calculus
calculus
calculus
calculus

ANALYTICAL GEOMETRY AND CALCULUS WITH VECTORS

calculus

CALCULUS

Analytic Geometry and Calculus, with Vectors

CALCULUS

Analytic Geometry and Calculus, with Vectors

Ralph Palmer Agnew
Professor of Mathematics
Department of Mathematics
Cornell University

McGraw-Hill Book Company, Inc.
New York San Francisco Toronto London

CALCULUS

Analytic Geometry and Calculus, with Vectors

Library of Congress Catalog Card Number: 61-18624

II

00595

Preface

There is an element of truth in the old saying that the Euler textbook *Introductio in Analysin Infinitorum* (Lausannae, 1748) was the first great calculus textbook, and that all elementary calculus textbooks published since that time have been copied from Euler or have been copied from books that were copied from Euler. Euler, the greatest mathematician of his day and in many respects the greatest mathematician of all time, held sway when, except where the geometry of Euclid was involved, it was not the fashion to try to base mathematical work upon accurately formulated basic concepts. Problems were the important things, and meaningful formulations of axioms, postulates, definitions, hypotheses, conclusions, and theorems either were not written or played minor roles.

Through most of the first half of the twentieth century, elementary textbooks in our subject taught unexplained but "well motivated" intuitive ideas along with their problems. Enthusiasm for this approach to calculus waned when it was realized that students were not nourished by stews in which problems, motivations, fuzzy definitions, and fuzzy theorems all boiled together while something approached something else without ever quite getting there. About the middle of the twentieth century, precise formulations of basic concepts began to occupy minor but increasingly important roles.

So far as calculus is concerned, this book attaches primary importance to basic concepts. These concepts comprise the solid foundation upon which advanced as well as elementary applications of calculus are based. Applications, including those that have great historical interest, occupy secondary roles. With this shift in our emphasis, we can remove the mystery from old mathematics and learn modern mathematics when we sometimes spend a day or two studying basic concepts and attaining mastery of ideas, language, and notation that are used. The mathematical counterparts of hydrogen and electrons are important, and we study them before trying to construct the mathematical counterparts of carbohydrates and television receivers.

This book contains just 76 sections, of which only a half dozen can be omitted without destroying the continuity of the course. In a three-term course meeting thrice weekly for fifteen weeks each term, times for reviews, tests, and occasional excursions remain when two sections are covered each week.

With few or no exceptions, each section presents each student an opportunity to make a thoroughly sound investment of time that will pay dividends in personal satisfaction, intellectual enlightenment, and scientific power. The material of the section is guaranteed to be worthy of study, it being stoutly maintained that nobody should study inept material. Each student is expected to read the text and problems of each section as carefully as an alert physicist reads an account of a newly developed nuclear reaction, and to learn as much as he can. In most cases a reasonable investment of time can produce satisfactory understanding of the text as well as solutions of several of the problems at the end of the section. Thus average students can make satisfactory progress. In some cases it is an almost superhuman task to digest all of the problems and remarks at the end of a section before additional mathematics has been studied. Thus superior students have ample opportunities to acquire large amounts of additional information and skill.

To a considerable extent, this book is a book about mathematics as well as a mathematics textbook that teaches formulas and procedures. The historical and philosophical aspects of our subject are not neglected. The text, problems, and remarks frequently give students quite unusual opportunities and incentives to think and to become genuine authorities on developments of ideas, terminologies, notations, and theories. The book strives to produce thoughtful articulate and perhaps even somewhat sophisticated students who will find that their course in calculus gives them admirable preparation for intellectual pursuits. It is frequently said that calculus textbooks contain so little of the spirit and content of modern mathematics that they do not enable students to decide whether they have the interests and the aptitudes required for life-long careers in pure mathematics or in another science in which mathematics plays a major role. Hopefully, this book does.

The first third of the book contains all or nearly all of the information about analytic geometry, vectors, and calculus that students normally need in their introductory full-year college and university courses in physics. One distinguishing feature of the book is the early introduction and continued use of vectors in three-dimensional space. These vectors simplify, clarify, and modernize our mathematics and, at the same time, make our course more interesting to teachers and vastly more interesting and immediately useful to students. Modern meaningful definitions and terminologies of the calculus are used, but we retain and explain the standard notations so students can be prepared to live in the parts of the world outside their own calculus classrooms.

The logical structure of the book should be explained. We make no effort to tell what points, lines, and planes are; we suppose that they exist, and use the axioms of the geometry of Euclid. Similarly we make no effort to tell what real numbers are; we suppose that the things exist and use the axioms that govern operations involving them. The book is based upon these axioms. If a theorem fails to have enough hypotheses to imply the conclusion, it is a blunder. If an assertion or definition is meaningless, it is a blunder. If an argument purported to be a proof or a derivation has a flaw, it is a blunder. If we pretend to prove a formula for something that has not been defined, this is a blunder. Being "rigorous" means, in mathematics, being free from blunders. "Giving all the rigor that a student can appreciate" means avoiding all the blunders that the student can detect. The author will not say that this book is rigorous because minor errors are inevitable and major blunders are possible, but he will say that he has tried to be rigorous. Thus students can be and should be invited to be critical. Detection of a blunder should be a major accomplishment.

The author will be delighted if students discover everything that is bad in the book and everything that is good in mathematics. Teachers need not and perhaps should not give as much attention to the theoretical aspects of the subject as the text does. Because the text contains many of those comments and explanations that teachers are normally called upon to supply as answers to questions, teachers are enabled to devote more of their attention to problems. Problems and applications are important and, particularly when tests and examinations consist almost exclusively of problems, major emphasis must be placed upon the problems. We will be unfair to our students if we behave like the president of a construction company who trains an employee to be an architect and then discharges him because he fails to lay bricks properly.

Ralph Palmer Agnew

Table of Contents

CALCULUS

Analytic Geometry and Calculus, with Vectors

1 Analytic geometry in two dimensions

1.1 Real numbers Without undertaking an exhaustive exposition of the subject, this preliminary section presents fundamental ideas about arithmetic, geometry, and algebra that are used throughout the book. While much of the material will be familiar, students are expected to read everything (including the problems) to assimilate information and points of view required for comprehension of later sections of the book. The rules of our game must be unmistakably clear at all times. We read everything, and we work some of the problems.

The numbers with which we are most familiar are the *positive real numbers*. These are the numbers, such as $\frac{1}{2}$, 1, $\sqrt{2}$, 2, π, 416, etcetera, that represent weights of material objects, distances between towns, etcetera. The *negative real numbers* are the negatives of these, examples being $-\frac{1}{2}$, -1, and $-\sqrt{2}$. These positive and negative real numbers, together with 0, which is neither positive nor negative, constitute the *set* *of real numbers* or the *real-number system.* Except where explicit statements to the contrary are made, the word *number* in this book always means *real number.* It is assumed that we are all familiar with the idea that numbers can be represented or approximated in decimal form. The equality $\frac{1}{2} = 0.5$ and the approximation

$$\pi = 3.14159\ 26535\ 89793 \tag{1.11}$$

must not frighten us. Searching questions about the possibility of "representing" π and other numbers by "infinite decimals" can be postponed. Our decimal system was devised by Hindus and was carried to Europe by Arabs in the twelfth century and earlier, but it took a few centuries to convince Europeans that they should and could teach the system to all of their children.

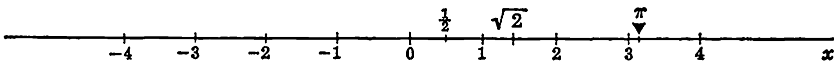


Figure 1.12

With each number x we associate a point on a line as in Figure 1.12. The line is called *the real line* or *the x axis*, and the point associated with 0 (zero) is called the *origin O* (oh). If x is positive, say 2, the point associated with x lies x, say 2, units to the right of the origin. If x is negative, say -3, the point lies $-x$, say 3, units to the left of the origin. This correspondence between numbers and points is one to one; that is, to each number there corresponds exactly one point and to each point there corresponds exactly one number. The number is called the *coordinate* of the point. While points and numbers are entities of different kinds, we sometimes find convenience in abbreviating our language by using "the point x" to mean "the point having coordinate x." The part of the x axis upon which positive numbers are plotted, or located, is called the positive x axis.

The statement $a = b$ is read "a equals b" or "a is equal to b." Similarly, the statement "$a \neq b$" is read "a is not equal to b" or "a is different from b." Thus the statements $2 = 2$ and $2 \neq 3$ are true. The statements $2 \neq 2$ and $2 = 3$ are false.

When two numbers a and b are so related that the point corresponding to a lies to the left of the point corresponding to b as in Figure 1.13, we say that a is less than b and write $a < b$. In this case we say also that b is greater than a and write $b > a$. For example, $2 < 6$, $-3 < 1$, and $-4 < -2$. This terminology agrees with common usage when temperatures are being compared; we say that a temperature $-3°$ is less, or lower, than a temperature $1°$. The inequality $-2 < 0$ means that -2 is less than 0 and that -2 is negative. The inequality $4 > 0$ means that 4 is greater than 0 and that 4 is positive. The statement that the weight w, measured in pounds, of a

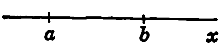


Figure 1.13

horse is greater than 1990 and less than 2010 becomes

$$1990 < w < 2010.$$

This can be read "1990 is less than w is less than 2010." In this book, Figure 1.12 precedes Figure 1.13 because $1.12 < 1.13$, and both precede Section 1.2 because $1.13 < 1.2$; the decimal system governs the numbering of all items except those appearing in the lists of problems at the ends of the sections. The basic idea that the number 1.131 or 1.135 can be assigned to an item which appears between items numbered 1.13 and 1.14 is often used to bring order out of chaos and has hordes of valuable applications. While facts of life are being considered and he still has his full complement of readers, the author can extend his best wishes to the canonical 20 per cent who will not complete this course. Those who abandon their studies to work in the design department of a sport shirt factory will be rewarded for commencement of their studies if they have learned that items in their stocks can be identified by numbers in one sequence and that numbers such as 416.35 and 416.351 can be assigned to items that should be listed between 416.3 and 416.4. Numbers assigned to items in books and factories are akin to numbers assigned to buildings beside streets and to doors inside skyscrapers. In the best of circumstances, these numbers are assigned in an informative way, and they are noticed and used when occasions arise. Persons who study Section 5.4 will be well aware that they are just getting started when they reach 5.41, that they are about halfway through the text of the section when they reach 5.45, and that they have reached the problems at the end of the section when they reach 5.49.

The inequality $a \leqq b$ is read "a is less than or equal to b." The inequalities $4 \leqq 5$ and $5 \leqq 5$ are both true, but the inequality $6 \leqq 5$ is false.

The *absolute value* of a number x is denoted by $|x|$. It is equal to x itself if $x > 0$; it is equal to 0 if $x = 0$; and it is equal to $-x$ if $x < 0$. Thus $|7| = 7$, $|0| = 0$, and $|-4| = 4$. For each x we have $|x| \geqq 0$. Moreover, $|x| = 0$ if and only if $x = 0$. For each x we have either $|x| = x$ or $|x| = -x$, and since x^2 and $(-x)^2$ are equal it follows that $|x|^2 = x^2$.

With the aid of Figure 1.12, we acquire the idea that the distance between the point with coordinate 1 and the point with coordinate 4 is 3. The distance between the points with coordinates -3 and 2 is seen to be 5, that is, $2 - (-3)$. By considering the different typical cases, we reach the conclusion that *the distance*† *between the two points with coordi-*

† It is far from easy to formulate and use enough axioms involving the geometry of Euclid and the set of real numbers to *prove* that the number $|b - a|$ is the distance between the points having coordinates a and b. To place our ideas upon a rigorous base, we can do what is usually done in more advanced mathematics: construct the foundations of ordinary geometry and analysis in such a way that the number $|b - a|$ is *defined* to be the distance between the two points having coordinates a and b.

nates a and b is $|b - a|$, that is, $b - a$ when $b \geqq a$ and $a - b$ when $b \leqq a$. The fundamental fact that the distance from a to b is less than or equal to the distance from a to 0 plus the distance from 0 to b is expressed by the inequality

$$|a - b| \leqq |a| + |b|. \tag{1.14}$$

Replacing b by $-b$ in this inequality gives the inequality

$$|a + b| \leqq |a| + |b|. \tag{1.15}$$

Problem 41 at the end of this section shows how this can be proved.

We learn in the arithmetic and algebra of real numbers that $x^2 = 0$ when $x = 0$ and that $x^2 > 0$ when $x \neq 0$. If N is a positive number, then there are two values of x for which $x^2 = N$; the positive one of these numbers is denoted by $\sqrt{N}$, and the negative one is denoted by $-\sqrt{N}$. Thus $4^2 = 16$, $(-4)^2 = 16$, $\sqrt{16} = 4$, and $-\sqrt{16} = -4$. Since $(-4)^2 = 16$ and $\sqrt{16} = 4$, we see that

$$\sqrt{(-4)^2} = \sqrt{16} = 4 = |-4|. \tag{1.16}$$

This is a special case of the formula $\sqrt{x^2} = |x|$, which holds for each real number x. In particular, $\sqrt{0} = 0$.

There are times when special properties of the number zero must be taken into account. The facts that $0 + a = a$ and $0 \cdot a = 0$ for each number a seem to be thoroughly understood by all arithmeticians, but the role of zero in division may require comment here. It is a fundamental fact that we write $x = b/a$ to represent the number x that satisfies the equation $ax = b$, provided there is one and only one number x that satisfies the equation. If $a \neq 0$ and $b = 0$, then 0 is the one and only number x that satisfies the equation and therefore

$$\frac{0}{a} = 0 \qquad (a \neq 0).$$

Thus $0/a = 0$ provided $a \neq 0$. If $a = b = 0$, then each number satisfies the equation and therefore

$$\frac{0}{0} \text{ is meaningless.}$$

If $a = 0$ and $b \neq 0$, then no number x satisfies the equation and therefore

$$\frac{b}{0} \text{ is meaningless}$$

when $b \neq 0$. Thus we see that b/a is meaningless when $a = 0$, whether b is 0 or not; division by zero is taboo. To look at the matter another way, we observe that if $a \neq 0$, then the equation $ax = ay$ implies that

$x = y$; but the equation $0{\cdot}2 = 0{\cdot}3$ does *not* imply that $2 = 3$. We must always be suspicious of results obtained by division unless we know that the divisor is not 0.

In order to pass literacy tests and to converse with our fellow men, it is necessary to know that the numbers

$$(1.17) \qquad \cdots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \cdots$$

are called *integers*. The numbers $1, 2, 3, \cdots$ are the *positive* integers. If m and n are integers and $n \neq 0$, the solution of the equation $nx = m$ is written in the form m/n and is called a *rational* (ratio-nal) number. Each integer m is a rational number because it is the solution of the equation $1x = m$ and is $m/1$. There is no greatest integer, because to each integer n there corresponds the greater integer $n + 1$. Likewise, there is no least integer, but 1 is the least positive integer. If ϵ (epsilon) and a are positive numbers, then there is a positive integer n for which $n\epsilon > a$; this is the *Archimedes property* of numbers. Another basic fact which is easier to comprehend than to prove is that if x is a number, then there is an integer n for which $n \leqq x < n + 1$.

As we near the end of this introductory section, we call attention to some additional terminology which is more important than beautiful and to which we shall slowly become accustomed as we proceed. When $a < b$, the set of points having coordinates x for which $a \leqq x \leqq b$ is called the *closed interval* of points (or numbers) from a to b. The points a and b are end points of the closed interval, and they belong to (or are points of) the closed interval. The set of points (or numbers) for which $a < x < b$ is called the *open interval* from a to b. The points a and b are still called end points of the interval, but they do not belong to the open interval. In each case, the number $b - a$ is called the *length* of the interval. Thus the length of an interval is the distance between its end points. When $a < b$, the relations

$$b - \frac{a+b}{2} = \frac{b-a}{2} > 0, \qquad \frac{a+b}{2} - a = \frac{b-a}{2} > 0$$

imply that $a < (a + b)/2 < b$ and that the point having the coordinate $(a + b)/2$ lies between and is equidistant from the points having coordinates a and b. This point is called the *mid-point* of the interval (open or closed) having its end points at a and b. If $b < a$, the above inequalities are reversed, but $(a + b)/2$ is still midway between a and b.

The following problems promote understanding of statements made by use of inequality and absolute-value signs. For example, the inequality $148 < x < 152$ says that the number x (which might be the weight of a man) is greater than 148 and less than 152. This means that x differs from 150 by a number with absolute value less than 2 and hence that $|x - 150| < 2$. It is just as easy to see that if $|x - 150| < 2$, then

$148 < x < 152$. This is a special instance illustrating the fact that if a and δ (delta) are numbers for which $\delta > 0$, then the set of numbers x for which $a - \delta < x < a + \delta$ is the same (see Figure 1.18) as the set of

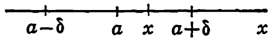


Figure 1.18

numbers x for which $|x - a| < \delta$. It is often convenient to allow δ, the Greek d, to make us think of a distance. When $\delta > 0$, the set of points (or numbers) x for which $a - \delta < x < a + \delta$ or $|x - a| < \delta$ is the set of points (or numbers) x having distance from a which is less than the distance δ. This set is an interval. A complete understanding of the nature of the assertion $|x - a| < \delta$ happens to be particularly important. There will be times when we shall use ϵ as well as δ. For example, Problem 42 of Problems 1.19 will invite attention to matters relating to the simple but important fact that if x and ϵ are numbers for which

$$|x - 2| + |x - 3| < \epsilon, \tag{1.181}$$

then ϵ cannot be 0.01.

Problems 1.19

Each of the following 40 statements is followed by a question mark, which indicates that the statement may be true or may be false. By drawing appropriate figures or otherwise, determine which of the statements are true and which are false. The answers (0 for false and 1 for true) are given at the end of the list of statements.

1 $2 < 5$?
2 $-2 > -3$?
3 $7 < 7$?
4 $7 = 7$?
5 $7 \leqq 7$?
6 $2 < -5$?
7 $-2 \leqq -1$?
8 $|-4| = 4$?
9 $|-3| < 2$?
10 $-5 = 5$?
11 $|-2| > 0$?
12 $|ab| = |a||b|$?

13 If p is the number of pages in this book, then $75 < p < 85$?
14 If C is the circumference of a circle of radius r, then $6r < C < 6.3r$? (Remember that $C = 2\pi r$, where $\pi = 3.14159+$.)
15 If $x = 5.4$, then $29 < x^2 < 30$?
16 If $x = 420$, then $20 < \sqrt{x} < 21$?
17 If $x = 4$, then $2 < x < 9$?
18 If $5 \leqq x \leqq 7$, then $4 \leqq x \leqq 9$?
19 If $5 \leqq x \leqq 7$, then $5 < x < 7$?
20 If $5 < x < 7$, then $5 \leqq x \leqq 7$?
21 If $5 \leqq x < 7$, then $5 < x < 7$?
22 If $a < b$, then $|a| < |b|$?
23 If $-2 < x < 2$, then $0 < x^2 < 4$?
24 If $x^2 < 4$, then $|x| \leqq 2$?

25 If $|x - 5| < 2$, then $3 < x < 7$?
26 If $|x - 2| < 0.01$, then $|x^2 - 4| < 0.0401$?
27 If $|x - 3| < 1$, then $|x^2 - 9| < 5$?
28 If $1 < x < 2$ and $1 < y < 2$, then $|x - y| < 1$?
29 $|(3.05)(3.06) - 9| < \frac{1}{8}$?
30 There is no real number x such that $x^2 = -1$?
31 If $0 < x < 1$, then $-1 < -x < 0$?
32 If $a < x < b$, then $-b < -x < -a$?
33 If $0 < x < 1$ and $0 < y < 1$, then $0 < x + y < 1$?
34 If $0 < x < 1$, then $0 < 2x < 1$?
35 If $|x - a| \leqq \delta$, then $|x - a| < \delta$?
36 If $|x - a| < \delta$, then $|x - a| \leqq \delta$?
37 If $\delta = \frac{1}{2}$ and $0 \leqq x \leqq 1$, then $|x^2 - \frac{1}{2}| < \delta$?
38 If $\delta = \frac{1}{2}$ and $0 \leqq x \leqq 1$, then $|x^2 - \frac{1}{2}| \leqq \delta$?
39 If $\delta = \frac{1}{2}$ and $0 < x < 1$, then $|x^2 - \frac{1}{2}| < \delta$?
40 If $\delta = \frac{1}{2}$ and $0 < x < 1$, then $|x^2 - \frac{1}{2}| \leqq \delta$?

Answers, 0 for false and 1 for true:

5	10	15	20	25	30	35	40
11011	01100	11011	11101	00011	10111	11000	10111

41 We learned while studying arithmetic and algebra that the product of either two positive numbers or two negative numbers is positive, while the product of a positive number and a negative number is negative. Supposing that x and y are nonnegative, use the identity

$$(y - x)(y + x) = y^2 - x^2 \tag{1}$$

to show that $x \leqq y$ if $x^2 \leqq y^2$. Hence show that the inequality

$$|a + b| \leqq |a| + |b| \tag{2}$$

will be true if $|a + b|^2 \leqq (|a| + |b|)^2$ or

$$(a + b)^2 \leqq a^2 + 2|a||b| + b^2. \tag{3}$$

Finally, show that (3) is true and hence that (2) is true.

42 Sketch several figures which lead to the conclusion that if x and ϵ are numbers for which

$$|x - 2| + |x - 3| < \epsilon, \tag{1}$$

then $\epsilon > 1$. *Remark:* Without using figures, we can prove the result by observing that if (1) holds, then

$$1 = |3 - 2| = |(x - 2) - (x - 3)| \leqq |x - 2| + |x - 3| < \epsilon$$

and hence $\epsilon > 1$.

43 Using the ideas of the preceding problem, prove that if a, b, x, and ϵ are numbers for which

$$|x - a| + |x - b| < \epsilon,$$

then $\epsilon > |b - a|$.

44 Let h be positive and let λ (lambda) be greater than 1. Observe that Figure 1.191 shows the correct positions of six points having the coordinates

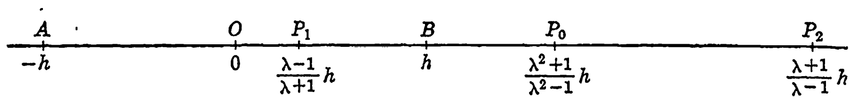


Figure 1.191

shown there when $\lambda = 2$. Make a new figure which shows where P_0, P_1, and P_2 should be when $\lambda = 10$. Make another new figure which shows where P_0, P_1, and P_2 should be when $\lambda = \frac{11}{9}$.

45 Referring to Figure 1.191 and supposing that $h > 0$ and $\lambda > 1$ as before, show that P_0 is the mid-point of the line segment with end points at P_1 and P_2.

46 When an appropriate time comes, we shall prove that there is a positive number, denoted by the symbol $\sqrt{2}$, whose square is 2. Everyone should know that $\sqrt{2}$ is not rational, and students possessing requisite time and acumen should become familiar with a proof. We prove the fact by obtaining a contradiction of the assumption that $\sqrt{2}$ is rational and hence that $\sqrt{2}$ is representable in the form $\sqrt{2} = m/n$, where m and n are positive integers. We use the fact that $28 = 2^2 \cdot 7$ and the more general fact that each positive integer n is representable in the form $n = 2^q s$, where q is a nonnegative integer and s is one of the odd integers 1, 3, 5, 7, $\cdots$. If we suppose that $\sqrt{2} = m/n$, then

$$(1) \qquad 2 = \left(\frac{m}{n}\right)^2 = \left(\frac{2^p r}{2^q s}\right)^2 = \frac{2^{2p}r^2}{2^{2q}s^2},$$

where p and q are nonnegative integers and r and s are odd integers. In case $q \geqq p$, (1) gives

$$(2) \qquad 2^{1+2q-2p}s^2 = r^2,$$

and this is false because the left side is divisible by 2, while the right side, being the square of an odd integer, is odd and is not divisible by 2. In case $p > q$, (1) gives

$$(3) \qquad s^2 = 2^{2p-2q-1}r^2,$$

and this is false because the right side is divisible by 2 while the left side is not. This proves that $\sqrt{2}$ is not rational; the assumption that $\sqrt{2}$ is rational leads to false conclusions. *Remark:* It is possible to give different proofs of this result and of the more general fact that if n is a positive integer which is not one of the perfect squares 1, 4, 9, 16, 25, 36, 49, $\cdots$, then $\sqrt{n}$ is irrational. The standard proofs depend, in one way or another, upon fundamental facts about factoring positive integers.

47 Persons who make desk calculators do their menial arithmetical chores can get very good approximations to square roots by use of an excellent method which involves some very interesting arithmetical ideas. When we want to approximate the square root of a positive number A given in decimal form, we put A in the form $A = 10^{2n}a$, where n is an integer and $1 \leqq a < 100$, and use the fact that $\sqrt{A} = 10^n\sqrt{a}$. To obtain good approximations to $\sqrt{a}$, we start with a given first approximation x_1 for which $1 \leqq x_1 \leqq 10$ and calculate some more elements of a sequence $x_1, x_2, x_3, \cdots$ of successively better approximations. If we suppose that x_n (where $n = 1$ when we start) is one of these numbers which is different from $\sqrt{a}$, we can absorb and prove the idea that $\sqrt{a}$ should lie between x_n and a/x_n. Do it. We then examine the tentative but sensible suggestion that the average of x_n and a/x_n may be a better approximation to $\sqrt{a}$. With this motivation, let

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right) \tag{1}$$

and prove that

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

and hence that

$$x_{n+1} - \sqrt{a} = \frac{1}{2x_n}(x_n - \sqrt{a})^2. \tag{3}$$

If we have not already picked up the idea that squares of small numbers are much smaller, we can start by observing that $(0.2)^2 = 0.04$, $(0.04)^2 = 0.0016$, $(0.0016)^2 = 0.00000\,256$, and $(0.00000\,3)^2 = 0.00000\,00000\,09$. This leads us to the idea that if x_n is a good approximation to $\sqrt{a}$, then x_{n+1} is much better. In fact, if one approximation x_n is correct to k decimal places, we expect the next approximation x_{n+1} to be correct to about $2k$ decimal places. Jumps from 3 to 6 to 12 to 24 are quite amazing. Calculations based upon (1) can be made very rapidly. When x_n and a/x_n agree to 10 decimal places and $\sqrt{a}$ lies between them, we have very solid information. The method has another feature that even professional computers like. Mistakes made before the final calculation do not produce an incorrect answer, because using an erroneously calculated approximation is equivalent to starting off with a different first approximation. There is even a possibility that mistakes may be helpful.

48 Supposing that $0 < a < b$, prove that

$$0 < \frac{a+b}{2} - \sqrt{ab} < \frac{(b-a)^2}{8a}.$$

Hint: Obtain and use the equality

$$\frac{\dfrac{a+b}{2} - \sqrt{ab}}{1} \cdot \frac{\dfrac{a+b}{2} + \sqrt{ab}}{\dfrac{a+b}{2} + \sqrt{ab}} = \frac{(b-a)^2}{2(a+b) + 4\sqrt{ab}}.$$

Use the fact that if a quotient has a positive numerator and a positive denominator, we obtain a greater quotient when we replace the denominator by a smaller positive number.

Remark: Persons who study science and philosophy can learn that noble but basically ineffective efforts have been made to *prove* that points, lines, planes, and numbers really exist in our physical universe, and to tell precisely what these things are. It is the opinion of the author that discussions of such matters have no place in a calculus textbook. As the preface says, we assume that these mathematical things exist (at least as "mathematical models") and we make the usual assumptions about them. We can hear many different and even contradictory tales about the world, but we can always be cheered by the fact that our assumptions are universally considered to be interesting enough and useful enough to be worthy of study. Absorbing these ideas may not keep us young and fair, but we need the ideas to be debonair.

1.2 Slopes and equations of lines When we study trigonometry, we learn about the plane rectangular coordinate system shown in Figures

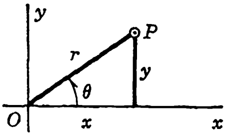


Figure 1.21

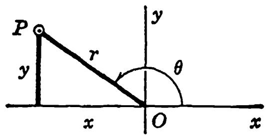


Figure 1.22

1.21 and 1.22 and we become familiar with the formulas

$$(1.23) \qquad \begin{cases} \sin\theta = \dfrac{y}{r}, & \tan\theta = \dfrac{y}{x}, & \sec\theta = \dfrac{r}{x} \\ \cos\theta = \dfrac{x}{r}, & \cot\theta = \dfrac{x}{y}, & \csc\theta = \dfrac{r}{y} \end{cases}$$

which define the six basic trigonometric functions.† In each figure, the horizontal axis is the *x axis*, the vertical axis is the *y axis*, and the intersection O of the two axes is the *origin* of the coordinate system. Since the matter will be of great importance to us, we review the standard procedure for plotting (or locating) points whose coordinates are given. To plot the point $A(3,2)$, the point A whose coordinates are the positive numbers 3 and 2, we start at the origin and go 3 units to the right (in the direction of the positive x axis) and then go 2 units up (in the direction of the positive y axis) to reach the point A of Figure 1.24. To plot the point B whose coordinates are the negative numbers -4 and -2, we start at the origin and go 4 units to the left (in the direction of the negative x axis) and then go 2 units down (in the direction of the negative y axis) to reach B. Everyone should examine Figure 1.24 to see that the other points are correctly plotted. The signs of the coordinates tell us which ways we go,

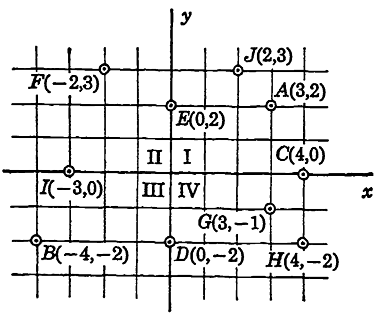


Figure 1.24

† Our rigorous presentation of angles and trigonometric functions will come in Chapter 8. Meanwhile we shall very often review and use facts about angles and trigonometric functions that are learned in trigonometry.

and the absolute values of the coordinates tell us how far we go. The quadrant (or subset) of the plane consisting of points having nonnegative coordinates is called the *closed first quadrant.* The quadrant (or subset) of the plane consisting of points having positive coordinates is called the *open first quadrant.* The Roman numerals of Figure 1.24 show us how the quadrants are numbered.

Figure 1.25 shows a line L which slopes upward to the right. The line L does not necessarily pass through the origin, but we suppose that it

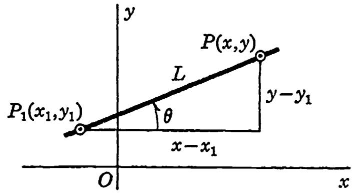


Figure 1.25

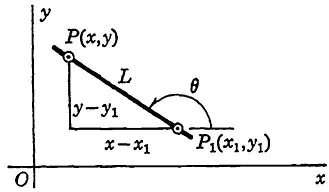


Figure 1.251

passes through a given point (x_1,y_1). The angle θ (theta) lies between 0 and $\pi/2$ (that is, between 0° and 90°), and $\tan\theta$ is called the *slope* of the line and is denoted by the letter m, so that

$$(1.252) \qquad m = \text{slope} = \tan\theta = \frac{y - y_1}{x - x_1}$$

when (x_1,y_1) and (x,y) are two different points on L. Figure 1.251 shows a line L which slopes downward to the right. This time $\tan\theta$ is negative, but it is still called the slope of the line. We must always remember that lines which slope upward to the right have positive slopes and lines which slope downward to the right have negative slopes. For horizontal lines, we have $\theta = 0$, so $\tan\theta = 0$ and $m = 0$; thus, horizontal lines have slope zero. For vertical lines, we have $\theta = \pi/2$ (or $\theta = 90°$), so $\tan\theta$ does not exist; thus, vertical lines do not have slopes. To locate a second point on a line which passes through a given point and has slope m, we start at the given point, go 1 unit to the right and then go m units in the direction of the positive y axis. When $m < 0$, a journey of m units in the direction of the positive y axis is always interpreted to be a journey of $|m|$ units in the direction of the negative y axis. Everyone should look at Figure 1.253 and think about this.

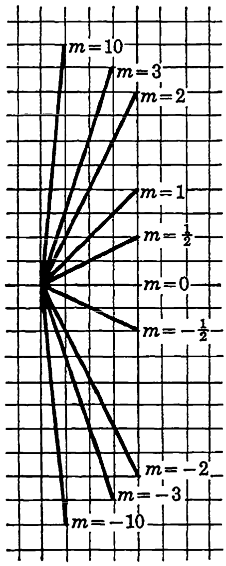


Figure 1.253

Theorem 1.26 *A point $P(x,y)$ lies on the line which contains the point $P_1(x_1,y_1)$ and has slope m if and only if its coordinates satisfy the point-*

slope equation (or formula)

$$y - y_1 = m(x - x_1). \tag{1.27}$$

Proof of this theorem is very simple. When $x = x_1$, the point $P(x,y)$ lies on the line if and only if $y = y_1$ and hence if and only if (1.27) holds. When $x \neq x_1$, the point $P(x,y)$ lies on the line if and only if (1.252) holds and hence if and only if (1.27) holds. This proves the theorem. Formula (1.27) is known as the *point-slope* formula, and it must be permanently remembered.

In accordance with general terminology which we shall introduce in Section 1.5, (1.27) is an *equation* of the line which passes through the point (x_1,y_1) and has slope m. Moreover, the line is the *graph* of the equation. When we are required to obtain an equation of the line which passes through the point $(\frac{1}{2},-\frac{1}{4})$ and has slope 3, we put $x_1 = \frac{1}{2}$, $y_1 = -\frac{1}{4}$, $m = 3$, and write immediately

$$y + \tfrac{1}{4} = 3(x - \tfrac{1}{2}).$$

Sometimes, but not always, it is desirable to put this equation in one of the forms

$$y = 3x - \tfrac{7}{4}, \qquad 12x - 4y - 7 = 0,$$

and we tolerate the custom which allows any one of the three equations to be called "the" equation of the line. Conversely, when we are required to draw or sketch the graph of the equation

$$y + \tfrac{1}{4} = 3(x - \tfrac{1}{2}),$$

we observe that the equation has the point-slope form with $x_1 = \frac{1}{2}$, $y_1 = -\frac{1}{4}$, $m = 3$ and then immediately draw the line through the point $(\frac{1}{2},-\frac{1}{4})$ having slope 3. Problems at the end of this section provide opportunities for practice in the art of using these ideas quickly, neatly, and correctly.

When, as sometimes happens, we want to find an equation of the line which passes through two given points $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ for which $x_2 \neq x_1$, we determine the slope m of the line from the formula

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{or} \quad \frac{y_1 - y_2}{x_1 - x_2} \tag{1.28}$$

and then use the point-slope formula. For example, the slope of the line passing through the points $(3,-4)$ and $(-2,1)$ is $-5/5$ or -1, and the equation of the line through these points is $y + 4 = -1(x - 3)$.

Problems 1.29

1 With Figure 1.24 out of sight, plot the points $A(3,2)$, $B(-4,-2)$, $C(4,0)$, $D(0,-2)$, and $F(-2,3)$. If correct results are not obtained, read the explanations of the text and try again.

2 Plot the points $A(-7,-1)$, $B(-5,0)$, $C(-3,1)$, $D(-1,2)$, and $E(5,5)$. (These points all lie on a line, and the figure should not contradict this fact.)

3 Plot the points $(6,2)$, $(2,6)$, $(-6,2)$, $(-6,-2)$, $(-2,-6)$, $(2,-6)$, and $(6,-2)$. (These points all lie on the circle with center at the origin and radius $\sqrt{40}$, and the figure should not contradict this fact.)

4 Three vertices of a rectangle are $(4,-1)$, $(-6,-1)$, and $(4,5)$. Sketch the rectangle and find the coordinates of the fourth vertex.

5 For each of several values of x, plot the point $P(x, 2-x)$. What can be said about the resulting set of points?

6 Plot the points $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $Q(x_1+x_2,\ y_1+y_2)$, $R\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$ and make an observation about the figure obtained by drawing the line segments from these points to each other and to the origin when

(*a*) $x_1 = 6, y_1 = 0, x_2 = 0, y_2 = 4$ (*b*) $x_1 = 2, y_1 = 5, x_2 = 6, y_2 = 3$
(*c*) $x_1 = -2, y_1 = -4, x_2 = 7, y_2 = 1$ (*d*) $x_1 = -1, y_1 = 1, x_2 = 1, y_2 = 0$

Ans.: The figure is a parallelogram together with its diagonals. *Remark:* Investing time in a good problem can produce dividends. Observe and remember that the mid-point of the line segment joining (x_1,y_1) and (x_2,y_2) is $\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$. More information about such matters will appear in the next chapter.

7 Draw the triangle having vertices at the points $P_1(-3,1)$, $P_2(7,-1)$, $P_3(1,5)$. For each $k = 1, 2, 3$, let m_k be the slope of the side opposite the vertex P_k. Work out the slopes shown in Figure 1.291 and observe that the answers look right.

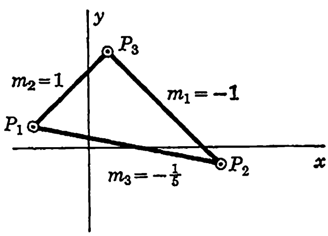


Figure 1.291

8 Show that the equation of the line P_1P_3 of the preceding problem is $y = x + 4$. Find the x coordinate of the point on this line for which $y = 0$. *Ans.:* The answer is -4, and inspection of Figure 1.291 shows that this answer looks right.

9 When numerical values are assigned to the coordinates (x_1,y_1) of a point P_1 and to m, it is possible to plot the point P_1, to sketch the line L through P_1 having slope m, and (provided $m \neq 0$) to estimate the x coordinate x_0 of the point $(x_0,0)$ on L for which $y = 0$. It is then possible to find the equation of L and determine x_0 algebraically. Do all this and make the results agree when

(*a*) $(x_1,y_1) = (1,2)$, $m = 1$ (*b*) $(x_1,y_1) = (1,2)$, $m = -1$
(*c*) $(x_1,y_1) = (-2,1)$, $m = 1$ (*d*) $(x_1,y_1) = (-2,1)$, $m = -1$
(*e*) $(x_1,y_1) = (-2,-3)$, $m = \frac{1}{2}$ (*f*) $(x_1,y_1) = (4,-2)$, $m = 2$
(*g*) $(x_1,y_1) = (-1,4)$, $m = \frac{1}{2}$ (*h*) $(x_1,y_1) = (1,1)$, $m = 1$

10 When numerical values are assigned to the coordinates (x_1,y_1) and (x_2,y_2) of two points P_1 and P_2, it is possible to plot these points, to sketch the line L through them, and (except when L is parallel to the x axis) to estimate the x coordinate of the point $(x_0,0)$ on L for which $y = 0$. It is then possible to find the slope m of L, to find the equation of L, and to determine x_0 algebraically. Do all this and make the results agree when

(*a*) $(x_1,y_1) = (1,1)$, $(x_2,y_2) = (3,2)$
(*b*) $(x_1,y_1) = (1,-1)$, $(x_2,y_2) = (3,1)$
(*c*) $(x_1,y_1) = (-4,-2)$, $(x_2,y_2) = (-1,-1)$
(*d*) $(x_1,y_1) = (0,4)$, $(x_2,y_2) = (1,2)$
(*e*) $(x_1,y_1) = (1,2)$, $(x_2,y_2) = (2,1)$
(*f*) $(x_1,y_1) = (-1,1)$, $(x_2,y_2) = (4,-1)$

11 Plot at least five points $P(x,y)$ whose coordinates satisfy the equation $y = 2x - 4$. The coordinates can be found by giving values such as -1, 0, $\frac{1}{2}$, 1 to x and calculating y. Observe that these points appear to lie on a line L. Show that the given equation can be written in the form $y - 0 = 2(x - 2)$ and hence that the points must lie on the line L through the point (2,0) which has slope 2. Make everything check.

12 Supposing that a and b are nonzero constants, find the point-slope form of the equation of the line L through the two points $(a,0)$ and $(0,b)$, and show that this equation can be put in the forms

$$bx + ay - ab = 0, \qquad \frac{x}{a} + \frac{y}{b} = 1.$$

The second form is the *intercept form* of the equation of L. Note that it is very easy to put $y = 0$ and see that L intersects (or intercepts) the x axis at the point for which $x = a$. It is equally easy to put $x = 0$ and see that L intersects (or intercepts) the y axis at the point for which $y = b$.

13 A line intersects the x axis at the point $(a,0)$ and cuts from the first quadrant a triangular region having area A. Find the equation of the line. *Ans.:* $2Ax + a^2y = 2aA$.

14 For each of the cases

(*a*) $P_1 = (1,1)$, $P_2 = (7,1)$, $P_3 = (7,7)$
(*b*) $P_1 = (2,2)$, $P_2 = (8,2)$, $P_3 = (8,8)$
(*c*) $P_1 = (-3,-1)$, $P_2 = (2,-7)$, $P_3 = (4,1)$
(*d*) $P_1 = (-4,-2)$, $P_2 = (1,-8)$, $P_3 = (3,0)$
(*e*) $P_1 = (-2,-4)$, $P_2 = (1,-5)$, $P_3 = (2,-3)$

sketch the triangle having vertices P_1, P_2, P_3 and the line L containing P_1 and the mid-point of the side opposite P_1. Use the figure to obtain an estimate of the x coordinate of the point where L intersects the x axis. Then find the equation of L and determine the coordinate algebraically. Produce results that have reasonable agreement. *Remark:* There is one respect in which many problems in pure and applied mathematics are like this one. Graphs or something else give more or less good approximations to answers, but we need equations to get correct answers. When equations give answers that seem to be wrong, the whole situation must be given close scrutiny. Mistakes in sign are particularly damaging, and we all make mistakes when we work too rapidly or too thoughtlessly.

15 A triangle with vertices A, B, C is placed upon a coordinate system in such a way that A is at the origin and the mid-point D of the opposite side is the point $(h,0)$ on the positive x axis as in Figure 1.292. Supposing that the coordinates of C are $(h + a, k)$, show that the coordinates of B must be $(h - a, -k)$. Find equations of the lines containing the sides of the triangles. Write the equations in the form $y = mx + b$ and check the answers by determining whether the coordinates of the vertices satisfy the equations.

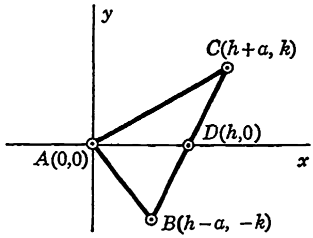


Figure 1.292

16 The vertices $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$ of a triangle are unknown, but it is known that the mid-points of the sides P_1P_2, P_2P_3, and P_3P_1 are respectively $(7,-1)$, $(4,3)$, and $(1,1)$. Find the unknowns and check the results by drawing an appropriate figure.

17 Formulate and solve a more general problem of which Problem 16 is a special case.

1.3 *Lines and linear equations; parallelism and perpendicularity*

When A, B, and C are constants† for which A and B are not both 0, the equation

$$Ax + By + C = 0 \tag{1.31}$$

is a *linear equation* and we must both prove and remember that its graph is a line. In case $B \neq 0$, we can put the equation in the point-slope form

$$y - \left(-\frac{C}{B}\right) = -\frac{A}{B}(x - 0)$$

and see that the graph is the line L which passes through the point $(0, -C/B)$ and has slope $-A/B$. In case $B = 0$, we must have $A \neq 0$, and we can put the equation in the form

$$x = -\frac{C}{A}.$$

The graph of this equation is the vertical line consisting of all those points (x,y) for which $x = -C/A$. This proves the result.

The equation

$$y = mx + b \tag{1.32}$$

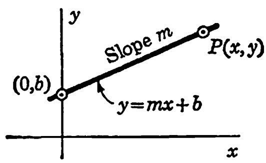


Figure 1.33

can be put in the form $y - b = m(x - 0)$, and hence it is the equation of the line L which passes through the point $(0,b)$ and has slope m. The equation (1.32) is called

† The hypothesis that A, B, and C are *constants* means merely that they are numbers that are "given" or "selected" or "fixed" in some way. There is no implication that other numbers are unstable in the sense that they are moving around. We shall hear more about this matter later.

the *slope-intercept formula.* The easiest way to find the slope m of the line having the equation $2x - 3y - 4 = 0$ is to solve the equation for y to obtain

$$y = \tfrac{2}{3}x - \tfrac{4}{3}$$

and see that $m = \frac{2}{3}$.

Let L_1 and L_2 be two lines which are neither horizontal nor vertical and let their slopes be m_1 and m_2. Figure 1.34 reminds us of the elementary fact in plane geometry that L_1 and L_2 are parallel if and only if θ_2 and θ_1 are equal and hence if and only if $\tan\theta_2 = \tan\theta_1$ and $m_2 = m_1$. *Thus L_1 and L_2 are parallel if and only if their slopes are equal.*

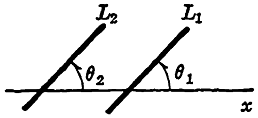


Figure 1.34

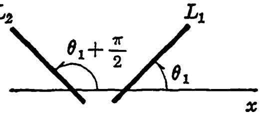


Figure 1.35

For perpendicular lines, the story is more complicated. *The lines L_1 and L_2 are perpendicular if and only if their slopes m_1 and m_2 are negative reciprocals*, that is, $m_2 = -1/m_1$ or $m_1 = -1/m_2$ or $m_1m_2 = -1$. To prove this, we observe that L_2 and L_1 are perpendicular if and only if $\theta_2 = \theta_1 + \pi/2$ as in Figure 1.35 or $\theta_1 = \theta_2 + \pi/2$ when the roles of L_1 and L_2 are reversed. In the first case we have

$$\text{(1.351)} \qquad m_2 = \tan\left(\theta_1 + \frac{\pi}{2}\right) = -\cot\theta_1 = -\frac{1}{\tan\theta_1} = -\frac{1}{m_1}$$

and the result follows. To get the result in the second case, we merely reverse the roles of L_1 and L_2.

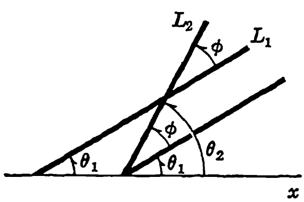


Figure 1.36

As in Figure 1.36, let L_1 and L_2 be two nonvertical lines and let ϕ (phi) be the angle through which a line must be rotated to bring it from coincidence with L_1 to coincidence (or parallelism) with L_2. When we can find the slopes m_1 and m_2 of L_1 and L_2, we can determine ϕ from the fact that $\phi = \pi/2$ when $m_1m_2 = -1$ and

$$\text{(1.37)} \qquad \tan\phi = \frac{m_2 - m_1}{1 + m_1m_2}$$

when $m_1m_2 \neq -1$. The latter formula is proved by the formula

$$\text{(1.371)} \qquad \tan\phi = \tan(\theta_2 - \theta_1) = \frac{\tan\theta_2 - \tan\theta_1}{1 + \tan\theta_1\tan\theta_2} = \frac{m_2 - m_1}{1 + m_1m_2}$$

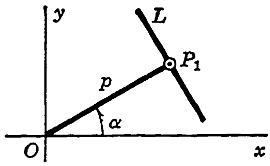


Figure 1.38

which employs the standard trigonometric formula for the tangent of the difference of two angles. When we are asked to find ϕ, we present $\tan \phi$ as the answer† to our problem.

As an application of some of the above ideas, we find the equation of the line L of Figure 1.38. The positive number p is the distance from the origin to the point P_1, and L is perpendicular to the line OP_1 at P_1. The coordinates of P_1 are $p \cos \alpha$ (alpha) and $p \sin \alpha$. The slope of OP_1 is $\tan \alpha$, or $\sin \alpha/\cos \alpha$, and the slope of L is the negative reciprocal $-\cos \alpha/\sin \alpha$. Use of the point-slope formula gives the equation of L in the form

$$(1.381) \qquad y - p \sin \alpha = -\frac{\cos \alpha}{\sin \alpha}(x - p \cos \alpha).$$

Multiplying by $\sin \alpha$ and using the identity $\sin^2 \alpha + \cos^2 \alpha = 1$ puts the equation in the more attractive form

$$(1.382) \qquad (\cos \alpha)x + (\sin \alpha)y = p.$$

The line OP_1, being a line perpendicular to L, is called a *normal* to L, and the equation (1.382) is called the *normal form* of the equation of L because it gives information about this normal. It is sometimes thought to be worthwhile to know a speedy way to put the equation $Ax + By + C = 0$ into normal form. We suppose that A and B are not both 0 and that $C \neq 0$. The trick is to transpose C to obtain $Ax + By = -C$ and then divide by one of $\pm\sqrt{A^2 + B^2}$ to obtain

$$(1.383) \qquad \frac{A}{\pm\sqrt{A^2+B^2}}x + \frac{B}{\pm\sqrt{A^2+B^2}}y = \frac{-C}{\pm\sqrt{A^2+B^2}},$$

where the sign is so chosen that the right side is positive. This equation has the normal form. The right side is the distance p from the origin to the line, and the coefficients of x and y are respectively the numbers $\cos \alpha$ and $\sin \alpha$ which determine the angle α of Figure 1.38.

Problems 1.39

1 Each of the following equations is the equation of a line L. In each case, find the slope m by finding the coordinates of the points in which the line intersects the coordinate axes and then finding the slope of the line through those

† Traditionally, students are required by tests and examinations to find $\tan \phi$. Accordingly, students who hope to pass examinations by learning a few formulas—and those who aspire to a substantial understanding of their subject—are well advised to learn the necessary ritual.

two points. Then find m by putting the equation in the form $y = mx + b$ Make the results agree.

