parts formula (Theorem 5.32). We will write  $\int_a^b f(x) dx$  for  $\int_{[a, b]} f dm$  when  $m$  is Lebesgue measure on  $\mathbf{R}$ , and denote Lebesgue measure on  $\mathbf{R}^2$  by  $m_2$ .

**10.53 Theorem.** *Let  $f$  and  $g$  be summable over the interval  $[a, b]$ , and suppose  $F(x) = F(a) + \int_a^x f(y) dy$ , and  $G(x) = G(a) + \int_a^x g(y) dy$ . Then*

$$\int_a^b F(x)g(x) dx = F(b)G(b) - F(a)G(a) - \int_a^b G(x)f(x) dx.$$

**Proof.** Let  $T = \{(x, y) : a \leq y \leq x \leq b\}$ . Then

$$\begin{aligned} \int_a^b F(x)g(x) dx &= \int_a^b \left( F(a) + \int_a^x f(y) dy \right) g(x) dx \\ &= \int \mathbf{1}_T(x, y) f(y) g(x) dm_2(x, y) + F(a) \int_a^b g(x) dx \\ &= \int_a^b \int_y^b g(x) dx f(y) dy + F(a)[G(b) - G(a)] \\ &= \int_a^b [G(b) - G(y)] f(y) dy + F(a)[G(b) - G(a)] \\ &= F(b)G(b) - F(a)G(a) - \int_a^b G(x)f(x) dx, \end{aligned}$$

as claimed. ■

## 10.9 Exercises

In the following exercises,  $\mathcal{A}$  is a  $\sigma$ -algebra of subsets of a set  $X$ , and  $\mu$  is a measure with domain  $\mathcal{A}$ . Measurability will refer to measurability with respect to  $\mathcal{A}$ , unless otherwise indicated.

1. Show that if  $f : X \rightarrow \mathbf{R}$  is measurable, and  $g : \mathbf{R} \rightarrow \mathbf{R}$  is continuous, then  $g \circ f$  is measurable.

2. Let  $C$  be the Cantor set, so  $x \in C$  if and only if there exist  $a_n \in \{0, 2\}$  such that  $x = \sum_{n=1}^{\infty} a_n 3^{-n}$ . Define  $f_0 : C \rightarrow [0, 1]$  by

$$f_0\left(\sum_{n=1}^{\infty} \frac{a_n}{3^n}\right) = \sum_{n=1}^{\infty} \frac{a_n}{2^{n+1}}.$$

Show that  $f_0$  is surjective. Show that  $f_0$  can be extended to a continuous nondecreasing map  $f$  of  $[0, 1]$  to itself. (The function  $f$  is known as the Cantor function.)

3. Use the example of Exercise 2 to show that if  $f$  is continuous on  $[0, 1]$  and  $A$  is a Lebesgue measurable subset of  $[0, 1]$ , it need not follow that  $f(A)$  is Lebesgue measurable.

4. Let  $F(x) = \frac{1}{2}(x + f(x))$ , where  $f$  still denotes the Cantor function.

- Show that  $F$  is a strictly increasing continuous function mapping  $[0, 1]$  onto itself, so that  $G = F^{-1}$  is a continuous map of  $[0, 1]$  onto itself.
- Show that  $B = G^{-1}(C)$  is a Borel set, and  $m(B) = \frac{1}{2}$ .
- Show that there exists a Lebesgue measurable set  $E \subset [0, 1]$  such that  $G^{-1}(E)$  is not measurable. This gives an example of a Lebesgue measurable set which is not a Borel set. Show that there exists a Lebesgue measurable function  $g$  such that  $g \circ G$  is not Lebesgue measurable. (Compare Exercise 1.)

5. Use Fatou's Lemma to do Exercise 4 of the last chapter. Use the monotone convergence theorem to do Exercise 5 of the last chapter.

6. Let  $f$  be a nonnegative measurable function, and let  $\nu(A) = \int_A f d\mu$  for each  $A \in \mathcal{A}$ . Suppose  $f$  is summable, i.e., that  $\nu(X) < \infty$ . Show that  $\nu$  is continuous with respect to  $\mu$  in the following sense: for each  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $\nu(A) < \epsilon$  for every  $A \in \mathcal{A}$  with  $\mu(A) < \delta$ . HINT: If not, there is some  $\epsilon > 0$  such that for every  $n$  there exists  $A_n \in \mathcal{A}$  with  $\mu(A_n) < 2^{-n}$  and  $\nu(A_n) \geq \epsilon$ . Let  $A = \limsup A_n$  (see the exercises in the last chapter). Show that  $\nu(A) \geq \epsilon$  and  $\mu(A) = 0$ , which is impossible.

7. Suppose  $\mu(X) < \infty$ . Let  $f$  be a measurable function, and define  $F : (0, \infty) \rightarrow [0, \infty]$  by  $F(t) = \mu(\{x : |f(x)| > t\})$ . ( $F$  is sometimes called the *distribution function* of  $f$ .)

- Show that if  $\int |f|^p d\mu < \infty$  for some  $p > 0$ , then  $F(t) \leq Ct^{-p}$  for some constant  $C$ , and all  $t > 0$ . (The assumption  $\mu(X) < \infty$  is not necessary for this part.)
- Show that  $\int |f| d\mu < \infty$  if and only if  $\sum_{n=1}^{\infty} F(n) < \infty$ .
- Show that if there exists a constant  $C$  such that  $F(t) \leq Ct^{-p}$  for all  $t > 0$ , then  $\int |f|^r d\mu < \infty$  for all  $r$ ,  $0 < r < p$ .

8. With the notation of the last exercise, show that if  $f \in L(\mu)$ , then  $tF(t) \rightarrow 0$  as  $t \rightarrow +\infty$ .

9. Let  $\{f_n\}$  be a sequence of measurable functions. Show that if

$$\sum_{n=1}^{\infty} \int |f_n| d\mu < \infty,$$

then  $\sum_{n=1}^{\infty} f_n$  converges a.e. to a summable function, and

$$\int \left( \sum_{n=1}^{\infty} f_n \right) d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

In particular, if  $f_n \geq 0$  for every  $n$ , then  $\sum \int f_n d\mu < \infty$  implies  $\sum f_n < \infty$  a.e.

10. Let  $f$  be integrable with respect to Lebesgue measure on  $\mathbf{R}$ . Define  $F(x) = \int_{[0,x]} f dm$  for  $x \geq 0$ , and  $F(x) = -\int_{[x,0]} f dm$  for  $x < 0$ .

(a) Show that  $F$  is continuous on  $\mathbf{R}$ .

(b) Show that if  $F(x) = 0$  for all  $x$ , then  $f = 0$  a.e.

11. Let  $\{f_n\}$  be a sequence of measurable functions. We say that  $(f_n)$  *converges in measure* to  $f$  if for every  $\epsilon > 0$ ,

$$\mu(\{x : |f(x) - f_n(x)| > \epsilon\}) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Show that if  $\mu$  is a finite measure, and  $f_n \rightarrow f$  a.e., then  $f_n \rightarrow f$  in measure. Give an example of a sequence which converges in measure, but does not converge a.e.

12. Let  $\{f_n\}$  be a sequence of measurable functions, and suppose that for every  $\epsilon > 0$ ,

$$\sum_{n=1}^{\infty} \mu(\{x : |f_n(x)| > \epsilon\}) < \infty.$$

Show that  $f_n \rightarrow 0$  a.e.

13. If  $\mathbf{f} : X \rightarrow \mathbf{R}^n$ , we say that  $\mathbf{f}$  is measurable if  $\mathbf{f}^{-1}(U) \in \mathcal{A}$  for every open  $U \subset \mathbf{R}^n$ . Show that  $\mathbf{f} = (f_1, \dots, f_n)$  is measurable if and only if  $f_k$  is measurable for  $k = 1, 2, \dots, n$ . We say that  $\mathbf{f}$  is integrable if each component  $f_k$  is integrable, and define  $\int \mathbf{f} d\mu = (\int f_1 d\mu, \dots, \int f_n d\mu)$ . Show that if  $\mathbf{f}$  is integrable, then  $|\int \mathbf{f} d\mu| \leq \int |\mathbf{f}| d\mu$ .

14. Let  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  be integrable, and  $\epsilon > 0$ . Show that there exist disjoint compact sets  $K_1, \dots, K_r$  in  $\mathbf{R}^n$ , and real constants  $c_1, \dots, c_r$ , such that if we put  $h = \sum_{j=1}^r c_j \mathbf{1}_{K_j}$ , then we have  $\int |f - h| dm < \epsilon$ . Show that there exists a continuous function  $g$ , vanishing outside a bounded subset of  $\mathbf{R}^n$ , with  $\int |f - g| dm < \epsilon$ .

15. Use the result of the last exercise to show that if  $f$  is integrable on  $\mathbf{R}^n$ , then

$$\int |f(x+t) - f(x)| dx \rightarrow 0 \quad \text{as } t \rightarrow 0.$$

16. Show that

$$\lim_{n \rightarrow \infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx = 1.$$

17. Show that if  $f_n, f$  are integrable, and  $f_n \rightarrow f$  a.e., then  $\int |f_n - f| d\mu \rightarrow 0$  as  $n \rightarrow \infty$  if and only if  $\int |f_n| d\mu \rightarrow \int |f| d\mu$  as  $n \rightarrow \infty$ . HINT: Reread the proof of the dominated convergence theorem.

18. Suppose  $\mu(X) = 1$ , and that  $f$  and  $g$  are positive measurable functions such that  $fg \geq 1$  a.e. Show that  $\int f d\mu \cdot \int g d\mu \geq 1$ .

19. Suppose  $\mu(X) < \infty$ . For each bounded measurable  $f$  on  $X$ , show that the function  $p \mapsto \log \|f\|_p$  is convex on  $(0, \infty)$ .

20. Suppose  $\mu(X) = 1$ . Let  $f$  be integrable, and let  $\varphi(r) = \|f\|_r$ . Show that  $\varphi$  is increasing on  $(0, \infty)$ , and that  $\lim_{r \rightarrow 0} \varphi(r) = \exp\left(\int \log |f| d\mu\right)$  and  $\lim_{r \rightarrow \infty} \varphi(r) = \|f\|_\infty$ .

21. Show (by a symmetry argument) that

$$m(\{\mathbf{x} : 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1\}) = \frac{1}{n!}$$

and deduce (by using a linear transformation) that

$$m(\{\mathbf{x} \in \mathbf{R}^n : x_j \geq 0, \sum x_j \leq 1\}) = \frac{1}{n!}.$$

22. Let  $\mathbf{x}_j = (x_j^1, \dots, x_j^n) \in \mathbf{R}^n$ ,  $j = 0, \dots, n$ . Let  $K$  be the convex set spanned by  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , i.e.,

$$K = \left\{ \sum t_j \mathbf{x}_j : t_j \geq 0, \sum t_j = 1 \right\}.$$

Show that  $m(K)$  is the absolute value of

$$\frac{1}{n!} \det \begin{bmatrix} x_0^1 & x_1^1 & \cdots & x_n^1 \\ \vdots & \vdots & & \vdots \\ x_0^n & x_1^n & \cdots & x_n^n \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

23. Explain the formula  $m(A) = \iint r dr d\theta$  for area in polar coordinates, in terms of the contents of this chapter.

24. Use the Fubini-Tonelli theorem to show that the graph of a measurable function has measure zero.

25. Let  $f(x, y) = (x^2 - y^2)/(x^2 + y^2)^2$  for  $0 < x, y \leq 1$ . Compute and compare the iterated integrals

$$\int_0^1 \int_0^1 f(x, y) dx dy \quad \text{and} \quad \int_0^1 \int_0^1 f(x, y) dy dx.$$

Reconcile your result with the Fubini-Tonelli theorem.

26. Let  $f$  be a real-valued function on  $\mathbf{R}^2$ , with the properties that  $f(\cdot, y)$  is Borel measurable for each  $y$ , and  $f(x, \cdot)$  is continuous for each  $x$ . Show that  $f$  is Borel measurable.

27. Let  $f$  be a measurable function, and  $E$  a measurable subset of  $\mathbf{R}^n$ . Use the Fubini-Tonelli theorem to show that

$$\int_E |f|^p dm = \int_0^\infty pt^{p-1} m\{x \in E : |f(x)| > t\} dt.$$

(The integrand involves the distribution function of Exercises 7 and 8.)

28. Let  $f$  and  $g$  be integrable functions on  $\mathbf{R}^n$ . Show that for almost all  $x \in \mathbf{R}^n$ , the function  $y \mapsto f(x - y)g(y)$  is integrable. Let  $h(x) = \int f(x - y)g(y) dm(y)$ . Show that  $h$  is integrable, and that  $\int |h| dm \leq (\int |f| dm)(\int |g| dm)$ . (The function  $h$  is called the *convolution* of  $f$  and  $g$ .)

29. Use the results of this chapter to compute  $\omega_n = m(B^n)$ , where  $B^n$  is the unit ball in  $\mathbf{R}^n$ . HINT: Obtain a formula for  $\omega_{n+2}$  in terms of  $\omega_n$ . Clearly,  $\omega_1 = 2$ , and  $\omega_0 = 1$ .

30. Let  $f : U \rightarrow \mathbf{R}$  be of class  $C^2$ , where  $U$  is an open set in  $\mathbf{R}^2$ . Let  $C = \{x \in U : df(x) = 0\}$ ;  $C$  is called the set of *critical points* of  $f$ . Show that  $f(C)$ , the set of *critical values* of  $f$ , has measure 0. HINT: Let  $C = A \cup B$ , where  $B = \{x \in C : D_j D_k f(x) = 0, i, j = 1, 2\}$ , and  $A = C \setminus B$ . Show that  $m(A) = 0$  by using the implicit function theorem, and that  $m(B) = 0$  by using Taylor's theorem.

## 10.10 Notes

The modern approach to integration began with Lebesgue's thesis, in 1902. The approach taken here is somewhat different from that of Lebesgue, but the underlying idea of measurable functions, and sums of the form  $\sum c_k \mu(E_k)$  was there from the beginning. Lebesgue dealt originally only

with functions on the line, and Lebesgue measure. The extension to integrals in any dimension space, even infinite-dimensional space, occurred gradually over the following decades.

There are other approaches to the Lebesgue integral which have their advantages. One approach, originating with the work of Daniell (1917-18), begins with a linear functional  $I$  defined on a space  $\mathcal{E}$  of functions on a set  $X$ , for instance, the Riemann integral thought of as a linear functional on the continuous functions on  $[a, b]$ , or perhaps just on the space of step functions. It is assumed that  $I$  is positive, i.e., that  $I(f) \geq 0$  for every  $f \in \mathcal{E}$  with  $f \geq 0$ , and that  $I$  is continuous in the following sense: if  $f_n \in \mathcal{E}$  and  $f_n$  decrease to 0 pointwise, then  $\lim(f_n) = 0$ . One then extends the functional to the class of limits of increasing sequences in  $\mathcal{E}$ , then to differences of functions in this class, etc. The integral is then defined on a large class of functions without explicitly invoking the concept of measure; finally, one can define the measure of a set as the integral of its indicator function. Such a program is carried out, for instance, in [12].

When  $n = 2$ , the statement of Corollary 10.26 reduces simply to the definition of a convex function; the original proof of this corollary (by the Danish mathematician J.W.L. Jensen) was by induction on  $n$ , and the original proof of Theorem 10.25 was by reducing it to the corollary. The integral inequality of Corollary 10.27 is also known as the inequality of the arithmetic and geometric means. Hölder proved his inequality for finite sums, deriving it from an equivalent inequality of Rogers, and F. Riesz extended it to integrals. Minkowski proved his inequality (actually, a different, less symmetrically stated inequality equivalent to what we call Minkowski's inequality) for finite sums; the deduction from Hölder's inequality, and the extension to integrals, are due to Riesz. Another proof of the Hölder inequality begins with the elementary inequality  $ab \leq a^p/p + b^q/q$ , valid for  $a, b \geq 0$ ; integrated, this yields  $\int |fg| d\mu \leq 1$  if  $\int |f|^p d\mu = \int |g|^q d\mu = 1$ , and the general case follows by homogeneity.

The change of variable formula (Theorem 10.46) is due to Jacobi.

As we mentioned, Theorem 10.52 is usually referred to as Fubini's theorem. It was well-known in the nineteenth century that  $\int_a^b \int_c^d f(x, y) dx dy = \int_c^d \int_a^b f(x, y) dy dx$  when  $f$  is continuous on the rectangle  $\{a \leq x \leq b, c \leq y \leq d\}$ ; it was proved by Lebesgue in 1904 for  $f$  a bounded measurable function. Fubini stated the theorem for not necessarily bounded measurable functions in 1907, but his proof was unconvincing. The first proof generally regarded as correct was given by Tonelli in 1909. Whose name should be attached? The appellation "Fubini-Tonelli theorem" may be coming into use.

In this chapter we formulate the notion of a *manifold*, which generalizes the familiar ideas of a (smooth) curve or surface. The intuitive idea of a curve in  $\mathbf{R}^2$  or  $\mathbf{R}^3$  is that of a subset  $C$  of  $\mathbf{R}^2$  or  $\mathbf{R}^3$  which locally looks like a segment of the real line; in other words, if  $\mathbf{p} \in C$ , there should be a neighborhood  $U$  of  $\mathbf{p}$  in  $\mathbf{R}^3$  and an interval  $(a, b)$  in  $\mathbf{R}$ , together with a bijective map  $\alpha : (a, b) \rightarrow C \cap U$  which is bicontinuous. If  $\mathbf{p}$  is an endpoint of  $C$ , the interval  $(a, b)$  should be replaced by an interval  $[a, b)$  or  $(a, b]$ . This formulation is purely topological, i.e., defined in terms of continuity; we want to restrict our attention to differentiable curves, which should mean that  $\alpha$  and  $\alpha^{-1}$  are required to be differentiable. We have to explain what we mean by  $\alpha^{-1}$  being differentiable, since its domain is not an open subset of  $\mathbf{R}^3$ . We will also formulate the notion of the tangent space to a differentiable manifold, generalizing the familiar ideas of tangent line to a curve or tangent plane to a surface, and discuss the idea of orientation.

## 11.1 Definitions

**Notation.** Throughout this and the following chapters, we will denote by  $\mathbf{x}$  the identity function on  $\mathbf{R}^n$  (or its restrictions), and denote the associated coordinate functions by  $x^1, x^2, \dots, x^n$ ; thus, for example,  $x^2(a_1, \dots, a_n) = a_2$ . This annoying use of superscripts will make it awkward to write polynomial functions, but will make some complicated expressions a little easier in later chapters. When  $n = 2$  or  $3$ , we will sometimes use  $x$ ,  $y$ , and  $z$  to denote the coordinate functions. In Chapter 8, we defined the notion of

differentiability of a function at a point which was interior to its domain. We now want to extend the idea of differentiability.

**11.1 Definition.** Let  $A$  be a subset of  $\mathbf{R}^k$ . A map  $\mathbf{f} : A \rightarrow \mathbf{R}^n$  is said to be of class  $C^r$  if for each  $\mathbf{p} \in A$ , there is an open neighborhood  $U$  of  $\mathbf{p}$  in  $\mathbf{R}^k$ , and  $\mathbf{F}$  of class  $C^r$  defined in  $U$ , such that  $\mathbf{F}(\mathbf{q}) = \mathbf{f}(\mathbf{q})$  for all  $\mathbf{q} \in U \cap A$ . We say that  $\mathbf{f}$  is smooth if it is of class  $C^\infty$ . A function  $\mathbf{f} : A \rightarrow B$  is called a diffeomorphism if  $\mathbf{f}$  is bijective, and both  $\mathbf{f}$  and  $\mathbf{f}^{-1}$  are smooth.

It is clear that if  $A$  is open, this definition gives the same meaning to the class  $C^r$  as we had before. In general, if  $\mathbf{f}$  is smooth on  $A \subset \mathbf{R}^k$ , the differential  $d\mathbf{f}$  of  $\mathbf{f}$  is not well-defined (except at interior points of  $A$ ), since the extended map  $\mathbf{F}$  is not uniquely determined. For instance, if  $A = \{(s, t) \in \mathbf{R}^2 : t = 0\}$ , the 0 function on  $A$  is the restriction to  $A$  of the coordinate function  $x^2$  (the function  $y$ ) as well as the 0 function on  $\mathbf{R}^2$ . But if  $A$  is contained in the closure of its interior, and  $\mathbf{f}$  is of class  $C^1$  on  $A$ , then  $d\mathbf{f}$ , and its matrix  $\mathbf{f}'$ , are unambiguously defined on  $A$ .

**Notation.** Let  $\mathbf{R}_+^k = \{\mathbf{u} \in \mathbf{R}^k : u_k \geq 0\}$ ; we refer to  $\mathbf{R}_+^k$  as the *upper halfspace* in  $\mathbf{R}^k$ .

It follows from the remark above that if  $\mathbf{f}$  is smooth in an open subset  $V$  of  $\mathbf{R}_+^k$ , then  $d\mathbf{f}$  is well-defined in  $V$ .

**11.2 Definition.** Let  $k$  be a positive integer. A  $k$ -dimensional manifold (or simply  $k$ -manifold) in  $\mathbf{R}^n$  is a subset  $M$  of  $\mathbf{R}^n$  with the following property: for each  $\mathbf{p} \in M$ , there exists a map  $\alpha$ , satisfying:

- (a) if  $k > 1$ , the domain  $V_\alpha$  of  $\alpha$  is an open subset of  $\mathbf{R}^k$  or  $\mathbf{R}_+^k$ ; if  $k = 1$ ,  $V_\alpha$  is an open subset of  $\mathbf{R}$  or  $\mathbf{R}_+$  or  $\mathbf{R}_- = (-\infty, 0]$ ;
- (b)  $\alpha$  is a smooth map of  $V_\alpha$  into  $\mathbf{R}^n$ , and its differential  $d\alpha_t$  is nonsingular at each point  $t \in V_\alpha$ ; and
- (c)  $\alpha$  is injective on  $V_\alpha$ , its image  $U_\alpha = \alpha(V_\alpha)$  is an open subset of  $M$  containing  $\mathbf{p}$  (i.e.,  $U_\alpha$  is open in the relative topology of  $M$ ), and  $\alpha^{-1}$  is continuous on  $U_\alpha$ .

The space  $\mathbf{R}^0$  could only be interpreted to mean a zero-dimensional vector space, thus consisting of one point,  $\{0\}$ . Then (a) and (b) above are automatically satisfied, and only (c) has content: it asserts that  $\alpha(0)$  is an open subset of  $M$ . Thus we arrive at the definition:

**11.3 Definition.** A manifold of dimension 0 in  $\mathbf{R}^n$  is a subset  $M$  of  $\mathbf{R}^n$  which consists entirely of isolated points.



If we replace the word "smooth" by  $C^r$  in Definition 11.2, we arrive at the idea of a manifold of class  $C^r$ . When  $r = 0$ , we have the notion of topological manifold. We will stay with the concept of  $C^\infty$  manifold.

Any map  $\alpha$  having the properties (a)–(c) of Definition 11.2 is called a *local coordinate for  $M$  at  $\mathbf{p}$* , and the image  $U_\alpha$  of  $\alpha$  will be referred to as a *coordinate patch at  $\mathbf{p}$* . Obviously, such an  $\alpha$  is also a local coordinate for  $M$  at  $\mathbf{q}$ , and  $U_\alpha$  is a coordinate patch at  $\mathbf{q}$ , for any  $\mathbf{q} \in U_\alpha$ . It is straightforward to see that any open subset of a coordinate patch is also a coordinate patch, or to put it another way, the restriction of a local coordinate to an open subset of its domain is also a local coordinate. It follows that a (relatively) open subset of a  $k$ -manifold in  $\mathbf{R}^n$  is again a  $k$ -manifold in  $\mathbf{R}^n$ .

**11.4 Lemma.** *If  $\alpha$  is a local coordinate for the manifold  $M$  in  $\mathbf{R}^n$ , then  $\alpha^{-1}$  is smooth on  $U_\alpha$ .*

**Proof.** Let  $\mathbf{p} \in U_\alpha$ , and suppose  $\mathbf{p} = \alpha(\mathbf{t})$ . Since  $\alpha'(\mathbf{t})$  is nonsingular, the vectors  $\alpha'(\mathbf{t})\mathbf{e}_1, \dots, \alpha'(\mathbf{t})\mathbf{e}_k$  are linearly independent; hence, there exist vectors  $\mathbf{v}_{k+1}, \dots, \mathbf{v}_n$  such that

$$\alpha'(\mathbf{t})\mathbf{e}_1, \dots, \alpha'(\mathbf{t})\mathbf{e}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n$$

form a basis for  $\mathbf{R}^n$ . Define a map  $\mathbf{G}$  of  $V_\alpha \times \mathbf{R}^{n-k} \rightarrow \mathbf{R}^n$  by

$$\mathbf{G}(u_1, \dots, u_n) = \alpha(u_1, \dots, u_k) + \sum_{j=k+1}^n u_j \mathbf{v}_j;$$

it is evident that  $\mathbf{G}$  is smooth, and that

$$\mathbf{G}'(u_1, \dots, u_n)\mathbf{e}_j = \begin{cases} \alpha'(u_1, \dots, u_k)\mathbf{e}_j & \text{if } 1 \leq j \leq k; \\ \mathbf{v}_j & \text{if } k+1 \leq j \leq n. \end{cases}$$

Thus  $\mathbf{G}'(\mathbf{t}, \mathbf{0})$  is nonsingular, so according to the inverse function theorem (Theorem 8.27), there exists an open neighborhood  $\tilde{V}$  of  $(\mathbf{t}, \mathbf{0})$  in  $\mathbf{R}^n$  which  $\mathbf{G}$  maps injectively onto an open set  $\tilde{U}$ , with  $\mathbf{p} \in \tilde{U}$ , and a smooth  $\mathbf{F}$  defined on  $\tilde{U}$  such that  $\mathbf{F} \circ \mathbf{G}$  is the identity on  $\tilde{V}$ . But if  $\mathbf{q} \in \tilde{U} \cap M$ , then  $\mathbf{q} = \alpha(u_1, \dots, u_k) = \mathbf{G}(u_1, \dots, u_k, 0, \dots, 0)$ , so  $\mathbf{F}(\mathbf{q}) = (u_1, \dots, u_k, 0, \dots, 0) = (\alpha^{-1}(\mathbf{q}), \mathbf{0})$ ; thus  $\alpha^{-1}$  agrees on  $\tilde{U} \cap M$  with the restriction of the smooth function  $(F_1, \dots, F_k)$ , i.e.,  $\alpha^{-1}$  is smooth. ■

**11.5 Corollary.** *Let  $M$  be a  $k$ -manifold in  $\mathbf{R}^n$ . A map  $\mathbf{f} : M \rightarrow \mathbf{R}^m$  is smooth on  $M$  if and only if  $\mathbf{f} \circ \alpha$  is smooth on  $V_\alpha$  for every local coordinate  $\alpha$  for  $M$ .*

**Proof.** If  $\mathbf{f}$  is smooth (see Definition 11.1) there exists, for each  $\mathbf{p} \in M$ , a neighborhood  $\tilde{U}$  of  $\mathbf{p}$  in  $\mathbf{R}^n$  and a smooth  $\mathbf{F} : \tilde{U} \rightarrow \mathbf{R}^m$  such that  $\mathbf{F} = \mathbf{f}$  on  $M \cap \tilde{U}$ . Then  $\mathbf{f} \circ \alpha = \mathbf{F} \circ \alpha$  is also smooth, for any local coordinate  $\alpha$ . On the other hand, if  $\mathbf{f} \circ \alpha$  is smooth in  $V_\alpha$ , then since  $\alpha^{-1}$  is smooth in  $U_\alpha$  by Lemma 11.4, it follows that  $\mathbf{f} = (\mathbf{f} \circ \alpha) \circ \alpha^{-1}$  is smooth in  $U_\alpha$ . ■

**11.6 Definition.** Let  $M$  be a manifold in  $\mathbf{R}^n$ . For each pair  $\alpha$  and  $\beta$  of local coordinates such that  $U_\alpha \cap U_\beta \neq \emptyset$ , let  $V_{\alpha\beta} = \beta^{-1}(U_\alpha \cap U_\beta)$ , and define the transition function  $\varphi_{\alpha\beta} : V_{\alpha\beta} \rightarrow V_{\beta\alpha}$  by  $\varphi_{\alpha\beta} = \alpha^{-1} \circ \beta$ .

It is clear that  $V_{\alpha\beta}$  is an open subset of  $\mathbf{R}^k$  or  $\mathbf{R}_+^k$ , and that  $\varphi_{\alpha\beta}$  maps  $V_{\alpha\beta}$  bijectively to  $V_{\beta\alpha}$ ; evidently,  $\varphi_{\alpha\beta}^{-1} = \varphi_{\beta\alpha}$ . Applying Lemma 11.4, and the chain rule, we see that  $\varphi_{\alpha\beta}$  is smooth, and since it has a smooth inverse, that  $\varphi'_{\alpha\beta}$  is always nonsingular. In other words,  $\varphi_{\alpha\beta}$  is a diffeomorphism of  $V_{\alpha\beta}$  onto  $V_{\beta\alpha}$ .

**11.7 Definition.** Let  $M$  be a manifold in  $\mathbf{R}^n$ . We say that  $\mathbf{p}$  lies on the boundary of  $M$ , and write  $\mathbf{p} \in \partial M$ , if  $\mathbf{p} \in M$  and there exists a local coordinate  $\alpha$  at  $\mathbf{p}$  such that  $V_\alpha \subset \mathbf{R}_+^k$ , and  $\mathbf{p} = \alpha(\mathbf{u})$  with  $\mathbf{u} = (u_1, \dots, u_{k-1}, 0)$ .

We observe that if  $\mathbf{p} \in \partial M$ , and  $\beta$  is any local coordinate at  $\mathbf{p}$ , then  $V_\beta \subset \mathbf{R}_+^k$ , and  $x^k(\beta^{-1}(\mathbf{p})) = 0$ . (Recall that  $x^k$  denotes the  $k$ th coordinate function in  $\mathbf{R}^k$ .) For if  $\mathbf{p} = \beta(\mathbf{t}) = \alpha(\mathbf{u})$ , and  $\mathbf{t}$  is in the interior (relative to  $\mathbf{R}^k$ ) of  $V_\beta$ , then  $\varphi_{\alpha\beta}(\mathbf{t}) = \mathbf{u}$ , and by the inverse function theorem, there is an open neighborhood  $N$  (in  $\mathbf{R}^k$ ) of  $\mathbf{t}$  which  $\varphi_{\alpha\beta}$  maps onto an open set (in  $\mathbf{R}^k$ ) containing  $\mathbf{u}$ ; but if  $\mathbf{t}$  lies in the interior (relative to  $\mathbf{R}^k$ ) of  $V_\beta$ , this implies that  $\mathbf{u}$  lies in the interior of  $V_\alpha$ , contrary to the definition of  $\partial M$ .

**11.8 Lemma.** If  $M$  is a manifold of dimension  $k$  in  $\mathbf{R}^n$ , then  $\partial M$  is a manifold of dimension  $k - 1$ , with empty boundary.

**Proof.** Let  $\mathbf{p} \in \partial M$ , and let  $\alpha$  be a local coordinate at  $\mathbf{p}$ . Let  $V_{\hat{\alpha}} = \{\mathbf{u} \in \mathbf{R}^{k-1} : (\mathbf{u}, 0) \in V_\alpha\}$ , so  $V_{\hat{\alpha}}$  is an open set in  $\mathbf{R}^{k-1}$ . Define  $\hat{\alpha} : V_{\hat{\alpha}} \rightarrow \partial M$  by  $\hat{\alpha}(\mathbf{u}) = \alpha(\mathbf{u}, 0)$ . It is easy to verify that  $\hat{\alpha}$  is a local coordinate at  $\mathbf{p}$  for  $\partial M$ . ■

**11.9 Example.** Any open subset  $U$  of  $\mathbf{R}^n$  is a manifold of dimension  $n$  in  $\mathbf{R}^n$ ; one coordinate patch suffices, with the identity map serving as local coordinate (global, in this case.) The halfspace  $\mathbf{R}_+^n$  is a manifold of dimension  $n$ ; again, one local coordinate suffices. The boundary  $\partial U$  of an open set  $U$  is empty; of course,  $\text{bdry } U$ , the boundary of  $U$  regarded as a subset of the metric space  $\mathbf{R}^n$ , is not empty unless  $U = \mathbf{R}^n$ . We see that  $\partial \mathbf{R}_+^n = \text{bdry } \mathbf{R}_+^n = \mathbf{R}^{n-1} \times \{0\}$ , a manifold of dimension  $n - 1$  in  $\mathbf{R}^n$ .

**11.10 Example.** If  $V$  is an open set in  $\mathbf{R}^k$ , and  $\mathbf{f} : V \rightarrow \mathbf{R}^{n-k}$  is a smooth map, then the graph of  $\mathbf{f}$ ,

$$\{(t, \mathbf{f}(t)) : t \in V\}$$

is a  $k$ -manifold in  $\mathbf{R}^n$ . For instance,  $M = \{(t, \sin(1/t)) : t \neq 0\}$  is a one-dimensional manifold in  $\mathbf{R}^2$ . This example shows that the closure of a manifold need not be a manifold.

**11.11 Example.** The closed ball  $B^n = \{\mathbf{u} \in \mathbf{R}^n : |\mathbf{u}| \leq 1\}$  is a manifold of dimension  $n$  in  $\mathbf{R}^n$ . This time more than one local coordinate is needed. Here is one way to cover  $B^n$  with coordinate patches: let  $V_0 = U_0 = \{\mathbf{u} : |\mathbf{u}| < 1\}$ , and let  $\alpha_0$  be the identity map. Define  $g : \mathbf{R}^n \rightarrow \mathbf{R}$  by  $g(\mathbf{u}) = (1 - u_n) - \sum_{j=1}^{n-1} u_j^2$ . Then

$$V = \{\mathbf{u} : g(\mathbf{u}) > 0, 0 \leq u_n < 1\}$$

is an open subset of  $\mathbf{R}_+^n$ . We put  $V_1 = \cdots = V_{2n} = V$ , and define

$$\begin{aligned}\alpha_1(\mathbf{u}) &= (\sqrt{g(\mathbf{u})}, u_1, \dots, u_{n-1}), \\ \alpha_2(\mathbf{u}) &= (-\sqrt{g(\mathbf{u})}, u_1, \dots, u_{n-1}), \\ \alpha_3(\mathbf{u}) &= (u_1, \sqrt{g(\mathbf{u})}, u_2, \dots, u_{n-1}), \\ &\dots \dots \dots \\ \alpha_{2n}(\mathbf{u}) &= (u_1, \dots, u_{n-1}, -\sqrt{g(\mathbf{u})}),\end{aligned}$$

and observe that  $|\alpha_j(\mathbf{u})|^2 = 1 - u_n$ , so  $\alpha_j$  maps  $V$  into  $B^n$ , and  $V \cap \{u_n = 0\}$  into  $S^{n-1} = \{\mathbf{u} \in \mathbf{R}^n : |\mathbf{u}| = 1\}$ . It is clear that each  $\alpha_j$  is injective. We verify that  $\alpha'_j$  is nonsingular in the case  $j = 2n$ : writing  $\alpha_{2n} = (f^1, \dots, f^n)$ , we have  $D_j f^k = \delta_j^k$  for  $1 \leq j \leq n$ ,  $1 \leq k < n$ , so  $\det(\alpha'_{2n})(\mathbf{u}) = D_n f^n(\mathbf{u}) = 1/(2\sqrt{g(\mathbf{u})}) > 0$ . Thus  $\alpha'_{2n}$  is nonsingular, and since each  $\alpha_j$  is the composition of  $\alpha_{2n}$  with a permutation of coordinates and possibly a reflection in the last coordinate, each  $\alpha'_j$  is nonsingular. Now the inverse function theorem tells us that each  $\alpha_j$  is an open map, so all the requirements for a local coordinate are satisfied. We note that if  $|\mathbf{u}| = 1$ , then  $u_j \neq 0$  for some  $j$ , so  $\mathbf{u} \in \alpha_k(V)$  for  $k = 2j - 1$  or  $k = 2j$ . In this example, we see that  $\partial B^n = S^{n-1} = \{\mathbf{u} \in \mathbf{R}^n : |\mathbf{u}| = 1\}$  is the boundary of  $B^n$  in the usual sense. We could have covered  $B^n$  with just two coordinate patches, instead of  $2n + 1$ , but the procedure chosen seems natural, and generalizes to a large class of manifolds. See Theorem 11.17 below.

