

Tom M. Apostol

---

# CALCULUS

VOLUME 1

One-Variable Calculus, with an  
Introduction to Linear Algebra

SECOND EDITION

John Wiley & Sons, Inc.  
New York • Santa Barbara • London • Sydney • Toronto

CONSULTING EDITOR

George Springer, *Indiana University*

XEROX ® is a trademark of Xerox Corporation.

Second Edition Copyright ©1967 by John Wiley & Sons, Inc.

First Edition copyright © 1961 by Xerox Corporation.

All rights reserved. Permission in writing must be obtained from the publisher before any part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system.

ISBN 0 471 00005 1

Library of Congress Catalog Card Number: 67-14605

Printed in the United States of America.

1 0 9 8 7 6 5 4 3 2

*To*  
*Jane and Stephen*



# PREFACE

---

## *Excerpts from the Preface to the First Edition*

There seems to be no general agreement as to what should constitute a first course in calculus and analytic geometry. Some people insist that the only way to really understand calculus is to start off with a thorough treatment of the real-number system and develop the subject step by step in a logical and rigorous fashion. Others argue that calculus is primarily a tool for engineers and physicists; they believe the course should stress applications of the calculus by appeal to intuition and by extensive drill on problems which develop manipulative skills. There is much that is sound in both these points of view. Calculus is a deductive science and a branch of pure mathematics. At the same time, it is very important to remember that calculus has strong roots in physical problems and that it derives much of its power and beauty from the variety of its applications. It is possible to combine a strong theoretical development with sound training in technique; this book represents an attempt to strike a sensible balance between the two. While treating the calculus as a deductive science, the book does not neglect applications to physical problems. Proofs of all the important theorems are presented as an essential part of the growth of mathematical ideas; the proofs are often preceded by a geometric or intuitive discussion to give the student some insight into why they take a particular form. Although these intuitive discussions will satisfy readers who are not interested in detailed proofs, the complete proofs are also included for those who prefer a more rigorous presentation.

The approach in this book has been suggested by the historical and philosophical development of calculus and analytic geometry. For example, integration is treated before differentiation. Although to some this may seem unusual, it is historically correct and pedagogically sound. Moreover, it is the best way to make meaningful the true connection between the integral and the derivative.

The concept of the integral is defined first for step functions. Since the integral of a step function is merely a finite sum, integration theory in this case is extremely simple. As the student learns the properties of the integral for step functions, he gains experience in the use of the summation notation and at the same time becomes familiar with the notation for integrals. This sets the stage so that the transition from step functions to more general functions seems easy and natural.

*Preface to the Second Edition*

The second **edition** differs from the first in **many** respects. Linear algebra has been incorporated, the mean-value theorems and routine applications of **calculus** are introduced at an earlier stage, and **many** new and easier **exercises** have been added. A glance at the table of contents reveals that the book has been divided into smaller **chapters**, **each** centering on an important concept. Several sections have been rewritten and reorganized to **provide** better motivation and to improve the flow of ideas.

As in the first **edition**, a historical introduction **precedes** **each** important new concept, tracing its development from an early intuitive physical notion to its **precise** mathematical formulation. The student is told something of the struggles of the past and of the triumphs of the men who contributed most to the subject. Thus the student becomes an active participant in the evolution of ideas rather than a passive observer of results.

The second **edition**, like the first, is divided into two volumes. The first two thirds of Volume I deals with the **calculus** of **functions** of **one** variable, including **infinite** series and an introduction to differential equations. The last third of Volume I introduces linear algebra with applications to geometry and analysis. **Much** of this material **leans** heavily on the **calculus** for examples that illustrate the general theory. It **provides** a natural blending of algebra and analysis and helps pave the way for the transition from **one**-variable **calculus** to multivariable calculus, discussed in Volume II. Further development of linear algebra **will occur** as needed in the second **edition** of Volume II.

Once **again** I acknowledge with pleasure my debt to Professors H. F. Bohnenblust, A. Erdélyi, F. B. Fuller, K. Hoffman, G. Springer, and H. S. Zuckerman. Their influence on the first **edition** **continued** into the second. In preparing the second **edition**, I received additional help from Professor Basil Gordon, who suggested **many** improvements. Thanks are also due George Springer and William P. Ziemer, who read the final draft. The staff of the Blaisdell Publishing Company has, as always, been helpful; I appreciate their **sympathetic** **consideration** of my wishes concerning format and typography.

Finally, it gives me **special** pleasure to express my gratitude to my wife for the **many** ways she has contributed **during** the preparation of both editions. In grateful acknowledgment I happily dedicate this book to her.

T. M. A.

**Pasadena, California**  
**September 16, 1966**

# CONTENTS

---

## 1. INTRODUCTION

### *Part 1. Historical Introduction*

11.1	The two <b>basic</b> concepts of <b>calculus</b>	1
1 1.2	Historical background	2
1 1.3	'The method of exhaustion for the <b>area</b> of a parabolic segment	3
*I 1.4	Exercises	8
1 1.5	A critical analysis of Archimedes' method	8
1 1.6	The approach to <b>calculus</b> to be used in this book	10

### *Part 2. Some Basic Concepts of the Theory of Sets*

12.1	Introduction to set theory	11
1 2.2	Notations for designating sets	12
12.3	Subsets	12
1 2.4	Unions, intersections, <b>complements</b>	13
1 2.5	Exercises	15

### *Part 3. A Set of Axioms for the Real-Number System*

13.1	Introduction	17
1 3.2	The <b>field</b> axioms	17
*I 3.3	Exercises	19
1 3.4	The <b>order</b> axioms	19
*I 3.5	Exercises	21
1 3.6	Integers and rational numbers	21
		ix

1 3.7	Geometric interpretation of real numbers as points on a line	22
1 3.8	Upper bound of a set, maximum element, least upper bound (supremum)	23
1 3.9	The least-Upper-bound axiom (completeness axiom)	25
1 3.10	The Archimedean property of the real-number system	25
1 3.11	Fundamental properties of the supremum and infimum	26
*1 3.12	Exercises	28
*1 3.13	Existence of square roots of nonnegative real numbers	29
*1 3.14	Roots of higher order. Rational powers	30
*1 3.15	Representation of real numbers by decimals	30

### *Part 4. Mathematical Induction, Summation Notation, and Related Topics*

14.1	An example of a <b>proof</b> by mathematical induction	32
1 4.2	The principle of mathematical induction	34
*1 4.3	The well-ordering principle	34
1 4.4	Exercises	35
*I 4.5	<b>Proof</b> of the well-ordering principle	37
1 4.6	The summation notation	37
1 4.7	Exercises	39
1 4.8	<b>Absolute</b> values and the triangle inequality	41
1 4.9	Exercises	43
*I 4.10	Miscellaneous <b>exercises</b> involving induction	44

## 1. THE CONCEPTS OF INTEGRAL CALCULUS

1.1	The <b>basic</b> ideas of Cartesian geometry	48
1.2	<b>Functions.</b> Informal description and examples	50
*1.3	<b>Functions.</b> Formal definition as a set of ordered pairs	53
1.4	More examples of real <b>functions</b>	54
1.5	Exercises	56
1.6	The concept of <b>area</b> as a set <b>function</b>	57
1.7	Exercises	60
1.8	Intervals and ordinate sets	60
1.9	Partitions and <b>step functions</b>	61
1.10	Sum and <b>product</b> of <b>step functions</b>	63
1.11	Exercises	63
1.12	The definition of the integral for <b>step functions</b>	64
1.13	Properties of the integral of a <b>step function</b>	66
1.14	Other notations for integrals	69



---

1.15 Exercises	70
1.16 The integral of more general functions	72
1.17 Upper and lower integrals	74
1.18 The <b>area</b> of an ordinate set expressed as an integral	75
1.19 Informal remarks on the theory and technique of integration	75
1.20 <b>Monotonic</b> and piecewise <b>monotonic</b> functions. Definitions and examples	76
1.21 Integrability of bounded <b>monotonic</b> functions	77
1.22 Calculation of the integral of a bounded <b>monotonic</b> function	79
1.23 Calculation of the integral $\int_0^b x^p dx$ when $p$ is a positive integer	79
1.24 The basic properties of the integral	80
1.25 Integration of polynomials	81
1.26 Exercises	83
1.27 Proofs of the basic properties of the integral	84

## 2. SOME APPLICATIONS OF INTEGRATION

2.1 Introduction	88
2.2 The <b>area</b> of a region between two graphs expressed as an integral	88
2.3 Worked examples	89
2.4 Exercises	94
2.5 The trigonometric functions	94
2.6 Integration formulas for the sine and cosine	97
2.7 A geometric description of the sine and cosine functions	102
2.8 Exercises	104
2.9 Polar coordinates	108
2.10 The integral for <b>area</b> in polar coordinates	109
2.11 Exercises	110
2.12 Application of integration to the calculation of volume	111
2.13 Exercises	114
2.14 Application of integration to the concept of work	115
2.15 Exercises	116
2.16 Average value of a function	117
2.17 Exercises	119
2.18 The integral as a function of the <b>upper</b> limit. Indefinite integrals	120
2.19 Exercises	124

## 3. CONTINUOUS FUNCTIONS

3.1 Informal description of <b>continuity</b>	126
3.2 The definition of the limit of a function	127

---

3.3	The definition of continuity of a function	130
3.4	The basic limit theorems. More examples of continuous functions	131
3.5	Proofs of the basic limit theorems	135
3.6	Exercises	138
3.7	Composite functions and continuity	140
3.8	Exercises	142
3.9	Bolzano's theorem for continuous functions	142
3.10	The intermediate-value theorem for continuous functions	144
3.11	Exercises	145
3.12	The process of inversion	146
3.13	Properties of functions preserved by inversion	147
3.14	Inverses of piecewise <b>monotonic</b> functions	148
3.15	Exercises	149
3.16	The extreme-value theorem for continuous functions	150
3.17	The small-span theorem for continuous functions (uniform continuity)	152
3.18	The integrability theorem for continuous functions	152
3.19	<b>Mean-value</b> theorems for integrals of continuous functions	154
3.20	Exercises	155

## 4. DIFFERENTIAL CALCULUS

4.1	Historical introduction	156
4.2	A problem involving velocity	157
4.3	The derivative of a function	159
4.4	Examples of derivatives	161
4.5	The algebra of derivatives	164
4.6	Exercises	167
4.7	Geometric interpretation of the derivative as a slope	169
4.8	Other notations for derivatives	171
4.9	Exercises	173
4.10	The <b>chain rule</b> for differentiating composite functions	174
4.11	Applications of the <b>chain rule</b> . Related rates and <b>implicit</b> differentiation	176
4.12	Exercises	179
4.13	Applications of differentiation to extreme values of functions	181
4.14	The <b>mean-value</b> theorem for derivatives	183
4.15	Exercises	186
4.16	Applications of the mean-value theorem to geometric properties of functions	187
4.17	Second-derivative test for extrema	188
4.18	Curve sketching	189
4.19	Exercises	191

4.20	Worked examples of extremum problems	191
4.21	Exercises	194
“4.22	Partial derivatives	196
“4.23	Exercises	201

## 5. THE RELATION BETWEEN INTEGRATION AND DIFFERENTIATION

5.1	The derivative of an indefinite integral. The first fundamental theorem of calculus	202
5.2	The zero-derivative theorem	204
5.3	Primitive functions and the second fundamental theorem of calculus	205
5.4	Properties of a function deduced from properties of its derivative	207
5.5	Exercises	208
5.6	The Leibniz notation for primitives	210
5.7	Integration by substitution	212
5.8	Exercises	216
5.9	Integration by parts	217
5.10	Exercises	220
*5.11	Miscellaneous review exercises	222

## 6. THE LOGARITHM, THE EXPONENTIAL, AND THE INVERSE TRIGONOMETRIC FUNCTIONS

6.1	Introduction	226
6.2	Motivation for the definition of the natural logarithm as an integral	227
6.3	The definition of the logarithm. Basic properties	229
6.4	The graph of the natural logarithm	230
6.5	Consequences of the functional equation $L(ab) = L(a) + L(b)$	230
6.6	Logarithms referred to any positive base $b \neq 1$	232
6.7	Differentiation and integration formulas involving logarithms	233
6.8	Logarithmic differentiation	235
6.9	Exercises	236
6.10	Polynomial approximations to the logarithm	238
6.11	Exercises	242
6.12	The exponential function	242
6.13	Exponentials expressed as powers of $e$	244
6.14	The definition of $e^x$ for arbitrary real $x$	244
6.15	The definition of $a^x$ for $a > 0$ and $x$ real	245

---

6.16	Differentiation and integration formulas involving exponentials	245
6.17	Exercises	248
6.18	The hyperbolic functions	251
6.19	Exercises	251
6.20	Derivatives of inverse functions	252
6.21	Inverses of the trigonometric functions	253
6.22	Exercises	256
6.23	Integration by partial fractions	258
6.24	Integrals which <b>can</b> be transformed into integrals of rational functions	264
6.25	Exercises	267
6.26	Miscellaneous review <b>exercises</b>	268

## 7. POLYNOMIAL APPROXIMATIONS TO FUNCTIONS

7.1	Introduction	272
7.2	The Taylor polynomials generated by a function	273
7.3	<b>Calculus</b> of Taylor polynomials	275
7.4	Exercises	278
7.5	<b>Taylor's</b> formula with remainder	278
7.6	Estimates for the error in Taylor's formula	280
*7.7	Other forms of the remainder in Taylor's formula	283
7.8	Exercises	284
7.9	Further remarks on the error in Taylor's formula. The o-notation	286
7.10	Applications to indeterminate forms	289
7.11	Exercises	290
7.12	<b>L'Hôpital's rule</b> for the indeterminate form $0/0$	292
7.13	Exercises	295
7.14	The symbols $+\infty$ and $-\infty$ , Extension of <b>L'Hôpital's rule</b>	296
7.15	<b>Infinite</b> limits	298
7.16	The behavior of $\log x$ and $e^x$ for large $x$	300
7.17	Exercises	303

## 8. INTRODUCTION TO DIFFERENTIAL EQUATIONS

8.1	Introduction	305
8.2	Terminology and notation	306
8.3	A first-order differential equation for the exponential function	307
8.4	First-order linear differential equations	308

8.5 Exercises	311
8.6 Some physical problems leading to first-order linear differential equations	313
8.7 Exercises	319
8.8 Linear equations of second order with constant coefficients	322
8.9 Existence of solutions of the equation $y'' + by = 0$	323
8.10 <b>Reduction</b> of the general equation to the <b>special</b> case $y'' + by = 0$	324
8.11 Uniqueness theorem for the equation $y'' + by = 0$	324
8.12 Complete solution of the equation $y'' + by = 0$	326
8.13 Complete solution of the equation $y'' + ay' + by = 0$	326
8.14 Exercises	328
8.15 Nonhomogeneous linear equations of second order with constant coefficients	329
8.16 <b>Special</b> methods for determining a particular solution of the nonhomogeneous equation $y'' + ay' + by = R$	332
8.17 Exercises	333
8.18 Examples of physical problems leading to linear second-order equations with constant coefficients	334
8.19 Exercises	339
8.20 Remarks concerning nonlinear differential equations	339
8.21 Integral <b>curves</b> and direction fields	341
8.22 Exercises	344
8.23 First-order separable equations	345
8.24 Exercises	347
8.25 Homogeneous first-order equations	347
8.26 Exercises	350
8.27 Some geometrical and physical problems leading to first-order equations	351
8.28 Miscellaneous review <b>exercises</b>	355

## 9. COMPLEX NUMBERS

9.1 Historical introduction	358
9.2 Definitions and field properties	358
9.3 The <b>complex</b> numbers as an extension of the real numbers	360
9.4 The imaginary unit $i$	361
9.5 Geometric interpretation. Modulus and argument	362
9.6 Exercises	365
9.7 Complex exponentials	366
9.8 Complex-valued <b>functions</b>	368
9.9 Examples of differentiation and integration formulas	369
9.10 Exercises	371

## 10. SEQUENCES, INFINITE SERIES, IMPROPER INTEGRALS

10.1 Zeno's paradox	374
10.2 Sequences	378
10.3 Monotonic sequences of real numbers	381
10.4 Exercises	382
10.5 Infinite series	383
10.6 The linearity property of convergent series	385
10.7 Telescoping series	386
10.8 The geometric series	388
10.9 Exercises	391
*10.10 Exercises on decimal expansions	393
10.11 Tests for convergence	394
10.12 Comparison tests for series of nonnegative terms	394
10.13 The integral test	397
10.14 Exercises	398
10.15 The root test and the ratio test for series of nonnegative terms	399
10.16 Exercises	402
10.17 Alternating series	403
10.18 Conditional and absolute convergence	406
10.19 The convergence tests of Dirichlet and Abel	407
10.20 Exercises	409
*10.21 Rearrangements of series	411
10.22 Miscellaneous review exercises	414
10.23 Improper integrals	416
10.24 Exercises	420

## 11. SEQUENCES AND SERIES OF FUNCTIONS

11.1 Pointwise convergence of sequences of functions	422
11.2 Uniform convergence of sequences of functions	423
11.3 Uniform convergence and continuity	424
11.4 Uniform convergence and integration	425
11.5 A sufficient condition for uniform convergence	427
11.6 Power series. Circle of convergence	428
11.7 Exercises	430
11.8 Properties of functions represented by real power series	431
11.9 The Taylor's series generated by a function	434
11.10 A sufficient condition for convergence of a Taylor's series	435

---

11.11	Power-series expansions for the exponential and trigonometric functions	435
*11.12	Bernstein's theorem	437
11.13	Exercises	438
11.14	Power series and differential equations	439
11.15	The binomial series	441
11.16	Exercises	443

## 12. VECTOR ALGEBRA

12.1	Historical introduction	445
12.2	The vector space of $n$ -tuples of real numbers.	446
12.3	Geometric interpretation for $n \leq 3$	448
12.4	Exercises	450
12.5	The dot product	451
12.6	Length or norm of a vector	453
12.7	Orthogonality of vectors	455
12.8	Exercises	456
12.9	Projections. Angle between vectors in $n$ -space	457
12.10	The unit coordinate vectors	458
12.11	Exercises	460
12.12	The linear span of a finite set of vectors	462
12.13	Linear independence	463
12.14	Bases	466
12.15	Exercises	467
12.16	The vector space $V_n(\mathbf{C})$ of $n$ -tuples of complex numbers	468
12.17	Exercises	470

## 13. APPLICATIONS OF VECTOR ALGEBRA TO ANALYTIC GEOMETRY

13.1	Introduction	471
13.2	Lines in $n$ -space	472
13.3	Some simple properties of straight lines	473
13.4	Lines and vector-valued functions	474
13.5	Exercises	477
13.6	Planes in Euclidean $n$ -space	478
13.7	Planes and vector-valued functions	481
13.8	Exercises	482
13.9	The cross product	483

---

13.10	The cross product expressed as a determinant	486
13.11	Exercises	487
13.12	The scalar triple product	488
13.13	Cramer's <b>rule</b> for solving a system of three linear equations	490
13.14	Exercises	491
13.15	Normal vectors to planes	493
13.16	Linear Cartesian equations for planes	494
13.17	Exercises	496
13.18	The conic sections	497
13.19	Eccentricity of conic sections	500
13.20	Polar equations for conic sections	501
13.21	Exercises	503
13.22	<b>Conic</b> sections symmetric <b>about</b> the origin	504
13.23	Cartesian equations for the conic sections	505
13.24	Exercises	508
13.25	Miscellaneous exercises on conic sections	509