(*a*) $x + y = 2$ (*b*) $x + y = 3$
(*c*) $2x + y = 2$ (*d*) $x + 2y = 2$
(*e*) $2x + 3y - 4 = 0$ (*f*) $2x + 3y + 4 = 0$
(*g*) $2x - 3y - 4 = 0$ (*h*) $2x - 3y + 4 = 0$
(*i*) $\frac{x}{1} + \frac{y}{2} = 1$ (*j*) $\frac{x}{1} - \frac{y}{2} = 1$

2 Draw the triangle having vertices at the points $P_1(-3,1)$, $P_2(7,-1)$, $P_3(1,5)$, and observe that P_3P_2 seems to be nearly perpendicular to P_1P_3. Find the equation of the line through P_3 perpendicular to P_1P_3 and show that this line does contain the point P_2. (The figure appears among the problems at the end of Section 1.2.)

3 For the points P_1, P_2, P_3 of the preceding problem, find the equation of the line through P_2 parallel to the line P_1P_3. Find the coordinates of the point in which this new line intersects the y axis. Put this new line into the figure, and make everything check.

4 For each of the following equations find numerical coordinates of three points P_1, P_2, P_3 whose coordinates satisfy the equation. If you cannot think of a better procedure, let $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ and calculate y_1, y_2, y_3. Plot the three points P_1, P_2, P_3 and notice that they seem to lie on a line. Calculate the slopes of P_1P_2 and P_1P_3 and observe that they are equal. Observe that there is ample opportunity to check all answers.

(*a*) $y = x + 1$ (*b*) $y = 2x + 3$
(*c*) $x + y = 5$ (*d*) $x + y + 2 = 0$
(*e*) $2x - 3y + 4 = 0$ (*f*) $2x + 3y + 4 = 0$

5 Plot the lines having the equations $y = 2x$ and $y = 3x$ and observe that the acute angle ϕ between them seems to be rather small. Find ϕ by finding $\tan \phi$, and then construct and examine an appropriate figure to see that your answer seems to be correct.

6 Supposing that k is a nonnegative number, find the acute angle between the lines having the equations $y = kx$ and $y = (k + 1)x$. Check the answer in at least one special case. Tell why the angle should be small when k is large.

7 Sketch the line L_1 which intersects the coordinate axes at the points $(0,-4)$ and $(5,0)$, and the line L_2, which intersects the coordinate axes at the points $(0,-5)$ and $(6,0)$. Find the acute angle between the lines. *Ans.*: $\tan \theta = \frac{1}{50}$.

8 While assembly lines and mass production reduce costs of manufactured items, there is an element of sanity in the idea that the total cost y of publishing x copies of a book is $ax + b$. Sketch a graph of the equation $y = ax + b$ and discover the significance of the numbers a and b.

9 Find the equation of the perpendicular bisector of the line segment joining the points $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$, putting the answer in the form $Ax + By = C$. *Ans.*:

$$(x_2 - x_1)x + (y_2 - y_1)y = \frac{x_2^2 - x_1^2}{2} + \frac{y_2^2 - y_1^2}{2}.$$

10 Any given rectangle can be placed upon the x, y coordinate system in such a way that its vertices are (0,0), $(0,a)$, $(b,0)$, and (b,a). Prove that if the diagonals are perpendicular, then the rectangle is a square.

11 Sketch a figure showing the triangle having vertices at the points $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, and $P_3(x_3,y_3)$. For each $k = 1, 2, 3$, mark the mid-point Q_k of the side opposite P_k and find the coordinates of Q_k. Supposing that the line Q_2Q_1 is not vertical, calculate its slope and show that it is parallel to the line P_1P_2.

12 Prove analytically (by calculating slopes) that the mid-points of the sides of a convex quadrilateral are vertices of a parallelogram. *Remark:* Taking vertices at (x_1,y_1), (x_2,y_2), (x_3,y_3), (x_4,y_4) produces "symmetric" formulas.

13 Show that the lines having the equations

$$a_1x + b_1y = c_1$$
$$a_2x + b_2y = c_2$$

are parallel if and only if $a_1b_2 - a_2b_1 = 0$. If the lines are not parallel, they must intersect at a point $P(x,y)$ whose coordinates satisfy both equations. Assuming that the lines are not parallel, solve the equations to obtain the formulas

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}, \qquad y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}$$

for the coordinates of the point of intersection. *Remark:* Those who have forgotten how to solve systems of linear equations can recover by noticing that we can multiply the first equation by b_2 and the second by $-b_1$ and then add the results to eliminate y and obtain an equation that can be solved for x. This process is known as the *process of successive elimination.*

14 Copy Figure 1.292 and then find the equations of the three medians of the triangle and show that these medians intersect at the point $(2h/3, 0)$. *Remark:* Since the median placed upon the x axis could have been any median of the triangle, this provides a proof that the three medians of a triangle intersect at the point which trisects each of them.

15 Show that the lines obtained by giving constant values to k in the equation

$$2x + 3y + k = 0$$

are all parallel. Show that the line L having the equation

$$2(x - x_1) + 3(y - y_1) = 0$$

belongs to this family and contains the point (x_1,y_1).

16 Show that if the lines AP and BP joining the points $A(1,2)$ and $B(5,-4)$ to $P(x,y)$ are perpendicular, then

$$(x - 1)(x - 5) + (y + 4)(y - 2) = 0.$$

Remark: Persons well acquainted with elementary geometry should know that P must lie on the circle having the line segment AB for a diameter.

17 Put the following equations into normal form and check the results by drawing graphs showing the lines having the given equations and the line segments through the origin perpendicular to these lines.

(*a*) $2x + 3y = 12$ (*b*) $2x - 3y = 12$
(*c*) $2x - 3y = -12$ (*d*) $2x + 3y = -12$
(*e*) $y = 2x + 3$ (*f*) $y = 0$

18 Let $k > 0$. For each of several values of w, draw the perpendicular bisector of the line segment joining the points $(0, 1/4k)$ and $(w, -1/4k)$. Then determine the condition which x and y must satisfy if the point $P(x,y)$ lies

(*a*) on exactly one of these bisectors,
(*b*) on more than one of these bisectors,
(*c*) on none of these bisectors.

Ans.: (*a*) $y = kx^2$, (*b*) $y < kx^2$, (*c*) $y > kx^2$.

19 For what pairs of values of b and c do the two equations

$$\begin{aligned} 3x + by + c &= 0 \\ cx - 2y + 12 &= 0 \end{aligned}$$

have the same graph? *Partial ans.:* There are two pairs which are easily checked after they have been found.

20 The points $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$ are vertices of a triangle. Find the coordinates (x,y) of the point Q_1 where the line through P_1 perpendicular to the line P_2P_3 intersects the line P_2P_3. *Partial ans.:*

$$x = \frac{(x_3 - x_2)^2x_1 + (y_3 - y_1)(y_3 - y_2)x_2 + (y_2 - y_1)(y_2 - y_3)\,x_3}{(x_3 - x_2)^2 + (y_3 - y_2)^2}.$$

Remark: It is possible to write the answer in different forms. This form enables us to check quickly that interchanging the subscripts 2 and 3 does not change the value of x. Such checks are often used to guard against clerical errors in deriving or copying formulas.

21 Let the vertices of a triangle be $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$. For each k, let L_k be the line containing P_k which is perpendicular to the line containing the other two vertices. Recognizing that altitudes are numbers (not line segments or lines), we call the lines L_1, L_2, L_3 the altitudinal lines of the triangle. Prove that these altitudinal lines are concurrent. *Remark:* The conclusion means that there is a point P_0, called the *orthocenter* of the triangle, at which the three altitudinal lines intersect. When asked to prove the conclusion by synthetic methods, we use our ingenuity (or that of some other people) in searches for appropriate figures and ideas upon which the proof can be based. When asked to prove the conclusion by analytic methods, we can proceed at once to apply a powerful method that cannot fail to produce results if we do the chores correctly We can find the equations of the three altitudinal lines and use two of the equations to find the coordinates of the point of intersection of two of the lines. If this point lies on the third line, the conclusion is true. If (for some triangle) the point fails to lie on the third line, the conclusion is false. The chores can be done in the following way. Considering separately the case in which the line P_2P_3 is neither horizontal nor vertical (so that this line and L_1 have slopes) and the cases in which P_2P_3 is horizontal or vertical, we can find that the equation of L_1 is the first of the equations

$$(1)\qquad (x_3 - x_2)x + (y_3 - y_2)y = (x_3 - x_2)x_1 + (y_3 - y_2)y_1$$
$$(2)\qquad (x_1 - x_3)x + (y_1 - y_3)y = (x_1 - x_3)x_2 + (y_1 - y_3)y_2$$
$$(3)\qquad (x_2 - x_1)x + (y_2 - y_1)y = (x_2 - x_1)x_3 + (y_2 - y_1)y_3.$$

It is possible to repeat the process to show that the equations of L_2 and L_3 are (2) and (3). It is, however, more fun to observe that we can convert the derivation of the equation of L_1 into a derivation of the equation of L_2 by making a "cyclic advance" of the subscripts so that 1 goes to (or is replaced by) 2, 2 goes to 3, and 3 goes to 1. The first cyclic advance converts (1) into (2), and another cyclic advance converts (2) into (3). The routine way to finish the proof is to solve (1) and (2) for x and y and show that these numbers (x,y) satisfy (3) However, if we do not want to obtain and preserve the formulas for x and y, we can finish the problem by observing that adding the members of (1), (2), and (3) gives $0 = 0$ and shows that the third equation is satisfied whenever the first two are satisfied. For the record, we note that solving (1) and (2) for x and y gives the formulas

$$(4)\qquad x = \frac{y_1(x_3 - x_2)x_1 + y_2(x_1 - x_3)x_2 + y_3(x_2 - x_1)x_3 - (y_2 - y_1)(y_3 - y_2)(y_1 - y_3)}{y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1)}$$

$$(5)\qquad y = \frac{x_1(y_3 - y_2)y_1 + x_2(y_1 - y_3)y_2 + x_3(y_2 - y_1)y_3 - (x_2 - x_1)(x_3 - x_2)(x_1 - x_3)}{x_1(y_3 - y_2) + x_2(y_1 - y_3) + x_3(y_2 - y_1)}$$

for the coordinates of the orthocenter. With the aid of the identity

$$(6)\qquad (y_2 - y_1)(y_3 - y_2)(y_1 - y_3) = y_1(y_3^2 - y_2^2) + y_2(y_1^2 - y_3^2) + y_3(y_2^2 - y_1^2),$$

we can put these formulas in the forms

$$(7)\qquad x = \frac{y_1[x_1(x_3 - x_2) + y_2^2 - y_3^2] + y_2[x_2(x_1 - x_3) + y_3^2 - y_1^2] + y_3[x_3(x_2 - x_1) + y_1^2 - y_2^2]}{y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1)}$$

$$(8)\qquad y = \frac{x_1[y_1(y_3 - y_2) + x_2^2 - x_3^2] + x_2[y_2(y_1 - y_3) + x_3^2 - x_1^2] + x_3[y_3(y_2 - y_1) + x_1^2 - x_2^2]}{x_1(y_3 - y_2) + x_2(y_1 - y_3) + x_3(y_2 - y_1)}$$

and in many other forms which look quite different. Interchanging the x's and y's in one of these formulas gives the other. Except for sign, the denominators are equal to each other and (as we shall see later) to twice the area of the triangle. As is to be expected, a cyclic advance of the subscripts does not alter the triangle and does not alter the formulas for the coordinates of the orthocenter.

22 Again let the vertices of a triangle be $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$. Find the coordinates of the point (x,y) of intersection of the three perpendicular bisectors of the sides of the triangle, and write also the result of making a cyclic advance (see the preceding problem) of the subscripts appearing in the answer. *Remark:* Elementary plane geometry shows that the point (x,y) is equidistant from the three vertices and hence is the center of the circle containing the vertices. The circle is called the *circumcircle* (or circumscribed circle) of the triangle, and its center is called the *circumcenter* of the triangle. The answer can be put in the forms

$$(1)\quad x = \frac{y_1(x_3 - x_2)(x_3 + x_2) + y_2(x_1 - x_3)(x_1 + x_3) + y_3(x_2 - x_1)(x_2 + x_1) + (y_2 - y_1)(y_3 - y_2)(y_1 - y_3)}{2[y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1)]}$$

$$(2)\quad y = \frac{x_1(y_3 - y_2)(y_3 + y_2) + x_2(y_1 - y_3)(y_1 + y_3) + x_3(y_2 - y_1)(y_2 + y_1) + (x_2 - x_1)(x_3 - x_2)(x_1 - x_3)}{2[x_1(y_3 - y_2) + x_2(y_1 - y_3) + x_3(y_2 - y_1)]}$$

and

$$(3)\quad x = \frac{y_1(x_3^2 + y_3^2 - x_2^2 - y_2^2) + y_2(x_1^2 + y_1^2 - x_3^2 - y_3^2) + y_3(x_2^2 + y_2^2 - x_1^2 - y_1^2)}{2[y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1)]}$$

$$(4)\quad y = \frac{x_1(x_3^2 + y_3^2 - x_2^2 - y_2^2) + x_2(x_1^2 + y_1^2 - x_3^2 - y_3^2) + x_3(x_2^2 + y_2^2 - x_1^2 - y_1^2)}{2[x_1(y_3 - y_2) + x_2(y_1 - y_3) + x_3(y_2 - y_1)]}$$

and in still other forms which look quite different.

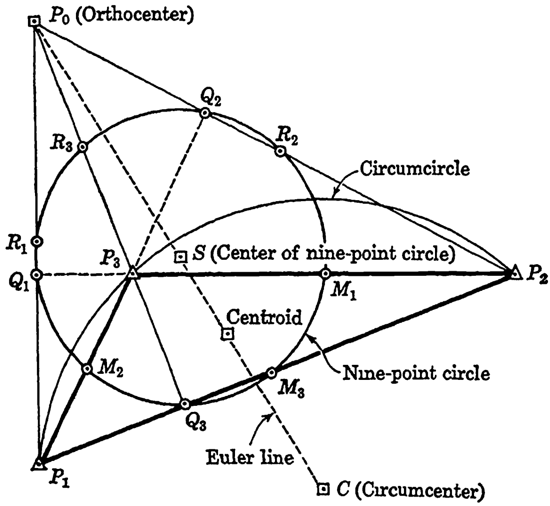


Figure 1.391

23 The triangle in Figure 1.391 has vertices at $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, and $P_3(x_3,y_3)$. The mid-points M_1, M_2, M_3 of the sides of this triangle are the vertices of the *mid-triangle* of the given triangle. With or without making use of the ideas and results of the preceding problem, find the coordinates of the circumcenter of this mid-triangle. *Remark:* The answer can be put in the form

$$(1)\quad x = \frac{y_1(x_3 - x_2)(2x_1 + x_2 + x_3) + y_2(x_1 - x_3)(2x_2 + x_3 + x_1) + y_3(x_2 - x_1)(2x_3 + x_1 + x_2) - (y_2 - y_1)(y_3 - y_2)(y_1 - y_3)}{4[y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1)]}$$

$$(2)\quad y = \frac{x_1(y_3 - y_2)(2y_1 + y_2 + y_3) + x_2(y_1 - y_3)(2y_2 + y_3 + y_1) + x_3(y_2 - y_1)(2y_3 + y_1 + y_2) - (x_2 - x_1)(x_3 - x_2)(x_1 - x_3)}{4[x_1(y_3 - y_2) + x_2(y_1 - y_3) + x_3(y_2 - y_1)]}$$

and in other forms which look quite different. As we shall see in the next problem, the circumcircle of the mid-triangle of the given triangle $P_1P_2P_3$ is the famous *nine-point circle* of the given triangle. The coordinates in (1) and (2) are therefore the coordinates of the center of this nine-point circle. The answers to this and the two preceding problems are written in such a way that it is very easy to see that the center of the nine-point circle is the mid-point of the line segment joining the orthocenter and the circumcenter of the given triangle.

24 To see that there are opportunities to use ideas of the preceding problems and the rest of this chapter in geometry, we look briefly at a triangle and its nine-point circle. Figure 1.391 shows a triangle $P_1P_2P_3$, the points Q_1, Q_2, Q_3 in which the altitudinal lines intersect the lines containing the sides of the triangle, and the orthocenter P_0. The points M_1, M_2, M_3 are the mid-points of the sides of the triangle, and the perpendiculars to the sides at these points intersect at a point C, the circumcenter of the given triangle. The points R_1, R_2, R_3 are the mid-points of the line segments P_1P_0, P_2P_0, P_3P_0. The famous *nine-point-circle theorem* says that *the nine points* M_1, M_2, M_3, Q_1, Q_2, Q_3, R_1, R_2, R_3 *all lie on a circle*. This circle, the nine-point circle, has its center at the mid-point S of the line segment joining the orthocenter P_0 and the circumcenter C. The radius of the nine-point circle is half the radius of the circumcircle. When the triangle is equilateral, the orthocenter, the circumcenter, the center of the nine-point circle, and the centroid (intersection of the medians) all coincide. When the triangle is not equilateral, the four points are distinct but are collinear, and the line upon which they lie is called the *Euler line* of the triangle.

25 In this problem we use results of Problems 21 and 23 to obtain a new formula and a proof of the nine-point-circle theorem. We know from Problem 23 that the mid-points M_1, M_2, M_3 of the sides of the triangle are on the circle; in fact, these three noncollinear points determine the nine-point circle. The remaining points R_1, R_2, R_3, Q_1, Q_2, Q_3, which we must prove to be on the nine-point circle, are not necessarily distinct from each other and from M_1, M_2, M_3, but our proof will not be a "partial proof" which covers only "general cases." Our proof will be a proof. Use a result of Problem 21 to show that the x coordinate of the point R_1 midway between the vertex P_1 and the orthocenter P_0 is

$$x = \frac{\begin{aligned} y_1(x_3 - x_2)(x_1 + x_1) + y_2(x_1 - x_3)(x_1 + x_2) + y_3(x_2 - x_1)(x_1 + x_3) \\ + (y_2 - y_1)(y_3 - y_2)(y_1 - y_3) \end{aligned}}{2[y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1)]}.$$

Use this result to show that the x coordinate of the point midway between R_1 and the mid-point M_1 of the line segment P_2P_3 is the x coordinate of the center of the nine-point circle given in Problem 23. *Remark:* This fact and the associated fact involving y coordinates imply that the points R_1 and M_1 are at opposite ends of a diameter (line segment, not number) of the nine-point circle. Similar proofs (which are attained by cyclic advances of subscripts) show that R_2 and M_2 are at opposite ends of a diameter and that R_3 and M_3 are at opposite ends of a diameter. This proves that R_1, R_2, R_3, lie on the circle. We recall that Q_1 is the point at which the altitudinal line through P_1 intersects the line containing the vertices P_2 and P_3, that R_1 is on the altitudinal line, and that M_1 is on the line containing P_2 and P_3. In case Q_1 coincides with M_1 or R_1, we conclude that Q_1 is on the circle. In the contrary case, the angle $R_1Q_1M_1$ is a right angle.

Since the line segment R_1M_1 is a diameter of the circle, this implies that Q_1 must be on the circle. Cyclic advances of subscripts prove that Q_2 and Q_3 lie on the circle. This completes the proof of the nine-point-circle theorem.

26 This problem involves the intersection of the medians of the triangle having vertices $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$. For each $k = 1, 2, 3$, let M_k be the mid-point of the side opposite P_k. Show that the equation of the line containing the median P_1M_1 can be put in the form

$$(y_2 + y_3 - 2y_1)(x - x_1) - (x_2 + x_3 - 2x_1)(y - y_1) = 0. \tag{1}$$

Show that if we define $\bar{x}$ and $\bar{y}$ by the formulas

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3}, \qquad \bar{y} = \frac{y_1 + y_2 + y_3}{3}, \tag{2}$$

then

$$\begin{aligned} x_2 + x_3 - 2x_1 &= x_1 + x_2 + x_3 - 3x_1 = 3(\bar{x} - x_1) \\ y_2 + y_3 - 2y_1 &= y_1 + y_2 + y_3 - 3y_1 = 3(\bar{y} - y_1) \end{aligned}$$

and (1) can be put in the form

$$(\bar{y} - y_1)(x - x_1) - (\bar{x} - x_1)(y - y_1) = 0. \tag{3}$$

Show that (3) implies that the point $(\bar{x},\bar{y})$ lies on the median P_1M_1. Finally, show how this work can be modified to prove that the point $(\bar{x},\bar{y})$ lies on the other two medians and hence is the point of intersection of the medians. *Remark:* One reason for interest in this matter can be understood when we know enough about centroids. The point $(\bar{x},\bar{y})$, the intersection of the medians, is the centroid of the triangular region bounded by the triangle. It is also the centroid of the set consisting of the three vertices of the triangle. Moreover, it is the centroid of the triangle itself, that is, the set consisting of the sides of the triangle. The coordinates of the intersection of the medians were obtained in a tricky way. It is possible to put the equation (1) of the median P_1M_1 and the equation of the median P_2M_2 in the forms

$$\begin{aligned} &(y_2 + y_3 - 2y_1)x - (x_2 + x_3 - 2x_1)y \\ &\qquad = (y_2 + y_3 - 2y_1)x_1 - (x_2 + x_3 - 2x_1)y_1 \end{aligned} \tag{4}$$

$$\begin{aligned} &(y_3 + y_1 - 2y_2)x - (x_3 + x_1 - 2x_2)y \\ &\qquad = (y_3 + y_1 - 2y_2)x_2 - (x_3 + x_1 - 2x_2)y_2 \end{aligned} \tag{5}$$

and obtain the coordinates of the intersection of the medians by solving these equations for x and y without using trickery. There is, however, no guarantee that time invested in a study of (4) and (5) will produce attractive dividends. It is easy to obtain ponderous formulas for x and y, but it is not so easy to reduce the formulas to the right members of the formulas (2).

1.4 Distances, circles, and parabolas

As we shall see, the *distance formula*

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1.41}$$

gives the distance d between two points $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ in the plane. To prove (1.41), we notice first that if $y_2 = y_1$, then the points P_1

and P_2 lie on the same horizontal line and the formula reduces to the correct formula $d = |x_2 - x_1|$. If $x_2 = x_1$, then P_1 and P_2 lie on the same vertical line and the formula reduces to the correct formula $d = |y_2 - y_1|$. When $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ are two given points for which $x_2 \neq x_1$ and $y_2 \neq y_1$, we can, as in Figures 1.42 and 1.421, let $Q(x_2,y_1)$ be the point on

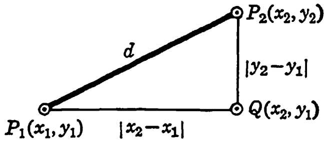


Figure 1.42

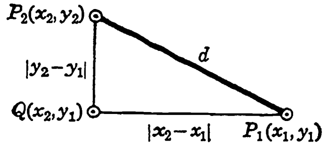


Figure 1.421

the horizontal line through P_1 and on the vertical line through P_2. The length of the horizontal line segment P_1Q is $x_2 - x_1$ if $x_2 - x_1 \geqq 0$, is $x_1 - x_2$ if $x_1 - x_2 \geqq 0$, and is $|x_2 - x_1|$ in each case. The length of the vertical line segment QP_2 is $y_2 - y_1$ if $y_2 - y_1 \geqq 0$, is $y_1 - y_2$ if $y_1 - y_2 \geqq 0$, and is $|y_2 - y_1|$ in each case. With the understanding that the distance d between P_1 and P_2 is the length of the line segment P_1P_2, we can therefore apply the Pythagoras theorem to the right triangle P_1QP_2 to obtain

$$d^2 = |x_2 - x_1|^2 + |y_2 - y_1|^2$$

and hence

$$d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2.$$

Since $d \geqq 0$, taking square roots gives the required formula (1.41).

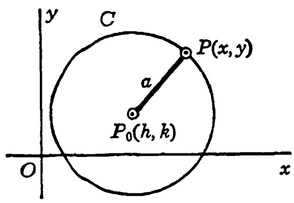


Figure 1.43

We are all familiar with the fact, illustrated in Figure 1.43, that the circle C with center at $P_0(h,k)$ and radius a is the set of points in the plane whose distances from P_0 are equal to the radius a. From the distance formula, we see that the point $P(x,y)$ lies on this circle if and only if

$$(x - h)^2 + (y - k)^2 = a^2. \tag{1.44}$$

This is therefore the equation of the circle with center at (h,k) and radius a. We must always remember this and the fact that

$$x^2 + y^2 = a^2 \tag{1.45}$$

is the equation of the circle with center at the origin and radius a.

The equation of the circle with center at $(-2,3)$ and radius 5 is

$$(x + 2)^2 + (y - 3)^2 = 25. \tag{1.451}$$

When the parentheses are removed and the constant terms are collected, this equation takes the less informative form

$$(1.452) \qquad x^2 + y^2 + 4x - 6y - 12 = 0.$$

This has the form

$$(1.453) \qquad x^2 + y^2 + Dx + Ey + F = 0,$$

where D, E, and F are constants. It turns out that for some sets of values of D, E, and F, (1.453) is the equation of a circle. To try to write (1.453) in the standard form (1.44), we begin by writing it in the form

$$(1.454) \qquad (x^2 + Dx + \quad) + (y^2 + Ey + \quad) = -F.$$

The next step is to add a constant to the term $x^2 + Dx$ so that the sum will be the square of a quantity of the form $(x + Q)$. What shall we add? A good look at the formula

$$(x + Q)^2 = x^2 + 2Qx + Q^2$$

provides the answer: divide the coefficient of x by 2 and square the result. Thus we add $D^2/4$ and $E^2/4$ to both sides of (1.454) to obtain

$$\left(x^2 + Dx + \frac{D^2}{4}\right) + \left(y^2 + Ey + \frac{E^2}{4}\right) = \frac{D^2 + E^2 - 4F}{4}$$

or

$$(1.46) \qquad \left(x + \frac{D}{2}\right)^2 + \left(y + \frac{E}{2}\right)^2 = \frac{D^2 + E^2 - 4F}{4}.$$

We can now see how the graph depends upon the constants D, E, and F. In case $D^2 + E^2 - 4F > 0$, then (1.46) is the equation of the circle with center at $(-\frac{1}{2}D, -\frac{1}{2}E)$ and radius $\frac{1}{2}\sqrt{D^2 + E^2 - 4F}$. In case $D^2 + E^2 - 4F = 0$, the equation becomes

$$(1.461) \qquad (x + \tfrac{1}{2}D)^2 + (y + \tfrac{1}{2}E)^2 = 0.$$

This equation is satisfied when and only when $x = -\frac{1}{2}D$ and $y = -\frac{1}{2}E$ so the graph is the single point $(-\frac{1}{2}D, -\frac{1}{2}E)$. In case $D^2 + E^2 - 4F < 0$, there are no pairs of values of x and y for which the equation is satisfied. One is tempted to say that the poor equation has no graph, but the graph is actually the empty set, that is, the set having no points in it. Thus, determination of the graph of the equation

$$x^2 + y^2 + 6x - 7y + 8 = 0$$

is made by *completing squares*. The process is important and must be remembered.

Before starting the next paragraph, we look at some algebra and ways in which it is printed. The quotient a/bc is called a *shilling quotient* and is often printed instead of the *built-up quotient* $\frac{a}{bc}$. Learning to read printed mathematics involving shilling quotients is an art that must be cultivated, and this is a good opportunity. Since "multiplication takes precedence over division" the quotient a/bc means $a/(bc)$ and does not mean $(a/b)c$. Thus, for example, $1/2k$ means $1/(2k)$ or $\frac{1}{2k}$ and does not mean $(1/2)k$ or $\frac{1}{2}k$. When the next paragraph is read, the quotients should be handwritten in built-up forms so the calculations can be made more easily. If troubles appear, the difficulty may be the canonical one that arises when a printer converts an author's $1/2k$ into $\frac{1}{2}k$. Everything should be checked.

We can get experience with the distance formula by starting to learn about parabolas. A parabola is, as we shall show in Section 6.2, the set of points (in a plane) equidistant from a fixed point F which is called the *focus* and a fixed line L which is called the *directrix* and which does not pass through the focus.† In order to obtain the equation of a parabola in an attractive form, we let $1/2k$ denote the distance from F to L so that $1/2k = p$ and $k = 1/2p$, where p is the distance (length of the "perpendicular") from F to L. Then we put the y axis through F perpendicular to L and put the x axis midway between F and L as in Figure 1.47.

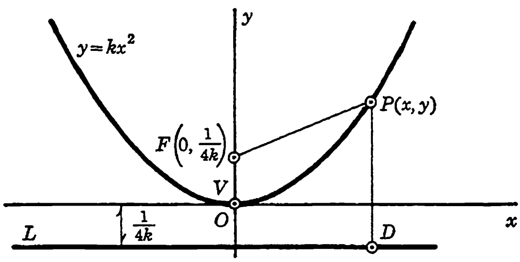


Figure 1.47

The parabola is the set of points $P(x,y)$ for which $FP = DP$. Using the distance formula and the fact that $y + 1/4k > 0$ when $FP = DP$ gives

$$\text{(1.471)} \qquad \sqrt{x^2 + \left(y - \frac{1}{4k}\right)^2} = y + \frac{1}{4k}.$$

† The assumption that F is a "fixed" point and L is a "fixed" line means merely that F and L are "given" or "selected" in some way. There is no implication that other points and lines are "unfixed" in the sense that they are moving. At one time the parabola was defined as the path (or locus) of a point P which moves in such a way that it is always equidistant from F and L. There are reasons why it is better to say that a parabola is a point set. Everybody knows that pencil points and numerous other things move, but even if we swallow the dubious idea that "mathematical points" can move we still find that the old-fashioned definition does not tell how a point P should move to trace the whole parabola and not merely a part of the parabola.

This holds if and only if

$$x^2 + \left(y - \frac{1}{4k}\right)^2 = \left(y + \frac{1}{4k}\right)^2.$$

Simplifying this gives the very simple and attractive equation

$$y = kx^2. \tag{1.472}$$

This is the equation of the parabola shown in Figure 1.47.

The point V on a parabola which lies midway between the focus and directrix of the parabola is called the *vertex* of the parabola. For example, when $k \neq 0$, the point (x_0,y_0) is the vertex of the parabola for which the point $F(x_0, y_0 + 1/4k)$ is the focus and the line L having the equation $y = y_0 - 1/4k$ is the directrix. As Problem 26 invites us to discover, the equation of this parabola is

$$y - y_0 = k(x - x_0)^2. \tag{1.473}$$

When $k \neq 0$, the equation

$$y = kx^2 + ax + b \tag{1.474}$$

can be put in the form (1.473) by completing a square and transposing. Thus, when $k \neq 0$, the graph of (1.474) is a parabola, and we must always remember the fact. The distance from the focus to the vertex is $\left|\frac{1}{4k}\right|$. When the positive y axis lies above the origin as it usually does, the focus is above the vertex when $k > 0$ and is below the vertex when $k < 0$.

It is possible to proceed in various ways to calculate the distance d from a given point $P(x_0,y_0)$ to the line L having the given equation $Ax + By + C = 0$. Problem 34 at the end of this section requires that the answer be worked out in a specified straightforward way. It is sometimes convenient to omit the calculations and use the result, which is set forth in the following theorem.

Theorem 1.48 *The distance d from the point $P(x_0,y_0)$ to the line L having the equation $Ax + By + C = 0$ is given by the formula*

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}. \tag{1.481}$$

In some applications of this theorem, we use the version obtained by deleting the subscripts.

Problems 1.49

1 Draw the triangle having vertices at the points $A(2,2)$, $B(-5,-2)$, and $C(-2,-4)$. Calculate the lengths a, b, and c of the three sides BC, CA, and AB and show that $c^2 = a^2 + b^2$. This implies that the triangle is a right triangle having a right angle at C, and hence that the lines BC and CA must be perpendicular. Calculate the slopes of these lines and verify the perpendicularity. Make everything check.

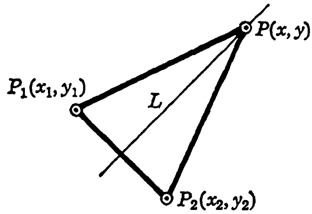


Figure 1.491

2 Figure 1.491 illustrates the familiar fact that, when $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ are two distinct points in a plane, the set of points $P(x,y)$ equidistant from P_1 and P_2 is the perpendicular bisector L of the line segment P_1P_2. Equate expressions for the distance PP_1 and PP_2 and simplify the result to obtain the equation of L in the form

$$(x_2 - x_1)\left(x - \frac{x_1 + x_2}{2}\right) + (y_2 - y_1)\left(y - \frac{y_1 + y_2}{2}\right) = 0.$$

Then show that this line passes through the mid-point

$$P\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right)$$

of the segment P_1P_2 and is perpendicular to the segment.

3 Sketch a figure showing the triangle having vertices at the three given points and then calculate distances to determine whether the triangle is isosceles (that is, has two sides of equal length):

(*a*) (1,0), (8,2), (3,−7) (*b*) (1,4), (6,−1), (7,6)
(*c*) (0,*a*), (*a*,0), (*b*,*b*) (*d*) (*a*,*a*), (−*a*,−*a*), (*b*,−*b*)

4 Find the length of the part of the x axis which lies inside the triangle having vertices at the points (−3,−1), (5,1), and (1,5). Use a figure to determine whether the answer is reasonable.

5 Find the point on the x axis equidistant from the two points $P_1(-2,-1)$ and $P_2(4,3)$ in two different ways. First, find the equation of the perpendicular bisector of the line segment P_1P_2 and find the point where this bisector intersects the x axis. Then, with the aid of the distance formula, determine x so that the distance from $(x,0)$ to P_1 is equal to the distance from $(x,0)$ to P_2.

6 Find the center and radius of the circle having the equation

$$(x - 1)(x - 5) + (y + 4)(y - 2) = 0.$$

Show that the center is the mid-point of the line segment joining the points $A(1,2)$ and $B(5,-4)$.

7 Find the center and radius of the circle having the equation

$$(x - x_1)(x - x_2) + (y - y_1)(y - y_2) = 0.$$

8 Show that the equation of the circle C with center at $P_0(x_0,y_0)$ and radius a can be put in the form

$$(x - x_0)(x - x_0) + (y - y_0)(y - y_0) = a^2$$

and that the equation of the tangent to C at a point $P_1(x_1,y_1)$ on C can be put in the form

$$(x_1 - x_0)(x - x_0) + (y_1 - y_0)(y - y_0) = a^2.$$

Hint: Draw a figure and notice that when $x_1 \neq x_0$, we can calculate the slope of the line P_0P_1 and use the fact (from plane geometry) that the tangent to C at P_1 is perpendicular to P_0P_1.

9 Show that the circle passing through the three points $A(0,2)$, $B(2,0)$, and $C(4,0)$ has its center at the point $(3,3)$ and has radius $\sqrt{10}$. *Hint:* The perpendicular bisectors of the segments AB and BC are easily found, and their intersection is the required center.

10 A circle passes through the points $(0,7)$ and $(0,9)$ and is tangent to the x axis at a point on the negative x axis. Find the radius, center, and equation of the circle.

11 Let $0 < a < b$ and find the radius r and center (h,k) of the circle which passes through the points $(0,a)$ and $(0,b)$ and which is tangent to the x axis at a point to the left of the origin. *Ans.:*

$$r = \frac{a+b}{2}, \qquad h = -\sqrt{ab}, \qquad k = \frac{a+b}{2}.$$

12 A circle has a diameter (line segment, not number) on the x axis. The circle contains the two points $(a,0)$ and (b,c) for which $c \neq 0$. Show that $b \neq a$ and find the center of the circle. *Ans.:*

$$\left(\frac{b^2 + c^2 - a^2}{2(b - a)}, 0\right).$$

13 The points $A(-a,0)$ and $B(a,0)$ are the ends of a diameter (line segment, not number) of a circle of radius a having its center at the origin. Write and simplify the equation which x and y must satisfy if A, B, and $P(x,y)$ are vertices of a right triangle the side AB of which is the hypotenuse.

14 An equilateral triangle has its center at the origin and has one vertex at the point $(a,0)$. Find the coordinates of the other vertices and check the results by use of the distance formula.

15 Sketch a figure which shows whether there are values of y for which the point $(0,y)$ is equidistant from the points $(-4,1)$ and $(7,-2)$. Then attack the problem analytically. Make everything check.

16 An equilateral triangle in the closed first quadrant has vertices at the origin and at $(a,0)$. Find the coordinates of the third vertex and the slopes of the sides.

17 An isosceles triangle is placed upon a coordinate system in such a way that its vertices are $(-a,0)$, $(a,0)$, and $(0,b)$. Prove analytically that two of the medians have equal lengths.

18 A triangle has vertices at $A(-a,0)$, $B(b,0)$, $C(0,c)$. Prove that if the medians drawn from A and B have equal lengths, then the triangle is isosceles.

19 Find the values of the constant b for which the line having the equation $y = 2x + b$ intersects the circle having the equation $x^2 + y^2 = 25$. *Ans.:* $|b| \leqq \sqrt{125}$.

20 A triangle has vertices $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$. Prove analytically that 4 times the sum of the squares of the lengths of the medians is equal to 3 times the sum of the squares of the lengths of the sides.

21 Discover for yourself that a part of a parabola can be drawn with the aid of the right triangle (or rectangle), string, and pin mechanism shown in Figure 1.492. A string of length ED has one end fastened to the triangle at E and has the other end fastened to a pin at the focus F. A pencil point at P keeps the string taut, so $FP = DP$, and traces a part of the parabola as the base of the triangle is moved along the directrix. Such constructions are taboo in the classical ruler-and-compass geometry of Euclid, but in analytic geometry we can recognize the existence of all kinds of machinery.

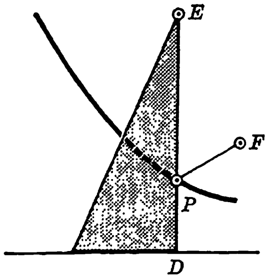


Figure 1.492

22 Supposing that $p > 0$, find and simplify the equation of the parabola whose focus is at the origin and whose directrix is the line having the equation $y = -p$. *Ans.:* $y = \frac{1}{2p}(x^2 - p^2)$. *Remark:* If we set $k = 1/2p$, then the equation takes the form $y = k(x^2 - 1/4k^2)$. The parabolas obtained by taking different values of p or k constitute a family of confocal parabolas; concentric circles have the same center and confocal parabolas have the same focus.

23 Supposing that $p < 0$, find and simplify the equation of the parabola whose focus is at the origin and whose directrix is the line having the equation $y = -p$.

24 Supposing that $p > 0$, find and simplify the equation of the parabola whose focus is at the origin and whose directrix is the line having the equation $x = -p$. *Ans.:* $x = \frac{1}{2p}(y^2 - p^2)$.

25 Find the equation of the parabola whose focus is the point (12,0) and whose directrix is the line having the equation $x = -12$. *Ans.:* $x = y^2/48$.

26 Supposing that $k \neq 0$, use the distance formula to obtain the equation satisfied by the coordinates (x,y) of points P equidistant from the point $F(x_0, y_0 + 1/4k)$ and the line L having the equation $y = y_0 - 1/4k$. *Outline of solution:* A point $P(x,y)$ lies on the parabola if and only if $FP = DP$, where D is the point $(x, y_0 - 1/4k)$. Writing FP and DP in terms of coordinates gives an equation which reduces to (1.473).

27 Supposing that $h > 0$ and $\lambda > 1$, find and simplify the equation satisfied by the coordinates of the points $P(x,y)$ whose distances from the point $A(-h,0)$ are λ times their distances from the point $B(h,0)$. *Ans.:*

$$\left(x - \frac{\lambda^2 + 1}{\lambda^2 - 1}h\right)^2 + y^2 = \left(\frac{2\lambda}{\lambda^2 - 1}h\right)^2.$$

28 Still supposing that $h > 0$ and $\lambda > 1$, show that the graph of the answer to Problem 27 is a circle having its center at a point P_0 on the x axis. Find the x coordinates of the points P_1 and P_2 where the circle intersects the x axis. *Ans.:* See Figure 1.191, which displays the x coordinates of the points and shows their correct positions relative to A, 0, B and to each other.

29 From the first two of the equations

$$(1) \qquad x^2 + y^2 + a_1x + b_1y + c_1 = 0$$
$$(2) \qquad x^2 + y^2 + a_2x + b_2y + c_2 = 0$$
$$(3) \qquad (a_2 - a_1)x + (b_2 - b_1)y + (c_2 - c_1) = 0$$

we can obtain the third by equating the left members of (1) and (2) and simplifying the result. Supposing that the graphs of (1) and (2) are nonconcentric circles, show that the graph of (3) is a line perpendicular to the line containing the centers of the circles. Show also that if these circles intersect in one or two points, then the line contains the point or points of intersection. *Remark:* The line is called the *radical axis* of the circles, it being named because some people want to talk about it.

30 The points $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, and $P_3(x_3,y_3)$ are, in positive or counterclockwise order, the vertices of an equilateral triangle. Find formulas which express x_3 and y_3 in terms of the coordinates of P_1 and P_2. *Solution:* While the problem can be attacked in other ways, we eliminate difficulties involving order relations by observing that if the half-line extending from P_1 through P_2 makes the angle θ with the positive x axis, then the half-line extending from P_1 through P_3 makes the angle $\theta + \pi/3$ with the positive x axis. Let a be the lengths of the sides of the equilateral triangle. The definitions of the trigonometric functions then give

$$(1) \qquad x_2 - x_1 = a\cos\theta, \qquad y_2 - y_1 = a\sin\theta$$

$$(2) \qquad x_3 - x_1 = a\cos\left(\theta + \frac{\pi}{3}\right) = \frac{1}{2}a\cos\theta - \frac{\sqrt{3}}{2}a\sin\theta$$

$$(3) \qquad y_3 - y_1 = a\sin\left(\theta + \frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}a\cos\theta + \frac{1}{2}a\sin\theta.$$

In obtaining the latter formulas, we use the "addition formulas"

$$(4) \qquad \cos(\theta + \phi) = \cos\theta\cos\phi - \sin\theta\sin\phi$$
$$(5) \qquad \sin(\theta + \phi) = \sin\theta\cos\phi + \cos\theta\sin\phi$$

and the values of the sine and cosine of $\pi/3$ which can be determined with the aid of Figures 1.493. From (2), (3), and (1), we obtain the answers

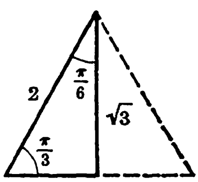


Figure 1.493

$$(6) \qquad x_3 = x_1 + \frac{1}{2}(x_2 - x_1) - \frac{\sqrt{3}}{2}(y_2 - y_1)$$

$$(7) \qquad y_3 = y_1 + \frac{\sqrt{3}}{2}(x_2 - x_1) + \frac{1}{2}(y_2 - y_1).$$

Remark: The points in the xy plane for which both coordinates are integers are called *lattice points*. Our results enable us to show very easily that triangles having vertices at lattice points cannot be equilateral. To prove this, let an equilateral triangle have two of its vertices, say P_1 and P_2, at lattice points. Then x_1, x_2, y_1, y_2 are integers and, since $\sqrt{3}$ is irrational, (6) shows that x_3 cannot be an integer unless $y_2 = y_1$. If $y_2 = y_1$, we must have $x_2 \neq x_1$, and then

(7) shows that y_3 cannot be an integer. This shows that if two vertices of an equilateral triangle are lattice points, then the third vertex cannot be a lattice point.

31 Let n be a positive integer. Let $m_1, m_2, \cdots, m_n, x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_n$ be numbers for which

$$(1) \qquad m_1 + m_2 + \cdots + m_n = M > 0.$$

Let

$$(2) \quad \bar{x} = \frac{m_1x_1 + m_2x_2 + \cdots + m_nx_n}{M}, \qquad \bar{y} = \frac{m_1y_1 + m_2y_2 + \cdots + m_ny_n}{M}.$$

For each $k = 1, 2, \cdots, n$, let r_k be the distance from (x,y) to (x_k,y_k) and let d_k be the distance from $(\bar{x},\bar{y})$ to (x_k,y_k). A timid person may be comforted by the special case in which $n = 4$, $m_1 = m_2 = m_3 = m_4 = 1$, and the points (x_k,y_k) are the vertices $(1,1)$, $(-1,1)$, $(-1,-1)$, $(1,-1)$ of a square. Confining attention to the special case if this be deemed desirable, prove that

$$(3) \quad m_1r_1^2 + m_2r_2^2 + \cdots + m_nr_n^2 = M[(x - \bar{x})^2 + (y - \bar{y})^2] + m_1d_1^2 + m_2d_2^2 + \cdots + m_nd_n^2.$$

With the aid of this result, let I be a constant and describe the set of points (x,y) for which

$$(4) \qquad m_1r_1^2 + m_2r_2^2 + \cdots + m_nr_n^2 = I.$$

32 Let P_1, P_2, P_3 have coordinates (x_1,y_1), (x_2y_2), (x_3,y_3), respectively. The *triangle inequality*

$$(1) \quad \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} \leqq \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} + \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$$

says that the distance from P_1 to P_3 is less than or equal to the sum of the distance from P_1 to P_2 and the distance from P_2 to P_3. In more advanced mathematics, analytic proofs of (1) and more or less similar inequalities are very important. Show that setting

$$(2) \quad a_1 = x_2 - x_1, \qquad a_2 = y_2 - y_1, \qquad b_1 = x_3 - x_2, \qquad b_2 = y_3 - y_2$$

puts (1) in the more agreeable form

$$(3) \qquad \sqrt{(a_1 + b_1)^2 + (a_2 + b_2)^2} \leqq \sqrt{a_1^2 + a_2^2} + \sqrt{b_1^2 + b_2^2}.$$

By squaring and simplifying, show that (3) holds if

$$(4) \qquad |a_1b_1 + a_2b_2| \leqq \sqrt{a_1^2 + a_2^2}\,\sqrt{b_1^2 + b_2^2}.$$

By squaring and simplifying again, show that (4) holds if

$$(5) \qquad 0 \leqq (a_1b_2 - a_2b_1)^2.$$

Finally, tell why (5) must hold. *Remark:* In order to appreciate the significance of this work, we must do a little thinking about "elementary" mathematics. It is sometimes said that a straight line is the shortest distance between two points. If this silly collection of words means anything it means that the length (a number) of the line segment (a point set) joining two points P_1 and P_2 is less than the length (a number) of each other path (a point set) joining P_1 and P_2. We must study more mathematics before we can learn what we mean by a path joining P_1 and P_2 and what we mean by the length of such a path. In some parts of "advanced" mathematics, the multifarious axioms of Euclid and the theorem of Pythagoras are bypassed and the number d in the formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

is *defined* to be the distance (in Euclid space of two dimensions) between the two points $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$. It is useful as well as possible to define d by other formulas to obtain spaces that are not Euclid spaces. In such situations it is necessary to use analytical methods instead of geometrical methods to determine whether triangle inequalities hold.

33 Four numbers a_{11}, a_{12}, a_{21}, and a_{22} determine the equations

$$\text{(1)} \qquad \begin{cases} x' = a_{11}x + a_{12}y \\ y' = a_{21}x + a_{22}y \end{cases}$$

into which we can substitute the coordinates of a given point (x,y) to obtain the coordinates of a *transform*, or transformed point, (x',y'). Supposing that (x_1,y_1) and (x_2,y_2) are two given points and that D is the distance between their transforms (x_1', y_1') and (x_2', y_2'), find D^2. *Ans.:*

$$\text{(2)} \quad D^2 = (a_{11}^2 + a_{21}^2)(x_2 - x_1)^2 + (a_{12}^2 + a_{22}^2)(y_2 - y_1)^2 + 2(a_{11}a_{12} + a_{21}a_{22})(x_2 - x_1)(y_2 - y_1).$$

Remark: The transformer is called *isometric* if the distance d between two points is always the same as the distance D between their transforms. If the transformer is isometric, we can put $x_2 - x_1 = 1$ and $y_2 - y_1 = 0$ to obtain

$$\text{(3)} \qquad a_{11}^2 + a_{21}^2 = 1,$$

we can put $x_2 - x_1 = 0$ and $y_2 - y_1 = 1$ to obtain

$$\text{(4)} \qquad a_{12}^2 + a_{22}^2 = 1,$$

and we can put $x_2 - x_1 = 1$ and $y_2 - y_1 = 1$ and use (3) and (4) to obtain

$$\text{(5)} \qquad a_{11}a_{12} + a_{21}a_{22} = 0.$$

On the other hand, if (3), (4), and (5) hold, then (2) shows that the transformer is isometric.

34 Supposing that the first of the two equations

$$\text{(1)} \qquad Ax + By = -C, \qquad Bx - Ay = Bx_0 - Ay_0$$

is the equation of a given line L and that $P_0(x_0,y_0)$ is a given point, find the equation of the line through P_0 perpendicular to L and show that it is equivalent to the second of the two equations. Solve these equations to find that the coordinates x_1, y_1 of the foot P_1 of the perpendicular from P_0 to L are

$$(2) \qquad x_1 = \frac{B^2x_0 - ABy_0 - AC}{A^2 + B^2}, \qquad y_1 = \frac{A^2y_0 - ABx_0 - BC}{A^2 + B^2}.$$

Show that

$$(3) \quad x_1 - x_0 = \frac{-A}{A^2 + B^2}(Ax_0 + By_0 + C),$$

$$y_1 - y_0 = \frac{-B}{A^2 + B^2}(Ax_0 + By_0 + C).$$

Finally, use the fact that the distance d from P_0 to L is the distance from P_0 to P_1 to obtain the formula

$$(4) \qquad d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}.$$

1.5 *Equations, statements, and graphs* The equation $y = x + 2$ can be regarded as a statement that is true for some pairs of values of x and y, for example, $x = 3$, $y = 5$, and is false for some other pairs of values of x and y, for example, $x = 7$, $y = 7$. A similar remark applies to each of the equations $x^2 + y^2 = 4$, $0x + 0y = 1$, and $0x + 0y = 0$, and to each of the inequalities $0 < x < 1$, $y < x$, and $x^2 + y^2 < 1$. Each is a statement that is true for some (or none or all) pairs of values of x and y and is false for the remaining ones. The *graph* of such a statement is the set or collection of points $P(x,y)$ whose coordinates are pairs of values of x and y for which the statement is true. For example, the graph of the statement (or equation) $y = x$ is a line L. We can always know that there is a substantial difference between an equation (or statement) and its graph (a point set). Hence, we may be carrying abbreviation of language a bit too far when we sometimes follow the old and misleading custom of referring to "the line $y = x$" instead of to "the line having the equation $y = x$." In any case, we should think about this matter enough to know that we are introducing analytic geometry and hopefully trying to make sense out of nonsense if we receive a mysterious order to "find the part of $y = x$ in $x^2 + y^2 = 1$" and proceed to find the length of the part of the line having the equation $y = x$ which lies inside the circle having the equation $x^2 + y^2 = 1$.†

Most of the graphs that appear in our work are graphs of equations. However, graphs of inequalities can be important, and we look at some simple examples. The graph of the inequality $xy > 0$ consists of those points $P(x,y)$ in the first quadrant (where x and y are both positive)

† Persons who start picking up clear ideas about these things may even enjoy studying statements and sets in mathematical logic and elsewhere.

together with those in the third quadrant (where x and y are both negative); see Figure 1.51. The graph of the inequality $y \leqq x$ consists of those points $P(x,y)$ which lie on and below the line $y = x$ of Figure 1.52.