**11.12 Example.** The set  $M = \{(u_1, u_2, u_3) : u_1^2 + u_2^2 = 1, -1 \leq u_3 \leq 1\}$  is an example of a 2-manifold (or *surface*) in  $\mathbf{R}^3$ ; it is easily covered by four coordinate patches. You can check that  $\partial M$  is the union of the two circles  $\{u_1^2 + u_2^2 = 1, u_3 = 1\}$  and  $\{u_1^2 + u_2^2 = 1, u_3 = -1\}$ , which, of course, is not the boundary of  $M$  in the usual sense:  $\text{bdry } M = M$ , as is easily seen.

**11.13 Example.** Let  $0 < a < b$  and consider the map  $\mathbf{f} : \mathbf{R}^2 \rightarrow \mathbf{R}^3$  defined by

$$\mathbf{f}(s, t) = (b \cos s, b \sin s, 0) + a \cos t (\cos s, \sin s, 0) + a \sin t (0, 0, 1);$$

the image of  $\mathbf{f}$  is a 2-manifold without boundary (a two-dimensional torus in  $\mathbf{R}^3$ ), and the restriction of  $\mathbf{f}$  to any small enough open set is a local coordinate. It is not hard to see that  $M$  is obtained by rotating the circle  $\{(b + a \cos t, 0, a \sin t) : 0 \leq t \leq 2\pi\}$  about the  $z$ -axis.

**11.14 Example.** Again, let  $0 < a < b$ , and now define  $\mathbf{g} : \mathbf{R} \times [-a, a] \rightarrow \mathbf{R}^3$  by

$$\mathbf{g}(s, t) = (b \cos s, b \sin s, 0) + t \left( \sin \frac{s}{2} (\cos s, \sin s, 0) + \cos \frac{s}{2} (0, 0, 1) \right);$$

the rectangle  $[0, 2\pi) \times [-a, a]$  is mapped by  $\mathbf{g}$  onto the image  $M$  of  $\mathbf{g}$  in a one-one way, the segment  $2\pi \times [-a, a]$  is mapped onto the same line segment  $\{(1, 0, t) : -a \leq t \leq a\}$  as the segment  $0 \times [-a, a]$ , but this segment is traced out now in the opposite direction. This compact surface is known as the *Möbius strip*, or *Möbius band*. You can check that  $\partial M$  is the 1-manifold  $\{\mathbf{g}(s, a) : 0 \leq s \leq 4\pi\}$ , which is the diffeomorphic image of a circle.

**11.15 Example.** Let  $\mathbf{f} : \mathbf{R} \rightarrow \mathbf{R}^2$  be defined by  $\mathbf{f}(t) = (t^2, t^3)$ . Then  $M = \mathbf{f}(\mathbf{R})$  is *not* a manifold, even though  $\mathbf{f}$  is a bijective smooth map of  $\mathbf{R}$  onto  $M$ , and has a continuous inverse;  $\mathbf{f}$  fails to be a coordinate because  $\mathbf{f}'(0) = (0, 0)$ , so that  $\mathbf{f}^{-1}$  is not smooth. In fact, if  $\alpha$  is any smooth map of an open  $V \subset \mathbf{R}$  into  $M$ , with, say,  $\alpha(0) = (0, 0)$ , then necessarily  $\alpha'(0) = (0, 0)$ . To see this, write  $\alpha = (g, h)$ ; since  $g(t) \geq 0$  for all  $t \in V$  and  $g(0) = 0$ , it follows that  $g'(0) = 0$ . Since  $h^2 = g^3$ , we have  $2h(t)h'(t) = 3g^2(t)g'(t)$  for  $t \in V$ , and hence  $2h'(t) = 3g^{1/2}(t)g'(t)$ , so  $h'(0) = 0$ . Thus  $\alpha'(0) = 0$ ; there exists no local coordinate for  $M$  at  $(0, 0)$ .

**11.16 Example.** One can construct a subset  $M$  of  $\mathbf{R}^2$ , such that  $M = \mathbf{f}(V)$ , where  $V$  is an open interval in  $\mathbf{R}$  and  $\mathbf{f}$  is a smooth mapping of  $V$  into  $\mathbf{R}^2$ , such that  $\mathbf{f}$  is one-one on  $V$ , and  $\mathbf{f}'$  is never  $0$  (thus, is nonsingular), yet  $M$  is not a manifold. Take  $M$  to be a *lemniscate*, for instance, (i.e., a figure eight, if you've forgotten). Let  $V = \mathbf{R}$ , and

$$\mathbf{f}(t) = \left( \frac{t(1+t^2)}{1+t^4}, \frac{t(1-t^2)}{1+t^4} \right);$$

verifying the assertions is not at all hard, but a sketch tells the story best. The problem here is that  $\mathbf{f}^{-1}$  is not continuous on  $M$ ; if we removed the origin from  $M$ , we would be left with a manifold. (By the way,  $M$  is described in polar coordinates by the equation  $r^2 = \cos 2\theta$ ; can you verify this?)

## 11.2 Constructing Manifolds

**11.17 Theorem.** Let  $\Omega$  be an open set in  $\mathbf{R}^n$ , and let  $\varphi$  be a smooth real-valued function on  $\Omega$ . Let  $M = \{\mathbf{p} \in \Omega : \varphi(\mathbf{p}) \geq 0\}$ , and  $\Gamma = \{\mathbf{p} \in \Omega : \varphi(\mathbf{p}) = 0\}$ . Suppose  $M \neq \emptyset$ , and that  $d\varphi_{\mathbf{p}} \neq 0$  for every  $\mathbf{p} \in \Gamma$ . Then  $M$  is an  $n$ -manifold, and  $\partial M = \Gamma$ . In particular,  $\Gamma$  is an  $(n-1)$ -manifold in  $\mathbf{R}^n$ .

**Proof.** Let  $U_0 = \{\mathbf{p} : \varphi(\mathbf{p}) > 0\}$ ; then  $U_0$  is open, and the identity map serves as a local coordinate. Let  $\mathbf{p} \in \Gamma$ ; since  $d\varphi_{\mathbf{p}} \neq 0$ , there exists  $k$ ,  $1 \leq k \leq n$ , such that  $D_k\varphi(\mathbf{p}) \neq 0$ , and we may suppose  $k = n$ . Let  $F(\mathbf{u}) = (u_1, \dots, u_{n-1}, \varphi(\mathbf{u}))$ ; then  $F$  is a smooth map of  $\Omega$  into  $\mathbf{R}^n$ , and  $dF_{\mathbf{p}}$  is nonsingular, since its determinant is  $D_n\varphi(\mathbf{p})$ . Hence, by the inverse function theorem, there exists a neighborhood  $\tilde{U}$  of  $\mathbf{p}$  such that  $F$  maps  $\tilde{U}$  diffeomorphically onto an open set  $\tilde{V}$  in  $\mathbf{R}^n$ ; if  $U = M \cap \tilde{U}$ , then  $F(U) = V = \tilde{V} \cap \mathbf{R}_+^n$ . Thus  $\alpha = F^{-1}$  is the desired local coordinate at  $\mathbf{p}$ . Since  $x^n(F(\mathbf{p})) = 0$ , we see that  $\mathbf{p} \in \partial M$ . ■

The reader should examine Example 11.11 above in the light of this theorem.

It is easy to see that  $\partial M \subset \text{bdry } M$  in general; in the setting of Theorem 11.17, we have equality in the important special case that  $M$  is compact. This is left as an exercise.

**11.18 Theorem.** Let  $\Omega$  be an open set in  $\mathbf{R}^n$ , let  $\mathbf{f}$  be a smooth mapping of  $\Omega$  into  $\mathbf{R}^{n-k}$ , and let  $M = \{\mathbf{p} \in \Omega : \mathbf{f}(\mathbf{p}) = \mathbf{0}\}$ . If  $\mathbf{f}'$  has maximal rank at each point of  $M$ , i.e.,  $\mathbf{f}'$  has rank  $n - k$  everywhere in  $M$ , then  $M$  is a manifold of dimension  $k$ , with empty boundary.

**Proof.** Let  $\mathbf{p} \in M$ . The hypothesis is that the linear map  $d_{\mathbf{p}}\mathbf{f}$  maps  $\mathbf{R}^n$  onto  $\mathbf{R}^{n-k}$ . Without loss of generality, we can suppose that  $\mathbf{e}_{k+1}, \dots, \mathbf{e}_n$  are mapped to a linearly independent set (i.e., a basis) in  $\mathbf{R}^{n-k}$  (i.e., the last  $n - k$  columns of  $\mathbf{f}'(\mathbf{p})$  are linearly independent). According to the implicit function theorem (Theorem 8.29), there exist an open neighborhood  $U$  of  $\mathbf{p}$  in  $\mathbf{R}^n$ , an open set  $V$  in  $\mathbf{R}^k$ , and a smooth function  $\mathbf{g} : V \rightarrow \mathbf{R}^{n-k}$ , such that  $M \cap U = \{(\mathbf{u}, \mathbf{g}(\mathbf{u})) : \mathbf{u} \in V\}$ . Thus the map  $\alpha : V \rightarrow \mathbf{R}^n$  defined by  $\alpha(\mathbf{u}) = (\mathbf{u}, \mathbf{g}(\mathbf{u}))$  is a local coordinate for  $M$  at  $\mathbf{p}$ . ■

Another look at the proof of Lemma 11.4 will convince you that any  $k$ -dimensional manifold without boundary in  $\mathbf{R}^n$  is *locally* of the form described in Theorem 11.18. In the course of proving that lemma, we found, for any  $\mathbf{p} \in M$ , a neighborhood  $\tilde{U}$  of  $\mathbf{p}$  in  $\mathbf{R}^n$ , and smooth functions  $F_1, \dots, F_n$  such that  $M \cap \tilde{U} = \{\mathbf{q} \in \tilde{U} : F_{k+1}(\mathbf{q}) = \dots = F_n(\mathbf{q}) = 0\}$ . The map  $(F_{k+1}, \dots, F_n)$  had a derivative of maximal rank since the map  $(F_1, \dots, F_n)$  had a nonsingular derivative.

If we drop the hypothesis in Theorem 11.18 that  $\mathbf{f}'$  has maximal rank, then  $M$  is not necessarily a manifold (of course, it might be: consider  $f(u, v) = u^2$ ), as shown by the example  $f(u, v) = uv$ , for instance. A more elaborate example is the set  $M$  of Example 11.16, which can also be described as  $\{(u, v) : (u^2 + v^2)^2 = u^2 - v^2\}$ . Another example:  $f(u, v) = u^3 - v^2$  (see Example 11.15).

## 11.3 Tangent Spaces

If  $M$  is a manifold of dimension  $k \geq 1$  in  $\mathbf{R}^n$ ,  $\mathbf{p} \in M$ , and  $\alpha$  a local coordinate at  $\mathbf{p}$ ,  $\alpha(\mathbf{t}) = \mathbf{p}$ , then  $d\alpha_{\mathbf{t}}(\mathbf{R}^k)$  is a  $k$ -dimensional subspace of  $\mathbf{R}^n$  (since  $d\alpha_{\mathbf{t}}$  is nonsingular.) Furthermore, this space does not depend on the choice of local coordinate  $\alpha$ ; for if  $\beta$  is another local coordinate at  $\mathbf{p}$ ,  $\beta(\mathbf{u}) = \mathbf{p}$ , then, as we have seen,  $\beta = \alpha \circ \varphi_{\alpha\beta}$ , where  $\varphi_{\alpha\beta}$  is a diffeomorphism of the neighborhood  $V_{\alpha\beta}$  of  $\mathbf{u}$  onto the neighborhood  $V_{\beta\alpha}$  of  $\mathbf{t}$ ; since  $(d\varphi_{\alpha\beta})_{\mathbf{u}}$  is nonsingular, it maps  $\mathbf{R}^k$  onto itself, so  $d\alpha_{\mathbf{t}}$  and  $d\beta_{\mathbf{u}} = d\alpha_{\mathbf{t}} \circ (d\varphi_{\alpha\beta})_{\mathbf{u}}$  have the same range.

**11.19 Definition.** Let  $M$  be a manifold of dimension  $k$  in  $\mathbf{R}^n$ , and let  $\mathbf{p} \in M$ . If  $\alpha$  is a local coordinate with  $\alpha(\mathbf{t}) = \mathbf{p}$ , we refer to the subspace  $d\alpha_{\mathbf{t}}(\mathbf{R}^k)$  as the tangent space to  $M$  at  $\mathbf{p}$ , and denote it by  $T_{\mathbf{p}}(M)$ . The elements of  $T_{\mathbf{p}}(M)$  are called tangent vectors at  $\mathbf{p}$  to  $M$ .

**11.20 Definition.** If  $\nu \in \mathbf{R}^n$  and for every  $\tau \in T_{\mathbf{p}}(M)$  we have  $\nu \perp \tau$  (i.e.,  $\langle \nu, \tau \rangle = 0$ , where  $\langle \cdot, \cdot \rangle$  refers to the usual inner product in  $\mathbf{R}^n$ ), we say that  $\nu$  is normal to  $M$  at  $\mathbf{p}$ ; the set of normal vectors thus forms a subspace of dimension  $n - k$  of  $\mathbf{R}^n$ .

Although we have defined the tangent space  $T_{\mathbf{p}}(M)$  to be a linear subspace of  $\mathbf{R}^n$ , we visualize it as the affine subspace  $\mathbf{p} + T_{\mathbf{p}}(M)$ , i.e., as having its origin located at  $\mathbf{p}$ . Thus,  $T_{\mathbf{p}}(M)$  is to be regarded as distinct from  $T_{\mathbf{q}}(M)$  if  $\mathbf{p} \neq \mathbf{q}$ , even if  $d\alpha$  and  $d\beta$  happen to have the same range for some local coordinates  $\alpha$  at  $\mathbf{p}$  and  $\beta$  at  $\mathbf{q}$ . A better definition of the tangent space would then be:  $T_{\mathbf{p}}(M) = (\mathbf{p}, d\alpha_{\mathbf{t}}(\mathbf{R}^k))$ . We have chosen the sloppier form in order to avoid excessive notation.

We give the following definition for more accuracy:

**11.21 Definition.** If  $M$  is a  $k$ -manifold in  $\mathbf{R}^n$ , we define the tangent bundle of  $M$  to be the set

$$T(M) = \{(\mathbf{p}, \mathbf{h}) : \mathbf{p} \in M, \mathbf{h} \in T_{\mathbf{p}}(M)\}.$$

Thus  $T(M)$  is a subset of  $\mathbf{R}^{2n}$ ; it is not hard to see that  $T(M)$  is a manifold of dimension  $2k$ . For instance, if  $\alpha$  is a local coordinate for  $M$  at  $\mathbf{p}$ , with  $\alpha(\mathbf{t}) = \mathbf{p}$ , then a local coordinate for  $T(M)$  at  $(\mathbf{p}, \mathbf{h})$  is furnished by the map  $(\alpha, d\alpha_{\mathbf{t}})$  of  $V_{\alpha} \times \mathbf{R}^k$  to  $T(M)$ . Thus,  $T(M)$  is locally diffeomorphic to the Cartesian product  $U_{\alpha} \times \mathbf{R}^k$ . In general, though, it is not true that  $T(M)$  is diffeomorphic to  $M \times \mathbf{R}^k$ . A map  $X : M \rightarrow T(M)$  with the property that  $X(\mathbf{p}) \in T_{\mathbf{p}}(M)$  for all  $\mathbf{p} \in M$  is called a *vector field* on  $M$ ; it may be continuous, or smooth, in the sense that these words apply to any map of  $M$  into  $\mathbf{R}^{2n}$ .

We can form the tangent space without explicit use of local coordinates, as the next proposition shows. It also provides some justification for the terminology, since it is a familiar idea that the tangent vector to a parametrized curve  $t \mapsto \gamma(t)$  is its derivative  $\gamma'(t)$ .

**11.22 Proposition.** *Let  $M$  be a  $k$ -manifold in  $\mathbf{R}^n$ ,  $k \geq 1$ , and suppose  $\mathbf{p} \in M$ ,  $\mathbf{p} \notin \partial M$ . If  $\mathbf{h} \in \mathbf{R}^n$ , then  $\mathbf{h} \in T_{\mathbf{p}}(M)$  if and only if there exists  $\epsilon > 0$  and a smooth  $\gamma : (-\epsilon, \epsilon) \rightarrow M$ , with  $\gamma(0) = \mathbf{p}$  and  $\gamma'(0) = \mathbf{h}$ .*

**Proof.** Recall that  $\gamma'(0)$  is the matrix of the linear map  $d\gamma_0 : \mathbf{R} \rightarrow \mathbf{R}^n$ , in other words, is the (column) vector  $d\gamma_0 1 = \lim_{s \rightarrow 0} (\gamma(s) - \gamma(0))/s$ , the usual derivative of a function of one variable. Let  $\alpha$  be a local coordinate at  $\mathbf{p}$ , with  $\alpha(\mathbf{t}) = \mathbf{p}$ . If  $\mathbf{h} \in T_{\mathbf{p}}(M)$ , so  $\mathbf{h} = \alpha'(\mathbf{t})\mathbf{k}$  for some  $\mathbf{k} \in \mathbf{R}^k$ , let  $\gamma(s) = \alpha(\mathbf{t} + s\mathbf{k})$ . Then  $\gamma'(0) = \alpha'(\mathbf{t})\mathbf{k}$  by the chain rule. On the other hand, if  $\mathbf{h} = \gamma'(0)$  for some smooth map  $\gamma : (-\epsilon, \epsilon) \rightarrow M$ , with  $\gamma(0) = \mathbf{p}$ , then put  $\Gamma = \alpha^{-1} \circ \gamma$  (we can assume  $\gamma((-\epsilon, \epsilon)) \subset U_{\alpha}$ , by using a smaller  $\epsilon$  if necessary). Since  $\gamma = \alpha \circ \Gamma$ , we have  $\gamma'(0) = \alpha'(\mathbf{t})\Gamma'(0) \in T_{\mathbf{p}}(M)$ . ■

When  $\mathbf{p} \in \partial M$ , this proposition is no longer true, since we can't be sure that  $\mathbf{t} + s\mathbf{k} \in V_{\alpha}$  for  $-\epsilon < s < \epsilon$ , or even for  $0 \leq s < \epsilon$ . However, we can show that if  $\mathbf{h} \in T_{\mathbf{p}}(M)$ , then either  $\mathbf{h}$  or  $-\mathbf{h}$  is of the form  $\gamma'(0)$ , for some smooth  $\gamma : [0, \epsilon) \rightarrow M$ , with  $\gamma(0) = \mathbf{p}$ . We omit the (easy) details.

Suppose  $f$  is a smooth function on the  $k$ -dimensional manifold  $M$  in  $\mathbf{R}^n$  (see Definition 11.1). Thus, given a point  $\mathbf{p} \in M$ , there is a neighborhood  $U$  of  $\mathbf{p}$  in  $\mathbf{R}^n$  and a smooth function  $F$  defined on  $U$ , such that  $F = f$  in  $U \cap M$ . When  $k < n$ ,  $F$  is not uniquely determined, and, in general, we may have  $F$  and  $G$  smooth in  $U$ , with  $F = G$  on  $M \cap U$  but  $dF_{\mathbf{p}} \neq dG_{\mathbf{p}}$ . The next proposition shows that  $df_{\mathbf{p}}$  makes sense as a linear function on  $T_{\mathbf{p}}(M)$ .

**11.23 Lemma.** *Let  $M$  be a  $k$ -dimensional manifold in  $\mathbf{R}^n$ , and  $\mathbf{p} \in M$ . If  $F$  is smooth in a neighborhood  $U$  of  $\mathbf{p}$ , and  $F = 0$  on  $U \cap M$ , then  $dF_{\mathbf{p}}(\mathbf{h}) = 0$  for every  $\mathbf{h} \in T_{\mathbf{p}}(M)$ .*

**Proof.** We can assume that  $U$  is a coordinate patch, say  $U = U_{\alpha}$ , and that  $\alpha(\mathbf{t}) = \mathbf{p}$ . Then  $F \circ \alpha$  vanishes identically on  $V_{\alpha}$ , so its derivative vanishes at  $\mathbf{t}$ ; by the chain rule, we then have  $dF_{\mathbf{p}} \circ d\alpha_{\mathbf{t}} = 0$ , i.e.,  $dF_{\mathbf{p}}(\mathbf{h}) = 0$  whenever  $\mathbf{h} = d\alpha_{\mathbf{t}}(\mathbf{k})$  for some  $\mathbf{k} \in \mathbf{R}^k$ ; this is exactly the statement of the Lemma. ■

**11.24 Proposition.** *Let  $M$  be a manifold in  $\mathbf{R}^n$ , and let  $f$  be a smooth real-valued function on  $M$ . There is a (unique) map  $df$  which assigns to each  $\mathbf{p} \in M$  a linear map  $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow \mathbf{R}$ , with the property that  $df_{\mathbf{p}}(\mathbf{h}) = dF_{\mathbf{p}}(\mathbf{h})$  for every  $\mathbf{h} \in T_{\mathbf{p}}(M)$  and every smooth function  $F$  defined in a neighborhood  $U$  of  $\mathbf{p}$  such that  $F = f$  on  $M \cap U$ .*

**Proof.** If  $F$  and  $G$  are smooth functions in a neighborhood  $U$  of  $\mathbf{p}$  such that  $F = G = f$  on  $M \cap U$ , then the lemma applied to  $F - G$  shows that  $dF_{\mathbf{p}}\mathbf{h} = dG_{\mathbf{p}}\mathbf{h}$  for every  $\mathbf{h} \in T_{\mathbf{p}}(M)$ . ■

If  $M$  is a manifold in  $\mathbf{R}^n$ , and  $\mathbf{f} : M \rightarrow \mathbf{R}^m$  is smooth, then applying Proposition 11.24 to each coordinate of  $\mathbf{f}$  shows that  $d\mathbf{f}_{\mathbf{p}}$  is a well-defined linear map of  $T_{\mathbf{p}}(M)$  into  $\mathbf{R}^m$ .

**11.25 Proposition.** Let  $M$  be a  $k$ -manifold in  $\mathbf{R}^n$ , and let  $N$  be a  $j$ -manifold in  $\mathbf{R}^m$ . If  $\mathbf{f} : M \rightarrow \mathbf{R}^m$  is a smooth map, with  $\mathbf{f}(M) \subset N$ , then  $d\mathbf{f}_{\mathbf{p}}(T_{\mathbf{p}}(M)) \subset T_{\mathbf{f}(\mathbf{p})}(N)$ , for every  $\mathbf{p} \in M$ .

**Proof.** See the exercises at the end of this chapter. ■

We can use the last lemma to obtain yet another characterization of the tangent space in an important special case.

**11.26 Proposition.** Let  $\Omega$  be an open set in  $\mathbf{R}^n$ , and let  $\mathbf{f}$  be a smooth mapping of  $\Omega$  into  $\mathbf{R}^{n-k}$ . If  $\mathbf{f}'$  has maximal rank at each point of  $M = \{\mathbf{p} \in \Omega : \mathbf{f}(\mathbf{p}) = \mathbf{0}\}$ , so that according to Theorem 11.18,  $M$  is a manifold of dimension  $k$ , then for each  $\mathbf{p} \in M$ , we have

$$T_{\mathbf{p}}(M) = \{\mathbf{h} \in \mathbf{R}^n : \mathbf{f}'(\mathbf{p})(\mathbf{h}) = \mathbf{0}\}.$$

**Proof.** See the exercises at the end of this chapter. ■

## 11.4 Orientation

**11.27 Definition.** Let  $V$  be a vector space of dimension  $k$  over  $\mathbf{R}$ . If  $(\mathbf{v}_1, \dots, \mathbf{v}_k)$  and  $(\mathbf{w}_1, \dots, \mathbf{w}_k)$  are (ordered) bases for  $V$ , we may write  $\mathbf{w}_j = \sum_{i=1}^k a_{ij}^i \mathbf{v}_i$ ; we say that  $(\mathbf{v}_1, \dots, \mathbf{v}_k)$  and  $(\mathbf{w}_1, \dots, \mathbf{w}_k)$  have the same orientation if  $\det(a_{ij}^i) > 0$ .

It is easy to see that this defines an equivalence relation on the set of all ordered bases, and that there are exactly two equivalence classes. We call these the two *orientations* of  $V$ .

Let us say that the basis  $(\mathbf{w}_1, \dots, \mathbf{w}_k)$  is a *deformation* of the basis  $(\mathbf{v}_1, \dots, \mathbf{v}_k)$  if there exists a continuous one-parameter family of bases to which these belong; in other words, if there exist continuous functions  $\mathbf{f}_j : [0, 1] \rightarrow V$ ,  $1 \leq j \leq k$ , such that  $\mathbf{f}_j(0) = \mathbf{v}_j$  and  $\mathbf{f}_j(1) = \mathbf{w}_j$  for each  $j$ , and such that  $(\mathbf{f}_1(t), \dots, \mathbf{f}_k(t))$  is a basis of  $V$  for each  $t$ ,  $0 \leq t \leq 1$ . Let  $\mathbf{f}_j(t) = \sum_{i=1}^k a_{ij}^i(t) \mathbf{v}_i$ , and let  $\Delta(t) = \det(a_{ij}^i(t))$ . Since the determinant of a matrix is a continuous function of its entries, and each  $a_{ij}^i$  is a continuous function,  $\Delta$  is a continuous function on  $[0, 1]$  which never vanishes, and



$\Delta(0) = 1$ , hence  $\Delta(1) > 0$ . Thus any basis to which  $(\mathbf{v}_1, \dots, \mathbf{v}_k)$  can be deformed has the same orientation. It is a good problem in linear algebra to show that in fact  $(\mathbf{v}_1, \dots, \mathbf{v}_k)$  can be deformed into any basis with the same orientation.

Consider the special case  $V = \mathbf{R}^k$ . The orientation to which the natural basis  $(\mathbf{e}_1, \dots, \mathbf{e}_k)$  belongs will be called the *usual* orientation of  $\mathbf{R}^k$ ; a basis with the same orientation as the natural basis is often said to be *positively oriented* or, when  $k = 3$ , to *define a right-handed system of coordinates*. If  $V$  is a proper subspace of  $\mathbf{R}^n$ , it is impossible to single out either orientation as the “usual” one.

If  $M$  is a  $k$ -manifold in  $\mathbf{R}^n$ , and  $\alpha$  is a local coordinate, then  $\alpha$  defines an orientation on each tangent space  $T_{\mathbf{p}}$ ,  $\mathbf{p} \in U_{\alpha}$ ; we refer, naturally enough, to the orientation of the basis  $(\alpha'(\mathbf{t})\mathbf{e}_1, \dots, \alpha'(\mathbf{t})\mathbf{e}_k)$ , where  $\mathbf{p} = \alpha(\mathbf{t})$ , of  $T_{\mathbf{p}}$ . This idea leads to the following definition:

**11.28 Definition.** We say that the manifold  $M$  is orientable if there is a collection  $\mathcal{O}$  of local coordinates for  $M$  such that:

- (a)  $\bigcup_{\alpha \in \mathcal{O}} U_{\alpha} = M$ ; and
- (b) if  $\alpha, \beta \in \mathcal{O}$  and  $U_{\alpha} \cap U_{\beta} \neq \emptyset$ , then  $\alpha$  and  $\beta$  define the same orientation on  $T_{\mathbf{p}}(M)$ , for each  $\mathbf{p} \in U_{\alpha} \cap U_{\beta}$ .

By an orientation of  $M$ , we will mean a collection  $\mathcal{O}$  satisfying (a), (b), and also:

- (c) if  $\beta$  is a local coordinate on  $M$  such that  $\beta$  and  $\alpha$  define the same orientation on  $T_{\mathbf{p}}$  for every  $\alpha \in \mathcal{O}$  such that  $U_{\alpha} \cap U_{\beta} \neq \emptyset$  and every  $\mathbf{p} \in U_{\alpha} \cap U_{\beta}$ , then  $\beta \in \mathcal{O}$ .

In other words, an orientation of  $M$  is a maximal collection of local coordinates satisfying (a) and (b). Finally, by an oriented manifold we mean a manifold  $M$ , together with an orientation of  $M$ .

This idea makes sense for manifolds of dimension  $\geq 1$ . We define an orientation on a zero-dimensional manifold  $M$  to be a function on  $M$  which takes the values  $\pm 1$ .

The next lemma shows that orientability and orientation can be defined without any explicit reference to the tangent space.

**11.29 Lemma.** The manifold  $M$  is orientable if and only if there is a collection  $\mathcal{O}$  of local coordinates for  $M$  such that:

- (a) for each  $\mathbf{p} \in M$ , there is an  $\alpha \in \mathcal{O}$  with  $\mathbf{p} \in U_{\alpha}$ ; and
- (b) if  $\alpha, \beta \in \mathcal{O}$  and  $U_{\alpha} \cap U_{\beta} \neq \emptyset$ , then  $\det \varphi'_{\alpha\beta} > 0$  in  $V_{\alpha\beta}$ .

**Proof.** Let  $\alpha$  and  $\beta$  be local coordinates, and suppose  $\mathbf{p} = \alpha(\mathbf{t}) = \beta(\mathbf{u})$ . Then, since the orientation of  $T_{\mathbf{p}}$  defined by  $\alpha$  is the orientation of the basis  $(A\mathbf{u}_1, \dots, A\mathbf{u}_k)$  for any positively oriented basis  $(\mathbf{u}_1, \dots, \mathbf{u}_k)$  of  $\mathbf{R}^k$ , where  $A = \alpha'(\mathbf{t})$ , and since  $\beta = \alpha \circ \varphi_{\alpha\beta}$  in a neighborhood of  $\mathbf{u}$ , we see by the chain rule that  $\alpha$  and  $\beta$  define the same orientation at  $\mathbf{p}$  if and only if  $\det \varphi'_{\alpha\beta}(\mathbf{u}) > 0$ . The equivalence of conditions (a) and (b) of the lemma with those of Definition 11.28 is now clear. ■

**11.30 Proposition.** *If  $M$  is an orientable manifold, there exist at least two distinct orientations of  $M$ ; if  $M$  is connected, there are exactly two.*

**Proof.** If  $\mathcal{O}$  is a collection of local coordinates satisfying (a) and (b) of Definition 11.28 (or Lemma 11.29), it can be enlarged to satisfy (c) of the definition in the obvious way. That is, there exists an orientation on any orientable manifold. If  $\alpha$  is a local coordinate on  $M$ , with domain  $V_\alpha$  a neighborhood of  $0 \in \mathbf{R}^k$  or  $\mathbf{R}^k_+$ , define  $\tilde{\alpha}$  with domain  $\tilde{V}_\alpha = \{\mathbf{t} : (-t_1, t_2, \dots, t_k) \in V_\alpha\}$ , by  $\tilde{\alpha}(t_1, \dots, t_k) = \alpha(-t_1, t_2, \dots, t_k)$  (here we use the special dispensation in Definition 11.2 for  $V_\alpha$  in the case  $k = 1$ ). It is not hard to see that  $U_\alpha = U_{\tilde{\alpha}}$  and that  $\alpha$  and  $\tilde{\alpha}$  define opposite orientations at each point of  $U_\alpha$ . Hence,  $\tilde{\mathcal{O}} = \{\tilde{\alpha} : \alpha \in \mathcal{O}\}$  is an orientation of  $M$  distinct from  $\mathcal{O}$ . We will call it the orientation opposite to  $\mathcal{O}$ . But suppose that  $M$  is connected, and that  $\mathcal{O}$  and  $\mathcal{O}'$  are orientations of  $M$ . For any  $\alpha \in \mathcal{O}$ ,  $\beta \in \mathcal{O}'$ , and  $\mathbf{p} \in U_\alpha \cap U_\beta$ , let  $\varepsilon_{\alpha\beta}(\mathbf{p}) = \text{sgn } \det \varphi'_{\alpha\beta}(\mathbf{u})$ , where  $\mathbf{p} = \beta(\mathbf{u})$ . Clearly,  $\varepsilon_{\alpha\beta}$  is a continuous function on  $U_\alpha \cap U_\beta$ . But if  $\gamma$  and  $\delta$  are any other local coordinates at  $\mathbf{p}$ , with  $\gamma \in \mathcal{O}$  and  $\delta \in \mathcal{O}'$ , then  $\varepsilon_{\alpha\beta}(\mathbf{p}) = \varepsilon_{\gamma\delta}(\mathbf{p})$ , since (as is easily seen)  $\varphi_{\gamma\delta} = \varphi_{\gamma\alpha} \circ \varphi_{\alpha\beta} \circ \varphi_{\beta\delta}$ . Thus we may define the continuous function  $\varepsilon : M \rightarrow \{-1, 1\}$  unambiguously by  $\varepsilon(\mathbf{p}) = \varepsilon_{\alpha\beta}(\mathbf{p})$ , where  $\alpha$  and  $\beta$  are local coordinates at  $\mathbf{p}$  with  $\alpha \in \mathcal{O}$  and  $\beta \in \mathcal{O}'$ . Since  $M$  is connected, either  $\varepsilon = +1$  on  $M$  or  $\varepsilon = -1$  on  $M$ . But this means that either  $\mathcal{O}' = \mathcal{O}$  or that  $\mathcal{O}' = \tilde{\mathcal{O}}$ . ■

**11.31 Lemma.** *If  $M$  is an orientable manifold, so is  $\partial M$ .*

**Proof.** We may assume that  $M$  has dimension  $k > 1$ . We proceed as in Lemma 11.8. For any local coordinate  $\alpha$  at  $\mathbf{p} \in \partial M$ , let  $\tilde{V}_\alpha = \{\mathbf{u} \in \mathbf{R}^{k-1} : (\mathbf{u}, 0) \in V_\alpha\}$  and  $\hat{\alpha}(\mathbf{u}) = \alpha(\mathbf{u}, 0)$ . Let  $\beta$  be another local coordinate at  $\mathbf{p}$ . Then  $V_{\alpha\beta}$  and  $V_{\beta\alpha}$  are open subsets of  $\mathbf{R}^k_+$ , so  $x_k(\varphi_{\alpha\beta}(\mathbf{t})) \geq 0$  for  $\mathbf{t} \in V_{\alpha\beta}$ , with equality holding if  $t_k = 0$ . It follows that  $D_k(x_k(\varphi_{\alpha\beta}))(\mathbf{t}) \geq 0$  if  $\mathbf{t} \in V_{\alpha\beta}$  and  $t_k = 0$ . Since

$$\det \varphi'_{\alpha\beta}(\mathbf{u}, 0) = \det \varphi'_{\hat{\alpha}\hat{\beta}}(\mathbf{u}) D_k(x_k(\varphi_{\alpha\beta}))(\mathbf{u}, 0) > 0,$$

it follows that  $\hat{\alpha}$  and  $\hat{\beta}$  induce the same orientation on  $T_{\mathbf{p}}(\partial M)$  whenever  $\alpha$  and  $\beta$  induce the same orientation on  $T_{\mathbf{p}}(M)$ . ■

**Convention.** If  $M$  is an oriented manifold of dimension  $k$ , we define the *positive* orientation of  $\partial M$  to be  $\hat{\mathcal{O}} = \{\hat{\alpha} : \alpha \in \mathcal{O}\}$  if  $k$  is even, and the

opposite orientation if  $k > 1$  is odd. (Here we are using the notation of Lemma 11.31.) If  $k = 1$ , we take the orientation on  $\partial M$  which is  $+1$  at points  $\mathbf{p}$  where  $V_\alpha \subset (-\infty, 0]$ , and  $-1$  at points  $\mathbf{p}$  where  $V_\alpha \subset [0, +\infty)$ , where  $\alpha$  is a local coordinate at  $\mathbf{p}$  belonging to the given orientation of  $M$ .