## 14. CALCULUS OF VECTOR-VALUED FUNCTIONS

14.1	Vector-valued <b>functions</b> of a real variable	512
14.2	Algebraic operations. Components	512
14.3	Limits, derivatives, and integrals	513
14.4	Exercises	516
14.5	Applications to <b>curves. Tangency</b>	517
14.6	Applications to curvilinear motion. Velocity, speed, and acceleration	520
14.7	Exercises	524
14.8	The unit tangent, the principal normal, and the osculating plane of a curve	525
14.9	Exercises	528
14.10	The definition of arc length	529
14.11	Additivity of arc length	532
14.12	The arc-length <b>function</b>	533
14.13	Exercises	535
14.14	Curvature of a curve	536
14.15	Exercises	538
14.16	Velocity and acceleration in polar coordinates	540
14.17	Plane motion with radial acceleration	542
14.18	Cylindrical coordinates	543
14.19	Exercises	543
14.20	Applications to planetary motion	545
14.2	1 Miscellaneous review exercises	549



## 15. LINEAR SPACES

15.1	Introduction	551
15.2	The definition of a linear space	551
15.3	Examples of linear spaces	552
15.4	Elementary consequences of the axioms	554
15.5	Exercises	555
15.6	Subspaces of a linear space	556
15.7	Dependent and independent sets in a linear space	557
15.8	Bases and dimension	559
15.9	Exercises	560
15.10	Inner products, Euclidean spaces, norms	561
15.11	Orthogonality in a Euclidean space	564
15.12	Exercises	566
15.13	Construction of orthogonal sets. The Gram-Schmidt process	568
15.14	Orthogonal complements. Projections	572
15.15	Best approximation of elements in a Euclidean space by elements in a finite-dimensional subspace	574
15.16	Exercises	576

## 16. LINEAR TRANSFORMATIONS AND MATRICES

16.1	Linear transformations	578
16.2	Null space and range	579
16.3	Nullity and rank	581
16.4	Exercises	582
16.5	Algebraic operations on linear transformations	583
16.6	Inverses	585
16.7	One-to-one linear transformations	587
16.8	Exercises	589
16.9	Linear transformations with prescribed values	590
16.10	Matrix representations of linear transformations	591
16.11	Construction of a matrix representation in diagonal form	594
16.12	Exercises	596
16.13	Linear spaces of matrices	597
16.14	Isomorphism between linear transformations and matrices	599
16.15	Multiplication of matrices	600
16.16	Exercises	603
16.17	Systems of linear equations	605

16.18 Computation techniques	607
16.19 Inverses <del>sq</del> <del>far</del> matrices	611
16.20 <b>Exercises</b>	613
16.21 Miscellaneous exercises on matrices	614
Answers to exercises	617
Index	657

# *Calculus*



# INTRODUCTION

## *Part 1. Historical Introduction*

### 11.1 The two basic concepts of calculus

The remarkable progress that has been made in science and technology during the last Century is due in large part to the development of mathematics. That branch of mathematics known as integral and differential calculus serves as a natural and powerful tool for attacking a variety of problems that arise in physics, astronomy, engineering, chemistry, geology, biology, and other fields including, rather recently, some of the social sciences.

To give the reader an idea of the many different types of problems that can be treated by the methods of calculus, we list here a few sample questions selected from the exercises that occur in later chapters of this book.

With what speed should a rocket be fired upward so that it never returns to earth? What is the radius of the smallest circular disk that can cover every isosceles triangle of a given perimeter  $L$ ? What volume of material is removed from a solid sphere of radius  $2r$  if a hole of radius  $r$  is drilled through the center? If a strain of bacteria grows at a rate proportional to the amount present and if the population doubles in one hour, by how much will it increase at the end of two hours? If a ten-pound force stretches an elastic spring one inch, how much work is required to stretch the spring one foot?

These examples, chosen from various fields, illustrate some of the technical questions that can be answered by more or less routine applications of calculus.

Calculus is more than a technical tool—it is a collection of fascinating and exciting ideas that have interested thinking men for centuries. These ideas have to do with *speed*, *area*, *volume*, *rate of growth*, *continuity*, *tangent line*, and other concepts from a variety of fields. Calculus forces us to stop and think carefully about the meanings of these concepts. Another remarkable feature of the subject is its unifying power. Most of these ideas can be formulated so that they revolve around two rather specialized problems of a geometric nature. We turn now to a brief description of these problems.

Consider a curve  $C$  which lies above a horizontal base line such as that shown in Figure 1.1. We assume this curve has the property that every vertical line intersects it once at most.

The shaded portion of the figure consists of those points which lie below the curve  $C$ , above the horizontal base, and between two parallel vertical segments joining  $C$  to the base. The first fundamental problem of calculus is this : **To assign a number which measures the area of this shaded region.**

Consider next a line drawn tangent to the curve, as shown in Figure 1.1. The second fundamental problem may be stated as follows: **To assign a number which measures the steepness of this line.**

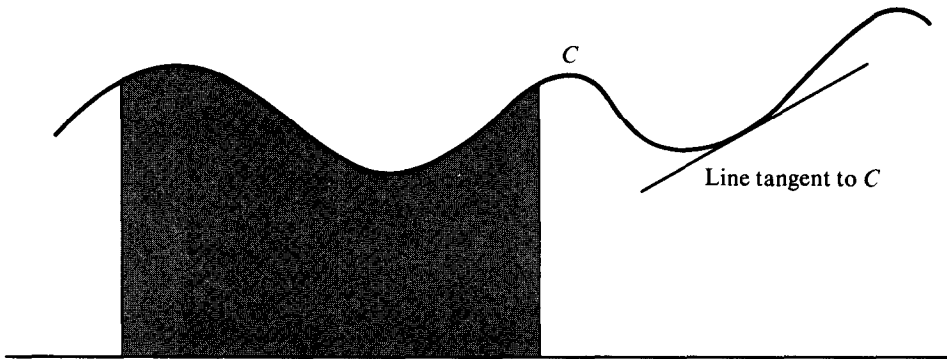


FIGURE 1.1

Basically, calculus has to do with the precise formulation and solution of these two special problems. It enables us to *define* the concepts of **area** and tangent line and *to calculate* the **area** of a given region or the steepness of a given tangent line. **Integral calculus** deals with the problem of **area** and will be discussed in Chapter 1. **Differential calculus** deals with the problem of tangents and will be introduced in Chapter 4.

The study of **calculus** requires a certain mathematical background. The present chapter deals with this background material and is divided into four parts : Part 1 provides historical perspective; Part 2 discusses some notation and terminology from the mathematics of sets; Part 3 deals with the real-number system; Part 4 treats mathematical induction and the summation notation. If the reader is acquainted with these topics, he **can** proceed directly to the development of integral calculus in Chapter 1. If not, he should become familiar with the material in the unstarred sections of this Introduction before proceeding to Chapter 1.

## II.2 Historical background

The birth of integral **calculus** occurred more than 2000 years ago when the Greeks attempted to determine **areas** by a process which they called the **method of exhaustion**. The essential ideas of this method are **very** simple and **can** be described briefly as follows: Given a region whose **area** is to be determined, we **inscribe** in it a polygonal region which **approximates** the given region and whose **area** we **can** easily compute. Then we **choose** another polygonal region which gives a better approximation, and we continue the **process**, taking **polygons** with more and more **sides** in an attempt to exhaust the given region. The method is illustrated for a semicircular region in Figure 1.2. It was used successfully by Archimedes (287-212 B.C.) to find exact formulas for the **area** of a **circle** and a few other special figures.

The development of the method of exhaustion beyond the point to which Archimedes carried it had to wait nearly eighteen centuries until the use of algebraic symbols and techniques became a standard part of mathematics. The elementary algebra that is familiar to most high-school students today was completely unknown in Archimedes' time, and it would have been next to impossible to extend his method to any general class of regions without some convenient way of expressing rather lengthy calculations in a compact and simplified form.

A slow but revolutionary change in the development of mathematical notations began in the 16th Century A.D. The cumbersome system of Roman numerals was gradually displaced by the Hindu-Arabic characters used today, the symbols  $+$  and  $-$  were introduced for the first time, and the advantages of the decimal notation began to be recognized. During this same period, the brilliant successes of the Italian mathematicians Tartaglia,

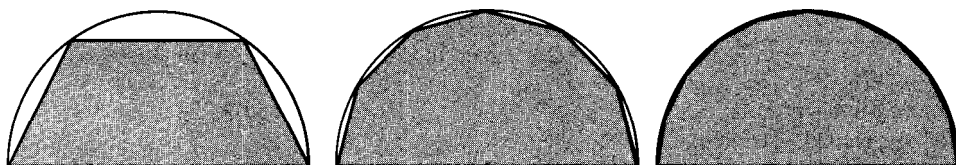


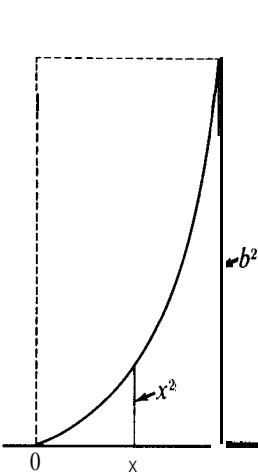
FIGURE 1.2 The method of exhaustion applied to a semicircular region.

Cardano, and Ferrari in finding algebraic solutions of cubic and quartic equations stimulated a great deal of activity in mathematics and encouraged the growth and acceptance of a new and superior algebraic language. With the widespread introduction of well-chosen algebraic symbols, interest was revived in the ancient method of exhaustion and a large number of fragmentary results were discovered in the 16th Century by such pioneers as Cavalieri, Toricelli, Roberval, Fermat, Pascal, and Wallis.

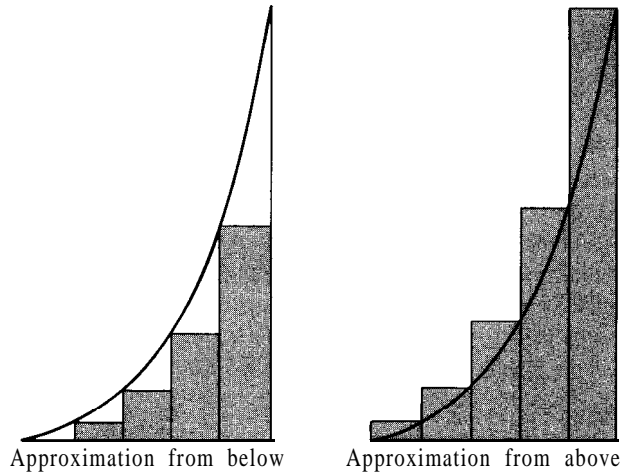
Gradually the method of exhaustion was transformed into the subject now called integral calculus, a new and powerful discipline with a large variety of applications, not only to geometrical problems concerned with areas and volumes but also to problems in other sciences. This branch of mathematics, which retained some of the original features of the method of exhaustion, received its biggest impetus in the 17th Century, largely due to the efforts of Isaac Newton (1642-1727) and Gottfried Leibniz (1646-1716), and its development continued well into the 19th Century before the subject was put on a firm mathematical basis by such men as Augustin-Louis Cauchy (1789-1857) and Bernhard Riemann (1826-1866). Further refinements and extensions of the theory are still being carried out in contemporary mathematics.

### 11.3 The method of exhaustion for the area of a parabolic segment

Before we proceed to a systematic treatment of integral calculus, it will be instructive to apply the method of exhaustion directly to one of the special figures treated by Archimedes himself. The region in question is shown in Figure 1.3 and can be described as follows: If we choose an arbitrary point on the base of this figure and denote its distance from 0 by  $x$ , then the vertical distance from this point to the curve is  $x^2$ . In particular, if the length of the base itself is  $b$ , the altitude of the figure is  $b^2$ . The vertical distance from  $x$  to the curve is called the "ordinate" at  $x$ . The curve itself is an example of what is known



**FIGURE 1.3** A parabolic segment.

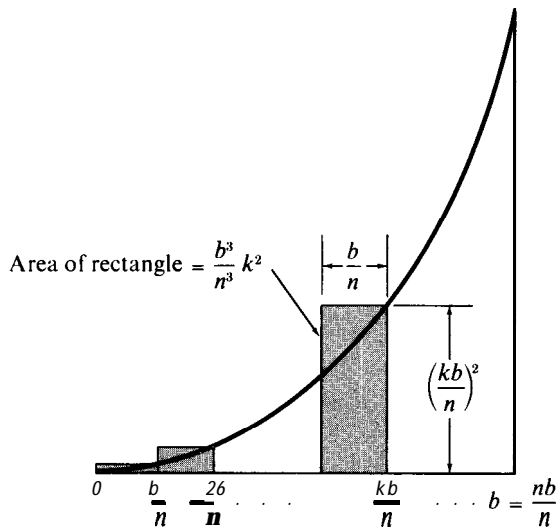


**FIGURE 1.4**

as a **parabola**. The region bounded by it and the two line segments is called a **parabolic segment**.

This figure may be enclosed in a rectangle of base  $b$  and altitude  $b^2$ , as shown in Figure 1.3. Examination of the figure suggests that the **area** of the parabolic segment is less than half the **area** of the rectangle. Archimedes made the surprising discovery that the **area** of the parabolic segment is exactly *one-third* that of the rectangle; that is to say,  $A = b^3/3$ , where  $A$  denotes the **area** of the parabolic segment. We shall show presently how to arrive at this result.

It should be pointed out that the parabolic segment in Figure 1.3 is not shown exactly as Archimedes drew it and the details that follow are not exactly the same as those used by him.



**FIGURE 1.5** Calculation of the **area** of a parabolic segment.



Nevertheless, the essential **ideas** are those of Archimedes; what is presented here is the method of exhaustion in modern notation.

The method is simply this: We **slice** the figure into a number of strips and obtain two approximations to the region, **one** from below and **one** from above, by using two sets of rectangles as illustrated in Figure 1.4. (We use rectangles rather than arbitrary polygons to simplify the computations.) The **area** of the parabolic segment is larger than the total **area** of the inner rectangles but smaller than that of the **outer** rectangles.

If **each** strip is further subdivided to obtain a new approximation with a larger number of strips, the total **area** of the inner rectangles **increases**, whereas the total **area** of the **outer** rectangles **decreases**. Archimedes realized that an approximation to the **area** within **any** desired degree of accuracy **could** be obtained by simply taking enough strips.

Let us carry out the actual computations that are required in this case. For the sake of **simplicity**, we subdivide the base into ***n*** equal parts, each of length  $b/n$  (see Figure 1.5). The points of subdivision correspond to the following values of  $x$ :

$$0, \frac{b}{n}, \frac{2b}{n}, \frac{3b}{n}, \dots, \frac{(n-1)b}{n}, \frac{nb}{n} = b$$

A typical point of subdivision corresponds to  $x = kb/n$ , where ***k*** takes the successive values ***k* = 0, 1, 2, 3, . . . , *n***. At each point  $kb/n$  we construct the outer rectangle of altitude  $(kb/n)^2$  as illustrated in Figure 1.5. The area of this rectangle is the product of its base and altitude and is equal to

$$\left(\frac{b}{n}\right)\left(\frac{kb}{n}\right)^2 = \frac{b^3}{n^3} k^2.$$

Let us denote by  $S_n$  the sum of the areas of all the outer rectangles. Then since the  $k$ th rectangle has area  $(b^3/n^3)k^2$ , we obtain the formula

$$(I.1) \quad S_n = \frac{b^3}{n^3} (1^2 + 2^2 + 3^2 + \dots + n^2).$$

In the same way we obtain a formula for the sum  $s_n$  of all the inner rectangles:

$$(I.2) \quad s_n = \frac{b^3}{n^3} [1^2 + 2^2 + 3^2 + \dots + (n-1)^2].$$

This brings us to a very important stage in the calculation. Notice that the factor multiplying  $b^3/n^3$  in Equation (I.1) is the sum of the squares of the first ***n*** integers:

$$1^2 + 2^2 + \dots + n^2.$$

[The corresponding factor in Equation (I.2) is similar except that the sum has only ***n* - 1** terms.] For a large value of ***n***, the computation of this sum by direct addition of its terms is tedious and inconvenient. Fortunately there is an interesting identity which makes it possible to evaluate this sum in a simpler way, namely,

$$(I.3) \quad 1^2 + 2^2 + \dots + n^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}.$$

This identity is valid for every integer  $n \geq 1$  and can be proved as follows: Start with the formula  $(k + 1)^3 = k^3 + 3k^2 + 3k + 1$  and rewrite it in the form

$$3k^2 + 3k + 1 = (k + 1)^3 - k^3.$$

Taking  $k = 1, 2, \dots, n - 1$ , we get  $n - 1$  formulas

$$3 \cdot 1^2 + 3 \cdot 1 + 1 = 2^3 - 1^3$$

$$3 \cdot 2^2 + 3 \cdot 2 + 1 = 3^3 - 2^3$$

$$3(n - 1)^2 + 3(n - 1) + 1 = n^3 - (n - 1)^3.$$

When we add these formulas, all the terms on the right cancel except two and we obtain

$$3[1^2 + 2^2 + \dots + (n - 1)^2] + 3[1 + 2 + \dots + (n - 1)] + (n - 1) = n^3 - 1^3.$$

The second sum on the left is the sum of terms in an arithmetic progression and it simplifies to  $\frac{1}{2}n(n - 1)$ . Therefore this last equation gives us

$$(I.4) \quad 1^2 + 2^2 + \dots + (n - 1)^2 = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6}.$$

Adding  $n^2$  to both members, we obtain (1.3).

For our purposes, we do not need the exact expressions given in the right-hand members of (1.3) and (1.4). All we need are the two *inequalities*

$$(I.5) \quad 1^2 + 2^2 + \dots + (n - 1)^2 < \frac{n^3}{3} < 1^2 + 2^2 + \dots + n^2$$

which are valid for every integer  $n \geq 1$ . These inequalities can be deduced easily as consequences of (1.3) and (I.4), or they can be proved directly by induction. (A proof by induction is given in Section 14.1.)

If we multiply both inequalities in (1.5) by  $b^3/n^3$  and make use of (1.1) and (I.2), we obtain

$$(I.6) \quad s_n < \frac{b^3}{3} < S_n$$

for every  $n$ . The inequalities in (1.6) tell us that  $b^3/3$  is a number which lies between  $s_n$  and  $S_n$  for every  $n$ . We will now prove that  $b^3/3$  is the *only* number which has this property. In other words, we assert that if  $A$  is any number which satisfies the inequalities

$$(I.7) \quad s_n < A < S_n$$

for every positive integer  $n$ , then  $A = b^3/3$ . It is because of this fact that Archimedes concluded that the area of the parabolic segment is  $b^3/3$ .

To prove that  $A = b^3/3$ , we use the inequalities in (1.5) once more. Adding  $n^2$  to both sides of the leftmost inequality in (1.5), we obtain

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

Multiplying this by  $b^3/n^3$  and using (I.1), we find

$$(I.8) \quad S_n < \frac{b^3}{3} + \frac{b^3}{n}.$$

Similarly, by subtracting  $n^2$  from both sides of the rightmost inequality in (1.5) and multiplying by  $b^3/n^3$ , we are led to the inequality

$$(I.9) \quad \frac{b^3}{3} - \frac{b^3}{n} < s_n.$$

Therefore, any number  $A$  satisfying (1.7) must also satisfy

$$(1.10) \quad \frac{b^3}{3} - \frac{b^3}{n} < A < \frac{b^3}{3} + \frac{b^3}{n}$$

for every integer  $n \geq 1$ . Now there are only three possibilities:

$$A > \frac{b^3}{3}, \quad A < \frac{b^3}{3}, \quad A = \frac{b^3}{3}.$$

If we show that each of the first two leads to a contradiction, then we must have  $A = b^3/3$ , since, in the manner of Sherlock Holmes, this exhausts all the possibilities.