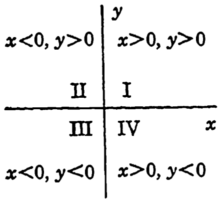


Figure 1.51

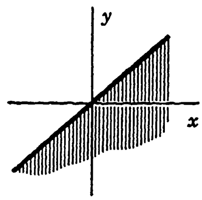


Figure 1.52

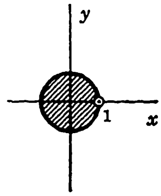


Figure 1.53

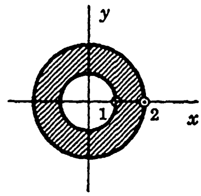


Figure 1.54

The graph of the inequality $x^2 + y^2 < 1$ consists of the points inside the circle with center at the origin and unit radius. This set of points is often called the *unit disk;* see Figure 1.53. The graph of the inequality

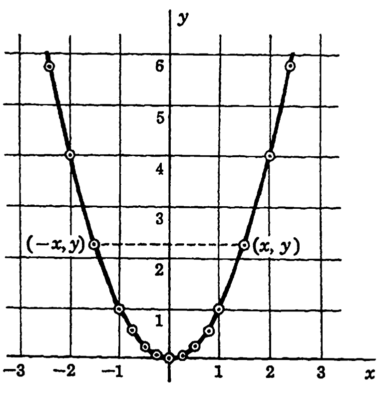


Figure 1.55

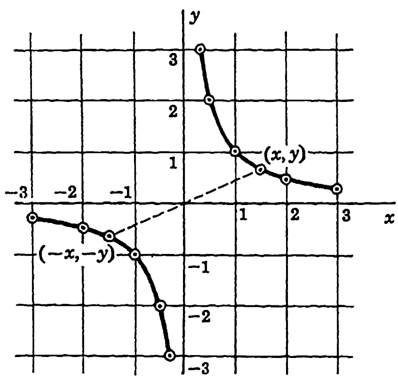


Figure 1.56

$1 < x^2 + y^2 < 4$ is the set of points in the annulus or ring between two circles; see Figure 1.54.

The equation $y = x^2$ is, as we saw in Section 1.4, the equation of a parabola. After plotting the points whose coordinates appear in the table

x	0	$\pm\frac{1}{4}$	$\pm\frac{1}{2}$	$\pm\frac{3}{4}$	± 1	$\pm\frac{3}{2}$	± 2	$\pm\frac{5}{2}$	± 3
y	0	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{9}{16}$	1	$\frac{9}{4}$	4	$\frac{25}{4}$	9

we are easily led to the correct conclusion that the graph of $y = x^2$ is the curve shown in Figure 1.55. It should be noted that the graph contains no point (x,y) for which $y < 0$; if $y < 0$, there is no x for which $y = x^2$. The y axis is an *axis of symmetry* of the graph, because if (x,y) is a point on the graph, then the point $(-x,y)$ is also on the graph.

The graph of the equivalent equations

$$(1.561) \qquad xy = 1, \qquad y = \frac{1}{x}$$

is more complex. As we shall see later, the graph is a rectangular hyperbola. It is easy to add more items to the table

x	$\frac{1}{10}$	$\frac{1}{2}$	1	2	10
y	10	2	1	$\frac{1}{2}$	$\frac{1}{10}$

and to sketch the part of the graph to the right of the y axis in Figure 1.56. A similar table in which x and y are both negative enables us to sketch the part lying to the left of the y axis. The graph contains no point (x,y) for which $x = 0$ or $y = 0$. The x and y axes are not axes of symmetry, but the origin is a *center of symmetry*, because if (x,y) is a point on the graph, then the point $(-x,-y)$ is also on the graph.

The symbol $[x]$ represents, when we are properly warned, the greatest integer in x, that is, the greatest integer n for which $n \leqq x$. Thus $[1.99] = 1$, $[3.14] = 3$, $[0.25] = 0$, $[-0.25] = -1$, $[-3.01] = -4$, and $[2] = 2$. It is not difficult to show that the graphs of $y = [x]$ and of the saw-tooth function $y = x - [x] - \frac{1}{2}$ have the forms shown in Figures 1.57 and 1.571.

Trigonometric functions will appear very often in our work, and there will be very many times when we must know the natures of the graphs of

Figure 1.57

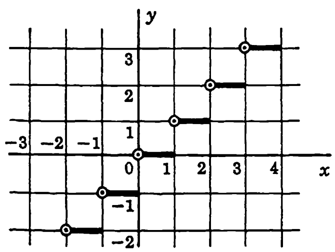
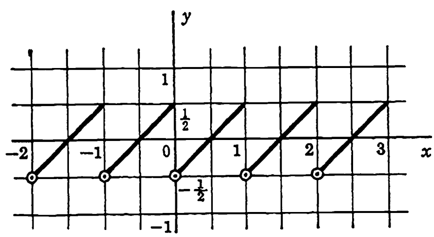


Figure 1.571



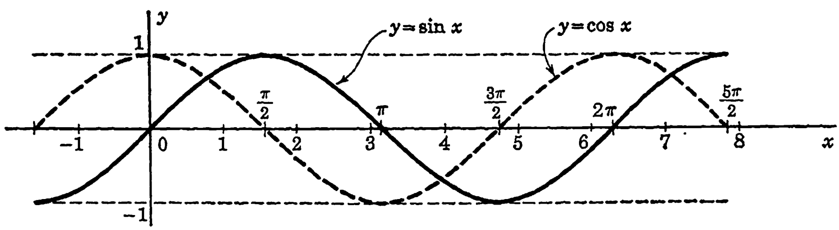


Figure 1.58

$y = \sin x$ and $y = \cos x$. The graphs are shown in Figure 1.58. We must always know that, except at the points of tangency, the graphs lie between the lines having the equations $y = -1$ and $y = 1$. Moreover, π is a little bit greater than 3, and this must be fully recognized when the graphs are sketched. When we want to sketch the graphs, the first step is to draw guide lines one unit above and one unit below the x axis. The next step is to hop three units and a bit more to the right of the origin to mark π, and make another such hop to mark 2π. We must be able to do this and sketch reasonably accurate graphs of $y = \sin x$ and $y = \cos x$ in a few seconds, and we must be able to look at the graphs and see answers to trigonometric questions just as we look at dogs and see answers to questions about canine structure. We cannot tolerate doubts about the assertions $\sin 0 = 0$, $\cos 0 = 1$, $\sin \pi/2 = 1$, $\cos \pi/2 = 0$, and dogs have two ears. The table on the back cover of this book can be used to produce very accurate graphs, but this is seldom necessary.

Finally, we are never too young to be informed that substantial parts of scientific lives are devoted to learning about and using equations akin to $y = e^x$ and $y = \log x$. Graphs of these equations are shown in Figures 1.581 and 1.582. The exponentials and logarithms have base e, and e is a number that we shall encounter very often. Here again the tables on the back cover of this book can be used. While we should have basic information about graphs before we start our study of functions, limits, and the

Figure 1.581

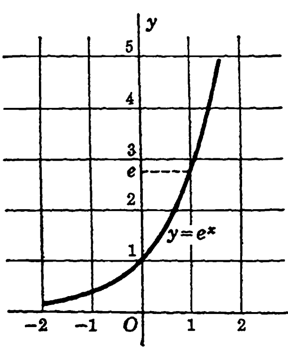
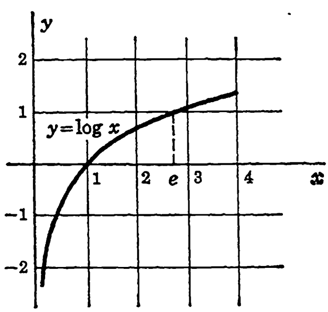


Figure 1.582



calculus, most of our work with equations and graphs will be done with the aid of the calculus.

Problems 1.59

Sketch graphs of the following equations and inequalities:

1 $y = x$	***2*** $x = 0$	***3*** $y = 0$
4 $y = (x - 1)^2$	***5*** $y = (x + 1)^2$	***6*** $y = x^3$
7 $y = 1 + x$	***8*** $y = 1 + x^2$	***9*** $xy = -1$
10 $y = \dfrac{1}{1 + x}$	***11*** $y = \dfrac{1}{(1 + x)^2}$	***12*** $y = \dfrac{1}{1 + x^2}$
13 $y = \dfrac{x}{1 + x^2}$	***14*** $y = x + \dfrac{1}{x}$	***15*** $y = x - \dfrac{1}{x}$
16 $y = \lvert x \rvert$	***17*** $y = \lvert x - 2 \rvert$	***18*** $y = \frac{1}{2}(x + \lvert x \rvert)$
19 $0 < x < 1$	***20*** $0 < y < 1$	***21*** $0 < x + y < 1$
22 $\lvert x \rvert < 1$	***23*** $\lvert x - 2 \rvert < \frac{1}{2}$	***24*** $\frac{3}{2} < x < \frac{5}{2}$
25 $y < x^2$	***26*** $\lvert y \rvert < \lvert x \rvert$	***27*** $\lvert y \rvert < x^2$
28 $\lvert x \rvert + \lvert y \rvert = 1$	***29*** $\lvert x \rvert + \lvert y \rvert < 1$	***30*** $\lvert x \rvert + \lvert y \rvert > 1$

31 With Figure 1.58 out of sight, sketch graphs of $y = \sin x$ and $y = \cos x$. If unsuccessful, glance at Figure 1.58 and try again.

32 Figure 1.591, which features half of an equilateral triangle each side of which has length 2, shows that

$$\sin\frac{\pi}{6} = \frac{1}{2}, \qquad \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}, \qquad \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}, \qquad \cos\frac{\pi}{3} = \frac{1}{2}.$$

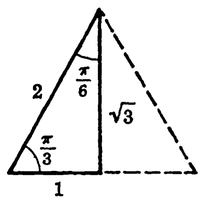


Figure 1.591

Cultivate the ability to sketch this figure quickly. Use the information obtained from it to locate points on the graphs of $y = \sin x$ and $y = \cos x$. Sketch a right triangle in which each leg has unit length and obtain more points on the graphs. Finally, sketch graphs of $y = \sin x$ and $y = \cos x$ again. *Remark:* We need familiarity with our graphs, and we need confidence in them.

33 Sketch graphs of

(*a*) $y = 3 \sin x$ (*b*) $y = \sin 2x$ (*c*) $y = \sin\left(x + \dfrac{\pi}{2}\right)$.

Remark: Graphs of equations of the form $y = E \sin (\omega x + \alpha)$ are called *sinusoids*, and we hear very often that E is the *amplitude*, ω is the *angular frequency*, and α is the *phase angle* of the sinusoid.

34 Where are the points (x,y) for which $0 \leqq x \leqq 2\pi$ and $\sin x \leqq y \leqq \cos x$?

35 Supposing that $h \neq 0$, find the slope m of the secant line (or chord) containing the two points of the graph of the equation $y = x^2$ having x coordinates x_1 and $x_1 + h$. *Ans.:* $2x_1 + h$.

36 It is sometimes quite important to have correct information about the graphs of $y = x^3$ and $y = x^{1/3}$. Sketch the graphs over the interval $-2 \leqq x \leqq 2$.

37 With the aid of the quadratic formula, show that the point (x,y) lies on the graph of the equation $x^2 + xy + y^2 = 3$ if and only if $-2 \leqq x \leqq 2$ and y is one of the two numbers

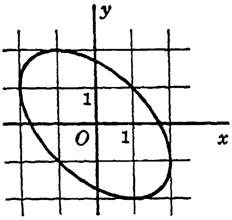


Figure 1.592

$$\frac{-x - \sqrt{3(4 - x^2)}}{2}, \qquad \frac{-x + \sqrt{3(4 - x^2)}}{2}$$

which are equal only when $x = -2$ and when $x = 2$. Formulate and prove an analogous statement in which the roles of x and y are interchanged. Find the coordinates of the eight points in which the graph intersects the lines having the equations $x = -2$, $x = 2$, $y = -2$, $y = 2$, $y = x$, and $y = -x$. *Remark:* The graph is an oval which is shown in Figure 1.592 and which is, as Chapter 6 will show us, an ellipse.

38 Sketch a graph of $y = \sin x$ over the interval $0 \leqq x \leqq 2\pi$ and then, with the aid of simple arithmetic facts like $0^2 = 0$, $(0.4)^2 = 0.16$, $(0.8)^2 = 0.64$, use the result to obtain a graph of $y = \sin^2 x$.

39 Sketch graphs of $y = \cos 2x$ and $y = (1 - \cos 2x)/2$. *Remark:* Because of the trigonometric identity

$$\sin^2 x = \frac{1 - \cos 2x}{2},$$

the answers to this and the preceding problem are the same.

40 Perhaps the classic guns-and-butter interpretation of the formula

$$x + y = M$$

should not be overlooked. It is supposed that a chief has control of M man-hours of human energy. The chief may preempt x man-hours to provide pressure and power to keep his subjects in line and to preserve or extend his authority. Then, even when $x = M$ and $y = 0$, there remain y man-hours part of which may be used for production of food, shelter, education, and sundries. Sketch a graph which shows how x and y are related. *Hint:* Do not ignore the basic idea that $x \geqq 0$ and $y \geqq 0$.

41 Sketch graphs of the three equations $\sqrt{y} = \sqrt{x}$, $y = x$, $y^2 = x^2$ and make some relevant comments.

42 Let $a > 0$. Show that the equation

$$\sqrt{x} + \sqrt{y} = \sqrt{a} \tag{1}$$

holds if and only if $0 \leqq x \leqq a$ and

$$y = (\sqrt{a} - \sqrt{x})^2 = a + x - 2\sqrt{ax}. \tag{2}$$

Without making onerous calculations, sketch rough graphs of (1), (2),

$$y = a + x + 2\sqrt{ax}, \tag{3}$$

and

$$(y - a - x)^2 = 4ax. \tag{4}$$

Remark: Chapter 6 will reveal the fact that the graph of (4) is a parabola. The graphs of (1), (2), and (3) are parts (subsets) of the parabola.

43 Let us suppose that a man who marries should select for his wife a woman whose age is 10 years more than half his age. Construct a graph for use of bachelors who are accustomed to picking information from graphs in the *Wall Street Journal* and everywhere else but are unaccustomed to making abstruse mathematical calculations.

44 Let time t be measured in seconds so that, as we can see by replacing x by t in Figure 1.58, $\sin t$ increases from 0 to 1 and decreases back to 0 in π (about 3) seconds. If you can acquire the ability to move your pencil point in the xy plane in such a way that its coordinates (x,y) at time t are $x = \sin t$ and $y = |\sin t|$, you will get a V for victory.

1.6 *Introduction to velocity and acceleration* Teachers of mathematics and physics are accustomed to difficulties involved in correlating studies of graphs, vectors, velocities, and accelerations in mathematics to studies of diagrams, forces, velocities, and accelerations in physics. There is a reason why it is not easy to achieve complete correlation. In order to be able to solve just one of his easiest problems involving motion of a body or particle, a physics student requires a little information about several basic concepts. This section is introduced at the end of our first chapter because it may be a desirable or even necessary part of some educational programs. Students can be advised to read it to obtain preliminary ideas about their external world but, so far as this course is concerned, can be advised to postpone the learning of the mathematics in it. Some and perhaps most teachers will proceed directly to the next chapter and will devote a classroom hour to this section only if and when their students face the prospect of studying falling bodies in their physics courses before they encounter derivatives and integrals in their mathematics courses. The next chapter, Chapter 2, treats vectors in space of three as well as fewer dimensions. While physicists can regret that this delays our full treatment of velocities and accelerations, they can also rejoice in the fact that the delay permits production of a much more useful treatment of the matter.†

As the preface states, the first third of this book contains all or nearly all of the analytic geometry and calculus that students normally encounter in their introductory full-year college and university courses in physics. In a few weeks, formulas like

$$s = \frac{1}{2}gt^2 + v_0t + s_0 \tag{1.611}$$

$$v = \frac{ds}{dt} = gt + v_0 \tag{1.612}$$

$$a = \frac{dv}{dt} = \frac{d^2s}{dt^2} = g \tag{1.613}$$

† This is a very conservative statement. Vectors, like numbers, are important things and there are many reasons why they should be encountered early and frequently when geometry and calculus are studied.

will be completely familiar and meaningful to us. Meanwhile, we make a preliminary study of ways in which they are related to experiments involving falling bodies. Chapters 3 and 4 will give much less information about the physics experiments but much more information about the mathematics.

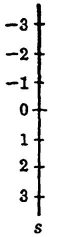


Figure 1.62

Suppose we have, as in Figure 1.62, a vertical s axis with the positive s axis below the origin. For laboratory experiments, we can place one meter stick above another and place minus signs in front of the numbers on the upper stick. We can suppose that a body is, at time $t = 0$, falling or just being dropped so that it travels past the markings on our meter sticks with increasing rapidity as time passes. On the other hand, we can suppose that the body is rising at time $t = 0$ so that it rises for a while before it begins its descent. We may suppose that distances are measured in centimeters, so that $s = 20$ when the body is 20 centimeters below the origin, and that t is measured in seconds, so that $t = 0.5$ when a timing device shows a half-second after our time origin or zero-hour.

Anyone who tosses a body upward and observes the ensuing motion must realize that it is not an easy matter to use an ordinary clock to obtain accurate data giving the coordinate s of the body at various times t. While solid information about such matters must be obtained from physicists, we can all recognize the possibility of getting useful data with the aid of apparatus so arranged that at each of the times $t = 0$, $t = 0.01$, $t = 0.02$, $t = 0.03$, $\cdots$ an electric spark jumps from a pointer on the falling body to burn a tiny hole in a long strip of paper attached to the meter sticks. When enough reasonably accurate information has been obtained in one way or another, we can use it to plot points (t,s) in a ts plane and obtain a graph more or less like that shown in Figure 1.63. For each t within the domain for which measurements are made, the s coordinate of the point $P(t,s)$ on our graph is a more or less good approximation to the coordinate or *displacement* of the body at time t.

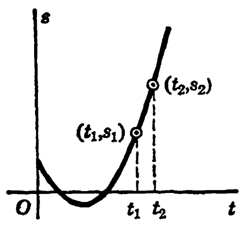


Figure 1.63

Some information should be in hand when we undertake to use our data and graph to obtain information about our falling body. Without pretending to have precise ideas yet, we can start with the rough idea that forces and velocities and accelerations exist and that these things are vectors or are represented by vectors. The reason our falling body plummets toward the center of the earth, with speed increasing when it is headed downward, is that the earth exerts a gravitational force upon it. The magnitude of this force is the weight of the body. Since we find no perceptible change in the weight of a body when we raise or lower it a few meters, we conclude that, so far as our problem is concerned, the magnitude of the gravitational force may be considered to be a constant, that is, the same at all places on our meter sticks. We can know that air resistance retards the motion of moving bodies but, when heavy bodies fall only a few meters, this produces consequences so small that our measurements are unaffected. Thus, so far as our measurements can tell, we are investigating the motion of a body which moves on a line through the center of the earth with only a constant gravitational force acting upon it.

In what follows, *vectors* are denoted by boldface letters as they usually are in printed scientific works.† Study of physics books or the next chapter reveals the meaning of the statement that the gravitational force $\mathbf{F}$ which the earth exerts upon our falling body is $mg\mathbf{u}$, where m is a positive number (the mass of the body), g is a positive number (the scalar acceleration of gravity), and $\mathbf{u}$ is a unit vector which lies on the line along which our body falls and is directed toward the center of the earth. The velocity $\mathbf{v}$ and the acceleration $\mathbf{a}$ of our falling body are vectors, but they are representable in the form $\mathbf{v} = v\mathbf{u}$ and $\mathbf{a} = a\mathbf{u}$, where v and a are real numbers that are not vectors and are sometimes called *scalars* to emphasize the fact that they are not vectors. Thus v is not a velocity, but it *is* the scalar component of a velocity. We call v a *scalar velocity*. Similarly, a is a *scalar acceleration*.

Fortified by at least a hazy understanding of the significance of our problem, we use experimental data of a table or of Figure 1.63 to learn about the scalar velocity v and the scalar acceleration a of our body. Let t_1 and t_2 be two different times and let s_1 and s_2 be the displacements of our body at these times. As the formula

$$\frac{s_2 - s_1}{t_2 - t_1} = \text{average scalar velocity} \tag{1.64}$$

indicates, the quotient on the left is called the *average scalar velocity* of our body over the time interval from the lesser to the greater of t_1 and t_2. In case $t_1 < t_2$ and $s_1 < s_2$, the quotient in (1.64) has a very familiar form. Except that the units may be different, the quotient is a positive number of miles divided by a positive number of hours and hence is a number of miles per hour that we normally call an average speed instead of an

† We pause to observe that boldface letters cannot be conveniently made with pencils, pens, crayons, and typewriters, and that a vector $\mathbf{F}$ (boldface) is often denoted by $\vec{F}$. Readers are advised to look at $\mathbf{F}$ (boldface) and imagine that there is an arrow on top of it so they will, in effect, see the $\vec{F}$ which they write when they want to emphasize the fact that the symbol is (or represents) a vector. Thus the formula $\mathbf{F} = m\mathbf{a}$ becomes $\vec{F} = m\vec{a}$ when it is transferred from printed material to handwritten hieroglyphics. Sometimes the arrows are printed to remove the necessity for use of imaginations, but we can, in effect, be paid for using our imaginations because printing the arrows increases costs of books.

average scalar velocity. After appropriate preliminary topics have been studied, Chapter 3 will tell precisely how the velocity **v** and the scalar velocity v at time t are defined. It turns out that the scalar velocity v and a number $\frac{ds}{dt}$, called the derivative of s with respect to t, are equal to each other and, moreover, that the quotient in (1.64) is nearly equal to v and $\frac{ds}{dt}$ whenever $t_1 = t$ and t_2 is nearly equal to t but $t_2 \neq t$. To obtain an estimate of the scalar velocity v at a particular time t from experimental data, which may be presented in a graph, it therefore suffices to calculate and use the average scalar velocity over a short time interval beginning or ending at t. Use of experimental data for this purpose is rendered difficult by the fact that, when t_1 and t_2 are nearly equal, small relative errors in measurements can produce huge errors in estimates of the value of the quotient $(s_2 - s_1)/(t_2 - t_1)$. It is a truly remarkable fact that when reasonably accurate data are collected and intelligently used, it is possible to estimate v for various values of t and to find that the points (t,v) in a tv plane come so close to lying on a line that all of the deviations can be attributed to errors in measurement and calculation. Thus our experimental work leads to the conclusion that, as in Figure 1.65, the graph of v versus t is either a part of a line or a very close approximation to a part of a line.

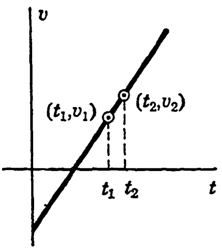


Figure 1.65

The scalar acceleration a of our falling body is defined in terms of the scalar velocity v in the same way that the scalar velocity v is defined in terms of the scalar displacement s. Thus, in addition to the basic formula (1.64), we have the basic formula

$$\text{(1.66)} \qquad \frac{v_2 - v_1}{t_2 - t_1} = \text{average scalar acceleration}$$

in which v_1 and v_2 are the scalar velocities at times t_1 and t_2. The scalar acceleration a at time t and the derivative $\frac{dv}{dt}$ are equal to each other and, moreover, the quotient in (1.66) is nearly equal to a and to $\frac{dv}{dt}$ whenever $t_1 = t$ and t_2 is nearly equal to t but $t_2 \neq t$. On the basis of the assumption that the graph of v versus t is a part of a line as in Figure 1.65, the average scalar acceleration is the slope m_1 of the part of the line. The hypothesis that each average scalar acceleration is the constant m_1 leads to the conclusion that, at each time t, the scalar acceleration is m_1; that is, $a = m_1$. Calculations from reasonably accurate data show that m_1 is about 980 when centimeters and seconds are used and about 32 when feet and seconds are used. This number m_1 is the gravitational constant g to which we have referred.

The simplest reasonable conclusion that can be drawn from data involving falling bodies is the following. To each place on the surface of the earth there corresponds a positive constant g, the scalar acceleration of gravity at that place, such that when a body moves on a vertical line near the surface of the earth with no appreciable external force other than the gravitational force exerted upon it, reasonable answers to problems can be based upon the assumption that the body is accelerated toward the center of the earth and that the scalar acceleration is g. Another similar but more lengthy conclusion involves the idea that the graph of v versus t is a line and that reasonable results are obtainable from the formula $v = gt + v_0$, where v_0 is a particular constant that depends upon choice of the time-origin used when studying a particular flight. Finally, it is possible to use the data and quite primitive mathematics to reach the more abstruse conclusion that there exist constants g, v_0, and s_0, the latter two of which depend upon the time-origin and the space-origin used in the study of a particular flight, such that reasonable results are obtainable from (1.611). A campaign to reach this conclusion can start with the observation that the graph in Figure 1.63 does look like a part of a parabola.

Mathematicians do not, except when they are behaving like physicists, actually perform physical experiments. Mathematicians cannot, unless they have physical laws or other information upon which proofs can be based, prove the formulas that are useful in mechanical dynamics and thermodynamics and hydrodynamics and aerodynamics and electrodynamics and economics and psychology and genetics and chemistry and cosmology. But mathematicians can, when they are given a few weeks, learn enough about derivatives and other things to enable them to start with given information and produce more information with astonishing ease. One who knows the content of Chapter 3 can start with the first of the three formulas

(1.671) $$s = \frac{1}{2}gt^2 + v_0t + s_0$$

(1.672) $$v = \frac{ds}{dt} = gt + v_0$$

(1 673) $$a = \frac{dv}{dt} = \frac{d^2s}{dt^2} = g$$

and produce the other two as fast as he can write. All he needs to do is apply standard rules for writing derivatives. The problems at the end of this section provide preliminary ideas about this matter. It is much more significant that one who knows the content of Chapter 4 can start with the last of the formulas and produce the other two as fast as he can write. All he needs to do is apply standard rules for writing integrals.

The physical significance of the constants in (1.671), (1.672), and (1.673) is worthy of notice. As we see by putting $t = 0$ in (1.671), s_0 is the value of s (the displacement) when $t = 0$, so s_0 is called the *initial displacement*. As we see by putting $t = 0$ in (1.672), v_0 is the scalar velocity when $t = 0$, so v_0 is called the *initial scalar velocity*. It is easily seen from (1.672) that $v = 0$ when $t = -v_0/g$. The values (if any) of t for which $s = 0$ can be obtained by putting $s = 0$ in (1.671) and solving the resulting quadratic equation for t. In many applications, the space and time coordinates are so chosen that the initial displacement s_0 and initial velocity v_0 are both 0. In this case (1.671) reduces to the simpler formula

$$s = \tfrac{1}{2}gt^2. \tag{1.674}$$

The related formulas

$$t = \sqrt{\frac{2s}{g}} \qquad v = gt = \sqrt{2gs}, \tag{1.675}$$

which give the time required for the body to fall a distance s and the speed attained when the body has fallen a distance s, are often useful.

We conclude with a remark about uniform circular motion. Suppose a particle starts at time $t = 0$ on the positive x axis and moves, with angular speed ω (omega) radians per second, in the positive (counterclockwise) direction around the circle of radius R having its center at the origin. Letting $\mathbf{r}$ denote the vector running from the origin to the particle P at time t gives the first of the formulas

$$\mathbf{r} = R(\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j}) \tag{1.681}$$

$$\mathbf{v} = \omega R(-\sin \omega t\mathbf{i} + \cos \omega t\mathbf{j}) \tag{1.682}$$

$$\mathbf{a} = -\omega^2 R(\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j}) \tag{1.683}$$

where, as in Figure 1.684, $\mathbf{i}$ and $\mathbf{j}$ are unit vectors having the directions of

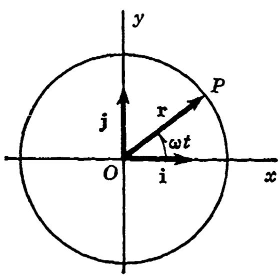


Figure 1.684

the positive x and y axes. Application of rules of Chapter 3 then gives (1.682) and (1.683) as rapidly as we can write them. Looking at (1.681) and (1.683) shows that $\mathbf{a} = -\omega^2\mathbf{r}$ and hence that P is always accelerated toward the center. The length of the vector $\mathbf{r}$ is always R, and the length of the vector $\mathbf{a}$ is always $\omega^2 R$. This shows that the magnitude of the acceleration is always $\omega^2 R$. These results are important in physics and engineering. Physics books that do not make effective use of good mathematics do not derive their results so efficiently.

Problems 1.69

1 Supposing that g, v_0, and s_0 are constants and that

$$s = \tfrac{1}{2}gt^2 + v_0 t + s_0 \tag{1}$$

at each time t, use notation like that in (1.64), so that $s = s_1$ when $t = t_1$ and $s = s_2$ when $t = t_2$, to obtain the formula

$$s_2 - s_1 = \tfrac{1}{2}g(t_2^2 - t_1^2) + v_0(t_2 - t_1) \tag{2}$$

and hence

$$\frac{s_2 - s_1}{t_2 - t_1} = \tfrac{1}{2}g(t_2 + t_1) + v_0 \tag{3}$$

when $t_2 \neq t_1$. *Remark:* Even though we have not yet encountered procedures by which such statements are made precise, we can temporarily accept without question the statement that the right side of (3) must be near $gt + v_0$ whenever t_1 and t_2 are both near t and hence that

$$v = gt + v_0. \tag{4}$$

2 Supposing that g and v_0 are constants such that

$$v = gt + v_0 \tag{1}$$

at each time t, use notation like that in (1.66), so that $v = v_1$ when $t = t_1$ and $v = v_2$ when $t = t_2$, to obtain the formula

$$v_2 - v_1 = g(t_2 - t_1) \tag{2}$$

and hence

$$\frac{v_2 - v_1}{t_2 - t_1} = g \tag{3}$$

when $t_2 \neq t_1$. *Remark:* A remark similar to that of the preceding problem is applicable here; the scalar acceleration a at time t is g.

3 We should now be well aware of the fact that Problems 9.29 will appear at the end of Chapter 9, Section 2. While the trick is not used in this book, we can use the numbers 9.2908 and 9.2922 to identify problems 8 and 22 of Problems 9.29. Now comes the problem. Write a single number to identify formula 15 of Problem 4 at the end of Section 6 of Chapter 12. *Ans.:* 12.690415. Persons who feel that this trick is complicated should think about the matter to capture some of the spirit of members of a research staff of a data processing department of IBM (International Business Machine Corporation) who find that such tricks keep them in business.

2 Vectors and geometry in three dimensions

2.1 Vectors in E_3 To facilitate discussions and solutions of problems in geometry and calculus, and for many other purposes in pure and applied mathematics, it is necessary to know about things called vectors. All points and vectors with which we are concerned are supposed to lie in ordinary Euclid space E_3 of three dimensions in which such things as points, lines, planes, cubes, spheres, and automobiles can exist. The definitions of this section do not depend upon a coordinate system and are therefore said to be *intrinsic definitions.* We shall hear more about this matter later.

Before introducing vectors, we observe the familiar fact that two distinct (that is, different) points P_1 and P_2 determine the line P_1P_2 which passes through P_1 and P_2 and extends beyond P_1 and P_2 in two directions as in Figure 2.11. Vectors are more like line segments than like lines. An ordered pair P_1, P_2 of distinct points, in which P_1 and P_2 are respectively the first point and the second point in the pair, determines the *vector* $\overline{P_1P_2}$ or "arrow" or "directed line segment" which runs (or extends) from the first point to the second point as in Figure 2.111. The purpose of the arrowhead is to show that the vector runs from P_1 to P_2. The vector shown in Figure 2.112 is not $\overrightarrow{P_1P_2}$, but is $\overrightarrow{P_2P_1}$. The *length* (or magnitude) $|\overrightarrow{P_1P_2}|$ of a vector $\overrightarrow{P_1P_2}$ is the length of the line segment upon which it lies, that is, the distance between the points P_1 and P_2. If P_2 and P_1 coincide, that is, $P_2 = P_1$, the points do not determine a line but they do determine the vector $\overrightarrow{P_1P_2}$, which has length 0 and which is called the *zero vector*.

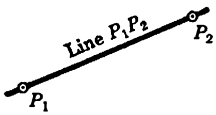


Figure 2.11

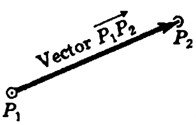


Figure 2.111

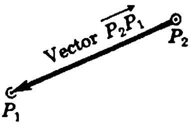


Figure 2.112

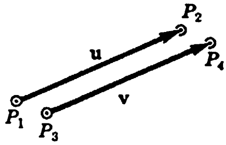
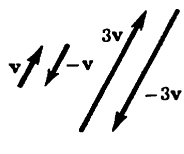


Figure 2.113

As indicated in Figure 2.113, vectors are often denoted by boldface letters which keep us informed that the symbols represent vectors rather than numbers or chemical elements. Thus we can set $\mathbf{u} = \overrightarrow{P_1P_2}$ and $\mathbf{v} = \overrightarrow{P_3P_4}$. Two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be *equal*, and we write $\mathbf{u} = \mathbf{v}$ when, as in Figure 2.113, they (i) lie on parallel lines, (ii) have equal lengths, and (iii) have the same (not opposite) directions. Two zero vectors $\mathbf{u}_0$ and $\mathbf{v}_0$ do not have directions, but we say that $\mathbf{u}_0 = \mathbf{v}_0$ anyway. If $\mathbf{u}$ is a nonzero vector and $\mathbf{v}$ is a zero vector, then $\mathbf{u} \neq \mathbf{v}$. We use the ordinary 0 (zero) to denote the zero vector; it turns out that we will not need an arrow or distinctive type face to tell us whether 0 is the number zero or a vector having length zero. The advice given in a the footnote on page 41 merits repetition here. Whenever we see **F**(boldface) or any other letter that is boldface, we recognize that it is a vector and imagine that there is an arrow above it so that we, in effect, see the symbols $\vec{F}$, $\vec{u}$, etcetera, that are made by pencils, pens, and chalk. Thus our imaginations convert what we see into what we write, and the disadvantage of boldface print has disappeared.

It is both interesting and important to know what is meant by the *product* $k\mathbf{v}$ of a number (real number or *scalar*) k and a vector $\mathbf{v}$ and by the sum $\mathbf{u} + \mathbf{v}$ of two vectors. The definitions will imply validity of the formula $2\mathbf{u} = \mathbf{u} + \mathbf{u}$ as well as other useful formulas. In case $k = 0$ or $\mathbf{v} = 0$ or both, the product $k\mathbf{v}$ is the zero vector, that is, $k\mathbf{v} = 0$. In case $k \neq 0$ and $\mathbf{v} \neq 0$, the vector $k\mathbf{v}$ is a vector such that (i) $\mathbf{v}$ and $k\mathbf{v}$ lie on the same or parallel lines, (ii) the length of $k\mathbf{v}$ is $|k|\ |\mathbf{v}|$, and (iii) $\mathbf{v}$ and $k\mathbf{v}$ have the same direction if $k > 0$ and opposite directions if $k < 0$. Figure 2.12 shows examples. This definition implies that if $\mathbf{v}$ is a nonzero vector, then the *unit vector* (vector one unit long) in the direction of $\mathbf{v}$ is $(1/|\mathbf{v}|)\mathbf{v}$ or $\mathbf{v}/|\mathbf{v}|$.

Figure 2.12



The *sum* $\mathbf{r}$ of two vectors $\mathbf{u}$ and $\mathbf{v}$ is, ~~as in Figure 2.13~~, the vector which runs from the tail of $\mathbf{u}$ to the head of $\mathbf{v}$ when the tail of $\mathbf{v}$ is placed at the head of $\mathbf{u}$. The figure shows that $\mathbf{v} + \mathbf{u} = \mathbf{u} + \mathbf{v}$. Because the sum of two vectors is, as in Figure 2.13, the diagonal of a parallelogram, the rule (or law) for addition of vectors is called the *parallelogram law.* Figure 2.131 shows the sum $\mathbf{r}$ of four vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{u}_4$. In applied mathematics the sum of two or more vectors is sometimes called their *resultant.*

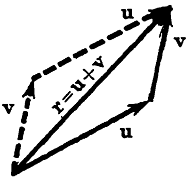


Figure 2.13

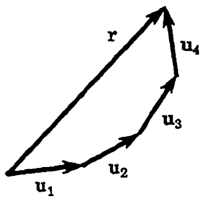


Figure 2.131

The difference $\mathbf{u} - \mathbf{v}$ is defined to be the sum of $\mathbf{u}$ and $-\mathbf{v}$, so that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$. The most obvious way to find $\mathbf{u} - \mathbf{v}$ is to find $-\mathbf{v}$ and add it to $\mathbf{u}$. In substantially all cases, it is quicker, easier, and more useful to observe that $\mathbf{u} - \mathbf{v}$ is the vector which we must add to $\mathbf{v}$ to obtain the sum $\mathbf{u}$. When the tails of $\mathbf{u}$ and $\mathbf{v}$ coincide, *the vector* $\mathbf{u} - \mathbf{v}$ *runs from the head of* $\mathbf{v}$ *to the head of* $\mathbf{u}$. It is worthwhile to look at the italicized statement and Figure 2.14 until both are thoroughly understood and remembered. The figure clearly says that

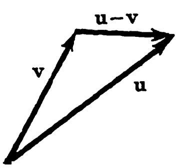


Figure 2.14

$$\mathbf{u} = \mathbf{v} + (\mathbf{u} - \mathbf{v}). \tag{2.141}$$

Since angles between vectors can be sources of confusion and misunderstanding, we give a little careful attention to the subject. In case one or the other of two vectors has length 0, there is no reasonable way to determine an angle that should be called the angle between them, and we say that the angle is undetermined or undefined. Two sharpened pencils of positive length represent vectors in the directions of their sharpened tips. In case these vectors do not intersect, we can choose any point O in E_3 and replace the vectors by equal vectors having their tails at O as in Figure 2.15. Suppose first that these vectors $\mathbf{u}$ and $\mathbf{v}$ have neither the same nor opposite directions. These vectors then determine the plane in which they lie. The angle θ is determined by the method used in trigonometry to introduce radian measure. The first step is to draw, in the plane of the vectors, a circle of radius a with center at O and to find the length s of the shorter of the two arcs into which the vectors cut the circle. The number θ defined by

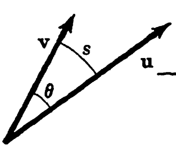


Figure 2.15

$$\theta = \frac{s}{a} \qquad \text{or} \qquad \text{angle} = \frac{\text{length of arc}}{\text{radius}} \tag{2.151}$$

is called the *angle between* **u** *and* **v** or *the angle which* **u** *makes with* **v** or *the angle which* **v** *makes with* **u**. Thus angles are numbers.† If **u** and **v** have the same (or opposite) directions, slight modifications of the above construction give $s = 0$ (or $s = \pi a$) and the same formula (2.151) is used to define θ. In each case we have $0 \leqq s \leqq \pi a$, and hence $0 \leqq \theta \leqq \pi$. When working with angles between vectors, we never have to bother with "negative angles" and "angles greater than straight angles." For perpendicular vectors, we have $\theta = \pi/2$. We, like electronic computers and some trigonometric tables, use radian measure and seldom bother with degrees, minutes, and seconds.

The remainder of the text (not problems) of this section gives basic information about products of vectors. The importance of the material will be revealed later in this book and by textbooks in other subjects in pure and applied mathematics. It is not necessary to presume that the material is difficult. In fact, students who do not have the good fortune to study this material calmly in mathematics sometimes find that their teachers in physics and engineering undertake to teach all of it in a few seconds.

There are two different kinds of elementary products of vectors **u** and **v** that turn out to be interesting and useful. These are the *scalar product* (or dot product) defined by the formula

$$\mathbf{u}\cdot\mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \cos\theta \tag{2.16}$$

and the *vector product* (or cross product) defined by the formula

$$\mathbf{u} \times \mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \sin\theta\,\mathbf{n}. \tag{2.17}$$

These formulas will now be discussed. If $\mathbf{u} = 0$ or $\mathbf{v} = 0$ or both, the angle θ appearing in the formulas is not determined by **u** and **v**, but the products $\mathbf{u}\cdot\mathbf{v}$ and $\mathbf{u} \times \mathbf{v}$ are defined to be 0 anyway. Henceforth, we consider cases in which $|\mathbf{u}| > 0$ and $|\mathbf{v}| > 0$, these being the lengths of **u** and **v**. Then, as in Figures 2.18 and 2.181, the two vectors determine an angle θ for which $0 \leqq \theta \leqq \pi$. In case $0 \leqq \theta \leqq \pi/2$, the number $|\mathbf{v}| \cos\theta$ is the length of the projection of the vector **v** on the vector **u** and the scalar product is therefore the product of the length of **u** and the length

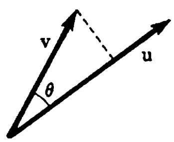


Figure 2.18

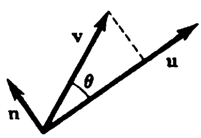


Figure 2.181

† Dictionaries convey assorted ideas akin to the ideas that an angle is the "enclosed space" or "corner" or "opening" near the point where two intersecting lines meet. While we need not expect to be injured by conflicting meanings of the word angle, we can use the term "geometric angle θ" to signify the "opening" between the two vectors of Figure 2.15. The number θ is then a measure of the size of the geometric angle θ, and we have satisfactory but somewhat awkward terminology.

of the projection of $\mathbf{v}$ on $\mathbf{u}$. In case $\pi/2 < \theta \leqq \pi$, the scalar product is the negative of this number. The definition of $\mathbf{u}\cdot\mathbf{v}$ implies that $\mathbf{u}\cdot\mathbf{v} = 0$ if and only if $\cos\theta = 0$. Thus $\mathbf{u}\cdot\mathbf{v} = 0$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* (that is, perpendicular to each other). Those who are or want to be conversant with principles of physics can note that if a particle P moves from the tail to the head of the vector $\mathbf{u}$ with the constant force $\mathbf{v}$ acting upon P during the motion, then $\mathbf{u}\cdot\mathbf{v}$ is the work done by the force during the motion.

Referring to Figure 2.181, we can see that if $\mathbf{u}$ and $\mathbf{v}$ are collinear vectors (vectors which lie on the same line), then θ is 0 or π, so $\sin\theta = 0$. In this case the vector $\mathbf{n}$ of the formula (2.17) is not determined, but $\mathbf{u} \times \mathbf{v}$ is defined to be 0 anyway. Henceforth, we suppose that $0 < \theta < \pi$. In this case, the vector $\mathbf{n}$ is the unit normal to the plane of $\mathbf{u}$ and $\mathbf{v}$ which is determined by the *right-hand rule*. A right hand is so placed that the thumb is perpendicular to the plane of $\mathbf{u}$ and $\mathbf{v}$ and the fingers are parallel to this plane and point in the direction that a line rotates in passing over the geometric angle θ from $\mathbf{u}$ to $\mathbf{v}$ (not $\mathbf{v}$ to $\mathbf{u}$). The *unit normal* $\mathbf{n}$ is then the vector which has the direction of the thumb and which is one unit long. From Figure 2.181 we see that $|\mathbf{v}| \sin\theta$ is the altitude of the triangle of which the vectors $\mathbf{u}$ and $\mathbf{v}$ form two sides. It follows from (2.17) that $\mathbf{u} \times \mathbf{v} = 2A\mathbf{n}$, where A is the area of this triangle. It must always be remembered that the vector product $\mathbf{u} \times \mathbf{v}$ is a vector which, when it is not 0, has the direction of the thumb when the right-hand rule is applied. Moreover, it is necessary to observe and remember that, except when $\mathbf{u} \times \mathbf{v} = 0$, the vector $\mathbf{v} \times \mathbf{u}$ is not the same as the vector $\mathbf{u} \times \mathbf{v}$. After having found $\mathbf{u} \times \mathbf{v}$ by the right-hand rule, we must flip the hand over so that the thumb points in the opposite direction to find $\mathbf{v} \times \mathbf{u}$, and it follows that

$$\mathbf{v} \times \mathbf{u} = -\mathbf{u} \times \mathbf{v}. \tag{2.182}$$

Anyone can attain complete understanding of these matters by making a few experiments in which two pencils (representing vectors) are held in the left hand while the right hand is used to determine the direction of their vector product. While vector products appear infrequently in this book, they have many important applications.

Finally, we call attention to some simple formulas that are easy to use but are not so easy to prove. The basic formula, which is proved in Problem 17 below, is

$$\begin{aligned}(\mathbf{u}_1 + \mathbf{u}_2)\cdot(\mathbf{v}_1 + \mathbf{v}_2) &= \mathbf{u}_1\cdot(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{u}_2\cdot(\mathbf{v}_1 + \mathbf{v}_2)\\ &= \mathbf{u}_1\cdot\mathbf{v}_1 + \mathbf{u}_1\cdot\mathbf{v}_2 + \mathbf{u}_2\cdot\mathbf{v}_1 + \mathbf{u}_2\cdot\mathbf{v}_2.\end{aligned} \tag{2.183}$$

Analogous formulas hold when the parentheses in the left member contain sums of more than two vectors. Moreover, correct formulas are obtained by replacing the dots by crosses. Proofs of this fact are given in textbooks on vector analysis.

Problems 2.19

1 As in Figure 2.191, let $A, B, C, \cdots, H$ be equally spaced points on the line P_1P_2 with $C = P_1$ and $G = P_2$. Apply appropriate definitions of the text to show that

$$\begin{aligned}
&\overrightarrow{P_1D} = \tfrac{1}{4}\overrightarrow{P_1P_2}, && \overrightarrow{P_1E} = \tfrac{1}{2}\overrightarrow{P_1P_2}, && \overrightarrow{P_1F} = \tfrac{3}{4}\overrightarrow{P_1P_2}, && \overrightarrow{P_1G} = \overrightarrow{P_1P_2} \\
&\overrightarrow{P_1H} = \tfrac{5}{4}\overrightarrow{P_1P_2}, && \overrightarrow{P_1B} = -\tfrac{1}{4}\overrightarrow{P_1P_2}, && \overrightarrow{P_1A} = -\tfrac{1}{2}\overrightarrow{P_1P_2}, && \overrightarrow{DE} = \tfrac{1}{4}\overrightarrow{P_1P_2}.
\end{aligned}$$

Observe that the vector $\overrightarrow{P_1P}$ lies on the line $\overrightarrow{P_1P_2}$ if and only if there is a scalar (or number or constant) λ (lambda) such that

$$\overrightarrow{P_1P} = \lambda\overrightarrow{P_1P_2}.$$

Observe that the points P_1 and P_2 separate the line into three parts and tell what values of λ correspond to points in the different parts.

2 Construct a figure similar to Figure 2.191 which shows points P_1 and P_2 and also points A, B, C, D for which

$$\begin{aligned}
&\overrightarrow{P_1A} = -\tfrac{1}{3}\overrightarrow{P_1P_2}, && \overrightarrow{P_1B} = \tfrac{1}{3}\overrightarrow{P_1P_2}, \\
&\overrightarrow{P_1C} = \tfrac{2}{3}\overrightarrow{P_1P_2}, && \overrightarrow{P_1D} = \tfrac{4}{3}\overrightarrow{P_1P_2}.
\end{aligned}$$

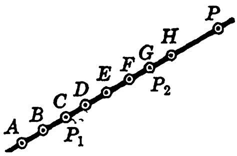


Figure 2.191

3 Let O (an origin), P_1, and P_2 be three points in E_3 with $P_1 \neq P_2$ as in Figure 2.192. Verify that if P is a point on the line P_1P_2, then there is a scalar λ for which

$$\overrightarrow{P_1P} = \lambda\overrightarrow{P_1P_2} = \lambda(\overrightarrow{OP_2} - \overrightarrow{OP_1})$$

and

$$\overrightarrow{OP} = \overrightarrow{OP_1} + \overrightarrow{P_1P} = \overrightarrow{OP_1} + \lambda(\overrightarrow{OP_2} - \overrightarrow{OP_1})$$

so

$$\overrightarrow{OP} = \lambda\overrightarrow{OP_2} + (1 - \lambda)\overrightarrow{OP_1}.$$

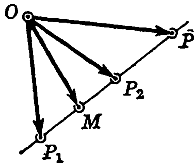


Figure 2.192

Show that if M is the mid-point of the line segment P_1P_2, then

$$\overrightarrow{OM} = \tfrac{1}{2}(\overrightarrow{OP_1} + \overrightarrow{OP_2}).$$

4 Let **i**, **j**, and **k** be mutually perpendicular vectors which run along bottom and back edges of a cube as in Figure 2.193. Let P_1, P_2, P_3, P_4 be the mid-points of the top edges upon which they lie. Show that

$$\overrightarrow{OP_1} = \tfrac{1}{2}\mathbf{i} + \mathbf{k}, \qquad \overrightarrow{OP_2} = \mathbf{i} + \tfrac{1}{2}\mathbf{j} + \mathbf{k},$$

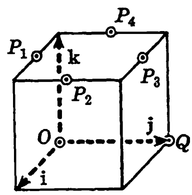


Figure 2.193

and write similar formulas for $\overrightarrow{OP_3}$, $\overrightarrow{OP_4}$, and $\overrightarrow{QP_2}$.