The reason for this mysterious convention will appear later.

Any  $n$ -manifold in  $\mathbf{R}^n$  is orientable; we take the *standard orientation* of  $M$  to be the collection of all local coordinates  $\alpha$  such that  $\det \alpha' > 0$  on  $V_\alpha$ . (In other words, the standard orientation is the orientation obtained by using the identity map as local coordinate.) But if  $k < n$ , a manifold of dimension  $k$  in  $\mathbf{R}^n$  need not be orientable. The basic example is the Möbius strip (Example 11.14 above). (Showing that the Möbius strip is not orientable is a good exercise.) Next we give a characterization of orientability for a *hypersurface*, as we call an  $(n-1)$ -manifold in  $\mathbf{R}^n$ .

**11.32 Proposition.** *If  $M$  is a manifold of dimension  $n-1$  in  $\mathbf{R}^n$ , then there is a one-to-one correspondence between orientations of  $M$  and continuous unit normal vector fields on  $M$ .*

**Proof.** Let  $\alpha$  be a local coordinate for  $M$ . We associate to each  $\mathbf{p} = \alpha(t) \in U_\alpha$  a unit normal vector  $\nu(\mathbf{p})$  as follows: since  $T_\mathbf{p}$  has dimension  $n-1$ , there are exactly two unit vectors orthogonal to it. Choose  $\nu(\mathbf{p})$  to be the one such that

$$(\nu(\mathbf{p}), \alpha'(t)\mathbf{e}_1, \dots, \alpha'(t)\mathbf{e}_{n-1})$$

is positively oriented. It is clear that the map  $\mathbf{p} \mapsto \nu(\mathbf{p})$  is continuous (in fact, smooth) on  $U_\alpha$ , and easy to see that we get the same  $\nu(\mathbf{p})$  if we replace  $\alpha$  by any local coordinate at  $\mathbf{p}$  which defines the same orientation on  $T_\mathbf{p}$ . Thus, if  $\mathcal{O}$  is an orientation, we get a globally defined continuous unit normal vector field  $\nu$ . Conversely, given such a normal vector field, we can define the associated orientation to consist of all local coordinates  $\alpha$  such that  $(\nu(\alpha(t)), \alpha'(t)\mathbf{e}_1, \dots, \alpha'(t)\mathbf{e}_{n-1})$  is positively oriented, for all  $t \in V_\alpha$ . It is straightforward to check that if  $\alpha$  and  $\beta$  satisfy this condition, then they define the same orientation on  $T_\mathbf{p}$  for any  $\mathbf{p} \in U_\alpha \cap U_\beta$ , so this class of local coordinates is indeed an orientation of  $M$ . ■

**Remark.** If  $M$  is a manifold of dimension  $n$  in  $\mathbf{R}^n$ , we consider  $M$  to be oriented with the standard orientation; then the associated positive orientation of  $\partial M$  (described by the convention above) is the one for which the associated unit normal vector always points out of  $M$ .

## 11.5 Exercises

1. Let  $0 < a < b$ , and let  $\mathbf{f} : \mathbf{R}^2 \rightarrow \mathbf{R}^3$  be the map defined in Example 11.13. Let  $\xi$  be an irrational real number, and define  $\mathbf{g} : \mathbf{R} \rightarrow \mathbf{R}^3$  by  $\mathbf{g}(t) = \mathbf{f}(t, \xi t)$ .

- (a) Show that  $\mathbf{g}$  is smooth and injective, with  $\mathbf{g}'(t) \neq \mathbf{0}$  for all  $t \in \mathbf{R}$ .
- (b) Show that  $\mathbf{g}(\mathbf{R})$  is a dense subset of the manifold of Example 11.13.
- (c) Show that  $\mathbf{g}(\mathbf{R})$  is not a manifold, but  $\mathbf{g}(J)$  is a 1-manifold in  $\mathbf{R}^3$  for any bounded interval  $J \subset \mathbf{R}$ .

2. Show that if  $M$  is a compact manifold in  $\mathbf{R}^n$ , then  $\partial M$  is also compact; if also  $M$  is  $n$ -dimensional, then  $\partial M = \text{bdry } M$ .

3. Let  $U$  be an open set in  $\mathbf{R}^n$ , and suppose that  $\Gamma = \text{bdry } U = \overline{U} \setminus U$  is an  $(n-1)$ -manifold, with  $\partial\Gamma = \emptyset$ . Show that  $M = U \cup \Gamma$  is an  $n$ -manifold. Give an example where  $\partial M \neq \Gamma$ .

4. Show that if  $M$  and  $N$  are  $k$ -manifolds in  $\mathbf{R}^n$ , it does not follow that  $M \cup N$  is a  $k$ -manifold. Give some sufficient conditions for  $M \cup N$  to be a manifold.

5. Prove Proposition 11.25. (You may find Proposition 11.22 helpful.)

6. Prove Proposition 11.26.

7. Let  $M$  be a  $k$ -dimensional manifold in  $\mathbf{R}^n$ , and  $g$  a smooth function on  $M$ . Show that if  $g$  has a local maximum (or minimum) at  $\mathbf{p} \in M$ , then  $dg_{\mathbf{p}} = 0$ .

8. Let  $\Omega$  be an open set in  $\mathbf{R}^n$ ,  $\mathbf{f} = (f^1, \dots, f^k) : \Omega \rightarrow \mathbf{R}^k$  a smooth map,  $M = \{\mathbf{p} \in \Omega : \mathbf{f}(\mathbf{p}) = \mathbf{0}\}$ , and suppose that  $\mathbf{f}'(\mathbf{p})$  has rank  $k$  at each  $\mathbf{p} \in M$ . Let  $G$  be a smooth function defined in an open set in  $\mathbf{R}^n$  containing  $M$ , and let  $g$  denote the restriction of  $G$  to  $M$ . Show that if  $g$  has a local maximum (or minimum) at  $\mathbf{p} \in M$ , then there exist scalars  $\lambda_1, \dots, \lambda_k$  such that  $dG_{\mathbf{p}} = \sum_{j=1}^k \lambda_j df_{\mathbf{p}}^j$ . (The numbers  $\lambda_j$  are known as *Lagrange multipliers*.)

9. Let  $A$  be a symmetric  $n \times n$  real matrix, and define  $G : \mathbf{R}^n \rightarrow \mathbf{R}$  by  $G(\mathbf{t}) = \langle A\mathbf{t}, \mathbf{t} \rangle$ ; let  $g : S^{n-1} \rightarrow \mathbf{R}$  be the restriction of  $G$  to the unit sphere  $S^{n-1} = \{\mathbf{t} \in \mathbf{R}^n : |\mathbf{t}| = 1\}$ . Use the last exercise to show that if  $g$  attains a maximum (or minimum) value at a point  $\mathbf{t}$ , then  $\mathbf{t}$  is an eigenvector for  $A$ , i.e., that there exists  $\lambda \in \mathbf{R}$  such that  $A\mathbf{t} = \lambda\mathbf{t}$ .

10. Let  $M$  and  $N$  be 2-manifolds in  $\mathbf{R}^3$ . Show that if the tangent spaces to  $M$  and  $N$  do not coincide at any point of  $M \cap N$ , then  $M \cap N$  is a 1-manifold, and the tangent line to  $M \cap N$  at  $\mathbf{p} \in M \cap N$  is the intersection of the tangent planes to  $M$  and  $N$  at  $\mathbf{p}$ .

11. Let  $M$  and  $N$  be manifolds in  $\mathbf{R}^n$ , with dimensions  $k$  and  $l$ , respectively, and suppose that  $k+l > n$ . Let  $\mathbf{p} \in M \cap N$ . We say that  $M$  and  $N$  *intersect transversally* at  $\mathbf{p}$  if  $\dim[T_{\mathbf{p}}(M) \cap T_{\mathbf{p}}(N)] = k + l - n$ . Show that if  $M$

and  $N$  intersect transversally at each point of  $M \cap N$ , then  $M \cap N$  is a manifold.

12. Identify the set  $M_n$  of  $n \times n$  matrices with  $\mathbf{R}^{n^2}$  in the obvious way, i.e., by listing the rows one after the other. Let  $O(n)$  be the set of all orthogonal matrices.

- (a) Show that  $O(n)$  is a compact manifold, without boundary, of dimension  $\frac{1}{2}n(n-1)$ .
- (b) Show that  $T$  is a tangent vector to  $O(n)$  at  $I$  (the identity matrix) if and only if  $T$  is skew-symmetric, i.e.,  $T^t = -T$ .

13. Show that the Möbius strip is not orientable.

14. Let  $M$  be a compact  $n$ -manifold in  $\mathbf{R}^n$ , and let  $\mathbf{p} \in \partial M$ . If  $\nu$  is a unit normal vector to  $\partial M$  at  $\mathbf{p}$ , show that one of the following alternatives holds:

- (i) there exists  $\delta > 0$  such that  $\mathbf{p} + t\nu \notin M$  for all  $0 < t < \delta$ ; or
- (ii) there exists  $\delta > 0$  such that  $\mathbf{p} + t\nu \in M$  for all  $0 < t < \delta$ .

We say  $\nu$  points out of  $M$  if (i) holds. Verify the remark made after Proposition 11.32.

## 11.6 Notes

Mathematicians first began the systematic study of curves and surfaces with the tools of analysis in the latter part of the eighteenth century. The pioneering figures in this new subject of differential geometry were above all Monge and (a little later) Gauss. The daring step of imagining higher-dimensional objects with somehow the essential character of surfaces, i.e., manifolds of dimension greater than two, was first taken by Riemann. Riemann had previously, in his investigations of complex function theory, come to the idea of an abstract surface, i.e., a surface not presented as a subset of  $\mathbf{R}^3$  (or any  $\mathbf{R}^n$  for that matter, but that issue did not arise).

It is not uncommon in the literature to see the word “manifold” used only in the sense of manifold with empty boundary; in such cases, what we have called simply a manifold is referred to explicitly as a *manifold with boundary*, or sometimes *bordered manifold*.

It can be shown that a connected 1-manifold is diffeomorphic to either the circle  $S^1 = \{(u, v) \in \mathbf{R}^2 : u^2 + v^2 = 1\}$  or to an interval  $J \subset \mathbf{R}$ . This is an elementary fact, but hardly trivial. The reader is invited to find a simple proof. A proof can be found in [9]. In particular, any compact connected 1-manifold with empty boundary is diffeomorphic to the circle. There is also a classification of compact 2-manifolds with empty boundary;

any such surface which is orientable is diffeomorphic to the 2-sphere, with some number  $g \geq 0$  of "handles" attached (attaching one handle to a sphere produces a torus, for example). Classification of manifolds of higher dimension is much more difficult (in general, impossible for dimension  $\geq 4$ ) and is a lively area of research.

# 12

## Multilinear Algebra

### 12.1 Vectors and Tensors

Throughout this chapter,  $V$  will denote a vector space of dimension  $n$  over the reals.

**12.1 Definition.** We denote by  $V^*$  the space  $\mathcal{L}(V, \mathbf{R})$  of all linear functionals on  $V$ . The elements of  $V^*$  are sometimes referred to as covectors.

Thus,  $\alpha \in V^*$  means that  $\alpha : V \rightarrow \mathbf{R}$ , and  $\alpha(av + bw) = a\alpha(v) + b\alpha(w)$  for every  $a, b \in \mathbf{R}$  and every  $v, w \in V$ . The space  $V^*$  is called the *dual space*, or *conjugate space*, of  $V$ ; it is itself a vector space, with the natural operations of functions (to wit,  $(\alpha + \beta)(v) = \alpha(v) + \beta(v)$ , etc.).

When  $V = \mathbf{R}^n$ , it is common practice to write the elements of  $V$  as column vectors ( $n \times 1$  matrices). Let  $e_j$  be the  $j$ th standard basis vector, i.e., the column having 1 in the  $j$ th place and zeros elsewhere. Given any  $\alpha \in \mathbf{R}^{n*}$  we put  $a_j = \alpha(e_j)$ , and find that for any  $x = \sum x^j e_j$ , we have  $\alpha(x) = \sum x^j \alpha(e_j) = \sum a_j x^j$ . As we saw in Chapter 8, if we write the  $n$ -tuple  $a = (a_j)$  as a row vector ( $1 \times n$  matrix), then  $\alpha(x)$  is just the matrix product  $ax$ . Let  $\tilde{e}^k$  denote the row vector with 1 in the  $k$ th place and 0 elsewhere; we see that  $\tilde{e}^1, \dots, \tilde{e}^n$  form a basis for  $\mathbf{R}^{n*}$  and that  $\tilde{e}^k(e_j) = \delta_j^k$ , where  $\delta_j^k$  is the Kronecker delta, standing for 1 if  $j = k$  and for 0 if  $j \neq k$ . We generalize this observation.

**12.2 Theorem.** If  $(u_1, \dots, u_n)$  is a basis of  $V$ , then there exists a basis  $(\tilde{u}^1, \dots, \tilde{u}^n)$  of  $V^*$ , called the dual basis to  $(u_1, \dots, u_n)$ , with  $\tilde{u}^j(u_k) = \delta_k^j$ .

**Proof.** Define the elements  $\tilde{u}^1, \dots, \tilde{u}^n$  of  $V^*$  as follows: if  $\mathbf{x} = \sum_1^n x^j \mathbf{u}_j$ , then  $\tilde{u}^k(\mathbf{x}) = x^k$  for  $k = 1, \dots, n$ . Evidently, each  $\tilde{u}^k$  is a well-defined linear function on  $V$ , and  $\tilde{u}^k(\mathbf{u}_j) = \delta_j^k$ . Suppose that  $\sum a_j \tilde{u}^j = 0$ ; then  $(\sum a_j \tilde{u}^j)(\mathbf{u}_k) = 0$ , which gives  $a_k = 0$  for every  $k$ . Thus  $\tilde{u}^1, \dots, \tilde{u}^n$  are linearly independent. If  $\alpha \in V^*$ , let  $a_j = \alpha(\mathbf{u}_j)$ , and observe that  $\alpha(\mathbf{x}) = \sum x^j \alpha(\mathbf{u}_j) = \sum a_j x^j = (\sum a_j \tilde{u}^j)(\mathbf{x})$  for every  $\mathbf{x} \in V$ , i.e., that  $\alpha = \sum_1^n a_j \tilde{u}^j$ . Thus  $(\tilde{u}^1, \dots, \tilde{u}^n)$  span  $V^*$ , and thus form a basis of  $V^*$ . ■

**12.3 Corollary.** *The spaces  $V$  and  $V^*$  have the same dimension.*

**12.4 Corollary.** *If  $\mathbf{x} \in V$ , and  $\mathbf{x} \neq 0$ , there exists  $\alpha \in V^*$  such that  $\alpha(\mathbf{x}) \neq 0$ .*

**12.5 Proposition.** *There is a natural isomorphism of  $V$  onto  $V^{**}$ .*

**Proof.** By "natural," we mean one that does not depend on the choice of a basis; the isomorphism of  $V$  and  $V^*$  found above was obtained by first choosing a basis. For each  $\mathbf{x} \in V$ , define  $f(\mathbf{x}) : V^* \rightarrow \mathbf{R}$  by the rule  $f(\mathbf{x})(\alpha) = \alpha(\mathbf{x})$ . That  $f(\mathbf{x})$  is a linear function on  $V^*$  is simply the definition of addition and scalar multiplication in  $V^*$ ; thus  $f(\mathbf{x}) \in V^{**}$ . It's easy to see that  $f : V \rightarrow V^{**}$  is linear:

$$f(a\mathbf{x} + b\mathbf{y})(\alpha) = \alpha(a\mathbf{x} + b\mathbf{y}) = a\alpha(\mathbf{x}) + b\alpha(\mathbf{y}) = af(\mathbf{x})(\alpha) + bf(\mathbf{y})(\alpha),$$

so  $f(a\mathbf{x} + b\mathbf{y}) = af(\mathbf{x}) + bf(\mathbf{y})$ . If  $f(\mathbf{x}) = 0$ , then  $\alpha(\mathbf{x}) = 0$  for every  $\alpha \in V^*$ , so  $\mathbf{x} = 0$  by the last corollary; thus,  $f$  is injective, and since  $\dim V^{**} = \dim V$  by the previous corollary,  $f$  is also surjective, i.e., an isomorphism. ■

**12.6 Corollary.** *If  $(\alpha^1, \dots, \alpha^n)$  is a basis of  $V^*$ , then there exists a basis  $(\mathbf{u}_1, \dots, \mathbf{u}_n)$  of  $V$  such that  $\alpha^j(\mathbf{u}_k) = \delta_k^j$ , i.e., such that  $\alpha^j = \tilde{u}^j$ .*

**12.7 Definition.** *Let  $r$  be a positive integer. By a covariant tensor of rank  $r$  on  $V$ , we mean a map  $\alpha : V^r \rightarrow \mathbf{R}$  which is linear in each variable separately. We denote the set of covariant tensors of rank  $r$  on  $V$  by the symbol  $T^r$ , or  $T^r(V^*)$  if there is more than one vector space under discussion.*

Thus, we ask that for each  $j$ ,  $1 \leq j \leq r$ , for any real  $t$ , and any set of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_j, \mathbf{v}'_j, \dots, \mathbf{v}_r$ , we have

$$\alpha(\mathbf{v}_1, \dots, \mathbf{v}_j + \mathbf{v}'_j, \dots, \mathbf{v}_r) = \alpha(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_r) + \alpha(\mathbf{v}_1, \dots, \mathbf{v}'_j, \dots, \mathbf{v}_r)$$

and

$$\alpha(\mathbf{v}_1, \dots, t\mathbf{v}_j, \dots, \mathbf{v}_r) = t\alpha(\mathbf{v}_1, \dots, \mathbf{v}_r).$$

We see that a covariant tensor of rank 1 is simply an element of  $V^*$ . A covariant tensor of rank 2 is a bilinear function on  $V \times V$ ; an example of



such a tensor is an *inner product* on  $V$ , i.e., an element  $\gamma \in T^2$  with the additional properties that  $\gamma(\mathbf{x}, \mathbf{y}) = \gamma(\mathbf{y}, \mathbf{x})$  for all  $\mathbf{x}, \mathbf{y} \in V$ , and  $\gamma(\mathbf{x}, \mathbf{x}) \geq 0$ , equality holding only if  $\mathbf{x} = \mathbf{0}$ .

There is also a notion of *contravariant tensor*. In this book, we will only consider covariant tensors, and will often omit the adjective covariant.

It is clear that  $T^r$  is a vector space, with the natural operations of functions. There is also a multiplication of tensors, in a natural way. If  $\alpha$  and  $\beta$  are elements of  $V^*$ , we can combine them to get a tensor of rank 2 as follows: let  $\gamma(\mathbf{v}, \mathbf{w}) = \alpha(\mathbf{v})\beta(\mathbf{w})$ ; we call  $\gamma$  the *tensor product* of  $\alpha$  and  $\beta$ , and write  $\gamma = \alpha \otimes \beta$ . A similar construction is possible with tensors of any rank.

**12.8 Definition.** If  $\alpha$  is a tensor of rank  $r$  and  $\beta$  is a tensor of rank  $s$ , we define their tensor product  $\alpha \otimes \beta$  by the formula

$$\alpha \otimes \beta(\mathbf{v}_1, \dots, \mathbf{v}_{r+s}) = \alpha(\mathbf{v}_1, \dots, \mathbf{v}_r)\beta(\mathbf{v}_{r+1}, \dots, \mathbf{v}_{r+s}).$$

It is clear that  $\alpha \otimes \beta$  is a tensor of rank  $r + s$ . It is quite obvious that the tensor product is associative, i.e., that  $(\alpha \otimes \beta) \otimes \gamma = \alpha \otimes (\beta \otimes \gamma)$  for any three tensors  $\alpha, \beta, \gamma$  of any ranks.

Another notation for the space  $T^r$  of tensors of rank  $r$  is

$$T^r = \underbrace{V^* \otimes \dots \otimes V^*}_{r \text{ times}}$$

but this does not mean that every tensor of rank  $r$  is the product of  $r$  tensors of rank 1 (linear functionals).

**12.9 Theorem.** If  $(\mathbf{u}_1, \dots, \mathbf{u}_n)$  is a basis for  $V$ , then the set

$$\{\tilde{\mathbf{u}}^{j_1} \otimes \dots \otimes \tilde{\mathbf{u}}^{j_r} : 1 \leq j_1 \leq n, \dots, 1 \leq j_r \leq n\} \quad (12.1)$$

is a basis for  $T^r$ . In particular, the dimension of  $T^r$  is  $n^r$ .

**Proof.** It is immediate from the definition of the tensor product that

$$\tilde{\mathbf{u}}^{j_1} \otimes \dots \otimes \tilde{\mathbf{u}}^{j_r}(\mathbf{u}_{k_1}, \dots, \mathbf{u}_{k_r}) = \delta_{k_1 \dots k_r}^{j_1 \dots j_r},$$

where the multi-index Kronecker delta is defined by

$$\delta_{k_1 \dots k_r}^{j_1 \dots j_r} = \delta_{k_1}^{j_1} \delta_{k_2}^{j_2} \dots \delta_{k_r}^{j_r} = \begin{cases} 1 & \text{if } (j_1, \dots, j_r) = (k_1, \dots, k_r); \\ 0 & \text{otherwise.} \end{cases}$$

If  $c_{j_1 \dots j_r}$  are scalars such that

$$\sum c_{j_1 \dots j_r} \tilde{\mathbf{u}}^{j_1} \otimes \dots \otimes \tilde{\mathbf{u}}^{j_r} = \mathbf{0},$$

it follows that  $c_{j_1 \dots j_r} = 0$  for all  $j_1, \dots, j_r$ . Thus the set of tensors (12.1) is linearly independent.

Let  $\alpha$  be a tensor of rank  $r$  and define, for each  $r$ -tuple  $j_1, \dots, j_r$ ,

$$a_{j_1 \dots j_r} = \alpha(\mathbf{u}_{j_1}, \dots, \mathbf{u}_{j_r}).$$

Then given any vectors  $\mathbf{v}_1, \dots, \mathbf{v}_r \in V$ , we have, for  $j = 1, \dots, r$ , an expansion  $\mathbf{v}_j = \sum_{k=1}^n v_j^k \mathbf{u}_k$ , where  $v_j^k = \tilde{\mathbf{u}}^k(\mathbf{v}_j)$ , from which multilinearity gives

$$\begin{aligned} \alpha(\mathbf{v}_1, \dots, \mathbf{v}_r) &= \sum v_1^{j_1} v_2^{j_2} \dots v_r^{j_r} \alpha(\mathbf{u}_{j_1}, \dots, \mathbf{u}_{j_r}) \\ &= \sum a_{j_1 \dots j_r} \tilde{\mathbf{u}}^{j_1} \otimes \dots \otimes \tilde{\mathbf{u}}^{j_r}(\mathbf{v}_1, \dots, \mathbf{v}_r), \end{aligned}$$

where the sums are taken over all possible  $r$ -tuples  $(j_1, \dots, j_r)$ . This shows that the set (12.1) spans the space of tensors of rank  $r$ . ■

It is convenient to extend the concept of tensor to the case  $r = 0$ : a tensor of rank 0 will mean a scalar. Thus tensors of rank 0 form a space of dimension  $n^0$ . The tensor product of a scalar with a tensor will be interpreted as the usual product.

## 12.2 Alternating Tensors

**12.10 Definition.** We say that  $\alpha \in T^r$  is alternating if  $\alpha(\mathbf{v}_1, \dots, \mathbf{v}_r) = 0$  whenever there exist  $j \neq k$  such that  $\mathbf{v}_j = \mathbf{v}_k$ .

We denote the set of all alternating tensors of rank  $r$  by the symbol  $\Lambda^r$ , or  $\Lambda^r(V^*)$  if there is any chance of confusion with tensors defined on some other vector space.

This definition allows every tensor of rank 0 or 1 to be classified as alternating. If  $\alpha$  and  $\beta$  are tensors of rank 1 (i.e., elements of  $V^*$ ), the 2-tensor  $\alpha \otimes \beta - \beta \otimes \alpha$  is alternating.

**12.11 Proposition.** If  $\alpha \in \Lambda^r$  and  $1 \leq j < k \leq r$ , then for all  $\mathbf{v}_1, \dots, \mathbf{v}_r$ , we have

$$\alpha(\mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{v}_k, \mathbf{v}_{j+1}, \dots, \mathbf{v}_{k-1}, \mathbf{v}_j, \mathbf{v}_{k+1}, \dots, \mathbf{v}_r) = -\alpha(\mathbf{v}_1, \dots, \mathbf{v}_r).$$

**Proof.** It is obviously sufficient to establish the formula in the case  $r = 2$ . We have, using linearity in the second variable,

$$\begin{aligned} \alpha(\mathbf{v}, \mathbf{w}) &= \alpha(\mathbf{v}, \mathbf{v}) + \alpha(\mathbf{v}, \mathbf{w}) = \alpha(\mathbf{v}, \mathbf{v} + \mathbf{w}), \\ \alpha(\mathbf{w}, \mathbf{v}) &= \alpha(\mathbf{w}, \mathbf{v}) + \alpha(\mathbf{w}, \mathbf{w}) = \alpha(\mathbf{w}, \mathbf{v} + \mathbf{w}), \end{aligned}$$

and hence, by linearity in the first variable,

$$\alpha(\mathbf{v}, \mathbf{w}) + \alpha(\mathbf{w}, \mathbf{v}) = \alpha(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) = 0,$$

whence  $\alpha(\mathbf{v}, \mathbf{w}) = -\alpha(\mathbf{w}, \mathbf{v})$ . ■

We remark that if a tensor has the “antisymmetry” property described in Proposition 12.11, it is evidently alternating; for if  $x \in \mathbf{R}$  and  $x = -x$ , then  $x = 0$ . If we were to consider vector spaces over a field of characteristic 2, however, the properties would not be equivalent.

**12.12 Proposition.** *If  $\alpha \in \Lambda^r$ , and  $\mathbf{v}_1, \dots, \mathbf{v}_r$  are linearly dependent elements of  $V$ , then  $\alpha(\mathbf{v}_1, \dots, \mathbf{v}_r) = 0$ .*

**Proof.** There exist scalars  $c^1, \dots, c^r$ , not all 0, such that  $\sum_{j=1}^r c^j \mathbf{v}_j = \mathbf{0}$ . Let  $k = \max\{j : c^j \neq 0\}$ . If  $k = 1$ , then  $\mathbf{v}_1 = \mathbf{0}$ , and the desired conclusion follows since  $\alpha$  is multilinear. If  $k > 1$ , we can write  $\mathbf{v}_k = \sum_{j=1}^{k-1} b^j \mathbf{v}_j$ , where  $b^j = -c^j/c^k$ , and hence

$$\begin{aligned} \alpha(\mathbf{v}_1, \dots, \mathbf{v}_r) &= \alpha(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \sum_{j=1}^{k-1} b^j \mathbf{v}_j, \mathbf{v}_{k+1}, \dots, \mathbf{v}_r) \\ &= \sum_{j=1}^{k-1} b^j \alpha(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_j, \mathbf{v}_{k+1}, \dots, \mathbf{v}_r) \\ &= 0, \end{aligned}$$

since  $\alpha$  is alternating. ■

**12.13 Corollary.** *If  $r > n$ , then  $\Lambda^r = (0)$ .*

**Proof.** If  $r > n$ , any set of  $r$  vectors in  $V$  is linearly dependent. ■

Let  $S_r$  denote the set of permutations on  $r$  letters, i.e., the set of all one-one mappings of  $\{1, 2, \dots, r\}$  onto itself. It is a familiar fact that  $S_r$  is a group with  $r!$  elements.

**12.14 Definition.** *Let  $\sigma \in S_r$ . We put*

$$\varepsilon(\sigma) = \prod_{j < k} \operatorname{sgn}(\sigma(k) - \sigma(j)),$$

*and call  $\varepsilon(\sigma)$  the sign of  $\sigma$ .*

We recall that  $\operatorname{sgn} x$  is the sign of  $x$ , i.e.,

$$\operatorname{sgn} x = \begin{cases} x/|x| & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

**12.15 Proposition.** *Suppose  $\tau \in S_r$  is a transposition, i.e., there exist  $1 \leq i < j \leq r$  such that  $\tau(i) = j$ ,  $\tau(j) = i$ , and  $\tau(k) = k$  unless  $k = i$  or  $j$ . Then  $\varepsilon(\tau) = -1$ .*

**Proof.** We note that  $\tau(k) < \tau(l)$  whenever  $k < l$ , except in the cases where  $k = i$  and  $l \leq j$ , or  $l = j$  and  $i \leq k$  (or both); there are  $2(j - i) + 1$  such cases, so  $\varepsilon(\tau) = (-1)^{2j-1} = -1$ . ■

**12.16 Proposition.** For any  $\sigma, \tau \in S_r$  we have  $\varepsilon(\sigma\tau) = \varepsilon(\sigma)\varepsilon(\tau)$ .

**Proof.** We observe that

$$\begin{aligned} \varepsilon(\sigma\tau) &= \prod_{j < k} \operatorname{sgn} \left( \frac{(\sigma\tau)(k) - (\sigma\tau)(j)}{k - j} \right) \\ &= \prod_{j < k} \operatorname{sgn} \left( \frac{\sigma(\tau(k)) - \sigma(\tau(j))}{\tau(k) - \tau(j)} \frac{\tau(k) - \tau(j)}{k - j} \right) \\ &= \prod_{j < k} \operatorname{sgn} \left( \frac{\sigma(\tau(k)) - \sigma(\tau(j))}{\tau(k) - \tau(j)} \right) \prod_{j < k} \operatorname{sgn} \left( \frac{\tau(k) - \tau(j)}{k - j} \right) \\ &= \prod_{l < m} \operatorname{sgn} \left( \frac{\sigma(m) - \sigma(l)}{m - l} \right) \prod_{j < k} \operatorname{sgn} \left( \frac{\tau(k) - \tau(j)}{k - j} \right) \\ &= \varepsilon(\sigma)\varepsilon(\tau), \end{aligned}$$

which was to be proved. ■

Suppose  $\sigma \in S_r$ , and  $\sigma = \tau_1 \cdots \tau_k$ , where each  $\tau_j$  is a transposition. Then by Propositions 12.15 and 12.16, we have  $\varepsilon(\sigma) = (-1)^k$ . Now any permutation can be expressed as a product of transpositions, and in many different ways; according to what we have just observed, it always takes an even number, or it always takes an odd number. This is not an obvious fact.

**12.17 Definition.** Let  $\alpha$  be a tensor of rank  $r$ , and  $\sigma \in S_r$ . We define a new tensor  ${}^\sigma\alpha$  of rank  $r$  by the formula

$${}^\sigma\alpha(\mathbf{v}_1, \dots, \mathbf{v}_r) = \alpha(\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(r)})$$

for all  $\mathbf{v}_1, \dots, \mathbf{v}_r \in V$ .

Evidently, the map  $\alpha \mapsto {}^\sigma\alpha$  is a linear mapping of  $T^r$  onto itself, for any  $\sigma \in S_r$ , and it is easy to see that  ${}^{\sigma\tau}\alpha = {}^\sigma({}^\tau\alpha)$  for any  $\sigma, \tau \in S_r$ . We can restate Proposition 12.11 above in the form: if  $\alpha \in \Lambda^r$ , ( $r > 1$ ), and  $\tau \in S_r$  is a transposition, then  ${}^\tau\alpha = -\alpha$ . Hence, if  $\sigma \in S_r$ , and  $\sigma = \tau_1 \cdots \tau_k$  where each  $\tau_j$  is a transposition, we can conclude that  ${}^\sigma\alpha = (-1)^k\alpha$ . In other words:

**12.18 Proposition.** If  $\alpha$  is an alternating tensor, and  $\sigma \in S_r$ , then  ${}^\sigma\alpha = \varepsilon(\sigma)\alpha$ .

**12.19 Definition.** If  $\alpha$  is a tensor of rank  $r$ , we define

$$A\alpha = \sum_{\sigma \in S_r} \varepsilon(\sigma) \sigma \alpha;$$

we call  $A\alpha$  the alternation of  $\alpha$ , and refer to  $A$  as the alternation operator.