Suppose the inequality  $A > b^3/3$  were true. From the second inequality in (1.10) we obtain

$$(1.11) \quad A - \frac{b^3}{3} < \frac{b^3}{n}$$

for every integer  $n \geq 1$ . Since  $A - b^3/3$  is positive, we may divide both sides of (1.11) by  $A - b^3/3$  and then multiply by  $n$  to obtain the equivalent statement

$$n < \frac{b^3}{A - b^3/3}$$

for every  $n$ . But this inequality is obviously false when  $n \geq b^3/(A - b^3/3)$ . Hence the inequality  $A > b^3/3$  leads to a contradiction. By a similar argument, we can show that the

inequality  $A < b^3/3$  also leads to a contradiction, and therefore we must have  $A = b^3/3$ , as asserted.

### \*I 1.4 Exercises

- (a) Modify the region in Figure 1.3 by assuming that the ordinate at each  $x$  is  $2x^2$  instead of  $x^2$ . Draw the new figure. Check through the principal steps in the foregoing section and find what effect this has on the calculation of the area. Do the same if the ordinate at each  $x$  is (b)  $3x^2$ , (c)  $\frac{1}{4}x^2$ , (d)  $2x^2 + 1$ , (e)  $ax^2 + c$ .
- Modify the region in Figure 1.3 by assuming that the ordinate at each  $x$  is  $x^3$  instead of  $x^2$ . Draw the new figure.
  - Use a construction similar to that illustrated in Figure 1.5 and show that the outer and inner sums  $S_n$  and  $s_n$  are given by

$$S_n = \frac{b^4}{n^4} (1^3 + 2^3 + \cdots + n^3), \quad s_n = \frac{b^4}{n^4} [1^3 + 2^3 + \cdots + (n-1)^3].$$

- Use the inequalities (which can be proved by mathematical induction; see Section 14.2)

$$(1.12) \quad 1^3 + 2^3 + \cdots + (n-1)^3 < \frac{n^4}{4} < 1^3 + 2^3 + \cdots + n^3$$

to show that  $s_n < b^4/4 < S_n$  for every  $n$ , and prove that  $b^4/4$  is the *only* number which lies between  $s_n$  and  $S_n$  for every  $n$ .

- What number takes the place of  $b^4/4$  if the ordinate at each  $x$  is  $ax^3 + c$ ?
- The inequalities (1.5) and (1.12) are special cases of the more general inequalities

$$(1.13) \quad 1'' + 2'' + \cdots + (n-1)'' < \frac{n^{k+1}}{k+1} < 1^k + 2'' + \cdots + n^k$$

that are valid for every integer  $n \geq 1$  and every integer  $k \geq 1$ . Assume the validity of (1.13) and generalize the results of Exercise 2.

### II.5 A critical analysis of Archimedes' method

From calculations similar to those in Section I 1.3, Archimedes concluded that the area of the parabolic segment in question is  $b^3/3$ . This fact was generally accepted as a mathematical theorem for nearly 2000 years before it was realized that one must re-examine the result from a more critical point of view. To understand why anyone would question the validity of Archimedes' conclusion, it is necessary to know something about the important changes that have taken place in the recent history of mathematics.

Every branch of knowledge is a collection of ideas described by means of words and symbols, and one cannot understand these ideas unless one knows the exact meanings of the words and symbols that are used. Certain branches of knowledge, known as **deductive systems**, are different from others in that a number of "undefined" concepts are chosen in advance and all other concepts in the system are defined in terms of these. Certain statements about these undefined concepts are taken as **axioms** or **postulates** and other

statements that **can** be deduced from the axioms are called *theorems*. The most familiar example of a deductive system is the Euclidean theory of elementary geometry that has been studied by well-educated men **since** the time of the **ancient** Greeks.

The spirit of early Greek mathematics, with its emphasis on the theoretical and **postu-**lational approach to geometry as presented in Euclid's *Elements*, dominated the thinking of mathematicians until the time of the Renaissance. A new and vigorous phase in the development of mathematics began with the **advent** of algebra in the 16th Century, and the next 300 years witnessed a flood of important discoveries. Conspicuously absent from this period was the logically **precise** reasoning of the deductive method with its use of axioms, definitions, and theorems. Instead, the pioneers in the **16th, 17th, and 18th cen-**turies resorted to a **curious** blend of deductive reasoning **combined** with intuition, pure guesswork, and **mysticism**, and it is not surprising to find that some of their work was **later** shown to be incorrect. However, a surprisingly large number of important discoveries emerged from this era, and a great deal of the work has survived the test of history—a **tribute** to the unusual skill and ingenuity of these pioneers.

As the flood of new discoveries began to **recede**, a new and more critical period emerged. Little by little, mathematicians felt **forced** to return to the classical ideals of the deductive method in an attempt to put the new mathematics on a firm foundation. This phase of the development, which began early in the 19th Century and has **continued** to the present day, has resulted in a degree of logical purity and abstraction that has surpassed **all** the traditions of Greek science. At the **same** time, it has brought **about** a clearer understanding of the foundations of not only **calculus** but of **all** of mathematics.

There are **many** ways to develop **calculus** as a deductive system. **One** possible approach is to take the real numbers as the undefined **objects**. Some of the **rules** governing the operations on real numbers **may** then be taken as axioms. **One** such set of axioms is listed in Part 3 of this Introduction. New concepts, such as *integral, limit, continuity, derivative*, must then be defined in terms of real numbers. Properties of these concepts are then deduced as theorems that follow from the axioms.

Looked at as part of the deductive system of calculus, Archimedes' result **about** the **area** of a parabolic segment **cannot** be **accepted** as a theorem until a satisfactory definition of **area** is given first. It is not clear whether Archimedes had ever formulated a **precise** definition of what he meant by **area**. He seems to have taken it for granted that every region has an **area** associated with it. On this assumption he then set **out** to calculate **areas** of particular regions. In his calculations he made use of certain **facts** about **area** that **cannot** be proved until we know what is *meant* by **area**. For instance, he assumed that if **one** region lies inside another, the **area** of the smaller region **cannot** exceed that of the **larger** region. Also, if a region is decomposed into two or more parts, the sum of the **areas** of the individual parts is equal to the **area** of the whole region. **All** these are properties we would like **area** to possess, and we shall insist that **any** definition of **area** should imply these properties. It is **quite** possible that Archimedes himself **may** have taken **area** to be an undefined concept and then used the properties we just mentioned as *axioms* **about** **area**.

Today we consider the work of Archimedes as being important not **so** **much** because it helps us to compute **areas** of particular figures, but rather because it suggests a reasonable way to *define* the concept of **area** for more or less *arbitrary* figures. As it turns **out**, the method of Archimedes suggests a way to **define** a **much** more general concept known as the *integral*. The integral, in turn, is used to compute not only **area** but also quantities **such** as arc length, volume, work and others.

If we look ahead and make use of the terminology of integral calculus, the result of the calculation carried out in Section 1.1.3 for the parabolic segment is often stated as follows :

“The integral of  $x^2$  from 0 to  $b$  is  $b^3/3$ .”

It is written symbolically as

$$\int_0^b x^2 dx = \frac{b^3}{3},$$

The symbol  $\int$  (an elongated  $S$ ) is called an *integral sign*, and it was introduced by Leibniz in 1675. The process which produces the number  $b^3/3$  is called *integration*. The numbers 0 and  $b$  which are **attached** to the integral sign are referred to as the *limits of integration*. The symbol  $\int_0^b x^2 dx$  must be regarded as a whole. Its definition will treat it as such, just as the dictionary describes the word “lapidate” without reference to “lap,” “id,” or “ate.”

Leibniz’ symbol for the integral was readily **accepted** by many early mathematicians because they liked to think of integration as a kind of “summation process” which enabled them to add together infinitely many “infinitesimally small quantities.” For example, the **area** of the parabolic segment was conceived of as a sum of infinitely many infinitesimally thin rectangles of height  $x^2$  and base  $dx$ . The integral sign represented the process of **adding** the **areas** of **all** these thin rectangles. This kind of thinking is suggestive and often **very** helpful, but it is not easy to assign a **precise** meaning to the idea of an “infinitesimally small quantity.” Today the integral is defined in terms of the notion of real number without using ideas like “infinitesimals.” This definition is given in Chapter 1.

## 11.6 The approach to calculus to be used in this book

A thorough and **complete** treatment of either integral or differential **calculus depends** ultimately on a careful study of the real number system. This study in itself, when carried out in full, is an interesting but somewhat lengthy program that requires a small volume for its **complete** exposition. The approach in this book is to begin with the real numbers as *undefined objects* and simply to list a number of fundamental properties of real numbers which we shall take as *axioms*. These axioms and some of the simplest theorems that **can** be deduced from them are discussed in Part 3 of this **chapter**.

Most of the properties of real numbers discussed here are probably familiar to the reader from his study of elementary algebra. However, there are a few properties of real numbers that do not ordinarily come into **consideration** in elementary algebra but which play an important role in the calculus. These properties stem from the so-called *least-upper-bound axiom* (also known as the *completeness* or *continuity axiom*) which is dealt with here in some detail. The reader **may** wish to study Part 3 before proceeding with the main body of the text, or he **may** postpone reading this material until **later** when he reaches those parts of the theory that make use of least-Upper-bound properties. Material in the text that **depends** on the least-Upper-bound axiom **will** be clearly indicated.

To develop **calculus** as a **complete**, formal mathematical theory, it would be necessary to state, in addition to the axioms for the real number system, a list of the various “methods of **proof**” which would be permitted for the **purpose** of deducing theorems from the axioms. Every statement in the theory would then have to be justified either as an “established law” (that is, an axiom, a definition, or a previously proved theorem) or as the result of applying

one of the acceptable methods of proof to an established law. A program of this sort would be extremely long and tedious and would add very little to a beginner's understanding of the subject. Fortunately, it is not necessary to proceed in this fashion in order to get a good understanding and a good working knowledge of calculus. In this book the subject is introduced in an informal way, and ample use is made of geometric intuition whenever it is convenient to do so. At the same time, the discussion proceeds in a manner that is consistent with modern standards of precision and clarity of thought. All the important theorems of the subject are explicitly stated and rigorously proved.

To avoid interrupting the principal flow of ideas, some of the proofs appear in separate starred sections. For the same reason, some of the chapters are accompanied by supplementary material in which certain important topics related to calculus are dealt with in detail. Some of these are also starred to indicate that they may be omitted or postponed without disrupting the continuity of the presentation. The extent to which the starred sections are taken up or not will depend partly on the reader's background and skill and partly on the depth of his interests. A person who is interested primarily in the basic techniques may skip the starred sections. Those who wish a more thorough course in calculus, including theory as well as technique, should read some of the starred sections.

## Part 2. Some Basic Concepts of the Theory of Sets

### 12.1 Introduction to set theory

In discussing any branch of mathematics, be it analysis, algebra, or geometry, it is helpful to use the notation and terminology of set theory. This subject, which was developed by Boole and Cantor<sup>†</sup> in the latter part of the 19th Century, has had a profound influence on the development of mathematics in the 20th Century. It has unified many seemingly disconnected ideas and has helped to reduce many mathematical concepts to their logical foundations in an elegant and systematic way. A thorough treatment of the theory of sets would require a lengthy discussion which we regard as outside the scope of this book. Fortunately, the basic notions are few in number, and it is possible to develop a working knowledge of the methods and ideas of set theory through an informal discussion. Actually, we shall discuss not so much a new theory as an agreement about the precise terminology that we wish to apply to more or less familiar ideas.

In mathematics, the word "set" is used to represent a collection of objects viewed as a single entity. The collections called to mind by such nouns as "flock," "tribe," "crowd," "team," and "electorate" are all examples of sets. The individual objects in the collection are called *elements* or *members* of the set, and they are said to *belong* to or to be *contained* in the set. The set, in turn, is said to *contain* or be *composed of* its elements.

---

<sup>†</sup> George Boole (1815-1864) was an English mathematician and logician. His book, *An Investigation of the Laws of Thought*, published in 1854, marked the creation of the first workable system of symbolic logic. Georg F. L. P. Cantor (1845-1918) and his school created the modern theory of sets during the period 1874-1895.

We shall be interested primarily in sets of mathematical objects: sets of numbers, sets of curves, sets of geometric figures, and so on. In many applications it is convenient to deal with sets in which nothing special is assumed about the nature of the individual objects in the collection. These are called abstract sets. Abstract set theory has been developed to deal with such collections of arbitrary objects, and from this generality the theory derives its power.

## 12.2 Notations for designating sets

Sets usually are denoted by capital letters:  $A, B, C, \dots, X, Y, Z$ ; elements are designated by lower-case letters:  $a, b, c, \dots, x, y, z$ . We use the special notation

$$x \in S$$

to mean that “ $x$  is an element of  $S$ ” or “ $x$  belongs to  $S$ .” If  $x$  does not belong to  $S$ , we write  $x \notin S$ . When convenient, we shall designate sets by displaying the elements in braces; for example, the set of positive even integers less than 10 is denoted by the symbol  $\{2, 4, 6, 8\}$  whereas the set of all positive even integers is displayed as  $\{2, 4, 6, \dots\}$ , the three dots taking the place of “and so on.” The dots are used only when the meaning of “and so on” is clear. The method of listing the members of a set within braces is sometimes referred to as the *roster notation*.

The first basic concept that relates one set to another is *equality* of sets:

**DEFINITION OF SET EQUALITY.** Two sets  $A$  and  $B$  are said to be *equal* (or *identical*) if they consist of exactly the same elements, in which case we write  $A = B$ . If one of the sets contains an element not in the other, we say the sets are *unequal* and we write  $A \neq B$ .

**EXAMPLE 1.** According to this definition, the two sets  $\{2, 4, 6, 8\}$  and  $\{2, 8, 6, 4\}$  are equal since they both consist of the four integers 2, 4, 6, and 8. Thus, when we use the roster notation to describe a set, the order in which the elements appear is irrelevant.

**EXAMPLE 2.** The sets  $\{2, 4, 6, 8\}$  and  $\{2, 2, 4, 4, 6, 8\}$  are equal even though, in the second set, each of the elements 2 and 4 is listed twice. Both sets contain the four elements 2, 4, 6, 8 and no others; therefore, the definition requires that we call these sets equal. This example shows that we do not insist that the objects listed in the roster notation be distinct. A similar example is the set of letters in the word *Mississippi*, which is equal to the set  $\{M, i, s, p\}$ , consisting of the four distinct letters  $M, i, s$ , and  $p$ .

## 12.3 Subsets

From a given set  $S$  we may form new sets, called *subsets* of  $S$ . For example, the set consisting of those positive integers less than 10 which are divisible by 4 (the set  $\{4, 8\}$ ) is a subset of the set of all even integers less than 10. In general, we have the following definition.

**DEFINITION OF A SUBSET.** A set  $A$  is said to be a subset of a set  $B$ , and we write

$$A \subseteq B,$$

whenever every element of  $A$  also belongs to  $B$ . We also say that  $A$  is *contained in*  $B$  or that  $B$  *contains*  $A$ . The relation  $\subseteq$  is referred to as *set inclusion*.



The statement  $A \subseteq B$  does not rule out the possibility that  $B \subseteq A$ . In fact, we may have both  $A \subseteq B$  and  $B \subseteq A$ , but this happens only if  $A$  and  $B$  have the same elements. In other words,

$$A = B \quad \text{if and only if} \quad A \subseteq B \text{ and } B \subseteq A.$$

This theorem is an immediate consequence of the foregoing definitions of equality and inclusion. If  $A \subseteq B$  but  $A \neq B$ , then we say that  $A$  is a *proper subset* of  $B$ ; we indicate this by writing  $A \subset B$ .

In all our applications of set theory, we have a fixed set  $S$  given in advance, and we are concerned only with subsets of this given set. The underlying set  $S$  may vary from one application to another; it will be referred to as the *universal set* of each particular discourse. The notation

$$\{x \mid x \in S \text{ and } x \text{ satisfies } P\}$$

will designate the set of all elements  $x$  in  $S$  which satisfy the property  $P$ . When the universal set to which we are referring is understood, we omit the reference to  $S$  and write simply  $\{x \mid x \text{ satisfies } P\}$ . This is read "the set of all  $x$  such that  $x$  satisfies  $P$ ." Sets designated in this way are said to be described by a defining property. For example, the set of all positive real numbers could be designated as  $\{x \mid x > 0\}$ ; the universal set  $S$  in this case is understood to be the set of all real numbers. Similarly, the set of all even positive integers  $\{2, 4, 6, \dots\}$  can be designated as  $\{x \mid x \text{ is a positive even integer}\}$ . Of course, the letter  $x$  is a dummy and may be replaced by any other convenient symbol. Thus, we may write

$$\{x \mid x > 0\} = \{y \mid y > 0\} = \{t \mid t > 0\}$$

and so on.

It is possible for a set to contain no elements whatever. This set is called the *empty set* or the *void set*, and will be denoted by the symbol  $\emptyset$ . We will consider  $\emptyset$  to be a subset of every set. Some people find it helpful to think of a set as analogous to a container (such as a bag or a box) containing certain objects, its elements. The empty set is then analogous to an empty container.

To avoid logical difficulties, we must distinguish between the element  $x$  and the set  $\{x\}$  whose only element is  $x$ . (A box with a hat in it is conceptually distinct from the hat itself.) In particular, the empty set  $\emptyset$  is not the same as the set  $\{\emptyset\}$ . In fact, the empty set  $\emptyset$  contains no elements, whereas the set  $\{\emptyset\}$  has one element,  $\emptyset$ . (A box which contains an empty box is not empty.) Sets consisting of exactly one element are sometimes called *one-element sets*.

Diagrams often help us visualize relations between sets. For example, we may think of a set  $S$  as a region in the plane and each of its elements as a point. Subsets of  $S$  may then be thought of as collections of points within  $S$ . For example, in Figure 1.6(b) the shaded portion is a subset of  $A$  and also a subset of  $B$ . Visual aids of this type, called *Venn diagrams*, are useful for testing the validity of theorems in set theory or for suggesting methods to prove them. Of course, the proofs themselves must rely only on the definitions of the concepts and not on the diagrams.

## 12.4 Unions, intersections, complements

From two given sets  $A$  and  $B$ , we can form a new set called the *union* of  $A$  and  $B$ . This new set is denoted by the symbol

$$A \cup B \quad (\text{read: "A union B"})$$

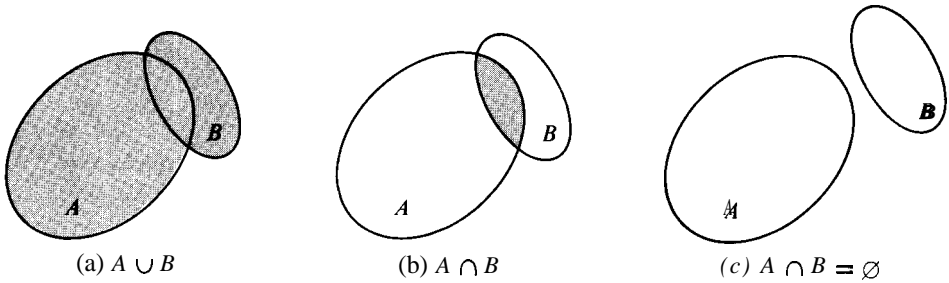


FIGURE 1.6 Unions and intersections.

and is defined as the set of those elements which are in  $A$ , in  $B$ , or in both. That is to say,  $A \cup B$  is the set of all elements which belong to at least one of the sets  $A$ ,  $B$ . An example is illustrated in Figure 1.6(a), where the shaded portion represents  $A \cup B$ .