5 Supposing that the vectors **i**, **j**, **k** of the preceding problem are unit vectors, apply the definitions of products of vectors to prove that

$$\mathbf{i}\cdot\mathbf{i} = 1, \qquad \mathbf{i}\cdot\mathbf{j} = 0, \qquad \mathbf{i} \times \mathbf{i} = 0, \qquad \mathbf{i} \times \mathbf{j} = \mathbf{k}.$$

Hint: In each case, write the definition of the product and use the angle correctly.

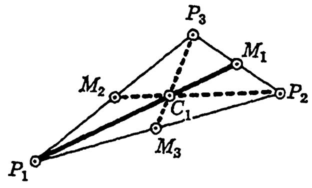


Figure 2.194

6 We are going to prove a theorem in geometry. As in Figure 2.194, let P_1, P_2, P_3 be vertices of a triangle and, for each $k = 1, 2, 3$, let M_k be the mid-point of the side opposite P_k. For each k, let C_k be the point of trisection of the segment P_kM_k for which $|\overrightarrow{P_kC_k}| = \frac{2}{3}|\overrightarrow{P_kM_k}|$. We will prove that the points C_1, C_2, C_3 coincide and we may put $C = C_1 = C_2 = C_3$. The line segments P_kM_k are, in geometry, called *medians* of the triangle Thus, our result shows that *the three medians intersect at a point C which trisects each of them.* For reasons which we shall not now discuss, the point C is the *centroid* of the triangular region T bounded by the sides of the triangle. To prove our result, let O be any point and show that

$$\begin{aligned}\overrightarrow{OC_1} &= \overrightarrow{OP_1} + \tfrac{2}{3}\overrightarrow{P_1M_1} = \overrightarrow{OP_1} + \tfrac{2}{3}(\overrightarrow{OM_1} - \overrightarrow{OP_1}) \\ &= \overrightarrow{OP_1} + \frac{2}{3}\left(\frac{\overrightarrow{OP_2} + \overrightarrow{OP_3}}{2} - \overrightarrow{OP_1}\right) = \frac{\overrightarrow{OP_1} + \overrightarrow{OP_2} + \overrightarrow{OP_3}}{3}.\end{aligned}$$

The way in which P_1, P_2, and P_3 appear in the result can make us feel sure that we must have

$$\overrightarrow{OC_k} = \frac{\overrightarrow{OP_1} + \overrightarrow{OP_2} + \overrightarrow{OP_3}}{3}$$

for each k. However, calculate $\overrightarrow{OC_2}$ and $\overrightarrow{OC_3}$ and show that it is so.

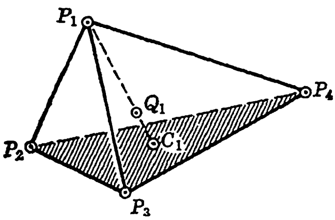


Figure 2.195

7 The line segment P_kC_k joining a vertex P_k of a tetrahedron to the centroid C_k of the opposite face, as in Figure 2.195, is called a *median* of the tetrahedron. For each k, let Q_k be the point of quadrisection of the median P_kC_k for which $|\overrightarrow{P_kQ_k}| = \frac{3}{4}|\overrightarrow{P_kC_k}|$. Let O be any point. When $k = 1$, prove the formula

$$\overrightarrow{OQ_k} = \frac{\overrightarrow{OP_1} + \overrightarrow{OP_2} + \overrightarrow{OP_3} + \overrightarrow{OP_4}}{4}$$

and then prove or guess that the formula is valid when $k = 1, 2, 3, 4$. The point Q for which

$$\overrightarrow{OQ} = \frac{\overrightarrow{OP_1} + \overrightarrow{OP_2} + \overrightarrow{OP_3} + \overrightarrow{OP_4}}{4}$$

is the centroid of the tetrahedron. Thus the four medians of a tetrahedron intersect at the centroid, and this centroid quadrisects each median.

8 Prove that the line segment joining the mid-points of two opposite edges of a tetrahedron contains and is bisected by the centroid of the tetrahedron.

9 Determine whether, in all cases, the two line segments joining mid-points of opposite edges of a quadrilateral must intersect and bisect each other. Be sure to recognize that a quadrilateral in E_3 need not have all of its vertices in the same plane.

10 Prove that if P_1, P_2, P_3, P_4 are the vertices of a square having its center at C, then

$$\overrightarrow{OC} = \frac{\overrightarrow{OP_1} + \overrightarrow{OP_2} + \overrightarrow{OP_3} + \overrightarrow{OP_4}}{4}.$$

Hint: For each k we can write $\overrightarrow{OP_k} = \overrightarrow{OC} + \overrightarrow{CP_k}$ and notice that something can be said about $\overrightarrow{CP_k}$ and $\overrightarrow{CP_j}$ when P_k and P_j are opposite vertices of the square.

11 Prove that if $P_1, P_2, \cdots, P_8$ are the vertices of a cube having its center at C, then

$$\overrightarrow{OC} = \frac{\overrightarrow{OP_1} + \overrightarrow{OP_2} + \cdots + \overrightarrow{OP_8}}{8}.$$

12 Figure 2 196 shows eight vectors $\mathbf{i}$, $\mathbf{j}$, $-\mathbf{i}$, $-\mathbf{j}$, $\mathbf{i}'$, etcetera which are unit vectors (vectors one unit long) having their tails at the origin of an xy plane and having their tips on the unit circle with center at the origin. Discover reasons why the first pair of equations

$$\begin{cases} \mathbf{i} + \mathbf{j} = \sqrt{2}\,\mathbf{i}' \\ -\mathbf{i} + \mathbf{j} = \sqrt{2}\,\mathbf{j}' \end{cases} \qquad \begin{cases} \mathbf{i} = \dfrac{1}{\sqrt{2}}(\mathbf{i}' - \mathbf{j}') \\ \mathbf{j} = \dfrac{1}{\sqrt{2}}(\mathbf{i}' + \mathbf{j}') \end{cases}$$

is valid, and then solve the first pair to obtain the second pair.

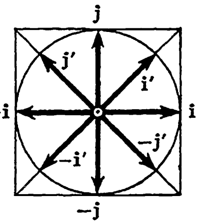


Figure 2.196

13 The vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ run in positive (counterclockwise) directions along three consecutive sides of a regular hexagon. Express $\mathbf{w}$ in terms of $\mathbf{u}$ and $\mathbf{v}$ *Hint:* Sketch the hexagon and the line segments from the center that separate it into six equilateral triangles. Perhaps the simplest observation that can be made is $\mathbf{u} + \mathbf{v} + \mathbf{w} = 2\mathbf{v}$.

14 From the vertices of a triangle, vectors are drawn to the mid-points of the opposite sides. Prove that the sum of the three vectors is zero.

15 What can be said about the location of Q if $\overrightarrow{PQ} = \overrightarrow{PA} + \overrightarrow{PB} + \overrightarrow{PC} + \overrightarrow{PD}$. where A, B, C, D are the vertices of a square and P is on a side of the square?

16 Abilities to sketch figures and construct formulas involving vectors must be cultivated. As in Figure 2.197, let $\mathbf{u}$ be a unit vector (vector of unit length) having its tail at O. Show that $\mathbf{u}\cdot\mathbf{u} = 1$. Let $\mathbf{v}$ be another vector having its tail at the same point O. Show that the vector $\mathbf{U}$ defined by $\mathbf{U} = (\mathbf{v}\cdot\mathbf{u})\mathbf{u}$ is the vector running from O to the projection of the tip of $\mathbf{v}$ on the line bearing the vector $\mathbf{u}$. Observe that the vector $\mathbf{V}$ defined by $\mathbf{V} = \mathbf{v} - \mathbf{U}$ or by $\mathbf{V} = \mathbf{v} - (\mathbf{v}\cdot\mathbf{u})\mathbf{u}$ runs from the tip of $\mathbf{U}$ to the tip of $\mathbf{v}$. Verify that $\mathbf{V}$ is perpendicular to $\mathbf{U}$ by showing that

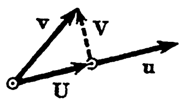


Figure 2.197

$$\mathbf{U}\cdot\mathbf{V} = (\mathbf{v}\cdot\mathbf{u})\mathbf{u}\cdot[\mathbf{v} - (\mathbf{v}\cdot\mathbf{u})\mathbf{u}] = 0.$$

Remark: More opportunities to become familiar with these things will appear later. Meanwhile, we can note that we have seen the (or a) standard procedure for resolving a given vector $\mathbf{v}$ into vector components parallel and perpendicular to a given unit vector $\mathbf{u}$.

17 Sketch a figure showing four points O, P_1, P_2, and Q in E_3 and suppose that $|\overrightarrow{OQ}| = 1$. Let Q_1 and Q_2 be the projections of P_1 and P_2 on the line OQ. Show that

$$(1) \qquad [\overrightarrow{OP_1}\cdot\overrightarrow{OQ}]\overrightarrow{OQ} = \overrightarrow{OQ_1}, \qquad [\overrightarrow{P_1P_2}\cdot\overrightarrow{OQ}]\overrightarrow{OQ} = \overrightarrow{Q_1Q_2},$$

$$(2) \qquad [(\overrightarrow{OP_1} + \overrightarrow{P_1P_2})\cdot\overrightarrow{OQ}]\overrightarrow{OQ} = \overrightarrow{OQ_2}$$

and hence that

$$(3) \qquad [(\overrightarrow{OP_1} + \overrightarrow{P_1P_2})\cdot\overrightarrow{OQ}]\overrightarrow{OQ} = (\overrightarrow{OP_1}\cdot\overrightarrow{OQ} + \overrightarrow{P_1P_2}\cdot\overrightarrow{OQ})\overrightarrow{OQ}$$

and therefore

$$(4) \qquad (\overrightarrow{OP_1} + \overrightarrow{P_1P_2})\cdot\overrightarrow{OQ} = \overrightarrow{OP_1}\cdot\overrightarrow{OQ} + \overrightarrow{P_1P_2}\cdot\overrightarrow{OQ}.$$

Remark: If we set $\mathbf{u}_1 = \overrightarrow{OP_1}$, $\mathbf{u}_2 = \overrightarrow{P_1P_2}$, and $\mathbf{v} = \overrightarrow{OQ}$, this shows that the formula

$$(5) \qquad (\mathbf{u}_1 + \mathbf{u}_2)\cdot\mathbf{v} = \mathbf{u}_1\cdot\mathbf{v} + \mathbf{u}_2\cdot\mathbf{v}$$

is valid when $\mathbf{u}_1$ and $\mathbf{u}_2$ are vectors and $\mathbf{v}$ is a unit vector. It follows from this that (5) is valid whenever $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{v}$ are vectors. With the aid of (5) and simpler properties of scalar products, we find that

$$(6) \qquad \begin{aligned}(\mathbf{u}_1 + \mathbf{u}_2)\cdot(\mathbf{v}_1 + \mathbf{v}_2) &= (\mathbf{v}_1 + \mathbf{v}_2)\cdot(\mathbf{u}_1 + \mathbf{u}_2)\\ &= \mathbf{v}_1\cdot(\mathbf{u}_1 + \mathbf{u}_2) + \mathbf{v}_2\cdot(\mathbf{u}_1 + \mathbf{u}_2)\\ &= (\mathbf{u}_1 + \mathbf{u}_2)\cdot\mathbf{v}_1 + (\mathbf{u}_1 + \mathbf{u}_2)\cdot\mathbf{v}_2\end{aligned}$$

and hence

$$(7) \qquad (\mathbf{u}_1 + \mathbf{u}_2)\cdot(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{u}_1\cdot\mathbf{v}_1 + \mathbf{u}_1\cdot\mathbf{v}_2 + \mathbf{u}_2\cdot\mathbf{v}_1 + \mathbf{u}_2\cdot\mathbf{v}_2.$$

This is the basic formula (2.183).

18 This problem and the next involve some very simple but very important ideas. Let $\mathbf{r}$ be the vector running from the origin to $P(x,y)$, the point P having coordinates x and y, as in Figure 2.198. Let $\mathbf{i}$ be a unit vector having the direction of the positive x axis. Considering separately the cases in which $x > 0$, $x = 0$, and $x < 0$, show that $x\mathbf{i}$ is the vector running from the origin to the projection of P upon the x axis. Then let $\mathbf{j}$ be a unit vector having the direction of the positive y axis and prove that $y\mathbf{j}$ is the vector running from the origin to the projection of P upon the y axis. *Hint:* All that is required is appropriate use of the definition of the product of a scalar and a vector.

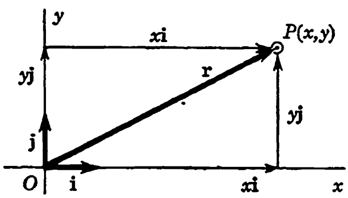


Figure 2.198

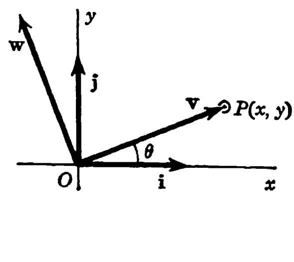


Figure 2.199

19 As in Figure 2.199, let $\mathbf{i}$ and $\mathbf{j}$ be unit vectors having the directions of the x and y axes of a plane coordinate system. Let $\mathbf{v}$ be a nonzero vector running from the origin to $P(x,y)$. Show that

$$(1) \qquad \mathbf{v} = x\mathbf{i} + y\mathbf{j}.$$

Show that if θ, not necessarily confined to the interval $0 \leqq \theta \leqq \pi$, is (as in trigonometry) an angle which the line from O to P makes with the positive x axis, then

$$\mathbf{v} = |\mathbf{v}| \cos \theta \mathbf{i} + |\mathbf{v}| \sin \theta \mathbf{j}. \tag{2}$$

Let w be the vector obtained by rotating the vector $\mathbf{v}$ through a right angle in the positive (counterclockwise) direction, so that (as in trigonometry) $\theta + \pi/2$ is one of the angles which the vector $\mathbf{w}$ makes with the positive x axis. Show that

$$\mathbf{w} = |\mathbf{v}| \cos\left(\theta + \frac{\pi}{2}\right)\mathbf{i} + |\mathbf{v}| \sin\left(\theta + \frac{\pi}{2}\right)\mathbf{j} \tag{3}$$

and hence that

$$\mathbf{w} = |\mathbf{v}|\,(-\sin \theta)\mathbf{i} + |\mathbf{v}|\,(\cos \theta)\mathbf{j} \tag{4}$$

and

$$\mathbf{w} = -y\mathbf{i} + x\mathbf{j}. \tag{5}$$

Remark: While our present interest lies in vectors, our result is equivalent to the fact that, whatever x and y may be, if we start at $P(x,y)$, the point P having coordinates x and y, and run in the positive direction along a quadrant of a circle having its center at the origin, we will stop at the point $Q(-y,x)$. This fact implies and is implied by the formulas

$$\cos\left(\theta + \frac{\pi}{2}\right) = -\sin \theta, \qquad \sin\left(\theta + \frac{\pi}{2}\right) = \cos \theta \tag{6}$$

which were used to obtain (4) from (3).

20 Sketch some figures and discover the circumstances under which two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ are such that $|\mathbf{u} + \mathbf{v}| = |\mathbf{u} - \mathbf{v}|$. Then prove that

$$|\mathbf{u} + \mathbf{v}|^2 = (\mathbf{u} + \mathbf{v})\cdot(\mathbf{u} + \mathbf{v}) = \mathbf{u}\cdot\mathbf{u} + 2\mathbf{u}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{v} \tag{1}$$

and

$$|\mathbf{u} + \mathbf{v}|^2 - |\mathbf{u} - \mathbf{v}|^2 = 4\mathbf{u}\cdot\mathbf{v}. \tag{2}$$

21 The *span* of the set of n vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is the set of vectors $\mathbf{v}$ representable in the form

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n,$$

where $c_1, c_2, \cdots, c_n$ are scalars. Show that the span of the set of three given vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is the same as the span of the set of three vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ defined by the system of equations

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{v}_1 + \mathbf{v}_2 \\ \mathbf{u}_2 &= \phantom{\mathbf{v}_1 + {}} \mathbf{v}_2 + \mathbf{v}_3 \\ \mathbf{u}_3 &= \mathbf{v}_1 \phantom{{} + \mathbf{v}_2} + \mathbf{v}_3. \end{aligned}$$

Hint: The proof consists of two parts. Suppose first that $\mathbf{w}$ belongs to the span of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ and seek an easy way to show that $\mathbf{w}$ must belong to the span of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. It remains to suppose that $\mathbf{w}$ belongs to the span of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ and then show that $\mathbf{w}$ must belong to the span of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$. As a start, solve the given system of equations for $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. One of the results is

$$\mathbf{v}_1 = \tfrac{1}{2}\mathbf{u}_1 - \tfrac{1}{2}\mathbf{u}_2 + \tfrac{1}{2}\mathbf{u}_3.$$

22 Using the definition of Problem 21, prove that if

$$(1) \qquad c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = 0,$$

where c_1, c_2, c_3 are scalars that are not all zero, then some one of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ belongs to the span of the other two. *Remark:* In this case the set of three vectors v_1, v_2, v_3 is said to be a *dependent* (or linearly dependent) set. In case (1) holds only when $c_1 = c_2 = c_3 = 0$, the three vectors are said to be *independent*. These concepts are very important in several branches of mathematics.

23 Perhaps we need a little experience drawing and adding vectors that all lie in the same plane. Start with a clean sheet of paper and draw unit vectors $\mathbf{u}$ and $\mathbf{v}$ headed, respectively, toward the right side and top of the page. Let P_0 be the point at the center of the page. More points in the sequence P_0, P_1, P_2, P_3, $\cdots$ are to be obtained in the following random way. Start with $k = 1$. Get two coins of different size and toss them so that each lands H (head) or T (tail).

If big coin is H and small coin is H, let $\overrightarrow{P_{k-1}P_k} = \mathbf{u}$.
If big coin is H and small coin is T, let $\overrightarrow{P_{k-1}P_k} = \mathbf{v}$.
If big coin is T and small coin is H, let $\overrightarrow{P_{k-1}P_k} = -\mathbf{u}$.
If big coin is T and small coin is T, let $\overrightarrow{P_{k-1}P_k} = -\mathbf{v}$.

Then draw $\overrightarrow{P_0P_1}$. With $k = 2$, repeat the coin tossing to locate P_2, and continue until P_{10} has been reached. It is not improper to become interested in the probability that all of the points $P_0, P_1, \cdots, P_{10}$ lie inside the circle with center at the origin and radius 5. This is a *random-walk problem* and such problems are of interest in the theory of diffusion. To prepare for investigation of these things, we must study analytic geometry, calculus, probability, and statistics.

24 Using one die (singular of dice, a cube with six numbered faces) instead of two coins, describe a procedure for obtaining paths for use in random-walk problems in E_3.

25 The problem here is to grasp the meanings of the following statements when n is 2 and 3 and perhaps even when n is a greater integer. When $P_1, P_2, \cdots, P_{n+1}$ are $n + 1$ points that lie in the same E_n but do not lie in an E_{n-1}, these points are the vertices of an n-dimensional simplex. A line segment which joins two of these points is an edge of the simplex, so the simplex has $n(n + 1)/2$ edges. To each vertex P_k there corresponds the opposite simplex of $n - 1$ dimensions having vertices at the remaining points. A median of a simplex is the line segment joining a vertex P_k to the centroid A_k of the opposite simplex. The $n + 1$ medians of the simplex all intersect at a point B, and for each k,

$$\overrightarrow{P_kB} = \frac{n}{n+1}\overrightarrow{P_kA_k}.$$

This point B is the centroid of the n-dimensional simplex and, when an origin O has been selected, the centroid B is determined by the formula

$$\overrightarrow{OB} = \frac{\overrightarrow{OP_1} + \overrightarrow{OP_2} + \overrightarrow{OP_3} + \cdots + \overrightarrow{OP_{n+1}}}{n+1}.$$

Remark: As the assertions may have suggested, simplexes of one, two, and three dimensions are, respectively, line segments, triangles, and tetrahedrons. When

n exceeds 3, the simplex does not have a plebeian name; it is an n-dimensional simplex.

2.2 *Coordinate systems and vectors in* E_3 To locate a point in a plane (Euclid space E_2 of two dimensions), it suffices to have a two-dimensional rectangular coordinate system involving the two mutually perpendicular x and y axes with which we are familiar. To locate a point in E_3 (Euclid space of three dimensions), it suffices to have a three-dimensional rectangular coordinate system involving the three mutually perpendicular x, y, and z axes of Figure 2.21. To partially overcome the difficulties involved in picturing three-dimensional objects on a flat piece of paper, we consider the y and z axes to be in the plane of the paper which, like a blackboard in a classroom, is vertical and consider the x axis to be perpendicular to the y and z axes and sticking out toward us. We can also consider the x and y axes to be wires on horizontal fences separating rectangular fields and consider the z axis to be a vertical post at their intersection.

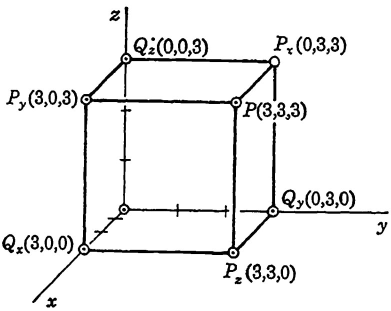


Figure 2.21

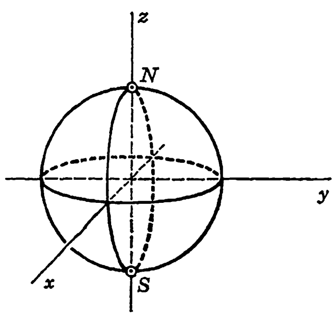


Figure 2.22

To locate the point $P(x,y,z)$ having nonnegative coordinates x, y, and z, we start at the origin, go x units forward (in the direction of the positive x axis), then go y units to the right (in the direction of the positive y axis) in the xy plane, and then go z units upward (in the direction of the positive z axis) to reach $P(x,y,z)$. If $x < 0$, we start by going $|x|$ units in the direction of the negative x axis. Similar rules apply when other coordinates are negative. Figure 2.21 shows the point $P(3,3,3)$ and, in addition, the projections P_x, P_y, P_z, Q_x, Q_y, and Q_z of this point on the three coordinate planes and coordinate axes. The figure is worth a little study. The eight encircled points lie at the vertices of a cube. Each of the edges is 3 units long, but in the flat figure the distance between two points 1 unit apart on the x axis is only a half or a third or a quarter of the distance between two such points on the y and z axes. Further information about the natures of figures involving rectangular coordinate systems in E_3 can be obtained by looking at Figure 2.22. This shows a sphere with center at the origin. The intersection (or section) of the sphere and the yz plane is a circle through the north and south poles which could be drawn with a compass. The intersection of the sphere and the xy plane is the equatorial circle which appears in the flat figure to be a flattened circle. The intersection of the sphere and the xz plane is a circle composed of two meridians passing through the poles. The three coordinate planes, the xy plane, the yz plane, and xz plane, cut E_3 into eight parts called *octants*. The octant containing points having only nonnegative coordinates is called the *first octant*, and most people neither know nor care whether the others are numbered.

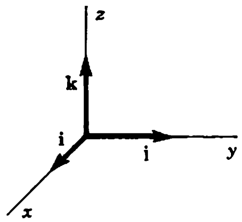


Figure 2.23

We can learn about coordinate systems and, at the same time, prepare ourselves to solve problems of many types in mathematics and other sciences by introducing vectors. As in Figure 2.23, let $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ denote unit vectors (vectors of length 1) in the directions of the positive x, y, and z axes. Since these vectors are *orthogonal* (which means that two different ones are orthogonal or perpendicular), *normalized* (which means that each one has unit length), and lie in E_3 (Euclid space of three dimensions), we say that they constitute an *orthonormal set* of vectors in E_3. The definition of scalar products given in (2.16) implies that

$$\mathbf{i}\cdot\mathbf{i} = 1, \qquad \mathbf{j}\cdot\mathbf{j} = 1, \qquad \mathbf{k}\cdot\mathbf{k} = 1 \tag{2.231}$$

and that $\mathbf{u}\cdot\mathbf{v} = 0$ when $\mathbf{u}$ and $\mathbf{v}$ are two different ones of the vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$. Similarly, the definition (2.17) of vector products implies that

$$\mathbf{i}\times\mathbf{i} = 0, \qquad \mathbf{j}\times\mathbf{j} = 0, \qquad \mathbf{k}\times\mathbf{k} = 0 \tag{2.232}$$

and that

$$\begin{cases} \mathbf{i}\times\mathbf{j} = \mathbf{k}, & \mathbf{j}\times\mathbf{k} = \mathbf{i}, \qquad \mathbf{k}\times\mathbf{i} = \mathbf{j} \\ \mathbf{j}\times\mathbf{i} = -\mathbf{k}, & \mathbf{k}\times\mathbf{j} = -\mathbf{i}, \qquad \mathbf{i}\times\mathbf{k} = -\mathbf{j}. \end{cases} \tag{2.233}$$

To help remember these formulas, we can notice that if we write the ordered set

$$\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{i}, \mathbf{j}, \mathbf{k} \tag{2.234}$$

of vectors, then the vector product of two consecutive ones in this order is the next but that changing the order of the factors changes the sign of the product. A rectangular coordinate system is said to be *right-handed* when the x, y, and z axes are so oriented (or arranged) that their orthonormal set $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ of vectors is such that the formulas (2.233) are correct; otherwise, the system is left-handed. We shall use only right-handed systems so that we can always use the formulas (2.233).

As in Figure 2.242, let $\overrightarrow{OP}$ be the vector running from the origin O to the point $P(x,y,z)$. The rules for multiplying vectors by scalars and for adding vectors imply that

$$\overrightarrow{OP} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}. \tag{2.24}$$

The three vectors $x\mathbf{i}$, $y\mathbf{j}$, and $z\mathbf{k}$ are the *vector components* of the vector $\overrightarrow{OP}$, and the three scalars x, y, and z are the *scalar components*.† We can start getting acquainted with scalar products by observing that the angle between a vector and itself is 0, so

$$\begin{aligned} |\overrightarrow{OP}|^2 &= |\overrightarrow{OP}|\,|\overrightarrow{OP}| \cos 0 \\ &= \overrightarrow{OP}\cdot\overrightarrow{OP} \\ &= (x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\cdot(x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) \\ &= x^2 + y^2 + z^2 \end{aligned}$$

and hence that

$$|\overrightarrow{OP}| = \sqrt{x^2 + y^2 + z^2}. \tag{2.241}$$

This important formula holds whether x, y, and z are positive or not. In case x, y, and z are all positive, we can give another proof of the formula by applying the Pythagoras theorem twice to the rectangular

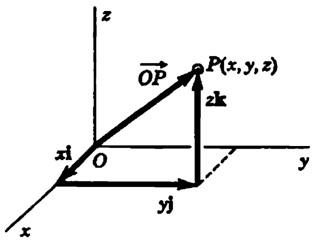


Figure 2.242

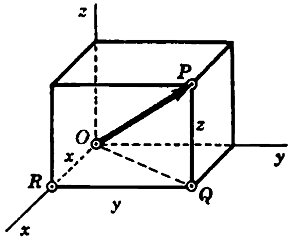


Figure 2.243

parallelepiped (or brick) of Figure 2.243. Because the angles OQP and ORQ are right angles, the Pythagoras theorem gives

$$\begin{aligned} |\overrightarrow{OP}|^2 &= |\overrightarrow{OQ}|^2 + |\overrightarrow{QP}|^2 \\ &= |\overrightarrow{OR}|^2 + |\overrightarrow{RQ}|^2 + |\overrightarrow{QP}|^2 \\ &= x^2 + y^2 + z^2 \end{aligned}$$

and (2.241) follows. The same methods give the *distance formula*

$$|\overrightarrow{P_1P_2}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{2.25}$$

† When physicists talk about the components of a vector, they often mean vector components. When mathematicians talk about components, they usually mean scalar components. Hence the unqualified term "components" is ambiguous. We will have quite a bonfire if we burn all the books that tell confusing tales about components and projections and directed distances.

for the length of the vector $\overrightarrow{P_1P_2}$, that is, for the distance between P_1 and P_2. This formula will be derived more carefully in the next section. A sphere is defined to be the set of points P in E_3 which lie at a fixed distance r (called the radius) from a fixed point P_0 (called the center). It follows from this definition and the distance formula that the equation

$$(2.26) \qquad (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$$

is the equation of the sphere with center at $P_0(x_0,y_0,z_0)$ and radius r.

A *paraboloid* should be something resembling a parabola, because the Greek suffix "oid" means "like" or "resembling." A paraboloid (or circular paraboloid) is defined to be the set of points $P(x,y,z)$ in E_3 equidistant from a fixed point F which is called the *focus* and a fixed plane π which is called the *directrix* and which does not contain the focus. In order to obtain the equation of a paraboloid in an attractive form, we let $1/2k$ denote the distance from F to π so that $1/2k = p$ and $k = 1/2p$, where p is the length of the perpendicular from F to π. Then we put the z axis through F perpendicular to π and put the origin midway between F and π as in Figure 2.27. The paraboloid is then the set of points $P(x,y,z)$ for which $|\overrightarrow{FP}| = |\overrightarrow{DP}|$, where D is the projection of P on the plane π and has coordinates $(x, y, -1/4k)$. The present situation is very similar to that in Section 1.4, where the equation of a parabola was worked out. A point $P(x,y,z)$ lies on the paraboloid if and only if $|\overrightarrow{FP}|^2 = |\overrightarrow{DP}|^2$ and hence, as use of the distance formula shows, if and only if

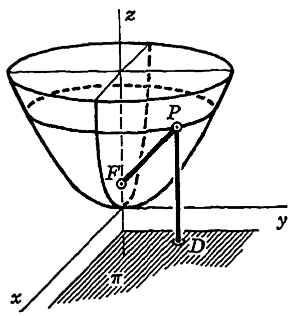


Figure 2.27

$$x^2 + y^2 + \left(z - \frac{1}{4k}\right)^2 = \left(z + \frac{1}{4k}\right)^2.$$

Simplifying this gives the more attractive equation

$$(2.28) \qquad z = k(x^2 + y^2).$$

This is the equation of the paraboloid shown in Figure 2.27.

Problems 2.29

1 Plot the points $A(1,1,0)$, $B(1,0,1)$, and $C(0,1,1)$.

2 In a new figure, repeat the construction of Problem 1 and insert the horizontal and vertical line segments running from A, B, and C to the coordinate axes.

3 In a new figure, repeat all of the construction of Problem 2. Then insert the point $D(1,1,1)$ and the line segments joining D to A, to B, and to C. ***Remark:***

The final figure should look much like Figure 2.21. It seems that we do not inherit abilities to do things like this neatly and correctly. A little practice is needed, and it often happens that the first figures we draw are very clumsy.

4 In the xy and xz planes, sketch circles of radius 3 having their centers at the origin. Then complete the sketch of the sphere of which these circles are great circles, that is, intersections of the sphere and planes through the center of the sphere.

5 A spherical ball of radius 3 has its center at the origin. Sketch the part of it that lies in the first octant.

6 Sketch a rectangular x, y, z coordinate system and observe that, in each case, the graph of the equation or system of equations on the left is the entity (point set) on the right:

$$\begin{array}{rl} x = 0 & yz \text{ plane} \\ y = 0 & xz \text{ plane} \\ z = 0 & xy \text{ plane} \\ x = y = 0 & z \text{ axis} \\ x = z = 0 & y \text{ axis} \\ y = z = 0 & x \text{ axis} \\ x = y = z & \text{line through } (0,0,0),\ (1,1,1) \end{array}$$

Remark: We make no effort to remember these facts, but whenever we see an x, y, z coordinate system, we should be able to observe and use these facts as they are needed.

7 There are many points $P(x,y,z)$ whose coordinates satisfy the equation $y = 3$. Some examples are $(0,3,0)$, $(0,3,1)$, $(1,3,0)$, $(1,3,1)$, and $(-40,3,416)$. Sketch a figure and become convinced that the graph of the equation $y = 3$ is the plane π which passes through the point $(0,3,0)$ and is both perpendicular to the y axis and parallel to the xz plane. Then, without so much attention to details, describe the graph of the equation $z = 2$.

8 Plot the points $(0,1,0)$ and $(0,0,1)$ and then draw the line L through these points. Show that if $P(x,y,z)$ lies on L, then $x = 0$ and $y + z = 1$. Show also that if $x = 0$ and $y + z = 1$, then $P(x,y,z)$ lies on L. *Remark:* It is possible to write a single equation equivalent to the system $x = 0$, $y + z = 1$. For example, each of the equations

$$|x| + |y + z - 1| = 0$$
$$x^2 + (y + z - 1)^2 = 0$$

does the trick. It is fashionable to keep the two equations, and one who wishes to do so may learn something by thinking about the matter.

9 Put the equation $x^2 + y^2 + z^2 - 2x - 4y + 8z = 0$ into the standard form (2.26) of the equation of a sphere and find the center and radius of the sphere. *Hint:* Complete squares. Check your result by observing that the coordinates of the origin satisfy the given equation and hence that the distance from the origin to the center of the sphere must be the radius of the sphere.

10 A set S consists of those points P in E_3 for which $|\overrightarrow{AP}|^2 + |\overrightarrow{BP}|^2 = 16$, where A is the origin and B is the point $(0,2,0)$. Show that S is the sphere of radius $\sqrt{7}$ having its center at the point $(0,1,0)$. Sketch the coordinate system and S.

11 Make a sketch showing an x, y, z coordinate system, the sphere S having the equation $x^2 + y^2 + z^2 = 9$ and the line L having the equations $x = 2$, $y = 2$. Find the length of the part of L that lies inside S. *Hint:* Do not depend upon your figure to obtain precise quantitative information. Find and use the coordinates of the points on the sphere for which $x = 2$ and $y = 2$. *Ans.:* 2.

12 By use of the distance formula, show that the equation of the set of points $P(x,y,z)$ equidistant from two given points $P_1(x_1,y_1,z_1)$ and $P_2(x_2,y_2,z_2)$ can be put in the form

$$(1)\quad (x_2 - x_1)\left(x - \frac{x_1 + x_2}{2}\right) + (y_2 - y_1)\left(y - \frac{y_1 + y_2}{2}\right) + (z_2 - z_1)\left(z - \frac{z_1 + z_2}{2}\right) = 0.$$

Remark: Our official introduction to planes in E_3 will come in Section 2.4. Meanwhile, we can observe that if P_1 and P_2 are distinct points, so that $x_2 \neq x_1$ or $y_2 \neq y_1$ or $z_2 \neq z_1$, then the set mentioned above is a plane and the equation which we have found is its equation. The equation has the form

$$(2)\qquad A(x - x_0) + B(y - y_0) + C(z - z_0) = 0,$$

where A, B, C are constants not all 0 and (x_0,y_0,z_0) is a point in (or on) the plane.

13 Supposing that A, B, C, x_0, y_0, z_0 are constants for which A, B, C are not all 0, show that the equation

$$(1)\qquad A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$$

is the equation of a plane. *Hint:* Taking cognizance of Problem 12, solve the equations

$$(2)\qquad x_2 - x_1 = A, \qquad y_2 - y_1 = B, \qquad z_2 - z_1 = C$$

$$(3)\qquad \frac{x_1 + x_2}{2} = x_0, \qquad \frac{y_1 + y_2}{2} = y_0, \qquad \frac{z_1 + z_2}{2} = z_0$$

to obtain two distinct points P_1 and P_2 such that the graph of (1) is the set of points equidistant from P_1 and P_2.

14 The base of a regular tetrahedron has its center at the origin and has vertices at the points $(2a,0,0)$, $(-a,\sqrt{3}\,a,0)$, $(-a,-\sqrt{3}\,a,0)$. The other vertex is on the positive z axis. Find the coordinates of this other vertex. Check the result by using the distance formula to determine whether the edges have equal lengths. Finally, sketch the tetrahedron.

15 Determine whether it is possible to multiply all of the coordinates of the points $(2a,0,0)$, $(-a, \sqrt{3}\,a,0)$, $(-a,-\sqrt{3}\,a,0)$, $(0,0,\sqrt{8}\,a)$ by the same constant λ to obtain new points that are vertices of a regular tetrahedron each edge of which has length a.

16 A set C of points in E_3 is called a *cone* with *vertex* V if whenever it contains a point P_0 different from V it also contains the whole line through V and P_0. Each of these lines is called a generator of the cone, the ancient idea being that if it moves in an appropriate way, it will "generate" the cone. A cone is called the circular cone whose vertex is V, whose axis is the line L, and whose central angle is α if V is on L and the cone consists of the points on those lines through V

which make the angle α with L. Supposing that $0 < \alpha < \pi/2$, sketch the circular cone whose vertex is the origin, whose axis is the z axis, and whose central angle is α. Then show that the equation of this cone is $z^2 = k^2(x^2 + y^2)$, where $k = \cot \alpha$.

17 Sketch a figure, similar to Figure 2.27, in which the paraboloid opens to the right along the y axis instead of upward along the z axis. Note that interchanging y and z in (2.28) gives the equation $y = k(x^2 + z^2)$ of the new paraboloid.

18 Sketch a figure, similar to Figure 2.27, in which the paraboloid opens forward along the x axis instead of upward along the z axis. Note that interchanging x and z in (2.28) gives the equation $x = k(y^2 + z^2)$ of the new paraboloid.

19 Plot the eight points $(\pm 2, \pm 2, \pm 2)$ obtained by taking all possible choices of the plus and minus signs. Then connect these points by line segments to obtain the edges of the cube of which the eight points are vertices. *Remark:* One who finds this problem to be unexpectedly difficult need not be disturbed. The problem is unexpectedly difficult.

20 We embark on a little excursion to learn more about our abilities to sketch graphs. The graph in E_2 of the equation $xy = 1$ does not intersect the coordinate axes, and it consists of two parts (or branches) that are easily sketched. The graph in E_3 of the equation $xyz = 1$ consists of those points in E_3 having coordinates (x,y,z) for which $xyz = 1$. The graph does not intersect the coordinate planes, and it consists of four parts, namely, the one containing some points for which $x > 0$, $y > 0$, $z > 0$, the one containing some points for which $x > 0$, $y < 0$, $z < 0$, the one containing some points for which $x < 0$, $y > 0$, $z < 0$, and the one containing some points for which $x < 0$, $y < 0$, $z > 0$. Everyone should discover for himself that it is surprisingly difficult (or hopelessly impossible) to draw x, y, z axes on a flat sheet of paper and sketch a figure which shows what these four parts look like and how they are situated relative to each other and to the coordinate system.

21 Draw the rectangular coordinate system obtained from that in Figure 2.23 by interchanging the x and y axes and the $\mathbf{i}$ and $\mathbf{j}$ vectors. Work out the formulas for the vector products of these vectors and show that the system is left-handed. As a safety measure, make a note on your figure that it is left-handed and be sure that your formulas for vector products are *not* remembered.

22 It is not necessarily true that our study of mathematical machinery is made more difficult when we pause briefly to look at a rather complicated application of it. Figure 2.291 shows a circle C in the yz plane which has its center at the point $(0,b,0)$ and has radius a. We suppose that $0 < a < b$. The surface T obtained by rotating this circle C about the z axis is called a *torus*. Thus a torus is the surface of a ring or hoop that is more or less closely approximated by an automobile tire. As the figure indicates, each point P on the torus T lies on the circle which (i) contains a point P' on C, (ii) lies in a plane parallel to the xy plane, and (iii) has its center at a point Q on the z axis. When the angles θ and

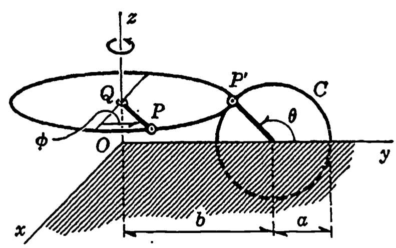


Figure 2.291

ϕ are determined by P' and P as in the figure, we see that

(1) $$|\overrightarrow{QP}| = |\overrightarrow{QP'}| = b + a\cos\theta$$

and that

(2) $$\overrightarrow{QP} = |\overrightarrow{QP}|\,(\cos\phi\mathbf{i} + \sin\phi\mathbf{j})$$

so

(3) $$\overrightarrow{QP} = (b + a\cos\theta)\cos\phi\mathbf{i} + (b + a\cos\theta)\sin\phi\mathbf{j}.$$

Letting $\mathbf{r}$ be the vector running from the origin to the point P on T, we see that

(4) $$\mathbf{r} = \overrightarrow{QP} + \overrightarrow{OQ} = \overrightarrow{QP} + a\sin\theta\mathbf{k}$$

and hence that

(5) $$\mathbf{r} = (b + a\cos\theta)\cos\phi\mathbf{i} + (b + a\cos\theta)\sin\phi\mathbf{j} + a\sin\theta\mathbf{k}.$$

Thus a vector $\mathbf{r}$ *having its tail at the origin has its tip on the torus* T *if and only if there exist angles* ϕ *and* θ *for which* (5) *holds.* Thus (5) is a vector equation of the torus. This implies that $P(x,y,z)$ lies on the torus T if and only if there exist angles ϕ and θ such that

(6) $$x = (b + a\cos\theta)\cos\phi, \qquad y = (b + a\cos\theta)\sin\phi, \qquad z = a\sin\theta.$$

These are parametric equations of the torus, the parameters being ϕ and θ. The torus is the graph of the parametric equations.

23 Show that the x, y, z equation of the torus of the preceding problem can be put in one or the other of the two equivalent forms

(1) $$(a^2 + b^2 - x^2 - y^2 - z^2)^2 = 4b^2(a^2 - z^2)$$

(2) $$(b^2 - a^2 + x^2 + y^2 + z^2)^2 = 4b^2(x^2 + y^2).$$

Hint: One way to start is to square and add the first two of the equations (6) of the preceding problem. *Remark:* In case $0 \leqq b \leqq a$, the equations of this and the preceding problem are not equations of a torus but they are equations of a surface.

24 Sketch the visible edges of the solid that remains when a cube having edges of length $a/2$ is removed from the upper right front corner of a cube having edges of length a. *Remark:* This figure can easily become a favorite doodle, and perfectly normal people can become very much interested in it.

25 Mr. T., a topologist, claims that a given right-handed rectangular coordinate system in E_3 can, when considered to consist of three stiff wires rigidly welded together at their origin, be moved around in E_3 in such a way that it will coincide with any other such system but cannot be similarly moved into coincidence with a left-handed system. It is not always easy to understand such assertions and to give proofs of them. It is sometimes easy to wave some arms around and give some more or less convincing arguments that could not be called proofs. The difficulty here is that those who wave arms and produce more or less convincing arguments sometimes reach erroneous conclusions. Anyone who wishes to do so may think about this matter.

2.3 *Scalar products, direction cosines, and lines in E_3* Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in E_3 having scalar components u_1, u_2, u_3 and v_1, v_2, v_3 so that

$$\begin{cases} \mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k} \\ \mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}. \end{cases} \tag{2.31}$$

It may be helpful to look at Figure 2.311, which shows vectors $\mathbf{u}$ and $\mathbf{v}$ and the projections of their tips on the xy plane.

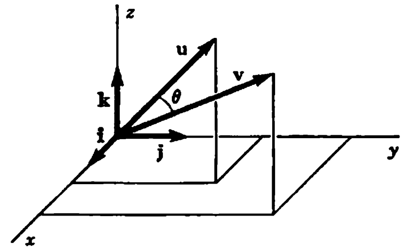


Figure 2.311

Since the scalar product $\mathbf{u}\cdot\mathbf{v}$ is defined by the formula

$$\mathbf{u}\cdot\mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \cos\theta, \tag{2.312}$$

it could be supposed that we should find $\cos\theta$ in order to find $\mathbf{u}\cdot\mathbf{v}$. It turns out, however, that there is a very simple and useful formula for $\mathbf{u}\cdot\mathbf{v}$, and we can use this formula to find $\cos\theta$ whenever we want it because we know that

$$|\mathbf{u}| = \sqrt{u_1^2 + u_2^2 + u_3^2}, \qquad |\mathbf{v}| = \sqrt{v_1^2 + v_2^2 + v_3^2}. \tag{2.313}$$

We find that

$$\begin{aligned} \mathbf{u}\cdot\mathbf{v} &= (u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k})\cdot(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \\ &= u_1\mathbf{i}\cdot(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) + u_2\mathbf{j}\cdot(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) + u_3\mathbf{k}\cdot(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}). \end{aligned}$$

With the aid of (2.231), we see that this reduces to the very important formula

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3. \tag{2.32}$$

In order to help remember this formula, we can remember that *the scalar product of two vectors is the sum of the products of their scalar components.* Use of (2.32) and (2.312) gives, when $|\mathbf{u}|\,|\mathbf{v}| \neq 0$, the formula

$$\cos\theta = \frac{u_1v_1 + u_2v_2 + u_3v_3}{|\mathbf{u}|\,|\mathbf{v}|} \tag{2.321}$$

which gives the angle θ between two vectors in terms of the scalar components of the vectors. This formula may be remembered. It is, in the long run, more sensible to concentrate upon the two formulas

$$\begin{cases} \mathbf{u}\cdot\mathbf{v} = |\mathbf{u}|\,|\mathbf{v}|\cos\theta \\ \mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3 \end{cases} \tag{2.33}$$

and to realize that the formula for $\cos\theta$ can be obtained very quickly by equating the two formulas for $\mathbf{u}\cdot\mathbf{v}$.

When, as above,

$$\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$$

and $|\mathbf{v}| > 0$, the vector in the parentheses in the right member of the formula

$$\mathbf{v} = |\mathbf{v}|\left(\frac{v_1}{|\mathbf{v}|}\mathbf{i} + \frac{v_2}{|\mathbf{v}|}\mathbf{j} + \frac{v_3}{|\mathbf{v}|}\mathbf{k}\right) \tag{2.34}$$

is the unit vector in the direction of $\mathbf{v}$. It is possible, and sometimes thought to be useful, to recognize that the scalar components of the unit vector are the cosines of the angles α, β, γ which the vector $\mathbf{v}$ makes with the unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. This is true because the formulas

$$\begin{array}{lll} \mathbf{v}\cdot\mathbf{i} = |\mathbf{v}|\cos\alpha, & \mathbf{v}\cdot\mathbf{j} = |\mathbf{v}|\cos\beta, & \mathbf{v}\cdot\mathbf{k} = |\mathbf{v}|\cos\gamma \\ \mathbf{v}\cdot\mathbf{i} = v_1 & \mathbf{v}\cdot\mathbf{j} = v_2 & \mathbf{v}\cdot\mathbf{k} = v_3 \end{array}$$

imply that

$$\frac{v_1}{|\mathbf{v}|} = \cos\alpha, \qquad \frac{v_2}{|\mathbf{v}|} = \cos\beta, \qquad \frac{v_3}{|\mathbf{v}|} = \cos\gamma. \tag{2.341}$$

The angles α, β, γ, shown in Figure 2.342, are called the *direction angles* of the vector $\mathbf{v}$. The cosines of these angles are called the *direction cosines* of the vector. Even those who do not like to prove formulas by use of special figures in E_3 should look at Figure 2.342 and see that the formulas

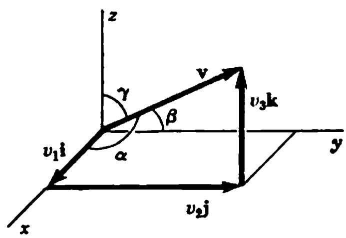


Figure 2.342

(2.341) can be proved by use of the formula that defines cosines in terms of coordinates and distances. Squaring and adding the members of the equations in (2.341) gives the formula

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1 \tag{2.343}$$

which provides an interesting way of making the more prosaic statement that the sum of the squares of the numerical components of a unit vector is 1. It is possible to put the formula (2.321) in the form

$$\text{(2.344)} \qquad \cos\theta = \cos\alpha_1 \cos\alpha_2 + \cos\beta_1 \cos\beta_2 + \cos\gamma_1 \cos\gamma_2,$$

where θ is the angle between two vectors making angles α_1, β_1, γ_1 and α_2, β_2, γ_2 with the unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. The formulas (2.343) and (2.344) are impressive formulas, but persons who know about vectors may hold the view that the formulas are antiquated and may concentrate their attention upon the formulas (2.33).

The following numerical example shows an application of the ideas of this paragraph. The vector in the left member of the formula

$$2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k} = \sqrt{29}\left(\frac{2}{\sqrt{29}}\mathbf{i} - \frac{3}{\sqrt{29}}\mathbf{j} + \frac{4}{\sqrt{29}}\mathbf{k}\right)$$

is the vector running from the origin O to the point $P(2,-3,4)$. The length of the vector is the square root of the sum of the squares of its numerical components and hence is $\sqrt{29}$. The vector in parentheses in the right member is a unit vector in the direction of $\overrightarrow{OP}$. The fact that its scalar components are cosines of direction angles is very often of no importance.