For instance, with  $V = \mathbf{R}^n$ , if  $\alpha = \tilde{\mathbf{e}}^j \otimes \tilde{\mathbf{e}}^k$ , then

$$A\alpha(\mathbf{v}, \mathbf{w}) = \alpha(\mathbf{v}, \mathbf{w}) - \alpha(\mathbf{w}, \mathbf{v}) = v^j w^k - v^k w^j,$$

i.e.,  $A(\tilde{\mathbf{e}}^j \otimes \tilde{\mathbf{e}}^k) = \tilde{\mathbf{e}}^j \otimes \tilde{\mathbf{e}}^k - \tilde{\mathbf{e}}^k \otimes \tilde{\mathbf{e}}^j$ . Similarly, we compute that

$$\begin{aligned} A(\tilde{\mathbf{e}}^1 \otimes \tilde{\mathbf{e}}^2 \otimes \tilde{\mathbf{e}}^3)(\mathbf{u}, \mathbf{v}, \mathbf{w}) &= u^1 v^2 w^3 - u^1 v^3 w^2 + u^2 v^3 w^1 \\ &\quad - u^2 v^1 w^3 + u^3 v^1 w^2 - u^3 v^2 w^1, \end{aligned}$$

which is the  $3 \times 3$  determinant made from the first three coordinates of  $\mathbf{u}$ ,  $\mathbf{v}$ , and  $\mathbf{w}$ .

The next proposition sums up the basic properties of the alternation operator; up to a constant factor, it is a projection of  $T^r$  onto  $\Lambda^r$ .

**12.20 Proposition.** For any  $\alpha \in T^r$ , and  $\tau \in S_r$ , we have:

- (a)  $\tau(A\alpha) = \varepsilon(\tau)A\alpha$ ;
- (b)  $A\alpha$  is an alternating tensor;
- (c) if  $\alpha$  is alternating, then  $A\alpha = r!\alpha$ ; and
- (d)  $A^\tau \alpha = \varepsilon(\tau)A\alpha$ .

**Proof.** For any  $\alpha \in T^r$ ,  $\tau \in S_r$ , we have

$$\begin{aligned} \tau(A\alpha) &= \tau \left( \sum_{\sigma \in S_r} \varepsilon(\sigma) \sigma \alpha \right) = \sum_{\sigma \in S_r} \varepsilon(\tau) \varepsilon(\tau \sigma) \tau \sigma \alpha \\ &= \varepsilon(\tau) \sum_{\sigma' \in S_r} \varepsilon(\sigma') \sigma' \alpha = \varepsilon(\tau) A\alpha. \end{aligned}$$

Thus (a) is proved. We observed earlier that (a) implies (b). If  $\alpha$  itself is alternating, then  $A\alpha = r!\alpha$ , since by Proposition 12.18 every one of the  $r!$  terms in the defining sum is equal to  $\alpha$ . Finally, (d) is proved in exactly the same way as (a).  $\blacksquare$

**Notation.** We will denote by a single capital letter a finite sequence (ordered  $r$ -tuple) of integers from  $\{1, \dots, n\}$ ; thus,  $I = (i_1, \dots, i_r)$ ; we put  $|I| = r$ , the length of  $I$ . We say that  $I$  is *increasing* if  $i_1 < i_2 < \dots < i_r$ .

**12.21 Definition.** Let  $(u_1, \dots, u_n)$  be a basis for  $V$ . We put

$$\tilde{u}^I = A(\tilde{u}^{i_1} \otimes \cdots \otimes \tilde{u}^{i_r}),$$

for each  $r$ -tuple  $I = (i_1, \dots, i_r)$ .

**12.22 Proposition.** The set

$$\{\tilde{u}^I : |I| = r, I \text{ increasing}\} \quad (12.2)$$

forms a basis for  $\Lambda^r(V^*)$ .

**Proof.** Observe that if  $I = (i_1, \dots, i_r)$  and  $J = (j_1, \dots, j_r)$  are increasing sequences, then

$$\begin{aligned} \tilde{u}^I(u_{j_1}, \dots, u_{j_r}) &= A(\tilde{u}^{i_1} \otimes \cdots \otimes \tilde{u}^{i_r})(u_{j_1}, \dots, u_{j_r}) \\ &= \sum_{\sigma \in S_r} \varepsilon(\sigma) \tilde{u}^{i_1} \otimes \cdots \otimes \tilde{u}^{i_r}(u_{j_{\sigma(1)}}, \dots, u_{j_{\sigma(r)}}) \\ &= \delta_J^I, \end{aligned}$$

the Kronecker delta, and hence, using Proposition 12.20, that for any  $r$ -tuples  $I$  and  $J$ , not necessarily increasing,

$$\tilde{u}^I(u_{j_1}, \dots, u_{j_r}) = \varepsilon_J^I,$$

where  $\varepsilon_J^I$ , the “Kronecker epsilon,” is defined to be 0 unless the sequence  $I$  is a rearrangement of the sequence  $J$ , and to be  $\varepsilon(\sigma)$ , if the permutation  $\sigma$  transforms  $I$  to  $J$ . It is now easy to see that the set (12.2) is linearly independent, for if  $\alpha = \sum c_I \tilde{u}^I$ , then  $\alpha(u_{j_1}, \dots, u_{j_r}) = c_J$ , for each increasing  $J = (j_1, \dots, j_r)$ ; it follows that  $\alpha = 0$  only if every  $c_J = 0$ . Using Theorem 12.9 and Proposition 12.20, we see that  $\{\tilde{u}^I : \text{all } r\text{-tuples } I\}$  spans  $\Lambda^r$ . By Proposition 12.20 (d), we see that  $\tilde{u}^I = 0$  unless  $I$  contains no index twice, i.e., unless  $I$  is a permutation of an increasing  $r$ -tuple; for if  $\alpha = \tilde{u}^I$  with  $i_j = i_k$  for some  $j \neq k$ , then  $\tau\alpha = \alpha$  for a transposition  $\tau$ , so  $A\alpha = \varepsilon(\tau)A\alpha = -A\alpha$ . The same proposition shows that if  $I = \sigma(J)$ , then  $\tilde{u}^I = \varepsilon(\sigma)\tilde{u}^J$ . Thus the set (12.2) spans  $\Lambda^r$ , and hence is a basis. ■

**12.23 Corollary.** The dimension of  $\Lambda^r$  is  $\binom{n}{r}$ , for  $r = 0, \dots, n$ .

**Proof.** There is a one-to-one correspondence between increasing sequences of length  $r$  in  $\{1, \dots, n\}$  and subsets of  $\{1, \dots, n\}$  with cardinality  $r$ ; the number of such subsets is  $\binom{n}{r}$ , so this is the dimension of  $\Lambda^r$  by the last proposition. ■

In particular, we see that  $\Lambda^n$  is one-dimensional, and is spanned by the alternating tensor  $\tilde{u}^{12 \cdots n}$ .

## 12.3 The Exterior Product

We next introduce a multiplication of alternating tensors; it is easy to see that the tensor product of alternating tensors is usually *not* alternating.

**12.24 Definition.** If  $\alpha \in \Lambda^r$  and  $\beta \in \Lambda^s$ , we define

$$\alpha \wedge \beta = A \left( \frac{\alpha}{r!} \otimes \frac{\beta}{s!} \right),$$

so  $\alpha \wedge \beta \in \Lambda^{r+s}$ . We refer to  $\alpha \wedge \beta$  as the exterior product or wedge product of  $\alpha$  and  $\beta$ .

If  $r = 0$  or  $s = 0$ , we see that  $\alpha \wedge \beta$  is just the usual product of a scalar and a tensor, since  $A\alpha = r!\alpha$ ,  $A\beta = s!\beta$ . We will soon see that the factors  $1/r!$  and  $1/s!$  serve another useful purpose.

**12.25 Proposition.** If  $\alpha$  is a tensor such that  $A\alpha = 0$ , then

$$A(\alpha \otimes \beta) = A(\beta \otimes \alpha) = 0$$

for every tensor  $\beta$ .

**Proof.** Let  $r$  be the rank of  $\alpha$  and let  $s$  be the rank of  $\beta$ . Let

$$T = \{\tau \in S_{r+s} : \tau(j) = j \text{ for } j = r+1, \dots, r+s\},$$

so  $T$  is a subgroup of  $S_{r+s}$ , isomorphic in an obvious way to  $S_r$ . Observe that for  $\tau \in T$ ,  $\tau(\alpha \otimes \beta) = \tau\alpha \otimes \beta$ . The left cosets of  $T$  partition  $S_{r+s}$ , so we may choose  $\sigma_1, \dots, \sigma_p$  such that

$$\sigma_j T \cap \sigma_k T = \emptyset \text{ if } j \neq k \quad \text{and} \quad S_{r+s} = \bigcup_{j=1}^p \sigma_j T.$$

Then

$$\begin{aligned} A(\alpha \otimes \beta) &= \sum_{\sigma \in S_{r+s}} \varepsilon(\sigma) \sigma(\alpha \otimes \beta) \\ &= \sum_{j=1}^p \sum_{\tau \in T} \varepsilon(\sigma_j \tau) \sigma_j \tau(\alpha \otimes \beta) \\ &= \sum_{j=1}^p \varepsilon(\sigma_j)^{\sigma_j} \left( \sum_{\tau \in T} \varepsilon(\tau) \tau(\alpha \otimes \beta) \right) \\ &= \sum_{j=1}^p \varepsilon(\sigma_j)^{\sigma_j} \left( \sum_{\tau \in S_r} \varepsilon(\tau) \tau\alpha \otimes \beta \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^p \varepsilon(\sigma_j)^{\sigma_j} \left( \left[ \sum_{\tau \in S_r} \varepsilon(\tau) \tau \alpha \right] \otimes \beta \right) \\
&= \sum_{j=1}^p \varepsilon(\sigma_j)^{\sigma_j} (A\alpha \otimes \beta) = 0,
\end{aligned}$$

as was to be proved.

To see that also  $A(\beta \otimes \alpha) = 0$ , we observe that  $\beta \otimes \alpha = {}^\sigma(\alpha \otimes \beta)$ , where  $\sigma$  is the permutation which sends  $(1, \dots, r+s)$  to  $(r+1, \dots, r+s, 1, \dots, r)$ , and recall that  $A {}^\sigma \gamma = \varepsilon(\sigma) A \gamma$  for any tensor  $\gamma$  of rank  $t$  and  $\sigma \in S_t$  (Proposition 12.20); thus  $A(\alpha \otimes \beta) = 0$  implies that  $A(\beta \otimes \alpha) = 0$ . ■

**12.26 Corollary.** Suppose that  $\alpha, \alpha'$  are tensors of rank  $r$  such that  $A\alpha = A\alpha'$ . Then  $A(\alpha \otimes \beta) = A(\alpha' \otimes \beta)$  for every tensor  $\beta$ .

**Proof.** We have  $A(\alpha \otimes \beta) - A(\alpha' \otimes \beta) = A((\alpha - \alpha') \otimes \beta) = 0$ , since  $A(\alpha - \alpha') = 0$ . ■

**12.27 Theorem.** The exterior product is associative: for any  $\alpha \in \Lambda^r$ ,  $\beta \in \Lambda^s$ , and  $\gamma \in \Lambda^t$ , we have

$$(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma) = \frac{1}{r!} \frac{1}{s!} \frac{1}{t!} A(\alpha \otimes \beta \otimes \gamma). \quad (12.3)$$

**Proof.** Since  $\alpha \wedge \beta$  is alternating, we have

$$A \frac{\alpha \wedge \beta}{(r+s)!} = \alpha \wedge \beta = A \frac{\alpha \otimes \beta}{r! s!};$$

it follows from the last corollary that

$$\begin{aligned}
(\alpha \wedge \beta) \wedge \gamma &= A \left( \frac{(\alpha \wedge \beta)}{(r+s)!} \otimes \frac{\gamma}{t!} \right) \\
&= A \left( \frac{\alpha \otimes \beta}{r! s!} \otimes \frac{\gamma}{t!} \right)
\end{aligned}$$

and, similarly, that

$$\alpha \wedge (\beta \wedge \gamma) = A \left( \frac{\alpha}{r!} \otimes \frac{\beta \otimes \gamma}{s! t!} \right),$$

and since the tensor product is associative, the proof is complete. ■

The formula (12.3) above obviously extends to a product of any number of factors; in particular, we have

$$\tilde{\mathbf{u}}^{i_1} \wedge \tilde{\mathbf{u}}^{i_2} \wedge \dots \wedge \tilde{\mathbf{u}}^{i_r} = A(\tilde{\mathbf{u}}^{i_1} \otimes \dots \otimes \tilde{\mathbf{u}}^{i_r}) = \tilde{\mathbf{u}}^{i_1 \dots i_r}.$$



**12.28 Proposition.** If  $\alpha \in \Lambda^r$ ,  $\beta \in \Lambda^s$ , then  $\alpha \wedge \beta = (-1)^{rs} \beta \wedge \alpha$ .

**Proof.** Let  $\sigma$  be the permutation

$$(1, \dots, r+s) \mapsto (r+1, \dots, r+s, 1, \dots, r)$$

so that  $\beta \otimes \alpha = \sigma(\alpha \otimes \beta)$ . It is easy to calculate that  $\varepsilon(\sigma) = (-1)^{rs}$ , and from Proposition 12.20 we deduce that  $\beta \wedge \alpha = \varepsilon(\sigma) \alpha \wedge \beta$ . ■

**12.29 Proposition.** A finite subset  $\{\alpha^1, \dots, \alpha^r\}$  of  $V^*$  is linearly independent if and only if  $\alpha^1 \wedge \dots \wedge \alpha^r \neq 0$ .

**Proof.** Suppose that  $\alpha^1, \dots, \alpha^r$  are linearly dependent. If  $\alpha^1 = 0$ , it is obvious that  $\alpha^1 \wedge \dots \wedge \alpha^r = 0$ . If  $\alpha^i = \alpha^j$  for some  $i < j$ , then  $\alpha^1 \wedge \dots \wedge \alpha^r = 0$  by the last proposition and associativity. If  $\alpha^1 \neq 0$ , there is some  $k$ ,  $1 < k \leq r$ , such that  $\alpha^k$  is a linear combination of  $\alpha^1, \dots, \alpha^{k-1}$ . Then

$$\begin{aligned} \alpha^1 \wedge \dots \wedge \alpha^k &= \alpha^1 \wedge \dots \wedge \alpha^{k-1} \wedge \left( \sum_{j=1}^{k-1} c_j \alpha^j \right) \\ &= \sum_{j=1}^{k-1} c_j \alpha^1 \wedge \dots \wedge \alpha^{k-1} \wedge \alpha^j = 0. \end{aligned}$$

Next suppose that  $\alpha^1, \dots, \alpha^r$  are linearly independent. Then they can be extended to a basis of  $V^*$ , and so, by Proposition 12.2, there exists a basis  $\mathbf{v}_1, \dots, \mathbf{v}_n$  of  $V$  such that  $\alpha^j = \tilde{\mathbf{v}}^j$  for  $j = 1, \dots, r$ . But then

$$\alpha^1 \wedge \dots \wedge \alpha^r(\mathbf{v}_1, \dots, \mathbf{v}_r) = 1,$$

so  $\alpha^1 \wedge \dots \wedge \alpha^r \neq 0$ . ■

**12.30 Example.** Let us see what all this algebra looks like when  $V = \mathbf{R}^3$ . The spaces  $\Lambda^1$  and  $\Lambda^2$  are both three-dimensional;  $\Lambda^1 = V^*$  can be identified with  $\mathbf{R}^3$  viewed as row vectors, and  $\Lambda^2$  has as a basis the set  $\eta^1 = \tilde{\mathbf{e}}^2 \wedge \tilde{\mathbf{e}}^3$ ,  $\eta^2 = \tilde{\mathbf{e}}^3 \wedge \tilde{\mathbf{e}}^1$ , and  $\eta^3 = \tilde{\mathbf{e}}^1 \wedge \tilde{\mathbf{e}}^2$ . We calculate easily that if  $\tilde{\mathbf{v}}$  and  $\tilde{\mathbf{w}}$  are elements of  $\Lambda^1$ , then the components of  $\tilde{\mathbf{v}} \wedge \tilde{\mathbf{w}}$  in terms of the basis  $\eta^1, \eta^2, \eta^3$  form a row vector which coincides with the familiar *cross product* of the row vectors  $\tilde{\mathbf{v}}, \tilde{\mathbf{w}}$ . We calculate also that if  $\tilde{\mathbf{v}}^j$  ( $j = 1, 2, 3$ ) are any three row vectors, then

$$\begin{aligned} \tilde{\mathbf{v}}^1 \wedge \tilde{\mathbf{v}}^2 \wedge \tilde{\mathbf{v}}^3 &= \sum_{i,j,k} v_i^1 v_j^2 v_k^3 \tilde{\mathbf{e}}^{ijk} \\ &= (\det v_j^i) \tilde{\mathbf{e}}^{123}, \end{aligned}$$

in other words, that the exterior product of three row vectors corresponds to the triple scalar product of the vectors.

## 12.4 Change of Coordinates

**12.31 Definition.** Let  $W$  and  $V$  be finite-dimensional vector spaces, and let  $A : W \rightarrow V$  be a linear transformation. The adjoint of  $A$  is the linear map  $A^*$  of  $T^r(V^*)$  into  $T^r(W^*)$  defined by

$$(A^*\alpha)(\mathbf{w}_1, \dots, \mathbf{w}_r) = \alpha(A\mathbf{w}_1, \dots, A\mathbf{w}_r)$$

for every  $\alpha \in T^r(V^*)$  and all  $\mathbf{w}_1, \dots, \mathbf{w}_r \in W$ .

Suppose that  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is a basis for  $V$ , that  $\mathbf{y}_1, \dots, \mathbf{y}_m$  is a basis for  $W$ , and that  $A\mathbf{y}_j = \sum_{k=1}^n a_j^k \mathbf{x}_k$  (i.e., that  $(a_j^k)$  is the matrix of  $A$  with respect to the given bases.) Let us calculate the matrix of  $A^*$  with respect to the induced bases for the spaces of tensors over  $V^*$  and  $W^*$ .

Let us take first the case  $r = 1$ . Suppose  $\alpha \in V^*$  and  $c_k = \alpha(\mathbf{x}_k)$ , so  $\alpha = \sum_{k=1}^n c_k \tilde{\mathbf{x}}^k$ . We can write  $A^*\alpha = \sum_{j=1}^m d_j \tilde{\mathbf{y}}^j$ . Our object is to express the coefficients  $d_j$  in terms of the coefficients  $c_k$ . Suppose  $\mathbf{w} \in W$ . Writing  $\mathbf{w} = \sum_{j=1}^m w^j \mathbf{y}_j$ , we have

$$\begin{aligned} (A^*\alpha)(\mathbf{w}) &= \alpha(A\mathbf{w}) = \sum_{j=1}^m w^j \alpha(A\mathbf{y}_j) \\ &= \sum_{j=1}^m w^j \alpha\left(\sum_{k=1}^n a_j^k \mathbf{x}_k\right) \\ &= \sum_{j=1}^m \sum_{k=1}^n a_j^k c_k w^j = \sum_{j=1}^m \sum_{k=1}^n a_j^k c_k \tilde{\mathbf{y}}^j(\mathbf{w}), \end{aligned}$$

in other words,  $d_j = \sum_{k=1}^n a_j^k c_k$  for  $j = 1, \dots, m$ .

The case of general  $r$  is just as easy to calculate, although the notation gets a little thick. Suppose now that  $\alpha$  is a tensor of rank  $r$ , and that

$$\alpha = \sum_{i_1, \dots, i_r} c_{i_1 \dots i_r} \tilde{\mathbf{x}}^{i_1} \otimes \dots \otimes \tilde{\mathbf{x}}^{i_r}.$$

We want to find the coefficients  $d_{j_1 \dots j_r}$  in the expression

$$A^*\alpha = \sum_{j_1, \dots, j_r} d_{j_1 \dots j_r} \tilde{\mathbf{y}}^{j_1} \otimes \dots \otimes \tilde{\mathbf{y}}^{j_r}$$

in terms of the coefficients  $c_{i_1 \dots i_r}$ . To this end, let  $\mathbf{w}_1, \dots, \mathbf{w}_r \in W$ , and write

$$\mathbf{w}_j = \sum_{k=1}^m w_j^k \mathbf{y}_k \quad (j = 1, \dots, r)$$

so that

$$A\mathbf{w}_j = \sum_{k=1}^m w_j^k A\mathbf{y}_k = \sum_{i=1}^n \sum_{k=1}^m w_j^k a_k^i \mathbf{x}_i,$$

and hence

$$\begin{aligned}
 \alpha(A\mathbf{w}_1, \dots, A\mathbf{w}_r) &= \alpha\left(\sum_{i_1=1}^n \sum_{k_1=1}^{m_1} w_1^{k_1} a_{k_1}^{i_1} \mathbf{x}_{i_1}, \dots, \sum_{i_r=1}^n \sum_{k_r=1}^{m_r} w_r^{k_r} a_{k_r}^{i_r} \mathbf{x}_{i_r}\right) \\
 &= \sum_{k_1, \dots, k_r} \sum_{i_1, \dots, i_r} w_1^{k_1} \dots w_r^{k_r} a_{k_1}^{i_1} \dots a_{k_r}^{i_r} \alpha(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_r}) \\
 &= \sum_{k_1, \dots, k_r} \sum_{i_1, \dots, i_r} c_{i_1, \dots, i_r} a_{k_1}^{i_1} \dots a_{k_r}^{i_r} w_1^{k_1} \dots w_r^{k_r},
 \end{aligned}$$

which expresses that

$$A^* \alpha = \sum_{k_1, \dots, k_r} d_{k_1, \dots, k_r} \tilde{\mathbf{y}}^{k_1} \otimes \dots \otimes \tilde{\mathbf{y}}^{k_r},$$

where

$$d_{k_1, \dots, k_r} = \sum_{i_1, \dots, i_r} a_{k_1}^{i_1} \dots a_{k_r}^{i_r} c_{i_1, \dots, i_r}, \quad (12.4)$$

the formula we were after.

Calculations such as we have just made would look cleaner, or at least would use less ink, if we adopted the *Einstein summation convention*: if the same index occurs as both a subscript and superscript in a monomial, summation over all values of that index is understood. Thus the transformation law we have just obtained for the components of covariant  $r$ -tensors would be written simply

$$d_{k_1, \dots, k_r} = a_{k_1}^{i_1} \dots a_{k_r}^{i_r} c_{i_1, \dots, i_r}$$

using the summation convention.

Let us next calculate the transformation rules for alternating tensors. If  $\alpha \in \Lambda^r(V^*)$ , we can write

$$\alpha = \sum_I' c_I \tilde{\mathbf{x}}^I, \quad A^* \alpha = \sum_J' d_J \tilde{\mathbf{y}}^J,$$

where the primes indicate summation over (strictly) increasing  $r$ -tuples. We want to express the coefficients  $d_J$  in terms of the  $c_I$ . Since  $A^* \alpha = \sum_I' c_I A^* \tilde{\mathbf{x}}^I$ , and

$$\begin{aligned}
 A^* \tilde{\mathbf{x}}^I &= A^*(A(\tilde{\mathbf{x}}^{i_1} \otimes \dots \otimes \tilde{\mathbf{x}}^{i_r})) \\
 &= A(A^*(\tilde{\mathbf{x}}^{i_1} \otimes \dots \otimes \tilde{\mathbf{x}}^{i_r})) \\
 &= A\left(\sum_J a_{j_1}^{i_1} \dots a_{j_r}^{i_r} \tilde{\mathbf{y}}^{j_1} \otimes \dots \otimes \tilde{\mathbf{y}}^{j_r}\right) \\
 &= \sum_J' \sum_{\sigma \in S_r} \varepsilon(\sigma) a_{j_{\sigma(1)}}^{i_1} \dots a_{j_{\sigma(r)}}^{i_r} \tilde{\mathbf{y}}^{j_1} \wedge \dots \wedge \tilde{\mathbf{y}}^{j_r},
 \end{aligned}$$

we conclude that  $d_J = \sum_I' c_I \Delta_J^I$ , where

$$\Delta_J^I = \sum_{\sigma \in S_r} \varepsilon(\sigma) a_{j_{\sigma(1)}}^{i_1} \cdots a_{j_{\sigma(r)}}^{i_r},$$

in other words,  $\Delta_J^I$  is the determinant of the  $r \times r$  matrix formed from the rows numbered  $i_1, \dots, i_r$  and the columns numbered  $j_1, \dots, j_r$  of the matrix  $(a_j^i)$ . Especially interesting to us is the case  $m = n = r$ ; the spaces  $\Lambda^n(V^*)$  and  $\Lambda^n(W^*)$  are one-dimensional, and the action of  $A^*$  is simply multiplication by  $\det(a_j^i)$ . Some authors define the determinant of a linear transformation  $A: V \rightarrow V$  as the adjoint map  $A^*$  restricted to  $\Lambda^n(V^*)$ .

The most important special case of all these transformation formulas is probably the case where  $W = V$  and  $A$  is the identity transformation. Then the formulas we have derived are those which express the components of a given tensor with respect to a new basis in terms of the old components and the matrix effecting the change of basis.

## 12.5 Exercises

1. Which of the following functions on  $\mathbf{R}^4 \times \mathbf{R}^4$  is a tensor?

- (a)  $\alpha(\mathbf{x}, \mathbf{y}) = x^1 y^1 + x^2 y^3 + x^4 y^1$ .
- (b)  $\alpha(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^2$ .
- (c)  $\alpha(\mathbf{x}, \mathbf{y}) = x^1 + x^2 + x^3 y^3 + x^4 y^4$ .

2. Explain how an inner product on the  $n$ -dimensional vector space  $V$  gives rise to an isomorphism of  $V$  with  $V^*$ . Use this to associate with any inner product  $\gamma$  on  $V$ , in a natural way an inner product  $\tilde{\gamma}$  on  $V^*$ .

3. Let  $V$  be an  $n$ -dimensional vector space with an inner product  $\gamma$ . Let  $\mathbf{v}_1, \dots, \mathbf{v}_n$  be a basis of  $V$ , let  $\tilde{\mathbf{v}}^1, \dots, \tilde{\mathbf{v}}^n$  be the associated dual basis, and let  $T$  be the isomorphism of  $V$  onto  $V^*$  that the inner product gives rise to. Let  $g_{jk} = \gamma(\mathbf{v}_j, \mathbf{v}_k)$  and let  $g^{jk} = \tilde{\gamma}(\tilde{\mathbf{v}}^j, \tilde{\mathbf{v}}^k)$ . Show that  $(g_{jk})$  is the matrix of  $T$  with respect to the given bases, and that  $(g^{jk})$  is the inverse matrix to  $(g_{jk})$ .

4. Compute  $\varepsilon(\sigma)$  when  $\sigma$  is the element of  $S_5$  defined by

$$\sigma(j) = 2j + 1 \pmod{5} \quad (1 \leq j \leq 5).$$

How many terms are there in the product defining  $\varepsilon(\sigma)$ ? Express  $\sigma$  as the product of transpositions, as economically as you can.

5. A tensor  $\alpha$  of rank  $r$  is called *symmetric* if  ${}^\sigma \alpha = \alpha$  for every  $\sigma \in S_r$ . Show that the set of all symmetric tensors of rank  $r$  forms a subspace of  $T^r$ , and find its dimension.

6. Let  $\alpha$  be the tensor of rank 3 on  $\mathbf{R}^4$  defined by

$$\alpha(\mathbf{x}, \mathbf{y}, \mathbf{z}) = x^2 y^2 z^4 + x^3 y^1 z^2.$$

Find  $A\alpha$ , and express in terms of the standard basis  $\tilde{\mathbf{e}}^I$  ( $|I| = 3$ ) for  $\Lambda^3$ .

7. Show that if  $\alpha \in \Lambda^1$ , then  $\alpha \wedge \alpha = 0$ . Give an example of  $\alpha \in \Lambda^2$  such that  $\alpha \wedge \alpha \neq 0$ .

8. Show that every covariant tensor  $\alpha$  has a unique expression of the form  $\alpha = \beta + \gamma$ , where  $\beta$  is alternating and  $A\gamma = 0$ .

9. An alternating  $r$ -tensor  $\alpha$  is called *elementary*, or an  $r$ -covector, if there exist  $\tilde{\mathbf{v}}^1, \dots, \tilde{\mathbf{v}}^r$  in  $V^*$  such that

$$\alpha = \tilde{\mathbf{v}}^1 \wedge \dots \wedge \tilde{\mathbf{v}}^r. \quad (12.5)$$

According to Proposition 12.29 such an  $\alpha \neq 0$  if and only if  $\tilde{\mathbf{v}}^1, \dots, \tilde{\mathbf{v}}^r$  are linearly independent. If  $\alpha \neq 0$  has the form (12.5), define

$$W_\alpha = \{\mathbf{x} \in V : \tilde{\mathbf{v}}^j(\mathbf{x}) = 0, j = 1, \dots, r\}.$$

Show that  $W_\alpha$  is a subspace of  $V$ , and that it depends only on  $\alpha$ , and not on the particular representation (12.5). Show that  $W_\alpha = W_\beta$  if and only if there exists a nonzero constant  $c$  such that  $\alpha = c\beta$ .

10. Let  $\omega$  be a nonzero element of  $\Lambda^n$ . Show that for each  $\alpha \in \Lambda^r$ ,  $\alpha \neq 0$ , there exists  $\beta \in \Lambda^{n-r}$  such that  $\alpha \wedge \beta = \omega$ . Is  $\beta$  unique?

11. Define an inner product on  $\Lambda^r = \Lambda^r(\mathbf{R}^{n*})$  by decreeing that the standard basis elements  $\tilde{\mathbf{e}}^I$  form an orthonormal set. Show that for each  $r$ ,  $0 \leq r \leq n$ , there is a linear map  $\omega \mapsto *\omega$  of  $\Lambda^r$  onto  $\Lambda^{n-r}$  with the properties:

(a)  $|*\omega| = |\omega|$  for every  $\omega \in \Lambda^r$ ;

(b)  $\omega \wedge *\omega = |\omega|^2 \tilde{\mathbf{e}}^{12 \dots n}$ ; and

(c)  $**\omega = (-1)^{r(n-r)}\omega$  for all  $\omega$ .

Here we use the usual notation:  $|\omega|^2 = \langle \omega, \omega \rangle$ .

## 12.6 Notes

Multilinear algebra seems to have begun in the middle of the nineteenth century in the work of Grassman and Möbius. Grassman introduced the exterior product for multivectors. His work remained obscure for many years, but the algebra of alternating multilinear forms (the direct sum of all the  $\Lambda^k$ ) is called the Grassman algebra today, a term first introduced by

E. Cartan in the 1920s. The theory of tensors, covariant and contravariant, developed gradually in the latter part of the century; it was systematized and popularized most notably by Ricci and Levi-Civita in their work on differential geometry at the turn of the century and later. The position of Grassman's multivectors as alternating tensors became clear only gradually.

Having studied tensors and alternating tensors from the purely algebraic point of view, we now consider functions whose values are tensors, or especially, alternating tensors. These are called tensor fields, and alternating tensor fields are called differential forms. They have many applications in geometry and analysis, as well as physics.

### 13.1 Tensor Fields

**13.1 Definition.** Let  $U$  be an open subset of  $\mathbf{R}^n$ . A vector field in  $U$  is a map of  $U$  into  $\mathbf{R}^n$ ; a tensor field of rank  $r$  is a map of  $U$  into  $T^r(\mathbf{R}^{n*})$ . A differential form of degree  $r$  (or briefly,  $r$ -form) in  $U$  is a map  $\omega$  of  $U$  into  $\Lambda^r(\mathbf{R}^{n*})$ ; in other words, a tensor field of rank  $r$  which is alternating at each point.

Thus, if  $\omega$  is a differential form of degree  $r$  in  $U$ , then  $\omega : \mathbf{p} \mapsto \omega_{\mathbf{p}}$ , where  $\omega_{\mathbf{p}}$  is an alternating  $r$ -linear function on  $\mathbf{R}^n$ . In particular, a tensor field of rank 0, or differential form of degree 0 in  $U$ , is simply a real-valued function on  $U$ . A tensor field  $\omega$  of rank 1, or differential form of rank 1, is simply a covector field: it assigns to each point  $\mathbf{p} \in U$  a linear function  $\omega_{\mathbf{p}}$  on  $\mathbf{R}^n$ , which may be identified with a row vector of length  $n$ .

**13.2 Definition.** Let  $\omega$  be a tensor field of rank  $r$  in  $U$ . We say that  $\omega$  is of class  $C^k$  if the real-valued function  $\mathbf{p} \mapsto \omega_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_r)$  is of class  $C^k$  for every  $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbf{R}^n$ .

**Notation.** Recall that  $x^j$  denotes the  $j$ th coordinate function on  $\mathbf{R}^n$ ; thus,  $x^j(a_1, \dots, a_n) = a_j$ . If  $f$  is a differentiable real-valued function on  $U$ , then  $df_{\mathbf{p}}$  is a linear mapping of  $\mathbf{R}^n$  to  $\mathbf{R}$ , for each  $\mathbf{p} \in U$ , given by

$$df_{\mathbf{p}}(h^1, \dots, h^n) = \sum_{j=1}^n (D_j f)(\mathbf{p}) h^j,$$

so  $df$  is a 1-form in  $U$ ; if  $f$  is of class  $C^k$ , then  $df$  is a 1-form of class  $C^{k-1}$ . We observe also that  $dx_{\mathbf{p}}^j(h^1, \dots, h^n) = h^j$ , or in other words,  $dx^j$  is a constant mapping of  $U$  into  $\mathbf{R}^{n*}$ : it assigns to each  $\mathbf{p} \in \mathbf{R}^n$  the element  $\tilde{\mathbf{e}}^j$  of the basis dual to the natural basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of  $\mathbf{R}^n$ . Thus if  $\omega$  is any 1-form in  $U$ , then  $\omega$  can be expressed uniquely in the form  $\omega = \sum_{j=1}^n f_j dx^j$ , where the  $f_j$  are real-valued functions in  $U$ . Since, as we saw in the last chapter, the alternating tensors  $\tilde{\mathbf{e}}^I$ , with  $I$  an increasing  $r$ -tuple, form a basis of  $\Lambda^r(\mathbf{R}^{n*})$ , we see that if  $\omega$  is an  $r$ -form in  $U$ , then  $\omega$  can be uniquely expressed in the form

$$\omega = \sum_I' a_I dx^I,$$

where the prime indicates that the sum is over all increasing  $r$ -tuples  $I = (i_1, \dots, i_r)$ , each  $a_I$  is a real-valued function on  $U$ , and  $dx^I = dx^{i_1} \wedge \dots \wedge dx^{i_r}$ . It is easy to see that  $\omega$  is of class  $C^k$  if and only if each function  $a_I$  is of class  $C^k$ .

In particular, if  $f$  is a differentiable function, we have the formula

$$df = \sum_{j=1}^n (D_j f) dx^j = \sum_{j=1}^n \frac{\partial f}{\partial x^j} dx^j.$$

It will be convenient for us to consider mainly functions and forms of class  $C^\infty$ ; as before, such functions or forms will be called *smooth*.

**Notation.** We denote the space of all smooth  $r$ -forms in the open set  $U$  of  $\mathbf{R}^n$  by  $\Omega^r(U)$ .