Similarly, the *intersection* of  $A$  and  $B$ , denoted by

$$A \cap B \quad (\text{read: "A intersection B"}),$$

is defined as the set of those elements common to *both*  $A$  and  $B$ . This is illustrated by the shaded portion of Figure 1.6(b). In Figure 1.6(c), the two sets  $A$  and  $B$  have no elements in common; in this case, their intersection is the empty set  $\emptyset$ . Two sets  $A$  and  $B$  are said to be *disjoint* if  $A \cap B = \emptyset$ .

If  $A$  and  $B$  are sets, the *difference*  $A - B$  (also called the *complement of  $B$  relative to  $A$* ) is defined to be the set of all elements of  $A$  which are not in  $B$ . Thus, by definition,

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}.$$

In Figure 1.6(b) the unshaded portion of  $A$  represents  $A - B$ ; the unshaded portion of  $B$  represents  $B - A$ .

The operations of union and intersection have many formal similarities to (as well as differences from) ordinary addition and multiplication of real numbers. For example, since there is no question of order involved in the definitions of union and intersection, it follows that  $A \cup B = B \cup A$  and that  $A \cap B = B \cap A$ . That is to say, union and intersection are *commutative* operations. The definitions are also phrased in such a way that the operations are *associative*:

$$(A \cup B) \cup C = A \cup (B \cup C) \quad \text{and} \quad (A \cap B) \cap C = A \cap (B \cap C).$$

These and other theorems related to the "algebra of sets" are listed as Exercises in Section 1.2.5. One of the best ways for the reader to become familiar with the terminology and notations introduced above is to carry out the proofs of each of these laws. A sample of the type of argument that is needed appears immediately after the Exercises.

The operations of union and intersection can be extended to finite or infinite collections of sets as follows: Let  $\mathcal{F}$  be a nonempty class† of sets. The union of all the sets in  $\mathcal{F}$  is

---

† To help simplify the language, we call a collection of sets a *class*. Capital script letters  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$  are used to denote classes. The usual terminology and notation of set theory applies, of course, to classes. Thus, for example,  $A \in \mathcal{F}$  means that  $A$  is one of the sets in the class  $\mathcal{F}$ , and  $\mathcal{A} \subseteq \mathcal{B}$  means that every set in  $\mathcal{A}$  is also in  $\mathcal{B}$ , and so forth.

defined as the set of those elements which belong to at least **one** of the sets in  $\mathcal{F}$  and is denoted by the symbol

$$\bigcup_{A \in \mathcal{F}} A.$$

If  $\mathcal{F}$  is a finite collection of sets, say  $\mathcal{F} = \{A_1, A_2, \dots, A_n\}$ , we write

$$\bigcup_{A \in \mathcal{F}} A = \bigcup_{k=1}^n A_k = A_1 \cup A_2 \cup \dots \cup A_n.$$

Similarly, the intersection of all the sets in  $\mathcal{F}$  is defined to be the set of those elements which belong to every **one** of the sets in  $\mathcal{F}$ ; it is denoted by the symbol

$$\bigcap_{A \in \mathcal{F}} A.$$

For finite collections (as above), we write

$$\bigcap_{A \in \mathcal{F}} A = \bigcap_{k=1}^n A_k = A_1 \cap A_2 \cap \dots \cap A_n.$$

Unions and intersections have been defined in such a way that the associative laws for these operations are automatically satisfied. Hence, there is no ambiguity when we write  $A_1 \cup A_2 \cup \dots \cup A_n$ , or  $A_1 \cap A_2 \cap \dots \cap A_n$ .

## 12.5 Exercises

1. Use the roster notation to designate the following sets of real numbers.

$$A = \{x \mid x^2 - 1 = 0\}, \quad D = \{x \mid x^3 - 2x^2 + x = 2\}.$$

$$B = \{x \mid (x - 1)^2 = 0\}, \quad E = \{x \mid (x + 8)^2 = 9^2\}.$$

$$C = \{x \mid x + 8 = 9\}, \quad F = \{x \mid (x^2 + 16x)^2 = 17^2\}.$$

2. For the sets in Exercise 1, note that  $B \subseteq A$ . List all the inclusion relations  $\subseteq$  that hold among the sets  $A, B, C, D, E, F$ .
3. Let  $A = \{1\}$ ,  $B = \{1, 2\}$ . Discuss the validity of the following statements (prove the ones that are true and explain why the others are not true).
  - (a)  $A \subset B$ .
  - (b)  $A \subseteq B$ .
  - (c)  $A \in B$ .
  - (d)  $1 \in A$ .
  - (e)  $1 \subseteq A$ .
  - (f)  $1 \subset B$ .
4. Solve Exercise 3 if  $A = \{1\}$  and  $B = \{\{1\}, 1\}$ .
5. Given the set  $S = \{1, 2, 3, 4\}$ . Display all subsets of  $S$ . There are 16 altogether, counting  $\emptyset$  and  $S$ .
6. Given the following four sets

$$A = \{1, 2\}, \quad B = \{\{1\}, \{2\}\}, \quad C = \{\{1\}, \{1, 2\}\}, \quad D = \{\{1\}, \{2\}, \{1, 2\}\},$$

discuss the validity of the following statements (prove the **ones** that are true and explain why the others are not true).

- (a)  $A = B$ .      (d)  $A \in C$ .      (g)  $B \subset D$ .  
 (b)  $A \subseteq B$ .      (e)  $A \subset D$ .      (h)  $BE D$ .  
 (c)  $A \subset c$ .      (f)  $B \subset C$ .      (i)  $A \in D$ .

7. Prove the following properties of set equality.

- (a)  $\{a, a\} = \{a\}$ .  
 (b)  $\{a, b\} = \{b, a\}$ .  
 (c)  $\{a\} = \{b, c\}$  if and only if  $a = b = c$ .

Prove the set relations in Exercises 8 through 19. (Sample proofs are given at the end of this section).

8. *Commutative laws:*  $A \cup B = B \cup A$ ,  $A \cap B = B \cap A$ .

9. *Associative laws:*  $A \cup (B \cap C) = (A \cup B) \cap C$ ,  $A \cap (B \cup C) = (A \cap B) \cup C$ .

10. *Distributive laws:*  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ ,  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ .

11.  $A \cup A = A$ ,  $A \cap A = A$ ,

12.  $A \subseteq A \cup B$ ,  $A \cap B \subseteq A$ .

13.  $A \cup \emptyset = A$ ,  $A \cap \emptyset = \emptyset$ .

14.  $A \cup (A \cap B) = A$ ,  $A \cap (A \cup B) = A$ .

15. If  $A \subseteq C$  and  $B \subseteq C$ , then  $A \cup B \subseteq C$ .

16. If  $C \subseteq A$  and  $C \subseteq B$ , then  $C \subseteq A \cap B$ .

17. (a) If  $A \subset B$  and  $B \subset C$ , prove that  $A \subset C$ .

(b) If  $A \subseteq B$  and  $B \subseteq C$ , prove that  $A \subseteq C$ .

(c) What can you conclude if  $A \subset B$  and  $B \subseteq C$ ?

(d) If  $x \in A$  and  $A \subseteq B$ , is it necessarily true that  $x \in B$ ?

(e) If  $x \in A$  and  $A \in B$ , is it necessarily true that  $x \in B$ ?

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

19. Let  $\mathcal{F}$  be a class of sets. Then

$$B - \bigcup_{A \in \mathcal{F}} A = \bigcap_{A \in \mathcal{F}} (B - A) \quad \text{and} \quad B - \bigcap_{A \in \mathcal{F}} A = \bigcup_{A \in \mathcal{F}} (B - A).$$

20. (a) Prove that **one** of the following two formulas is always right and the other **one** is sometimes wrong :

$$(i) A - (B - C) = (A - B) \cup C,$$

$$(ii) A - (B \cup C) = (A - B) - C.$$

(b) State an additional necessary and sufficient condition for the formula which is sometimes incorrect to be always right.

*Proof of the commutative law  $A \cup B = B \cup A$ .* Let  $X = A \cup B$ ,  $Y = B \cup A$ . To prove that  $X = Y$  we prove that  $X \subseteq Y$  and  $Y \subseteq X$ . Suppose that  $x \in X$ . Then  $x$  is in at least **one** of  $A$  or  $B$ . Hence,  $x$  is in at least **one** of  $B$  or  $A$ ; so  $x \in Y$ . Thus, every element of  $X$  is also in  $Y$ , so  $X \subseteq Y$ . Similarly, we find that  $Y \subseteq X$ , so  $X = Y$ .

*Proof of  $A \cap B \subseteq A$ .* If  $x \in A \cap B$ , then  $x$  is in both  $A$  and  $B$ . In particular,  $x \in A$ . Thus, every element of  $A \cap B$  is also in  $A$ ; therefore,  $A \cap B \subseteq A$ .

## Part 3. A Set of Axioms for the Real-Number System

### 13.1 Introduction

There are many ways to introduce the real-number system. One popular method is to begin with the positive integers 1, 2, 3, . . . and use them as building blocks to construct a more comprehensive system having the properties desired. Briefly, the idea of this method is to take the positive integers as undefined concepts, state some axioms concerning them, and then use the positive integers to build a larger system consisting of the positive **rational** numbers (quotients of positive integers). The positive rational numbers, in turn, may then be used as a basis for constructing the positive **irrational** numbers (real numbers like  $\sqrt{2}$  and  $\pi$  that are not rational). The final step is the introduction of the negative real numbers and zero. The most difficult part of the whole process is the transition from the rational numbers to the irrational numbers.

Although the need for irrational numbers was apparent to the ancient Greeks from their study of geometry, satisfactory methods for constructing irrational numbers from rational numbers were not introduced until late in the 19th Century. At that time, three different theories were outlined by Karl Weierstrass (1815–1897), Georg Cantor (1845–1918), and Richard Dedekind (1831–1916). In 1889, the Italian mathematician Guiseppe Peano (1858–1932) listed five axioms for the positive integers that could be used as the starting point of the whole construction. A detailed account of this construction, beginning with the Peano postulates and using the method of Dedekind to introduce irrational numbers, may be found in a book by E. Landau, **Foundations of Analysis** (New York, Chelsea Publishing Co., 1951).

The point of view we shall adopt here is nonconstructive. We shall start rather far out in the process, taking the real numbers themselves as undefined objects satisfying a number of properties that we use as axioms. That is to say, we shall assume there exists a set **R** of objects, called real numbers, which satisfy the 10 axioms listed in the next few sections. All the properties of real numbers can be deduced from the axioms in the list. When the real numbers are defined by a constructive process, the properties we list as axioms must be proved as theorems.

In the axioms that appear below, lower-case letters **a, b, c, . . .**, x, y, z represent arbitrary real numbers unless something is said to the contrary. The axioms fall in a natural way into three groups which we refer to as the **field axioms**, the **order axioms**, and the **least-upper-bound axiom** (also called the **axiom of continuity** or the **completeness axiom**).

### 13.2 The field axioms

Along with the set **R** of real numbers we assume the existence of two operations called **addition** and **multiplication**, such that for every pair of real numbers x and y we can form the **sum** of x and y, which is another real number denoted by  $x + y$ , and the **product** of x and y, denoted by  $xy$  or by  $x \cdot y$ . It is assumed that the sum  $x + y$  and the product **xy** are **uniquely determined** by x and y. In other words, given x and y, there is exactly one real number  $x + y$  and exactly one real number  $xy$ . We attach no special meanings to the symbols + and  $\cdot$  other than those contained in the axioms.

AXIOM 1. COMMUTATIVE LAWS.  $x + y = y + x$ ,  $xy = yx$ .

AXIOM 2. ASSOCIATIVE LAWS.  $x + (y + z) = (x + y) + z$ ,  $x(yz) = (xy)z$ .

AXIOM 3. DISTRIBUTIVE LAW.  $x(y + z) = xy + xz$ .

AXIOM 4. EXISTENCE OF IDENTITY ELEMENTS. *There exist two distinct real numbers, which we denote by 0 and 1, such that for every real  $x$  we have  $x + 0 = x$  and  $1 \cdot x = x$ .*

AXIOM 5. EXISTENCE OF NEGATIVES. *For every real number  $x$  there is a real number  $y$  such that  $x + y = 0$ .*

AXIOM 6. EXISTENCE OF RECIPROCAL. *For every real number  $x \neq 0$  there is a real number  $y$  such that  $xy = 1$ .*

*Note:* The numbers 0 and 1 in Axioms 5 and 6 are those of Axiom 4.

From the above axioms we can deduce all the usual laws of elementary algebra. The most important of these laws are collected here as a list of theorems. In all these theorems the symbols **a**, **b**, **c**, **d** represent arbitrary real numbers.

THEOREM 1.1. CANCELLATION LAW FOR ADDITION. *If  $a + b = a + c$ , then  $b = c$ . (In particular, this shows that the number 0 of Axiom 4 is unique.)*

THEOREM 1.2. POSSIBILITY OF SUBTRACTION. *Given  $a$  and  $b$ , there is exactly one  $x$  such that  $a + x = b$ . This  $x$  is denoted by  $b - a$ . In particular,  $0 - a$  is written simply  $-a$  and is called the negative of  $a$ .*

THEOREM 1.3.  $b - a = b + (-a)$ .

THEOREM 1.4.  $-(-a) = a$ .

THEOREM 1.5.  $a(b - c) = ab - ac$ .

THEOREM 1.6.  $0 \cdot a = a \cdot 0 = 0$ .

THEOREM 1.7. CANCELLATION LAW FOR MULTIPLICATION. *If  $ab = ac$  and  $a \neq 0$ , then  $b = c$ . (In particular, this shows that the number 1 of Axiom 4 is unique.)*

THEOREM 1.8. POSSIBILITY OF DIVISION. *Given  $a$  and  $b$  with  $a \neq 0$ , there is exactly one  $x$  such that  $ax = b$ . This  $x$  is denoted by  $b/a$  or  $\frac{b}{a}$  and is called the quotient of  $b$  and  $a$ . In particular,  $1/a$  is also written  $a^{-1}$  and is called the reciprocal of  $a$ .*

THEOREM 1.9. *If  $a \neq 0$ , then  $b/a = b \cdot a^{-1}$ .*

THEOREM 1.10. *If  $a \neq 0$ , then  $(a^{-1})^{-1} = a$ .*

THEOREM 1.11. *If  $ab = 0$ , then  $a = 0$  or  $b = 0$ .*

THEOREM 1.12.  $(-a)b = -(ab)$  and  $(-a)(-b) = ab$ .

THEOREM 1.13.  $(a/b) + (c/d) = (ad + bc)/(bd)$  if  $b \neq 0$  and  $d \neq 0$ .

THEOREM 1.14.  $(a/b)(c/d) = (ac)/(bd)$  if  $b \neq 0$  and  $d \neq 0$ .

THEOREM 1.15.  $(a/b)/(c/d) = (ad)/(bc)$  if  $b \neq 0$ ,  $c \neq 0$ , and  $d \neq 0$ .

To illustrate how these statements may be obtained as consequences of the axioms, we shall present proofs of Theorems 1.1 through 1.4. Those readers who are interested may find it instructive to carry out proofs of the remaining theorems.

**Proof of 1.1.** Given  $a + b = a + c$ . By Axiom 5, there is a number  $y$  such that  $y + a = 0$ . Since sums are uniquely determined, we have  $y + (a + b) = y + (a + c)$ . Using the associative law, we obtain  $(y + a) + b = (y + a) + c$  or  $0 + b = 0 + c$ . But by Axiom 4 we have  $0 + b = b$  and  $0 + c = c$ , so that  $b = c$ . Notice that this theorem shows that there is only one real number having the property of 0 in Axiom 4. In fact, if 0 and 0' both have this property, then  $0 + 0' = 0$  and  $0 + 0 = 0$ . Hence  $0 + 0' = 0 + 0$  and, by the cancellation law,  $0 = 0'$ .

**Proof of 1.2.** Given  $a$  and  $b$ , choose  $y$  so that  $a + y = 0$  and let  $x = y + b$ . Then  $a + x = a + (y + b) = (a + y) + b = 0 + b = b$ . Therefore there is at least one  $x$  such that  $a + x = b$ . But by Theorem 1.1 there is at most one such  $x$ . Hence there is exactly one.

**Proof of 1.3.** Let  $x = b - a$  and let  $y = b + (-a)$ . We wish to prove that  $x = y$ . Now  $x + a = b$  (by the definition of  $b - a$ ) and

$$y + a = [b + (-a)] + a = b + [(-a) + a] = b + 0 = b.$$

Therefore  $x + a = y + a$  and hence, by Theorem 1.1,  $x = y$ .

**Proof of 1.4.** We have  $a + (-a) = 0$  by the definition of  $-a$ . But this equation tells us that  $a$  is the negative of  $-a$ . That is,  $a = -(-a)$ , as asserted.

### \*I 3.3 Exercises

1. Prove Theorems 1.5 through 1.15, using Axioms 1 through 6 and Theorems 1.1 through 1.4.

In Exercises 2 through 10, prove the given statements or establish the given equations. You may use Axioms 1 through 6 and Theorems 1.1 through 1.15.

2.  $-0 = 0$ .
3.  $1^{-1} = 1$ .
4. Zero has no reciprocal.
5.  $-(a + b) = -a - b$ .
6.  $-(a - b) = -a + b$ .
7.  $(a - b) + (b - c) = a - c$ .
8. If  $a \neq 0$  and  $b \neq 0$ , then  $(ab)^{-1} = a^{-1}b^{-1}$ .
9.  $-(a/b) = (-a)/b = a/(-b)$  if  $b \neq 0$ .
10.  $(a/b) - (c/d) = (ad - bc)/(bd)$  if  $b \neq 0$  and  $d \neq 0$ .

### 13.4 The order axioms

This group of axioms has to do with a concept which establishes an **ordering** among the real numbers. This ordering enables us to make statements about one real number being larger or smaller than another. We choose to introduce the order properties as a set of

axioms about a new undefined concept called **positiveness** and then to define terms like **less than** and **greater than** in terms of positiveness.

We shall assume that there exists a certain subset  $\mathbf{R}^+ \subset \mathbf{R}$ , called the set of **positive** numbers, which satisfies the following three order axioms :

AXIOM 7. **If  $x$  and  $y$  are in  $\mathbf{R}^+$ , so are  $x + y$  and  $xy$ .**

AXIOM 8. **For every real  $x \neq 0$ , either  $x \in \mathbf{R}^+$  or  $-x \in \mathbf{R}^+$ , but not both.**

AXIOM 9.  **$0 \notin \mathbf{R}^+$ .**

Now we can define the symbols  $<$ ,  $>$ ,  $\leq$ , and  $\geq$ , called, respectively, **less than**, **greater than**, **less than or equal to**, and **greater than or equal to**, as follows:

$x < y$  means that  $y - x$  is positive;

$y > x$  means that  $x < y$ ;

$x \leq y$  means that either  $x < y$  or  $x = y$ ;

$y \geq x$  means that  $x \leq y$ .