It is easy to adapt these ideas to obtain information about the vector $\overrightarrow{P_1P_2}$ running from $P_1(x_1,y_1,z_1)$ to $P_2(x_2,y_2,z_2)$ as in Figure 2.36. Starting with the formulas

$$\text{(2.35)} \qquad \overrightarrow{OP_1} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}, \qquad \overrightarrow{OP_2} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k},$$

it is easy to see that the rules for addition and subtraction of vectors give the formulas

$$\begin{aligned} \overrightarrow{OP_2} + \overrightarrow{OP_1} &= (x_2 + x_1)\mathbf{i} + (y_2 + y_1)\mathbf{j} + (z_2 + z_1)\mathbf{k} \\ \overrightarrow{OP_2} - \overrightarrow{OP_1} &= (x_2 - x_1)\mathbf{i} + (y_2 - y_1)\mathbf{j} + (z_2 - z_1)\mathbf{k}. \end{aligned}$$

But $\overrightarrow{OP_2} - \overrightarrow{OP_1} = \overrightarrow{P_1P_2}$, and hence

$$\text{(2.351)} \qquad \overrightarrow{P_1P_2} = (x_2 - x_1)\mathbf{i} + (y_2 - y_1)\mathbf{j} + (z_2 - z_1)\mathbf{k}.$$

Therefore,

$$\text{(2.352)} \qquad |\overrightarrow{P_1P_2}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

If to simplify writing we let d denote this distance between P_1 and P_2, then we can put (2.351) in the form

$$\text{(2.353)} \qquad \overrightarrow{P_1P_2} = d\left(\frac{x_2 - x_1}{d}\mathbf{i} + \frac{y_2 - y_1}{d}\mathbf{j} + \frac{z_2 - z_1}{d}\mathbf{k}\right).$$

The vector in parentheses is then the unit vector in the direction of $\overrightarrow{P_1P_2}$, and its scalar components are the direction cosines of $\overrightarrow{P_1P_2}$.

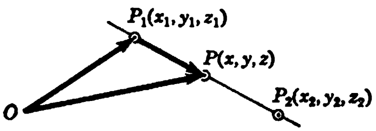


Figure 2.36

Before introducing coordinate systems, we called attention to the fact that a point P lies on the line P_1P_2 if and only if there is a scalar λ such that $\overrightarrow{P_1P} = \lambda\overrightarrow{P_1P_2}$ and

$$\overrightarrow{OP} = \overrightarrow{OP_1} + \lambda(\overrightarrow{OP_2} - \overrightarrow{OP_1}).$$

When the points have coordinates as in Figure 2.36, this equation can be put in the forms

$$(2.361)\quad (x - x_1)\mathbf{i} + (y - y_1)\mathbf{j} + (z - z_1)\mathbf{k} = \lambda(x_2 - x_1)\mathbf{i} + \lambda(y_2 - y_1)\mathbf{j} + \lambda(z_2 - z_1)\mathbf{k}$$

and

$$(2.362)\quad x - x_1 = \lambda(x_2 - x_1), \qquad y - y_1 = \lambda(y_2 - y_1), \qquad z - z_1 = \lambda(z_2 - z_1).$$

In case $x_2 \neq x_1$, $y_2 \neq y_1$, and $z_2 \neq z_1$, these equations hold for some λ if and only if

$$(2.37)\qquad \frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}.$$

In case $x_2 = x_1$, the condition on x is to be replaced by the condition $x = x_1$, and similar remarks apply to y_2 and z_2. With this understanding *the equations* (2.37) *are equations of the line* P_1P_2, that is, equations that are satisfied by x, y, z when and only when $P(x,y,z)$ lies on the line. The numbers $x_2 - x_1$, $y_2 - y_1$, $z_2 - z_1$ are the numerical components of a vector lying on the line P_1P_2, and we know how to find the direction cosines of this vector.

It can be claimed that the equations

$$(2.38)\qquad \frac{x - x_1}{a} = \frac{y - y_1}{b} = \frac{z - z_1}{c}$$

do not look like the equations (2.37) of a line, but we can put these equations in the form

$$\frac{x - x_1}{(x_1 + a) - x_1} = \frac{y - y_1}{(y_1 + b) - y_1} = \frac{z - z_1}{(z_1 + c) - z_1}$$

which does have the form (2.37). Thus the equations (2.38) are in fact equations of the line L which passes through the points $P_1(x_1,y_1,z_1)$ and $P_2(x_1 + a,\ y_1 + b,\ z_1 + c)$. The numbers a, b, c are the scalar components of the vector $\overrightarrow{P_1P_2}$ on L, and they determine the direction cosines of $\overrightarrow{P_1P_2}$ in the usual way. The equations (2.37) and (2.38) are called *point-direction* equations of lines because they reveal coordinates of points and scalar components of vectors lying on the lines.

Problems 2.39

1 Write the intrinsic (not depending upon coordinate system) formula for the scalar product of two vectors $\mathbf{u}$ and $\mathbf{v}$ and be prepared to rewrite it at any time.

2 Write the coordinate-dependent formula for the scalar product of the two vectors

$$\mathbf{u} = a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}, \qquad \mathbf{v} = a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}$$

and be prepared to rewrite it at any time.

3 Use the results of the first two problems to obtain a formula for the cosine of the angle between the two vectors in Problem 2 and be prepared to repeat the process at any time.

4 Find the scalar product of the two given vectors and use it to find $\cos\theta$, the cosine of the angle between the vectors:

(*a*) $\mathbf{u} = 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$, $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ *Ans.:* $\cos\theta = \frac{11}{29}$

(*b*) $\mathbf{u} = 2\mathbf{i} - 3\mathbf{j} - 4\mathbf{k}$, $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ *Ans.:* $\cos\theta = \dfrac{-21}{29}$

(*c*) $\mathbf{u} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$, $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ *Ans.:* $\cos\theta = 1$

(*d*) $\mathbf{u} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$, $\mathbf{v} = -2\mathbf{i} - 3\mathbf{j} - 4\mathbf{k}$ *Ans.:* $\cos\theta = -1$

(*e*) $\mathbf{u} = 2\mathbf{i} + 3\mathbf{j}$, $\mathbf{v} = 3\mathbf{i} + 4\mathbf{j}$ *Ans.:* $\cos\theta = \dfrac{18}{\sqrt{325}} = \sqrt{\dfrac{324}{325}}$

5 Determine c so that the two given vectors will be orthogonal (or perpendicular).

(*a*) $\mathbf{u} = 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$, $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + c\mathbf{k}$ *Ans.:* $c = \frac{5}{4}$

(*b*) $\mathbf{u} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{v} = \mathbf{i} + \mathbf{j} + c\mathbf{k}$ *Ans.:* $c = -2$

6 For each vector $\mathbf{v}$, find the unit vector $\mathbf{v}_1$ in the direction of $\mathbf{v}$.

(*a*) $\mathbf{v} = 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$ *Ans.:* $\mathbf{v}_1 = \dfrac{2}{\sqrt{29}}\mathbf{i} - \dfrac{3}{\sqrt{29}}\mathbf{j} + \dfrac{4}{\sqrt{29}}\mathbf{k}$

(*b*) $\mathbf{v} = 7\mathbf{i}$ *Ans.:* $\mathbf{v}_1 = \mathbf{i}$

(*c*) $\mathbf{v} = \mathbf{i} + \mathbf{j}$ *Ans.:* $\mathbf{v}_1 = \dfrac{1}{\sqrt{2}}\mathbf{i} + \dfrac{1}{\sqrt{2}}\mathbf{j}$

7 Supposing that $\mathbf{v}$ and $\mathbf{w}$ are orthonormal vectors and that $\mathbf{u} = a\mathbf{v} + b\mathbf{w}$, where a and b are not both zero, find the angle between $\mathbf{u}$ and $\mathbf{w}$. *Hint:* Use the basic formulas

$$\mathbf{u}\cdot\mathbf{w} = |\mathbf{u}|\,|\mathbf{w}|\cos\theta, \qquad \mathbf{u}\cdot\mathbf{w} = (a\mathbf{v} + b\mathbf{w})\cdot\mathbf{w}.$$

Ans.: $\cos\theta = b/\sqrt{a^2 + b^2}$.

8 With the text of this section out of sight, sketch a figure showing points $(0,0,0)$, $P_1(x_1,y_1,z_1)$, and $P_2(x_2,y_2,z_2)$. Starting with the assumption that $P(x,y,z)$

lies on the line P_1P_2 and hence that there is a number λ for which $\overrightarrow{P_1P} = \lambda\overrightarrow{P_1P_2}$, show how vectors are used to obtain the coordinate equation

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1} = \lambda.$$

Refer to Figure 2.36 and the text only if assistance is needed.

9 Show that the equations

$$(1) \quad x - x_1 = \lambda(x_2 - x_1), \qquad y - y_1 = \lambda(y_2 - y_1), \qquad z - z_1 = \lambda(z_2 - z_1),$$

from which the text of this section obtained the equations

$$(2) \qquad \frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1},$$

can be put in the form

$$(3) \quad x = (x_2 - x_1)\lambda + x_1, \qquad y = (y_2 - y_1)\lambda + y_1, \qquad z = (z_2 - z_1)\lambda + z_1.$$

Then show that the equations

$$(4) \qquad x = a_1\lambda + b_1, \qquad y = a_2\lambda + b_2, \qquad z = a_3\lambda + b_3$$

can be put in the form

$$(5) \qquad \frac{x - b_1}{a_1} = \frac{y - b_2}{a_2} = \frac{z - b_3}{a_3} = \lambda.$$

Remark: The equations in (1), (3), and (4) are called *parametric equations* of lines; different values of the parameter λ yield different points on the lines. It is fashionable to use the letter λ for a parameter because people who work with Lagrange (French mathematician) parameters (or multipliers) get the habit; the letter t is used when time is involved and sometimes when time is not involved.

10 Look at the equations

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

of the line L which passes through the points $P_1(x_1,y_1,z_1)$ and $P_2(x_2,y_2,z_2)$ and tell why these equations are satisfied when $x = x_1$, $y = y_1$, $z = z_1$ and when $x = x_2$, $y = y_2$, $z = z_2$. With this is mind tell how we can quickly find the coordinates of two points on the line having equations

$$\frac{x - 2}{1 - 2} = \frac{y + 3}{2 + 3} = \frac{z + 1}{3 + 1}.$$

11 Find equations of the line L which passes through the points $P_1(2,-3,-1)$ and $P_2(1,2,3)$ and then, after putting $z = 0$, find the coordinates of the point $(x,y,0)$ in which L intersects the xy plane. *Ans.:* $x = \frac{7}{4}$, $y = -\frac{7}{4}$, $z = 0$.

12 This problem involves a little introductory information that everyone should have. The right-handed coordinate system shown in Figure 2.391 is the one we ordinarily use. When we are wholly or primarily interested in vectors lying in the same plane, it is sometimes convenient to think of this plane as being the plane of the paper upon which we print or write and to use the right-handed coordinate system shown in Figure 2.392. Vectors of the form $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ for which $z = 0$ then lie in the plane of the paper and figures showing them are

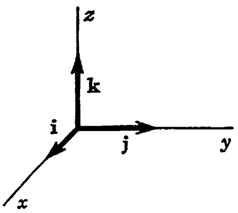


Figure 2.391

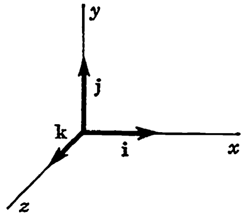


Figure 2.392

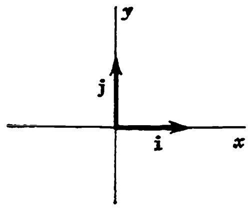


Figure 2.393

undistorted. When we are interested only in vectors lying in one plane, we may leave the z axis out of the figure and use Figure 2.393. The introduction is finished, and we come to our problem. Use the method, involving slopes, of Section 1.3 to show that $\tan\theta = \frac{1}{18}$ when θ is the angle between the vectors

$$\mathbf{u} = 2\mathbf{i} + 3\mathbf{j}, \qquad \mathbf{v} = 3\mathbf{i} + 4\mathbf{j}$$

running from the origin to the points (2,3) and (3,4). Use the method of this section to show (or to show again) that

$$\cos\theta = \frac{18}{\sqrt{325}} = \sqrt{\frac{324}{325}}.$$

Then construct and use a modest but appropriate figure to show that the two results agree. To conclude with another story, we can remark that the method involving slopes may sometimes be preferred because it often gives answers without radicals when easy problems in E_2 are solved. However, the scalar-product method is the more powerful method which works in E_2 and in E_3 and, as some people learn, in E_n when $n > 3$.

13 A vector $\mathbf{v}$ makes equal acute angles δ with the three positive coordinate axes. Find δ (to find $\cos\delta$ is enough) (i) by use of an identity involving direction cosines and (ii) by using the edges and the diagonals of a cube having opposite vertices at (0,0,0) and (1,1,1). Make everything check.

14 Referring to Figure 2.21, find the angle θ between the two vectors running from the origin to the mid-points of the top edges of the cube that pass through $P(3,3,3)$. *Ans.:* $\cos\theta = \frac{8}{9}$.

15 Let $\mathbf{v}$ be a nonzero vector and $\mathbf{w}$ a unit vector having their tails at the same point O. Show with the aid of a figure that the vector $(\mathbf{v}\cdot\mathbf{w})\mathbf{w}$ is the vector component of $\mathbf{v}$ in the direction of $\mathbf{w}$, and that the vector $\mathbf{v} - (\mathbf{v}\cdot\mathbf{w})\mathbf{w}$ is the vector component of $\mathbf{v}$ orthogonal (or perpendicular) to $\mathbf{w}$. *Remark:* This problem appeared among Problems 2.19 with different notation and additional information.

16 When $\mathbf{u} \neq 0$, each vector $\mathbf{v}$ is representable as the sum of a vector component $c\mathbf{u}$ and a vector component $\mathbf{q}$ orthogonal to $\mathbf{u}$. Find $\mathbf{q}$ and find a way of checking the answer when

(*a*) $\mathbf{u} = 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$, $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$
(*b*) $\mathbf{u} = \mathbf{i} + \mathbf{j}$, $\quad\mathbf{v} = \mathbf{i}$
(*c*) $\mathbf{u} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, $\quad\mathbf{v} = \mathbf{i}$

17 Show that the two vectors

$$\begin{aligned} \mathbf{u}_1 &= \cos\theta\mathbf{i} + \sin\theta\mathbf{j} \\ \mathbf{u}_2 &= -\sin\theta\mathbf{i} + \cos\theta\mathbf{j} \end{aligned}$$

constitute an orthonormal set, that is,

$$|\mathbf{u}_1| = 1, \qquad |\mathbf{u}_2| = 1, \qquad \text{and} \qquad \mathbf{u}_1\cdot\mathbf{u}_2 = 0.$$

18 Show that the three vectors

$$\begin{aligned} \mathbf{u}_1 &= \cos\phi\sin\theta\mathbf{i} + \sin\phi\sin\theta\mathbf{j} + \cos\theta\mathbf{k} \\ \mathbf{u}_2 &= \cos\phi\cos\theta\mathbf{i} + \sin\phi\cos\theta\mathbf{j} - \sin\theta\mathbf{k} \\ \mathbf{u}_3 &= -\sin\phi\mathbf{i} + \cos\phi\mathbf{j} \end{aligned}$$

constitute an orthonormal set, that is,

$$\mathbf{u}_p\cdot\mathbf{u}_q \text{ is } 1 \text{ when } p = q \text{ and is } 0 \text{ when } p \neq q.$$

19 This problem requires that we learn the procedure by which we obtain a useful formula for the gravitational force $\mathbf{F}$ which is exerted upon a particle P_1^* having mass m_1 and situated at the point $P_1(x_1,y_1,z_1)$ by another particle P_2^* having mass m_2 and situated at the point $P_2(x_2,y_2,z_2)$. The very modest Figure 2.394 can help us understand what we do. We start with the *Newton law of universal gravitation*, which is an "intrinsic law" that does not depend upon coordinate systems. This law says that there is a "universal constant" G, which depends only upon the units used to measure force, distance, and mass, such that a particle P_1^* of mass m_1 at P_1 is attracted toward a second particle P_2^* of mass m_2 at P_2 by a force $\mathbf{F}$ having magnitude

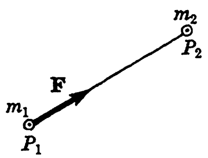


Figure 2.394

$$|\mathbf{F}| = G\frac{m_1m_2}{d^2}, \tag{1}$$

where d is the distance between the points. Our next step is to put (1) in the form

$$|\mathbf{F}| = G\frac{m_1m_2}{|\overrightarrow{P_1P_2}|^2}. \tag{2}$$

According to the Newton law, the direction of $\mathbf{F}$ is the direction of the vector $\overrightarrow{P_1P_2}$ (not $\overrightarrow{P_2P_1}$). We have learned (and can relearn if the fact has been forgotten) that each nonzero vector $\mathbf{F}$ is the product of $|\mathbf{F}|$, the length or magnitude of $\mathbf{F}$, and the unit vector $\mathbf{F}/|\mathbf{F}|$ in the direction of $\mathbf{F}$. When $\mathbf{F}$ has the direction of $\overrightarrow{P_1P_2}$, this unit vector is $\overrightarrow{P_1P_2}/|\overrightarrow{P_1P_2}|$. Therefore,

$$\mathbf{F} = G\frac{m_1m_2}{|\overrightarrow{P_1P_2}|^2}\frac{\overrightarrow{P_1P_2}}{|\overrightarrow{P_1P_2}|} = Gm_1m_2\frac{\overrightarrow{P_1P_2}}{|\overrightarrow{P_1P_2}|^3}. \tag{3}$$

Thus when we know about vectors, we can put the Newton law in the following much more useful form. There is a constant G such that a particle P_1^* of mass m_1 at P_1 is attracted toward a particle P_2^* of mass m_2 at P_2 by the force $\mathbf{F}$ defined by (3). The problem which we originally proposed involved rectangular coordinates, and it should now be completely obvious that the answer to our problem is

$$\mathbf{F} = Gm_1m_2\frac{(x_2 - x_1)\mathbf{i} + (y_2 - y_1)\mathbf{j} + (z_2 - z_1)\mathbf{k}}{[(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{3/2}}. \tag{4}$$

Remarks: Persons who work with these things normally recognize the fact that we do not, in our physical world, have "particles" concentrated at points. It is sometimes possible, however, to obtain useful results from calculations based on the assumption that particles are concentrated at points. When this assumption has been made, we can complicate ideas and simplify language by replacing the concept of "a particle P_1^* at the point P_1 having mass m_1" by the concept of "a point P_1 having mass m_1." It can be insisted that this linguistic antic should be explained; otherwise, a serious student of mathematics is entitled to ask where the postulates of Euclid provide for the possibility that some points can be heavier than others. Finally, it can be insisted that the formulas (3) and (4) should *not* be remembered; it is better to know (2) and to understand the very simple process by which we use it to obtain the more useful formulas (3) and (4) whenever we want them.

20 Let $P_1(x_1,y_1,z_1)$ and $P_2(x_2,y_2,z_2)$ be two distinct points in E_3 and recall (or prove again) that to each point $P(x,y,z)$ on the line L containing P_1 and P_2 there corresponds a number λ for which

$$x = x_1 + \lambda(x_2 - x_1), \qquad y = y_2 + \lambda(y_2 - y_1), \qquad z = z_2 + \lambda(z_2 - z_1). \tag{1}$$

While we may not yet know why such matters are important, we can observe that the equations

$$\begin{cases} x' = a_{11}x + a_{12}y + a_{13}z + b_1 \\ y' = a_{21}x + a_{22}y + a_{23}z + b_2 \\ z' = a_{31}x + a_{32}y + a_{33}z + b_3, \end{cases} \tag{2}$$

in which the a's and b's are constants, provide mathematical machinery into which we can substitute the coordinates (x,y,z) of a point in E_3 and thereby obtain the coordinates (x',y',z') of a point (usually another point) in E_3. It is very helpful to think of the equations (2) as being a transformer which transforms a given point $P(x,y,z)$ into a transform (or transformed point) $P'(x',y',z')$. Thus the transformer goes to work on $P(x,y,z)$, which engineers and others call an *input*, and produces $P'(x',y',z')$, which they call an *output*. This problem concerns transforms of points that lie on the given line L. Find, for each λ, the transform $P'(x',y',z')$ of the point $P(x,y,z)$ on L whose coordinates are given by (1). *Ans.:* The result can be put in the form

$$x' = x_1' + \lambda(x_2' - x_1'), \qquad y' = y_1' + \lambda(y_2' - y_1'), \qquad z' = z_1' + \lambda(z_2' - z_1'),$$

where

$$\begin{aligned} x_k' &= a_{11}x_k + a_{12}y_k + a_{13}z_k + b_1 \\ y_k' &= a_{21}x_k + a_{22}y_k + a_{23}z_k + b_2 \\ z_k' &= a_{31}x_k + a_{32}y_k + a_{33}z_k + b_3 \end{aligned}$$

when $k = 1$ and when $k = 2$. *Remark:* In case $x_2' = x_1'$, $y_2' = y_1'$, $z_2' = z_1'$, the transforms of the points on L all coincide with the point (x_1',y_1',z_1'). In case $x_2' \neq x_1'$ or $y_2' \neq y_1'$ or $z_2' \neq z_1'$, the transforms of the points on L constitute the line L' containing the distinct points $P_1'(x_1',y_1',z_1')$ and $P_2'(x_2',y_2',z_2')$ and, moreover, the transforms of points on L between P_1 and P_2 lie on L' between P_1' and P_2'.

21 Without use of a figure, suppose that θ is a given number for which $0 < \theta < \pi/2$ and find and simplify the condition that numbers x, y, z (not all zero) must satisfy if the vector

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

makes one of the angles θ and $\pi - \theta$ with the positive z axis. *Ans.:*

$$\mathbf{r}\cdot\mathbf{k} = \pm|\mathbf{r}|\,|\mathbf{k}| \cos\theta$$

and $z^2 = c^2(x^2 + y^2)$, where $c = \cot\theta$.

22 Two distinct (different) points P_0 and P_1, together with a number θ for which $0 < \theta < \pi/2$, determine a right circular cone consisting of P_0 (the vertex) and those points P for which the vector $\overrightarrow{PP_0}$ makes the angle θ or $\pi - \theta$ with the vector $\overrightarrow{P_0P_1}$. Show that the intrinsic equation of the cone is

$$(\overrightarrow{P_0P_1}\cdot\overrightarrow{P_0P})^2 = |\overrightarrow{P_0P_1}|^2|\overrightarrow{P_0P}|^2 \cos^2\theta.$$

Supposing that P_0, P_1, and P have coordinates (x_0,y_0,z_0), (x_1,y_1,z_1), (x,y,z), and that

$$(x_1 - x_0)\mathbf{i} + (y_1 - y_0)\mathbf{j} + (z_1 - z_0)\mathbf{k} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k},$$

find the coordinate equation of the cone. *Ans.:*

$$\begin{aligned}&[A(x - x_0) + B(y - y_0) + C(z - z_0)]^2\\ &\qquad = (A^2 + B^2 + C^2)[(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2] \cos^2\theta.\end{aligned}$$

23 The vertex of a right circular cone is at the point $(0,0,h)$, the axis of the cone is parallel to the vector $\mathbf{i} + \mathbf{j}$, and the lines on the cone make the angle $\pi/4$ with the axis of the cone. Find and simplify the equation of the cone. *Ans.:* $2xy = (z - h)^2$. *Remark:* Putting $z = 0$ shows that the graph in the xy plane having the equation $xy = h^2/2$ is a conic section, that is, the intersection of a right circular cone and a plane.

24 Prove that the graphs of the equations

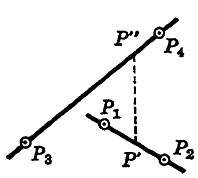
$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz = 0$$

and

$$x^3 + xyz + y^3 = 0$$

are cones with vertices at the origin. (See Problem 16 of Problems 2.29 for information about cones.)

Figure 2.395



25 Let $P_k(x_k,y_k,z_k)$, $k = 1, 2, 3, 4$, be four given points determining two skew (noncoplanar) lines P_1P_2 and P_3P_4 as in Figure 2.395. Many persons, including some who are not easily excited, become quite interested in the problem of determining a point $P'(x',y',z')$ on P_1P_2 and a point $P''(x'',y'',z'')$ on P_3P_4 such that the line $P'P''$ is perpendicular to both P_1P_2 and P_3P_4. Figure 2.395 may seem to be quite mysterious until it is realized that the lines P_1P_2 and P_3P_4 might be horizontal, while $P'P''$ might be a vertical line perpendicular to both of them. We shall solve this problem with the aid of vectors. The first step is to observe that the coordinates of P' and P'' can be written down easily if we find constants λ (lambda) and μ (mu) such that $\overrightarrow{P_1P'} = \lambda\overrightarrow{P_1P_2}$ and $\overrightarrow{P_3P''} = \mu\overrightarrow{P_3P_4}$. Therefore, our real problem is to find λ and μ. Starting with the formula

$$\overrightarrow{P'P''} = \overrightarrow{P'P_1} + \overrightarrow{P_1P_3} + \overrightarrow{P_3P''} \tag{1}$$

we find that

$$\overrightarrow{P'P''} = -\lambda\overrightarrow{P_1P_2} + \mu\overrightarrow{P_3P_4} + \overrightarrow{P_1P_3}. \tag{2}$$

The requirement that $\overrightarrow{P'P''}$ be perpendicular to $\overrightarrow{P_1P_2}$ and to $\overrightarrow{P_3P_4}$ is equivalent to the requirement that $\overrightarrow{P'P''}\cdot\overrightarrow{P_1P_2} = 0$ and $\overrightarrow{P'P''}\cdot\overrightarrow{P_3P_4} = 0$. This is equivalent to the requirement that λ and μ satisfy the equations

$$\begin{aligned} \lambda\overrightarrow{P_1P_2}\cdot\overrightarrow{P_1P_2} - \mu\overrightarrow{P_3P_4}\cdot\overrightarrow{P_1P_2} &= \overrightarrow{P_1P_3}\cdot\overrightarrow{P_1P_2} \\ \lambda\overrightarrow{P_1P_2}\cdot\overrightarrow{P_3P_4} - \mu\overrightarrow{P_3P_4}\cdot\overrightarrow{P_3P_4} &= \overrightarrow{P_1P_3}\cdot\overrightarrow{P_3P_4}. \end{aligned} \tag{3}$$

The question whether these equations have solutions for λ and μ is now critical. Such questions are so often critical that Theorem 2.57 will soon appear. For the record, and perhaps for future reference, it can be noted that the determinant of the coefficients of λ and μ is

$$-\,|\overrightarrow{P_1P_2}|^2\,|\overrightarrow{P_3P_4}|^2 \sin^2\theta, \tag{4}$$

where θ is the angle between the lines P_1P_2 and P_3P_4. If these lines are not parallel, then the determinant is different from 0 and the equations uniquely determine λ and μ. Our skew lines are not parallel. Therefore, λ and μ and hence P' and P'' are uniquely determined.

26 This problem requires that we pick up the idea that vectors and scalar products are not unrelated to problems in statistics. Let n be a positive integer and suppose at first that $n = 3$. Suppose n students took examinations in English and Mathematics and got grades $e_1, e_2, \cdots, e_n$ and $m_1, m_2, \cdots, m_n$. Let the mean (or average) of the English grades be E and let the mean of the Mathematics grades be M. For each $k = 1, 2, \cdots, n$ let

$$u_k = e_k - E, \qquad v_k = m_k - M. \tag{1}$$

The numbers $u_1, u_2, \cdots, u_n$ and $v_1, v_2, \cdots, v_n$ can be regarded as scalar components of vectors $\mathbf{u}$ and $\mathbf{v}$ in Euclid space E_n of n dimensions. Show that, except in the trivial case in which $|\mathbf{u}| = 0$ or $|\mathbf{v}| = 0$, the cosine of the angle between these vectors is determined by the formula

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}|\,|\mathbf{v}|} = \frac{u_1v_1 + u_2v_2 + \cdots + u_nv_n}{\sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}}. \tag{2}$$

Remark: It is not difficult to develop enough geometry of E_n to show that (2) is valid when $n > 3$ as well as when $n = 3$. In statistics, the last member of (2) is called the *correlation coefficient* of the English and Mathematics grades. Show that if this coefficient is 1, then the vectors $\mathbf{u}$ and $\mathbf{v}$ must have the same direction and hence that there must be a positive constant λ such that $v_k = \lambda u_k$ for each k. If possible, draw at least one substantial conclusion from this. Tell what we would conclude if the correlation coefficient turned out to be -1.

27 Show how something in the preceding problem can be used to prove that

$$|u_1v_1 + u_2v_2 + \cdots + u_nv_n| \leq \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

This is the *Schwarz inequality*. It is both interesting and useful.

2.4 *Planes and lines in E_3* Planes are important things, and we must think about them and the natures of their equations. To start the proceedings, we can think of the top surface of a flat horizontal sheet of paper as being a part of a plane π. Let P_1 be a point in π. A vertical pencil then represents a vector **V** which is a normal to the plane. Without bothering to decide how the fact is related to this or that set of postulates and definitions in Euclid geometry, we shall use the fact that a point P different from P_1 lies in π if and only if the vector $\overrightarrow{P_1P}$ is horizontal, that is, perpendicular to **V**. Our next step is to apply this idea to a plane π, shown schematically in Figure 2.43, which is not necessarily horizontal. Let **V** be a vector of positive length which is perpendicular to π and which runs from the origin to a point (A,B,C) not necessarily in π. Let $P_1(x_1,y_1,z_1)$ be a point in π. A point P then lies in π if and only if

$$\mathbf{V}\cdot\overrightarrow{P_1P} = 0. \tag{2.401}$$

This means that either $P = P_1$ or $\overrightarrow{P_1P}$ is a vector of positive length which is perpendicular to V. Thus a point $P(x,y,z)$ lies in π if and only if

$$[A\mathbf{i} + B\mathbf{j} + C\mathbf{k}]\cdot[(x - x_1)\mathbf{i} + (y - y_1)\mathbf{j} + (z - z_1)\mathbf{k}] = 0 \tag{2.402}$$

or

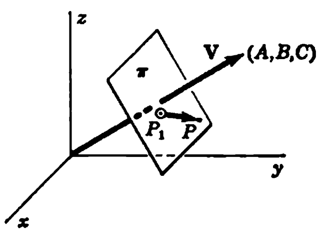
$$A(x - x_1) + B(y - y_1) + C(z - z_1) = 0 \tag{2.41}$$

or

$$Ax + By + Cz + D = 0, \tag{2.42}$$

where D is the constant defined by $D = -Ax_1 - By_1 - Cz_1$. It is easy to remember that the equation of a plane can always be put in the form (2.42), where A, B, C, D are constants of which A, B, C are not all zero. It is not so easy to remember that (2.41) is the equation of a plane which passes through the point (x_1,y_1,z_1) and is normal to the vector having scalar components A, B, C, but this should be done. To complete this little story, we must prove that if A, B, C, D are constants for which

Figure 2.43



A, B, C are not all 0, then the equation

$$Ax + By + Cz + D = 0$$

is the equation of a plane. In case $A \neq 0$, we can accomplish the result by putting the equation in the form

$$A\left(x - \frac{-D}{A}\right) + B(y - 0) + C(z - 0) = 0$$

and noticing that it is the equation of the plane which passes through the point $(-D/A, 0, 0)$ and is normal to the vector having scalar components A, B, C. In case $B \neq 0$ or $C \neq 0$, the proof is similar.

The information which we have obtained can be useful in various ways. Suppose, for example, we are required to find the equation of the plane π which contains three given noncollinear (not on a line) points $P_1(x_1,y_1,z_1)$, $P_2(x_2,y_2,z_2)$, $P_3(x_3,y_3,z_3)$. The obvious way to solve this problem is to use the fact that the equation of π must have the form of the first equation in the system

$$\left\{\begin{aligned} Ax + By + Cz + D &= 0 \\ Ax_1 + By_1 + Cz_1 + D &= 0 \\ Ax_2 + By_2 + Cz_2 + D &= 0 \\ Ax_3 + By_3 + Cz_3 + D &= 0, \end{aligned}\right. \tag{2.44}$$

where A, B, C, D are constants for which A, B, C are not all 0. Since π contains P_1, P_2, P_3, the remaining three equations must be satisfied. In case $A \neq 0$, the equation of π can be obtained by solving the last three equations for B, C, D in terms of A and substituting the results in the first equation. In case $A = 0$, we must have either $B \neq 0$ or $C \neq 0$, and a similar procedure will work. Except for cases in which some of the coordinates of the given points are zero, solving the problem in this way can be tedious. The next section will show how answers to this and other problems can be expressed in terms of determinants.

We now look at an interesting procedure which often provides a good way to find the equation of a plane π_1 which contains two given points $P_1(x_1,y_1,z_1)$ and $P_2(x_2,y_2,z_2)$ and satisfies another condition. We simplify matters by supposing that $x_2 \neq x_1$, $y_2 \neq y_1$, and $z_2 \neq z_1$. We know that P_1 and P_2 determine a line L and that the family F of planes π that contain P_1 and P_2 is identical with the family of planes π that contain L. We capitalize this fact. If a point $P(x,y,z)$ lies on L, then

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

and hence

$$\lambda\left(\frac{x - x_1}{x_2 - x_1} - \frac{z - z_1}{z_2 - z_1}\right) + \mu\left(\frac{y - y_1}{y_2 - y_1} - \frac{z - z_1}{z_2 - z_1}\right) = 0 \tag{2.45}$$

or

$$\frac{\lambda}{x_2 - x_1}(x - x_1) + \frac{\mu}{y_2 - y_1}(y - y_1) - \frac{\lambda + \mu}{z_2 - z_1}(z - z_1) = 0, \tag{2.451}$$

where λ (lambda) and μ (mu) are constants. Conversely, if λ and μ are not both 0, then (2.45) and (2.451) are equations of a plane containing L. To see this, we notice that they have the form $Ax + By + Cz + D = 0$, where A and B are not both 0, and that they are satisfied when $x = x_1$, $y = y_1$, $z = z_1$ and when $x = x_2$, $y = y_2$, $z = z_2$. Now we can solve problems. Suppose we want to find the equation of the plane π_1 which contains P_1 and P_2 and also a third point $P_3(x_3,y_3,z_3)$ not on the line P_1P_2. Our answer will be (2.45) or (2.451), and λ and μ are determined such that they are not both zero and the formulas hold when $x = x_3$, $y = y_3$, $z = z_3$. If the coefficient of λ in (2.45) is zero, we can take $\lambda = 1$ and $\mu = 0$; otherwise, we can set $\mu = 1$ and solve for λ. Suppose next that we want to find the equation of the plane π which contains P_1 and P_2 and is perpendicular to a given plane π'. Our answer will be (2.451) when λ and μ are determined such that they are not both 0 and a normal to π is perpendicular to a normal to π'. Supposing that the equation of π' is

$$Ax + By + Cz + D = 0,$$

we find that the normals are perpendicular when

$$\frac{\lambda A}{x_2 - x_1} + \frac{\mu B}{y_2 - y_1} - \frac{(\lambda + \mu)C}{z_2 - z_1} = 0. \tag{2.46}$$

It is possible to find values of λ and μ which satisfy this equation.

Let d be the distance from a given point $P_1(x_1,y_1,z_1)$ to a given plane π having the equation

$$Ax + By + Cz + D = 0. \tag{2.47}$$

One way to find d is to find the point P_0 where the line through P_1 perpendicular to π intersects π and then find the distance from P_0 to P_1. Whether this method is tedious or not can be a matter of opinion, but it is quite lengthy even when A, B, C, D, x_1, y_1, z_1 are given to be nice little integers. With the aid of vectors, we can very easily find d in terms of A, B, C, D, x_1, y_1, z_1. Let $P(x,y,z)$ be any point in π, and let $\mathbf{n}$ be a unit normal to π. Then, as we can see with the aid of the schematic Figure 2.471 in which P_0 and P are on π and P_1P_0 is a normal to π,

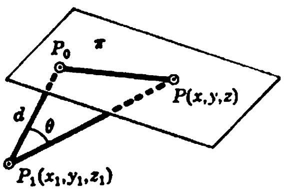


Figure 2.471

$$d = |\,|\overrightarrow{PP_1}|\cos\theta\,| = |\mathbf{n}\cdot\overrightarrow{PP_1}|. \tag{2.472}$$

But

$$\mathbf{n} = \frac{A\mathbf{i} + B\mathbf{j} + C\mathbf{k}}{\sqrt{A^2 + B^2 + C^2}}, \qquad \overrightarrow{PP_1} = (x_1 - x)\mathbf{i} + (y_1 - y)\mathbf{j} + (z_1 - z)\mathbf{k}.$$

Therefore,

$$d = \left| \frac{A(x_1 - x) + B(y_1 - y) + C(z_1 - z)}{\sqrt{A^2 + B^2 + C^2}} \right|. \tag{2.473}$$

This looks quite simple but, since P is in π, the equation of π shows that

$$-Ax - By - Cz = D$$

and we obtain the more useful formula

$$d = \left| \frac{Ax_1 + By_1 + Cz_1 + D}{\sqrt{A^2 + B^2 + C^2}} \right|. \tag{2.48}$$

One who must teach his little sister to start with (2.47) and get (2.48) can cook up a new five-step rule: (i) rub out the "$=0$"; (ii) put subscripts on x, y, z; (iii) divide by $\sqrt{A^2 + B^2 + C^2}$; (iv) stick on absolute-value signs; and (v) equate the result to d. It is, as a matter of fact, useful to know when and how it is possible to prepare instructions so explicit that routine chores can be performed mechanically and can even be performed by persons and machines unfamiliar with processes by which formulas are derived and combined to accomplish their purposes.

Problems 2.49

1 Give geometric interpretations of the numbers x_0, y_0, z_0, A, B, C appearing in the equation $A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$ of a plane π and be prepared to repeat the process at any time. *Ans.:* See text.

2 Write an intrinsic (not depending upon coordinates) equation of the plane π which contains a given point P_0 and is normal to a given vector $\mathbf{V}$. *Hint:* If P is in π, then $\overrightarrow{P_0P}$ must be perpendicular (or normal or orthogonal) to $\mathbf{V}$. *Ans.:* $\mathbf{V}\cdot\overrightarrow{P_0P} = 0$.

3 How can we derive the coordinate equation of Problem 1 from the intrinsic equation of Problem 2? *Ans.:* Set $\mathbf{V} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ and

$$\overrightarrow{P_0P} = (x - x_0)\mathbf{i} + (y - y_0)\mathbf{j} + (z - z_0)\mathbf{k}$$

so that

$$\mathbf{V}\cdot\overrightarrow{P_0P} = A(x - x_0) + B(y - y_0) + C(z - z_0).$$

4 Write an intrinsic (not depending upon coordinates) formula for the distance d from a point P_1 to a plane π which contains a point P and is normal to a unit vector $\mathbf{n}$. *Hint:* Construct an appropriate schematic figure and refer to Figure 2.471 and formula (2.472) only if assistance is needed.

5 In each case, find the (or an) equation of the plane π which contains the given point and is perpendicular to a vector having the given scalar components (or, in other words, perpendicular to a line having the given direction numbers).

(*a*) (0,0,0); 1, 1, 1 — *Ans.:* $x + y + z = 0$
(*b*) (1,1,1); 1, 1, 1 — *Ans.:* $x + y + z = 3$
(*c*) (1,2,3); 4, 5, 6 — *Ans.:* $4(x - 1) + 5(y - 2) + 6(z - 3) = 0$
(*d*) (x_1,y_1,z_1); A, B, C — *Ans.:* $A(x - x_1) + B(y - y_1) + C(z - z_1) = 0$

6 Because several of the coordinates are zero, it is relatively easy to determine A, B, C, D such that the equation $Ax + By + Cz + D = 0$ is satisfied by the coordinates of the three points (3,0,0), (0,4,0), (3,4,5). Do it and thereby find the equation of the plane containing the three points. *Ans.:* $20x + 15y - 12z - 60 = 0$.

7 A sphere of radius 3 has its center at the origin. Observe that it is not easy to sketch a figure showing the plane π tangent to the sphere at the point (2,2,1) and the point T where π intersects the z axis. Find the coordinates of T. *Hint:* The plane π is normal to the line from the center of the sphere to the point of tangency. *Ans.:* (0,0,9).

8 If a, b, and c are nonzero constants, show that the equation

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1$$

is the equation of the plane which intersects the coordinate axes at the points $(a,0,0)$, $(0,b,0)$, and $(0,0,c)$. Find the scalar components of a normal to the plane.

9 Find the distance from the origin to the plane of the preceding problem. Your answer is wrong if it does not reduce to $1/\sqrt{3}$ when $a = b = c = 1$.

10 A plane π_1 intersects the positive x, y, and z axes 1, 2, and 3 units, respectively, from the origin. A second plane π_2 intersects each positive axis one unit farther from the origin. Would you suppose that π_1 and π_2 are parallel? Find the acute angle θ between normals to the planes. *Ans.:* $\cos\theta = \sqrt{2916/2989}$.

11 Find the equation of the plane which contains the point (1,2,3) and is parallel to the plane having the equation $3x + 2y + z - 1 = 0$. Check the answer.

12 Find the equation of the plane π_1 which contains the point (1,3,1) and is perpendicular to the line L having the equations

$$(1) \qquad x = t, \qquad y = t, \qquad z = t + 2.$$

Hint: If you do not know what else to do, let $t = 0$ to get a point P_1 on L and let $t = 1$ to get another point P_2 on L. Then π must be perpendicular to $\overrightarrow{P_1P_2}$. Your answer can be worked out neatly by solving the individual equations in (1) for t to obtain

$$\frac{x-0}{1} = \frac{y-0}{1} = \frac{z-2}{1} = t.$$

This shows that a normal to π has scalar components 1, 1, 1 and hence that the equation of π is

$$(x - 1) + (y - 3) + (z - 1) = 0$$

or $x + y + z = 5$.

13 Find the equation of the plane π containing the point $P_1(1,1,1)$ which is perpendicular to the line L containing the points (0,1,0) and (0,0,1). Find the coordinates of the point P_2 where L intersects π. Then find $|\overrightarrow{P_1P_2}|$. Observe that the last number (which should be $\sqrt{\frac{3}{2}}$) is the distance from a vertex of a unit cube to a diagonal of a face not containing this vertex. Look at a figure and discover very simple reasons why $1 < |\overrightarrow{P_1P_2}| < 2$.

14 Determine the value of the parameter λ for which the two planes which have the equations

$$\begin{aligned} 2x + 3y + 4z + 5 &= 0 \\ 2x - 3y - \lambda z - 5 &= 0 \end{aligned}$$

are orthogonal. *Hint:* Modest experiments with two sheets of paper enable us to capture or recapture the idea that two planes are orthogonal (or normal or perpendicular to each other) if and only if their normals are orthogonal. *Ans.:* $\lambda = -\frac{5}{4}$.

15 If B, C, D are constants for which B and C are not both 0, then the equation $By + Cz + D = 0$ is the equation of a plane π. Show that the vector $\mathbf{V}_1$ with scalar components $0, B, C$ is normal to π. Show that the vector $\mathbf{V}_2$ with scalar components $1,0,0$, is normal to the yz plane. Show that $\mathbf{V}_1$ and $\mathbf{V}_2$ are perpendicular and hence that π is perpendicular to the yz plane.

16 Consider again the equation $By + Cz + D = 0$ or any other equation involving y and z but not x. Let us agree (this is an important definition) that a set S_1 in E_3 is a *cylinder* parallel to a line L if, whenever it contains a point P_0, it also contains all of the points on the line L_0 through P_0 parallel to L. Use this to show that the graph of the given equation is a cylinder parallel to the x axis. *Solution:* Let $P_0(x_0,y_0,z_0)$ be any point on the graph of the given equation. Then the numbers x_0, y_0, z_0 satisfy the equation. Since x does not appear in the equation, it follows that the numbers x, y_0, z_0 satisfy the equation for each x. This means that all of the points (x,y_0,z_0) on the line L_0 through (x_0,y_0,z_0) parallel to the x axis lie on the graph. Therefore, the graph is a cylinder parallel to the x axis.

17 Supposing that B, C, and D are constants for which B and C are not both 0 and $D \neq 0$, show that there is no number x for which the numbers $x, 0, 0$ satisfy the equation $By + Cz + D = 0$. What is the geometric significance of this result?

18 Look at the equations

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

of the line containing two points $P_1(x_1,y_1,z_1)$ and $P_2(x_2,y_2,z_2)$. Describe completely the graph of the equation obtained by deleting one of the members of this equality. Sketch the three graphs obtained in this way.

19 Problem 12 of Problems 2.29 is of interest here. Solve the problem again and think about it.

20 Take a good look at the equation

$$\frac{\lambda}{x_2 - x_1}(x - x_1) + \frac{\mu}{y_2 - y_1}(y - y_1) - \frac{\lambda + \mu}{z_2 - z_1}(z - z_1) = 0. \tag{1}$$

Supposing that $P_1(x_1,y_1,z_1)$ and $P_2(x_2,y_2,z_2)$ are two points for which $x_2 \neq x_1$, $y_2 \neq y_1$, $z_2 \neq z_1$ and that λ and μ are numbers not both zero, tell why (1) is the equation of a plane containing P_1 and P_2. Then study the text again and attain a better understanding of matters relating to (1).

21 Supposing that A, B, C are not all zero and that $D_2 \neq D_1$, let π_1 and π_2 be the planes having the equations

$$(1) \qquad Ax + By + Cz + D_1 = 0$$
$$(2) \qquad Ax + By + Cz + D_2 = 0.$$

Show that the planes have parallel normals and hence that the planes are parallel. Show again that the planes must be parallel by showing that they have no point of intersection; if the coordinates of a point $P(x,y,z)$ satisfy (1), they certainly cannot satisfy (2). Supposing that λ and μ are constants not both zero, show that the equation

$$(3) \qquad \lambda(Ax + By + Cz + D_1) + \mu(Ax + By + Cz + D_2) = 0$$

is the equation of a plane parallel to π_1 and π_2 unless $\lambda + \mu = 0$. Supposing that $P_0(x_0,y_0,z_0)$ is a given point, show that λ and μ can be so determined that the graph of (3) contains P_0. *Solution of last part:* The graph of (3) will be a plane containing P_0 if and only if $\lambda + \mu \neq 0$ and

$$(4) \qquad \lambda(Ax_0 + By_0 + Cz_0 + D_1) + \mu(Ax_0 + By_0 + Cz_0 + D_2) = 0.$$

Since $D_1 \neq D_2$, the coefficients of λ and μ are different numbers that we can call E and F. We can then put (4) in the form

$$(5) \qquad \lambda E + \mu F = 0$$

and obtain a solution of our problem by setting $\lambda = F$ and $\mu = -E$ because, in this case, $\lambda + \mu = F - E \neq 0$ and (4) holds.

22 Supposing that the graphs of the equations

$$(1) \qquad A_1x + B_1y + C_1z + D_1 = 0$$
$$(2) \qquad A_2x + B_2y + C_2z + D_2 = 0$$

are distinct (that is, different) planes π_1 and π_2 that intersect in a line L and that λ and μ are constants not both zero, show that the equation

$$(3) \qquad \lambda(A_1x + B_1y + C_1z + D_1) + \mu(A_2x + B_2y + C_2z + D_2) = 0$$

is the equation of a plane π containing the line L. *Solution:* It is clear that if $P(x,y,z)$ lies on L, then the coordinates of P satisfy both (1) and (2) and hence (3) To prove that (3) is the equation of a plane, we must prove that the three equations

$$\lambda A_1 + \mu A_2 = 0, \qquad \lambda B_1 + \mu B_2 = 0, \qquad \lambda C_1 + \mu C_2 = 0$$

cannot be simultaneously satisfied when λ and μ are not both zero. This matter is more delicate. If we suppose that the three equations are satisfied and $\lambda \neq 0$, we find that

$$A_1 = (-\mu/\lambda)A_2, \qquad B_1 = (-\mu/\lambda)B_2, \qquad C_1 = (-\mu/\lambda)C_2$$

and obtain a contradiction of the hypothesis that π_1 and π_2 are not parallel. The case $\mu \neq 0$ is similar.

23 Using the hypotheses and equations of the preceding problem, show how to determine λ and μ such that (3) will be the equation of a plane π containing L

and a given point $P_0(x_0,y_0,z_0)$. *Hint:* Consider separately the case in which P_0 is on L and the case in which P_0 is not on L.