## 13.2 The Calculus of Forms

The algebraic operations on alternating tensors naturally give rise to the analogous operations on forms; thus, if  $\omega$  is an  $r$ -form in  $U$ , and  $f$  is a real-valued function in  $U$ , then  $f\omega$  is the  $r$ -form defined by  $(f\omega)_{\mathbf{p}} = f(\mathbf{p})\omega_{\mathbf{p}}$ ; if  $\omega$  and  $\eta$  are  $r$ -forms, then  $\omega + \eta$  is again an  $r$ -form, defined by  $(\omega + \eta)_{\mathbf{p}} = \omega_{\mathbf{p}} + \eta_{\mathbf{p}}$ ; if  $\omega$  is an  $r$ -form, and  $\eta$  is an  $s$ -form, then  $\omega \wedge \eta$  is the  $(r+s)$ -form defined by  $(\omega \wedge \eta)_{\mathbf{p}} = \omega_{\mathbf{p}} \wedge \eta_{\mathbf{p}}$ . The algebraic operations on smooth forms result in smooth forms, obviously. Our next definition extends the mapping  $f \mapsto df$  of  $\Omega^0(U)$  into  $\Omega^1(U)$ .



**13.3 Definition.** Let  $\omega = \sum' a_I dx^I$  be a smooth  $r$ -form in  $U$ . We define the  $(r+1)$ -form  $d\omega$  by

$$d\omega = \sum' (da_I) \wedge dx^I$$

and call it the exterior differential of  $\omega$ .

**13.4 Proposition.** The exterior differential has the following properties:

(a)  $d$  is a linear map of  $\Omega^r(U)$  into  $\Omega^{r+1}(U)$ ;

(b) if  $\omega \in \Omega^r(U)$  and  $\eta \in \Omega^s(U)$ , then

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^r \omega \wedge d\eta;$$

(c) for any  $\omega \in \Omega^r(U)$ ,  $d(d\omega) = 0$ .

**Proof.** The linearity of  $d$  is quite obvious. We prove (b) first for the case  $r = s = 0$ : if  $a$  and  $b$  are smooth functions in  $U$ , then

$$d(ab) = \sum_j D_j(ab) dx^j = \sum_j (bD_j a + aD_j b) dx^j = b da + a db$$

which is (b) for this case (recall that the wedge product when one factor is of rank 0 is taken to be the ordinary product by a scalar). Now, in general, if  $\omega = \sum' a_I dx^I \in \Omega^r(U)$ , and  $\eta = \sum' b_J dx^J \in \Omega^s(U)$ , we have

$$\begin{aligned} d(\omega \wedge \eta) &= d\left(\left(\sum' a_I dx^I\right) \wedge \left(\sum' b_J dx^J\right)\right) \\ &= \sum' \sum' d(a_I b_J dx^I \wedge dx^J) \end{aligned}$$

(using the linearity of  $d$ ), and for each  $I, J$  we have

$$\begin{aligned} d(a_I b_J dx^I \wedge dx^J) &= d(a_I b_J) \wedge dx^I \wedge dx^J \\ &= (b_J da_I + a_I db_J) \wedge dx^I \wedge dx^J \\ &= b_J da_I \wedge dx^I \wedge dx^J + a_I db_J \wedge dx^I \wedge dx^J \\ &= (da_I \wedge dx^I) \wedge (b_J dx^J) \\ &\quad + (-1)^r (a_I dx^I) \wedge (db_J \wedge dx^J), \end{aligned}$$

where we used Proposition 12.28 for the last equality, which gives (b).

Assertion (c) essentially is a formulation of the equality of mixed partial derivatives. If  $f$  is a smooth function, then

$$\begin{aligned} d(df) &= d\left(\sum_{k=1}^n (D_k f) dx^k\right) = \sum_{k=1}^n d(D_k f) \wedge dx^k \\ &= \sum_{k=1}^n \left(\sum_{j=1}^n D_j (D_k f) dx^j\right) \wedge dx^k \end{aligned}$$

$$= \sum_{j < k} (D_j D_k - D_k D_j) f dx^j \wedge dx^k = 0,$$

thus establishing (c) for the case  $r = 0$ . In the general case, we have

$$d(da_I \wedge dx^I) = d(da_I) \wedge dx^I + (-1) da_I \wedge d(dx^I)$$

from (b); but  $d(da_I) = 0$  as we have just seen, and  $d(dx^I) = 0$  by the definition of  $d$ . It follows that  $d(d\omega) = 0$  for any smooth  $\omega$ . ■

### 13.3 Forms and Vector Fields

Since  $\Lambda^n(\mathbf{R}^{n*})$  is one-dimensional, to each  $n$ -form on  $U$  we can associate a real-valued function on  $U$ . Since  $\Lambda^{n-1}(\mathbf{R}^{n*})$ , as well as  $\Lambda^1(\mathbf{R}^{n*}) = \mathbf{R}^{n*}$ , is a space of dimension  $n$ , to each  $(n-1)$  form on  $U$ , as well as to each 1-form on  $U$ , we can associate a *vector field* on  $U$ . We next spell out such a correspondence, and relate it to the differential operator  $d$ .

**13.5 Definition.** Let  $\omega_0 = dx^1 \wedge \cdots \wedge dx^n$ ; for each  $j$ ,  $1 \leq j \leq n$ , we put

$$\eta^j = (-1)^j dx^1 \wedge \cdots \wedge dx^{j-1} \wedge dx^{j+1} \wedge \cdots \wedge dx^n.$$

The mysterious factor  $(-1)^j$  can be justified by the relation  $dx^j \wedge \eta^j = \omega_0$ , for each  $j$ ,  $1 \leq j \leq n$ . We note that  $dx^j \wedge \eta^k = 0$  if  $j \neq k$ .

Every  $\omega \in \Omega^n(U)$  can be uniquely expressed in the form  $\omega = g\omega_0$ , where  $g$  is a smooth function on  $U$ . We denote  $g$  by  $\Phi(\omega)$ . Every  $\omega \in \Omega^{n-1}(U)$  can be uniquely expressed in the form  $\omega = \sum_{j=1}^n g_j \eta^j$ , where each  $g_j$  is a smooth function on  $U$ . Let us then define the map  $\Phi$  of  $\Omega^{n-1}(U)$  to the space  $\mathcal{V}$  of smooth vector fields on  $U$  by  $\Phi(\omega) = (g_1, \dots, g_n)$ . Similarly for 1-forms: if  $\omega = \sum_1^n g_j dx^j$ , let  $\Phi(\omega)$  be the vector field  $(g_1, \dots, g_n)$ . The correspondence  $\Phi$  lets us interpret the differentiation operator in terms of vector fields. For instance, if  $\omega = \sum_1^n g_j \eta^j$ , then

$$d\omega = \sum_{j=1}^n dg_j \wedge \eta^j = \sum_{j=1}^n \left( \sum_{k=1}^n (D_k g_j) dx^k \right) \wedge \eta^j = \left( \sum_{j=1}^n D_j g_j \right) \omega_0.$$

In vector analysis, one defines the *gradient* of the function  $g$  to be the vector field

$$\text{grad } g = (D_1 g, \dots, D_n g),$$

and the *divergence* of the vector field  $\mathbf{g} = (g_1, \dots, g_n)$  to be the function

$$\text{div } \mathbf{g} = \sum_{j=1}^n D_j g_j;$$

thus we have

$$\Phi(dg) = \text{grad } g, \quad \Phi(d\omega) = \text{div } \Phi(\omega)$$

for smooth functions  $g$  and smooth  $(n-1)$ -forms  $\omega$ . We summarize in the diagrams:

$$\begin{array}{ccc} \Omega^0 & \xrightarrow{d} & \Omega^1 \\ \downarrow \text{id} & & \downarrow \Phi \\ C^\infty & \xrightarrow{\text{grad}} & \mathcal{V} \end{array} \qquad \begin{array}{ccc} \Omega^{n-1} & \xrightarrow{d} & \Omega^n \\ \downarrow \Phi & & \downarrow \Phi \\ \mathcal{V} & \xrightarrow{\text{div}} & C^\infty \end{array}$$

In the special case  $n = 3$ , the mapping  $d : \Omega^1(U) \rightarrow \Omega^2(U)$  can also be interpreted in terms of vector fields. If  $\omega = \sum_1^3 g_k dx^k$ , then

$$\begin{aligned} d\omega &= \sum_{k=1}^3 \left( \sum_{j=1}^3 (D_j g_k) dx^j \right) \wedge dx^k = \sum_{j < k} (D_j g_k - D_k g_j) dx^j \wedge dx^k \\ &= (D_2 g_3 - D_3 g_2) dx^2 \wedge dx^3 - (D_1 g_3 - D_3 g_1) dx^1 \wedge dx^3 \\ &\quad + (D_1 g_2 - D_2 g_1) dx^1 \wedge dx^2; \end{aligned}$$

thus, we see that when  $\omega$  is a smooth 1-form in  $R^3$ , we have

$$\Phi(d\omega) = \text{curl } \Phi(\omega),$$

where the *curl* of a vector field  $\mathbf{g} = (g_1, g_2, g_3)$  in  $R^3$  is defined by

$$\text{curl } \mathbf{g} = (D_2 g_3 - D_3 g_2, D_3 g_1 - D_1 g_3, D_1 g_2 - D_2 g_1)$$

(sometimes denoted by  $\text{Rot } \mathbf{g}$ ). The diagram for  $n = 3$  is thus

$$\begin{array}{cccc} \Omega^0 & \xrightarrow{d} & \Omega^1 & \xrightarrow{d} & \Omega^2 & \xrightarrow{d} & \Omega^3 \\ \downarrow \text{id} & & \downarrow \Phi & & \downarrow \Phi & & \downarrow \Phi \\ C^\infty & \xrightarrow{\text{grad}} & \mathcal{V} & \xrightarrow{\text{curl}} & \mathcal{V} & \xrightarrow{\text{div}} & C^\infty \end{array}$$

Other notations use the vector operator “nabla”

$$\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right);$$

using this symbol, we can write

$$\text{grad } g = \nabla g, \quad \text{div } \mathbf{g} = \nabla \cdot \mathbf{g}, \quad \text{curl } \mathbf{g} = \nabla \times \mathbf{g},$$

where the  $\cdot$  is the formal *dot product* or inner product, and  $\times$  refers to the *cross product* of vectors in  $R^3$ ; in our terminology,

$$\mathbf{g} \times \mathbf{h} = \Phi(\Phi^{-1}(g) \wedge \Phi^{-1}(h)).$$

## 13.4 Induced Mappings

Suppose now that  $V$  is an open set in  $\mathbf{R}^m$ , that  $U$  is an open set in  $\mathbf{R}^n$ , and that  $\mathbf{f}$  is a smooth map of  $V$  into  $U$ . Then  $\mathbf{f}$  induces a map  $\mathbf{f}^* : \Omega^r(U) \rightarrow \Omega^r(V)$ , as follows:

$$(\mathbf{f}^*\omega)_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_r) = \omega_{\mathbf{f}(\mathbf{p})}(\mathbf{f}'(\mathbf{p})\mathbf{v}_1, \dots, \mathbf{f}'(\mathbf{p})\mathbf{v}_r).$$

In other words,  $\mathbf{f}^*$  is obtained by putting together the adjoint maps  $[\mathbf{f}'(\mathbf{p})]^*$  of  $\Lambda^r(\mathbf{R}^{n*}) \rightarrow \Lambda^r(\mathbf{R}^{m*})$  induced by the linear maps  $\mathbf{f}'(\mathbf{p})$  of  $\mathbf{R}^n \rightarrow \mathbf{R}^m$  at each point  $\mathbf{p}$  of  $V$  (see Definition 12.31).

Let us calculate the meaning of  $\mathbf{f}^*$  in terms of coordinates. Note that

$$\begin{aligned} (\mathbf{f}^*dx^I)_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_r) &= dx^I(\mathbf{f}'(\mathbf{p})\mathbf{v}_1, \dots, \mathbf{f}'(\mathbf{p})\mathbf{v}_r) \\ &= (df^{i_1} \wedge \dots \wedge df^{i_r})_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_r), \end{aligned}$$

since  $\mathbf{f}'(\mathbf{p})\mathbf{v}$  is the vector with coordinates  $df_{\mathbf{p}}^1(\mathbf{v}), \dots, df_{\mathbf{p}}^n(\mathbf{v})$ . It follows that

$$\mathbf{f}^*\left(\sum_I' c_I dx^I\right) = \sum_I' (c_I \circ \mathbf{f}) df^I, \quad (13.1)$$

where we write  $df^I$  for  $df^{i_1} \wedge \dots \wedge df^{i_r}$  when  $I = (i_1, \dots, i_r)$ . Consider, in particular, the special case  $r = n = m$ . Now  $df^i = \sum_j (D_j f^i) dx^j$ , so

$$\begin{aligned} df^1 \wedge \dots \wedge df^n &= \left( \sum_{j_1=1}^n \frac{\partial f^1}{\partial x^{j_1}} dx^{j_1} \right) \wedge \dots \wedge \left( \sum_{j_n=1}^n \frac{\partial f^n}{\partial x^{j_n}} dx^{j_n} \right) \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) \frac{\partial f^1}{\partial x^{\sigma(1)}} \dots \frac{\partial f^n}{\partial x^{\sigma(n)}} dx^1 \wedge \dots \wedge dx^n \\ &= \det \left( \frac{\partial f^i}{\partial x^j} \right) dx^1 \wedge \dots \wedge dx^n \\ &= (\det \mathbf{f}') dx^1 \wedge \dots \wedge dx^n, \end{aligned}$$

so that in view of (13.1) we have

$$\mathbf{f}^*(\omega) = (\det \mathbf{f}')\omega \circ \mathbf{f} = J_{\mathbf{f}}\omega \circ \mathbf{f}, \quad (13.2)$$

a formula which will be basic.

**13.6 Proposition.** *The map  $\mathbf{f}^*$  is linear, and satisfies:*

- (a)  $\mathbf{f}^*(\omega \wedge \eta) = (\mathbf{f}^*\omega) \wedge (\mathbf{f}^*\eta)$  for any  $\omega \in \Omega^r(U)$ ,  $\eta \in \Omega^s(U)$ ; and
- (b)  $\mathbf{f}^*(d\omega) = d(\mathbf{f}^*\omega)$  for any  $\omega \in \Omega^r(U)$ .

**Proof.** If  $A : V \rightarrow W$  is a linear map of vector spaces, the induced map  $A^*$  of tensors is linear, and preserves the tensor product (and hence the wedge product on alternating tensors); this establishes the linearity of  $f^*$ , and assertion (a). But to verify (b) we only need to observe that  $d(df^{i_1} \wedge \cdots \wedge df^{i_r}) = 0$  by Proposition 13.4, that (b) holds when  $\omega$  is a 0-form by the chain rule, and use the formula (13.1) above for  $f^*\omega$ . ■

The next proposition follows easily from the definition, and the chain rule. We omit the proof.

**13.7 Proposition.** *If  $f : V \rightarrow U$  and  $g : W \rightarrow V$  are smooth, where  $U, V$ , and  $W$  are open sets in  $\mathbf{R}^n, \mathbf{R}^m$ , and  $\mathbf{R}^l$ , respectively, then  $(f \circ g)^* = g^* \circ f^*$ .*

## 13.5 Closed and Exact Forms

**13.8 Definition.** *Let  $\omega \in \Omega^r(U)$ . We say that  $\omega$  is closed if  $d\omega = 0$ . We say that  $\omega$  is exact if there exists  $\eta \in \Omega^{r-1}(U)$  such that  $\omega = d\eta$ .*

Proposition 13.4 shows that the closed  $r$ -forms make up a subspace  $\Omega_c^r$  of the vector space  $\Omega^r(U)$ , and that the exact  $r$ -forms constitute a subspace  $\Omega_e^r = d(\Omega^{r-1})$  of  $\Omega_c^r$ . The quotient space  $\Omega_c^r/\Omega_e^r$  is called the *de Rham cohomology group* of  $U$ . We will not discuss this topic beyond giving the following fundamental theorem, known as Poincaré's lemma.

**13.9 Theorem.** *Let  $U$  be a convex open set in  $\mathbf{R}^n$ , and let  $r$  be a positive integer. Then every closed  $r$ -form in  $U$  is exact; in other words, if  $\omega \in \Omega^r(U)$  and  $d\omega = 0$ , then there exists  $\eta \in \Omega^{r-1}(U)$  such that  $d\eta = \omega$ .*

**Proof.** Fix a point  $\mathbf{x}_0 \in U$ . Let

$$\tilde{U} = \{(\mathbf{x}, t) \in \mathbf{R}^{n+1} : t\mathbf{x} + (1-t)\mathbf{x}_0 \in U\}.$$

Then  $\tilde{U}$  is an open set, since  $U$  is open, and  $\tilde{U} \supset U \times [0, 1]$ , since  $U$  is convex. We begin by constructing a map  $\mathcal{J} : \Omega^{r+1}(\tilde{U}) \rightarrow \Omega^r(U)$  as follows: each  $\tilde{\omega} \in \Omega^{r+1}(\tilde{U})$  has a unique expression in the form

$$\tilde{\omega} = \sum'_{|I|=r+1} a_I dx^I + \sum'_{|J|=r} b_J dx^J \wedge dx^{n+1}, \quad (13.3)$$

where the primes indicate that the sums are taken over all increasing indices, and  $I, J$  are from  $\{1, \dots, n\}$ . We put

$$\mathcal{J}\tilde{\omega} = \sum_J' \left( \int_0^1 b_J(\cdot, t) dt \right) dx^J.$$

Because  $\tilde{U} \supset U \times [0, 1]$ , the integrals involved are well-defined, and are smooth functions on  $U$ , so  $\mathcal{J}\tilde{\omega} \in \Omega^r(U)$ . We note that

$$\begin{aligned} d(\mathcal{J}\tilde{\omega}) &= \sum'_J \sum_{j=1}^n \left( D_j \int_0^1 b_J(\cdot, t) dt \right) dx^j \wedge dx^J \\ &= \sum'_J \sum_{j=1}^n \left( \int_0^1 (D_j b_J)(\cdot, t) dt \right) dx^j \wedge dx^J, \end{aligned} \quad (13.4)$$

the interchange of differentiation and integration being justified by the bounded (even uniform) convergence of the difference quotients to the derivative. On the other hand, if  $\tilde{\omega} \in \Omega^r(\tilde{U})$ , then

$$\begin{aligned} \mathcal{J}(d\tilde{\omega}) &= \mathcal{J} \left( \sum'_I da_I \wedge dx^I \right. \\ &\quad \left. + \sum'_J db_J \wedge dx^J \wedge dx^{n+1} \right) \\ &= \mathcal{J} \left( \sum'_I (D_{n+1} a_I) dx^{n+1} \wedge dx^I \right. \\ &\quad \left. + \sum'_J \sum_{j=1}^n (D_j b_J) dx^j \wedge dx^J \wedge dx^{n+1} \right) \\ &= (-1)^r \sum'_I \left( \int_0^1 (D_{n+1} a_I)(\cdot, t) dt \right) dx^I \\ &\quad + \sum'_J \sum_{j=1}^n \left( \int_0^1 (D_j b_J)(\cdot, t) dt \right) dx^j \wedge dx^J \\ &= (-1)^r \sum'_I [a_I(\cdot, 1) - a_I(\cdot, 0)] dx^I + d(\mathcal{J}\tilde{\omega}), \end{aligned} \quad (13.5)$$

where we have used the fundamental theorem of calculus and equation (13.4) above. Define the maps  $g_i : U \rightarrow \tilde{U}$  ( $i = 0, 1$ ) by  $g_i(\mathbf{x}) = (\mathbf{x}, i)$ ; then  $g_i^*(dx^j) = dx^j$  for  $1 \leq j \leq n$ , and  $g_i^*(dx^{n+1}) = 0$ , so  $g_i^*\tilde{\omega} = \sum'_I (a_I \circ g_i) dx^I$ , if  $\tilde{\omega}$  has the form of (13.3) above. Thus equation (13.5) can be rewritten as

$$\mathcal{J}(d\tilde{\omega}) = d(\mathcal{J}\tilde{\omega}) + (-1)^r [g_1^*\tilde{\omega} - g_0^*\tilde{\omega}]. \quad (13.6)$$

Now let  $F : \tilde{U} \rightarrow U$  be defined by  $F(\mathbf{x}, t) = t\mathbf{x} + (1-t)\mathbf{x}_0$ , so  $F^* : \Omega^r(U) \rightarrow \Omega^r(\tilde{U})$ . Given  $\omega \in \Omega^r(U)$  with  $d\omega = 0$ , set  $\tilde{\omega} = F^*\omega$  and  $\eta = (-1)^{r-1}\mathcal{J}\tilde{\omega}$ . By equation (13.6) above,

$$d\eta = \pm \mathcal{J}(d\tilde{\omega}) + g_1^*\tilde{\omega} - g_0^*\tilde{\omega}. \quad (13.7)$$

Since  $d\tilde{\omega} = d(F^*\omega) = F^*(d\omega) = 0$ , and since  $g_i^*\tilde{\omega} = g_i^*F^*\omega = (F \circ g_i)^*\omega$ , we see that  $g_1^*\tilde{\omega} = \omega$  (since  $F \circ g_1$  is the identity map of  $U$ ) and  $g_0^*\tilde{\omega} = 0$  (since

$F \circ g_0$  is the constant map  $\mathbf{x} \mapsto \mathbf{x}_0$ ). Hence we conclude from equation (13.7) that  $d\eta = \omega$ . ■

## 13.6 Tensor Fields on Manifolds

Let  $M$  be a  $k$ -manifold in  $\mathbf{R}^n$ . By a *tensor field* of rank  $r$ , or a *differential form* of degree  $r$ , on  $M$  we mean a function  $\omega$  which assigns to each  $\mathbf{p} \in M$  an element  $\omega_{\mathbf{p}}$  of  $T^r(T_{\mathbf{p}}^*(M))$ , or respectively, of  $\Lambda^r(T_{\mathbf{p}}^*(M))$ . We say  $\omega$  is smooth if for every local coordinate  $\alpha$  for  $M$  the induced tensor field (or differential form)  $\alpha^*(\omega)$  is smooth in  $V_{\alpha}$ . We denote by  $\Omega^r = \Omega^r(M)$  the set of all smooth  $r$ -forms on  $M$ . It is clear that  $\Omega^r$  is a vector space, and the wedge product  $\omega \wedge \eta \in \Omega^{r+s}$  whenever  $\omega \in \Omega^r$  and  $\eta \in \Omega^s$ .

Let us observe that to check smoothness, it suffices to verify that for every  $\mathbf{p} \in M$  there exists a local coordinate  $\alpha$  with  $\mathbf{p} \in U_{\alpha}$  such that  $\alpha^*(\omega)$  is smooth; for if  $\beta$  is another local coordinate with  $\mathbf{p} \in U_{\beta}$ , then  $\beta = \alpha \circ \varphi_{\alpha\beta}$  in  $V_{\alpha\beta}$ , where  $\varphi_{\alpha\beta}$  is a diffeomorphism of  $V_{\alpha\beta}$  onto  $V_{\beta\alpha}$ , so  $\beta^* = \varphi_{\alpha\beta}^* \alpha^*$ , and thus  $\beta^*(\omega)$  is smooth if and only if  $\alpha^*(\omega)$  is smooth.

Let  $\alpha$  be a local coordinate for  $M$  and let  $\mathbf{y} = \alpha^{-1}$ , so  $\mathbf{y} = (y^1, \dots, y^k)$  is a smooth mapping of  $U_{\alpha}$  to  $V_{\alpha}$ . Since  $\mathbf{y} \circ \alpha$  is the identity map of  $V_{\alpha}$ ,  $d\mathbf{y} \circ d\alpha$  is the identity map of  $\mathbf{R}^k$ , so  $dy^i(\alpha'(t)\mathbf{e}_j) = \delta_j^i$  for  $i, j = 1, \dots, k$ . That is,  $dy^1, \dots, dy^k$  form a basis of  $T_{\mathbf{p}}^*(M)$ , dual to the basis  $\alpha'(t)\mathbf{e}^1, \dots, \alpha'(t)\mathbf{e}^k$  for  $T_{\mathbf{p}}(M)$  (here,  $\mathbf{p} = \alpha(t)$ ). Thus, every  $r$ -form  $\omega$  in  $U_{\alpha}$  can be expressed in the form  $\omega = \sum' c_I dy^I$ , where the sum is over all increasing  $r$ -tuples  $I = (i_1, \dots, i_r)$ , each  $c_I$  is a real function in  $U_{\alpha}$ , and  $dy^I$  is the abbreviation for  $dy^{i_1} \wedge dy^{i_2} \wedge \dots \wedge dy^{i_r}$ . It is clear that  $\omega$  is smooth if and only if each  $c_I$  is a smooth function; indeed,  $\alpha^*(\omega) = \sum'(c_I \circ \alpha) dt^1 \wedge \dots \wedge dt^k$  in  $V_{\alpha}$ . We define  $d\omega$  by putting it equal in  $U_{\alpha}$  to  $\sum' dc_I \wedge dy^I$ ; the verification that the resulting  $(r+1)$ -form is independent of the choice of  $\alpha$  is routine, and will be omitted. The notions of closed and exact forms are applicable to forms on manifolds. Let us call a coordinate patch  $U_{\alpha}$  on  $M$  *convex* if the corresponding  $V_{\alpha}$  is convex. We see from the Poincaré lemma that if  $\omega$  is a closed form on  $M$ , then the restriction of  $\omega$  to any convex coordinate patch is exact; this fact can be phrased as: closed forms are locally exact.

**13.10 Example.** Besides differential forms, other tensor fields on manifolds are of interest. We mention here only one kind of example. A *Riemannian structure* on the manifold  $M$  is a smooth tensor field  $\gamma$  of rank 2 on  $M$  such that  $\gamma_{\mathbf{p}}$  is an inner product on  $T_{\mathbf{p}}(M)$  for each  $\mathbf{p} \in M$ . Any smooth tensor field of rank 2 can be expressed in local coordinates in the form  $\gamma = \sum g_{ij} dy^i \otimes dy^j$ , where each  $g_{ij}$  is a smooth function in the coordinate patch  $U_{\alpha}$ ;  $\gamma$  gives an inner product at each point if and only if the matrix  $(g_{ij})$  is symmetric and positive definite at each point. We note that  $g_{ij}(\mathbf{p}) = \gamma(\alpha'(t)\mathbf{e}_i, \alpha'(t)\mathbf{e}_j)$ , where  $\mathbf{p} = \alpha(t)$ . If  $\gamma$  is the natural Riemannian

nian structure, i.e., the restriction to  $T_p(M)$  of the usual inner product on  $\mathbf{R}^n$ , we see that

$$g_{ij} = \langle D_i \alpha', D_j \alpha' \rangle = \sum_{k=1}^n \frac{\partial \alpha^k}{\partial t^i} \frac{\partial \alpha^k}{\partial t^j},$$

where  $\alpha = (\alpha^1, \dots, \alpha^n)$ .

## 13.7 Integration of Forms in $\mathbf{R}^n$

If  $U$  is an open set in  $\mathbf{R}^n$ , each  $\omega \in \Omega^n(U)$  can be, as we have remarked earlier, expressed uniquely in the form  $\omega = g dx^1 \wedge dx^2 \wedge \dots \wedge dx^n$ , where  $g$  is a smooth function in  $U$ ; so it seems that talking about  $n$ -forms in  $U$  is an unnecessarily complicated way of talking about functions in  $U$ . The advantage of forms over functions appears only when we consider coordinate systems other than the usual rectangular coordinates; it appears in the context of integration.

**13.11 Definition.** Let  $\omega$  be an  $n$ -form in the open set  $U \subset \mathbf{R}^n$ , so  $\omega = g dx^1 \wedge \dots \wedge dx^n = g \omega_0$ ; let  $A$  be a measurable subset of  $U$ . We define  $\int_A \omega$  to be  $\int_A g dm$ , where  $m$  is Lebesgue measure on  $\mathbf{R}^n$ , provided this integral exists.

Thus, for instance,  $\int_K \omega$  is well-defined whenever  $\omega$  is continuous and  $K$  is a compact subset of  $U$ .

**13.12 Lemma.** Let  $U$  and  $V$  be open sets in  $\mathbf{R}^n$ , and let  $\varphi$  be a one-one smooth mapping of  $U$  onto  $V$ , such that  $\det \varphi' > 0$  everywhere on  $U$ . Then for every continuous  $n$ -form  $\omega$  on  $V$ , and every compact subset  $K$  of  $U$ , we have

$$\int_{\varphi(K)} \omega = \int_K \varphi^* \omega.$$

**Proof.** If  $\omega = g dx^1 \wedge \dots \wedge dx^n$ , then

$$\varphi^* \omega = (g \circ \varphi) d\varphi^1 \wedge \dots \wedge d\varphi^n = (g \circ \varphi) (\det \varphi') dx^1 \wedge \dots \wedge dx^n$$

by formula (13.2) above, so the lemma is just a restatement of the change of variables formula (Theorem 10.46) from Chapter 10:

$$\int_{f(K)} g dm = \int_K (g \circ \varphi) |\det \varphi'| dm,$$

since the Jacobian of  $\varphi$ ,  $\det \varphi'$ , is assumed positive. ■

In the next chapter, we will develop integration of forms over manifolds, based on this lemma.



## 13.8 Exercises

1. Characterize the smooth functions  $f$  on  $\mathbf{R}^n$  with the property  $dx^1 \wedge df = 0$ .
2. Let  $g$  be a smooth function in the open set  $U \subset \mathbf{R}^n$ . Characterize the smooth functions  $f$  in  $U$  with the property  $df \wedge dg = 0$ .
3. Find the exterior differential of:
  - (a)  $x^1 dx^1 + x^2 dx^2 + \cdots x^n dx^n$ ;
  - (b)  $x^2 dx^1 - x^1 dx^2$ ;
  - (c)  $x dy \wedge dz + y dz \wedge dx + z dx \wedge dy$ ;
  - (d)  $f(x, y) dx$ .
4. Show that if  $\omega$  and  $\eta$  are closed forms, then so is  $\omega \wedge \eta$ . Show that if  $\omega$  is closed and  $\eta$  is exact, then  $\omega \wedge \eta$  is exact.
5. Suppose that  $\omega \in \Omega^1(U)$ . Show that if there exists a function  $f$ , smooth and never zero in  $U$ , such that  $f\omega$  is closed, then  $\omega \wedge d\omega = 0$ . [Such an  $f$  is called an integrating factor for  $\omega$ .]
6. Show that a closed 0-form in a connected open set  $U \subset \mathbf{R}^n$  must be a constant function. Conclude that if  $\eta$  is an exact 1-form, and  $\mathbf{p} \in U$ , that there exists a *unique* function  $g$  such that  $dg = \eta$  and  $g(\mathbf{p}) = 0$ .
7. Let  $U = \mathbf{R}^2 \setminus \{0\}$ , and consider the following elements of  $\Omega^1(U)$ :

$$\omega = \frac{x dx + y dy}{x^2 + y^2}; \quad \eta = \frac{x dy - y dx}{x^2 + y^2}.$$

Show that  $\omega$  and  $\eta$  are closed, that  $\omega$  is exact, and that  $\eta$  is not exact.

8. Translate the formula  $d(d\omega) = 0$  and the Poincaré lemma into theorems about vector fields in  $\mathbf{R}^3$ .
9. Let  $\omega$  be a smooth  $k$ -form in the open set  $U \subset \mathbf{R}^n$ . Suppose that  $\mathbf{v}_0, \dots, \mathbf{v}_k \in \mathbf{R}^n$ , and define the smooth functions  $f$  and  $g_j$  ( $j = 0, \dots, k$ ) in  $U$  by

$$f(\mathbf{p}) = d\omega_{\mathbf{p}}(\mathbf{v}_0, \dots, \mathbf{v}_k)$$

and

$$g_j(\mathbf{p}) = \omega_{\mathbf{p}}(\mathbf{v}_0, \dots, \widehat{\mathbf{v}_j}, \dots, \mathbf{v}_k),$$

where the hat indicates that the vector with index  $j$  is to be omitted. Show that

$$f(\mathbf{p}) = \sum_{j=0}^k (-1)^j \nabla g_j(\mathbf{p}) \cdot \mathbf{v}_j.$$

10. Let  $\omega$  be the 2-form on  $\mathbf{R}^3$  given by

$$\omega = x \, dy \wedge dz + y \, dz \wedge dx + z \, dx \wedge dy,$$

and let  $\mathbf{f}$  be the mapping of  $\mathbf{R}^2$  into  $\mathbf{R}^3$  defined by

$$\mathbf{f}(u, v) = (\sin u \cos v, \sin u \sin v, \cos u).$$

Calculate  $\mathbf{f}^*(\omega)$ , and then calculate  $\int_D \mathbf{f}^*(\omega)$ , where

$$D = \{(u, v) : 0 \leq u \leq \pi, 0 \leq v \leq 2\pi\}.$$

## 13.9 Notes

While Grassman's work of 1844 used his new algebra to study geometry (the first geometry of higher dimensions), and Cayley, in the same period, used his more accessible algebra for similar purposes, the theory of tensor fields began with Riemann's general theory of manifolds (1854), and was developed by Beltrami and Christoffel in the following decades. The formalism of tensors (originally called the "absolute differential calculus") is due to Ricci, and was developed by Ricci and his student Levi-Civita. The name "tensor calculus" is apparently due to Einstein, whose general theory of relativity made tensor calculus famous all over town. The use of differential forms as a fundamental concept of differential geometry is due to Elie Cartan, in our century.

# Integration on Manifolds

In this chapter, we define the integral of a  $k$ -form over a compact oriented  $k$ -manifold, and prove the important *generalized Stokes' theorem*, which can be regarded as a far-reaching generalization of the fundamental theorem of calculus. We also define the integral of a function over a (not necessarily oriented) manifold, and describe the integral of a form in terms of the integral of a function. The classical theorems of vector analysis (Green's theorem, divergence theorem, Stokes' theorem) appear as special cases of the general Stokes' theorem. Applications are made to topology (the Brouwer fixed point theorem) and to the study of harmonic functions (the mean value property, the maximum principle, Liouville's theorem, and the Dirichlet principle).

## 14.1 Partitions of Unity

We begin by constructing a very useful gadget, the purpose of which is to help us pass from the local to the global.

**14.1 Theorem.** *Let  $\{U_\alpha\}_{\alpha \in A}$  be a collection of open sets in  $\mathbf{R}^n$  and let  $U = \bigcup_{\alpha \in A} U_\alpha$ . There exists a sequence  $\{\phi_j\}_{j=1}^\infty$  of real-valued functions with the following properties:*

- (a) *each  $\phi_j \in C^\infty(\mathbf{R}^n)$ , and vanishes outside a compact subset of  $U_\alpha$ , for some  $\alpha = \alpha(j) \in A$ ;*

- (b) for each compact subset  $K$  of  $U$ ,  $\phi_j = 0$  on  $K$  for all but finitely many  $j$ ; and
- (c) for each  $j$ ,  $0 \leq \phi_j \leq 1$ , and  $\sum_j \phi_j = 1$ .

**Proof.** We carry out the proof with a sequence of lemmas.

**14.2 Lemma.** For any open set  $U \subset \mathbf{R}^n$ , there exists a sequence of compact sets  $(K_i)_{i=1}^\infty$  such that  $K_i \subset \text{int } K_{i+1}$  for every  $i$  and  $\bigcup_{i=1}^\infty K_i = U$ .