Thus, we have  $x > 0$  if and only if  $x$  is positive. If  $x < 0$ , we say that  $x$  is **negative**; if  $x \geq 0$ , we say that  $x$  is **nonnegative**. A pair of simultaneous inequalities such as  $x < y$ ,  $y < z$  is usually written more briefly as  $x < y < z$ ; similar interpretations are given to the compound inequalities  $x \leq y < z$ ,  $x < y \leq z$ , and  $x \leq y \leq z$ .

From the order axioms we can derive all the usual rules for calculating with inequalities. The most important of these are listed here as theorems.

THEOREM 1.16. TRICHOTOMY LAW. **For arbitrary real numbers  $a$  and  $b$ , exactly one of the three relations  $a < b$ ,  $b < a$ ,  $a = b$  holds.**

THEOREM 1.17. TRANSITIVE LAW. **If  $a < b$  and  $b < c$ , then  $a < c$ .**

THEOREM 1.18. **If  $a < b$ , then  $a + c < b + c$ .**

THEOREM 1.19. **If  $a < b$  and  $c > 0$ , then  $ac < bc$ .**

THEOREM 1.20. **If  $a \neq 0$ , then  $a^2 > 0$ .**

THEOREM 1.21.  **$1 > 0$ .**

THEOREM 1.22. **If  $a < b$  and  $c < 0$ , then  $ac > bc$ .**

THEOREM 1.23. **If  $a < b$ , then  $-a > -b$ . In particular, if  $a < 0$ , then  $-a > 0$ .**

THEOREM 1.24. **If  $ab > 0$ , then both  $a$  and  $b$  are positive or both are negative.**

THEOREM 1.25. **If  $a < c$  and  $b < d$ , then  $a + b < c + d$ .**

Again, we shall prove only a few of these theorems as samples to indicate how the proofs may be carried out. Proofs of the others are left as exercises.



**Proof of 1.16.** Let  $x = b - a$ . If  $x = 0$ , then  $b - a = a - b = 0$ , and hence, by Axiom 9, we cannot have  $a > b$  or  $b > a$ . If  $x \neq 0$ , Axiom 8 tells us that either  $x > 0$  or  $x < 0$ , but not both; that is, either  $a < b$  or  $b < a$ , but not both. Therefore, exactly one of the three relations,  $a = b$ ,  $a < b$ ,  $b < a$ , holds.

**Proof of 1.17.** If  $a < b$  and  $b < c$ , then  $b - a > 0$  and  $c - b > 0$ . By Axiom 7 we may add to obtain  $(b - a) + (c - b) > 0$ . That is,  $c - a > 0$ , and hence  $a < c$ .

**Proof of 1.18.** Let  $x = a + c$ ,  $y = b + c$ . Then  $y - x = b - a$ . But  $b - a > 0$  since  $a < b$ . Hence  $y - x > 0$ , and this means that  $x < y$ .

**Proof of 1.19.** If  $a < b$ , then  $b - a > 0$ . If  $c > 0$ , then by Axiom 7 we may multiply  $c$  by  $(b - a)$  to obtain  $(b - a)c > 0$ . But  $(b - a)c = bc - ac$ . Hence  $bc - ac > 0$ , and this means that  $ac < bc$ , as asserted.

**Proof of 1.20.** If  $a > 0$ , then  $a \cdot a > 0$  by Axiom 7. If  $a < 0$ , then  $-a > 0$ , and hence  $(-a) \cdot (-a) > 0$  by Axiom 7. In either case we have  $a^2 > 0$ .

**Proof of 1.21.** Apply Theorem 1.20 with  $a = 1$ .

### \*I 3.5 Exercises

1. Prove Theorems 1.22 through 1.25, using the earlier theorems and Axioms 1 through 9.

In Exercises 2 through 10, prove the given statements or establish the given inequalities. You may use Axioms 1 through 9 and Theorems 1.1 through 1.25.

2. There is no real number  $x$  such that  $x^2 + 1 = 0$ .
3. The sum of two negative numbers is negative.
4. If  $a > 0$ , then  $1/a > 0$ ; if  $a < 0$ , then  $1/a < 0$ .
5. If  $0 < a < b$ , then  $0 < b^{-1} < a^{-1}$ .
6. If  $a \leq b$  and  $b \leq c$ , then  $a \leq c$ .
7. If  $a \leq b$  and  $b \leq c$ , and  $a = c$ , then  $b = c$ .
8. For all real  $a$  and  $b$  we have  $a^2 + b^2 \geq 0$ . If  $a$  and  $b$  are not both 0, then  $a^2 + b^2 > 0$ .
9. There is no real number  $a$  such that  $x \leq a$  for all real  $x$ .
10. If  $x$  has the property that  $0 \leq x < h$  for every positive real number  $h$ , then  $x = 0$ .

### 13.6 Integers and rational numbers

There exist certain subsets of  $\mathbf{R}$  which are distinguished because they have special properties not shared by all real numbers. In this section we shall discuss two such subsets, the **integers** and the **rational numbers**.

To introduce the positive integers we begin with the number 1, whose existence is guaranteed by Axiom 4. The number  $1 + 1$  is denoted by 2, the number  $2 + 1$  by 3, and so on. The numbers 1, 2, 3, . . . , obtained in this way by repeated addition of 1 are all positive, and they are called the **positive integers**. Strictly speaking, this description of the positive integers is not entirely complete because we have not explained in detail what we mean by the expressions "and so on," or "repeated addition of 1." Although the intuitive meaning

of these expressions may seem clear, in a careful treatment of the real-number system it is necessary to give a more precise definition of the positive integers. There are many ways to do this. One convenient method is to introduce first the notion of an *inductive set*.

DEFINITION OF AN INDUCTIVE SET. A set of real numbers is called an inductive set if it has the following two properties:

- (a) The number 1 is in the set.
- (b) For every  $x$  in the set, the number  $x + 1$  is also in the set.

For example,  $\mathbf{R}$  is an inductive set. So is the set  $\mathbf{R}^+$ . Now we shall define the positive integers to be those real numbers which belong to every inductive set.

DEFINITION OF POSITIVE INTEGERS. A real number is called a positive integer if it belongs to every inductive set.

Let  $\mathbf{P}$  denote the set of all positive integers. Then  $\mathbf{P}$  is itself an inductive set because (a) it contains 1, and (b) it contains  $x + 1$  whenever it contains  $x$ . Since the members of  $\mathbf{P}$  belong to every inductive set, we refer to  $\mathbf{P}$  as the *smallest* inductive set. This property of the set  $\mathbf{P}$  forms the logical basis for a type of reasoning that mathematicians call *proof by induction*, a detailed discussion of which is given in Part 4 of this Introduction.

The negatives of the positive integers are called the *negative integers*. The positive integers, together with the negative integers and 0 (zero), form a set  $\mathbf{Z}$  which we call simply the *set of integers*.

In a thorough treatment of the real-number system, it would be necessary at this stage to prove certain theorems about integers. For example, the sum, difference, or product of two integers is an integer, but the quotient of two integers need not be an integer. However, we shall not enter into the details of such proofs.

Quotients of integers  $a/b$  (where  $b \neq 0$ ) are called *rational numbers*. The set of rational numbers, denoted by  $\mathbf{Q}$ , contains  $\mathbf{Z}$  as a subset. The reader should realize that all the field axioms and the order axioms are satisfied by  $\mathbf{Q}$ . For this reason, we say that the set of rational numbers is an *ordered field*. Real numbers that are not in  $\mathbf{Q}$  are called *irrational*.

### 13.7 Geometric interpretation of real numbers as points on a line

The reader is undoubtedly familiar with the geometric representation of real numbers by means of points on a straight line. A point is selected to represent 0 and another, to the right of 0, to represent 1, as illustrated in Figure 1.7. This choice determines the *scale*. If one adopts an appropriate set of axioms for Euclidean geometry, then each real number corresponds to exactly one point on this line and, conversely, each point on the line corresponds to one and only one real number. For this reason the line is often called the *real line* or the *real axis*, and it is customary to use the words *real number* and *point interchangeably*. Thus we often speak of the *point*  $x$  rather than the point corresponding to the real number  $x$ .

The ordering relation among the real numbers has a simple geometric interpretation. If  $x < y$ , the point  $x$  lies to the left of the point  $y$ , as shown in Figure 1.7. Positive numbers

lie to the right of 0 and negative numbers to the left of 0. If  $a < b$ , a point  $x$  satisfies the inequalities  $a < x < b$  if and only if  $x$  is **between**  $a$  and  $b$ .

This device for representing real numbers geometrically is a very worthwhile aid that helps us to discover and understand better certain properties of real numbers. However, the reader should realize that all properties of real numbers that are to be accepted as theorems must be deducible from the axioms without any reference to geometry. This does not mean that one should not make use of geometry in studying properties of real numbers. On the contrary, the geometry often suggests the method of proof of a particular theorem, and sometimes a geometric argument is more illuminating than a purely **analytic** proof (one depending entirely on the axioms for the real numbers). In this book, geometric

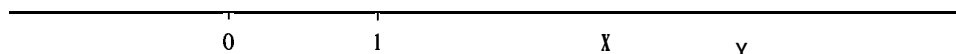


FIGURE 1.7 Real numbers represented geometrically on a line.

arguments are used to a large extent to help motivate or clarify a particular discussion. Nevertheless, the proofs of all the important theorems are presented in analytic form.

### 13.8 Upper bound of a set, maximum element, least upper bound (supremum)

The nine axioms listed above contain all the properties of real numbers usually discussed in elementary algebra. There is another axiom of fundamental importance in calculus that is ordinarily not discussed in elementary algebra courses. This axiom (or some property equivalent to it) is used to establish the existence of irrational numbers.

Irrational numbers arise in elementary algebra when we try to solve certain quadratic equations. For example, it is desirable to have a real number  $x$  such that  $x^2 = 2$ . From the nine axioms above, we cannot prove that such an  $x$  exists in  $\mathbf{R}$ , because these nine axioms are also satisfied by  $\mathbf{Q}$ , and there is no rational number  $x$  whose square is 2. (A proof of this statement is outlined in Exercise 11 of Section 1 3.12.) Axiom 10 allows us to introduce irrational numbers in the real-number system, and it gives the real-number system a property of continuity that is a keystone in the logical structure of calculus.

Before we describe Axiom 10, it is convenient to introduce some more terminology and notation. Suppose  $S$  is a nonempty set of real numbers and suppose there is a number  $B$  such that

$$x \leq B$$

for every  $x$  in  $S$ . Then  $S$  is said to be **bounded above** by  $B$ . The number  $B$  is called an **upper bound** for  $S$ . We say **an** upper bound because every number greater than  $B$  will also be an upper bound. If an upper bound  $B$  is also a member of  $S$ , then  $B$  is called the **largest member** or the **maximum element** of  $S$ . There can be at most one such  $B$ . If it exists, we write

$$B = \max x \in S.$$

Thus,  $B = \max S$  if  $B \in S$  and  $x \leq B$  for all  $x$  in  $S$ . A set with no upper bound is said to be **unbounded above**.

The following examples serve to illustrate the meaning of these terms.

**EXAMPLE 1.** Let  $S$  be the set of all positive real numbers. This set is unbounded above. It has no upper bounds and it has no maximum element.

**EXAMPLE 2.** Let  $S$  be the set of all real  $x$  satisfying  $0 \leq x \leq 1$ . This set is bounded above by 1. In fact, 1 is its maximum element.

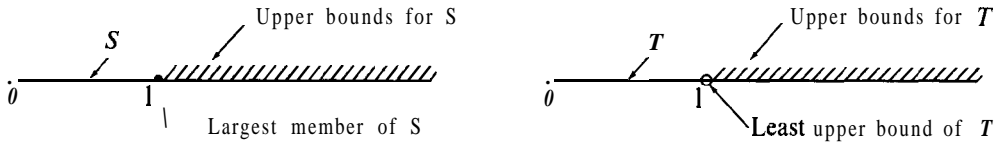
**EXAMPLE 3.** Let  $T$  be the set of all real  $x$  satisfying  $0 \leq x < 1$ . This is like the set in Example 2 except that the point 1 is not included. This set is bounded above by 1 but it has no maximum element.

Some sets, like the one in Example 3, are bounded above but have no maximum element. For these sets there is a concept which takes the place of the maximum element. This is called the **least upper bound** of the set and it is defined as follows:

**DEFINITION OF LEAST UPPER BOUND.** A number  $B$  is called a **least upper bound** of a nonempty set  $S$  if  $B$  has the following two properties:

- (a)  $B$  is an upper bound for  $S$ .
- (b) No number less than  $B$  is an upper bound for  $S$ .

If  $S$  has a maximum element, this maximum is also a least upper bound for  $S$ . But if  $S$  does not have a maximum element, it may still have a least upper bound. In Example 3 above, the number 1 is a least upper bound for  $T$  although  $T$  has no maximum element. (See Figure 1.8.)



(a)  $S$  has a largest member:  
 $\max S = 1$

(b)  $T$  has no largest member, but it has  
a least upper bound:  $\sup T = 1$

**FIGURE 1.8** Upper bounds, maximum element, supremum.

**THEOREM 1.26.** Two different numbers cannot be least upper bounds for the same set.

*Proof.* Suppose that  $B$  and  $C$  are two least upper bounds for a set  $S$ . Property (b) implies that  $C \geq B$  since  $B$  is a least upper bound; similarly,  $B \geq C$  since  $C$  is a least upper bound. Hence, we have  $B = C$ .

This theorem tells us that if there is a least upper bound for a set  $S$ , there is only one and we may speak of **the** least upper bound.

It is common practice to refer to the least upper bound of a set by the more concise term **supremum**, abbreviated **sup**. We shall adopt this convention and write

$$B = \sup S$$

to express the fact that  $B$  is the least upper bound, or supremum, of  $S$ .

### 13.9 The least-Upper-bound axiom (completeness axiom)

Now we are ready to state the least-Upper-bound axiom for the real-number system.

**AXIOM 10.** Every nonempty set  $S$  of real numbers which is bounded above has a supremum; that is, there is a real number  $B$  such that  $B = \sup S$ .

We emphasize once more that the supremum of  $S$  need not be a member of  $S$ . In fact,  $\sup S$  belongs to  $S$  if and only if  $S$  has a maximum element, in which case  $\max S = \sup S$ .

Definitions of the terms *lower bound*, *bounded below*, *smallest member* (or *minimum element*) may be similarly formulated. The reader should formulate these for himself. If  $S$  has a minimum element, we denote it by  $\min S$ .

A number  $L$  is called a *greatest lower bound* (or *infimum*) of  $S$  if (a)  $L$  is a lower bound for  $S$ , and (b) no number greater than  $L$  is a lower bound for  $S$ . The infimum of  $S$ , when it exists, is uniquely determined and we denote it by  $\inf S$ . If  $S$  has a minimum element, then  $\min S = \inf S$ .

Using Axiom 10, we can prove the following.

**THEOREM 1.27.** Every nonempty set  $S$  that is bounded below has a greatest lower bound; that is, there is a real number  $L$  such that  $L = \inf S$ .

*Proof.* Let  $-S$  denote the set of negatives of numbers in  $S$ . Then  $-S$  is nonempty and bounded above. Axiom 10 tells us that there is a number  $B$  which is a supremum for  $-S$ . It is easy to verify that  $-B = \inf S$ .

Let us refer once more to the examples in the foregoing section. In Example 1, the set of all positive real numbers, the number 0 is the infimum of  $S$ . This set has no minimum element. In Examples 2 and 3, the number 0 is the minimum element.

In all these examples it was easy to decide whether or not the set  $S$  was bounded above or below, and it was also easy to determine the numbers  $\sup S$  and  $\inf S$ . The next example shows that it may be difficult to determine whether upper or lower bounds exist.

**EXAMPLE 4.** Let  $S$  be the set of all numbers of the form  $(1 + 1/n)^n$ , where  $n = 1, 2, 3, \dots$ . For example, taking  $n = 1, 2$ , and  $3$ , we find that the numbers  $2$ ,  $\frac{9}{4}$ , and  $\frac{64}{27}$  are in  $S$ . Every number in the set is greater than 1, so the set is bounded below and hence has an infimum. With a little effort we can show that 2 is the smallest element of  $S$  so  $\inf S = \min S = 2$ . The set  $S$  is also bounded above, although this fact is not as easy to prove. (Try it!) Once we know that  $S$  is bounded above, Axiom 10 tells us that there is a number which is the supremum of  $S$ . In this case it is not easy to determine the value of  $\sup S$  from the description of  $S$ . In a later chapter we will learn that  $\sup S$  is an irrational number approximately equal to 2.718. It is an important number in calculus called the Euler number  $e$ .

### 13.10 The Archimedean property of the real-number system

This section contains a number of important properties of the real-number system which are consequences of the least-Upper-bound axiom.

**THEOREM 1.28.** *The set  $\mathbf{P}$  of positive integers  $1, 2, 3, \dots$  is unbounded above.*

*Proof.* Assume  $\mathbf{P}$  is bounded above. We shall show that this leads to a contradiction. Since  $\mathbf{P}$  is nonempty, Axiom 10 tells us that  $\mathbf{P}$  has a least upper bound, say  $b$ . The number  $b - 1$ , being less than  $b$ , cannot be an upper bound for  $\mathbf{P}$ . Hence, there is at least one positive integer  $n$  such that  $n > b - 1$ . For this  $n$  we have  $n + 1 > b$ . Since  $n + 1$  is in  $\mathbf{P}$ , this contradicts the fact that  $b$  is an upper bound for  $\mathbf{P}$ .

As corollaries of Theorem 1.28, we immediately obtain the following consequences:

**THEOREM 1.29.** *For every real  $x$  there exists a positive integer  $n$  such that  $n > x$ .*

*Proof.* If this were not so, some  $x$  would be an upper bound for  $\mathbf{P}$ , contradicting Theorem 1.28.

**THEOREM 1.30.** *If  $x > 0$  and if  $y$  is an arbitrary real number, there exists a positive integer  $n$  such that  $nx > y$ .*

*Proof.* Apply Theorem 1.29 with  $x$  replaced by  $y/x$ .

The property described in Theorem 1.30 is called the **Archimedean property** of the real-number system. Geometrically it means that any line segment, no matter how long, may be covered by a finite number of line segments of a given positive length, no matter how small. In other words, a small ruler used often enough can measure arbitrarily large distances. Archimedes realized that this was a fundamental property of the straight line and stated it explicitly as one of the axioms of geometry. In the 19th and 20th centuries, non-Archimedean geometries have been constructed in which this axiom is rejected.

From the Archimedean property, we can prove the following theorem, which will be useful in our discussion of integral calculus.

**THEOREM 1.31.** *If three real numbers  $a$ ,  $x$ , and  $y$  satisfy the inequalities*

$$(1.14) \quad a \leq x \leq a + \frac{y}{n}$$

*for every integer  $n \geq 1$ , then  $x = a$ .*

*Proof.* If  $x > a$ , Theorem 1.30 tells us that there is a positive integer  $n$  satisfying  $n(x - a) > y$ , contradicting (1.14). Hence we cannot have  $x > a$ , so we must have  $x = a$ .

### 13.11 Fundamental properties of the supremum and infimum

This section discusses three fundamental properties of the supremum and infimum that we shall use in our development of calculus. The first property states that any set of numbers with a supremum contains points arbitrarily close to its supremum; similarly, a set with an infimum contains points arbitrarily close to its infimum.