24 Let π_1 and π_2 be the planes having the equations

$$\begin{cases} 2x + 3y + 4z - 5 = 0 \\ x - 2y + 3z - 4 = 0 \end{cases} \tag{1}$$

Verify that π_1 and π_2 do not have parallel normals. This implies that π_1 and π_2 must intersect in a line L. Observe that a point $P(x,y,z)$ lies on L if and only if it lies on both π_1 and π_2 and hence if and only if its coordinates satisfy both of the equations (1). We should be able to find point-direction equations of L by finding the coordinates of two points P_1 and P_2 on L and using them. Do this by finding x and y such that the point $(x,y,0)$ lies on L and then finding x and y such that the point $(x,y,1)$ lies on L. We now develop a simpler and more interesting method for finding equations of L. Show that if $P(x,y,z)$ lies on L, then

$$7x + 17z - 22 = 0. \tag{2}$$

Observe that (2), the result of eliminating y from the equations (1), is obtained by multiplying the equations (1) by 2 and 3, respectively, and adding the results. Observe that (2) is the equation of the plane which passes through L and is perpendicular to the xz plane. Show also that if $P(x,y,z)$ lies on L, then

$$7y - 2z + 3 = 0. \tag{3}$$

Discuss this matter. By solving the equations (2) and (3) for z, show that

$$\frac{7x - 22}{-17} = \frac{7y + 3}{2} = z \tag{4}$$

and that these are equations of L. Observe that dividing these equations by 7 puts them in the point-direction form

$$\frac{x - \frac{22}{7}}{-17} = \frac{y + \frac{3}{7}}{2} = \frac{z - 0}{7}. \tag{5}$$

25 Find the equation of the plane π which contains the point $(1,3,1)$ and the line L having the equations

$$x = t, \qquad y = t, \qquad z = t + 2. \tag{1}$$

First solution: The equation of π has the form

$$A(x - 1) + B(y - 3) + C(z - 1) = 0. \tag{2}$$

Since π contains the given line, we must have

$$A(t - 1) + B(t - 3) + C(t + 1) = 0 \tag{3}$$

for each t. Putting $t = 3$ and then $t = 1$ shows that

$$2A + 4C = 0, \qquad -2B + 2C = 0. \tag{4}$$

Solving these equations for A and B and putting the results in (2) gives

$$-2C(x - 1) + C(y - 3) + C(z - 1) = 0. \tag{5}$$

Dividing by C and multiplying by -1 gives

(6) $$2(x - 1) - (y - 3) - (z - 1) = 0$$

or

(7) $$2x - y - z + 2 = 0.$$

Second solution: Eliminating t from the first two and then from the first and last of the equations (1) shows that L is the intersection of the planes having the equations

(8) $$x - y = 0, \qquad x - z + 2 = 0.$$

Each plane containing L has an equation of the form

(9) $$\lambda(x - y) + \mu(x - z + 2) = 0,$$

where λ and μ are constants. The plane having the equation (9) contains the point (1,3,1) if and only if $-2\lambda + 2\mu = 0$. Putting $\mu = \lambda$ gives the required equation

(10) $$2x - y - z + 2 = 0.$$

26 There are nontrivial applications of the idea that a point which lies in each of two nonparallel planes must lie on their line of intersection. Show that the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$

will be satisfied if, for some constant λ, the two equations

$$\frac{x}{a} + \frac{z}{c} = \lambda\left(1 + \frac{y}{b}\right), \qquad \lambda\left(\frac{x}{a} - \frac{z}{c}\right) = 1 - \frac{y}{b}$$

both hold. Try to determine a geometrical interpretation of this result. Try to obtain another very similar result.

27 Let L and L' be the lines of intersection of the planes having the equations

$$L\begin{cases} \dfrac{x}{a} + \dfrac{z}{c} = \lambda\left(1 + \dfrac{y}{b}\right) \\ \lambda\left(\dfrac{x}{a} - \dfrac{z}{c}\right) = 1 - \dfrac{y}{b}, \end{cases} \qquad L'\begin{cases} \dfrac{x}{a} + \dfrac{z}{c} = \mu\left(1 - \dfrac{y}{b}\right) \\ \mu\left(\dfrac{x}{a} - \dfrac{z}{c}\right) = 1 + \dfrac{y}{b}. \end{cases}$$

Work out the point-direction forms

$$L: \qquad \frac{x - \dfrac{2\lambda a}{1 + \lambda^2}}{-(1 - \lambda^2)a} = \frac{y - \dfrac{(1 - \lambda^2)b}{1 + \lambda^2}}{2\lambda b} = \frac{z - 0}{(1 + \lambda^2)c}$$

$$L': \qquad \frac{x - \dfrac{2\mu a}{1 + \mu^2}}{-(1 - \mu^2)a} = \frac{y + \dfrac{(1 - \mu^2)b}{1 + \mu^2}}{-2\mu b} = \frac{z - 0}{(1 + \mu^2)c}$$

of equations of L and L'.

28 What can be said about the tip of the vector $\overrightarrow{OP}$ if

$$\overrightarrow{OP} = c_1\overrightarrow{OP}_1 + c_2\overrightarrow{OP}_2 + c_3\overrightarrow{OP}_3,$$

where c_1, c_2, c_3 are scalars for which $c_1 + c_2 + c_3 = 1$? *Ans.:* It lies in the plane (or each plane) which contains P_1, P_2, and P_3.

2.5 *Determinants and applications*

Rectangular arrays of elements such as

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \qquad \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}$$

are called *matrices*. For the present we may think of the elements a_{jk} as being numbers. The middle matrix has three *rows*, the elements of the second row being a_{21}, a_{22}, a_{23}, and three *columns* (columns are things that stand in vertical positions), the elements of the third column being a_{13}, a_{23}, a_{33}. A matrix is *square* if it contains as many columns as rows, and in this case the number of rows is called the *order* of the matrix.

With each square matrix we associate a number which is called the *determinant of the matrix* or simply a *determinant*. The symbols Δ_2, Δ_3, and Δ_4 appearing in

$$(2.51)\qquad \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \qquad \Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$

$$\Delta_4 = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

are numbers, not matrices. It will, however, be a convenience to say that a_{32} is the element in the third row and second column of the determinant Δ_3 instead of saying that it is the element in the third row and second column of the matrix of which Δ_3 is the determinant. A little time spent learning about determinants can pay very handsome dividends.

The number Δ_2 is defined by the formula

$$(2.52)\qquad \Delta_2 = a_{11}a_{22} - a_{12}a_{21}.$$

This shows how to evaluate determinants of order 2. For example,

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1, \qquad \begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ab = 0, \qquad \begin{vmatrix} 1 & 2 \\ -3 & 4 \end{vmatrix} = 4 - (-6) = 10.$$

The definitions of Δ_3 and Δ_4 are more complicated, and we introduce helpful words and notations. To each element a_{jk} of a determinant there corresponds the *minor* A_{jk} which remains after the row and column containing a_{jk} have been covered or removed. Thus for the determinant Δ_3, we have

$$(2.53)\quad A_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \qquad A_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, \qquad A_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$$

and for the determinant Δ_4 we have

$$A_{11} = \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix}, \quad A_{12} = \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix}, \quad A_{23} = \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix}.$$

While it is possible to give more gruesome definitions, the number Δ_3 can be defined by the formula

$$(2.54)\qquad \Delta_3 = a_{11}A_{11} - a_{12}A_{12} + a_{13}A_{13}$$

or

$$\Delta_3 = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix},$$

which makes sense because we know how to evaluate determinants of order 2. The above formulas give *the expansion of Δ_3 in terms of the elements of the first row*. It can be proved that the same number Δ_3 is furnished by expansions in terms of another row or any column. Thus

$$\begin{aligned}
\Delta_3 &= -a_{21}A_{21} + a_{22}A_{22} - a_{23}A_{23} \\
\Delta_3 &= a_{31}A_{31} - a_{32}A_{32} + a_{33}A_{33} \\
\Delta_3 &= a_{11}A_{11} - a_{21}A_{21} + a_{31}A_{31} \\
\Delta_3 &= -a_{12}A_{12} + a_{22}A_{22} - a_{32}A_{32} \\
\Delta_3 &= a_{13}A_{13} - a_{23}A_{23} + a_{33}A_{33}.
\end{aligned}$$

In these expansions, the sign of the term involving a_{jk} is plus whenever the sum (or difference) of the subscripts is an even number like 0, ± 2, ± 4, $\cdots$ and is minus when the sum (or difference) is odd like ± 1, ± 3, ± 5, $\cdots$. To put this in other words, we can say that we get a plus sign whenever a_{jk} lies on the main diagonal (running from the upper left corner to the lower right corner), and that we get a change in sign whenever we move one step right, left, down, or up.

Progress to determinants of order 4 and more is now easy. The expansion of Δ_4 in terms of the elements of the first row is

$$(2.55)\qquad \Delta_4 = a_{11}A_{11} - a_{12}A_{12} + a_{13}A_{13} - a_{14}A_{14}$$

and Δ_4 has seven more expansions, in terms of the other rows and the columns, all of which yield the same number Δ_4. The usefulness of the possibility of expanding a determinant of order 4 in terms of elements of the third row is demonstrated by the expansion

$$\begin{vmatrix} 2 & 3 & -1 & 6 \\ 1 & 3 & 7 & -2 \\ 3 & 0 & 0 & 0 \\ 1 & -1 & 5 & 4 \end{vmatrix} = 3 \begin{vmatrix} 3 & -1 & 6 \\ 3 & 7 & -2 \\ -1 & 5 & 4 \end{vmatrix} + 0$$

which reduces the problem of evaluating a determinant of order 4 to the problem of evaluating a single determinant of order 3.

It should be known and remembered that *a determinant is* 0 *if two of its rows (or two of its columns) are identical.* For determinants of order 2, this is obvious because

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ab = 0, \qquad \begin{vmatrix} a & a \\ b & b \end{vmatrix} = ab - ab = 0.$$

That

$$\begin{vmatrix} a & b & c \\ a & b & c \\ d & e & f \end{vmatrix} = 0, \qquad \begin{vmatrix} a & d & a \\ b & e & b \\ c & f & c \end{vmatrix} = 0$$

can be seen by expanding the first determinant in terms of elements of the bottom row and by expanding the second in terms of elements of the middle column. When the result has been established for determinants of order 3, the same trick enables us to establish the result for determinants of order 4, and so on. A simple modification of the above procedure shows that if two adjacent rows (or columns) of a determinant are interchanged, then the value of the determinant is multiplied by -1, that is, the sign of the determinant is changed. It is sometimes useful to know the formulas

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ kb_1 & kb_2 & kb_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = k \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \tag{2.56}$$

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 + c_1 & b_2 + c_2 & b_3 + c_3 \\ d_1 & d_2 & d_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ d_1 & d_2 & d_3 \end{vmatrix} + \begin{vmatrix} a_1 & a_2 & a_3 \\ c_1 & c_2 & c_3 \\ d_1 & d_2 & d_3 \end{vmatrix} \tag{2.561}$$

and others like them. They can be proved by expanding the determinants in terms of elements of the middle row. The results are particularly useful when, for some constant k, we have $c_1 = kd_1$, $c_2 = kd_2$, and $c_3 = kd_3$. In this case the last determinant in the above formula is zero and we obtain the formula

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 + kd_1 & b_2 + kd_2 & b_3 + kd_3 \\ d_1 & d_2 & d_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ d_1 & d_2 & d_3 \end{vmatrix}. \tag{2.562}$$

What this and similar formulas say is that we do not change the value of a determinant when we add a constant multiple of the elements of one row (or column) to the elements of another row (or column). For example, we obtain the first equality in

$$\begin{vmatrix} 2 & -3 & 1 \\ 1 & -2 & 3 \\ 3 & 3 & -5 \end{vmatrix} = \begin{vmatrix} 2 & 1 & 1 \\ 1 & 0 & 3 \\ 3 & 9 & -5 \end{vmatrix} = \begin{vmatrix} 2 & 1 & -5 \\ 1 & 0 & 0 \\ 3 & 9 & -14 \end{vmatrix}$$

by adding 2 times the elements of the first column to the elements of the second column, and then we obtain the second equality by adding -3 times the elements of the first column to the elements of the last column. As we have seen, this reduces the problem of evaluating a determinant of order 3 to the problem of evaluating a determinant of order 2.

The following two theorems, and their obvious modifications involving systems containing two or more than three equations, are very important.

Theorem 2.57 *The system of equations*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= y_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= y_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= y_3 \end{aligned}$$

has a unique solution (is satisfied by one and only one set of numbers x_1, x_2, x_3) *if and only if*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \neq 0,$$

that is, the determinant of the coefficients is different from 0.

Theorem 2.58 *The system of equations*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= 0 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= 0 \end{aligned}$$

has a nontrivial solution (a solution for which x_1, x_2, x_3 are not all 0) *if and only if*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0,$$

that is, the determinant of the coefficients is 0.

Proofs of these theorems belong in books and courses in algebra, but everybody can observe that the first system of equations

$$\begin{aligned} 2x_1 + 4x_2 &= 8 \\ x_1 - 2x_2 &= 1, \end{aligned} \qquad \begin{aligned} 2x_1 + 4x_2 &= 8 \\ x_1 + 2x_2 &= 0, \end{aligned} \qquad \begin{aligned} 2x_1 + 4x_2 &= 0 \\ x_1 + 2x_2 &= 0 \end{aligned}$$

has a unique solution, the second system has no solutions, and the third system has many solutions including the nontrivial one $x_1 = 2, x_2 = -1$. Partly because of these two theorems, determinants are important. Determinants were originally devised to speed the process of solving systems of equations whose coefficients are given in decimal form. It is sometimes said that young algebra students should not be taught to solve systems of equations by use of determinants because the method is inefficient and yields too many errors; the method of successive eliminations is much better. This argument is vulnerable, because students who solve systems of equations by use of determinants acquire facility in use of determinants. In this course, it is recommended that determinants be used only for puiposes for which they are useful.

Problems 2.59

1 Show that

$$\begin{vmatrix} A & B & C \\ 1 & 2 & -1 \\ 3 & 4 & 5 \end{vmatrix} = 14A - 8B - 2C.$$

2 Supposing that $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ are fixed points in E_2, show that the equation

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0 \tag{1}$$

has the form

$$Ax + By + C = 0 \tag{2}$$

and that the graph contains P_1 and P_2. Comment upon the result. *Solution:* Expanding (1) in terms of the elements of the first row gives (2). The equation (1) is satisfied when $x = x_1$, $y = y_1$ and when $x = x_2$, $y = y_2$ because in each case the determinant has two identical rows. The equation (2) is the equation of a line unless $A = B = 0$, that is, unless P_1 and P_2 coincide. If P_1 and P_2 do coincide, the equation (2) becomes $0x + 0y + 0 = 0$ and the graph is the whole plane.

3 Letting Δ be the first determinant in the formula

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = \begin{vmatrix} x - x_1 & y - y_1 & z - z_1 & 0 \\ x_1 & y_1 & z_1 & 1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 & 0 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 & 0 \end{vmatrix},$$

show how the formula can be obtained and show that

$$\Delta = \begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix}.$$

Now write formulas for the coefficients A, B, C in the expansion

$$\Delta = A(x - x_1) + B(y - y_1) + C(z - z_1).$$

Show that the graph of the equation $\Delta = 0$ contains the three points $P_1(x_1,y_1,z_1)$, $P_2(x_2,y_2,z_2)$, $P_3(x_3,y_3,z_3)$. Comment upon the result.

4 Let $|T|$ denote the area of the triangle T having vertices at the points P_1, P_2, P of Figure 2.591. With an eye on the figure, discover a way in which the formula

$$(1) \qquad |T| = \frac{y_1 + y}{2}(x - x_1) + \frac{y + y_2}{2}(x_2 - x) - \frac{y_1 + y_2}{2}(x_2 - x_1)$$

can be obtained. After expanding the products, show that some of the terms cancel out and that the formula can be put in the form

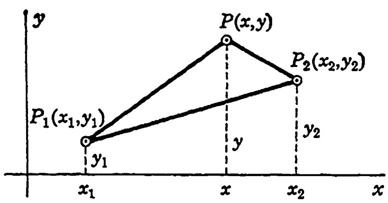


Figure 2.591

$$(2) \qquad |T| = \pm\tfrac{1}{2}\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}$$

with the plus sign. It can be shown that (2) is correct with the plus sign when the vertices P, P_1, P_2 occur in positive (counterclockwise) order, and that (2) is correct with the minus sign when the vertices P, P_1, P_2 occur in negative (clockwise) order. The members of (2) are 0 when the points are collinear. Many people remember this.

5 This and the next two problems, together with Problem 11 at the end of the next section, show that if V is the volume of the tetrahedron (or simplex) in E_3 having vertices (x,y,z), (x_1,y_1,z_1), (x_2,y_2,z_2), (x_3,y_3,z_3), then

$$V = \pm\tfrac{1}{6}\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix},$$

where the sign is chosen such that $V > 0$ (or $V \geqq 0$ if we allow degenerate tetrahedrons to be called tetrahedrons). Verify that this formula is correct with the negative sign for the special case in which a, b, c are positive constants and the four vertices are, in order, $(0,0,0)$, $(a,0,0)$, $(0,b,0)$, and $(0,0,c)$. *Hint:* Remember or learn that the volume of a simplex (tetrahedron) in E_3 is one-third of the product of the altitude (number, not line segment) and the area of the triangular base.

6 Letting D be the determinant of the preceding problem, show that

$$D = -\begin{vmatrix} x & y & z & 1 \\ -x_1 & -y_1 & -z_1 & -1 \\ -x_2 & -y_2 & -z_2 & -1 \\ -x_3 & -y_3 & -z_3 & -1 \end{vmatrix} = -\begin{vmatrix} x & y & z & 1 \\ x - x_1 & y - y_1 & z - z_1 & 0 \\ x - x_2 & y - y_2 & z - z_2 & 0 \\ x - x_3 & y - y_3 & z - z_3 & 0 \end{vmatrix}$$

and hence that the formula for the volume of the tetrahedron can be put in the form

$$V = \pm\tfrac{1}{6}\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x - x_2 & y - y_2 & z - z_2 \\ x - x_3 & y - y_3 & z - z_3 \end{vmatrix}.$$

7 Three vectors **u** and **v** and **w** with scalar components u_1, u_2, u_3 and v_1, v_2, v_3 and w_1, w_2, w_3 have their tails at the same point P and are edges of a tetrahedron having volume V. Show that the formulas for V in the two preceding problems can be put in the form

$$V = \pm\tfrac{1}{6}\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

8 Supposing that $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, and $P_3(x_3,y_3)$ are three noncollinear (not on a line) points in E_2, show that the equation

$$\begin{vmatrix} x^2 + y^2 & x & y & 1 \\ x_1^2 + y_1^2 & x_1 & y_1 & 1 \\ x_2^2 + y_2^2 & x_2 & y_2 & 1 \\ x_3^2 + y_3^2 & x_3 & y_3 & 1 \end{vmatrix} = 0$$

has the form

$$A(x^2 + y^2) + Bx + Cy + D = 0$$

and that the graph contains the three points. Comment upon this result.

9 Show that if $P_1(x_1,y_1)$, $\cdots$, $P_5(x_5,y_5)$ are five different points in E_2, then the equation

$$\begin{vmatrix} x^2 & xy & y^2 & x & y & 1 \\ x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & x_3y_3 & y_3^2 & x_3 & y_3 & 1 \\ x_4^2 & x_4y_4 & y_4^2 & x_4 & y_4 & 1 \\ x_5^2 & x_5y_5 & y_5^2 & x_5 & y_5 & 1 \end{vmatrix} = 0$$

has the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

and that its graph contains the five points.

10 Use the ideas of the above examples to obtain the equation of a sphere in E_3 which contains (or passes through) an appropriate collection of given points.

11 Let $P_1(x_1,y_1,z_1)$, $\cdots$, $P_5(x_5,y_5,z_5)$ be five given points (the coordinates are supposed to be known numbers) such that no four of them lie in the same plane. Tell how to decide whether the line L containing P_1, P_2 is parallel to the plane π containing P_3, P_4, P_5 and tell how to find the point of intersection when L intersects π. *Solution:* There are many ways to attack this problem. The following method gives answers in terms of the known coordinates with relatively little calculation. A point $P(x,y,z)$ lies on the line L if and only if, for some constant λ,

$$(1)\quad x = x_1 + \lambda(x_2 - x_1), \qquad y = y_1 + \lambda(y_2 - y_1), \qquad z = z_1 + \lambda(z_2 - z_1).$$

This point P lies on π if and only if $D = 0$, where

$$(2)\qquad D = \begin{vmatrix} x_1 + \lambda(x_2 - x_1) & y_1 + \lambda(y_2 - y_1) & z_1 + \lambda(z_2 - z_1) & 1 + 0 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \\ x_5 & y_5 & z_5 & 1 \end{vmatrix}.$$

But

$$(3)\qquad D = D_1 + \lambda D_2,$$

where

$$(4)\qquad D_1 = \begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \\ x_5 & y_5 & z_5 & 1 \end{vmatrix}, \qquad D_2 = \begin{vmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 & 0 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \\ x_5 & y_5 & z_5 & 1 \end{vmatrix}.$$

Now $D_1 \neq 0$ because P_1, P_3, P_4, P_5 do not lie in the same plane. In case $D_2 = 0$, there will be no λ for which $D = 0$ and hence no point P on L which lies in π, so L and π must be parallel. In case $D_2 \neq 0$, there will be exactly one λ for which $D = 0$, that is, $\lambda = -D_1/D_2$. Putting this in (1) then gives the coordinates of the point of intersection of L and π.

12 Find out whether the plane containing two vertices of a tetrahedron and the mid-point of the opposite edge must contain the centroid of the tetrahedron.

13 This story can be read by anyone. It interests nearly everyone, but students who do not have a lot of time at their disposal can postpone the pleasures and benefits enjoyed by those who fully understand it. Matrices are often denoted by single letters. For examples, we can set

$$(1)\quad P = \begin{pmatrix} 2 & 1 & 3 \\ -3 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 3 & 1 & -2 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad R = \begin{pmatrix} 7 & 4 & -6 \\ -7 & -6 & 9 \\ 2 & 3 & -4 \end{pmatrix}$$

$$S = \begin{pmatrix} 1 & 7 & 6 \\ 6 & -2 & 5 \\ -4 & 3 & -2 \end{pmatrix}, \quad T = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad U = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \quad V = \begin{pmatrix} -3 \\ 8 \\ -4 \end{pmatrix}.$$

We shall, among other things, develop enough algebra of matrices to enable us to understand and verify the fact that these matrices satisfy the conditions

$$(2)\qquad \begin{gathered} PQ = R \qquad QP = S \qquad PQ \neq QP, \\ QT = U, \quad PU = V, \quad P(QT) = V, \quad (PQ)T = V \\ \det(PQ) = \det(QP) = (\det P)(\det Q). \end{gathered}$$

If A is a square matrix, we can denote its determinant by $\det A$. Two matrices A and B are said to be *equal*, and we write $A = B$, if they (i) have the same number of rows, (ii) have the same number of columns, and (iii) have equal elements in corresponding positions. We can multiply matrices by scalars (numbers), and we can add two matrices which have the same number of rows and the same number of columns. For example, when P and Q are the matrices displayed above and k is a scalar, we have

(3) $$kP = \begin{pmatrix} 2k & k & 3k \\ -3k & 2k & -k \\ k & -k & k \end{pmatrix}, \quad P + Q = \begin{pmatrix} 5 & 2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

When we multiply a matrix by a scalar k, we multiply each element by k. When we add two matrices, we add them elementwise. These rules are very different from those applicable to determinants. The really crucial step in the development of a useful algebra of matrices is the determination, in terms of two suitable matrices A and B, of a third matrix C which we shall call the product AB of A and B. Let A have n rows and p columns and let B have q rows and n columns, so that the number of rows of A is the same as the number of columns of B. The product AB is then a matrix C having p rows and q columns, the element c_{jk} in the jth row and kth column of C being determined by the formula

(4) $$c_{jk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk},$$

where $a_{j1}, a_{j2}, \cdots, a_{jn}$ are the elements of the jth row of A and $b_{1k}, b_{2k}, \cdots, b_{nk}$ are the elements of the kth column of B. To demonstrate that applications of this ritual are not fearsome, we let A, B, C be the matrices P, Q, R defined above and see how the result

(5) $$\begin{pmatrix} 2 & 1 & 3 \\ -3 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & -2 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 7 & 4 & -6 \\ -7 & -6 & 9 \\ 2 & 3 & -4 \end{pmatrix}$$

is obtained. To obtain the element 7 in the first row and the first column of the last matrix, we run one finger across the first row of the first matrix, and at the same time, run another finger down the first column of the second matrix and (with regret that we do not have three hands) write the sum

(6) $$2\cdot 3 + 1\cdot 1 + 3\cdot 0 = 7$$

of the products of the elements that our fingers encounter. To obtain the element 4 in the *first row* and *second column* of the last matrix, we apply the fingers to the *first row* of the first matrix and the *second column* of the second matrix and obtain

(7) $$2\cdot 1 + 1\cdot(-1) + 3\cdot 1 = 4.$$

To obtain the element 3 in the *third row* and *second column* of the third matrix, we follow the *third row* of the first matrix and the *second column* of the second matrix to obtain

(8) $$1\cdot 1 + (-1)(-1) + 1\cdot 1 = 3.$$

Nine such excursions suffice to work out the product of two matrices of order 3. Only three such excursions suffice to give the general formula

(9) $$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{pmatrix}$$

or the special case

$$\begin{pmatrix} 3 & 1 & -2 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}. \tag{10}$$

If we let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \qquad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \tag{11}$$

then (9) shows that the whole system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= y_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= y_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= y_3 \end{aligned} \tag{12}$$

is equivalent to the single matrix equation $AX = Y$. It is standard practice to think of X and Y as being vectors having scalar components x_1, x_2, x_3 and y_1, y_2, y_3 and to think of the matrix A as being an *operator* (or transformer) which *transforms* (or carries or converts) the vector X into the vector Y. In this and other contexts, matrices have very many important applications. One who wishes an easy exercise can prove that the formula

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{13}$$

is valid whenever the rows of the first matrix are the scalar components of three orthonormal vectors. The last matrix in (13) is called the identity matrix I (of order 3) because $AI = IA = A$ whenever A is a square matrix (of order 3). If A is a square matrix of order n and $\det A \neq 0$, then there is a unique (that is, one and only one) matrix B such that $AB = I$, where I is the identity matrix of order n. This matrix B is such that $AB = BA = I$. It is called the *inverse* of A and is denoted by A^{-1}. If $AX = Y$ and $\det A \neq 0$, then $X = A^{-1}Y$. If $\det A = 0$, the matrix A cannot have an inverse because

$$\det (AB) = (\det A)(\det B) \tag{14}$$

when A and B are square matrices of the same order and, moreover, $\det I = 1$. One who is really ambitious may attack the famous eigenvalue problem for matrices. The problem is to start with a given square matrix A and learn about the scalars λ (the eigenvalues) and the nonzero vectors X (the eigenvectors) for which $AX = \lambda X$.

14 This problem involves square matrices of order 2; analogous results hold for square matrices of greater order. Suppose

$$\begin{aligned} z_1 &= a_{11}y_1 + a_{12}y_2 \qquad & y_1 &= b_{11}x_1 + b_{12}x_2 \\ z_2 &= a_{21}y_1 + a_{22}y_2, \qquad & y_2 &= b_{21}x_1 + b_{22}x_2. \end{aligned}$$

Let

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},$$

so that $z = Ay$, $y = Bx$, and $z = A(Bx)$. Show that $z = Cx$, where the matrix C is the product of A and B, that is, $C = AB$. *Remark:* The result shows that products of square matrices are defined in such a way that $A(Bx) = (AB)x$.

15 A two-by-two matrix of numbers a_{nk} determines the system of equations

$$\begin{aligned} x' &= a_{11}x + a_{12}y \\ y' &= a_{21}x + a_{22}y \end{aligned}$$

which transforms a given point (x,y) into its transform (x',y'). Let three points (x_1,y_1), (x_2,y_2), (x_3,y_3) have transforms (x_1',y_1'), (x_2',y_2'), (x_3',y_3'). As an exercise in multiplying matrices (or determinants), prove that

$$\begin{aligned} \begin{vmatrix} x_1' & y_1' & 1 \\ x_2' & y_2' & 1 \\ x_3' & y_3' & 1 \end{vmatrix} &= \begin{vmatrix} a_{11}x_1 + a_{12}y_1 & a_{21}x_1 + a_{22}y_1 & 1 \\ a_{11}x_2 + a_{12}y_2 & a_{21}x_2 + a_{22}y_2 & 1 \\ a_{11}x_3 + a_{12}y_3 & a_{21}x_3 + a_{22}y_3 & 1 \end{vmatrix} \\ &= \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \begin{vmatrix} a_{11} & a_{21} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix}. \end{aligned}$$

Remark: With the aid of the results of this problem, we can prove some theorems in geometry. Let $D = a_{11}a_{22} - a_{12}a_{21}$, so that D is the determinant of the matrix of the transformer. The area $|T'|$ of the triangular region having vertices at the transforms is equal to $|D|$ times the area $|T|$ of the triangular region having vertices at the original points. The orientation (clockwise or counterclockwise) is preserved if $D > 0$ and is reversed if $D < 0$. If $D = 0$, the transforms are collinear. The transformer conserves areas if and only if $|D| = 1$, that is, $D = 1$ or $D = -1$. If the transformer is isometric (conserves distances), then it also conserves areas, and hence $|D| = 1$. These results can be extended to give information about transforms of oriented simplexes having four ordered vertices (x_k,y_k,z_k), $k = 1, 2, 3, 4$, in E_3. When

$$D_1 = \begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}, \qquad D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$

the simplex is (by definition) positively oriented when $D_1 > 0$ and negatively oriented when $D_1 < 0$. The transformer conserves volumes if and only if $D = 1$ or $D = -1$, and it preserves orientation if and only if $D > 0$.

2.6 *Vector products and changes of coordinates in* E_3

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in E_3 having scalar components u_1, u_2, u_3 and v_1, v_2, v_3 with respect to a right-handed x, y, z coordinate system endowed with the usual unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. Then

$$\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}, \qquad \mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}. \tag{2.61}$$

The vector (or cross) product has been defined by the formula

$$\mathbf{u} \times \mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \sin\theta\mathbf{n} \tag{2.611}$$

involving the angle θ and the unit normal (thumb vector) $\mathbf{n}$ of Figure 2.612. We now work out a convenient formula which gives $\mathbf{u} \times \mathbf{v}$ in

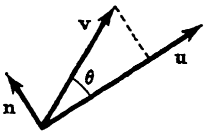


Figure 2.612

terms of the scalar components of $\mathbf{u}$ and $\mathbf{v}$. Remembering that the vector product of two vectors depends upon the order of the factors, we shall be very careful. We have

$$\begin{aligned}\mathbf{u} \times \mathbf{v} &= (u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}) \times (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \\ &= u_1\mathbf{i} \times (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \\ &\quad + u_2\mathbf{j} \times (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \\ &\quad + u_3\mathbf{k} \times (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k})\end{aligned}$$

so

$$\begin{aligned}\mathbf{u} \times \mathbf{v} = {} & u_1v_1\mathbf{i} \times \mathbf{i} + u_1v_2\mathbf{i} \times \mathbf{j} + u_1v_3\mathbf{i} \times \mathbf{k} \\ & + u_2v_1\mathbf{j} \times \mathbf{i} + u_2v_2\mathbf{j} \times \mathbf{j} + u_2v_3\mathbf{j} \times \mathbf{k} \\ & + u_3v_1\mathbf{k} \times \mathbf{i} + u_3v_2\mathbf{k} \times \mathbf{j} + u_3v_3\mathbf{k} \times \mathbf{k}.\end{aligned}$$

With the aid of the helpful fact about the vector product of two consecutive vectors in the parade **ijkijk**, given in (2.234), we obtain the unlovely formula

(2.613) $\mathbf{u} \times \mathbf{v} = \mathbf{i}(u_2v_3 - u_3v_2) - \mathbf{j}(u_1v_3 - u_3v_1) + \mathbf{k}(u_1v_2 - u_2v_1).$

This looks better when we put it in the form

$$\mathbf{u} \times \mathbf{v} = \mathbf{i}\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} - \mathbf{j}\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \mathbf{k}\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}.$$

Our next step is to allow ourselves the liberty of putting vectors into the first row of a determinant so we may put this in the form

(2.62)
$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

which is very easily remembered. It is the fashion to remember (2.62) and to expand the determinant in terms of elements of the first row whenever this is desirable. When we must find the vector product of two vectors $\mathbf{u}$ and $\mathbf{v}$ defined by

$$\mathbf{u} = \mathbf{i} - 2\mathbf{j} + 3\mathbf{k}, \qquad \mathbf{v} = 2\mathbf{i} - \mathbf{j} - \mathbf{k},$$

all persons except typists and printers are happy to solve the problem neatly by writing

$$\mathbf{u}\times\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & -2 & 3 \\ 2 & -1 & -1 \end{vmatrix} = 5\mathbf{i} + 7\mathbf{j} + 3\mathbf{k}.$$

We should all know enough to be able to guard against computational errors by observing that our answer is perpendicular to $\mathbf{u}$ because $5 - 14 + 9 = 0$ and is perpendicular to $\mathbf{v}$ because $10 - 7 - 3 = 0$.

To exhibit an application of vector products to a problem in geometry, we suppose that we are given two orthonormal vectors $\mathbf{i}'$ and $\mathbf{j}'$ in E_3 (this means that $\mathbf{i}'$ and $\mathbf{j}'$ are unit vectors and are orthogonal or perpendicular) and are required to find a third vector $\mathbf{k}'$ such that the three vectors $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ constitute a right-handed orthonormal system. The answer is given by the formula $\mathbf{k}' = \mathbf{i}'\times\mathbf{j}'$. This is so because

$$\mathbf{i}'\times\mathbf{j}' = |\mathbf{i}'|\,|\mathbf{j}'|\sin\theta\mathbf{n} = \mathbf{n}, \tag{2.621}$$

where $\mathbf{n}$ is the unit thumb vector, and this is exactly what $\mathbf{k}'$ must be. This problem can be put in a different form. Suppose we are given a right-handed rectangular x, y, z coordinate system endowed with the usual orthonormal vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. Suppose further we know the scalar components a_{11}, a_{12}, a_{13} and a_{21}, a_{22}, a_{23} of two orthonormal vectors $\mathbf{i}'$ and $\mathbf{j}'$ so that the coefficients in the first two of the equations

$$\begin{aligned} \mathbf{i}' &= a_{11}\mathbf{i} + a_{12}\mathbf{j} + a_{13}\mathbf{k} \\ \mathbf{j}' &= a_{21}\mathbf{i} + a_{22}\mathbf{j} + a_{23}\mathbf{k} \\ \mathbf{k}' &= a_{31}\mathbf{i} + a_{32}\mathbf{j} + a_{33}\mathbf{k} \end{aligned} \tag{2.63}$$

are known. The problem is to determine the numbers a_{31}, a_{32}, a_{33} so that the three vectors $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ will form a right-handed orthonormal system. These numbers are determined from the formula

$$a_{31}\mathbf{i} + a_{32}\mathbf{j} + a_{33}\mathbf{k} = \mathbf{k}' = \mathbf{i}'\times\mathbf{j}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{vmatrix}$$

and the problem is solved or almost solved. To write more formulas is somewhat anticlimactic, but we can do it. Equating coefficients of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ gives

$$\begin{aligned} a_{31} &= a_{12}a_{23} - a_{13}a_{22} \\ a_{32} &= a_{13}a_{21} - a_{11}a_{23} \\ a_{33} &= a_{11}a_{22} - a_{12}a_{21} \end{aligned} \tag{2.631}$$

and then the problem is surely solved.

Everyone should read the remainder of this section, but teachers who want to confine attention to other topics may inform their charges that hasty reading and preliminary ideas will be satisfactory preparation for future encounters with the material.

Here we begin to explore some of the reasons why the system (2.63) of equations is important. Suppose we have, as in Figure 2.64, two

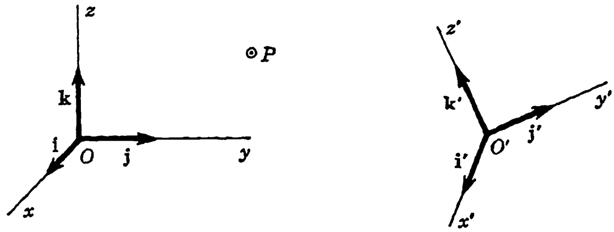


Figure 2.64

right-handed rectangular coordinate systems in E_3. The x, y, z coordinate system having origin at O and bearing an orthonormal set $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ of vectors is shown on the left. The x', y', z' coordinate system having origin at O' and bearing an orthonormal set $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ of vectors is shown on the right. Our first task is to study the important systems of equations

$$
(2.65)\qquad
\begin{aligned}
\mathbf{i}' &= a_{11}\mathbf{i} + a_{12}\mathbf{j} + a_{13}\mathbf{k}, \qquad & \mathbf{i} &= a_{11}\mathbf{i}' + a_{21}\mathbf{j}' + a_{31}\mathbf{k}'\\
\mathbf{j}' &= a_{21}\mathbf{i} + a_{22}\mathbf{j} + a_{23}\mathbf{k}, & \mathbf{j} &= a_{12}\mathbf{i}' + a_{22}\mathbf{j}' + a_{32}\mathbf{k}'\\
\mathbf{k}' &= a_{31}\mathbf{i} + a_{32}\mathbf{j} + a_{33}\mathbf{k}, & \mathbf{k} &= a_{13}\mathbf{i}' + a_{23}\mathbf{j}' + a_{33}\mathbf{k}'
\end{aligned}
$$

that relate the vectors in the two orthonormal sets. We observe a fact that can be considered to be remarkable even when we know the reason for it: the coefficients in the system of equations obtained by solving the first system for $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are easily written down by interchanging the rows and columns in the first system. The reason is simple. The orthonormality of the vectors implies that, in each system, a_{11} is $\mathbf{i}'\cdot\mathbf{i}$, a_{12} is $\mathbf{i}'\cdot\mathbf{j}$, a_{13} is $\mathbf{i}'\cdot\mathbf{k}$, a_{21} is $\mathbf{j}'\cdot\mathbf{i}$, and so on until, finally, a_{33} is $\mathbf{k}'\cdot\mathbf{k}$. The numerical coefficients in each row (and hence also in each column) in the right member of each system are the scalar components of a unit vector. The nine coefficients are cosines of direction angles, but we carefully avoid attempts to work out formulas by means of figures showing the nine angles.

As soon as we look at the point P of Figure 2.64, we realize that P has two sets of coordinates, there being a set x, y, z for the unprimed or *old* coordinate system and another set x', y', z' for the primed or *new* coordinate system. If we know enough about the relative positions of the two coordinate systems, we should be able to find one set of coordinates when we know the other set. We shall solve this problem with the aid of vectors. To specify the relation between the two coordinate systems, we suppose that, with reference to the unprimed coordinate system, the coordinates of O' are x_0, y_0, z_0, so that

$$\overrightarrow{OO'} = x_0\mathbf{i} + y_0\mathbf{j} + z_0\mathbf{k}, \tag{2.66}$$

and that the unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ and $\mathbf{i}', \mathbf{j}', \mathbf{k}'$ of the two coordinate systems are related by the fundamental formulas (2.65). It is now surprisingly easy to solve our problem. Let P be a point in E_3 having unprimed coordinates x, y, z and primed coordinates x', y', z', so that

$$\overrightarrow{OP} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \qquad \overrightarrow{O'P} = x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'. \tag{2.661}$$

Putting the formula $\overrightarrow{OP} = \overrightarrow{OO'} + \overrightarrow{O'P}$ in the form $\overrightarrow{OP} - \overrightarrow{OO'} = \overrightarrow{O'P}$ then gives

$$(x - x_0)\mathbf{i} + (y - y_0)\mathbf{j} + (z - z_0)\mathbf{k} = x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'. \tag{2.662}$$

An expression giving the right side in terms of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ is obtained by multiplying the members of the first three equations in (2.65) by x', y', z', respectively, and adding the results. The coefficients of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ turn out to be, respectively, the right members of the equations

$$\begin{aligned} x - x_0 &= a_{11}x' + a_{21}y' + a_{31}z' \\ y - y_0 &= a_{12}x' + a_{22}y' + a_{32}z' \\ z - z_0 &= a_{13}x' + a_{23}y' + a_{33}z'. \end{aligned}$$

These equations therefore result from equating the coefficients of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ in (2.662). Transposing gives the formulas

$$\begin{aligned} x &= x_0 + a_{11}x' + a_{21}y' + a_{31}z' \\ y &= y_0 + a_{12}x' + a_{22}y' + a_{32}z' \\ z &= z_0 + a_{13}x' + a_{23}y' + a_{33}z' \end{aligned} \tag{2.67}$$

which express the unprimed coordinates of a point in terms of the primed coordinates of the point. A very similar procedure gives the formulas

$$\begin{aligned} x' &= x_0' + a_{11}x + a_{12}y + a_{13}z \\ y' &= y_0' + a_{21}x + a_{22}y + a_{23}z \\ z' &= z_0' + a_{31}x + a_{32}y + a_{33}z \end{aligned} \tag{2.671}$$

which express the primed coordinates of a point in terms of the unprimed coordinates of the point. The formulas (2.67) and (2.671) are known as the formulas for changes of coordinates. The formulas (2.67) are often used to convert an equation involving coordinates x, y, z into a new equation involving new coordinates x', y', z'. As can be expected, it is sometimes far from easy to so determine the numbers x_0, y_0, z_0 and a_{pq} in (2.67) that the new equation will have the simplest possible form. For the present we do not need to know much about these matters, but we should know that there are situations in which one particular coordinate system is better than others and that there exist formulas relating the coordinates in two different coordinate systems.

It is sometimes said that fundamental problems in analytic geometry are not adequately covered in textbooks that combine the study of analytic geometry and calculus. Much more analytic geometry will appear later in this textbook. Meanwhile, we consider a fundamental problem in analytic geometry that is sometimes ignored in elementary geometry books. Suppose we say, with reference to some rectangular x, y, z coordinate system, that a set S in E_3 is a *quadric surface* if it is the set whose points $P(x,y,z)$ satisfy an equation of the form

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0, \tag{2.68}$$

where the coefficients A, B, C, D, E, F are not all zero. Our big question is the following. Can it happen that Miss White chooses a particular x, y, z coordinate system and finds that a particular set S^* is a quadric surface because there do exist coefficients A, B, $\cdots$, F not all 0 such that S^* is the set of points $P(x,y,z)$ for which $Ax^2 + By^2 + \cdots = 0$, while, at the same time, Mr. Black chooses another x, y, z coordinate system and finds that the same set S^* is *not* a quadric surface because for his system the required coefficients do not exist? If the answer is affirmative, then the above definition of quadric surface and the above set S^* should be placed in the museum of the SPC (Society for the Promotion of Confusion). It can be shown that the answer is negative and hence that the definition of quadric surface does make sense. To do this, we let x', y', z' denote the coordinates of Mr. Black and substitute the values of x, y, z from (2.67) into (2.68) to find what the equation of S^* will be in the coordinates of Mr. Black. The critical equation turns out to be

$$\begin{aligned} A'x'^2 + B'y'^2 + C'z'^2 + D'x'y' + E'x'z' + F'y'z' \\ + G'x' + H'y' + I'z' + J' = 0, \end{aligned} \tag{2.681}$$

where

$$A' = Aa_{11}^2 + Ba_{12}^2 + Ca_{13}^2 + Da_{11}a_{12} + Ea_{11}a_{13} + Fa_{12}a_{13},$$

and formulas for the other coefficients can be written out. Proof that the coefficients A', B', C', D', E', F' are not all zero can be based upon the fact that substituting the expression for x', y', z' from (2.671) into (2.681) must yield the original equation (2.68). If A', B', C', D', E', F' were all zero, this substitution would show that A, B, C, D, E, F are all zero, and this is not so. The principle involved is the following: As we see from (2.671), a change from coordinates x, y, z to x', y', z' cannot increase the degree of a polynomial in x, y, z. Moreover, the change cannot decrease the degree because, as we see from (2.67), the change from x', y', z' back to x, y, z cannot bring a polynomial of lower degree back to the original polynomial.

So far as we know, some quadric surfaces may be rather complicated things, and it is of interest to know what the intersection S_1 of a quadric surface and a plane π may be. Such a set S_1 is a *quadric section.* A little thought can save us a lot of trouble. We can introduce a coordinate system in such a way that the plane π is the plane having the equation $z = 0$. The equation of the quadric surface must have the form (2.68). Putting $z = 0$ in (2.68) shows that the quadric section must be the set of points $P(x,y)$ in the xy plane whose coordinates satisfy the equation

$$Ax^2 + By^2 + Dxy + Gx + Hy + J = 0. \tag{2.682}$$

Quadric surfaces and quadric sections will be studied later. Meanwhile, we make some remarks that may be at least partially understood. The family of quadric surfaces includes spheres, circular cylinders, circular cones, ellipsoids, various kinds of paraboloids and hyperboloids, and, in addition, assorted degenerate things such as empty sets, lines, planes, and pairs of planes. The equations $z^2 + 1 = 0$, $x^2 + y^2 = 0$, $z^2 = 0$, and $z^2 - 1 = 0$ do have the form (2.68). The family of quadric sections includes circles, parabolas, ellipses, hyperbolas, and, in addition, such degenerate things as the empty set, points, lines, pairs of lines, and the whole plane.

Problems 2.69

1 Supposing that $\mathbf{u}$ and $\mathbf{v}$ are vectors in E_3 having scalar components u_1, u_2, u_3 and v_1 v_2, v_3, tell how $\mathbf{u} \times \mathbf{v}$ can be expressed as a determinant. *Ans.:* (2.62).

2 Calculate $\mathbf{w} = \mathbf{u} \times \mathbf{v}$ and check your answer by showing that $\mathbf{w}\cdot\mathbf{u} = 0$ and $\mathbf{w}\cdot\mathbf{v} = 0$ when

(*a*) $\mathbf{u} = 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$, $\quad \mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$

(*b*) $\mathbf{u} = 2\mathbf{i} - 3\mathbf{j} + -4\mathbf{k}$, $\quad \mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$

(*c*) $\mathbf{u} = \mathbf{i} + \mathbf{j}$, $\quad \mathbf{v} = \mathbf{i} + \mathbf{j} + \mathbf{k}$

(*d*) $\mathbf{u} = \mathbf{i}$, $\quad \mathbf{v} = \mathbf{i} + \mathbf{j}$

3 Two unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ have their tails at the same point P on a sphere which we consider to be the surface of an idealized earth. Suppose that P is neither the north pole nor the south pole of the earth, that $\mathbf{u}_1$ points east, and that $\mathbf{u}_2$ points north. Find the direction of $\mathbf{u}_1 \times \mathbf{u}_2$. *Ans.:* Up.

4 Show that the vectors $\mathbf{i}'$ and $\mathbf{j}'$ defined by

$$\begin{aligned} \mathbf{i}' &= \tfrac{1}{3}(\mathbf{i} - 2\mathbf{j} + 2\mathbf{k}) \\ \mathbf{j}' &= \tfrac{1}{3}(2\mathbf{i} - \mathbf{j} - 2\mathbf{k}) \\ \mathbf{k}' &= a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \end{aligned}$$

constitute a two-dimensional orthonormal system and then so determine a, b, c that the three vectors constitute a right-handed three-dimensional orthonormal system. *Ans.:* $\mathbf{k}' = \frac{1}{3}(2\mathbf{i} + 2\mathbf{j} + \mathbf{k})$.

5 Show that the vectors

$$\begin{aligned}\mathbf{u}_1 &= \cos\phi\sin\theta\,\mathbf{i} + \sin\phi\sin\theta\,\mathbf{j} + \cos\theta\,\mathbf{k}\\ \mathbf{u}_2 &= \cos\phi\cos\theta\,\mathbf{i} + \sin\phi\cos\theta\,\mathbf{j} - \sin\theta\,\mathbf{k}\\ \mathbf{u}_3 &= -\sin\phi\,\mathbf{i} + \cos\phi\,\mathbf{j}\end{aligned}$$

in the order $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ constitute a right-handed orthonormal system.

6 If

$$\begin{aligned}\overrightarrow{OP_1} &= 2\mathbf{i} + 3\mathbf{j} + 2\mathbf{k}\\ \overrightarrow{OP_2} &= 3\mathbf{i} + 2\mathbf{j} + \mathbf{k}\\ \overrightarrow{OP_3} &= 2\mathbf{i} + \mathbf{j} + \mathbf{k},\end{aligned}$$

find a vector orthogonal to the plane containing P_1, P_2, P_3. *Hint:* The vector $\mathbf{v}$ defined by $\mathbf{v} = \overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}$ is an answer, and $\mathbf{v} = -\mathbf{i} + \mathbf{j} - 2\mathbf{k}$.

7 Check the answer to the preceding problem in the following way. Write the equation of the plane through P_1 orthogonal to $\mathbf{v}$ and then show that P_2 and P_3 lie in this plane.

8 Show that the lines having the equations

$$\frac{x-1}{1} = \frac{y+6}{2} = \frac{z+10}{3}, \qquad \frac{x-6}{2} = \frac{y+1}{-1} = \frac{z+5}{-4}$$

intersect. Then find equations of the line perpendicular to both at their point of intersection. *Solution:* The vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ running from the origin to points P and Q on the two lines are

$$\begin{aligned}\overrightarrow{OP} &= (1+t)\mathbf{i} + (-6+2t)\mathbf{j} + (-10+3t)\mathbf{k}\\ \overrightarrow{OQ} &= (6+2u)\mathbf{i} + (-1-u)\mathbf{j} + (-5-4u)\mathbf{k},\end{aligned}$$

where t and u are scalars that depend upon P and Q. Equating these vectors shows that they coincide when $t = 3$, $u = -1$ and hence that the given lines intersect at the point R for which

$$\overrightarrow{OR} = 4\mathbf{i} - \mathbf{k}.$$

Thus R is the point $(4,0,-1)$. This result is easily checked. The vector

$$\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 1 & 2 & 3\\ 2 & -1 & -4\end{vmatrix} = -5\mathbf{i} + 10\mathbf{j} - 5\mathbf{k}$$

is orthogonal to vectors on the given lines and hence the equations

$$\frac{x-4}{1} = \frac{y-0}{-2} = \frac{z+1}{1}$$

are equations of the required perpendicular line.