**Proof.** Let  $K_i$  be the set of all points  $\mathbf{p}$  satisfying the two conditions:  $B(\mathbf{p}, 1/i) \subset U$  and  $|\mathbf{p}| \leq i$ . The verification that  $(K_i)$  has the desired properties is easy. ■

**14.3 Lemma.** There exists a sequence  $(V_j)_{j=1}^\infty$  of open balls, such that:

- (a)  $U = \bigcup_j V_j$ ;
- (b) for each  $j$ , there is some  $\alpha \in A$  such that  $\bar{V}_j \subset U_\alpha$ ; and
- (c) for each compact  $K \subset U$ ,  $K \cap V_j = \emptyset$  for all but finitely many  $j$ .

**Proof.** Let  $(K_i)$  be the sequence of Lemma 14.2. Let  $L_0 = K_1$  and let  $L_i = K_{i+1} \setminus \text{int } K_i$  for  $i \geq 1$ ; thus  $U = \bigcup_i L_i$ , and  $L_i \cap K_{i-1} = \emptyset$ . For each  $\mathbf{p} \in L_i$ , there exists an  $\alpha \in A$  with  $\mathbf{p} \in U_\alpha$ . Since  $U_\alpha$  and  $K_{i-1}^c$  are open, there exists  $\delta > 0$  such that the open ball  $V_{\mathbf{p}} = B(\mathbf{p}, \delta)$  satisfies  $\bar{V}_{\mathbf{p}} \subset U_\alpha$  and  $V_{\mathbf{p}} \cap K_{i-1} = \emptyset$ . Since  $L_i$  is compact, we can choose a finite set  $F_i$  such that  $L_i \subset \bigcup_{\mathbf{p} \in F_i} V_{\mathbf{p}}$ . Then  $\{V_{\mathbf{p}} : \mathbf{p} \in \bigcup_{i=0}^\infty F_i\}$  is a countable collection of open balls, which may be enumerated as  $(V_j)_{j=1}^\infty$ . Properties (a) and (b) are immediate. If  $K$  is any compact subset of  $U$ , then  $K \subset K_m$  for some  $m$ , since  $\bigcup \text{int } K_m = U$ . Since  $V_{\mathbf{p}} \cap K_m = \emptyset$  if  $\mathbf{p} \in F_i$  with  $i > m$ , we see that  $V_j \cap K = \emptyset$  for all but finitely many  $j$ . ■

**14.4 Lemma.** Let  $V = \{\mathbf{u} : |\mathbf{u} - \mathbf{p}| < \delta\}$  be an open ball in  $\mathbf{R}^n$ . There exists a function  $\psi$  of class  $C^\infty$  on  $\mathbf{R}^n$  such that  $\psi(\mathbf{u}) > 0$  if  $\mathbf{u} \in V$ , and  $\psi(\mathbf{u}) = 0$  whenever  $\mathbf{u} \notin V$ .

**Proof.** Consider the function  $g$  on  $\mathbf{R}$ , defined by

$$g(t) = \begin{cases} 0 & \text{if } t \leq 0, \\ e^{-(1/t)} & \text{if } t > 0. \end{cases}$$

Then  $g(t) > 0$  for all  $t > 0$ , and  $g$  is of class  $C^\infty$ , as we saw in Example 4.39. Let  $f(t) = g(1-t)$ ; then  $f$  is of class  $C^\infty$ ,  $f(t) = 0$  for  $t \geq 1$ , and  $f(t) > 0$  for  $t < 1$ . Finally, put

$$\psi(\mathbf{u}) = f\left(\frac{|\mathbf{u} - \mathbf{p}|^2}{\delta^2}\right).$$

It is clear that  $\psi$  has the desired properties. ■

We now return to the proof of Theorem 14.1. Let  $(V_j)_{j=1}^\infty$  be the sequence of open balls of Lemma 14.3. For each  $j$ , let  $\psi_j$  be a function as described in

Lemma 14.4, so  $\psi_j$  is of class  $C^\infty$ , vanishes outside  $V_j$  (which is contained in a compact subset of some  $U_\alpha$ ), and is strictly positive in  $V_j$ . If  $N$  is any bounded open set with closure contained in  $U$ , only finitely many  $V_j$  meet  $N$ , so  $\psi = \sum_j \psi_j$  is well-defined, and of class  $C^\infty$ , on  $U$ . Furthermore,  $\psi > 0$  everywhere on  $U$ . We put  $\phi_j = \psi_j/\psi$ . It is clear that  $(\phi_j)_{j=1}^\infty$  has the desired properties. ■

**14.5 Definition.** A collection of functions  $(\phi_j)_{j=1}^\infty$  with the properties of Theorem 14.1 is called a  $C^\infty$  partition of unity, subordinate to the covering  $\{U_\alpha\}_{\alpha \in A}$  of  $U$ .

A related idea is the *cut-off function*.

**14.6 Corollary.** Let  $K$  be a compact subset of the open set  $U$  in  $\mathbf{R}^n$ . There exists a function  $\phi \in C^\infty(\mathbf{R}^n)$  such that  $\phi = 1$  on  $K$  and vanishes outside a compact subset of  $U$ , with  $0 \leq \phi \leq 1$  everywhere.

**Proof.** Let  $(\phi_j)_{j=1}^\infty$  be a partition of unity, subordinate to the trivial cover  $\{U\}$ . Since  $K$  is compact,  $J = \{j : \phi_j(\mathbf{u}) \neq 0 \text{ for some } \mathbf{u} \in K\}$  is finite, by property (b) of Theorem 14.1. Then  $\phi = \sum_{j \in J} \phi_j$  has the desired properties. ■

Here are two typical applications of partitions of unity.

**14.7 Proposition.** Let  $A$  be a subset of  $\mathbf{R}^n$ . If  $f$  is a smooth function on  $A$ , then there exists an open set  $U$  and a smooth function  $F$  on  $U$ , such that  $A \subset U$  and  $F(\mathbf{t}) = f(\mathbf{t})$  for all  $\mathbf{t} \in A$ .

**Proof.** According to Definition 11.1, for each  $\mathbf{p} \in A$  there exists an open set  $U_{\mathbf{p}}$  and a function  $F_{\mathbf{p}}$  which is smooth in  $U_{\mathbf{p}}$  and agrees with  $f$  on  $A \cap U_{\mathbf{p}}$ . Let  $U = \bigcup_{\mathbf{p} \in A} U_{\mathbf{p}}$  and let  $(\phi_j)_{j=1}^\infty$  be a  $C^\infty$  partition of unity subordinate to the cover  $\{U_{\mathbf{p}} : \mathbf{p} \in A\}$  of  $U$ . For each  $j$ , there exists  $\mathbf{p} = \mathbf{p}(j)$  such that  $\phi_j$  vanishes outside a compact subset of  $U_{\mathbf{p}}$ ; thus the function  $\phi_j F_{\mathbf{p}}$  vanishes outside a compact subset of  $U_{\mathbf{p}}$ , and so may be regarded as a smooth function on  $U$  (i.e., identify it with its extension by 0 to all of  $U$ ). Put  $F = \sum \phi_j F_{\mathbf{p}(j)}$ ; it is straightforward to see that  $F$  has the desired properties. ■

**14.8 Proposition.** Let  $M$  be a manifold of dimension  $k$  in  $\mathbf{R}^n$ . Then  $M$  is orientable if and only if there exists a nonvanishing  $k$ -form on  $M$ ; more precisely, if and only if there exists an open set  $W$  containing  $M$ , and a smooth  $k$ -form  $\omega$  in  $W$ , such that at each  $\mathbf{p} \in M$  the restriction of  $\omega(\mathbf{p})$  to  $T_{\mathbf{p}}(M)$  (an element of  $\Lambda^k(T_{\mathbf{p}}^*(M))$ ) is not 0.

**Proof.** If there is such an  $\omega$ , we can define an orientation  $\mathcal{O}$  of  $M$  as follows: if  $\alpha$  is a local coordinate at  $\mathbf{p} \in M$ ,  $\alpha(\mathbf{t}) = \mathbf{p}$ , let  $\alpha \in \mathcal{O}$  if

$\omega(\alpha'(t)e_1, \dots, \alpha'(t)e_k) > 0$ . (If  $k = n$ , we already used this method to define an orientation of  $M$ , having available in that case the form  $\omega = dx_1 \wedge \dots \wedge dx_n$ .)

Now suppose that  $M$  is orientable; let  $\mathcal{O}$  be an orientation of  $M$ . For each  $p \in M$ , let  $\alpha \in \mathcal{O}$  be a local coordinate at  $p$ , with  $\alpha(t) = p$ . Then  $\alpha'(t)e_1, \dots, \alpha'(t)e_k$  are linearly independent vectors in  $\mathbf{R}^n$ , so there exists  $\omega_p \in \Lambda^k(\mathbf{R}^{n*})$  such that  $\omega_p(\alpha'(t)e_1, \dots, \alpha'(t)e_k) = 1$ . (If  $v_1, \dots, v_k$  are linearly independent, there exist  $\tilde{v}^j \in \mathbf{R}^{n*}$  ( $1 \leq j \leq k$ ) such that  $\tilde{v}^j(v_i) = \delta_i^j$ ; then  $\tilde{v}^1 \wedge \dots \wedge \tilde{v}^k(v_1, \dots, v_k) = 1$ .) Since  $\alpha$  is smooth, there is a neighborhood  $V_p$  of  $t$ , contained in  $V_\alpha$ , such that  $\omega_p(\alpha'(u)e_1, \dots, \alpha'(u)e_k) > 0$  for all  $u \in V_p$ . Let  $W_p$  be a neighborhood of  $p$  in  $\mathbf{R}^n$  such that  $\alpha^{-1}(W_p) \subset V_p$ . If  $q \in W_p \cap M$ , and  $\beta \in \mathcal{O}$  is a local coordinate at  $q$ , say  $\beta(\tilde{u}) = q = \alpha(u)$ , then

$$\omega_p(\beta'(\tilde{u})e_1, \dots, \beta'(\tilde{u})e_k) = \det \varphi'_{\alpha\beta}(\tilde{u}) \omega_p(\alpha'(u)e_1, \dots, \alpha'(u)e_k),$$

where  $\varphi_{\alpha\beta}$  is the transition function, so  $\beta = \alpha \circ \varphi_{\alpha\beta}$ ; since  $\det(\varphi'_{\alpha\beta}) > 0$ , we see that  $\omega_p(\beta'(\tilde{u})e_1, \dots, \beta'(\tilde{u})e_k) > 0$ . Let  $W = \bigcup_{p \in M} W_p$ . Let  $(\phi_j)_{j=1}^\infty$  be a partition of unity subordinate to  $\{W_p : p \in M\}$ ; thus for each  $j$ ,  $\phi_j$  is a smooth function vanishing outside a compact subset of  $W_p$ , for some  $p = p(j)$ . Define  $\omega$  by  $\omega = \sum_j \phi_j \omega_{p(j)}$ . Since the sum is locally finite,  $\omega$  is a smooth  $k$ -form in  $W$ . For any  $q \in M$ , and any local coordinate  $\beta \in \mathcal{O}$  at  $q$ , say with  $\beta(u) = q$ , we have  $\phi_j(q) \geq 0$  and  $\omega_p(\beta'(u)e_1, \dots, \beta'(u)e_k) > 0$  whenever  $q \in W_p$ ; since  $\sum_j \phi_j = 1$  everywhere, we conclude that  $\omega(q)(\beta'(u)e_1, \dots, \beta'(u)e_k) > 0$ . Thus  $\omega$  is never 0 on any  $T_q(M)$ , as was to be proved. ■

## 14.2 Integrating $k$ -Forms

Next we turn to the main goal of this chapter: to define the integral of a  $k$ -form over an oriented  $k$ -manifold, and to prove the generalized Stokes' theorem.

Recall that we defined (at the end of the last chapter) the integral of a  $k$ -form  $\eta$ , defined in an open subset  $V$  of  $\mathbf{R}^k$ , by  $\int_V \eta = \int_V g \, dm$ , where  $\eta = g \, dx^1 \wedge \dots \wedge dx^k$  and  $m$  denotes Lebesgue measure on  $\mathbf{R}^k$ , provided this latter integral exists. We can use the same definition with  $V$  open in  $\mathbf{R}_+^k$ , instead of  $\mathbf{R}^k$ .

**14.9 Lemma.** *Let  $M$  be an oriented  $k$ -manifold in  $\mathbf{R}^n$ , where  $1 \leq k \leq n$ . Let  $\omega$  be a continuous  $k$ -form in  $\mathbf{R}^n$ , vanishing outside a compact set  $K$ . Suppose  $\alpha$  and  $\beta$  are local coordinates for  $M$ , and  $K \subset \tilde{U}_\alpha \cap \tilde{U}_\beta$ . Then*

$$\int_{V_\alpha} \alpha^*(\omega) = \int_{V_\beta} \beta^*(\omega).$$

**Proof.** The hypothesis on  $\omega$  assures us that the integrals in question are well-defined. We observe that

$$\begin{aligned}\int_{V_\beta} \beta^*(\omega) &= \int_{V_{\alpha\beta}} \beta^*(\omega) = \int_{V_{\alpha\beta}} (\alpha \circ \varphi_{\alpha\beta})^*(\omega) \\ &= \int_{V_{\alpha\beta}} \varphi_{\alpha\beta}^*(\alpha^*(\omega)) \\ &= \int_{V_{\beta\alpha}} \alpha^*(\omega) = \int_{V_\alpha} \alpha^*(\omega),\end{aligned}$$

using Lemma 13.12, and taking account of the fact that  $\det \varphi'_{\alpha\beta} > 0$  when  $\alpha$  and  $\beta$  are coordinates for the *oriented* manifold  $M$ . ■

This lemma makes the following definition possible:

**14.10 Definition.** Let  $1 \leq k \leq n$ . Let  $M$  be a  $k$ -dimensional oriented manifold in  $\mathbf{R}^n$  and  $\alpha$  a local coordinate for  $M$ . If  $\omega$  is a continuous  $k$ -form on  $\mathbf{R}^n$  which vanishes outside a compact subset of  $\tilde{U}_\alpha$ , we define  $\int_M \omega = \int_{V_\alpha} \alpha^*(\omega)$ .

A function or form which vanishes outside a compact set is said to have *compact support*. (The support of a function or form  $f$  is the smallest closed set  $F$  such that  $f$  vanishes outside  $F$ .) We can now use a partition of unity to extend this local definition to a global one.

**14.11 Definition.** Let  $M$  be an oriented  $k$ -manifold in  $\mathbf{R}^n$  and let  $\omega$  be a continuous  $k$ -form which vanishes outside a compact set. Let  $(\phi_j)_{j=1}^\infty$  be a partition of unity subordinate to the collection  $\{\tilde{U}_\alpha : \alpha \in \mathcal{O}\}$ , where  $\mathcal{O}$  is the orientation of  $M$ . We define

$$\int_M \omega = \sum_j \int_M \phi_j \omega.$$

We observe that the sum has only finitely many nonzero terms, since  $\omega$  has compact support. Let us also observe that this definition makes sense, i.e., that it does not depend on the choice of partition of unity. Suppose that  $(\psi_k)_{k=1}^\infty$  is another partition of unity subordinate to  $\{\tilde{U}_\alpha : \alpha \in \mathcal{O}\}$ . Then

$$\begin{aligned}\sum_j \int_M \phi_j \omega &= \sum_j \int_M \left( \sum_k \psi_k \right) \phi_j \omega \\ &= \sum_j \sum_k \int_M \psi_k \phi_j \omega \\ &= \sum_k \int_M \sum_j (\phi_j \psi_k \omega) = \sum_k \int_M \psi_k \omega,\end{aligned}$$

the interchanges of summation with summation and integration being justified by the fact that all sums are finite sums. Thus  $\int_M \omega$  is well-defined.

It is clear that  $\int_M \omega$  depends only on the values of  $\omega$  on  $M$ : if  $\omega(\mathbf{p}) = \eta(\mathbf{p})$  for every  $\mathbf{p} \in M$ , then  $\int_M \omega = \int_M \eta$ . In fact, we see that more is true; the integral depends only on the action of  $\omega$  on the tangent space of  $M$ . That is, if  $\omega$  and  $\eta$  are  $k$ -forms on  $\mathbf{R}^n$  such that

$$\omega(\mathbf{p})(\tau_1, \dots, \tau_k) = \eta(\mathbf{p})(\tau_1, \dots, \tau_k)$$

for every  $\mathbf{p} \in M$  and every  $\tau_1, \dots, \tau_k \in T_{\mathbf{p}}(M)$ , then  $\int_M \omega = \int_M \eta$ . (For in this case, we have  $\alpha^*(\omega) = \alpha^*(\eta)$  for any local coordinate  $\alpha$ .) Thus we have actually defined the integral of a  $k$ -form on  $M$ , although we have always referred above to forms defined in open sets containing  $M$ .

The next theorem generalizes Lemma 14.9.

**14.12 Theorem.** *Let  $M$  and  $N$  be oriented  $k$ -manifolds in  $\mathbf{R}^n$  and suppose that  $\mathbf{g}$  is a diffeomorphism of  $M$  onto  $N$ , i.e., that  $\mathbf{g} : M \rightarrow N$  is smooth and bijective and that  $\mathbf{g}^{-1}$  is also smooth. Then for any smooth  $k$ -form  $\omega$  on  $N$ , vanishing outside a compact set,  $\mathbf{g}^*\omega$  is a smooth  $k$ -form on  $M$ , and we have*

$$\int_N \omega = \int_M \mathbf{g}^*\omega.$$

**Proof.** If  $\alpha$  is a local coordinate for  $N$ , then  $\alpha = \mathbf{g} \circ \beta$  for some local coordinate  $\beta$  for  $M$ , with  $V_\beta = V_\alpha$  (namely,  $\beta = \mathbf{g}^{-1} \circ \alpha$ ). We may assume that  $\omega$  vanishes outside a compact subset of  $U_\alpha$ . Then

$$\int_N \omega = \int_{V_\alpha} \alpha^*\omega = \int_{V_\alpha} (\mathbf{g} \circ \beta)^*\omega = \int_{V_\beta} \beta^*(\mathbf{g}^*\omega) = \int_M \mathbf{g}^*\omega,$$

as claimed. ■

We have put aside the case  $k = 0$ . Recall that a 0-manifold is just a subset  $M$  of  $\mathbf{R}^n$  such that no  $\mathbf{p} \in M$  is a limit point of  $M$ ; then a subset of  $M$  is compact if and only if it is finite. Recall that an orientation on  $M$  is a function  $\varepsilon : M \rightarrow \{1, -1\}$ . If  $g$  is a 0-form (i.e., a function) on  $M$  which vanishes outside a compact subset of  $M$ , i.e., outside a finite subset of  $M$ , we make the natural definition:

$$\int_M g = \sum_j \varepsilon(\mathbf{p}_j) g(\mathbf{p}_j).$$

Suppose that  $\mathbf{f}$  is a one-one smooth mapping of the closed interval  $[a, b]$  into  $\mathbf{R}^n$ , such that  $\mathbf{f}'(t)$  is nonsingular for all  $t \in [a, b]$ . Then  $M = \mathbf{f}([a, b])$  is a compact 1-manifold and  $\partial M = \{\mathbf{p}, \mathbf{q}\}$ , where  $\mathbf{p} = \mathbf{f}(a)$  and  $\mathbf{q} = \mathbf{f}(b)$ . (It can be shown that every connected compact 1-manifold in  $\mathbf{R}^n$  with



nonempty boundary arises this way.) We can take as local coordinates on  $M$

$$\begin{aligned}\alpha(t) &= \mathbf{f}(a+t), & t \in V_\alpha &= [0, b-a], \\ \beta(t) &= \mathbf{f}(b+t), & t \in V_\beta &= (a-b, 0],\end{aligned}$$

and it is easy to check that  $\varphi'_{\alpha\beta}(t) = 1$  for all  $t$ , so these local coordinates define an orientation on  $M$ . With this orientation, the induced orientation on  $\partial M$  is given by  $\varepsilon(\mathbf{q}) = +1$ ,  $\varepsilon(\mathbf{p}) = -1$ . If  $g$  is any smooth 0-form (i.e., smooth function) on  $M$ , then

$$\begin{aligned}\int_M dg &= \int_{[a,b]} \mathbf{f}^*(dg) = \int_{[a,b]} d(\mathbf{f}^*g) \\ &= \int_a^b (g \circ \mathbf{f})'(t) dt = g(\mathbf{q}) - g(\mathbf{p}) = \int_{\partial M} g.\end{aligned}$$

This is the case  $k = 1$  of the next result, which is known as *Stokes' theorem for manifolds*.

**14.13 Theorem.** *Let  $M$  be a compact oriented manifold of dimension  $k$ , and let  $\omega$  be a smooth  $(k-1)$ -form in an open set containing  $M$ . Then*

$$\int_M d\omega = \int_{\partial M} \omega.$$

If  $\partial M = \emptyset$ , we take the right-hand side of this equation to be 0.

**Proof.** Let  $\mathcal{O}$  be the orientation of  $M$  and let  $(\phi_j)_{j=1}^\infty$  be a partition of unity subordinate to  $\{\tilde{U}_\alpha : \alpha \in \mathcal{O}\}$ . Then, since  $\omega = \sum_j \phi_j \omega$ , we have

$$\int_M d\omega = \int_M d\left(\sum_j \phi_j \omega\right) = \int_M \sum_j d(\phi_j \omega) = \sum_j \int_M d(\phi_j \omega),$$

the interchange of summation with integration or with exterior differentiation being justified by the fact that the sums are finite. Also,

$$\int_{\partial M} \omega = \sum_j \int_{\partial M} \phi_j \omega$$

according to Definition 14.11. Thus it suffices to prove the theorem for the special case that  $\omega$  vanishes outside a compact subset of  $\tilde{U}_\alpha$  for some local coordinate  $\alpha \in \mathcal{O}$ . Let  $\eta = \alpha^*(\omega)$ , so  $\eta$  is a smooth  $(k-1)$ -form in  $V_\alpha$ , vanishing outside a compact subset of  $V_\alpha$ , and  $d\eta = d(\alpha^*\omega) = \alpha^*(d\omega)$ . We may regard  $\eta$  as a smooth  $(k-1)$ -form on  $\mathbf{R}^k$  or  $\mathbf{R}_+^k$ , by extending it to be 0 outside  $V_\alpha$ . Then

$$\int_M d\omega = \int_{\mathbf{R}^k} d\eta \quad \text{or} \quad \int_{\mathbf{R}_+^k} d\eta, \quad \int_{\partial M} \omega = \int_{\partial \mathbf{R}_+^k} \eta.$$

Thus we have reduced the proof of Stokes' theorem to verifying the following special cases:

**14.14 Lemma.** *If  $\eta \in \Omega^{k-1}(\mathbf{R}^k)$  vanishes outside a compact subset of  $\mathbf{R}^k$ , then  $\int_{\mathbf{R}^k} d\eta = 0$ . If  $\eta \in \Omega^{k-1}(\mathbf{R}_+^k)$  vanishes outside a compact subset of  $\mathbf{R}_+^k$ , then  $\int_{\mathbf{R}_+^k} d\eta = \int_{\partial\mathbf{R}_+^k} \eta$ .*

**Proof.** Let

$$\eta^j = (-1)^{k-1} dx^1 \wedge \cdots \wedge dx^{j-1} \wedge dx^{j+1} \wedge \cdots \wedge dx^k$$

for  $1 \leq j \leq k$ . Then  $\eta_1, \dots, \eta_k$  provide a basis for  $\Lambda^{k-1}$  at each point of  $\mathbf{R}^k$ , so we can write

$$\eta = \sum_{j=1}^k g_j \eta^j,$$

where  $g_j$  is a smooth function for each  $j$ . Since  $dx^j \wedge \eta^j = dx^1 \wedge \cdots \wedge dx^k$ , we see that

$$d\eta = \sum_{j=1}^k (D_j g_j) dx^1 \wedge \cdots \wedge dx^k.$$

So if  $\eta \in \Omega^{k-1}(\mathbf{R}^k)$  vanishes outside a compact set, we have

$$\int_{\mathbf{R}^k} d\eta = \sum_{j=1}^k \int_{\mathbf{R}^k} (D_j g_j) dm,$$

and this integral can be computed by iterated integrals over  $\mathbf{R}$ , in any order. But, by the fundamental theorem of calculus,

$$\int_{-\infty}^{\infty} (D_j g_j)(\mathbf{u}, t, \mathbf{v}) dt = g_j(\mathbf{u}, +\infty, \mathbf{v}) - g_j(\mathbf{u}, -\infty, \mathbf{v}) = 0,$$

for any  $j$ ,  $1 \leq j \leq k$ , and any  $\mathbf{u} = (u_1, \dots, u_{j-1})$ ,  $\mathbf{v} = (u_{j+1}, \dots, u_k)$ . (For  $g_j(\mathbf{u}, t, \mathbf{v}) = 0$  for all  $\mathbf{u}, \mathbf{v}$  if  $|t|$  is sufficiently large.) It follows that  $\int_{\mathbf{R}^k} d\eta = 0$ .

Similarly, if  $\eta \in \Omega^{k-1}(\mathbf{R}_+^k)$  vanishes outside a compact subset of  $\mathbf{R}_+^k$ , we have

$$\int_{\mathbf{R}_+^k} d\eta = \int_0^{+\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{j=1}^k (D_j g_j)(\mathbf{u}) du_1 \cdots du_k,$$

where again the integrals can be computed in any order. Integrating the term  $D_j g_j$  first with respect to  $u_j$ , we get 0 as above, for any  $j$ ,  $1 \leq j < k$ ; the remaining integral is

$$\int_{\mathbf{R}^{k-1}} \int_0^{\infty} (D_k g_k)(\mathbf{u}, t) dt dm(\mathbf{u}),$$

and

$$\int_0^\infty (D_k g_k)(\mathbf{u}, t) dt = g_k(\mathbf{u}, \infty) - g_k(\mathbf{u}, 0) = -g_k(\mathbf{u}, 0),$$

since  $g_k$  vanishes outside a bounded set. Thus we have

$$\int_{\mathbf{R}_+^k} d\eta = - \int_{\mathbf{R}^{k-1}} g_k(\mathbf{u}, 0) dm(\mathbf{u}).$$

Now we compute

$$\int_{\partial \mathbf{R}_+^k} \eta = \sum_{j=1}^k \int_{\partial \mathbf{R}_+^k} g_j \eta^j.$$

Since  $dx^k = 0$  on  $\partial \mathbf{R}_+^k$ , we see that  $\eta^j = 0$  on  $\partial \mathbf{R}_+^k$  for  $1 \leq j < k$ , and thus the integral reduces to one term:

$$\int_{\partial \mathbf{R}_+^k} \eta = \int_{\partial \mathbf{R}_+^k} g_k \eta^k = \pm (-1)^{k-1} \int_{\mathbf{R}^{k-1}} g_k(\mathbf{u}, 0) dm(\mathbf{u}),$$

where the sign  $\pm 1$  is determined by the orientation given to  $\partial \mathbf{R}_+^k$ . Now recall the convention made in Chapter 11 about the orientation of  $\partial M$ ; it was designed for the express purpose of having  $\int_{\partial \mathbf{R}_+^k} g \eta_k = - \int_{\mathbf{R}^{k-1}} g dm$ , so that the lemma is proved, and with it the formula of Stokes is established. ■

## 14.3 The Brouwer Fixed Point Theorem

In this section, we use Stokes' theorem to prove the following theorem, known as Brouwer's fixed point theorem. Recall from Chapter 11 the notations  $B^n = \{\mathbf{x} \in \mathbf{R}^n : |\mathbf{x}| \leq 1\}$  and  $S^{n-1} = \{\mathbf{x} \in \mathbf{R}^n : |\mathbf{x}| = 1\}$ ; we observed there that  $B^n$  is an  $n$ -manifold and  $S^{n-1} = \partial B^n$ .

**14.15 Theorem.** *Let  $f$  be a continuous map of  $B^n$  into itself. Then  $f$  has a fixed point: there exists  $\mathbf{p} \in B^n$  such that  $f(\mathbf{p}) = \mathbf{p}$ .*

We shall approach this theorem with some preliminary results.

**14.16 Lemma.** *There is no smooth map  $g : B^n \rightarrow S^{n-1}$  with the property that  $g(\mathbf{p}) = \mathbf{p}$  for every  $\mathbf{p} \in S^{n-1}$ .*

**Proof.** Suppose that such a  $g = (g^1, \dots, g^n)$  exists. Let

$$\omega = \sum_{j=1}^n (-1)^{j-1} x^j dx^1 \wedge \cdots \wedge \widehat{dx^j} \wedge \cdots \wedge dx^n,$$

so  $\omega$  is a smooth  $(n-1)$ -form in  $\mathbf{R}^n$ , with  $d\omega = ndx^1 \wedge \cdots \wedge dx^n$ . We observe that

$$\mathbf{g}^*(\omega) = \sum_{j=1}^n (-1)^{j-1} g^j dg^1 \wedge \cdots \wedge \widehat{dg^j} \wedge \cdots \wedge dg^n,$$

so that  $\mathbf{g}^*(\omega) = \omega$  on  $S^{n-1}$  (since  $\mathbf{g}$  is the identity map on  $S^{n-1}$ ), while

$$d(\mathbf{g}^*(\omega)) = \mathbf{g}^*(d\omega) = n dg^1 \wedge \cdots \wedge dg^n.$$

But the equation

$$\sum_{j=1}^n (g^j)^2 = 1$$

on  $B^n$ , which expresses the fact that  $\mathbf{g}$  maps  $B^n$  into  $S^{n-1}$ , implies that

$$\sum_{j=1}^n g^j dg^j = 0$$

throughout  $B^n$ , so that the differentials  $dg^1, \dots, dg^n$  are linearly dependent at each point of  $B^n$ , and hence (Proposition 12.29)  $dg^1 \wedge \cdots \wedge dg^n = 0$  identically in  $B^n$ . Applying Stokes' theorem, we have

$$\int_{S^{n-1}} \omega = \int_{B^n} d\omega = n \int_{B^n} dx^1 \wedge \cdots \wedge dx^n = nm(B^n),$$

but, on the other hand,

$$\begin{aligned} \int_{S^{n-1}} \omega &= \int_{S^{n-1}} \mathbf{g}^*(\omega) \\ &= \int_{B^n} d(\mathbf{g}^*(\omega)) \\ &= \int_{B^n} ndg^1 \wedge \cdots \wedge dg^n = 0, \end{aligned}$$

a contradiction. ■

**14.17 Theorem.** Let  $\mathbf{f} : B^n \rightarrow B^n$  be a smooth mapping. There exists  $\mathbf{p} \in B^n$  such that  $\mathbf{f}(\mathbf{p}) = \mathbf{p}$ .

**Proof.** If  $\mathbf{f}(\mathbf{p}) \neq \mathbf{p}$  for every  $\mathbf{p} \in B^n$ , we can construct a smooth  $\mathbf{g} : B^n \rightarrow S^{n-1}$  such that  $\mathbf{g}(\mathbf{p}) = \mathbf{p}$  whenever  $\mathbf{p} \in S^{n-1}$ , as follows: for each  $\mathbf{p} \in B^n$ , we put

$$\mathbf{g}(\mathbf{p}) = \mathbf{p} + t(\mathbf{p} - \mathbf{f}(\mathbf{p})),$$

where the nonnegative number  $t = t(\mathbf{p})$  is chosen so that  $|\mathbf{g}(\mathbf{p})| = 1$ . An examination of this condition shows that it is equivalent to a quadratic

equation in  $t$  which has exactly one nonnegative solution  $t$ ; furthermore,  $\mathbf{g}(\mathbf{p}) = \mathbf{p}$  if and only if  $t = 0$  if (and only if)  $|\mathbf{p}| = 1$ . One also sees that the discriminant of this quadratic is strictly positive, so that the solution  $t(\mathbf{p})$  is a smooth function of  $\mathbf{p}$ , and hence the resulting function  $\mathbf{g}$  is smooth. Applying the last lemma, we have reached a contradiction. ■

To deduce Theorem 14.15 from Theorem 14.17, we use the next two lemmas.

**14.18 Lemma.** *If  $\mathbf{f} : B^n \rightarrow B^n$  is continuous and  $\epsilon > 0$ , there exists a smooth  $\mathbf{g} : B^n \rightarrow B^n$  such that  $|\mathbf{g}(\mathbf{p}) - \mathbf{f}(\mathbf{p})| < 2\epsilon$  for every  $\mathbf{p} \in B^n$ .*

**Proof.** Apply the Weierstrass polynomial approximation theorem to each component of  $\mathbf{f}$  to obtain  $\mathbf{h} : B^n \rightarrow \mathbb{R}^n$  with each component of  $\mathbf{h}$  a polynomial, such that  $|\mathbf{h}(\mathbf{p}) - \mathbf{f}(\mathbf{p})| < \epsilon$  for every  $\mathbf{p} \in B^n$ . Then put  $\mathbf{g}(\mathbf{p}) = \mathbf{h}(\mathbf{p})/(1 + \epsilon)$ , and we have  $\mathbf{g}(B^n) \subset B^n$  and  $|\mathbf{f}(\mathbf{p}) - \mathbf{g}(\mathbf{p})| < 2\epsilon$  for every  $\mathbf{p} \in B^n$ . ■

**14.19 Lemma.** *Suppose that  $\mathbf{g}_k$  is a continuous map of  $B^n$  into itself for each positive integer  $k$ , and that  $\mathbf{g}_k(\mathbf{p}_k) = \mathbf{p}_k$  for some  $\mathbf{p}_k \in B^n$ . If  $(\mathbf{g}_k)$  converges uniformly to  $\mathbf{f}$ , then  $\mathbf{f}(\mathbf{p}) = \mathbf{p}$  for any limit point  $\mathbf{p}$  of  $(\mathbf{p}_k)$ .*

**Proof.** Let  $\epsilon > 0$ . There exists a neighborhood  $N$  of  $\mathbf{p}$  such that  $|\mathbf{f}(\mathbf{p}) - \mathbf{f}(\mathbf{q})| < \epsilon$  for all  $\mathbf{q} \in N$ , and there exists  $m$  such that  $|\mathbf{f}(\mathbf{q}) - \mathbf{g}_k(\mathbf{q})| < \epsilon$  for all  $\mathbf{q} \in B^n$ , whenever  $k \geq m$ . If  $k \geq m$  is chosen so that  $|\mathbf{p}_k - \mathbf{p}| < \epsilon$  and  $\mathbf{p}_k \in N$ , it follows that

$$\begin{aligned} |\mathbf{f}(\mathbf{p}) - \mathbf{p}| &= |\mathbf{f}(\mathbf{p}) - \mathbf{f}(\mathbf{p}_k) + \mathbf{f}(\mathbf{p}_k) - \mathbf{g}_k(\mathbf{p}_k) + \mathbf{g}_k(\mathbf{p}_k) - \mathbf{p}| \\ &\leq |\mathbf{f}(\mathbf{p}) - \mathbf{f}(\mathbf{p}_k)| + |\mathbf{f}(\mathbf{p}_k) - \mathbf{g}_k(\mathbf{p}_k)| + |\mathbf{p}_k - \mathbf{p}| \\ &< \epsilon + \epsilon + \epsilon = 3\epsilon, \end{aligned}$$

and since  $\epsilon > 0$  was arbitrary, it follows that  $\mathbf{f}(\mathbf{p}) = \mathbf{p}$ . ■

**Proof of Theorem 14.15.** By Lemma 14.18, there exists a sequence  $(\mathbf{g}_k)_{k=1}^\infty$  of smooth maps of  $B^n$  into itself, such that  $\mathbf{g}_k$  converges uniformly to  $\mathbf{f}$  on  $B^n$ . By Theorem 14.17, for each  $k$  there exists  $\mathbf{p}_k \in B^n$  with  $\mathbf{g}_k(\mathbf{p}_k) = \mathbf{p}_k$ . Since  $B^n$  is compact, there exists a limit point  $\mathbf{p}$  of  $\{\mathbf{p}_k\}$ , and the theorem follows by applying Lemma 14.19. ■

## 14.4 Integrating Functions on a Manifold

In this section, our goal is to define the notion of  $k$ -dimensional volume in a manifold of dimension  $k$ . The simplest case is a  $k$ -dimensional linear subspace of  $\mathbb{R}^n$ . Let  $\mathbf{a}_1, \dots, \mathbf{a}_k$  be a set of  $k$  vectors in  $\mathbb{R}^n$ . We begin by

finding a formula for the “ $k$ -dimensional volume” of the “box” spanned by  $\mathbf{a}_1, \dots, \mathbf{a}_k$ ,

$$B(\mathbf{a}_1, \dots, \mathbf{a}_k) = \left\{ \sum_{j=1}^k t_j \mathbf{a}_j : 0 \leq t_j \leq 1 \right\}$$

which meets our minimum expectations. These expectations are: if each of the vectors  $\mathbf{a}_j$  lies in the subspace  $\{\mathbf{p} : x^j(\mathbf{p}) = 0 \text{ if } j > k\}$ , then we want the volume of  $B(\mathbf{a}_1, \dots, \mathbf{a}_k)$  to be the usual volume (i.e., Lebesgue measure) of the corresponding box in  $\mathbf{R}^k$ ; furthermore, the volume of a box should be invariant under rotations and reflections. The desired formula is found in the following proposition. Let  $A$  be the  $n \times k$  matrix  $[\mathbf{a}_1, \dots, \mathbf{a}_k]$ .