**THEOREM 1.32.** *Let  $h$  be a given positive number and let  $S$  be a set of real numbers.*

(a) *If  $S$  has a supremum, then for some  $x$  in  $S$  we have*

$$x > \sup S - h.$$

(b) *If  $S$  has an infimum, then for some  $x$  in  $S$  we have*

$$x < \inf S + h.$$

*Proof of (a).* If we had  $x \leq \sup S - h$  for *all*  $x$  in  $S$ , then  $\sup S - h$  would be an upper bound for  $S$  smaller than its least upper bound. Therefore we must have  $x > \sup S - h$  for at least one  $x$  in  $S$ . This proves (a). The proof of (b) is similar.

**THEOREM 1.33. ADDITIVE PROPERTY.** *Given nonempty subsets  $A$  and  $B$  of  $\mathbb{R}$ , let  $C$  denote the set*

$$C = \{a + b \mid a \in A, b \in B\}.$$

(a) *If each of  $A$  and  $B$  has a supremum, then  $C$  has a supremum, and*

$$\sup C = \sup A + \sup B.$$

(b) *If each of  $A$  and  $B$  has an infimum, then  $C$  has an infimum, and*

$$\inf C = \inf A + \inf B.$$

*Proof.* Assume each of  $A$  and  $B$  has a supremum. If  $c \in C$ , then  $c = a + b$ , where  $a \in A$  and  $b \in B$ . Therefore  $c \leq \sup A + \sup B$ ; so  $\sup A + \sup B$  is an upper bound for  $C$ . This shows that  $C$  has a supremum and that

$$\sup C \leq \sup A + \sup B.$$

Now let  $n$  be any positive integer. By Theorem 1.32 (with  $h = 1/n$ ) there is an  $a$  in  $A$  and a  $b$  in  $B$  such that

$$a > \sup A - \frac{1}{n}, \quad b > \sup B - \frac{1}{n}.$$

Adding these inequalities, we obtain

$$a + b > \sup A + \sup B - \frac{2}{n}, \quad \text{or} \quad \sup A + \sup B < a + b + \frac{2}{n} \leq \sup C + \frac{2}{n},$$

since  $a + b \leq \sup C$ . Therefore we have shown that

$$\sup C \leq \sup A + \sup B < \sup C + \frac{2}{n}$$

for every integer  $n \geq 1$ . By Theorem 1.31, we must have  $\sup C = \sup A + \sup B$ . This proves (a), and the proof of (b) is similar.

**THEOREM 1.34.** *Given two nonempty subsets  $S$  and  $T$  of  $\mathbb{R}$  such that*

$$s \leq t$$

*for every  $s$  in  $S$  and every  $t$  in  $T$ . Then  $S$  has a supremum, and  $T$  has an infimum, and they satisfy the inequality*

$$\sup S \leq \inf T.$$

*Proof.* Each  $t$  in  $T$  is an upper bound for  $S$ . Therefore  $S$  has a supremum which satisfies the inequality  $\sup S \leq t$  for all  $t$  in  $T$ . Hence  $\sup S$  is a lower bound for  $T$ , so  $T$  has an infimum which cannot be less than  $\sup S$ . In other words, we have  $\sup S \leq \inf T$ , as asserted.

### \*I 3.12 Exercises

1. If  $x$  and  $y$  are arbitrary real numbers with  $x < y$ , prove that there is at least **one** real  $z$  satisfying  $x < z < y$ .
2. If  $x$  is an arbitrary real number, prove that there are integers  $m$  and  $n$  such that  $m < x < n$ .
3. If  $x > 0$ , prove that there is a positive integer  $n$  such that  $1/n < x$ .
4. If  $x$  is an arbitrary real number, prove that there is exactly **one** integer  $n$  which satisfies the inequalities  $n \leq x < n + 1$ . This  $n$  is called the greatest integer in  $x$  and is denoted by  $[x]$ . For example,  $[5] = 5$ ,  $[\frac{5}{2}] = 2$ ,  $[-\frac{8}{3}] = -3$ .
5. If  $x$  is an arbitrary real number, prove that there is exactly **one** integer  $n$  which satisfies  $x \leq n < x + 1$ .
6. If  $x$  and  $y$  are arbitrary real numbers,  $x < y$ , prove that there exists at least **one** rational number  $r$  satisfying  $x < r < y$ , and **hence** infinitely many. This property is often described by saying that the rational numbers are **dense** in the real-number system.
7. If  $x$  is rational,  $x \neq 0$ , and  $y$  irrational, prove that  $x + y$ ,  $x - y$ ,  $xy$ ,  $x/y$ , and  $y/x$  are all irrational.
8. Is the sum or product of two irrational numbers always irrational?
9. If  $x$  and  $y$  are arbitrary real numbers,  $x < y$ , prove that there exists at least **one** irrational number  $z$  satisfying  $x < z < y$ , and **hence** infinitely many.
10. An integer  $n$  is called even if  $n = 2m$  for some integer  $m$ , and **odd** if  $n + 1$  is even. Prove the following statements :
  - (a) An integer cannot be both even and odd.
  - (b) Every integer is either even or odd.
  - (c) The sum or product of two even integers is even. What can you say about the sum or product of two odd integers?
  - (d) If  $n^2$  is even, so is  $n$ . If  $a^2 = 2b^2$ , where  $a$  and  $b$  are integers, then both  $a$  and  $b$  are even.
  - (e) Every rational number can be expressed in the form  $a/b$ , where  $a$  and  $b$  are integers, at least **one** of which is odd.
11. Prove that there is no rational number whose square is 2.

[Hint: Argue by contradiction. Assume  $(a/b)^2 = 2$ , where  $a$  and  $b$  are integers, at least **one** of which is odd. Use parts of Exercise 10 to deduce a contradiction.]



12. The Archimedean property of the real-number system was deduced as a consequence of the least-Upper-bound axiom. Prove that the set of rational numbers satisfies the Archimedean property but not the least-Upper-bound property. This shows that the Archimedean property does not imply the least-Upper-bound axiom.

### \*I 3.13 Existence of square roots of nonnegative real numbers

It was pointed out earlier that the equation  $x^2 = 2$  has no solutions among the rational numbers. With the help of Axiom 10, we can prove that the equation  $x^2 = a$  has a solution among the **real** numbers if  $a \geq 0$ . Each such  $x$  is called a **square root** of  $a$ .

First, let us see what we can say about square roots without using Axiom 10. Negative numbers cannot have square roots because if  $x^2 = a$ , then  $a$ , being a square, must be nonnegative (by Theorem 1.20). Moreover, if  $a = 0$ , then  $x = 0$  is the only square root (by Theorem 1.11). Suppose, then, that  $a > 0$ . If  $x^2 = a$ , then  $x \neq 0$  and  $(-x)^2 = a$ , so both  $x$  and its negative are square roots. In other words, if  $a$  has a square root, then it has two square roots, one positive and one negative. Also, it has **at most two** because if  $x^2 = a$  and  $y^2 = a$ , then  $x^2 = y^2$  and  $(x - y)(x + y) = 0$ , and so, by Theorem 1.11, either  $x = y$  or  $x = -y$ . Thus, if  $a$  has a square root, it has **exactly two**.

The existence of at least one square root can be deduced from an important theorem in calculus known as the intermediate-value theorem for continuous functions, but it may be instructive to see how the existence of a square root can be proved directly from Axiom 10.

**THEOREM 1.35. Every nonnegative real number  $a$  has a unique nonnegative square root.**

**Note:** If  $a \geq 0$ , we denote its nonnegative square root by  $a^{1/2}$  or by  $\sqrt{a}$ . If  $a > 0$ , the negative square root is  $-a^{1/2}$  or  $-\sqrt{a}$ .

**Proof.** If  $a = 0$ , then 0 is the only square root. Assume, then, that  $a > 0$ . Let  $S$  be the set of all positive  $x$  such that  $x^2 \leq a$ . Since  $(1 + a)^2 > a$ , the number  $1 + a$  is an upper bound for  $S$ . Also,  $S$  is nonempty because the number  $a/(1 + a)$  is in  $S$ ; in fact,  $a^2 \leq a(1 + a)^2$  and hence  $a^2/(1 + a)^2 \leq a$ . By Axiom 10,  $S$  has a least upper bound which we shall call  $b$ . Note that  $b \geq a/(1 + a)$  so  $b > 0$ . There are only three possibilities:  $b^2 > a$ ,  $b^2 < a$ , or  $b^2 = a$ .

Suppose  $b^2 > a$  and let  $c = b - (b^2 - a)/(2b) = \frac{1}{2}(b + a/b)$ . Then  $0 < c < b$  and  $c^2 = b^2 - (b^2 - a) + (b^2 - a)^2/(4b^2) = a + (b^2 - a)^2/(4b^2) > a$ . Therefore  $c^2 > x^2$  for each  $x$  in  $S$ , and hence  $c > x$  for each  $x$  in  $S$ . This means that  $c$  is an upper bound for  $S$ . Since  $c < b$ , we have a contradiction because  $b$  was the **least** upper bound for  $S$ . Therefore the inequality  $b^2 > a$  is impossible.

Suppose  $b^2 < a$ . Since  $b > 0$ , we may choose a positive number  $c$  such that  $c < b$  and such that  $c < (a - b^2)/(3b)$ . Then we have

$$(b + c)^2 = b^2 + c(2b + c) < b^2 + 3bc < b^2 + (a - b^2) = a$$

Therefore  $b + c$  is in  $S$ . Since  $b + c > b$ , this contradicts the fact that  $b$  is an upper bound for  $S$ . Therefore the inequality  $b^2 < a$  is impossible, and the only remaining alternative is  $b^2 = a$ .

### \*I 3.14 Roots of higher order. Rational powers

The least-Upper-bound axiom can also be used to show the existence of roots of higher order. For example, if  $n$  is a positive **odd** integer, then for **each** real  $x$  there is exactly **one** real  $y$  such that  $y^n = x$ . This  $y$  is called the  $n$ th **root** of  $x$  and is denoted by

$$(1.15) \quad y = x^{1/n} \quad \text{or} \quad y = \sqrt[n]{x}.$$

When  $n$  is **even**, the situation is slightly different. In this case, if  $x$  is negative, there is no real  $y$  such that  $y^n = x$  because  $y^n \geq 0$  for **all** real  $y$ . However, if  $x$  is positive, it can be shown that there is **one and only one** positive  $y$  such that  $y^n = x$ . This  $y$  is called **the positive  $n$ th root** of  $x$  and is denoted by the symbols in (1.15). Since  $n$  is even,  $(-y)^n = y^n$  and hence each  $x > 0$  has two real  $n$ th roots,  $y$  and  $-y$ . However, the symbols  $x^{1/n}$  and  $\sqrt[n]{x}$  are reserved for the **positive  $n$ th** root. We do not discuss the proofs of these statements here because they will be deduced later as consequences of the intermediate-value theorem for continuous functions (see Section 3.10).

If  $r$  is a positive rational number, say  $r = m/n$ , where  $m$  and  $n$  are positive integers, we define  $x^r$  to be  $(x^m)^{1/n}$ , the  $n$ th root of  $x^m$ , whenever this exists. If  $x \neq 0$ , we define  $x^{-r} = 1/x^r$  whenever  $x^r$  is defined. From these definitions, it is easy to verify that the usual laws of exponents are valid for rational exponents:  $x^r \cdot x^s = x^{r+s}$ ,  $(x^r)^s = x^{rs}$ , and  $(xy)^r = x^r y^r$ .

### \*I 3.15 Representation of real numbers by decimals

A real number of the form

$$(1.16) \quad r = a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_n}{10^n},$$

where  $a_0$  is a nonnegative integer and  $a_1, a_2, \dots, a_n$  are integers satisfying  $0 \leq a_i \leq 9$ , is usually written more briefly as follows:

$$r = a_0.a_1a_2 \cdots a_n.$$

This is said to be a **finite decimal representation** of  $r$ . For example,

$$\frac{1}{2} = \frac{5}{10} = 0.5, \quad \frac{1}{50} = \frac{2}{10^2} = 0.02, \quad \frac{29}{4} = 7 + \frac{2}{10} + \frac{5}{10^2} = 7.25.$$

Real numbers like these are necessarily rational and, in fact, they all have the form  $r = a/10^n$ , where  $a$  is an integer. However, not all rational numbers can be expressed with finite decimal representations. For example, if  $\frac{1}{3}$  could be so expressed, then we would have  $\frac{1}{3} = a/10^n$  or  $3a = 10^n$  for some integer  $a$ . But this is impossible since 3 is not a factor of any power of 10.

Nevertheless, we can approximate an arbitrary real number  $x > 0$  to any desired degree of accuracy by a sum of the form (1.16) if we take  $n$  large enough. The reason for this may be seen by the following geometric argument: If  $x$  is not an integer, then  $x$  lies between two consecutive integers, say  $a < x < a + 1$ . The segment joining  $a$  and  $a + 1$  may be

subdivided into ten equal parts. If  $x$  is not one of the subdivision points, then  $x$  must lie between two consecutive subdivision points. This gives us a pair of inequalities of the form

$$a_0 + \frac{a_1}{10} < x < a_0 + \frac{a_1 + 1}{10},$$

where  $a_1$  is an integer ( $0 \leq a_1 \leq 9$ ). Next we divide the segment joining  $a_0 + a_1/10$  and  $a_0 + (a_1 + 1)/10$  into ten equal parts (each of length  $10^{-2}$ ) and continue the process. If after a finite number of steps a subdivision point coincides with  $x$ , then  $x$  is a number of the form (1.16). Otherwise the process continues indefinitely, and it generates an infinite set of integers  $a_1, a_2, a_3, \dots$ . In this case, we say that  $x$  has the infinite decimal representation

$$x = a_0.a_1a_2a_3\cdots$$

At the  $n$ th stage,  $x$  satisfies the inequalities

$$a_0 + \frac{a_1}{10} + \cdots + \frac{a_n}{10^n} < x < a_0 + \frac{a_1}{10} + \cdots + \frac{a_n + 1}{10^n},$$

This gives us two approximations to  $x$ , one from above and one from below, by finite decimals that differ by  $10^{-n}$ . Therefore we can achieve any desired degree of accuracy in our approximations by taking  $n$  large enough.

When  $x = \frac{1}{3}$ , it is easy to verify that  $a_0 = 0$  and  $a_n = 3$  for all  $n \geq 1$ , and hence the corresponding infinite decimal expansion is

$$\frac{1}{3} = 0.333\cdots$$

Every irrational number has an infinite decimal representation. For example, when  $x = \sqrt{2}$  we may calculate by trial and error as many digits in the expansion as we wish. Thus,  $\sqrt{2}$  lies between 1.4 and 1.5, because  $(1.4)^2 < 2 < (1.5)^2$ . Similarly, by squaring and comparing with 2, we find the following further approximations:

$$1.41 < \sqrt{2} < 1.42, \quad 1.414 < \sqrt{2} < 1.415, \quad 1.4142 < \sqrt{2} < 1.4143.$$

Note that the foregoing process generates a succession of intervals of lengths  $10^{-1}, 10^{-2}, 10^{-3}, \dots$ , each contained in the preceding and each containing the point  $x$ . This is an example of what is known as a sequence of **nested intervals**, a concept that is sometimes used as a basis for constructing the irrational numbers from the rational numbers.

Since we shall do very little with decimals in this book, we shall not develop their properties in any further detail except to mention how decimal expansions may be defined analytically with the help of the least-upper-bound axiom.

If  $x$  is a given positive real number, let  $a_0$  denote the largest integer  $\leq x$ . Having chosen  $a_0$ , we let  $a_1$  denote the largest integer such that

$$a_0 + \frac{a_1}{10} \leq x.$$

More generally, having chosen  $a_0, a_1, \dots, a_{n-1}$ , we let  $a_n$  denote the largest integer such that

$$(1.17) \quad a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_n}{10^n} \leq x.$$

Let  $S$  denote the set of all numbers

$$(1.18) \quad a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_n}{10^n}$$

obtained in this way for  $n = 0, 1, 2, \dots$ . Then  $S$  is nonempty and bounded above, and it is easy to verify that  $x$  is actually the least upper bound of  $S$ . The integers  $a_0, a_1, a_2, \dots$  so obtained may be used to define a decimal expansion of  $x$  if we write

$$x = a_0.a_1a_2a_3 \cdots$$

to mean that the  $n$ th digit  $a_n$  is the largest integer satisfying (1.17). For example, if  $x = \frac{1}{8}$ , we find  $a_0 = 0, a_1 = 1, a_2 = 2, a_3 = 5$ , and  $a_n = 0$  for all  $n \geq 4$ . Therefore we may write

$$\frac{1}{8} = 0.125000 \cdots$$

If in (1.17) we replace the inequality sign  $\leq$  by  $<$ , we obtain a slightly different definition of decimal expansions. The least upper bound of all numbers of the form (1.18) is again  $x$ , although the integers  $a_0, a_1, a_2, \dots$  need not be the same as those which satisfy (1.17). For example, if this second definition is applied to  $x = \frac{1}{8}$ , we find  $a_0 = 0, a_1 = 1, a_2 = 2, a_3 = 4$ , and  $a_n = 9$  for all  $n \geq 4$ . This leads to the infinite decimal representation

$$\frac{1}{8} = 0.124999 \cdots$$

The fact that a real number might have two different decimal representations is merely a reflection of the fact that two different sets of real numbers can have the same supremum.

## Part 4. Mathematical Induction, Summation Notation, and Related Topics

### 14.1 An example of a proof by mathematical induction

There is no **largest** integer because when we add 1 to an integer  $k$ , we obtain  $k + 1$ , which is larger than  $k$ . Nevertheless, starting with the number 1, we can reach any positive integer whatever in a finite number of steps, passing successively from  $k$  to  $k + 1$  at each step. This is the basis for a type of reasoning that mathematicians call **proof by induction**. We shall illustrate the use of this method by proving the pair of inequalities used in Section

11.3 in the computation of the area of a parabolic segment, namely

$$(1.19) \quad 1^2 + 2^2 + \cdots + (n-1)^2 < \frac{n^3}{3} < 1^2 + 2^2 + \cdots + n^2.$$

Consider the leftmost inequality first, and let us refer to this formula as  $A(n)$  (an assertion involving  $n$ ). It is easy to verify this assertion directly for the first few values of  $n$ . Thus, for example, when  $n$  takes the values 1, 2, and 3, the assertion becomes

$$A(1): 0 < \frac{1^3}{3}, \quad A(2): 1^2 < \frac{2^3}{3}, \quad A(3): 1^2 + 2^2 < \frac{3^3}{3},$$

provided we agree to interpret the sum on the left as 0 when  $n = 1$ .

Our object is to prove that  $A(n)$  is true for every positive integer  $n$ . The procedure is as follows: Assume the assertion has been proved for a particular value of  $n$ , say for  $n = k$ . That is, assume we have proved

$$A(k): 1^2 + 2^2 + \cdots + (k-1)^2 < \frac{k^3}{3}$$

for a fixed  $k \geq 1$ . Now using this, we shall deduce the corresponding result for  $k+1$ :

$$A(k+1): 1^2 + 2^2 + \cdots + k^2 < \frac{(k+1)^3}{3}.$$

Start with  $A(k)$  and add  $k^2$  to both sides. This gives the inequality

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

To obtain  $A(k+1)$  as a consequence of this, it suffices to show that

$$\frac{k^3}{3} + k^2 < \frac{(k+1)^3}{3}.$$

But this follows at once from the equation

$$\frac{(k+1)^3}{3} = \frac{k^3 + 3k^2 + 3k + 1}{3} = \frac{k^3}{3} + k^2 + k + \frac{1}{3}.$$

Therefore we have shown that  $A(k+1)$  follows from  $A(k)$ . Now, since  $A(1)$  has been verified directly, we conclude that  $A(2)$  is also true. Knowing that  $A(2)$  is true, we conclude that  $A(3)$  is true, and so on. Since every integer can be reached in this way,  $A(n)$  is true for all positive integers  $n$ . This proves the leftmost inequality in (1.19). The rightmost inequality can be proved in the same way.