9 Prove that each vector $\mathbf{v}$ satisfies the equation

$$\mathbf{i} \times (\mathbf{v} \times \mathbf{i}) + \mathbf{j} \times (\mathbf{v} \times \mathbf{j}) + \mathbf{k} \times (\mathbf{v} \times \mathbf{k}) = 2\mathbf{v}.$$

10 Show that $(\mathbf{i} \times \mathbf{j}) \times \mathbf{j} \neq \mathbf{i} \times (\mathbf{j} \times \mathbf{j})$.

11 Prove that if **u** and **v** and **w** are vectors having scalar components u_1, u_2, u_3 and v_1, v_2, v_3 and w_1, w_2, w_3, then

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) = (u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k})\cdot\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \tag{1}$$

and hence

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}. \tag{2}$$

Remark: The number $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ is called the *scalar triple product* of the three vectors, and we shall see how it is related to volumes of tetrahedrons and parallelepipeds. Let **v** and **w** be nonzero nonparallel vectors having their tails at a point A as in Figure 2.691. Then

$$\mathbf{v}\times\mathbf{w} = |\mathbf{v}|\,|\mathbf{w}|\sin\theta\,\mathbf{n} = 2|T_2|\mathbf{n}, \tag{3}$$

where θ is the angle between **v** and **w**, **n** is the unit normal determined by the right-hand rule, and $|T_2|$ is the area of the two dimensional triangle T_2 of which

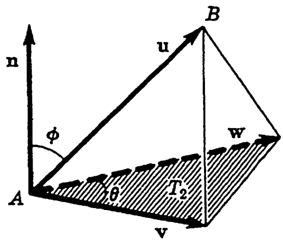


Figure 2.691

v and **w** form two sides. Let **u** be a third vector which has its tail at A and makes the angle ϕ with **n**. In case $0 \leqq \phi < \pi/2$, the number $\mathbf{u}\cdot\mathbf{n}$ is the distance from the tip of **u** to the plane of the vectors **v** and **w**. The volume V of the tetrahedron having base T_2 and opposite vertex B is therefore given by the formula

$$V = \tfrac{1}{3}(\mathbf{u}\cdot\mathbf{n})|T_2|. \tag{4}$$

Hence

$$V = \tfrac{1}{6}\mathbf{u}\cdot(2T_2\mathbf{n}) = \tfrac{1}{6}\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}). \tag{5}$$

The volume of the tetrahedron is half the volume of the pyramid whose vertex is B and whose base is the parallelogram of which **v** and **w** are two adjacent sides. The volume of this pyramid is one-third the volume V_P of the parallelepiped of which **u**, **v**, **w** are adjacent edges. Therefore,

$$V_P = \mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}). \tag{6}$$

Thus, when $0 \leqq \phi < \pi/2$, the scalar triple product $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ is the volume of the parallelepiped of which **u**, **v**, **w** are three adjacent edges. When $\phi = \pi/2$, the vector **u** lies in the plane of **v** and **w** and the scalar triple product is zero.

When $\pi/2 < \phi < \pi$, the scalar triple product is the negative of the volume of the parallelepiped.

12 Prove that if $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ are two right-handed orthonormal sets of vectors, then the determinant of the coefficients in the system

$$\begin{aligned} \vec{\mathbf{i}'} &= a_{11}\mathbf{i} + a_{12}\mathbf{j} + a_{13}\mathbf{k} \\ \vec{\mathbf{j}'} &= a_{21}\mathbf{i} + a_{22}\mathbf{j} + a_{23}\mathbf{k} \\ \vec{\mathbf{k}'} &= a_{31}\mathbf{i} + a_{32}\mathbf{j} + a_{33}\mathbf{k} \end{aligned}$$

must be $+1$. *Outline of solution:* The hypotheses imply that $\mathbf{j}' \times \mathbf{k}' = \mathbf{i}'$ and hence that $\mathbf{i}' \cdot (\mathbf{j}' \times \mathbf{k}') = 1$. But

$$\mathbf{i}' \cdot (\mathbf{j}' \times \mathbf{k}') = (a_{11}\mathbf{i} + a_{12}\mathbf{j} + a_{13}\mathbf{k}) \cdot \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \Delta,$$

where Δ is the determinant of the coefficients, and the result follows.

13 It is clear from geometrical considerations that if $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ are two right-handed orthonormal sets of vectors in E_3 for which $\mathbf{k}' = \mathbf{k}$, then there must be an angle ϕ such that the "new" vectors $\mathbf{i}'$ and $\mathbf{j}'$ are related to the "old" vectors $\mathbf{i}$ and $\mathbf{j}$ as in Figure 2.692. The new vectors and new $\mathbf{x}'$ axis and new $\mathbf{y}'$

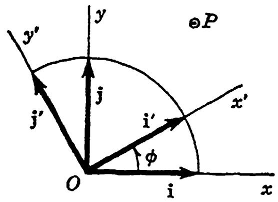


Figure 2.692

axis are dashed in the figure, and the vectors $\mathbf{k}$ and $\mathbf{k}'$ are not shown. It is easy to see, with the aid of the figure, that

$$(1) \qquad \begin{aligned} \mathbf{i}' &= (\cos\phi)\mathbf{i} + (\sin\phi)\mathbf{j} \\ \mathbf{j}' &= -(\sin\phi)\mathbf{i} + (\cos\phi)\mathbf{j}. \end{aligned}$$

Check up on this story by using (1) and vector methods to prove that, whatever ϕ may be, it is actually true that $|\mathbf{i}'| = 1$, $|\mathbf{j}'| = 1$, $\mathbf{i}' \cdot \mathbf{j}' = 0$, and $\mathbf{i}' \times \mathbf{j}' = \mathbf{k}$. Solve the equations (1) for $\mathbf{i}$ and $\mathbf{j}$ to obtain the formulas

$$(2) \qquad \begin{aligned} \mathbf{i} &= (\cos\phi)\mathbf{i}' - (\sin\phi)\mathbf{j}' \\ \mathbf{j} &= (\sin\phi)\mathbf{i}' + (\cos\phi)\mathbf{j}' \end{aligned}$$

which give the new vectors in terms of the old. Observe that changing the sign of ϕ converts one system of equations into the other. Finally, show that if P lies in the plane of Figure 2.692 and if

$$(3) \qquad \overrightarrow{OP} = x\mathbf{i} + y\mathbf{j} = x'\mathbf{i}' + y'\mathbf{j}',$$

then

$$(4) \qquad \begin{aligned} \overrightarrow{OP} &= x'[(\cos\phi)\mathbf{i} + (\sin\phi)\mathbf{j}] + y'[-(\sin\phi)\mathbf{i} + (\cos\phi)\mathbf{j}] \\ &= [x'\cos\phi - y'\sin\phi]\mathbf{i} + [x'\sin\phi + y'\cos\phi]\mathbf{j} \end{aligned}$$

and therefore

$$x = x' \cos \phi - y' \sin \phi \qquad (5)$$
$$y = x' \sin \phi + y' \cos \phi.$$

Remark: The formulas (5) are, perhaps unwisely, called "formulas for rotation of axes" in E_2. Actually, they are used to convert equations involving "old" coordinates x, y into new (and sometimes simpler) equations involving new coordinates x', y'.

14 Supposing that $\mathbf{u}$ and $\mathbf{v}$ are nonzero noncollinear vectors, show that the vector

$$\frac{(\mathbf{u} \times \mathbf{v}) \times \mathbf{u}}{|(\mathbf{u} \times \mathbf{v}) \times \mathbf{u}|}$$

is a unit vector which lies in the plane of $\mathbf{u}$ and $\mathbf{v}$ and is orthogonal (or perpendicular) to $\mathbf{u}$.

15 Cultivate some useful skills by following instructions and paying particular attention to steps that seem to be worthy of notice. Draw vectors $\overrightarrow{PP_2}$ and $\overrightarrow{PP_3}$ and then draw the angle θ and the unit normal $\mathbf{n}$ that appear in the definition of $\overrightarrow{PP_2} \times \overrightarrow{PP_3}$. Show that the area A of the parallelogram having adjacent sides on $\overrightarrow{PP_2}$ and $\overrightarrow{PP_3}$ is

$$A = |\overrightarrow{PP_2}|\,|\overrightarrow{PP_3}| \sin \theta.$$

Draw another vector $\overrightarrow{PP_1}$ and show that the distance d from P_1 to the plane of $\overrightarrow{PP_2}$ and $\overrightarrow{PP_3}$ is

$$d = \pm \overrightarrow{PP_1}\cdot\mathbf{n},$$

where the sign is so taken that $d \geqq 0$. Then, depending upon circumstances, remember or learn that $V = Ad$, where V is the volume of the parallelepiped having adjacent edges on $\overrightarrow{PP_1}$, $\overrightarrow{PP_2}$, $\overrightarrow{PP_3}$ if $d \neq 0$ and $V = 0$ if P_1 lies in the plane of $\overrightarrow{PP_2}$ and $\overrightarrow{PP_3}$. Use this to show that

$$V = \pm \overrightarrow{PP_1}\cdot(\overrightarrow{PP_2} \times \overrightarrow{PP_3}).$$

Supposing that P has coordinates x, y, z and that the points P_k have coordinates x_k, y_k, z_k, show that

$$V = \pm[(x_1 - x)\mathbf{i} + (y_1 - y)\mathbf{j} + (z_1 - z)\mathbf{k}]\cdot\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_2 - x & y_2 - y & z_2 - z \\ x_3 - x & y_3 - y & z_3 - z \end{vmatrix}$$

and hence that

$$V = \pm \begin{vmatrix} x_1 - x & y_1 - y & z_1 - z \\ x_2 - x & y_2 - y & z_2 - z \\ x_3 - x & y_3 - y & z_3 - z \end{vmatrix}.$$

Use this to show that

$$V = \mp \begin{vmatrix} x & y & z & 1 \\ x_1 - x & y_1 - y & z_1 - z & 0 \\ x_2 - x & y_2 - y & z_2 - z & 0 \\ x_3 - x & y_3 - y & z_3 - z & 0 \end{vmatrix} = \mp \begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix}.$$

16 Prove that a line which is not completely contained in a quadric surface can intersect the quadric surface at most twice. *Solution:* Suppose the coordinate system is so chosen that the given line has the equations $y = z = 0$, and let the equation of the quadric surface be (2.68). A point $(x,0,0)$ then lies on the surface if and only if $Ax^2 + Gx + J = 0$. If there is an x for which this equation is not satisfied, then at least one of A, G, J must be different from 0 and there are at most two values of x for which the equation is satisfied.

17 The purpose of this long problem is to develop ideas about the transversals of three given skew (no two lying in the same plane) lines P_1P_2, P_3P_4, P_5P_6. A line L is called a transversal of the given lines if it intersects the lines P_1P_2, P_3P_4, P_5P_6 at points Q, Q_1, Q_2 as in Figure 2.693. For each $k = 1, 2, \cdots, 6$ the coordi-

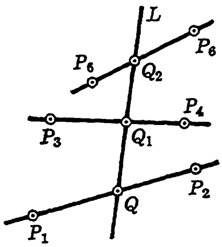


Figure 2.693

nates (x_k,y_k,z_k) of P_k are given numbers and we want information about the coordinates of Q, Q_1, Q_2. The latter coordinates are determined with the aid of numbers λ, λ_1, λ_2 for which

$$(1) \qquad \overrightarrow{P_1Q} = \lambda\overrightarrow{P_1P_2}, \qquad \overrightarrow{P_3Q_1} = \lambda_1\overrightarrow{P_3Q_4}, \qquad \overrightarrow{P_5Q_2} = \lambda_2\overrightarrow{P_5P_6}$$

Our first step is to select a number λ (or to think of λ as being "fixed") and ask whether a transversal through Q exists. If λ is so chosen that the plane π_1 containing Q, P_3, P_4 does not intersect the line P_5P_6, or intersects the line P_5P_6 at a point Q_2 for which the line QQ_2 is parallel to the line P_3P_4, then no transversal through Q exists. Henceforth, we suppose that λ is not so unhappily chosen. A transversal through Q is then obtained by drawing the line L through Q and the point Q_2 where π_1 intersects the line Q_5Q_6. The value of λ_2 can be calculated in terms of λ from the equation

$$(2) \qquad \begin{vmatrix} x_5 + \lambda_2(x_6 - x_5) & y_5 + \lambda_2(y_6 - y_5) & z_5 + \lambda_2(z_6 - z_5) & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \\ x_1 + \lambda(x_2 - x_1) & y_1 + \lambda(y_2 - y_1) & z_1 + \lambda(z_2 - z_1) & 1 \end{vmatrix} = 0$$

which says that Q_2 is the point on the line P_5P_6 which lies in the plane π_1 containing P_3, P_4, Q. Similarly, the value of λ_1 can be calculated in terms of λ from the equation

$$(3) \qquad \begin{vmatrix} x_3 + \lambda_1(x_4 - x_3) & y_3 + \lambda_1(y_4 - y_3) & z_3 + \lambda_1(z_4 - z_3) & 1 \\ x_5 & y_5 & z_5 & 1 \\ x_6 & y_6 & z_6 & 1 \\ x_1 + \lambda(x_2 - x_1) & y_1 + \lambda(y_2 - y_1) & z_1 + \lambda(z_2 - z_1) & 1 \end{vmatrix} = 0$$

which says that Q_1 is the point on the line P_3P_4 which lies in the plane π_2 containing P_5, P_6, Q. Thus the required coordinates of Q, Q_1, Q_2 are determined in terms of λ. *Remark:* Study of the set S of points $P(x,y,z)$ that lie upon transversals can be very interesting. If P lies upon the transversal through Q, then, for some scalar μ,

$$\overrightarrow{OP} = (1 - \mu)\overrightarrow{OQ} + \mu\overrightarrow{OQ_1}. \tag{4}$$

But

$$\overrightarrow{OQ} = (1 - \lambda)\overrightarrow{OP_1} + \lambda\overrightarrow{OP_2}, \tag{5}$$

and, since (3) shows that there are constants A and B for which $\lambda_1 = A\lambda + B$,

$$\overrightarrow{OQ_1} = (1 - [A\lambda + B])\overrightarrow{OP_3} + (A\lambda + B)\overrightarrow{OP_4}. \tag{6}$$

Therefore,

$$\begin{aligned}\overrightarrow{OP} = (1 - \mu)(1 - \lambda)\overrightarrow{OP_1} + (1 - \mu)\lambda\overrightarrow{OP_2} + \mu\overrightarrow{OP_3} \\ - \mu(A\lambda + B)\overrightarrow{OP_3} + \mu(A\lambda + B)\overrightarrow{OP_4}.\end{aligned} \tag{7}$$

Hence there are vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ such that

$$\overrightarrow{OP} = \mathbf{v}_1 + \lambda\mathbf{v}_2 + \mu\mathbf{v}_3 + \lambda\mu\mathbf{v}_4. \tag{8}$$

It follows that there are scalars, $a_1, \cdots, d_3$ such that

$$\begin{cases} x = a_1 + b_1\lambda + c_1\mu + d_1\lambda\mu \\ y = a_2 + b_2\lambda + c_2\mu + d_2\lambda\mu \\ z = a_3 + b_3\lambda + c_3\mu + d_3\lambda\mu. \end{cases} \tag{9}$$

It can be shown that the equations (9) are parametric equations of a quadric surface. In fact, eliminating λ and μ from the equations (9) shows that x, y, z must satisfy an equation of the form (2.68). Thus S is a quadric surface, and we have a quite straightforward procedure for determining its equation in terms of the eighteen given coordinates of the six given points $P_1, P_2, \cdots, P_6$. Students who attain full comprehension of this matter will have passed far beyond the minimum requirements of this course, and they can find the experience to be both enjoyable and beneficial.

18 Those who wish to extend acquaintance with matrix theory should copy the systems of equations in (2.65) and look at them while reading this. Let U and U^T denote the matrices of the coefficients (or scalar components or direction cosines) of the systems so that

$$U = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \qquad U^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}.$$

The matrix U^T is called the *transpose* (or *transposed matrix*) of the matrix U, and this invites us to realize that U^T can be obtained from U by transposing (interchanging) the rows and columns of U or by transposing the elements of U across its main diagonal. The rows of U are scalar components of orthonormal vectors, and the matrix is square. Such matrices are called *unitary* (or *orthonormal*) matrices. Therefore, U is unitary. When U is unitary, an application of the rule for multiplying matrices shows that $UU^T = I$, where I is the unit matrix. Therefore $U^{-1} = U^T$. This is important; the inverse of a unitary matrix U is U^T.

19 This problem requires us to agree with Miss Garnett that methods of analytic geometry can be used to solve a challenging problem that may baffle those who seek more elementary solutions. It is supposed that a, b, c are positive numbers and that the points $A(0,0)$, $B(c,0)$, $C(a,b)$, $D(a + c, b)$ are vertices of a parallelogram. Let a point $E(u,0)$ on the bottom side of the parallelogram be joined to the top vertices C and D and let a point $F(t,b)$ on the top side of the parallelogram be joined to the bottom vertices A and B. The lines EC and FA intersect at P_1, and the lines ED and FB intersect at P_2. The line P_1P_2 meets the side AC at Q_1 and meets the side BD at Q_2. The question then arises whether the distance from Q_1 to A is equal to the distance from Q_2 to D. Elementary geometrical considerations show that the answer will be affirmative if the line P_1P_2 contains the center P_3 of the parallelogram, this center P_3 being the intersection of the diagonals of the parallelogram. Show that, for each $k = 1, 2, 3$, the elements of the kth row of the determinant

$$\begin{vmatrix} \dfrac{tu}{t+u-a} & \dfrac{bu}{t+u-a} & 1 \\ \dfrac{ac+c^2-tu}{a+2c-t-u} & \dfrac{b(c-u)}{a+2c-t-u} & 1 \\ \dfrac{a+c}{2} & \dfrac{b}{2} & 1 \end{vmatrix}$$

are x_k, y_k, 1, where x_k and y_k are the coordinates of P_k. Then prove that the determinant is 0 and hence that the line P_1P_2 actually does contain P_3.

20 Let $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$ be vertices of a triangle such that no two of the vertices lie on a line through the origin. Let λ_1, λ_2, λ_3 be three different numbers, and for each $k = 1, 2, 3$, let Q_k be the point $(\lambda_k x_k, \lambda_k y_k)$. The two triangles $P_1P_2P_3$ and $Q_1Q_2Q_3$ are then *perspective*, the *center of perspectivity* being the origin. The lines P_1P_2 and Q_1Q_2 intersect at a point R_3, the lines P_2P_3 and Q_2Q_3 intersect at a point R_1, and the lines P_3P_1 and Q_3Q_1 intersect at a point R_2. The famous *Desargues theorem* says that the three points R_1, R_2, R_3 lie on a line L. It is easy to sketch figures illustrating the theorem, but proofs are not easily originated. Possessors of sufficient time, paper, and technique may cultivate additional technique by finding the x coordinate of R_3, and then interchange x and y and advance subscripts to discover that the coordinates of R_2 and R_3 are the first two elements of the bottom rows of the determinant in the equation

$$\begin{vmatrix} x & y & 1 \\ \dfrac{(\lambda_3-1)\lambda_1x_1-(\lambda_1-1)\lambda_3x_3}{\lambda_3-\lambda_1} & \dfrac{(\lambda_3-1)\lambda_1y_1-(\lambda_1-1)\lambda_3y_3}{\lambda_3-\lambda_1} & 1 \\ \dfrac{(\lambda_1-1)\lambda_2x_2-(\lambda_2-1)\lambda_1x_1}{\lambda_1-\lambda_2} & \dfrac{(\lambda_1-1)\lambda_2y_2-(\lambda_2-1)\lambda_1x_1}{\lambda_1-\lambda_2} & 1 \end{vmatrix} = 0.$$

This is, therefore, the equation of the line L through R_2 and R_3. Considerable courage is required to show that the coordinates of R_1 satisfy this equation and thus obtain an analytic proof of the Desargues theorem.

3 *Functions, limits, derivatives*

3.1 Functional notation As we progress in a study of a science, it is necessary to become familiar with terminology and notation used for conveying information. One of the most important mathematical words is the word *function*. We may look at Figure 3.11, in which x, y, and z are the lengths of the sides of a triangle and θ is the angle at the vertex V opposite the side of length z, and say that z is a function of x, y, and θ which we shall denote by $f(x,y,\theta)$. By this we mean that when x, y, and θ are given numbers for which $x > 0$, $y > 0$, and $0 < \theta < \pi$, the number z is completely determined and has a value which we denote by the symbol in the right member of the formula

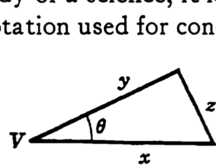


Figure 3.11

$$z = f(x,y,\theta). \tag{3.12}$$

This equation is read "z equals f of x and y and θ." It happens that the law of cosines, which involves one of the more important formulas which should be learned in trigonometry, gives the formula

$$f(x,y,\theta) = \sqrt{x^2 + y^2 - 2xy \cos \theta} \tag{3.13}$$

from which we can compute $f(x,y,\theta)$ when x, y, θ are given numbers. In spite of the fact that numbers do not move, it is sometimes a convenience to think of x, y, θ, z as being "variables" and to think of z as being the "dependent variable" which is a function of the three "independent variables" x, y, θ.

Many examples are more complicated than this, and we can broaden our intellectual horizons by thinking briefly about one of them. It is standard practice to write

$$\begin{aligned} \mathbf{v} &= \mathbf{f}(x,y,z,t) \\ &= f_1(x,y,z,t)\mathbf{i} + f_2(x,y,z,t)\mathbf{j} + f_3(x,y,z,t)\mathbf{k}, \end{aligned} \tag{3.14}$$

where $\mathbf{v}$, a vector, is the velocity of a fluid (which might be air) at the place having rectangular coordinates x, y, z and at time t. We say that $\mathbf{v}$ and its scalar components are functions of the four variables x, y, z, t. We mean that when x, y, z are coordinates of a point in the region being considered and when t is a time (measured in specified units from a specified zero hour) in the time interval being considered, the velocity $\mathbf{v}$ and its scalar components at that place and time are completely determined and that $\mathbf{f}(x,y,z,t)$ denotes the velocity and $f_1(x,y,z,t)$, $f_2(x,y,z,t)$, $f_3(x,y,z,t)$ denote the scalar components.

There are two useful and more or less modern ways of attaching meanings to the symbols $\mathbf{f}$ and f_1 appearing in the above example. One is the dynamic approach and the other is the static approach. In the dynamic approach, $\mathbf{f}$ and f_1 are regarded as *operators* or *transformers* (like machines) to which we can feed appropriate ordered sets x, y, z, t of numbers. Then (after mechanical squeaking or electronic flashing or what not) $\mathbf{f}$ and f_1 produce the required vector $\mathbf{f}(x,y,z,t)$ and the required number $f_1(x,y,z,t)$. In the static approach, $\mathbf{f}$ is regarded as being the *set of ordered quintuples* $(x, y, z, t, \mathbf{f}(x,y,z,t))$ of four numbers and a vector in which the allowable independent variables come first in the appropriate order and the vector $\mathbf{f}(x,y,z,t)$ comes last. In this static approach, f_1 is a set of quintuples of numbers. It is a common but not universal practice to consider these ideas to be more tangible and useful than the idea that $\mathbf{f}$ is a *law* or *rule* by means of which $\mathbf{f}(x,y,z,t)$ can be calculated when x, y, z, t are given. A simpler example may partially clarify these matters. As soon as we know that the area y of a circular disk is determined by its radius x (x being positive because radii of disks are positive numbers), we can say that y is a function of x and write $y = g(x)$. Then $g(2)$ is the area of a disk of radius 2 and $g(2.03)$ is the area of a disk of radius 2.03. In each case $g(x) = \pi x^2$. We can think of g as being the operator which converts x into πx^2 when $x > 0$ or as being the set of ordered pairs $(x,\pi x^2)$ for which $x > 0$.

It is important to know about a particular special way in which a scalar function of one scalar variable can be determined. Suppose we have a given set S of ordered pairs (x,y) of numbers such that the set does not contain two pairs (x_1,y_1) and (x_2,y_2) for which $x_2 = x_1$ and $y_2 \neq y_1$. To each number x_0 that appears as the first number in one of the pairs (x,y), there is then one and only one number y_0 such that the pair (x_0,y_0) appears in the set. We may let $f(x_0)$ denote this number y_0, and we have $y_0 = f(x_0)$. Thus the given set S of ordered pairs (x,y) becomes the set of ordered pairs $(x, f(x))$. When the pairs of numbers in the set are associated with points and are plotted in the usual way, an example being shown in Figure 3.15, the condition on the ordered pairs means that no two points fall on the same vertical line. In the example of the figure, we see that $f(x) = 2$ when $x = 0$, that $f(x) = 1$ when $1 \leqq x < 2$, that $f(x) = 2$ when $x = 2$, and that $f(x) = x - 2$ when $2 < x \leqq 3$. When x has a value different from 0 and not in the interval $1 \leqq x \leqq 3$, no meaning has been attached to $f(x)$ and we say that $f(x)$ is undefined. In this and other cases, the set of values of x for which $f(x)$ is defined is called the *domain* of the function, and the set of values attained by $f(x)$ is called the *range* of the function. All this is perfectly explicit and precise, and it should be thoroughly understood by everyone. One who wishes to regard f as an operator must realize that each set S of the type described above completely determines the number $f(x)$ that f must produce when it operates upon a given number x in the domain of f. Likewise, one who wishes to regard f as a set S must realize that the operator f determines his set S of pairs $(x, f(x))$. Everyone must realize that a set S^* of points in a plane endowed with an x, y coordinate system determines both the operator f and the set S, provided no two points of S^* lie on the same vertical line. As sometimes happens in mathematics and elsewhere, we have a situation in which different individuals can hold different personal preferences. For example, a person who wishes to regard f as an operator can take a dim view of the idea that an appropriate set S of ordered pairs of numbers "is" a function because it determines a function. He can feel that this is too much like saying that a social security number "is" a worker because it determines a worker, and he can object to the idea that social security numbers eat mashed potatoes.

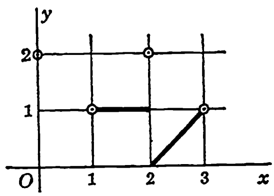


Figure 3.15

The contraption in the central part of Figure 3.151 is guaranteed to make nearly everybody imagine a more or less complicated process by which f might operate upon a given input x (an element of the domain of f) to produce the corresponding output y (an element of the range of f). The last problem of this section provides ideas about functions, operators, and transformers that are needed in advanced mathematics and are helpful in elementary mathematics.

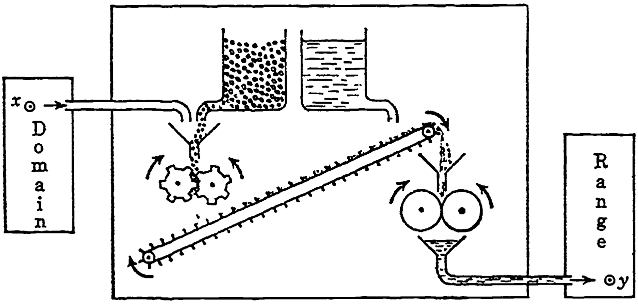


Figure 3.151

If we know that y is always positive and that x and y are always related by the formula $x^2 + y^2 = 9$, we can discover that $y = \sqrt{9 - x^2}$ when $-3 < x < 3$. Thus y is determined as a function of x which is defined over the interval $-3 < x < 3$, and the graph is as shown in Figure 3.16. Similarly, if we know that y is negative and $x^2 + y^2 = 9$, we can conclude that $y = -\sqrt{9 - x^2}$ and we have a function whose graph appears in

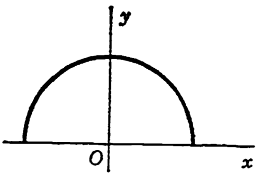


Figure 3.16

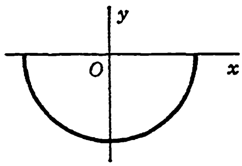


Figure 3.161

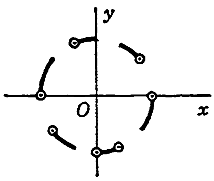


Figure 3.162

Figure 3.161. If we know that $x^2 + y^2 = 9$ but do not know whether y is positive or negative, we cannot determine y in terms of x. The best we can do is say that, for each x in the interval $-3 < x < 3$, y is one or the other of $\sqrt{9 - x^2}$ and $-\sqrt{9 - x^2}$. Figure 3.162 shows the graph of a function f for which $x^2 + [f(x)]^2 = 1$, it being true that $f(x) > 0$ for some values of x and $f(x) < 0$ for other values of x. Observe that the equation $x^2 + y^2 = 1$ does not, by itself, determine y as a function of x, but that there do exist functions f for which $x^2 + [f(x)]^2 = 1$.

One purpose of all this discussion is to emphasize the fact that our ideas about functions must be both broad and precise. We must remain calm when someone says that the temperature u at the north pole of our earth is a function of the time t and, without bothering to introduce a new letter whose significance must be remembered, uses the symbol $u(t)$ to denote the temperature at time t. Many problems in pure and applied mathematics involve functions about which we have some information and seek more. Moreover, we must allow ourselves freedom to use standard terminology that everyone else uses to convey ideas and information. We say that a function f is *increasing* over an interval $a \leqq x \leqq b$ if, as Figure 3.163 indicates, $f(x_1) < f(x_2)$ whenever $a \leqq x_1 < x_2 \leqq b$. Similarly, f is decreasing over the interval if

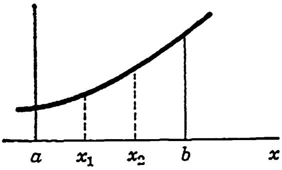


Figure 3.163

$$f(x_1) > f(x_2) \qquad (a \leqq x_1 < x_2 \leqq b). \tag{3.164}$$

In this displayed statement, the "whenever" is omitted. The line can be read "$f(x_1) > f(x_2)$ whenever $a \leqq x_1 < x_2 \leqq b$." If, as Figure 3.163 indicates, f is increasing over the interval $a \leqq x \leqq b$ and if $f(a) = A$ and $f(b) = B$, we say that $f(x)$ increases from A to B as x increases from a to b. While we use this convenient terminology, we need not be gullible people who are easily persuaded that numbers x and $f(x)$ can actually increase. To see 6 increase and say hello to 7 as it proceeds toward 8 could be quite amusing, but we make no pretense that such things actually happen. To avoid misunderstandings, the author wishes to publicly proclaim that he is not recommending rejection of the good old terminology; he is merely insisting that we know what we mean when we say that y or $f(x)$ increases as x increases from a to b.

Problem 15 at the end of this section deals with a famous number-theoretic function. From some points of view, a perfect definition of this function can be phrased as follows. Let π be the function whose domain is the set of real numbers and which is such that, for each x in the domain, $\pi(x)$ is the number of primes less than or equal to x. This makes the "law" or "rule" concept sound very good. We can easily make the pretense that a sufficiently dynamic operator could produce the numbers $\pi(x)$ that we need to form the set S of pairs $(x, \pi(x))$ needed for the static concept. It will be observed that, in Problem 15, the function is defined in fewer words.

Trigonometric functions and polynomials are simpler examples of functions that are important in advanced as well as in elementary science. A *polynomial* (or *polynomial in x*) is a function P having values defined by

$$P(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n \tag{3.17}$$

or by

$$P(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n, \tag{3.171}$$

where n and $a_0, a_1, \cdots, a_n$ and $b_0, b_1, \cdots, b_n$ are constants, n being a nonnegative integer. A *rational* (ratio-nal) *function* is a quotient of polynomials, an example being the function Q for which

$$Q(x) = \frac{x+1}{x^2+x-12} \tag{3.172}$$

for each x for which the denominator is different from zero. When we define a function by a formula more or less like (3.172), we ordinarily understand that the domain of the function is the whole set of numbers x for which the formula actually determines a number. We must, however, recognize the fact that the function g for which $g(x) = \sqrt{x}$ when $1 \leqq x \leqq 4$ is different from the function h for which $h(x) = \sqrt{x}$ when $x \geqq 0$; the domains of the functions are different, and the functions are therefore different. To clarify this point, we can recognize that a machine which is capable of cracking only medium-sized nuts is different from a machine that is capable of cracking nuts of all sizes.

If we have a load of coal of weight w and we toss a lump of coal on or off the load, then the new weight will be a new number which we can call $w + \Delta w$. Thus Δw, which may be either positive or negative, is the difference of two weights (the new minus the old). The number Δw, read "delta w," is a single number (not the product of two numbers Δ and w). This simple notational device turns out to be unexpectedly convenient. In physical chemistry Δp is the difference of two pressures, Δv is the difference of two volumes, and Δt is the difference of two times. In physics, $\Delta \mathbf{v}$ is the difference of two (vector) velocities, and ΔV is the difference of two potentials. In economics ΔP is the difference of two prices, and in

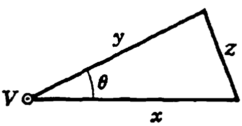


Figure 3.181

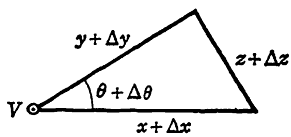


Figure 3.182

biology ΔP may be the difference of two populations. If, as in the discussion involving Figure 3.181, the left member of the formula

$$z = f(x,y,\theta) \tag{3.183}$$

represents the length of the side opposite the vertex V, then the left member of the formula

$$z + \Delta z = f(x + \Delta x,\ y + \Delta y,\ \theta + \Delta\theta) \tag{3.184}$$

represents the length of the side opposite the vertex V in the triangle of Figure 3.182. Of course we can and sometimes shall use shorter symbols such as h, k, p, q for Δx, Δy, $\Delta\theta$, Δz, but very often the extra labor involved in writing the more elaborate delta symbols is a small price to pay for the elimination of the superfluous symbols whose meanings may be forgotten and confused.

At the conclusion of the text of this section, the author makes some remarks that he would have made at the beginning if he had thought that they could have been understood. The old word "function" has been and is and will be used in many different ways. Students who get around will have serious difficulties unless they are so well informed and tolerant that they can accumulate and dispense information by reading and hearing and talking quite different languages. It is like being able to play football with those who play football and to play basketball with those who play basketball; one who knows only ping-pong is sometimes handicapped. This, of course, does not imply that a particular teacher is required to stand by while many different games are played simultaneously in his classroom. Each individual teacher may, with the full backing of the author, go as far as he likes in prescribing the rules of the game to be played in his own classroom.

Problems 3.19

1 If

$$f(x) = x, \qquad g(x) = x^2, \qquad h(x) = 2^x, \qquad \phi(x) = \frac{1}{1+x^2},$$

verify the following assertions and replace the question marks by appropriate answers.

(*a*) $f(0) = 0, f(-3) = -3, f(2) = ?$
(*b*) $g(0) = 0, g(-2) = 4, g(5) = ?$
(*c*) $h(-1) = \frac{1}{2}, h(0) = 1, h(2) = 4, h(\frac{1}{2}) = 1.4142$
(*d*) $h(5) = ?, h(-2) = ?, h(-\frac{1}{2}) = ?$
(*e*) $\phi(\frac{1}{2}) = \frac{4}{5}, \phi(2) = ?, \phi(-2) = ?, \phi(\frac{1}{10}) = ?$
(*f*) $f(8) - f(5) = 3, f(2.1) - f(2) = ?$
(*g*) $g(3) - g(2) = 5, g(2.1) - g(2) = ?$
(*h*) $h(3) - h(2) = 4, h(1) - h(0) = ?$
(*i*) $\phi(1.1) - \phi(1) = -.0475, \phi(0.2) - \phi(0) = ?$
(*j*) $\dfrac{f(4.5) - f(4)}{0.5} = 1, \dfrac{f(2.8) - f(2.7)}{0.1} = ?$
(*k*) $\dfrac{g(4.1) - g(4)}{0.1} = 8.1, \dfrac{g(2.8) - g(2.7)}{0.1} = ?$
(*l*) $\dfrac{\phi(2.1) - \phi(2)}{0.1} = -.152, \dfrac{\phi(0.2) - \phi(0)}{0.2} = ?$
(*m*) $\dfrac{g(x+2) - g(x)}{2} = 2x + 2, \dfrac{g(x+0.5) - g(x)}{0.5} = ?$
(*n*) $\dfrac{g(1+\Delta x) - g(1)}{\Delta x} = 2 + \Delta x, \dfrac{g(\Delta x) - g(0)}{\Delta x} = ?$

(o) $\dfrac{f(x + \Delta x) - f(x)}{\Delta x} = ?, \dfrac{g(x + \Delta x) - g(x)}{\Delta x} = 2x + \Delta x$

(p) $\dfrac{\phi(x + \Delta x) - \phi(x)}{\Delta x} = \dfrac{-2x - \Delta x}{[1 + x^2][1 + (x + \Delta x)^2]}$

2 The *signum function* having values sgn x (read signum x, almost like sine x) is defined by the formula

$$\begin{aligned} \operatorname{sgn} x &= 1 && (x > 0) \\ \operatorname{sgn} x &= 0 && (x = 0) \\ \operatorname{sgn} x &= -1 && (x < 0). \end{aligned}$$

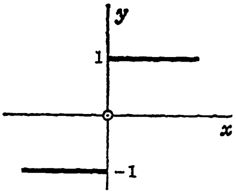


Figure 3.191

Show that Figure 3.191 displays the graph of sgn x and then draw the graph of sgn $(x - 2)$. Show that $|x| = x$ sgn x. *Hint:* Consider separately the cases in which $x > 0$, $x = 0$, and $x < 0$.

3 The *Heaviside* (1850–1925) *unit function* having values defined by

$$\begin{aligned} H(x) &= 1 && (x > 0) \\ H(x) &= \tfrac{1}{2} && (x = 0) \\ H(x) &= 0 && (x < 0) \end{aligned}$$

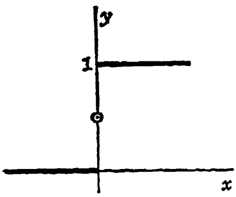


Figure 3.192

is named for the mighty electrical engineer who popularized its use. Show that Figure 3.192 displays the graph of $H(x)$ and then draw the graph of $H(x - 2)$. Show that

$$H(x) = \frac{1 + \operatorname{sgn} x}{2}, \qquad \operatorname{sgn} x = 2H(x) - 1.$$

4 We need more evidence that not all functions have simple graphs that are easily drawn. Let D be the *dizzy dancer function*, defined over the interval $0 \leqq x \leqq 1$, for which

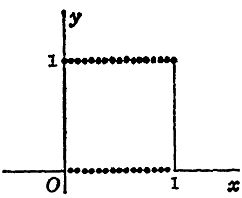


Figure 3.193

$$\begin{aligned} D(x) &= 0 && (x \text{ irrational}) \\ D(x) &= 1 && (x \text{ rational}). \end{aligned}$$

Think about this matter and acquire the ability to make a figure more or less like Figure 3.193 to "represent" the graph of D.

5 A function g is defined by the formulas

$$\begin{aligned} g(x) &= x^2 && (0 \leqq x \leqq 1) \\ g(x) &= x && (\text{otherwise}). \end{aligned}$$

Plot its graph.

6 A function f is said to be an *even function* if $f(-x) = f(x)$ whenever x belongs to the domain of f and is said to be an *odd function* if $f(-x) = -f(x)$ whenever x belongs to the domain of f. Prove that the polynomial having values $2 - 3x^2 + 5x^4$ (with only even exponents appearing) is even. Prove that the polynomial having values $x - 7x^5 + 2x^7$ (with only odd exponents appearing) is odd. Prove that the polynomial having values $1 - 2x + 3x^2$ is neither even nor odd.

7 If $h(x) = x + 1/x$ when $x \neq 0$, show that $h(1/t) = h(t)$ when $t \neq 0$ and that $[h(x)]^2 = h(x^2) + 2$. Work out a formula for $h(h(x))$ and check the formula by setting $x = 2$.

8 If $f(x) = x^2 + 3x + 1$, show that $f(-3) = 1$, $f(-1) = -1$, $f(0) = 1$, $f(\frac{1}{2}) = \frac{11}{4}$, $f(2) = 11$, and

$$f(x + \Delta x) = x^2 + 3x + 1 + (2x + 3)\,\Delta x + \Delta x^2$$

when Δx^2 means $(\Delta x)^2$. It is quite appropriate to use this formula as a basis for a feeling that, when x has a particular fixed value such as 0 or -2 or π, the value of $f(x + \Delta x)$ is nearly the same as the value of $f(x)$ whenever Δx is nearly 0.

9 If $f(x) = mx + b$, show that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = m$$

whenever $x_2 \neq x_1$. Sketch a figure and comment upon the result.

10 If $f(x) = x^2$, show that

$$\frac{f(x + h) - f(x)}{h} = 2x + h$$

when, as always when we make calculations of this kind, $h \neq 0$. Sketch a graph of the function and use the above formula to find the slope of the line L passing through the two points on the graph for which $x = 1$ and $x = 1.001$. The answer is 2.001, and it is quite appropriate to have a feeling that this is nearly the slope of the line tangent to the graph at the point (1,1).

11 If $f(x) = x^2$, simplify

$$\frac{f(x + \Delta x) + f(x - \Delta x) - 2f(x)}{\Delta x^2}.$$

12 If $f(x) = 1/x$, and if x and $x + \Delta x$ are both different from 0, show that

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = \frac{-1}{x(x + \Delta x)}.$$

13 Make appropriate use of the trigonometric formulas

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$$
$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

to obtain the formulas

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

14 Show that y will be a function of x for which

$$x^2 + xy(x) + [y(x)]^2 = 3$$

if $-2 \leqq x \leqq 2$ and, for each such x, $y(x)$ is one or the other of the two numbers

$$\frac{-x - \sqrt{3(4 - x^2)}}{2}, \qquad \frac{-x + \sqrt{3(4 - x^2)}}{2}$$

which are equal only when $x = -2$ and when $x = 2$. *Hint:* Use the quadratic formula.

15 An integer n greater than 1 is said to be *composite* if, like 39, it is representable as the product of two integers each greater than 1 and is said to be a *prime* if, like 29, it is not so representable. The primes are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, $\cdots$, there being an infinite set of them. One of the famous functions of number theory is $\pi(x)$, the number of primes less than or equal to x. It is easy to see that $\pi(8.27) = 4$. It has been proved that $\pi(10^3) = 168$, $\pi(10^6) = 78{,}498$, and $\pi(10^9) = 50{,}847{,}478$. To graph $\pi(x)$ over the whole interval $0 \leqq x \leqq 10^9$ would be quite a chore. However, draw the graph over a shorter interval, say $0 \leqq x \leqq 40$, and try to pick up some ideas.

16 Another famous number-theoretic function has, for each positive integer n, the value $d(n)$, where $d(n)$ is the number of positive integer divisors (including 1 and n) of n. For example, the divisors of 6 are 1, 2, 3, and 6. Verify the entries in the little table

$$\begin{array}{rcccccccccccccc} n = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ d(n) = & 1 & 2 & 2 & 3 & 2 & 4 & 2 & 4 & 3 & 4 & 2 & 6 & 2 \end{array}$$

and calculate $d(2^3 3^2 5^2)$.

17 We take a brief preliminary look at some functions that play fundamental roles in physics, mechanics, and statistics. Let n be a positive integer. For each $k = 1, 2, 3, \cdots, n$, let a particle P_k^* of mass m_k be concentrated at the point $P_k(x_k, y_k)$. In what follows we use ξ (xi, the Greek x) to denote a number which can easily be considered to be the x coordinate of a point, and we use M with a superscript to make us think of a moment. For each number ξ, the number $M^{(1)}_{x=\xi}$ defined by

(1) $$M^{(1)}_{x=\xi} = m_1(x_1 - \xi) + m_2(x_2 - \xi) + \cdots + m_n(x_n - \xi)$$

is called the *first moment* of the mass system about the line having the equation $x = \xi$. Supposing that the total mass

(2) $$M = m_1 + m_2 + \cdots + m_n$$

of the system is positive, we can put (1) in the form

(3) $$M^{(1)}_{x=\xi} = M\left(\frac{m_1x_1 + m_2x_2 + \cdots + m_nx_n}{M} - \xi\right).$$

Similarly, for each number η (eta) the number $M^{(1)}_{y=\eta}$ defined by

(4) $$M^{(1)}_{y=\eta} = m_1(y_1 - \eta) + m_2(y_2 - \eta) + \cdots + m_n(y_n - \eta)$$

is called the first moment of the mass system about the line having the equation $y = \eta$, and

(5) $$M^{(1)}_{y=\eta} = M\left(\frac{m_1y_1 + m_2y_2 + \cdots + m_ny_n}{M} - \eta\right).$$

The particular point (ξ,η) for which $M^{(1)}_{x=\xi} = 0$ and $M^{(1)}_{y=\eta} = 0$ is called the *centroid*† (thing like a center) of the mass system. The coordinates of the centroid are denoted by $\bar{x}$ and $\bar{y}$. Thus, $M^{(1)}_{x=\bar{x}} = 0$ and $M^{(1)}_{y=\bar{y}} = 0$, and it follows from (3) and (5) that

$$(6) \quad \bar{x} = \frac{m_1x_1 + m_2x_2 + \cdots + m_nx_n}{M}, \qquad \bar{y} = \frac{m_1y_1 + m_2y_2 + \cdots + m_ny_n}{M}.$$

In case $m_k = 1$ for each k, (2) shows that $M = n$ and the formulas (6) reduce to

$$(7) \qquad \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \qquad \bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}.$$

In this case the centroid $(\bar{x},\bar{y})$ is called the centroid of the set of points $P_1, P_2, \cdots, P_n$. To prepare us for Section 4.7 and other sections where less simple mass systems are considered, we should take brief cognizance of a more general definition. Let p be a nonnegative integer. The number $M^{(p)}_{x=\xi}$ is defined by

$$(8) \qquad M^{(p)}_{x=\xi} = m_1(x_1 - \xi)^p + m_2(x_2 - \xi)^p + \cdots + m_n(x_n - \xi)^p$$

and is called the pth *moment* of the mass system about the line having the equation $x = \xi$. Similarly, the number $M^{(p)}_{y=\eta}$ defined by

$$(9) \qquad M^{(p)}_{y=\eta} = m_1(y_1 - \eta)^p + m_2(y_2 - \eta)^p + \cdots + m_n(y_n - \eta)^p$$

is called the pth moment of the mass system about the line having the equation $y = \eta$. In physics and mechanics (but not so often in statistics) the second moment is called *moment of inertia.* Since we are studying functions, we can observe that, if our mass system contains 40 particles, there is a sense in which the moments in (8) and (9) are functions of 82 variables of which two are p and ξ. While this textbook does not require calculations of these moments, we can recognize that there are many situations in which calculations must be made, and this is one of the reasons why the world contains so many calculating machines and computers of assorted mechanical and electronic varieties.

18 If $f(x) = 1 + x + x^2 + x^3 + x^4$, show that $f(1) = 5$ and

$$f(x) = \frac{x^5 - 1}{x - 1} = \frac{1 - x^5}{1 - x}$$

when $x \neq 1$.

19 *Remark:* This remark invites more complete comprehension of ideas and terminology involving functions. The left-hand part of Figure 3.194 represents

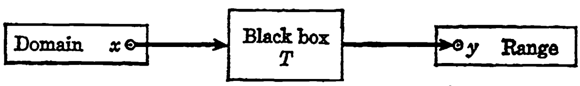


Figure 3.194

a set D of numbers or of vectors or of entities of some other kind that is called a *domain.* The central part of the figure represents a mechanism that is sometimes

† We are being rather unrealistic if we suppose that everybody always chooses the same coordinate system when studying a given system of particles. The coordinates of the centroid depend upon the coordinate system used, but (we omit the proof) the location of the centroid relative to the system of particles is the same for all coordinate systems. For example, if three particles of equal mass lie at the vertices of a triangle, then the centroid lies at the intersection of the medians of the triangle.

called a *black box* and is sometimes called a *transformer* T. When an element x of the domain is selected and fed into the black box or transformer, the x is called an *input* and the black box or transformer is supposed to produce an *output* which is an element y of a set R which is called a *range*. Thus to each x in D there corresponds exactly one y in R which is called the *transform* of x and is denoted by $T(x)$ so that $y = T(x)$. Thus we have a transformer T which transforms each x in D into a transform $T(x)$ in R. So far we have used the words "transformer" and "transform," but we have not used the word "transformation." Our domain and transformer and range determine and are determined by the set S of ordered pairs (x,y) for which x is an element of D and y is the element of R for which $y = T(x)$, and we call this set S a *transformation* T_S. The domain (set of inputs) and range (set of outputs) of the transformer T are also the domain and range of the transformation T_S. We now have adequate terminology and notation. The transformer T is the active "operator" that converts each element x of D (or each first element of one of the pairs in T_S) into the transform $T(x)$ in R (which is the appropriate second element of a pair in the set S which constitutes T_S). The transformer T and the transformation T_S are inherently different things, and there can be no doubt that our science is inadequately developed when we apply the same name and the same symbol to the two things. The worst of it is that, when the word "function" is used, this one word sometimes means a transform $T(x)$, sometimes means the transformer T, and sometimes means the transformation T_S. Perhaps an assertion involving the word "function" will help us to see why we must make a rather serious study of terminology before we can be intelligent readers and listeners. It is the function (see the nonmathematical meanings given in a dictionary) of a function (transformer) to carry an element of the domain D of the function (transformer or transformation) into the function (transform) in the range R of the function (transformer or transformation). Commenting upon this matter from the point of view of mathematical logic, Professor Rosser remarked to the author that some of our terminological difficulties are due to the fact that the already overburdened old word "function" was used as a name for the set S of ordered pairs. It is possible that terminology will slowly improve, but meanwhile we can be comforted by the fact that the bad terminology rarely if ever actually injures us. We can be irked by the fact that a "diameter" of a circle is sometimes a line segment (a point set) and is sometimes a number (the length of the line segment), but we are rarely if ever injured and there seems to be no overwhelming demand for improvement of the terminology.

3.2 Limits When we were infants learning to walk and to talk, and perhaps even after that, we heard many statements that we could not comprehend. When an explorer tells us that he found a complete set of normalized Legendre polynomials in an ancient cave in Peru and that the carbon test shows that the set is 24,500 years old, it may be difficult for us to learn what he is talking about and whether he is telling a truth. Moreover, statements involving erudite technical terminology are not the only ones that can be troublesome. Sometimes we must do considerable working and thinking before we can fully understand statements that involve only simple words and may seem, at first sight, to be childishly simple.