**14.20 Proposition.** Let  $\mathcal{M}_{n,k}$  denote the set of  $n \times k$  real matrices, where  $1 \leq k \leq n$ . Let  $V(A) = [\det A^t A]^{1/2}$  for  $A \in \mathcal{M}_{n,k}$ . Then  $V$  is the unique nonnegative function on  $\mathcal{M}_{n,k}$  which satisfies the conditions:

- (a)  $V(TA) = V(A)$  whenever  $T$  is an orthogonal  $n \times n$  matrix; and
- (b) if

$$A = \begin{bmatrix} B \\ 0 \end{bmatrix},$$

where  $B \in \mathcal{M}_{k,k}$  and  $0$  is the  $(n - k) \times k$  matrix of zeros, then  $V(A) = |\det B|$ .

**Proof.** Note that  $\det A^t A$  is a positive semi-definite matrix, since

$$\langle A^t A \mathbf{x}, \mathbf{x} \rangle = \langle A \mathbf{x}, A \mathbf{x} \rangle \geq 0$$

for every  $\mathbf{x} \in \mathbf{R}^k$ , so that  $\det A^t A \geq 0$ , and  $V$  is well-defined. (If  $k = n$ , then, of course,  $\det A^t A = (\det A)^2$ , so  $V(A) = |\det A|$ .) If  $T \in \mathcal{M}_{n,n}$  is orthogonal, then  $T^t T = I$ , the  $n \times n$  identity matrix, so  $\det[(TA)^t(TA)] = \det(A^t T^t T A) = \det(A^t A)$ , and property (a) follows. If  $A$  has the form described in (b), then  $A^t A = B^t B$ , so  $\det(A^t A) = \det(B^t B) = \det B^t \det B = (\det B)^2$ , and property (b) holds. Now if  $F : \mathcal{M}_{n,k} \rightarrow [0, \infty)$  has the properties (a) and (b), and  $A \in \mathcal{M}_{n,k}$ , we can always find an orthogonal  $T$  which maps the space spanned by the  $k$  columns of  $A$  into  $\mathbf{R}^k \times \{0\}$ , i.e., such that

$$TA = \begin{bmatrix} B \\ 0 \end{bmatrix}$$

so  $F^2(A) = F^2(TA) = |\det B|^2 = \det(B^t B) = \det(A^t A)$ . ■

We note that the matrix  $A^t A$  has the inner product  $\langle \mathbf{a}_i, \mathbf{a}_j \rangle$  as the entry in the  $i$ th row and  $j$ th column;  $\det A^t A$  is called the *Gram determinant*.

**14.21 Lemma.** Let  $C \in \mathcal{M}_{k,k}$ ,  $A \in \mathcal{M}_{n,k}$ , and let  $B = AC$ . Then  $V(B) = |\det C| V(A)$ .

**Proof.** We have

$$\begin{aligned} V^2(B) &= \det(B^t B) = \det((AC)^t(AC)) \\ &= \det(C^t(A^t A)C) \\ &= (\det C)^2 \det(A^t A) = (\det C)^2 V^2(A), \end{aligned}$$

and the lemma follows. ■

**14.22 Definition.** Let  $A$  be an  $n \times k$  matrix, and let  $I = (i_1, \dots, i_k)$ , where  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ . We denote by  $A^I$  the  $k \times k$  matrix formed with the rows of  $A$  numbered  $i_1, \dots, i_k$ . Thus,  $(A^I)_m^j = a_{i_m}^{j_m}$  for  $1 \leq j, m \leq k$ .

**14.23 Proposition.** The volume  $V(A)$  of the box spanned by the columns of  $A$  can also be computed by the formula

$$V^2(A) = \sum_{|I|=k}' \det[(A^I)^t A^I] = \sum_{|I|=k}' (\det A^I)^2,$$

where the sum is taken over all increasing  $k$ -tuples  $I$ .

**Proof.** The formula is trivial for  $k = 1$  or  $k = n$ , but not so obvious for  $1 < k < n$ . Let  $A \in \mathcal{M}_{n,k}$ , and define  $F : (\mathbf{R}^n)^k \rightarrow \mathbf{R}$  by  $F(\mathbf{b}_1, \dots, \mathbf{b}_k) = F(B) = \det B^t A$ . Similarly, define  $G : (\mathbf{R}^n)^k \rightarrow \mathbf{R}$  by

$$G(\mathbf{b}_1, \dots, \mathbf{b}_k) = G(B) = \sum_{|I|=k}' \det[(B^I)^t A^I] = \sum_I' (\det B^I)(\det A^I).$$

It is obvious that  $F$  and  $G$  are each linear in each variable separately, i.e., are tensors of rank  $k$ . Suppose  $\mathbf{b}_i = \mathbf{e}_{j_i}$ , for some  $\{j_1, \dots, j_k\} = J \subset \{1, \dots, n\}$ . Then  $B^I$  has a row of zeros for every  $I \neq J$ , so  $\det B^I = 0$  for  $I \neq J$ , and thus  $G(B) = \det((B^J)^t A^J)$ . But, in fact,  $B^t A = (B^J)^t A^J$  for such  $B$ , so we conclude  $F(B) = G(B)$  for all such  $B$ , and hence, by linearity, for every  $B \in \mathcal{M}_{n,k}$ . In particular,  $F(A) = G(A)$ . ■

The motivation for the next definition should be clear; if  $\alpha$  is a local coordinate for the manifold  $M$ , the image under  $\alpha$  of the  $k$ -dimensional box

$$\{\mathbf{t} : t_0^j \leq t^j \leq t_0^j + h^j\} = \mathbf{t}_0 + B(h^1 \mathbf{e}_1, \dots, h^k \mathbf{e}_k)$$

in  $V_\alpha$  is “approximated” in  $\mathbf{R}^n$  by the  $k$ -dimensional box

$$\alpha(\mathbf{t}_0) + B(h^1 \alpha'(\mathbf{t}_0) \mathbf{e}_1, \dots, h^k \alpha'(\mathbf{t}_0) \mathbf{e}_k),$$

which has  $k$ -dimensional volume

$$V(B(h^1 \alpha'(\mathbf{t}_0) \mathbf{e}_1, \dots, h^k \alpha'(\mathbf{t}_0) \mathbf{e}_k)) = h^1 \cdots h^k V(\alpha'(\mathbf{t}_0)).$$

**14.24 Definition.** Let  $M$  be a  $k$ -manifold in  $\mathbf{R}^n$ , and let  $\alpha$  be a local coordinate for  $M$ . If  $\psi : M \rightarrow \mathbf{R}$  is a Borel function vanishing outside a compact subset of  $U_\alpha$ , such that  $\psi$  is either bounded or nonnegative, we define

$$\int_M \psi dV = \int_{U_\alpha} \psi dV = \int_{V_\alpha} \psi(\alpha(t))V(\alpha'(t)) dm(t).$$

Since  $V(\alpha')$  is a positive continuous function on  $V_\alpha$ , the right-hand side of the above equation is well-defined under the hypothesis on  $\psi$ . It is not hard to verify that the value does not depend on the choice of local coordinate  $\alpha$ . This follows from Lemma 14.21 above, and the usual change of variables formula for integration in  $\mathbf{R}^n$ . It is trivial that the map  $\psi \rightarrow \int_M \psi dV$  is linear, and order-preserving.

**14.25 Definition.** Let  $M$  be a  $k$ -manifold in  $\mathbf{R}^n$ , and suppose  $\psi : M \rightarrow \mathbf{R}$  is either a nonnegative Borel function, or a bounded Borel function which vanishes outside a compact subset of  $M$ . Let  $(\phi_j)_{j=1}^\infty$  be a  $C^\infty$  partition of unity subordinate to the collection of all coordinate patches for  $M$ . We define

$$\int_M \psi dV = \sum_j \int_M \phi_j \psi dV.$$

The right-hand side above makes sense since each  $\phi_j \psi$  vanishes outside a compact subset of some coordinate patch, and since there are only finitely many nonzero terms, except in the case when all terms are nonnegative. The verification that the result is independent of the choice of partition of unity is exactly the same as the corresponding fact concerning the definition of  $\int_M \omega$  when  $\omega$  is a  $k$ -form. Taking  $\psi$  to be an indicator function, we can define the volume of any Borel subset of  $M$ : if  $A$  is a Borel subset of  $M$ ,

$$V(A) = \int_M \mathbf{1}_A dV.$$

We will sometimes write  $V_k$  instead of  $V$  if we are in a context where there is more than one dimension to consider.

Note that we have defined a volume measure on any manifold, and can integrate any positive Borel function, whether or not the manifold is orientable; to integrate forms, an orientation is essential. We next show that the integral of a  $k$ -form can, in fact, be expressed as the integral of a function with respect to  $k$ -dimensional volume.

**14.26 Lemma.** Let  $T$  be a  $k$ -dimensional linear subspace of  $\mathbf{R}^n$ , and let  $(\mathbf{u}_1, \dots, \mathbf{u}_k)$  be an orthonormal basis of  $T$ . If  $\omega \in \Lambda^k(\mathbf{R}^{n*})$  and  $\mathbf{a}_1, \dots, \mathbf{a}_k \in T$ , we have

$$\omega(\mathbf{a}_1, \dots, \mathbf{a}_k) = \pm V(\mathbf{a}_1, \dots, \mathbf{a}_k) \omega(\mathbf{u}_1, \dots, \mathbf{u}_k);$$



the sign is positive if  $(\mathbf{a}_1, \dots, \mathbf{a}_k)$  is a basis of  $T$  having the same orientation as  $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ , and negative if  $(\mathbf{a}_1, \dots, \mathbf{a}_k)$  is a basis of  $T$  having the opposite orientation to that of  $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ . If  $(\mathbf{a}_1, \dots, \mathbf{a}_k)$  is linearly dependent, the sign is irrelevant, since both sides are zero. In particular, if  $(\mathbf{a}_1, \dots, \mathbf{a}_k)$  is also orthonormal, then  $\omega(\mathbf{a}_1, \dots, \mathbf{a}_k) = \pm\omega(\mathbf{u}_1, \dots, \mathbf{u}_k)$ , the sign being determined by the orientations.

**Proof.** Express each  $\mathbf{a}_j$  as a linear combination of  $\mathbf{u}_1, \dots, \mathbf{u}_k$ :

$$\mathbf{a}_j = \sum_i b_j^i \mathbf{u}_i \quad (j = 1, \dots, k)$$

so that

$$\begin{aligned} \omega(\mathbf{a}_1, \dots, \mathbf{a}_k) &= \sum b_1^{j_1} b_2^{j_2} \cdots b_k^{j_k} \omega(\mathbf{u}_{j_1}, \dots, \mathbf{u}_{j_k}) \\ &= \sum \varepsilon_{j_1 \dots j_k}^{1 \dots k} b_1^{j_1} b_2^{j_2} \cdots b_k^{j_k} \omega(\mathbf{u}_1, \dots, \mathbf{u}_k) \\ &= \det B \omega(\mathbf{u}_1, \dots, \mathbf{u}_k), \end{aligned}$$

where, of course,  $B$  is the  $k \times k$  matrix  $(b_j^i)$ , and we used the “Kronecker epsilon” symbol:

$$\varepsilon_{1 \dots k}^{j_1 \dots j_k} = \varepsilon(\sigma)$$

if there exists a permutation  $\sigma$  such that  $j_i = \sigma(i)$  for  $i = 1, \dots, k$ , and otherwise equals 0. If  $(\mathbf{a}_1, \dots, \mathbf{a}_k)$  is a linearly independent sequence, then  $\det B > 0$  if and only if it has the same orientation as  $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ . In particular, if  $(\mathbf{a}_1, \dots, \mathbf{a}_k)$  is orthonormal, then  $B$  is an orthogonal  $k \times k$  matrix, and  $\det B = \pm 1$ , according to whether it preserves orientation or not. In general, we can find an orthogonal  $n \times n$  matrix  $O$  such that  $O\mathbf{u}_j = \mathbf{e}_j$  for  $j = 1, \dots, k$ . Then

$$(\det B)^2 = \det(B^t B) = V^2(O\mathbf{a}_1, \dots, O\mathbf{a}_k) = V^2(\mathbf{a}_1, \dots, \mathbf{a}_k).$$

Thus  $|\det B| = V(\mathbf{a}_1, \dots, \mathbf{a}_k)$ . ■

Now suppose that  $M$  is a compact oriented  $k$ -manifold in  $\mathbf{R}^n$ , and  $\omega$  a continuous  $k$ -form on  $M$ . Let  $F_\omega$  be the real-valued function on  $M$  defined by  $F_\omega(\mathbf{p}) = \omega(\mathbf{u}_1, \dots, \mathbf{u}_k)$ , where  $(\mathbf{u}_1, \dots, \mathbf{u}_k)$  is any positively oriented orthonormal base for  $T_{\mathbf{p}}(M)$ . The last lemma shows that this number is independent of the choice of basis. Since we can find  $\mathbf{u}_j(\mathbf{p})$  which depend smoothly on  $\mathbf{p} \in M$  and are orthonormal at each  $\mathbf{p}$  in a coordinate patch  $U_\alpha$  by applying the Gram-Schmidt process to the columns of  $\alpha'(t)$  (with  $\alpha(t) = \mathbf{p}$ ), we see that  $F_\omega$  is a continuous function on  $M$ . The last lemma shows that if  $\alpha$  is a local coordinate for  $M$ , then

$$\omega_{\alpha(t)}(\alpha'(t)\mathbf{e}_1, \dots, \alpha'(t)\mathbf{e}_k) = V(\alpha'(t))F_\omega(\alpha(t)),$$

and it follows that

$$\int_M \omega = \int_M F_\omega dV,$$

so that integrals of forms can be viewed as integrals of functions.

## 14.5 Vector Analysis

Let  $\Gamma$  be an oriented  $(n-1)$ -manifold in  $\mathbf{R}^n$  (also called a hypersurface). Let  $\alpha$  be a (positively oriented) local coordinate for  $\Gamma$ . In Chapter 11 we showed there exists a smooth unit normal vector field  $\nu$  on  $\Gamma$ ; now we construct  $\nu$  more explicitly. Let  $\mathbf{a}_j = \mathbf{a}_j(\mathbf{p}) = \alpha'(\mathbf{t})\mathbf{e}_j$  for  $\mathbf{p} = \alpha(\mathbf{t})$ ,  $\mathbf{p} \in U_\alpha$ , and let  $A$  be the  $n \times (n-1)$  matrix  $[\mathbf{a}_1 \cdots \mathbf{a}_{n-1}]$ . Thus  $A$  is a smooth matrix-valued function in  $U_\alpha$ . Let  $A^i$  be the  $(n-1) \times (n-1)$  matrix obtained by deleting the  $i$ th row of  $A$  (thus,  $A^i = A^j$  with our previous notation, where  $I = (1, \dots, \hat{i}, \dots, k)$ .) Define the vector field  $\mathbf{v}_\alpha$  on  $U_\alpha$  by

$$\mathbf{v}_\alpha = \sum_{j=1}^n (-1)^{j-1} (\det A^j) \mathbf{e}_j.$$

**14.27 Proposition.** *The vector field  $\mathbf{v} = \mathbf{v}_\alpha$  is smooth in  $U_\alpha$  and  $|\mathbf{v}| = V(A) > 0$ ; for each  $\mathbf{p} \in U_\alpha$ ,  $\mathbf{v}(\mathbf{p}) \perp T_{\mathbf{p}}(\Gamma)$ ;  $(\mathbf{v}, \mathbf{a}_1, \dots, \mathbf{a}_{n-1})$  is a positively oriented basis of  $\mathbf{R}^n$  at each point of  $U_\alpha$ .*

**Proof.** The smoothness of  $\mathbf{v}$  is obvious, and  $|\mathbf{v}|^2 = \sum_j |\det A^j|^2 = V(A)$  by Proposition 14.23. Let  $B = [\mathbf{v}, \mathbf{a}_1, \dots, \mathbf{a}_{n-1}]$ . Expanding by cofactors of the first column, we find

$$\det B = \sum_{j=1}^n (-1)^{j-1} v^j \det A^j = \sum_{j=1}^n (v^j)^2 = |\mathbf{v}|^2 > 0,$$

so  $(\mathbf{v}, \mathbf{a}_1, \dots, \mathbf{a}_{n-1})$  is positively oriented. For any  $i$ ,  $1 \leq i \leq n-1$ , we have

$$\mathbf{a}_i \cdot \mathbf{v} = \sum_{j=1}^n a_i^j (-1)^{j-1} \det A^j = \det [\mathbf{a}_i \quad \mathbf{a}_1 \quad \cdots \quad \mathbf{a}_{n-1}] = 0.$$

Since  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$  form a basis for  $T_{\mathbf{p}}(\Gamma)$ , we have  $\mathbf{v} \perp T_{\mathbf{p}}(\Gamma)$ . ■

It follows that  $\nu = \mathbf{v}/|\mathbf{v}|$  is the unique unit normal vector to  $\Gamma$  which makes  $(\nu, \mathbf{a}_1, \dots, \mathbf{a}_{n-1})$  a positively oriented basis for  $\mathbf{R}^n$ . We have now prepared the machinery to interpret Stokes' theorem in terms of vector fields when  $M$  is an  $n$ -manifold in  $\mathbf{R}^n$ . The following theorem is known as the *divergence theorem*:

**14.28 Theorem.** *If  $M$  is a compact  $n$ -manifold in  $\mathbf{R}^n$  and  $\mathbf{g}$  is a smooth vector field on  $M$ , then*

$$\int_M \operatorname{div} \mathbf{g} \, dV = \int_{\partial M} \mathbf{g} \cdot \nu \, dV,$$

where  $\nu$  is the outward unit normal vector on  $\partial M$ .

**Proof.** Let (as before)

$$\eta^j = (-1)^{j-1} dx^1 \wedge \cdots \wedge \widehat{dx^j} \wedge \cdots \wedge dx^n,$$

for  $j = 1, \dots, n$  and let  $\omega = \sum_j g_j \eta^j$ . Thus  $\omega$  is a smooth  $(n-1)$ -form on  $M$ , and

$$d\omega = \sum_{j=1}^n dg_j \wedge \eta^j = \left( \sum_{j=1}^n \frac{\partial g_j}{\partial x^j} \right) dx^1 \wedge \cdots \wedge dx^n = (\operatorname{div} \mathbf{g}) dx^1 \wedge \cdots \wedge dx^n.$$

Thus  $\int_M \operatorname{div} \mathbf{g} dV = \int_M d\omega$ . We may suppose that the restriction to  $\partial M$  of  $\mathbf{g} = (g_1, \dots, g_n)$  vanishes outside a compact subset of a coordinate patch  $U_\alpha$ , where  $\alpha$  is a local coordinate for  $\partial M$ . We observe that

$$\eta^j(\mathbf{a}_1, \dots, \mathbf{a}_{n-1}) = (-1)^{j-1} \det A^j$$

for any  $(n-1)$  vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$  in  $\mathbf{R}^n$ ; we take  $\mathbf{a}_j = \alpha'(t)\mathbf{e}_j$ , so  $A = \alpha'(t)$ . Then we have

$$\begin{aligned} \int_{\partial M} \omega &= \int_{\partial M} \left( \sum_j g_j \eta^j \right) \\ &= \int_{V_\alpha} \sum_j g_j(\alpha(t)) \eta^j(\alpha'(t)) dm(t) \\ &= \int_{V_\alpha} \sum_j g_j(\alpha(t)) (-1)^{j-1} \det A^j(t) dm(t) \\ &= \int_{V_\alpha} \sum_j (g_j(\alpha(t)) v^j(\alpha(t))) dm(t) \\ &= \int_{\partial M} \mathbf{g} \cdot \nu dV. \end{aligned}$$

According to Stokes' theorem,  $\int_M d\omega = \int_{\partial M} \omega$ . ■

We close this section with an application of the divergence theorem to the Laplace operator.

**14.29 Definition.** Let  $f$  be a function of class  $C^2$  in an open set in  $\mathbf{R}^n$ . We define  $\Delta f$ , the Laplacian of  $f$ , by

$$\Delta f = \sum_{j=1}^n D_j^2 f = \operatorname{div} \operatorname{grad} f.$$

We say  $f$  is harmonic if  $\Delta f = 0$ .

The following formulas are known as *Green's identities*.

**14.30 Theorem.** Let  $M$  be a compact  $n$ -manifold in  $\mathbf{R}^n$ , and let  $u$  and  $v$  be smooth functions in an open set containing  $M$ . Let  $\nu$  denote the unit normal vector field on  $\partial M$ . Then

$$\int_{\partial M} v \frac{\partial u}{\partial \nu} dV = \int_M (v \Delta u + \nabla v \cdot \nabla u) dV, \quad (14.1)$$

and

$$\int_{\partial M} \left( u \frac{\partial v}{\partial \nu} - v \frac{\partial u}{\partial \nu} \right) dV = \int_M (u \Delta v - v \Delta u) dV. \quad (14.2)$$

Here,  $\partial u / \partial \nu = \nabla u \cdot \nu$ ; this is called the normal derivative of  $u$ .

**Proof.** Observe that  $\operatorname{div}(v \nabla u) = \nabla v \cdot \nabla u + v \Delta u$ . Apply the divergence theorem (Theorem 14.28) to the vector field  $v \nabla u$  to get the first formula; interchange the roles of  $u$  and  $v$ , and subtract, to get the second. ■

## 14.6 Harmonic Functions

Throughout this section, we let  $M$  denote a compact  $n$ -manifold in  $\mathbf{R}^n$ , so that  $D = \operatorname{int} M$  is a bounded open set, and  $\Gamma = \partial M$  is a compact  $(n-1)$ -manifold. We will develop the basic properties of functions which are harmonic in  $D$  and smooth on  $\bar{D} = M$ .

**14.31 Lemma.** If  $u$  is a smooth function on  $M$  which is harmonic in  $D$ , then

$$\int_{\Gamma} \frac{\partial u}{\partial \nu} dV = 0.$$

**Proof.** Take  $v = 1$  in Green's formula (14.1). ■

Let  $\omega_n = V(B^n) = m(B^n)$  be the  $n$ -dimensional volume of the unit ball in  $\mathbf{R}^n$ . (We computed the value of  $\omega_n$  in an exercise in Chapter 10.) Let  $\sigma_n = V_{n-1}(S^{n-1})$  be the  $(n-1)$ -dimensional volume of the unit sphere in  $\mathbf{R}^n$ . We will write  $B(\mathbf{p}, r)$  for the closed ball of center  $\mathbf{p}$  and radius  $r$ , and  $S(\mathbf{p}, r)$  for its boundary; we will further abbreviate  $B(\mathbf{0}, r)$  to  $B(r)$ , and  $S(\mathbf{0}, r)$  to  $S(r)$ . We note that  $V(B(\mathbf{p}, r)) = \omega_n r^n$  and  $V(S(\mathbf{p}, r)) = \sigma_n r^{n-1}$ .

**14.32 Lemma.** For each  $n \in \mathbf{N}$ ,  $\sigma_n = n\omega_n$ . If  $u$  is harmonic in an open set containing  $B(r)$ , then

$$\frac{1}{V(S(r))} \int_{S(r)} u dV_{n-1} = \frac{1}{V(B(r))} \int_{B(r)} u dV_n.$$

**Proof.** Let  $v(\mathbf{x}) = |\mathbf{x}|^2$ . Applying Green's identity (14.2) with these functions, taking  $M = B(r)$ , and observing that  $\Delta v = 2n$ , and that  $\partial v / \partial \nu = 2r$ , we find that

$$\int_{S(r)} \left( 2ru - r^2 \frac{\partial u}{\partial \nu} \right) dV_{n-1} = \int_{B(r)} (2nu - v \Delta u) dV_n,$$

and taking hold of  $\Delta u = 0$  and Lemma 14.31, we see that

$$r \int_{S(r)} u dV_{n-1} = n \int_{B(r)} u dV_n. \quad (14.3)$$

Taking  $u = 1$  and  $r = 1$  here, we have  $\sigma_n = n\omega_n$ , and dividing both sides of (14.3) by  $V(B(r)) = \omega_n r^n$  we obtain the lemma. ■

The next result is known as the mean value theorem for harmonic functions. We write  $d\sigma$  for  $dV_{n-1}$  on  $S(r)$ .

**14.33 Theorem.** *Let  $u$  be harmonic in  $D$ , and suppose  $\mathbf{p} \in D$  and  $r > 0$ , with  $\overline{B}(\mathbf{p}, r) \subset D$ . Then*

$$u(\mathbf{p}) = \frac{1}{\sigma_n r^{n-1}} \int_{S(\mathbf{p}, r)} u d\sigma = \frac{1}{\omega_n r^n} \int_{B(\mathbf{p}, r)} u dm.$$

**Proof.** We may assume without loss of generality that  $\mathbf{p} = \mathbf{0}$ . Suppose  $n > 2$ . Let  $v(\mathbf{x}) = |\mathbf{x}|^{2-n}$ ; it is not hard to verify that  $\Delta v = 0$  in  $\mathbf{R}^n \setminus \{\mathbf{0}\}$ . Fix  $\epsilon$ ,  $0 < \epsilon < r$ , and let  $M_\epsilon = B(r) \setminus \text{int } B(\epsilon)$ , so that  $M_\epsilon$  is an  $n$ -manifold in  $\mathbf{R}^n$  and  $\partial M_\epsilon = S(r) \cup S(\epsilon)$ . Applying Green's identity (14.2), we get

$$\int_{\partial M_\epsilon} \left( u \frac{\partial v}{\partial \nu} - v \frac{\partial u}{\partial \nu} \right) d\sigma = 0,$$

since  $u$  and  $v$  are each harmonic in  $M_\epsilon$ . Since

$$\int_{S(t)} \frac{\partial u}{\partial \nu} d\sigma = 0$$

for  $0 < t \leq r$  by Lemma 14.31, and  $v$  is constant on  $S(r)$  and on  $S(\epsilon)$ , we have

$$\int_{\partial M_\epsilon} v \frac{\partial u}{\partial \nu} d\sigma = 0.$$

Now the outward normal vector  $\nu$  is equal to  $\mathbf{x}/|\mathbf{x}|$  on  $S(r)$  and  $-\mathbf{x}/|\mathbf{x}|$  on  $S(\epsilon)$ , so  $\partial v / \partial \nu = (2-n)r^{1-n}$  on  $S(r)$  and  $-(2-n)\epsilon^{1-n}$  on  $S(\epsilon)$ . Thus we have

$$r^{1-n} \int_{S(r)} u d\sigma = \epsilon^{1-n} \int_{S(\epsilon)} u d\sigma,$$

for every  $\epsilon \in (0, r)$ , or

$$\frac{1}{\sigma(S(r))} \int_{S(r)} u \, d\sigma = \frac{1}{\sigma(S(\epsilon))} \int_{S(\epsilon)} u \, d\sigma \quad (14.4)$$

for every  $\epsilon \in (0, r)$ . But

$$\begin{aligned} \left| \frac{1}{\sigma(S(\epsilon))} \int_{S(\epsilon)} u \, d\sigma - u(\mathbf{0}) \right| &= \left| \frac{1}{\sigma(S(\epsilon))} \int_{S(\epsilon)} (u - u(\mathbf{0})) \, d\sigma \right| \\ &\leq \sup_{|\mathbf{x}| \leq \epsilon} |u(\mathbf{x}) - u(\mathbf{0})|, \end{aligned}$$

and since  $u$  is continuous at  $\mathbf{0}$  we conclude that

$$\frac{1}{\sigma(S(\epsilon))} \int_{S(\epsilon)} u \, d\sigma \rightarrow u(\mathbf{0}) \quad \text{as } \epsilon \rightarrow 0.$$

This, combined with (14.4) above, proves the first equation asserted by the theorem, and the second follows from Lemma 14.32. Thus the theorem is proved for  $n > 2$ . If  $n = 2$ , then for  $v = |x|^{2-n} = 1$ , we have unfortunately that the normal derivatives of  $v$  vanish identically, so this argument fails; but taking  $v = \log |x|$ , we can repeat the argument word for word to get the desired conclusion. ■

One consequence of the mean value property of harmonic functions, expressed in Theorem 14.33, is the *maximum principle*.

**14.34 Theorem.** *Let  $u$  be harmonic in the connected open subset  $D$  of  $\mathbf{R}^n$ , and suppose that  $u$  attains a maximum or a minimum at some point  $\mathbf{p}$  of  $D$ . Then  $u$  is constant in  $D$ .*

**Proof.** We may assume that  $u$  has a minimum at  $\mathbf{p}$  (else consider  $-u$ ), and that  $u(\mathbf{p}) = 0$  (else consider  $u - u(\mathbf{p})$ ). Let  $V = \{\mathbf{q} \in D : u(\mathbf{q}) = 0\}$ . Then  $V$  is closed in  $D$  since  $u$  is continuous. Furthermore, if  $\mathbf{q} \in V$ , and  $\delta > 0$  is small enough so that  $B(\mathbf{q}, \delta) \subset D$ , we have

$$\int_{B(\mathbf{q}, \delta)} u \, dm = 0$$

by Theorem 14.33, but  $u \geq 0$  in  $B(\mathbf{q}, \delta)$  (indeed, throughout  $D$ ), so we have  $u = 0$  a.e. ( $m$ ) in  $B(\mathbf{q}, \delta)$ . Since  $u$  is continuous,  $\{\mathbf{t} \in B(\mathbf{q}, \delta) : u(\mathbf{t}) > 0\}$  is open, so we conclude this set is empty, i.e.,  $B(\mathbf{q}, \delta) \subset V$ . Thus,  $V$  is open as well as closed in  $D$ , and since  $D$  is connected and  $V$  is nonempty, we conclude that  $V = D$ . ■

Another important consequence of the mean value property is the following, known as *Liouville's theorem*:

**14.35 Theorem.** Suppose  $u$  is harmonic and bounded in all of  $\mathbf{R}^n$ . Then  $u$  is a constant function.

**Proof.** Let  $\mathbf{p} \in \mathbf{R}^n$ ; for each  $r > 0$ , we have

$$\begin{aligned} |u(\mathbf{p}) - u(\mathbf{0})| &= \frac{1}{\omega_n r^n} \left| \int_{B(\mathbf{p}, r)} u \, dm - \int_{B(\mathbf{0}, r)} u \, dm \right| \\ &= \frac{1}{\omega_n r^n} \left| \int_{B(\mathbf{p}, r) \setminus B(\mathbf{0}, r)} u \, dm - \int_{B(\mathbf{0}, r) \setminus B(\mathbf{p}, r)} u \, dm \right| \\ &\leq \frac{1}{\omega_n r^n} 2C\lambda(r), \end{aligned}$$

where  $C$  is an upper bound for  $|u|$  and we put  $\lambda(r) = m(B(\mathbf{0}, r) \setminus B(\mathbf{p}, r))$ . (So  $\lambda(r) = m(B(\mathbf{p}, r) \setminus B(\mathbf{0}, r))$  as well.) Now for all  $r > \rho = |\mathbf{p}|$ , we have  $B(\mathbf{0}, r) \supset B(\mathbf{p}, r - \rho)$ , so  $\lambda(r) \leq m(B(r)) - m(B(r - \rho)) = \omega_n(r^n - (r - \rho)^n)$ . Thus

$$|u(\mathbf{p}) - u(\mathbf{0})| \leq 2C \frac{r^n - (r - \rho)^n}{r^n} = 2C[1 - (1 - (\rho/r)^n)]$$

for every  $r > |\mathbf{p}|$ , so  $u(\mathbf{p}) = u(\mathbf{0})$  for all  $\mathbf{p} \in \mathbf{R}^n$ . ■

The next result gives a characterization of harmonicity in terms of the first-order partial derivatives of a function.

**14.36 Lemma.** Let  $u$  be a smooth function on  $M$ . Then  $u$  is harmonic in  $D$  if and only if

$$\int_M \nabla u \cdot \nabla v \, dV = \int_D \nabla u \cdot \nabla v \, dm = 0$$

for every smooth  $v$  which vanishes on  $\Gamma$ .