## 14.2 The principle of mathematical induction

The reader should make certain that he understands the *pattern* of the foregoing proof. First we proved the assertion  $A(n)$  for  $n = 1$ . Next we showed that *if* the assertion is true for a particular integer, *then* it is *also* true for the next integer. From this, we concluded that the assertion is true for *all* positive integers.

The idea of induction may be illustrated in many nonmathematical ways. For example, imagine a row of toy soldiers, numbered consecutively, and suppose they are so arranged that if *any one* of them falls, say the *one* labeled  $k$ , it will knock over the next *one*, labeled  $k + 1$ . Then anyone *can* visualize what would happen if soldier number 1 were toppled backward. It is also clear that if a later soldier were knocked over first, say the *one* labeled  $n_1$ , then *all* soldiers behind *him* would fall. This illustrates a slight generalization of the method of induction which *can* be described in the following way.

*Method of proof by induction.* Let  $A(n)$  be an assertion involving an integer  $n$ . We conclude that  $A(n)$  is true for every  $n \geq n_1$  if we can perform the following two steps:

- (a) Prove that  $A(n_1)$  is true.
- (b) Let  $k$  be an arbitrary but fixed integer  $\geq n_1$ . Assume that  $A(k)$  is true and prove that  $A(k + 1)$  is *also* true.

In actual practice  $n_1$  is usually 1. The logical justification for this method of proof is the following theorem about real numbers.

**THEOREM 1.36. PRINCIPLE OF MATHEMATICAL INDUCTION.** *Let  $S$  be a set of positive integers which has the following two properties:*

- (a) *The number 1 is in the set  $S$ .*
- (b) *If an integer  $k$  is in  $S$ , then so is  $k + 1$ .*

*Then every positive integer is in the set  $S$ .*

*Proof.* Properties (a) and (b) tell us that  $S$  is an inductive set. But the positive integers were defined to be exactly those real numbers which belong to every inductive set. (See Section 1 3.6.) Therefore  $S$  contains every positive integer.

Whenever we carry out a proof of an assertion  $A(n)$  for *all*  $n \geq 1$  by mathematical induction, we are applying Theorem 1.36 to the set  $S$  of *all* the integers for which the assertion is true. If we want to prove that  $A(n)$  is true only for  $n \geq n_1$ , we apply Theorem 1.36 to the set of  $n$  for which  $A(n + n_1 - 1)$  is true.

## \*I 4.3 The well-ordering principle

There is another important property of the positive integers, called the well-ordering principle, that is *also* used as a basis for proofs by induction. It *can* be stated as follows.

**THEOREM 1.37. WELL-ORDERING PRINCIPLE.** *Every nonempty set of positive integers contains a smallest member.*

Note that the well-ordering principle refers to sets of *positive* integers. The theorem is not true for arbitrary sets of integers. For example, the set of *all* integers has no smallest member.

The well-ordering principle can be deduced from the principle of induction. This is demonstrated in Section 14.5. We conclude this section with an example showing how the well-ordering principle can be used to prove theorems about positive integers.

Let  $A(n)$  denote the following assertion:

$$A(n): 1^2 + 2^2 + \dots + n^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}.$$

Again, we note that  $A(1)$  is true, since

$$1^2 = \frac{1}{3} + \frac{1}{2} + \frac{1}{6}.$$

Now there are only two possibilities. We have either

- (i)  $A(n)$  is true for every positive integer  $n$ , or
- (ii) there is at least one positive integer  $n$  for which  $A(n)$  is false.

We shall prove that alternative (ii) leads to a contradiction. Assume (ii) holds. Then by the well-ordering principle, there must be a *smallest* positive integer, say  $k$ , for which  $A(k)$  is false. (We apply the well-ordering principle to the set of all positive integers  $n$  for which  $A(n)$  is false. Statement (ii) says that this set is nonempty.) This  $k$  must be greater than 1, because we have verified that  $A(1)$  is true. Also, the assertion must be true for  $k - 1$ , since  $k$  was the smallest integer for which  $A(k)$  is false; therefore we may write

$$A(k-1): 1^2 + 2^2 + \dots + (k-1)^2 = \frac{(k-1)^3}{3} + \frac{(k-1)^2}{2} + \frac{k-1}{6}.$$

Adding  $k^2$  to both sides and simplifying the right-hand side, we find

$$1^2 + 2^2 + \dots + k^2 = \frac{k^3}{3} + \frac{k^2}{2} + \frac{k}{6}.$$

But this equation states that  $A(k)$  is true; therefore we have a contradiction, because  $k$  is an integer for which  $A(k)$  is false. In other words, statement (ii) leads to a contradiction. Therefore (i) holds, and this proves that the identity in question is valid for all values of  $n \geq 1$ . An immediate consequence of this identity is the rightmost inequality in (1.19).

A proof like this which makes use of the well-ordering principle is also referred to as a proof by induction. Of course, the proof could also be put in the more usual form in which we verify  $A(1)$  and then pass from  $A(k)$  to  $A(k+1)$ .

## 14.4 Exercises

1. Prove the following formulas by induction :

(a)  $1 + 2 + 3 + \dots + n = n(n+1)/2$ .

(b)  $1 + 3 + 5 + \dots + (2n-1) = n^2$ .

(c)  $1^3 + 2^3 + 3^3 + \dots + n^3 = (1 + 2 + 3 + \dots + n)^2$ .

(d)  $1^3 + 2^3 + \dots + (n-1)^3 < n^4/4 < 1^3 + 2^3 + \dots + n^3$ .

2. Note that

$$\begin{aligned}1 &= 1, \\1 - 4 &= -(1 + 2), \\1 - 4 + 9 &= 1 + 2 + 3, \\1 - 4 + 9 - 16 &= -(1 + 2 + 3 + 4).\end{aligned}$$

Guess the general law suggested and prove it by induction.

3. Note that

$$\begin{aligned}1 + \frac{1}{2} &= 2 - \frac{1}{2}, \\1 + \frac{1}{2} + \frac{1}{4} &= 2 - \frac{1}{4}, \\1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} &= 2 - \frac{1}{8}.\end{aligned}$$

Guess the general law suggested and prove it by induction.

4. Note that

$$\begin{aligned}1 - \frac{1}{2} &= \frac{1}{2}, \\(1 - \frac{1}{2})(1 - \frac{1}{3}) &= \frac{1}{3}, \\(1 - \frac{1}{2})(1 - \frac{1}{3})(1 - \frac{1}{4}) &= \frac{1}{4}.\end{aligned}$$

Guess the general law suggested and prove it by induction.

5. Guess a general law which simplifies the product

$$\left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{9}\right)\left(1 - \frac{1}{16}\right) \cdots \left(1 - \frac{1}{n^2}\right)$$

and prove it by induction.

6. Let  $A(n)$  denote the statement:  $1 + 2 + \dots + n = \frac{1}{8}(2n + 1)^2$ .  
 (a) Prove that if  $A(k)$  is true for an integer  $k$ , then  $A(k + 1)$  is also true.  
 (b) Criticize the statement: "By induction it follows that  $A(n)$  is true for all  $n$ ."  
 (c) Amend  $A(n)$  by changing the equality to an inequality that is true for all positive integers  $n$ .  
 7. Let  $n_1$  be the smallest positive integer  $n$  for which the inequality  $(1 + x)^n > 1 + nx + nx^2$  is true for all  $x > 0$ . Compute  $n_1$ , and prove that the inequality is true for all integers  $n \geq n_1$ .  
 8. Given positive real numbers  $a_1, a_2, a_3, \dots$ , such that  $a_n \leq ca_{n-1}$  for all  $n \geq 2$ , where  $c$  is a fixed positive number, use induction to prove that  $a_n \leq a_1 c^{n-1}$  for all  $n \geq 1$ .  
 9. Prove the following statement by induction: If a line of unit length is given, then a line of length  $\sqrt{n}$  can be constructed with straightedge and compass for each positive integer  $n$ .  
 10. Let  $b$  denote a fixed positive integer. Prove the following statement by induction: For every integer  $n \geq 0$ , there exist nonnegative integers  $q$  and  $r$  such that

$$n = qb + r, \quad 0 \leq r < b.$$

11. Let  $n$  and  $d$  denote integers. We say that  $d$  is a **divisor** of  $n$  if  $n = cd$  for some integer  $c$ . An integer  $n$  is called a **prime** if  $n > 1$  and if the only positive divisors of  $n$  are 1 and  $n$ . Prove, by induction, that every integer  $n > 1$  is either a prime or a product of primes.  
 12. Describe the fallacy in the following "proof" by induction:

**Statement.** Given any collection of  $n$  blonde girls. If at least one of the girls has blue eyes, then all  $n$  of them have blue eyes.

**"Proof."** The statement is obviously true when  $n = 1$ . The step from  $k$  to  $k + 1$  can be illustrated by going from  $n = 3$  to  $n = 4$ . Assume, therefore, that the statement is true



when  $n = 3$  and let  $G_1, G_2, G_3, G_4$  be four blonde girls, at least one of which, say  $G_1$ , has blue eyes. Taking  $G_1, G_2$ , and  $G_3$  together and using the fact that the statement is true when  $n = 3$ , we find that  $G_2$  and  $G_3$  also have blue eyes. Repeating the process with  $G_1, G_2$ , and  $G_4$ , we find that  $G_4$  has blue eyes. Thus all four have blue eyes. A similar argument allows us to make the step from  $k$  to  $k + 1$  in general.

*Corollary.* All blonde girls have blue eyes.

*Proof.* Since there exists at least one blonde girl with blue eyes, we can apply the foregoing result to the collection consisting of all blonde girls.

*Note:* This example is from G. Pólya, who suggests that the reader may want to test the validity of the statement by experiment.

### \*14.5 Proof of the well-ordering principle

In this section we deduce the well-ordering principle from the principle of induction.

Let  $T$  be a nonempty collection of positive integers. We want to prove that  $T$  has a smallest member, that is, that there is a positive integer  $t_0$  in  $T$  such that  $t_0 \leq t$  for all  $t$  in  $T$ .

Suppose  $T$  has no smallest member. We shall show that this leads to a contradiction. The integer 1 cannot be in  $T$  (otherwise it would be the smallest member of  $T$ ). Let  $S$  denote the collection of all positive integers  $n$  such that  $n < t$  for all  $t$  in  $T$ . Now 1 is in  $S$  because  $1 < t$  for all  $t$  in  $T$ . Next, let  $k$  be a positive integer in  $S$ . Then  $k < t$  for all  $t$  in  $T$ . We shall prove that  $k + 1$  is also in  $S$ . If this were not so, then for some  $t_1$  in  $T$  we would have  $t_1 \leq k + 1$ . Since  $T$  has no smallest member, there is an integer  $t_2$  in  $T$  such that  $t_2 < t_1$ , and hence  $t_2 < k + 1$ . But this means that  $t_2 \leq k$ , contradicting the fact that  $k < t$  for all  $t$  in  $T$ . Therefore  $k + 1$  is in  $S$ . By the induction principle,  $S$  contains all positive integers. Since  $T$  is nonempty, there is a positive integer  $t$  in  $T$ . But this  $t$  must also be in  $S$  (since  $S$  contains all positive integers). It follows from the definition of  $S$  that  $t < t$ , which is a contradiction. Therefore, the assumption that  $T$  has no smallest member leads to a contradiction. It follows that  $T$  must have a smallest member, and in turn this proves that the well-ordering principle is a consequence of the principle of induction.

### 14.6 The summation notation

In the calculations for the area of the parabolic segment, we encountered the sum

$$(1.20) \quad 1^2 + 2^2 + 3^2 + \cdots + n^2.$$

Note that a typical term in this sum is of the form  $k^2$ , and we get all the terms by letting  $k$  run through the values  $1, 2, 3, \dots, n$ . There is a very useful and convenient notation which enables us to write sums like this in a more compact form. This is called the *summation notation* and it makes use of the Greek letter sigma,  $\sum$ . Using summation notation, we can write the sum in (1.20) as follows:

$$\sum_{k=1}^n k^2.$$

This symbol is read: "The sum of  $k^2$  for  $k$  running from 1 to  $n$ ." The numbers appearing under and above the sigma tell us the range of values taken by  $k$ . The letter  $k$  itself is

referred to as the *index of summation*. Of course, it is not important that we use the letter  $k$ ; any other convenient letter may take its place. For example, instead of  $\sum_{k=1}^n k^2$  we could write  $\sum_{i=1}^n i^2$ ,  $\sum_{j=1}^n j^2$ ,  $\sum_{m=1}^n m^2$ , etc., all of which are considered as alternative notations for the same thing. The letters  $i, j, k, m$ , etc. that are used in this way are called *dummy indices*. It would not be a good idea to use the letter  $n$  for the dummy index in this particular example because  $n$  is already being used for the number of terms.

More generally, when we want to form the sum of several real numbers, say  $a_1, a_2, \dots, a_n$ , we denote such a sum by the symbol

$$(1.21) \quad a_1 + a_2 + \dots + a_n,$$

which, using summation notation, can be written as follows:

$$(1.22) \quad \sum_{k=1}^n a_k.$$

For example, we have

$$\begin{aligned} \sum_{k=1}^4 a_k &= a_1 + a_2 + a_3 + a_4, \\ \sum_{i=1}^5 x_i &= x_1 + x_2 + x_3 + x_4 + x_5 \end{aligned}$$

Sometimes it is convenient to begin summations from 0 or from some value of the index beyond 1. For example, we have

$$\begin{aligned} \sum_{i=0}^4 x_i &= x_0 + x_1 + x_2 + x_3 + x_4, \\ \sum_{n=2}^5 n^3 &= 2^3 + 3^3 + 4^3 + 5^3. \end{aligned}$$

Other uses of the summation notation are illustrated below:

$$\begin{aligned} \sum_{m=0}^4 x^{m+1} &= x + x^2 + x^3 + x^4 + x^5, \\ \sum_{j=1}^6 2^{j-1} &= 1 + 2 + 2^2 + 2^3 + 2^4 + 2^5. \end{aligned}$$

To emphasize once more that the choice of dummy index is unimportant, we note that the last sum may also be written in each of the following forms:

$$\sum_{q=1}^6 2^{q-1} = \sum_{r=0}^5 2^r = \sum_{n=0}^5 2^{5-n} = \sum_{k=1}^6 2^{6-k}.$$

*Note:* From a strictly logical standpoint, the symbols in (1.21) and (1.22) do not appear among the primitive symbols for the real-number system. In a more careful treatment, we could **define** these new symbols in terms of the primitive undefined symbols of our system.

This may be done by a process known as *definition by induction* which, like proof by induction, consists of two parts:

(a) We define

$$\sum_{k=1}^1 a_k = a_1.$$

(b) Assuming that we have defined  $\sum_{k=1}^n a_k$  for a fixed  $n \geq 1$ , we further define

$$\sum_{k=1}^{n+1} a_k = \left( \sum_{k=1}^n a_k \right) + a_{n+1}.$$

To illustrate, we may take  $n = 1$  in (b) and use (a) to obtain

$$\sum_{k=1}^2 a_k = \sum_{k=1}^1 a_k + a_2 = a_1 + a_2.$$

Now, having defined  $\sum_{k=1}^2 a_k$ , we can use (b) again with  $n = 2$  to obtain

$$\sum_{k=1}^3 a_k = \sum_{k=1}^2 a_k + a_3 = (a_1 + a_2) + a_3.$$

By the associative law for addition (Axiom 2), the sum  $(a_1 + a_2) + a_3$  is the same as  $a_1 + (a_2 + a_3)$ , and therefore there is no danger of confusion if we drop the parentheses and simply write  $a_1 + a_2 + a_3$  for  $\sum_{k=1}^3 a_k$ . Similarly, we have

$$\sum_{k=1}^4 a_k = \sum_{k=1}^3 a_k + a_4 = (a_1 + a_2 + a_3) + a_4.$$

In this case we can **prove** that the sum  $(a_1 + a_2 + a_3) + a_4$  is the same as  $(a_1 + a_2) + (a_3 + a_4)$  or  $a_1 + (a_2 + a_3 + a_4)$ , and therefore the parentheses can be dropped again without danger of ambiguity, and we agree to write

$$\sum_{k=1}^4 a_k = a_1 + a_2 + a_3 + a_4.$$

Continuing in this way, we find that (a) and (b) together give us a **complete** definition of the symbol in (1.22). The notation in (1.21) is considered to be merely an alternative way of writing (1.22). It is justified by a general associative law for addition which we shall not attempt to state or to prove here.

The reader should notice that *definition by induction* and *proof by induction* involve the same underlying idea. A definition by induction is also called a **recursive definition**.

## 14.7 Exercises

1. Find the numerical values of the following sums :

$$\begin{array}{lll} \text{(a)} \sum_{k=1}^4 k, & \text{(c)} \sum_{r=0}^3 2^{2r+1}, & \text{(e)} \sum_{i=0}^5 (2i + 1), \\ \text{(b)} \sum_{n=2}^5 2^{n-2}, & \text{(d)} \sum_{n=1}^4 n^n, & \text{(f)} \sum_{k=1}^5 \frac{1}{k(k+1)}. \end{array}$$

2. Establish the following properties of the summation notation:

$$(a) \sum_{k=1}^n (a_k + b_k) = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k \quad (\text{additive property}).$$

$$(b) \sum_{k=1}^n (ca_k) = c \sum_{k=1}^n a_k \quad (\text{homogeneous property}).$$

$$(c) \sum_{k=1}^n (a_k - a_{k-1}) = a_n - a_0 \quad (\text{telescoping property}).$$

Use the properties in Exercise 2 whenever possible to **derive** the formulas in Exercises 3 through 8.

$$3. \sum_{k=1}^n 1 = n. \quad (\text{This means } \sum_{k=1}^n a_k, \text{ where each } a_k = 1.)$$

$$4. \sum_{k=1}^n (2k - 1) = n^2. \quad [\text{Hint: } 2k - 1 = k^2 - (k - 1)^2.]$$

$$5. \sum_{k=1}^n k = \frac{n^2}{2} + \frac{n}{2}. \quad [\text{Hint: Use Exercises 3 and 4.}]$$

$$6. \sum_{k=1}^n k^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}. \quad [\text{Hint: } k^3 - (k - 1)^3 = 3k^2 - 3k + 1.]$$

$$7. \sum_{k=1}^n k^3 = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4}.$$

$$8. (a) \sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x} \quad \text{if } x \neq 1. \quad \text{Note: } x^0 \text{ is defined to be } 1.$$

[Hint: Apply Exercise 2 to  $(1 - x) \sum_{k=0}^n x^k$ .]

(b) What is the sum equal to when  $x = 1$ ?

9. Prove, by induction, that the sum  $\sum_{k=1}^{2n} (-1)^k (2k + 1)$  is proportional to  $n$ , and find the constant of proportionality.