**Proof.** Suppose  $u$  is harmonic, and  $v$  is smooth and vanishes on  $\Gamma$ . Then from Green's formula (14.1)

$$\int_{\Gamma} v \frac{\partial u}{\partial \nu} \, dV_{n-1} = \int_M (\nabla u \cdot \nabla v + v \Delta u) \, dV_n$$

we deduce that  $\int_M \nabla u \cdot \nabla v \, dV_n = 0$ . Conversely, if  $u$  is smooth and this equation holds whenever  $v$  is smooth and vanishes on  $\Gamma$ , we conclude that  $\int_M v \Delta u \, dV_n = 0$  for every smooth  $v$  vanishing outside a compact subset of  $D$ . If  $\Delta u$  is positive (say) at some point of  $D$ , it is positive throughout some disk  $A \subset D$ . But as we have seen, there exists a smooth function  $v$  which is positive in  $A$  and vanishes outside  $A$ , which gives  $\int_D v \Delta u \, dm > 0$ . It follows that  $\Delta u = 0$  everywhere in  $D$ . ■

Our final theorem is usually referred to as the *Dirichlet principle*. It seems to be due to Riemann.

**14.37 Theorem.** Let  $f$  be a smooth function defined on  $\Gamma$ , and let  $\mathcal{F}$  be the class of all smooth functions  $u$  on  $M$  whose restriction to  $\Gamma$  is  $f$ . If  $u \in \mathcal{F}$ , then  $u$  is harmonic in  $D$  (i.e.,  $u$  solves the Dirichlet problem with data  $f$ ) if and only if  $u$  minimizes the Dirichlet integral in the class  $\mathcal{F}$ :

$$\int_M |\nabla u|^2 dm \leq \int_M |\nabla v|^2 dm \quad \text{for every } v \in \mathcal{F}.$$

**Proof.** Let  $u \in \mathcal{F}$ . If  $v$  is smooth on  $M$ , then  $v \in \mathcal{F}$  if and only if  $w = v - u$  vanishes on  $\Gamma$ . We have

$$\int_M |\nabla v|^2 dm = \int_M |\nabla(u + w)|^2 dm = \int_M (|\nabla u|^2 + |\nabla w|^2 + 2\nabla u \cdot \nabla w) dm$$

so that if  $u$  is harmonic, we see by Lemma 14.36 that the last term on the right above vanishes, so

$$\int |\nabla u|^2 dm \leq \int |\nabla v|^2 dm$$

for every  $v \in \mathcal{F}$ . Conversely, suppose that this last inequality holds. Then for every  $w$  vanishing on  $\Gamma$ , the function

$$F(t) = \int_M |\nabla(u + tw)|^2 dm$$

attains its minimum value at  $t = 0$ . But

$$F(t) = \int_M (|\nabla u|^2 + 2t\nabla u \cdot \nabla w + t^2|\nabla w|^2) dm,$$

so  $F'(0) = 2 \int_M \nabla u \cdot \nabla w dm = 0$ , and since this holds for every smooth  $w$  vanishing on  $\Gamma$ , Lemma 14.36 tells us that  $u$  is harmonic in  $D$ . ■

## 14.7 Exercises

1. Let  $M$  be a one-dimensional oriented manifold in  $\mathbf{R}^n$ . (In fact, every one-dimensional manifold can be oriented, but we don't prove this here.) Show that there exists a unit tangent vector field  $\mathbf{t}$  on  $M$ ; show that any such field defines an orientation of  $M$ .
2. Let  $M$  be an oriented one-dimensional manifold in  $\mathbf{R}^n$ . It is usual to write  $ds$  instead of  $dV$  when  $k = 1$ . Let  $\omega$  be a smooth 1-form defined in an open set containing  $M$ , and let  $\mathbf{g}$  be the associated vector field ( $\mathbf{g} = \Phi(\omega)$  in the notation of Chapter 11). Show that

$$\int_M \omega = \int_M \mathbf{g} \cdot \mathbf{t} ds.$$



3. Let  $M$  be a compact oriented 2-manifold in  $\mathbf{R}^3$ , and let  $\mathbf{g}$  be a smooth vector field defined in an open set containing  $M$ . Prove the classical theorem of Stokes:

$$\int_{\partial M} \mathbf{g} \cdot \mathbf{t} \, ds = \int_M (\operatorname{curl} \mathbf{g}) \cdot \boldsymbol{\nu} \, d\sigma,$$

where  $ds$  is the one-dimensional volume element on  $\partial M$ ,  $d\sigma$  is the two-dimensional volume element on  $M$ , and  $\mathbf{t}$  and  $\boldsymbol{\nu}$  are the canonical unit tangent vector on  $\partial M$  and unit normal vector on  $M$ , respectively.

4. Let  $M$  be a compact oriented  $k$ -manifold in  $\mathbf{R}^n$ . Suppose that  $\omega$  is a smooth  $(k-1)$ -form on  $M$ , with  $d\omega = 0$ , and  $u$  is a smooth function on  $M$  which vanishes on  $\partial M$ . Show that

$$\int_M du \wedge \omega = 0.$$

5. Let  $M$  be a compact oriented  $k$ -manifold in  $\mathbf{R}^n$ . Suppose that  $\omega \in \Omega^i(M)$  and  $\eta \in \Omega^j(M)$ , where  $i + j + 1 = k$ , and suppose that  $\eta$  vanishes on  $\partial M$ . Show that

$$\int_M \omega \wedge d\eta = (-1)^{i-1} \int_M d\omega \wedge \eta.$$

6. Let  $U = \mathbf{R}^n \setminus \{0\}$ , and let  $S = S^{n-1}$  be the unit sphere in  $\mathbf{R}^n$ .

(a) Show that if  $\omega \in \Omega^{n-1}(U)$  is exact, then  $\int_S \omega = 0$ .

(b) Construct an example of a closed  $\omega \in \Omega^{n-1}(U)$  such that  $\int_S \omega \neq 0$ .

7. Let  $M$  be the Möbius strip, as described in Example 11.14 in the chapter on manifolds. Let

$$\omega = \frac{x \, dy - y \, dx}{x^2 + y^2},$$

so  $\omega$  is a smooth 1-form in an open set of  $\mathbf{R}^3$  which contains  $M$ . Show that  $d\omega = 0$ , but that  $\int_{\partial M} \omega = 4\pi$ . Reconcile this result with Stokes' theorem.

8. Use the Brouwer fixed point theorem to prove the following theorem of Perron-Frobenius: if  $A = [a_{ij}]$  is an  $n \times n$  matrix with nonnegative entries, then there exists  $\mathbf{v} \in \mathbf{R}^n$  with  $v^j \geq 0$  for all  $j$  and  $\sum_{j=1}^n v^j = 1$  such that  $A\mathbf{v} = \mathbf{v}$ . (It might be convenient to prove this first with the additional assumption that  $\sum_{j=1}^n a_{ij} = 1$  for every  $i$ .)

9. Use the formulas of this chapter to find the two-dimensional volume (i.e., surface area) of the torus described in the chapter on manifolds.

10. Let  $M$  be a compact  $n$ -manifold in  $\mathbf{R}^n$ , so  $M$  is the closure of a bounded open set  $D$ , and  $\Gamma = \partial M$  is a compact hypersurface in  $\mathbf{R}^n$ . Evaluate  $\int_{\Gamma} \mathbf{x} \cdot \boldsymbol{\nu}$ .

11. Find the compact 2-manifold  $M$  in  $\mathbf{R}^2$  with area  $\pi$  for which

$$\int_{\partial M} y^3 dx + (3x - x^3) dy$$

is maximal.

12. Find all functions  $f$  which are harmonic in  $\mathbf{R}^n \setminus \{0\}$  and which have the form  $f(|\mathbf{x}|) = g(|\mathbf{x}|)$  for some  $g$  which is smooth on  $(0, \infty)$ . (Functions of this form are called “radial.”)

13. Let  $u$  be a smooth function with compact support on  $\mathbf{R}^2$ . Show that for all  $\mathbf{x} \in \mathbf{R}^2$ , we have

$$u(\mathbf{x}) = \frac{1}{2\pi} \int_{\mathbf{R}^2} \Delta u(\mathbf{y}) \log |\mathbf{x} - \mathbf{y}| dm(\mathbf{y}).$$

HINT: Apply Green’s theorem, with  $M = \{\mathbf{y} : \delta \leq |\mathbf{y} - \mathbf{x}| \leq R\}$ , where  $R$  is large and  $\delta > 0$  is small; let  $\delta \rightarrow 0$ .

14. Let  $g$  be a smooth function on  $\mathbf{R}^2$  with compact support, and define  $u$  by

$$u(\mathbf{x}) = \frac{1}{2\pi} \int_{\mathbf{R}^2} g(\mathbf{y}) \log |\mathbf{x} - \mathbf{y}| dm(\mathbf{y}).$$

(a) Show that  $u$  is smooth on  $\mathbf{R}^2$ . HINT. Use dominated convergence to justify differentiating under the integral sign.

(b) Show that  $\Delta u = g$ .

15. Let  $n \geq 3$ , and let  $u$  be a smooth function vanishing outside a compact subset of  $\mathbf{R}^n$ . Show that for all  $\mathbf{x} \in \mathbf{R}^n$ ,

$$u(\mathbf{x}) = c_n \int_{\mathbf{R}^n} \frac{\Delta u(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|^{n-2}} dm(\mathbf{y});$$

here  $c_n$  is a constant depending only on  $n$ . HINT: Apply Green’s theorem, with  $M = \{\mathbf{y} : \delta \leq |\mathbf{y} - \mathbf{x}| \leq R\}$ , where  $R$  is large and  $\delta > 0$  is small; let  $\delta \rightarrow 0$ .

16. Let  $g$  be a smooth function on  $\mathbf{R}^n$ ,  $n \geq 3$ , vanishing outside a compact set, and define  $u$  by

$$u(\mathbf{x}) = c_n \int_{\mathbf{R}^n} \frac{g(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|^{n-2}} dm(\mathbf{y}),$$

where  $c_n$  is the constant of the last problem.

(a) Show that  $u$  is smooth in  $\mathbf{R}^n$ . HINT: Use dominated convergence to justify differentiating under the integral sign.

(b) Show that  $\Delta u = g$ .

## 14.8 Notes

The earliest form of what we know as Stokes' theorem (in the language of this chapter, the case of a 2-manifold in  $\mathbf{R}^2$ ), is due to George Green, a self-taught mathematician, who published it in 1828. It was rediscovered 11 years later by Gauss (Green's work was not widely known before Thomson had it reprinted in 1846). It was also rediscovered by Ostrogradski, who seems to be the first to publish the divergence theorem for domains in 3-space. According to Spivak [13], the first statement of the classical Stokes' theorem (the case of a 2-manifold in  $\mathbf{R}^3$ ) is found in a letter from Thomson (later Lord Kelvin) to Stokes, in 1850. Stokes gave the theorem as a problem in a prize examination at Cambridge, beginning in 1854, and by the time of his death the result was universally known as Stokes' theorem. Proofs were published by Thomson, and by Maxwell in his treatise on electricity and magnetism (the subject also of Green's essay.) The theorem is also attributed to Ampère.

The concept of partition of unity is due to Dieudonné; it is I think a very elegant way to avoid the difficulties of partitioning the manifold into small pieces; in specific instances, computation is however more practical by the latter procedure.

Brouwer proved his fixed point theorem around the year 1910. It was one of the starting points for the twentieth century development of topology, and has found many applications in analysis.

It can be shown that the  $k$ -dimensional volume measure we defined on a  $k$ -manifold  $M$  in  $\mathbf{R}^n$  is in fact (up to a constant multiple) the  $k$ -dimensional Hausdorff measure restricted to  $M$ .

The study of harmonic functions is also known as potential theory (the word potential in this context originating again with Green). It originated in the eighteenth century, flourished in the nineteenth century with the results described in this chapter and many others, and is still going strong. The proof of Liouville's theorem that we gave was found in the 1960s by Nelson.



# References

- [1] D. Bressoud. *A Radical Approach to Real Analysis*. Mathematical Association of America, Washington, D.C., 1994.
- [2] P. Halmos. *Naive Set Theory*. D. Van Nostrand, Princeton, NJ, 1960.
- [3] G. H. Hardy. *A Course of Pure Mathematics*. Cambridge University Press, Cambridge, UK, 9th edition, 1949.
- [4] G.H. Hardy and E.M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, New York, 4th edition, 1960.
- [5] I.N. Herstein. *Topics in Algebra*. Wiley, New York, 2nd edition, 1975.
- [6] J.L. Kelley. *General Topology*. Van Nostrand, Princeton, NJ, 1955.
- [7] A.Ya. Khinchin. *Continued Fractions*. University of Chicago Press, Chicago, 3rd edition, 1964.
- [8] T.W. Körner. *Fourier Analysis*. Cambridge University Press, Cambridge, UK, 1988.
- [9] J.W. Milnor. *Topology from the Differentiable Viewpoint*. University Press of Virginia, Charlottesville, 1965.
- [10] J.R. Munkres. *Analysis on Manifolds*. Addison-Wesley, Redwood City, CA, 1991.
- [11] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, 3rd edition, 1976.

- [12] G.E. Shilov and B.L. Gurevich. *Integral, Measure and Derivative: A Unified Approach*. Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [13] M. Spivak. *Calculus on Manifolds*. Addison-Wesley, Redwood City, CA, 1965.
- [14] O. Toeplitz. *Calculus: A Genetic Approach*. University of Chicago Press, Chicago, 1963.
- [15] E.T. Whittaker and G.N. Watson. *A Course of Modern Analysis*. Cambridge University Press, Cambridge, UK, 4th edition, 1962.

# Index

- a.a., *see* almost all
- Abel, N., 53, 72
- absolutely convergent
  - integral, 116
  - series, 40
- additive
  - countably, 206
  - finitely, 203
- adjoint map, 280
- a.e., *see* almost everywhere
- Agnesi, M., 96
- Alexandroff, P., 153
- algebra
  - generated by, 205
  - of sets, 202
- almost all, 221, 226
- almost everywhere, 226
- alternating tensor, 272
- alternating tensors
  - basis for, 276
  - dimension of, 276
  - in  $\mathbf{R}^3$ , 279
- alternation operator, 275
- Ampère, A.M., 321
- antiderivative, 110
- antisymmetry, 273
- Arbogast, L., 95
- Archimedean property, 15
- Archimedes, 26, 95
- Arzelà, C., 154
- Arzelà–Ascoli theorem, 145, 146
- Ascoli, G., 154
- axiom of choice, 217, 218
- $B^n$ , 257
- Baire, R., 153
- Baire theorem, 136
- ball, 128
- Barrow, I., 122
- basis, 176
  - natural, 176
- Bernoulli, Jakob, 53, 121
- Bernoulli, Johann, 96
- Bernstein, S., 174
- bijective, 3
- Bolzano, B., 53, 72, 153
- Bolzano–Weierstrass property, 144
- Bolzano–Weierstrass theorem, 35
- Bonnet, O., 96

Borel, É., 202, 221

Borel

algebra, 205

measure, 206

set, 205, 214, 215, 220, 224

Borel measure

translation invariant,  
uniqueness of, 239

Borel-Cantelli lemma, 222

bound

lower, 13

upper, 13

boundary, 125

of manifold, 256

bounded, 142

totally, 142

Brouwer, L.E.J., 321

Bunyakovsky, V., 109

$C^n$ ,  $C^\infty$ , 88

$C^r$ , 254

Cantor, G., 26, 27, 153, 174

Cantor function, 248

Cantor set, 149, 201, 248

dimension of, 211

measure of, 213

Carathéodory, C., 210, 221

cardinality, 6

Cartan, E., 284

Cartesian product, 2, 133

Cauchy, A., 72, 96, 121, 122

Cauchy condensation test, 43

Cauchy sequence, 36, 237

chain rule, 82, 184

classical notation, 187

change of variables

linear, 240

smooth, 112, 241

class  $C^r$ , 190

closed, 124

closed ball, 257

closure, 125

column vector, 177

compact, 140, 215

relatively, 144

sequentially, 144

compact set

continuous image of, 141

complement, 2

relative, 1

complete, 236

metric space, 136

ordered field, 14

component, 148

condensation point, 150

conjugate space, 269

connected, 147, 186

connected component, 148

connected set

continuous image, 148

content, 121

continued fraction, 36-39

continuity from below, 206

continuous function, 57

continuous map, 126

contraction map, 137, 193

convergence

almost everywhere, 249

in measure, 249

mean square, 166

monotone, 225

pointwise, 62, 226, 229

uniform, 62, 226

convex coordinate patch, 293

convex function, 70, 78

convex set, 185, 250

convolution, 251

coordinate patch, 255

coset, 218

cosets, 277

countable, 6-8

countable additivity, 219

countable base, 131

countable subadditivity, 219

countably additive, 206

countably infinite, 6

countably subadditive, 209

covector, 269, 283

Cramer's rule, 181

critical point, 251



- critical value, 251
- cross product, 289
- cube, 242
- curl, 289
- curve, 253
- cut-off function, 299
- $\Delta$  (Laplace operator), 314
- $\delta_j^I$ , *see* Kronecker  $\delta$
- $\delta_j^K$ , *see* Kronecker  $\delta$
- Daniell, P.J., 252
- Darboux, G., 84, 96
- Dedekind, R., 21, 26
- deformation, 262
- DeMorgan's laws, 2, 212
- dense, 16, 126, 218, 220
- de Rham cohomology, 291
- derivative, 74, 182
  - directional, 197
  - higher order, 88
  - of convex function, 80, 81
  - of exponential, 77
  - of polynomial, 76
  - of trigonometric function, 77
- determinant, 181, 275
  - Jacobian, 182
- diagonal matrix, 240
- diameter, 142
- diffeomorphism, 254
- differentiable, 74, 182, 254
- differential, 76, 182, 261
  - of form, *see* exterior differential
- differential calculus, 81
- differential form, 285
  - closed, 291
  - exact, 291
- differentiation
  - of series, 114
- Dini, U., 73
- Dini's theorem, 64
- Dirichlet, P.G.L., 53, 72, 174
- Dirichlet principle, 317
- Dirichlet test
  - for integrals, 117
  - for series, 43
- Dirichlet's theorem, 16, 218
- disconnected, 147
  - totally, 149
- distribution function, 248, 251
- divergence, 288
- divergence theorem, 312
- domain, 3
- dominated convergence theorem, 230
- dual basis, 270, 282
- dual space, 269
- dummy variable, 245
- Dynkin's  $\pi$ - $\lambda$  theorem, 238
- $\epsilon$ -net, 143
- $\epsilon(\sigma)$ , *see* sign of permutation
- $\epsilon_j^I$ , *see* Kronecker epsilon
- $e$ , 33, 66
- Einstein, A., 296
- empty set, 2
- equicontinuous, 145
- equivalence classes, 2, 217, 236
- Eudoxus, 26
- Euler, L., 53, 73, 96
- Euler's constant, 113
- expansion
  - binary, 31
  - continued fraction, 38
  - decimal, 31
- exterior differential, 287
- exterior product, 277
  - (anti)commutativity of, 279
  - associativity of, 278
- Fatou's lemma, 230, 248
- Fermat, P., 121
- Fibonacci numbers, 54
- field, 10-13
  - of sets, 221
  - ordered, 12
- finite induction, 4
- finite intersection property, 151
- finitely additive, 219
- fixed point theorem

- Brouwer. 305 307
  - contraction map, 137
- Fourier, J., 121
- Fourier coefficients, 164
  - uniqueness of, 165
- Fourier series, 164
- Fréchet, M., 153, 182, 221
- Fréchet derivative, 182
- Fubini, G., 252
- Fubini-Tonelli theorem, 245-247
- function, 3, 16
  - absolute value, 17
  - Borel measurable, 244
  - characteristic, 225
  - complex, 96
  - continuous, 55, 57
  - convex, 70, 78, 96, 233, 250
  - Dirichlet, 56, 100
  - exponential, 65-66
  - graph of, 3, 251
  - greatest integer, 58
  - harmonic, 314
  - indicator, 225
  - inverse, 85
  - limit of, 55
  - logarithm, 66
  - lower semicontinuous, 151
  - measurable, 223
    - Borel, 226
    - Lebesgue, 226
  - monotone, 56
  - nowhere differentiable
    - continuous, 75
  - periodic, 162
  - piecewise linear, 61
  - polynomial, 17, 58
  - radial, 320
  - rational, 17, 58
  - sign, 17
  - simple, 225-228, 232
  - step, 232
  - trigonometric, 66-69
  - uniformly continuous, 61
  - upper semicontinuous, 70, 151
- fundamental theorem of calculus, 110, 111
- Gauss, C.F., 54, 121, 267
- geodesics, 161
- gradient, 288
- Gram determinant, 308
- graph, 256
- Grassman, H., 283
- Green, G., 321
- Green's identities, 314
- group, 273
- Hahn, H., 221
- Halmos, P., 25
- Hardy, G.H., 27
- harmonic function, 314 318
- Hausdorff, F., 153, 154, 211
- Hausdorff
  - dimension, 211
  - measure, 210
  - space, 135
- Heine, H.E., 72
- Heine-Borel theorem, 140, 144, 210
- Hermite, C., 27
- Herstein, I., 27
- Hölder, O., 252
- homeomorphic, 127
- homeomorphism, 127
- hypersurface, 265, 312
- image, 3, 178
- implicit function theorem, 194, 259
  - classical notation, 196
- incommensurability, 8
- indexed family, 4
- induced mapping of forms, 290-291
- inequality
  - arithmetic-geometric means, 234
  - Bernoulli, 30
  - Bessel, 168

- Bunyakovsky, *see* Schwarz
- Cauchy, 129
- Cauchy-Schwarz-
  - Bunyakovsky, 235
- Chebyshev, 233
- Hölder, 235
- Jensen, 233
- Liapounov, 234
- Minkowski, 235
- Schwarz, 108, 165
- infimum, 15
- injective, 3, 178
- inner product, 128, 178, 271, 282
- integers, 4
- integrable, 230
- integral
  - change of variables, 112
  - improper, 114
  - is a measure, 227, 228
  - linear change of variables, 241
  - linearity of, 227, 229
  - of differential form, 294
  - of form over manifold, 301
  - of nonnegative function, 228
  - of simple function, 227
  - Riemann, 98-114, 201, 231-233
    - existence, 101-105
  - smooth change of variables, 241-244
- integration
  - by parts, 112, 247
  - interchange of limit and, 113
  - linearity of, 230
  - of series, 113
- interior, 125
- interior point, 124
- intermediate value theorem, 59
- intersection, 1
- interval, 18-19
  - closed, 19
  - in  $\mathbf{R}^n$ , 204
  - open, 19
- intervals
  - nested sequences, 19
- inverse function, 60, 191
- inverse function theorem, 192, 257
- isolated point, 125
- isomorphism, 282
  - natural, 270
- Jacobi, C.G.J., 252
- Jacobian, 182, 294
  - matrix, 196
- Jensen, J.W.L., 252
- Jordan, C., 121
- kernel, 178
- Kolmogorov, A., 221
- Kronecker, L., 54
- Kronecker  $\delta$ , 269, 276
  - multi-index, 271
- Kronecker epsilon, 276, 311
- $L(\mu)$ , 230
- $L^p(\mu)$ , 235
  - completeness of, 237
- Lagrange, J.-L., 95, 96
- Lagrange multipliers, 266
- $\Lambda^r$  (alternating tensors of rank  $r$ ), 272
- $\lambda$ -system, 237
- Laplacian (Laplace operator), 314
- largest integer in, 15
- Lebesgue, H., 173, 221, 252
- Leibniz, G., 53, 82, 95, 121
- lemniscate, 258
- level surface, 194
- Levi, B., 228
- Levi-Civita, T., 284
- L'Hospital, G.-F.-A., 96
- L'Hospital's rule, 86
- $\liminf$ , 34
  - of sets, 219
- $\limsup$ , 34
  - of sets, 219

- limit point, 125
- Lindemann, F., 27
- line segment, 185
- linear algebra, 176–181
- linear functional, 269
- linear transformation, 176, 240
  - metrics on, 179
- Liouville, J., 20, 27, 219
- Liouville's theorem
  - on algebraic numbers, 20
  - on harmonic functions, 317
- Lipschitz condition, 138
- local coordinate, 255
- locally connected, 152
- lower semicontinuous, 70, 159
- lower sum, 100
- manifold, 253, 254
  - construction of, 258–259
  - examples of, 256–258
  - orientable, 263
  - orientation of, 263
  - oriented, 263
- map, 3
  - closed, 134
  - linear, 176
  - open, 134
- mapping. *see* map
- matrix, 176–178
- matrix product, 177
- maximal, 13
- maximal rank, 259
- maximum
  - existence of, 59, 142
  - local, 83, 184, 266
- maximum principle, 316
- Maxwell, J.C., 321
- mean value inequality, 186
- mean value theorem, 83, 185
  - for harmonic functions, 315
  - for integrals, 108
- measurability criteria, 224
- measurable
  - Lebesgue, 213, 215–218
- measurable function, 223
- measure, 206
  - Borel, 206
  - counting, 203
  - determined by  $\pi$ -system, 238, 239
  - Hausdorff, 215, 321
  - inner, 220, 222
  - Lebesgue, 213, 215
- Mertens, F., 53
- metric, 127
  - product, 132
  - standard on  $\mathbf{R}$ , 128
  - standard on  $\mathbf{R}^d$ , 129
  - uniform on  $C(X, Y)$ , 145
- metric space, 128–131, 214, 236
- minimal, 13
- minimum
  - existence of, 142
  - local, 83, 184, 266
- Minkowski, H., 252
- Möbius, A.F., 283
- Möbius strip, 258, 267, 319
- Monge, G., 267
- monotone
  - convergence theorem, 228, 248
  - function, 56
  - sequence, 33, 64
  - set function, 203, 209
- $\mu^*$ -measurable, 211
- multinomial coefficient, 190
- Müntz-Szasz theorem, 174
- nabla ( $\nabla$ ), 289
- neighborhood, 56, 124
- nested compact sets, 141
- net, 153
- Newton, I., 53, 95, 121
- nonsingular, 178
- norm, 129
- normal derivative, 314
- normal vector, 260
- normal vector field, 312
- null space, 178
- numbers

- algebraic, 20
  - complex, 25
  - extended real, 15, 203
    - operations, 223
  - natural, 2, 4
  - rational, 5
    - countability of, 7
  - real, 15
    - existence of, 21-23
    - uncountability of, 19
  - satanic, 221
  - transcendental, 20
- open cover, 140
- open set, 123, 215
- operator norm, 179
- orientable, 263
  - criterion for, 299
- orientation, 262-265
  - positive
    - of boundary, 264
  - standard, 265
  - usual, 263
- oriented
  - positively, 263
- orthogonal transformation, 240
- Osgood, W.F., 137, 153
- Ostrogradski, M.V., 321
- outer measure, 209
  - Hausdorff, 210, 221
  - Lebesgue, 210
  - Lebesgue-Stieltjes, 210, 221
  - metric, 214, 221
- $\pi$ -system, 237
- Parseval's relation, 169
- partial derivative
  - Leibniz notation, 197
  - multi-index notation, 190
- partial derivatives, 186-191
  - higher order, 189
  - mixed, 189
- partial order, 3
- partition, 2, 98
- partition of unity, 298, 299
- path, 159
  - length of, 159
- permutations, 273
- Perron Frobenius theorem, 319
- Picard's theorem, 138
- piecewise smooth, 168
- Plato, 26
- Poincaré lemma, 291
- point mass, 203
- power series, 90
  - of matrices, 181
- power set, 2
- precompact, 144
- primitive, 110
- projection map, 133
- pseudometric, 166, 220, 236
- radius of convergence, 90
- Radon, J., 221
- range, 3, 178
- refinement, 98
- relation, 2
  - equivalence, 2, 217
- Ricci, G., 284
- Riemann, B., 53, 121, 267
- Riemann integral, *see* integral
- Riemann sum, 99
- Riemann-Lebesgue lemma, 168
- Riemannian structure, 293
- Riesz, F., 252
- Riesz-Fischer theorem, 237
- Rogers, L.J., 252
- Rolle, M., 96
- root, existence of, 18
- row vector, 183, 269
- Russell, B., 27
- Schwarz, H.A., 109
- second axiom of countability, 131
- Seidel, P., 73
- selection, 98
- separable, 126, 220
- sequence, 4, 28
  - Cauchy, 36, 136
  - convergent, 28

- divergent, 28
- in topological space, 135
- monotone, 33
- of functions, 62
- of measures, 208
- uniformly Cauchy, 63
- series, 39
  - absolutely convergent, 40
  - alternating, 43
  - Cauchy product of, 49
  - comparison test, 40
  - conditionally convergent, 40
  - engineers test, 39
  - for  $\epsilon$ , 42
  - Fourier, 122, 164
  - geometric, 41
  - harmonic, 30, 41, 42
  - partial sum, 39
  - ratio test, 42
  - rearrangement of, 45
  - root test, 41
  - telescoping, 41
  - trigonometric, 174
  - unordered, 47
- set, 1
  - Borel, 205, 214, 215, 220, 224
  - closed, 124
  - convex, 96, 152
  - directed, 53
  - finite, 6
  - infinite, 6
  - Lebesgue measurable, 248
  - unmeasurable, 217
- Sierpiński, W., 154, 174
- $\sigma$ -algebra, 204, 238
  - generated by, 205
- $\sigma$ -field, 221
- sign of permutation, 273
- simple function, 225
  - canonical representation, 225
- smooth, 254, 255
  - forms, 286
- space
  - door, 150
  - Euclidean, 123
  - Hausdorff, 135
  - Hilbert, 130
  - metric, 128
  - topological, 123
- square root algorithm, 31
- Stokes, G., 73, 321
- Stokes' theorem, 303
  - classical case, 319
- Stone, M., 158, 174
- Stone's theorem, 157
- Stone-Weierstrass theorem, 158
- subgroup, 218
- subsequence, 34
- subspace, 131
- summable, 230
- summation by parts, 44
- summation convention, 281
- support line, 81
- supremum, 15
- surface, 253, 257
- surjective, 3, 178
- symmetric tensor, 282
- $T^r$  (tensors of rank  $r$ ), 270
- tangent bundle, 260
- tangent space, 253, 260
- tangent vector, 260
  - characterization of, 261, 262
- Taylor, B., 96
- Taylor polynomial, 89, 191
- Taylor's theorem, 89, 96, 120
  - in  $n$  variables, 190
- tensor
  - alternating, 272
  - covariant, 270
  - elementary, 283
  - rank of, 270
- tensor field, 285
  - of class  $C^k$ , 286
- tensor product, 271
- tensors of rank  $r$ 
  - basis for, 271
- Thomson, W. (Kelvin), 321

- Toeplitz, O., 26
- Tonelli, L., 252
- topological space, 220
- topologies
  - partial ordering of, 124
- topology, 123
  - base for, 131
  - discrete, 124
  - metric, 128
  - of uniform convergence, 232
  - product, 132, 134
  - relative, 131
  - stronger, finer, 124
  - trivial, 124
  - usual of  $\mathbf{R}$ , 124
  - weaker, coarser, 124
- torus, 257
  - dense line in, 265
  - dense line on, 266
- total order, 3
- totally bounded, 142
- trace norm, 179
- transition function, 256
- translation invariant, 217
- translation operator, 97
- transpose, 178
- transposition, 274
- trigonometric polynomials, 162
- triple scalar product, 279
- Tychonoff, A., 153
- uncountable, 6
- uniform continuity, 142
- uniform metric, 145
- uniformly continuous, 131
- union, 1
- unmeasurable set, 218
- unordered sums, 208
- upper halfspace, 254
- upper sum, 100
- Urysohn, P., 153
- Van Vleck, E.B., 222
- vector field, 260, 288
  - normal, 265
- Vitali, G., 222
- volume, 307-311
- wedge product, *see* exterior product
- Weierstrass, K., 27, 72, 73, 153, 173
- Weierstrass
  - $M$ -test, 64
  - approximation theorem, 155
- well separated, 214
- well-ordered, 4
- Weyl, H., 170

# Undergraduate Texts in Mathematics

---

(continued from page ii)

Second edition.

**Lang:** Linear Algebra. Third edition.

**Lang:** Undergraduate Algebra. Second edition.

**Lang:** Undergraduate Analysis.

**Lax/Burstein/Lax:** Calculus with Applications and Computing. Volume 1.

**LeCuyer:** College Mathematics with APL.

**Lidl/Pilz:** Applied Abstract Algebra.

**Macki-Strauss:** Introduction to Optimal Control Theory.

**Malitz:** Introduction to Mathematical Logic.

**Marsden/Weinstein:** Calculus I, II, III. Second edition.

**Martin:** The Foundations of Geometry and the Non-Euclidean Plane.

**Martin:** Transformation Geometry: An Introduction to Symmetry.

**Millman/Parker:** Geometry: A Metric Approach with Models. Second edition.

**Moschovakis:** Notes on Set Theory.

**Owen:** A First Course in the Mathematical Foundations of Thermodynamics.

**Palka:** An Introduction to Complex Function Theory.

**Pedrick:** A First Course in Analysis.

**Peressini/Sullivan/Uhl:** The Mathematics of Nonlinear Programming.

**Prenowitz/Jantosciak:** Join Geometries.

**Priestley:** Calculus: An Historical Approach.

**Protter/Morrey:** A First Course in

Real Analysis. Second edition.

**Protter/Morrey:** Intermediate Calculus. Second edition.

**Ross:** Elementary Analysis: The Theory of Calculus.

**Samuel:** Projective Geometry. *Readings in Mathematics.*

**Scharlau/Opolka:** From Fermat to Minkowski.

**Sigler:** Algebra.

**Silverman/Tate:** Rational Points on Elliptic Curves.

**Simmonds:** A Brief on Tensor Analysis. Second edition.

**Singer/Thorpe:** Lecture Notes on Elementary Topology and Geometry.

**Smith:** Linear Algebra. Second edition.

**Smith:** Primer of Modern Analysis. Second edition.

**Stanton/White:** Constructive Combinatorics.

**Stillwell:** Elements of Algebra: Geometry, Numbers, Equations.

**Stillwell:** Mathematics and Its History.

**Strayer:** Linear Programming and Its Applications.

**Thorpe:** Elementary Topics in Differential Geometry.

**Troutman:** Variational Calculus and Optimal Control. Second edition.

**Valenza:** Linear Algebra: An Introduction to Abstract Mathematics.

**Whyburn/Duda:** Dynamic Topology.

**Wilson:** Much Ado About Calculus.



*Mathematical Analysis: An Introduction* is a textbook containing more than enough material for a year-long course in analysis at the advanced undergraduate or beginning graduate level.

The book begins with a brief discussion of sets and mappings, describes the real number field, and proceeds to a treatment of real-valued functions of a real variable. Separate chapters are devoted to the ideas of convergent sequences and series, continuous functions, differentiation, and the Riemann integral. The middle chapters cover general topology and a miscellany of applications: the Weierstrass and Stone-Weierstrass approximation theorems, the existence of geodesics in compact metric spaces, elements of Fourier analysis, and the Weyl equidistribution theorem. Next comes a discussion of differentiation of vector-valued functions of several real variables, followed by a brief treatment of measure and integration (in a general setting, but with emphasis on Lebesgue theory in Euclidean space). The final part of the book deals with manifolds, differential forms, and Stokes' theorem, which is applied to prove Brouwer's fixed point theorem and to derive the basic properties of harmonic functions, such as the Dirichlet principle.

ISBN 0-387-94614-4



ISBN 0-387-94614-4