10. (a) Give a reasonable definition of the symbol  $\sum_{k=m}^{m+n} a_k$ .  
 (b) Prove, by induction, that for  $n \geq 1$  we have

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

11. Determine whether **each** of the following statements is true or false. In **each** case give a reason for your decision.

$$(a) \sum_{n=0}^{100} n^4 = \sum_{n=1}^{100} n^4.$$

$$(d) \sum_{i=1}^{100} (i + 1)^2 = \sum_{i=0}^{99} i^2.$$

$$(b) \sum_{j=0}^{100} 2 = 200.$$

$$(e) \sum_{k=1}^{100} k^3 = \left( \sum_{k=1}^{100} k \right) \cdot \left( \sum_{k=1}^{100} k^2 \right).$$

$$(c) \sum_{k=0}^{100} (2 + k) = 2 + \sum_{k=0}^{100} k.$$

$$(f) \sum_{k=0}^{100} k^3 = \left( \sum_{k=0}^{100} k \right)^3.$$

12. Guess and prove a general rule which simplifies the sum

$$\sum_{k=1}^n \frac{1}{k(k+1)}.$$

13. Prove that  $2(\sqrt{n+1} - \sqrt{n}) < \frac{1}{\sqrt{n}} < 2(\sqrt{n} - \sqrt{n-1})$  if  $n \geq 1$ . Then use this to prove that

$$2\sqrt{m} - 2 < \sum_{n=1}^m \frac{1}{\sqrt{n}} < 2\sqrt{m} - 1$$

if  $m \geq 2$ . In particular, when  $m = 10^6$ , the sum lies between 1998 and 1999.

## 14.8 Absolute values and the triangle inequality

Calculations with inequalities arise quite frequently in calculus. They are of particular importance in dealing with the notion of **absolute value**. If  $x$  is a real number, the **absolute value** of  $x$  is a nonnegative real number denoted by  $|x|$  and defined as follows:

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

Note that  $-|x| \leq x \leq |x|$ . When real numbers are represented geometrically on a real axis, the number  $|x|$  is called the **distance** of  $x$  from 0. If  $a > 0$  and if a point  $x$  lies between  $-a$  and  $a$ , then  $|x|$  is nearer to 0 than  $a$  is. The analytic statement of this fact is given by the following theorem.

**THEOREM 1.38.** *If  $a \geq 0$ , then  $|x| \leq a$  if and only if  $-a \leq x \leq a$ .*

**Proof.** There are two statements to prove: first, that the inequality  $|x| \leq a$  implies the two inequalities  $-a \leq x \leq a$  and, conversely, that  $-a \leq x \leq a$  implies  $|x| \leq a$ .

Suppose  $|x| \leq a$ . Then we also have  $-a \leq -|x|$ . But either  $x = |x|$  or  $x = -|x|$  and hence  $-a \leq -|x| \leq x \leq |x| \leq a$ . This proves the first statement.

To prove the converse, assume  $-a \leq x \leq a$ . Then if  $x \geq 0$ , we have  $|x| = x \leq a$ , whereas if  $x \leq 0$ , we have  $|x| = -x \leq a$ . In either case we have  $|x| \leq a$ , and this completes the proof.

Figure 1.9 illustrates the geometrical significance of this theorem.

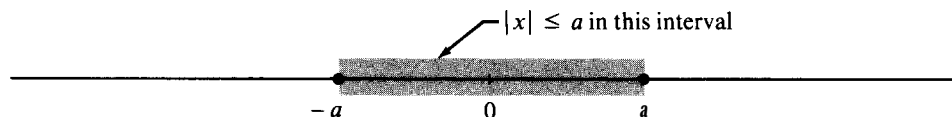


FIGURE 1.9 Geometrical significance of Theorem 1.38.

As a consequence of Theorem 1.38, it is easy to derive an important inequality which states that the **absolute value** of a sum of two real numbers **cannot** exceed the sum of their **absolute values**.

**THEOREM 1.39.** *For arbitrary real numbers  $x$  and  $y$ , we have*

$$|x + y| \leq |x| + |y|,$$

*Note:* This property is called the **triangle inequality**, because when it is generalized to vectors it states that the length of any side of a triangle is less than or equal to the sum of the lengths of the other two sides.

*Proof.* Adding the inequalities  $-|x| \leq x \leq |x|$  and  $-|y| \leq y \leq |y|$ , we obtain

$$-(|x| + |y|) \leq x + y \leq |x| + |y|,$$

and hence, by Theorem 1.38, we conclude that  $|x + y| \leq |x| + |y|$ .

If we take  $x = a - c$  and  $y = c - b$ , then  $x + y = a - b$  and the triangle inequality becomes

$$|a - b| \leq |a - c| + |b - c|.$$

This form of the triangle inequality is often used in practice.

Using mathematical induction, we may extend the triangle inequality as follows:

**THEOREM 1.40.** *For arbitrary real numbers  $a_1, a_2, \dots, a_n$ , we have*

$$\left| \sum_{k=1}^n a_k \right| \leq \sum_{k=1}^n |a_k|.$$

*Proof.* When  $n = 1$  the inequality is trivial, and when  $n = 2$  it is the triangle inequality. Assume, then, that it is true for  $n$  real numbers. Then for  $n + 1$  real numbers  $a_1, a_2, \dots, a_{n+1}$ , we have

$$\left| \sum_{k=1}^{n+1} a_k \right| = \left| \sum_{k=1}^n a_k + a_{n+1} \right| \leq \left| \sum_{k=1}^n a_k \right| + |a_{n+1}| \leq \sum_{k=1}^n |a_k| + |a_{n+1}| = \sum_{k=1}^{n+1} |a_k|.$$

Hence the theorem is true for  $n + 1$  numbers if it is true for  $n$ . By induction, it is true for every positive integer  $n$ .

The next theorem describes an important inequality that we shall use later in connection with our study of vector algebra.

**THEOREM 1.41. THE CAUCHY-SCHWARZ INEQUALITY.** *If  $a_1, \dots, a_n$  and  $b_1, \dots, b_n$  are arbitrary real numbers, we have*

$$(1.23) \quad \left( \sum_{k=1}^n a_k b_k \right)^2 \leq \left( \sum_{k=1}^n a_k^2 \right) \left( \sum_{k=1}^n b_k^2 \right).$$

**The equality sign holds if and only if there is a real number  $x$  such that  $a_k x + b_k = 0$  for each  $k = 1, 2, \dots, n$ .**

**Proof.** We have  $\sum_{k=1}^n (a_k x + b_k)^2 \geq 0$  for every real  $x$  because a sum of squares can never be negative. This may be written in the form

$$(1.24) \quad Ax^2 + 2Bx + C \geq 0,$$

where

$$A = \sum_{k=1}^n a_k^2, \quad B = \sum_{k=1}^n a_k b_k, \quad C = \sum_{k=1}^n b_k^2.$$

We wish to prove that  $B^2 \leq AC$ . If  $A = 0$ , then each  $a_k = 0$ , so  $B = 0$  and the result is trivial. If  $A \neq 0$ , we may complete the square and write

$$Ax^2 + 2Bx + C = A \left( x + \frac{B}{A} \right)^2 + \frac{AC - B^2}{A}.$$

The right side has its smallest value when  $x = -B/A$ . Putting  $x = -B/A$  in (1.24), we obtain  $B^2 \leq AC$ . This proves (1.23). The reader should verify that the equality sign holds if and only if there is an  $x$  such that  $a_k x + b_k = 0$  for each  $k$ .

## 14.9 Exercises

- Prove each of the following properties of absolute values.
 

(a) $ x  = 0$ if and only if $x = 0$ .	(f) $ xy  =  x   y $ .
(b) $ -x  =  x $ .	(g) $ x/y  =  x / y $ if $y \neq 0$ .
(c) $ x - y  =  y - x $ .	(h) $ x - y  \leq  x  +  y $ .
(d) $ x ^2 = x^2$ .	(i) $ x  -  y  \leq  x - y $ .
(e) $ x  = \sqrt{x^2}$ .	(j) $  x  -  y   \leq  x - y $ .
- Each inequality  $(a_i)$ , listed below, is equivalent to exactly one inequality  $(b_j)$ . For example,  $|x| < 3$  if and only if  $-3 < x < 3$ , and hence  $(a_1)$  is equivalent to  $(b_2)$ . Determine all equivalent pairs.
 

$(a_1)  x  < 3$ .	$(b_1) 4 < x < 6$ .
$(a_2)  x - 1  < 3$ .	$(b_2) -3 < x < 3$ .
$(a_3)  3 - 2x  < 1$ .	$(b_3) x > 3$ or $x < -1$ .
$(a_4)  1 + 2x  \leq 1$ .	$(b_4) x > 2$ .
$(a_5)  x - 1  > 2$ .	$(b_5) -2 < x < 4$ .
$(a_6)  x + 2  \geq 5$ .	$(b_6) -\sqrt{3} \leq x \leq -1$ or $1 \leq x \leq \sqrt{3}$ .
$(a_7)  5 - x^{-1}  < 1$ .	$(b_7) 1 < x < 2$ .
$(a_8)  x - 5  <  x + 1 $ .	$(b_8) x \leq -7$ or $x \geq 3$ .
$(a_9)  x^2 - 2  \leq 1$ .	$(b_9) \frac{1}{6} < x < \frac{1}{4}$ .
$(a_{10}) x < x^2 - 12 < 4x$ .	$(b_{10}) -1 \leq x \leq 0$ .
- Determine whether each of the following is true or false. In each case give a reason for your decision.
  - $x < 5$  implies  $|x| < 5$ .
  - $|x - 5| < 2$  implies  $3 < x < 7$ .
  - $|1 + 3x| \leq 1$  implies  $x \geq -\frac{2}{3}$ .
  - There is no real  $x$  for which  $|x - 1| = |x - 2|$ .
  - For every  $x > 0$  there is a  $y > 0$  such that  $|2x + y| = 5$ .
- Show that the equality sign holds in the Cauchy-Schwarz inequality if and only if there is a real number  $x$  such that  $a_k x + b_k = 0$  for every  $k = 1, 2, \dots, n$ .

### \*I 4.10 Miscellaneous exercises involving induction

In this section we assemble a number of miscellaneous facts whose proofs are good exercises in the use of mathematical induction. Some of these exercises may serve as a basis for supplementary classroom discussion.

**Factorials and binomial coefficients.** The symbol  $n!$  (read “ $n$  factorial”) may be defined by induction as follows:  $0! = 1$ ,  $n! = (n-1)!n$  if  $n \geq 1$ . Note that  $n! = 1 \cdot 2 \cdot 3 \cdots n$ .

If  $0 \leq k \leq n$ , the **binomial coefficient**  $\binom{n}{k}$  is defined as follows:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

**Note:** Sometimes  ${}_nC_k$  is written for  $\binom{n}{k}$ . These numbers appear as coefficients in the binomial theorem. (See Exercise 4 below.)

- Compute the values of the following binomial coefficients :  
 (a)  $\binom{5}{3}$ , (b)  $\binom{7}{0}$ , (c)  $\binom{7}{1}$ , (d)  $\binom{7}{2}$ , (e)  $\binom{17}{14}$ , (f)  $\binom{0}{0}$ .
- (a) Show that  $\binom{n}{k} = \binom{n}{n-k}$ . (c) Find  $k$ , given that  $\binom{14}{k} = \binom{14}{4}$ .  
 (b) Find  $n$ , given that  $\binom{n}{10} = \binom{n}{7}$ . (d) Is there a  $k$  such that  $\binom{12}{k} = \binom{12}{k-3}$ ?
- Prove that  $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$ . This is called the **law of Pascal's triangle** and it provides a rapid way of computing binomial coefficients successively. Pascal's triangle is illustrated here for  $n \leq 6$ .

$$\begin{array}{ccccccc}
 & & & & 1 & & \\
 & & & 1 & & 1 & \\
 & & 1 & & 2 & & 1 \\
 & 1 & & 3 & & 3 & & 1 \\
 & 1 & 4 & & 6 & & 4 & 1 \\
 1 & & 5 & 10 & & 10 & 5 & 1 \\
 1 & 6 & 15 & 20 & 15 & 6 & 1
 \end{array}$$

- Use induction to prove the binomial theorem

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

Then use the theorem to derive the formulas

$$\sum_{k=0}^n \binom{n}{k} = 2^n \quad \text{and} \quad \sum_{k=0}^n (-1)^k \binom{n}{k} = 0, \quad \text{if } n > 0.$$

**The product notation.** The product of  $n$  real numbers  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  is denoted by the symbol  $\prod_{k=1}^n \mathbf{a}_k$ , which may be defined by induction. The symbol  $\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_n$  is an alternative notation for this product. Note that

$$n! = \prod_{k=1}^n k.$$

- Give a definition by induction for the product  $\prod_{k=1}^n a_k$ .



Prove the following properties of products by induction:

$$6. \prod_{k=1}^n (a_k b_k) = \left( \prod_{k=1}^n a_k \right) \left( \prod_{k=1}^n b_k \right) \quad (\text{multiplicative property}).$$

An important special case is the relation  $\prod_{k=1}^n (ca_k) = c^n \prod_{k=1}^n a_k$ .

$$7. \prod_{k=1}^n \frac{a_k}{a_{k-1}} = \frac{a_n}{a_0} \quad \text{if each } a_k \neq 0 \quad (\text{telescoping property}).$$

8. If  $x \neq 1$ , show that

$$\prod_{k=1}^n (1 + x^{2^{k-1}}) = \frac{1 - x^{2^n}}{1 - x}.$$

What is the value of the product when  $x = 1$ ?

9. If  $a_k < b_k$  for each  $k = 1, 2, \dots, n$ , it is easy to prove by induction that  $\sum_{k=1}^n a_k < \sum_{k=1}^n b_k$ . Discuss the corresponding inequality for products:

$$\prod_{k=1}^n a_k < \prod_{k=1}^n b_k.$$

### Some special inequalities

10. If  $x > 1$ , prove by induction that  $x^n > x$  for every integer  $n \geq 2$ . If  $0 < x < 1$ , prove that  $x^n < x$  for every integer  $n \geq 2$ .

11. Determine all positive integers  $n$  for which  $2^n < n!$ .

12. (a) Use the binomial theorem to prove that for  $n$  a positive integer we have

$$\left(1 + \frac{1}{n}\right)^n = 1 + \sum_{k=1}^n \left\{ \frac{1}{k!} \prod_{r=0}^{k-1} \left(1 - \frac{r}{n}\right) \right\}.$$

(b) If  $n > 1$ , use part (a) and Exercise 11 to deduce the inequalities

$$2 < \left(1 + \frac{1}{n}\right)^n < 1 + \sum_{k=1}^n \frac{1}{k!} < 3.$$

13. (a) Let  $p$  be a positive integer. Prove that

$$b^p - a^p = (b - a)(b^{p-1} + b^{p-2}a + b^{p-3}a^2 + \dots + ba^{p-2} + a^{p-1}).$$

[Hint: Use the telescoping property for sums.]

(b) Let  $p$  and  $n$  denote positive integers. Use part (a) to show that

$$n^p < \frac{(n+1)^{p+1} - n^{p+1}}{p+1} < (n+1)^p$$

(c) Use induction to prove that

$$\sum_{k=1}^{n-1} k^p < \frac{n^{p+1}}{p+1} < \sum_{k=1}^n k^p.$$

Part (b) will assist in making the inductive step from  $n$  to  $n+1$ .

14. Let  $a_1, \dots, a_n$  be  $n$  real numbers, all having the same sign and all greater than  $-1$ . Use induction to prove that

$$(1 + a_1)(1 + a_2) \cdots (1 + a_n) \geq 1 + a_1 + a_2 + \cdots + a_n.$$

In particular, when  $a_1 = a_2 = \cdots = a_n = x$ , where  $x > -1$ , this yields

$$(1.25) \quad (1 + x)^n \geq 1 + nx \quad (\text{Bernoulli's inequality}).$$

Show that when  $n > 1$  the equality sign holds in (1.25) only for  $x = 0$ .

15. If  $n \geq 2$ , prove that  $n!/n^n \leq (\frac{1}{2})^k$ , where  $k$  is the greatest integer  $\leq n/2$ .  
 16. The numbers 1, 2, 3, 5, 8, 13, 21,  $\dots$ , in which each term after the second is the sum of its two predecessors, are called *Fibonacci numbers*. They may be defined by induction as follows:

$$a_1 = 1, \quad a_2 = 2, \quad a_{n+1} = a_n + a_{n-1} \quad \text{if } n \geq 2.$$

Prove that

$$a_n < \left( \frac{1 + \sqrt{5}}{2} \right)^n$$

for every  $n \geq 1$ .

*Unequalities relating different types of averages.* Let  $x_1, x_2, \dots, x_n$  be  $n$  positive real numbers. If  $p$  is a nonzero integer, the  $p$ th-power mean  $M_p$  of the  $n$  numbers is defined as follows:

$$M_p = \left( \frac{x_1^p + \cdots + x_n^p}{n} \right)^{1/p}.$$

The number  $M_1$  is also called the *arithmetic mean*,  $M_2$  the *root mean square*, and  $M_{-1}$  the *harmonic mean*.

17. If  $p > 0$ , prove that  $M_p < M_{2p}$  when  $x_1, x_2, \dots, x_n$  are not all equal.

[Hint: Apply the Cauchy-Schwarz inequality with  $a_k = x_k^p$  and  $b_k = 1$ .]

18. Use the result of Exercise 17 to prove that

$$a^4 + b^4 + c^4 \geq \frac{64}{3}$$

if  $a^2 + b^2 + c^2 = 8$  and  $a > 0, b > 0, c > 0$ .

19. Let  $a_1, \dots, a_n$  be  $n$  positive real numbers whose product is equal to 1. Prove that  $a_1 + \cdots + a_n \geq n$  and that the equality sign holds only if every  $a_k = 1$ .

[Hint: Consider two cases: (a) All  $a_k = 1$ ; (b) not all  $a_k = 1$ . Use induction. In case (b) notice that if  $a_1 a_2 \cdots a_{n+1} = 1$ , then at least one factor, say  $a_1$ , exceeds 1 and at least one factor, say  $a_{n+1}$ , is less than 1. Let  $b_1 = a_1 a_{n+1}$  and apply the induction hypothesis to the product  $b_1 a_2 \cdots a_n$ , using the fact that  $(a_1 - 1)(a_{n+1} - 1) < 0$ .]

20. The **geometric mean**  $G$  of  $n$  positive real numbers  $x_1, \dots, x_n$  is defined by the formula  $G = (x_1 x_2 \dots x_n)^{1/n}$ .
- (a) Let  $M_p$  denote the  $p$ th power mean. Prove that  $G \leq M_1$  and that  $G = M_1$  only when  $x_1 = x_2 = \dots = x_n$ .
- (b) Let  $p$  and  $q$  be integers,  $q < 0 < p$ . From part (a) deduce that  $M_q < G < M_p$  when  $x_1, x_2, \dots, x_n$  are not all equal.
21. Use the result of Exercise 20 to prove the following statement : If  $a, b$ , and  $c$  are positive real numbers such that  $abc = 8$ , then  $a + b + c \geq 6$  and  $ab + ac + bc \geq 12$ .
22. If  $x_1, \dots, x_n$  are positive numbers and if  $y_k = 1/x_k$ , prove that

$$\left( \sum_{k=1}^n x_k \right) \left( \sum_{k=1}^n y_k \right) \geq n^2.$$

23. If  $a, b$ , and  $c$  are positive and if  $a + b + c = 1$ , prove that  $(1 - a)(1 - b)(1 - c) \geq 8abc$ .