It is reasonable to suppose that the harangue of the previous paragraph is leading up to something, and that the lightning is about to strike. It is. We are going to undertake to make a sane appraisal of the assertion

(3.21) $\boxed{x^2 \text{ is near } 9 \text{ whenever } x \text{ is near } 3 \text{ but } x \neq 3}$

which we shall call the assertion in the first box. The assertion does not say anything about the value of x^2 when $x = 3$. It does not say that x^2 is 9 when $x = 3$, and hence it does not pretend to tell the whole truth. There is a fundamental reason why it is not completely easy to tell what the assertion does mean. The reason is that it simply does not make precise mathematical sense to say that a number x is near 3. Whether 416 or 4 or 3.01 or 3.00001 or 2.98 is considered to be near 3 or not can be a matter of opinion and can depend upon circumstances. Likewise, it does not make precise mathematical sense to say that x^2 is near 9. Discouraging as this may be, we must recognize that it may be possible to attach a precise meaning to the assertion in the first box without attaching meanings to the "assertions" x is near 3 and x^2 is near 9. After all, the word "attaching" can mean something even when "atta" and "ching" do not. It should be possible to tell precisely what the assertion does mean, because the assertion uses words in a thoroughly serious attempt to convey information. A fundamental idea is involved.

Our first attempt to make sense out of (3.21) is to replace it by the assertion

(3.211) $\boxed{\begin{array}{l} x^2 \text{ is a good approximation to } 9 \text{ whenever } x \text{ is} \\ \text{a good approximation to } 3 \text{ but } x \neq 3 \end{array}}$

in the second box. This change in the wording can be psychologically satisfying, and we started with (3.21) only because it is shorter than (3.211). We have not, however, conquered our fundamental difficulty, because the statement that one number is a good approximation to another is neither more nor less illuminating than the statement that one is near the other.

It is a remarkable fact that much of the mathematical progress of the past century is based upon the development and use of a particular special method of attaching meaning to the statements in the first two boxes. The method is called the epsilon-delta method because it traditionally employs the two Greek letters ϵ (epsilon) and δ (delta). The meaning of the assertions in the first two boxes is, by this method, defined to be the same as that of the assertion

(3.22)
$$\boxed{\begin{array}{l}\text{To each positive number } \epsilon \text{ there corresponds} \\ \text{a positive number } \delta \text{ such that} \\ \quad |x^2 - 9| < \epsilon \text{ whenever } 0 < |x - 3| < \delta\end{array}}$$

which we shall call the epsilon-delta assertion.† When we first see the epsilon-delta assertion, we are entitled to feel that it lacks the intuitional appeal of the preceding assertions, but it turns out to be the fully meaningful assertion which can be proved if it is true and can be disproved if it is false.

Before further discussion of the assertions in the boxes, we can note that they are so long that it is tedious to write them very often and that they are universally abbreviated by the efficient and effective shorthand

(3.23)
$$\boxed{\lim_{x\to 3} x^2 = 9}$$

in our fourth and final box. Thus the assertions in the four boxes are equivalent; if one is true, then all four are true; and if one is false, then all four are false. They all mean the same thing.

The only possible discordant phrase in the symphony is the noise we make when we read the assertion in the last box. We say that the limit as x approaches 3 of x^2 is 9. Thus we have another technical statement couched in terms of the dubious concept of moving numbers. Stephen Leacock (1869–1944) was wise enough to realize that if a number x really could approach 3 from more than one direction, then it should be able to reverse the process and go away from 3 in more than one direction. In any case, Leacock enabled Lord Ronald (a character in "Nonsense Novels, Gertrude the Governess: or, Simple Seventeen") to fling himself upon his horse and ride madly off in all directions. We make no attempt to explain our basic concept in terms of moving numbers. Such attempts are much too mystic and vague for advanced technical books, and we can hold the view that they are at least a little bit too mystic and vague for elementary books. In our book, the collection of words "the limit as x approaches 3 of x^2 is 9" does not suggest that numbers jump around; it suggests that "x^2 is near 9 whenever x is near 3 but x is different from 3," and this basic concept is made precise by the epsilon-delta assertion.

† In this and similar assertions, we avoid difficulties by using the word "each" in preference to "any" because the troublesome word "any" often means "at least one." Residents of Los Angeles can be expected to give lusty affirmative answers when asked whether any major city of the United States lies west of the Mississippi River. If $\epsilon_1 < \epsilon_2$ and $|x^2 - 9| < \epsilon_1$, then $|x^2 - 9| < \epsilon_2$. For this reason the epsilon-delta assertion will be true if to *each* ϵ for which $0 < \epsilon < 0.001$ there corresponds a positive number δ such that $|x^2 - 9| < \epsilon$ whenever $0 < |x - 3| < \delta$.

It is both easy and customary to adopt the absurd view that everybody has spent huge amounts of time squaring all sorts of numbers near 3 and has somehow picked up positive knowledge that the assertions in the boxes are true. Instead of trying to discover how uncertain we should be, we eliminate uncertainties by proving the epsilon-delta assertion. Let $\epsilon > 0$. If $0 < \delta < 1$ and $\delta < \epsilon/7$, then

$$(3.24)\quad |x^2 - 9| = |(x+3)(x-3)| = |x+3|\,|x-3| \leqq 7|x-3| < 7\delta < \epsilon$$

whenever $0 < |x - 3| < \delta$. To obtain the inequality $|x + 3| < 7$ which was used in (3.24), we can use an appropriate figure or, alternatively, use the fact that if $|x - 3| < \delta$ and $\delta < 1$, then

$$|x+3| = |x-3+6| \leqq |x-3| + 6 < \delta + 6 < 7.$$

Thus an appropriate δ can be found and the assertion is true. It is not inappropriate to think about this matter for a few minutes or perhaps longer.†

We are familiar with the nature of the graph of the equation $y = x^2$, and it is comforting to see that the epsilon-delta assertion has a simple geometric interpretation. It says that if $\epsilon > 0$ and if horizontal lines are drawn through the points with y coordinates $9 - \epsilon$ and $9 + \epsilon$ as in Figure 3.241, then there exist vertical lines (dotted in the figure) such that, with the possible exception of a single point for which $x = 3$, *the part of the graph between the vertical lines is also between the horizontal lines.* The little sister we mention occasionally might be irked by the possible exceptional point, but she certainly would be clever enough to put in the dotted lines after we had shown her a figure containing the horizontal lines; the process is thoroughly elementary and we need not require that efforts be made to seek the greatest δ that serves the purpose. Even though it does not make precise mathematical sense to say

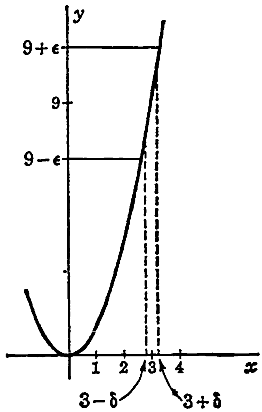


Figure 3.241

† The famous flea assertion "each flea has a smaller flea to bite him" is, in some respects, similar to the epsilon-delta assertion. We recognize that the "each flea" at the front of the assertion invites us to think about fleas one at a time, not every flea or all fleas at once. To prove the flea assertion, we would be required to start with a given flea, say Mr. F. (who could be any flea but would not be every flea or the collection of all fleas), and show that there is a smaller flea which is so related to Mr. F. (and which may be said to correspond to Mr. F.) that it bites him. To prove the epsilon-delta assertion, it suffices to start with a given positive number ϵ (which could be 416 or $\frac{1}{2}$ or 0.00001 or any other positive number but naturally cannot be all of these things at once) and then show that there is a positive number δ so related to ϵ that $|x^2 - 9| < \epsilon$ whenever $x \neq 3$ and $|x - 3| < \delta$.

that ϵ is small, we need not deny ourselves the satisfaction of the feeling that, when the given ϵ is small, the dotted lines must be close together and the δ must be small.

For the case in which $f(x) = x^2$ and $a = 3$, we have been discussing questions involving values of $f(x)$ when x is near a. Our serious interest often lies in such questions when $f(x)$ has a more complicated expression, say one of

$$\frac{1}{x}, \qquad \frac{\sqrt{2+x}-\sqrt{2}}{x}, \qquad \frac{\sin x}{x}, \qquad (1+x)^{1/x}.$$

We should therefore know that the assertions in the four boxes

(3.25) $\boxed{f(x) \text{ is near } L \text{ whenever } x \text{ is near } a \text{ but } x \neq a.}$

(3.251) $\boxed{\begin{array}{l} f(x) \text{ is a good approximation to } L \text{ whenever} \\ x \text{ is a good approximation to } a \text{ but } x \neq a. \end{array}}$

(3.26) $\boxed{\begin{array}{c} \text{To each } \epsilon > 0 \text{ there corresponds a } \delta > 0 \text{ such that} \\ |f(x) - L| < \epsilon \text{ whenever } 0 < |x - a| < \delta. \end{array}}$

(3.27) $\boxed{\lim_{x\to a} f(x) = L.}$

have identical meanings. When we have plenty of time, we can always replace the epsilon-delta assertion by the following more ponderous but psychologically satisfying one. To each positive number ϵ there corresponds a positive number δ such that $f(x)$ approximates L so closely that $|f(x) - L| < \epsilon$ whenever x is different from a but approximates a so closely that $|x - a| < \delta$. The assertion (3.27) is read "the limit as x approaches a of $f(x)$ is L."

If f and a are such that there is no L for which the four assertions are true, then we say that

$$\lim_{x\to a} f(x)$$

does not exist. Complete comprehension of this matter is essential; otherwise, we must be eternally confused by a statement that a thing at which we are looking does not exist.

Some assertions involving limits are not completely simple. There will come a day when we must know there is a number e, having the approximate value in

(3.271) $$e = 2.71828\ 18284\ 59045,$$

such that

$$\lim_{x\to 0} (1 + x)^{1/x} = e. \tag{3.272}$$

Anybody can collect a little evidence in support of this assertion by making calculations when x has such values as $\pm\frac{1}{2}$, $\pm\frac{1}{3}$, $\pm\frac{1}{4}$, and $\pm\frac{1}{5}$, but it is not so easy to prove the assertion. In fact we must have very substantial information about limits before we can, in Chapter 9, define functions having values e^x and, when $x > 0$, $\log x$. Meanwhile, many of the problems that confront us will be solved very quickly and easily with the aid of the following fundamental theorems. We call them limit theorems, but they are nothing but basic theorems in the theory of approximation.

Theorem 3.281 *If*

$$\lim_{x\to a} f(x) = L_1, \qquad \lim_{x\to a} f(x) = L_2,$$

then $L_2 = L_1$.

Theorem 3.282 *If b is a constant, then*

$$\lim_{x\to a} b = b.$$

Theorem 3.283

$$\lim_{x\to a} x = a.$$

Theorem 3.284 *If c is a constant, then*

$$\lim_{x\to a} cf(x) = c \lim_{x\to a} f(x)$$

provided the limit on the right exists.

Theorem 3.285 *The formulas*

$$\lim_{x\to a} [f(x) + g(x)] = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$

$$\lim_{x\to a} [f(x)g(x)] = [\lim_{x\to a} f(x)][\lim_{x\to a} g(x)]$$

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$$

are valid provided the limits on the right exist and, in the case of the last formula, $\lim_{x\to a} g(x) \neq 0$.

Theorem 3.286 *If*

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

then

$$\lim_{x\to a} |f(x) - L| = 0$$

and conversely.

Theorem 3.287 (sandwich theorem or flyswatter theorem) *If for some positive number p*

$$g(x) \leqq f(x) \leqq h(x)$$

when $a - p < x < a$ and when $a < x < a + p$, and if

$$\lim_{x \to a} g(x) = L, \qquad \lim_{x \to a} h(x) = L,$$

then

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

Theorem 3.288 *If p is a constant positive exponent, then*† *the first of the formulas*

$$\lim_{x \to a} x^p = a^p, \qquad \lim_{x \to a+} x^p = a^p$$

holds when $a > 0$ and the second holds when $a \geqq 0$.

These theorems are easily understood and will turn out to be very useful. Unless his teacher rules otherwise, each individual student has three options. He can claim that the theorems are so obvious that they do not need proof and, even though this is surely a precarious way to start a successful mathematical career, he may even be right. He can claim that they are not obviously true but he will accept them because they are printed and the teacher says there are no misprints. Finally, he may want to see proof because he is suspicious or inquisitive or wants to develop abilities to prove things. In the latter case he may attack Appendix 1 at the end of this book. Whatever we do, we should always believe that if $f(x)$ lies between $g(x)$ and $h(x)$ and if $g(x)$ and $h(x)$ are both near L whenever x is near a but $x \neq a$, then $f(x)$ must be near L whenever x is near a but $x \neq a$. This is what the sandwich theorem says, and the meanings of the other theorems are also simple.

The first two of the following problems are designed to promote understanding of the epsilon-delta assertion (3.26). We must always remember that if the epsilon-delta assertion is true, then to each (not all or every) epsilon that is positive there corresponds a delta that is positive such that $|f(x) - L| < \epsilon$ whenever x is different from a but so near a that $|x - a| < \delta$. It is not asserted that there is a delta which corresponds to every epsilon. It is asserted that to each epsilon there corresponds a delta. The epsilon comes first, and the delta follows.

† The meaning of the second of these statements is explained in Section 3.3. The theorem is, as Appendix 1 says, proved in Chapter 9 after the theory of exponentials and logarithms has been developed. See Theorem 9.271.

Problems 3.29

1 It is not enough to be able to read the four assertions which involve $f(x)$ when x is near a but $x \neq a$. We must be able to write them. Try to write them with the text out of sight and, if unsuccessful, read the text some more and try again.

2 In terms of epsilons and deltas, write a complete statement giving the exact meaning of each of the following true statements:

(*a*) x^3 is near 27 whenever x is near 3 but $x \neq 3$.
(*b*) $\sin x$ is near 0 whenever x is near 0 and $x \neq 0$.
(*c*) $\dfrac{\sin x}{x}$ is near 1 whenever x is near 0 and $x \neq 0$.
(*d*) $\dfrac{1 - \cos x}{x}$ is near 0 whenever x is near 0 and $x \neq 0$.
(*e*) x is near 1 whenever x is near 1 and $x \neq 1$.

(*f*) $\displaystyle\lim_{x\to 0} \frac{\sin x}{x} = 1$ (*g*) $\displaystyle\lim_{x\to 0} \frac{1 - \cos x}{x} = 0$

(*h*) $\displaystyle\lim_{x\to 0} \frac{1 - \cos x}{x^2} = \frac{1}{2}$ (*i*) $\displaystyle\lim_{x\to 2} e^x = e^2$

(*j*) $\displaystyle\lim_{x\to 0} (1 + x)^{1/x} = e$ (*k*) $\displaystyle\lim_{x\to 0} \frac{e^x - 1}{x} = 1$

(*l*) $\displaystyle\lim_{\Delta x\to 0} \frac{\sin (x + \Delta x) - \sin x}{\Delta x} = \cos x$

(*m*) $\displaystyle\lim_{\Delta x\to 0} \frac{\cos (x + \Delta x) - \cos x}{\Delta x} = -\sin x$

(*n*) $\displaystyle\lim_{x\to 1} \frac{nx^{n+1} - (n + 1)x^n + 1}{(x - 1)^2} = \frac{n(n + 1)}{2}$

Answer to last part: To each $\epsilon > 0$ corresponds a $\delta > 0$ such that

$$\left| \frac{nx^{n+1} - (n + 1)x^n + 1}{(x - 1)^2} - \frac{n(n + 1)}{2} \right| < \epsilon$$

whenever $0 < |x - 1| < \delta$.

3 The first formula of Theorem 3.285 assures us that if $f(x)$ is near 3 and $g(x)$ is near 5 when x is near a but $x \neq a$, then $f(x) + g(x)$ is near 8 whenever x is near a but $x \neq a$. Give similar applications of the other two formulas in the theorem.

4 Tell whether you would like to learn and use new notation by which one or the other of the "formulas"

$$(1) \qquad \operatorname*{approx}_{\epsilon, 0<|x-a|<\delta} f(x) = L, \qquad \operatorname*{approx}_{x\sim a} f(x) = L$$

is used to abbreviate the epsilon-delta assertion: to each positive number ϵ there corresponds a positive number δ such that $f(x)$ approximates L so closely that $|f(x) - L| < \epsilon$ whenever $0 < |x - a| < \delta$. If you have no opinion, think about the matter and get one. *Remark:* A person who thinks that this is a silly question may be thoroughly mistaken. It is not unreasonable to suppose that scientists of the future will adopt notation like that in (1) and that their historians will wonder why on earth people ever concocted tales about moving numbers and converted a few basic theorems in the theory of approximation into a mystic "theory of limits" that kept the world agog for several centuries. In our book, the theory of limits sometimes sounds like a theory of moving numbers but it is in fact a part of the theory of approximation. Let us get on with it.

5 Verify the following assertions and replace the question marks by appropriate answers. The basic limit theorems may be used.

(*a*) $\lim_{x\to 5} 3x = 15$ (*b*) $\lim_{x\to 2} 3x = ?$

(*c*) $\lim_{x\to 0} (3 - 2x) = 3$ (*d*) $\lim_{x\to 0} (4x - 5) = ?$

(*e*) $\lim_{y\to 3} (y + 1)(y + 2) = 20$ (*f*) $\lim_{x\to 4} (x + 2)^2 = ?$

(*g*) $\lim_{x\to 5} \dfrac{x^2 - x + 1}{x^2 + x + 1} = \dfrac{21}{31}$ (*h*) $\lim_{x\to 2} \dfrac{x^2 - 2}{x^2 + 2} = ?$

6 Pay very close attention to the problem of evaluating

$$\lim_{h\to 0} \frac{\sqrt{2+h} - \sqrt{2}}{h},$$

because the process involves some troublesome points. Tell why the last part of Theorem 3.285 cannot be used here. Look at the problem and observe that it is difficult or impossible to guess what the answer (if any) is. Observe that we must put the quotient in a more manageable form before we can find its limit. The next step is to remember from experiences in algebra or to learn right now that the numerator and denominator of the quotient should be multiplied by the "conjugate" of the numerator. Thus

$$\begin{aligned}\lim_{h\to 0} \frac{\sqrt{2+h} - \sqrt{2}}{h} &= \lim_{h\to 0} \frac{\sqrt{2+h} - \sqrt{2}}{h} \frac{\sqrt{2+h} + \sqrt{2}}{\sqrt{2+h} + \sqrt{2}} \\ &= \lim_{h\to 0} \frac{2 + h - 2}{h(\sqrt{2+h} + \sqrt{2})} = \lim_{h\to 0} \frac{1}{\sqrt{2+h} + \sqrt{2}} = \frac{1}{2\sqrt{2}}.\end{aligned}$$

Tell which of the theorems of this section are used in making the last step. To be sure that this process is thoroughly understood, make the small notational adjustments necessary to obtain the formula

$$\lim_{\Delta x\to 0} \frac{\sqrt{x + \Delta x} - \sqrt{x}}{\Delta x} = \frac{1}{2\sqrt{x}} \qquad (x > 0).$$

Put in at least as many steps as appear in the special case.

7 Supposing that $a > 0$, show that

$$\lim_{x\to 0} \frac{x}{\sqrt{x + a^2} - a} = 2a.$$

8 Show that

$$\lim_{x\to 0} \frac{\sqrt{1 + x^2} - 1}{x} = 0.$$

9 Supposing that $y = x^2$ and $y + \Delta y = (x + \Delta x)^2$, show that

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 2x.$$

10 Prove that

$$\lim_{\Delta x \to 0} \frac{(x + \Delta x)^3 - x^3}{\Delta x} = 3x^2.$$

11 Prove that, when $x > 0$,

$$\lim_{\Delta x \to 0} \frac{\sqrt{(x + \Delta x)^3} - \sqrt{x^3}}{\Delta x} = \frac{3}{2}\sqrt{x}.$$

12 We have shown that

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

and we shall learn that

$$\lim_{h \to 0} \frac{\sin h}{h} = 1, \qquad \lim_{h \to 0} \frac{1 - \cos h}{h} = 0.$$

Use these facts to find that

$$\lim_{h \to 0} \frac{\sin (x + h) - \sin x}{h} = \cos x.$$

13 Supposing that $y \neq 0$, prove that

$$\lim_{x \to 0} \frac{x^2 - y^2}{x^2 + y^2} = -1.$$

14 Supposing that $y = 0$, prove that

$$\lim_{x \to 0} \frac{x^2 - y^2}{x^2 + y^2} = 1.$$

15 Prove that if $\lim_{x \to a} f(x) = L$, then to each positive number ϵ there corresponds a positive number δ such that

$$|f(x_2) - f(x_1)| < \epsilon$$

whenever $0 < |x_2 - a| < \delta$ and $0 < |x_1 - a| < \delta$. *Remark:* Proof of this result depends upon the idea that it two things are near the same place, then the things must be near each other. The details require careful attention, however. To prove the result, let ϵ be a positive number. Then $\epsilon/2$ is a positive number. Hence there is a positive number δ such that $|f(x) - L| < \epsilon/2$ whenever $0 < |x - a| < \delta$. Therefore,

$$\begin{aligned} |f(x_2) - f(x_1)| &= |[f(x_2) - L] - [f(x_1) - L]| \\ &\leq |f(x_2) - L| + |f(x_1) - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

whenever $0 < |x_1 - a| < \delta$ and $0 < |x_2 - a| < \delta$.

16 Recall that the signum function having values sgn x (read signum x) is defined by the formula

$$\begin{aligned} \operatorname{sgn} x &= 1 && (x > 0) \\ \operatorname{sgn} x &= 0 && (x = 0) \\ \operatorname{sgn} x &= -1 && (x < 0). \end{aligned}$$

Show that

$$\lim_{x\to 0} \operatorname{sgn} x$$

does not exist. *Solution:* To prove this without the aid of the result of Problem 15, we let $f(x) = \operatorname{sgn} x$ and prove that there is no number L for which the epsilon-delta assertion is true. To do this we assume (intending to show that the assumption must be false) that there is a number L for which the assertion is true. Let ϵ be a number for which $0 < \epsilon < 1$, and let δ be a corresponding positive number such that $|f(x) - L| < \epsilon$ whenever $0 < |x| < \delta$. If $0 < x < \delta$, then $f(x) = 1$ and hence $|1 - L| < \epsilon$. If $-\delta < x < 0$, then $f(x) = -1$ and hence $|-1 - L| < \epsilon$. Therefore,

$$2 = |1 + 1| = |1 - L + 1 + L| \leqq |1 - L| + |1 + L| < 2\epsilon$$

and hence $\epsilon > 1$. This contradicts the inequality $\epsilon < 1$ and establishes our result.

17 Show that if $f(x) = |x|$, then

$$\lim_{h\to 0} \frac{f(0 + h) - f(0)}{h}$$

does not exist. *Solution:* Let $g(h)$ denote the above quotient. When $h > 0$, we find that $g(h) = h/h = 1$, and when $h < 0$, we find that $g(h) = -h/h = -1$. The result then follows from the preceding problem.

18 Prove that the first of the assertions

$$\lim_{x\to 2} x^2 = 4, \qquad \lim_{x\to 2} x^2 = 5 \ (?)$$

is true and that the second is false.

19 If D is the dizzy dancer function for which

$$\begin{aligned} D(x) &= 0 && (x \text{ irrational}) \\ D(x) &= 1 && (x \text{ rational}), \end{aligned}$$

prove that there is no a for which $\lim\limits_{x\to a} D(x)$ exists.

20 Suppose that, in some vast universe, it really is true that each flea has a smaller flea to bite him. Suppose also that the universe contains at least one flea. Do these hypotheses imply that there exist fleas having mass less than 1 milligram? *Ans.:* No. The hypotheses would be satisfied if to each positive integer n there corresponds a flea whose mass in milligrams is $1 + 1/n$, and the flea of mass $1 + 1/n$ is bitten by the flea of mass $1 + \dfrac{1}{n + 1}$.

3.3 Unilateral limits and asymptotes When we are talking about the function f for which $f(x) = \operatorname{sgn}(x - a)$ and see the graph shown in Figure 3.31, and in some other cases as well, we can cheerfully assert that λ_R (lambda sub R) is a number such that $f(x)$ is near λ_R whenever x is near a and $x > a$. We can feel sure that we know the meaning of the assertion, but we must nevertheless know that the epsilon-delta version of the assertion is the following.

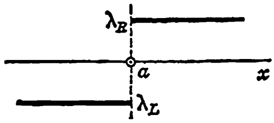


Figure 3.31

To each $\epsilon > 0$ there corresponds a $\delta > 0$ such that $|f(x) - \lambda_R| < \epsilon$ whenever $a < x < a + \delta$. This time the condition $x \neq a$ does not enter the assertion to bother our little sister and everything is very simple. The abbreviated version of the assertion is

$$\lim_{x \to a+} f(x) = \lambda_R. \tag{3.32}$$

The new thing in this symbol is the plus sign that follows the a. Perhaps the best way to read this is "the right-hand limit as x approaches a of $f(x)$ is λ_R," but it is always awkward to write one thing and say another, so the reading usually boils down to "the limit as x approaches a plus of $f(x)$ is λ_R." In case there is no number for which the assertion is valid, we say that the right-hand limit does not exist. A similar succession of ideas leads to the symbol

$$\lim_{x \to a-} f(x) = \lambda_L, \tag{3.321}$$

which says that the left-hand limit as x approaches a of $f(x)$ is λ_L.

If a function f and a number x_0 are such that the unilateral limits λ_R and λ_L in

$$\lim_{x \to x_0+} f(x) = \lambda_R, \qquad \lim_{x \to x_0-} f(x) = \lambda_L \tag{3.33}$$

exist and are different, then the function f is said to have a *jump* (or an *ordinary discontinuity*) at the point x_0. The *magnitude* of the jump is $|\lambda_R - \lambda_L|$. If $\lambda_R > \lambda_L$, then f has an upward jump, and if $\lambda_R < \lambda_L$, then f has a downward jump.

Another assertion that turns out to be both interesting and important is the assertion that a function f and a number L may be such that $f(x)$ is near L whenever x is large. When making this assertion precise, we do not use the letters ϵ and δ but, instead, use ϵ and some other letter, say N, that we can easily regard as a "large" number. The assertion means that to each $\epsilon > 0$ there corresponds a number N such that

$$|f(x) - L| < \epsilon \qquad (x > N). \tag{3.34}$$

By tossing in some surplus verbiage, we can put this in terms that may be psychologically satisfying. Whenever a positive number ϵ is selected, we can find a positive number N so large that $f(x)$ approximates L so closely that $|f(x) - L| < \epsilon$ whenever x is so large that $x > N$. The abbreviated version of this assertion is

$$\lim_{x \to \infty} f(x) = L. \tag{3.341}$$

This is read "the limit as x approaches infinity of eff of ex is ell," or "the limit as x becomes infinite of eff of ex is ell." This does not mean that "infinity" is a place toward which numbers can gallop. All tales about infinity† and galloping numbers are completely irrelevant, and there is no sense in which x really "becomes infinite." The assertion (3.341) means that $f(x)$ is near L whenever x is large. We examine an example. Everyone who has an appreciation of the magnitudes of the numbers $1/2$, $1/416$, $1/7{,}528{,}432$, and $1/10^{20}$ must believe that $1/x$ is near 0 whenever x is large, that is,

$$\lim_{x \to \infty} \frac{1}{x} = 0. \tag{3.342}$$

To prove this, let $\epsilon > 0$. Let $N = 1/\epsilon$. Then the inequality

$$\left| \frac{1}{x} - 0 \right| < \epsilon$$

is valid whenever $1 < \epsilon x$ and hence whenever $x > 1/\epsilon$ and hence whenever $x > N$. Thus when a positive number ϵ is given, we are able to find a number N for which the ϵ, N assertion is true. Therefore, (3.342) is a true assertion. It is equally easy to attach a meaning to the assertion that $f(x)$ is near L whenever x is negative and has a large absolute value. The abbreviated version of this assertion is

$$\lim_{x \to -\infty} f(x) = L. \tag{3.343}$$

We say that the limit as x approaches minus infinity of $f(x)$ is L.

There are some important modifications of these ideas that should now be easily understood. In case $f(x) = 1/(x - a)$ and also in some other cases, we can cheerfully assert that $f(x)$ is large whenever x is near a and $x > a$. This assertion is abbreviated to

$$\lim_{x \to a+} f(x) = \infty. \tag{3.35}$$

It means that to each number M there corresponds a $\delta > 0$ such that

$$f(x) > M \qquad (a < x < a + \delta), \tag{3.351}$$

† For those who are really interested in infinity, a remark appears at the end of the problems of this section.

that is, $f(x)$ exceeds M whenever $a < x < a + \delta$. The assertion that $f(x)$ is large whenever x is large is abbreviated to

$$\lim_{x\to\infty} f(x) = \infty. \tag{3.352}$$

It means that to each number M there corresponds a number N such that

$$f(x) > M \qquad (x > N). \tag{3.353}$$

It is quite appropriate to recognize that ideas akin to those of this section sometimes appear in elementary geometry books when information about lengths of circles is being sought. Let C be a circle having radius $\frac{1}{2}$ and diameter 1. We can imagine that, for each integer $n \geqq 3$, we have inscribed a regular polygon P_n with n sides and have found its length L_n. We can assume (or perhaps prove) that there is a number, which we can call π, such that L_n is near π whenever n is large. By this we mean that to each $\epsilon > 0$ there corresponds an integer N such that $|L_n - \pi| < \epsilon$ whenever $n > N$. The abbreviated form of the assertion is

$$\lim_{n\to\infty} L_n = \pi. \tag{3.354}$$

It is not necessary to try to explain how a polygon (which *is* something but cannot *do* anything) can sprout more sides and approach the circle as the number of "its" sides becomes infinite. The number π appearing in this way is the length of circle of diameter 1. We are all familiar with the fact that the length of a circle having radius r and diameter d is $2\pi r$, or πd.

The ideas of this section have swarms of applications. In particular, we can use them to introduce some ideas and terminology of analytic geometry. We begin by considering the graph of the equation $y = f(x)$, where f is a given function. If

$$\lim_{x\to\infty} f(x) = L \qquad \text{or} \qquad \lim_{x\to-\infty} f(x) = L, \tag{3.36}$$

then the line having the equation $y = L$ is called a *horizontal asymptote* of the graph. If

$$\begin{aligned} &\lim_{x\to a+} f(x) = \infty \qquad \text{or} \qquad \lim_{x\to a+} f(x) = -\infty \\ \text{or}\quad &\lim_{x\to a-} f(x) = \infty \qquad \text{or} \qquad \lim_{x\to a-} f(x) = -\infty, \end{aligned} \tag{3.361}$$

then the line having the equation $x = a$ is called a vertical asymptote of the graph. Employing a modification of these ideas, we consider a case in which A and B are numbers such that

$$\lim_{x\to\infty} [f(x) - (Ax + B)] = 0 \qquad \text{or} \qquad \lim_{x\to-\infty} [f(x) - (Ax + B)] = 0. \tag{3.37}$$

In this case, the line having the equation $y = Ax + B$ is called an asymptote of the graph. This asymptote is horizontal if $A = 0$. We want to be able to apply similar jargon to graphs of equations, such as

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{3.38}$$

which are not graphs of functions. When we start with an equation of the form (3.38) and transpose all of the terms to the left side, we obtain an equation of the form

$$F(x,y) = 0. \tag{3.381}$$

If f is a function such that (3.381) is true when $y = f(x)$, then each asymptote of the graph of $y = f(x)$ is also an asymptote of the graph of (3.381). Problem 7 at the end of this section involves the famous equation (3.38).

Problems 3.39

1 Using epsilons appropriately, give a full statement of the meaning of each of the following truthful assertions. In case an assertion is so subtle that we are not yet prepared to prove it and appreciate its consequences, we need not be disturbed. Scientists can, for example, understand the assertion "there is helium in the sun" before they are able to prove the fact and understand the role of helium in the production of energy radiated by the sun.

(*a*) $\lim_{x\to\infty} \frac{1}{x} = 0$ (*b*) $\lim_{x\to-\infty} \frac{1}{x} = 0$

(*c*) $\lim_{x\to0+} \frac{1}{x} = \infty$ (*d*) $\lim_{x\to0-} \frac{1}{x} = -\infty$

(*e*) $\lim_{x\to\infty} \{x - \sqrt{x^2 - 1}\} = 0$ (*f*) $\lim_{x\to0+} \sqrt{x} = 0$

(*g*) $\lim_{x\to\pi/2-} \tan x = \infty$ (*h*) $\lim_{x\to0+} \log x = -\infty$

(*i*) $\lim_{x\to\infty} \log x = \infty$ (*j*) $\lim_{x\to\infty} \frac{\sin x}{x} = 0$

(*k*) $\lim_{x\to\infty} \frac{x+1}{x-1} = 1$ (*l*) $\lim_{x\to\infty} \left(1 + \frac{1}{x}\right)^x = e$

(*m*) $\lim_{x\to-\infty} \left(1 + \frac{1}{x}\right)^x = e$ (*n*) $\lim_{x\to\infty} \frac{\log x}{x} = 0$

(*o*) $\lim_{x\to0+} x \log x = 0$ (*p*) $\lim_{x\to\infty} x \log\left(1 + \frac{1}{x}\right) = 1$

(*q*) $\lim_{n\to\infty} 2^n = \infty$ (*r*) $\lim_{x\to\infty} e^x = \infty$

(*s*) $\lim_{x\to\infty} e^{-x} = 0$ (*t*) $\lim_{n\to\infty} |x|^n = 0, \ (|x| < 1)$

(*u*) $e^x = \lim_{n\to\infty} \left[1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}\right]$

(*v*) $\cos x = \lim_{n\to\infty}\left[1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n\frac{x^{2n}}{(2n)!}\right]$

(*w*) $\sin x = \lim_{n\to\infty}\left[x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^n\frac{x^{2n+1}}{(2n+1)!}\right]$

(*x*) $\sin \pi x = \lim_{n\to\infty}\left[\pi x\left(1 - \frac{x^2}{1^2}\right)\left(1 - \frac{x^2}{2^2}\right)\left(1 - \frac{x^2}{3^2}\right)\cdots\left(1 - \frac{x^2}{n^2}\right)\right]$

(*y*) $x! = \lim_{n\to\infty}\frac{n!n^x}{(x+1)(x+2)(x+3)\cdots(x+n)}\quad (x \neq -1, -2, -3, \cdots)$

(*z*) $\lim_{x\to -1+} x! = \infty$

2 Does the statement

$$\operatorname*{approx}_{\epsilon, n>N}\frac{1}{n} = 0$$

abbreviate the statement that to each positive number ϵ there corresponds an integer N such that $|1/n| < \epsilon$ whenever $n > N$? *Ans.:* It can, but it does only if we agree that it does. *Remark:* Whether the above abbreviation is better than the abbreviation

$$\lim_{n\to\infty}\frac{1}{n} = 0$$

is purely a matter of opinion. If a person has the habit of using one notation, the other must seem to be quite absurd, awkward, and unteachable.

3 Draw a graph of the equation $y = x^2$. Then, supposing that M is a given number, show how the figure can be used to support the assertion that

$$\lim_{x\to\infty} x^2 = \infty.$$

4 One of the assertions

$$\lim_{x\to 0}\frac{1}{x} = \infty\ (?),\qquad \lim_{x\to 0}\frac{1}{x^2} = \infty\ (?)$$

is true and the other is false. Give a full discussion of this matter. *Remark:* Here and elsewhere, displayed assertions followed by question marks may be false assertions.

5 With the aid of the idea that the numerator and denominator of the first quotient can be divided by x, show that

(*a*) $\lim_{x\to\infty}\frac{x-1}{x+1} = 1$ (*b*) $\lim_{x\to\infty}\frac{2x^2-1}{x^2+1} = 2$ (*c*) $\lim_{x\to-\infty}\frac{3x^3-1}{x^3+1} = 3$

(*d*) $\lim_{x\to\infty}\frac{x^2+2x+3}{2x^2-2x+3} = \frac{1}{2}$ (*e*) $\lim_{x\to\infty}\frac{x^2+2x+3}{x^3-2x+3} = 0$

(*f*) $\lim_{x\to\infty}\frac{x}{\sqrt{1+x^2}} = 1$ (*g*) $\lim_{x\to\infty}\frac{x}{\sqrt{1+x}} = \infty$

6 Show that both coordinate axes are asymptotes of the graph of the equation $y = 1/x$.

7 There will come a day when we must learn that the graph of the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

in which a and b are positive constants, is a hyperbola. Show that if the point $(x, y(x))$ lies on the hyperbola and $y(x) > 0$, then

$$y(x) = \frac{b}{a}\sqrt{x^2 - a^2}.$$

Show that

$$\lim_{x\to\infty}\left[\frac{b}{a}x - y(x)\right] = 0$$

and hence that the line having the equation $y = (b/a)x$ is an asymptote of the hyperbola. *Hint:* The formula

$$x - \sqrt{x^2 - a^2} = \frac{x - \sqrt{x^2 - a^2}}{1}\,\frac{x + \sqrt{x^2 - a^2}}{x + \sqrt{x^2 - a^2}}$$

turns out to be a useful source of information.

8 Find the equations of the asymptotes of the graphs of the equations

(a) $y = \dfrac{x+2}{x-1}$ — *Ans.:* $x = 1$, $y = 1$

(b) $y = \left(\dfrac{x+1}{x-2}\right)^2$ — *Ans.:* $x = 2$, $y = 1$

(c) $y = x + \dfrac{1}{x}$ — *Ans.:* $x = 0$, $y = x$

(d) $y = \left(x + \dfrac{1}{x}\right)^2$ — *Ans.:* $x = 0$

(e) $xy = x + y$ — *Ans.:* $x = 1$, $y = 1$

(f) $x^2 + y^2 = x + y$ — *Ans.:* None

(g) $y = \sqrt{x+1} - \sqrt{x}$ — *Ans.:* $y = 0$

9 According to part f of Problem 1, the first of the statements

$$\lim_{x\to 0+}\sqrt{x} = 0\ (?),\qquad \lim_{x\to 0}\sqrt{x} = 0\ (?)$$

is true. Is the second statement also true?

√ **10** Prove that

$$\lim_{n\to\infty}\left(\frac{1}{n^2} + \frac{2}{n^2} + \cdots + \frac{n}{n^2}\right) = \frac{1}{2}.$$

Remark: This remark is dedicated to unfortunate individuals who never knew or have forgotten that if

$$S_n = 1 + 2 + 3 + 4 + \cdots + (n-1) + n,$$

then

$$S_n = n + (n-1) + (n-2) + (n-3) + \cdots + 2 + 1$$

and addition gives $2S_n = n(n+1)$, so $S_n = n(n+1)/2$.

11 Starting with the definition

$$n! = 1\cdot 2\cdot 3 \cdots n, \tag{1}$$

which is applicable when n is a positive integer, show that $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, $6! = 720$, and $7! = 5040$. Then give a full statement of the reason or reasons why it is true that, when z is a positive integer,

$$z! = \lim_{n\to\infty} 1\cdot 2\cdot 3 \cdots z \tag{2}$$

$$z! = \lim_{n\to\infty} \frac{1\cdot 2\cdot 3 \cdots z(z+1)(z+2)\cdots(z+n)}{(z+1)(z+2)\cdots(z+n)} \tag{3}$$

$$z! = \lim_{n\to\infty} \frac{n!n^z}{(z+1)(z+2)\cdots(z+n)}\,\frac{n+1}{n}\,\frac{n+2}{n}\cdots\frac{n+z}{n} \tag{4}$$

$$z! = \lim_{n\to\infty} \frac{n!n^z}{(z+1)(z+2)\cdots(z+n)}. \tag{5}$$

Remark: To show that the above manipulations serve a useful purpose, we take a little mental excursion. A *complex number* z is a number of the form $x + iy$, where x and y are real numbers and i is the *imaginary unit* for which $i^2 = -1$. While this book neither develops nor uses the algebra and calculus of complex numbers, we remark that $x + iy$ is the real number x if $y = 0$ and that $x + iy$ is a real integer if $y = 0$ and x is a real integer. We are now ready to look at (5). We have seen that (5) is correct if z is a positive integer and the definition (1) is applicable. While proof of the fact lies far beyond our present capabilities, it can be proved that the limit in the right member of (5) exists and is a complex number whenever z is a complex number which is not a negative integer. Moreover, when z is a complex number which is not a negative integer, $z!$ *is defined to be this limit.* It follows from the definition that $z!$ is a real number whenever z is a real number which is not a negative integer. Carl Friedrich Gauss (1777–1855), who had the habit of knowing how things should be done, made very effective use of (5). The index can always show us where this and other information about factorials is concealed.

12 If the preceding problem and remark have been digested, prove that $0! = 1$. *Remark:* Proof of the more esoteric facts that $(-\frac{1}{2})! = \sqrt{\pi}$ and $(\frac{1}{2})! = \sqrt{\pi}/2$ will not be too difficult when more mathematics of the right kind has been learned.

13 Observe that $8! = 8(7!)$. Then, assuming that the limits exist, prove that

$$\lim_{n\to\infty} \frac{n!n^{z+1}}{(z+1+1)(z+1+2)\cdots(z+1+n)} = (z+1)\lim_{n\to\infty} \frac{n!n^z}{(z+1)(z+2)\cdots(z+n)}.$$

Finally, use the remark of Problem 11 to prove that

$$(z+1)! = (z+1)(z!)$$

when z is not a negative integer.

14 For what pairs of numbers n and k does it make sense to define the *binomial coefficient function* by the formula

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

Hint: If necessary, read Problem 11. *Ans.:* When n, k, and $n - k$ are numbers (real or complex) that are not negative integers.

15 Try to make friends of the contents of the preceding problems by proving that

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

when n, $k - 1$, and $n - k$ are not negative integers. *Remark:* As some people learn while studying algebra, the ordinary binomial coefficients (in which n and k are integers for which $0 \leqq k \leqq n$) are the coefficients appearing in the formulas

$$(a + b)^0 = 1 \tag{2}$$
$$(a + b)^1 = a + b \tag{3}$$
$$(a + b)^2 = a^2 + 2ab + b^2 \tag{4}$$
$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 \tag{5}$$

and, in general, in the *binomial formula*

$$(a + b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \cdots + \binom{n}{n} a^0 b^n. \tag{6}$$

With the aid of (1), it is easy to fill in the rows of the *Pascal triangle*

$$\begin{array}{ccccccccccc} &&&&& 1 &&&&& \\ &&&& 1 && 1 &&&& \\ &&& 1 && 2 && 1 &&& \\ && 1 && 3 && 3 && 1 && \\ & 1 && 4 && 6 && 4 && 1 & \\ 1 && 5 && 10 && 10 && 5 && 1 \end{array}$$

which displays binomial coefficients. The sum of two consecutive elements of one row gives the element that lies below the space between them, and more rows of the Pascal triangle are easily written.

16 We can feel sure that if $x > 1$, then x^n must be large whenever n is large, but it is nevertheless worthwhile to be able to prove the precise version of the statement. When $x > 1$, there is a positive number h such that

$$x = 1 + h; \tag{1}$$

in fact, $h = x - 1$. Observe that

$$x^2 = 1 + 2h + h^2 > 1 + 2h \tag{2}$$
$$x^3 = 1 + 3h + 3h^2 + h^3 > 1 + 3h \tag{3}$$

and that the binomial formula shows that

$$x^n > 1 + nh \tag{4}$$

when $n \geqq 2$. It follows that if M is a given number and we choose a number N such that $N \geqq 2$ and $N > M/h$, then we will have

$$x^n > 1 + nh > nh > M \tag{5}$$

whenever $n > N$. Therefore,

$$\lim_{n\to\infty} x^n = \infty \qquad (x > 1). \tag{6}$$

17 We can feel sure that if $|x| < 1$, then x^n is near 0 whenever n is large. How can we prove it? *Solution:* Let $\epsilon > 0$. Suppose first that $x = 0$. Then $|x^n| < \epsilon$ when $n > 1$. Suppose finally that $0 < |x| < 1$. Let $y = 1/|x|$ so that $y > 1$. Then the preceding problem shows that

$$\lim_{n\to\infty} y^n = \infty.$$

If we choose an index N such that $y^n > 1/\epsilon$ when $n > N$, then

$$|x^n| = \frac{1}{y^n} < \epsilon$$

when $n > N$. Therefore,

$$\lim_{n\to\infty} x^n = 0 \qquad (|x| < 1).$$

18 Prove that

$$\lim_{n\to\infty}\left[\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^n}\right] = 1$$

$$\lim_{n\to\infty}\,[1 + x + x^2 + \cdots + x^n] = \frac{1}{1-x} \qquad (|x| < 1).$$

Hint: Long division (or factoring) shows that

$$1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x},$$

and we may use the fact that $\lim_{n\to\infty} x^n = 0$ when $|x| < 1$.

19 Once again, let the "bracket symbol" $[x]$ denote the "greatest integer in x," that is, the greatest integer less than or equal to x, so that $[8] = 8$ and $[15.359] = 15$. Show that, for each integer n,

$$\lim_{x\to n+} [x] = n, \qquad \lim_{x\to n-} [x] = n - 1.$$

20 Prove that if g is a function and A and B are numbers such that $|g(x)| \leqq A$ whenever $x \geqq B$, then

$$\lim_{x\to\infty} \frac{g(x)}{x} = 0$$

and that, if L is a number, then

$$\lim_{x\to\infty}\left(L + \frac{g(x)}{x}\right) = L.$$

21 Prove that

$$\lim_{x\to\infty} \frac{[x]}{x} = 1.$$

Hint: Let $\theta(x)$, read "theta of x," denote the "fractional part" of x so that $\theta(x) = x - [x]$ and $[x] = x - \theta(x)$.

22 Sketch a graph of the function h for which $h(x) = [x]/x$ when $x \geqq 1$, and observe that $h(x)$ really is near 1 whenever x is large.

23 Sometime we will learn that

$$\lim_{n\to\infty} \frac{n^3}{2^n} = 0. \tag{1}$$

Hence there must be an integer N such that

$$\frac{n^3}{2^n} < \frac{1}{100} \tag{2}$$

when $n > N$. Some numerical calculations can make us quite sure that (2) is valid when $n > 20$. Even though the author considers the problem to be too difficult for assignment at this time, it may be worthwhile to seek a way to determine whether (2) is valid when $n > 20$.

24 Prove that if x is a rational number, say p/q, where p and q are integers, then $\sin n!\pi x = 0$ for each sufficiently great integer n. Prove that if x is an irrational number, then $\sin n!\pi x \neq 0$ for each integer n. Using these results, show that

$$1 - \lim_{n\to\infty} \operatorname{sgn} \sin^2 n!\pi x = D(x)$$

where D is the dizzy dancer function for which $D(x) = 1$ when x is rational and $D(x) = 0$ when x is irrational.

25 Some old analytic geometry books pretend to prove that if n is a positive integer and $P_0, P_1, \cdots, P_n$ are polynomials in x, then the line having the equation $x = x_1$ will be an asymptote of the graph of the equation

$$P_0(x)y^n + P_1(x)y^{n-1} + \cdots + P_{n-1}(x)y + P_n(x) = 0 \tag{1}$$

provided $P_0(x_1) = 0$. These old books present unclear and unreliable treatments of matters involving limits and asymptotes, however, and the stated result is false. Prove that the line having the equation $x = 0$ is not an asymptote of the graph of the equation

$$x^2y^2 + x^2y + 1 = 0. \tag{2}$$

Remark: An example which establishes falsity of an assertion is called a counterexample. Persons who speak German (and many others also) call it a *Gegenbeispiel*. The simpler equation $x^2y^2 + 1 = 0$ serves the present purpose; the graph of this equation is the empty set.

26 Prove that if f_1, f_2, f_3 are continuous at a, if

$$\lim_{x\to a+} y(x) = \infty, \tag{1}$$

and if, for some positive number δ,

$$f_1(x)[y(x)]^2 + f_2(x)y(x) + f_3(x) = 0 \qquad (a < x < a + \delta), \tag{2}$$

then $f_1(a) = 0$. *Hint:* Choose a positive number δ_1 such that $\delta_1 < \delta$ and $y(x) > 1$ when $a < x < a + \delta_1$. Then, supposing that $a < x < a + \delta_1$, divide the members of (2) by $[y(x)]^2$ to obtain

$$f_1(x) + \frac{f_2(x)}{y(x)} + \frac{f_3(x)}{[y(x)]^2} = 0.$$

27 For hundreds of years, people have been interested in the magnitude of $\pi(x)$, the number of primes less than or equal to x, when x is large. About the year 1900, mathematicians succeeded in proving a remarkable fact that had been surmised since the time of Euler (1707–1783). It was proved that

$$(*) \qquad \lim_{x\to\infty} \frac{\pi(x)}{\dfrac{x}{\log x}} = 1.$$

We may know very little about logarithms and may not yet have learned that, in mathematics above the level of elementary trigonometry, $\log_{10} x$ denotes the logarithm of x with base 10 and $\log x$ denotes the logarithm of x with base e. We may not yet know how to calculate $\log x$ when x is a given positive number. Nevertheless we should be able to tell the meaning of the star formula. Do it. *Remark:* Anyone who wishes to make a very modest calculation may use the fact that log 20 is approximately 3 and may determine $\pi(20)$. When working on chalk boards and scratch pads, many people make effective use of stars and daggers and other things (instead of numbers) to designate significant formulas. The valuable idea is illustrated only occasionally in this book.

28 It is sometimes said that mathematics is a language. Perhaps it would be more sensible to say that mathematics is a collection of ideas and that mathematics books use language in more or less successful attempts to reveal the ideas. In any case, language is important and definitions constitute a basic part of this language. To help us realize this fact, we consider an example involving regular polygons. A *regular polygon* is a set in E_2 consisting of the points on the line segments $P_0P_1, P_1P_2, \cdots, P_{n-1}P_n, P_nP_0$, where the points $P_0, P_1, \cdots, P_n, P_0$ are equally spaced on a circle, n being an integer for which $n \geqq 3$. Under this definition, a circle is *not* a regular polygon. We do not have pencils sharp enough to draw regular polygons having a million sides, but we can nevertheless tolerate the idea that if we could draw one on an ordinary sheet of paper, then the result would *look* like a circle. We cannot, however, tolerate the ancient collection of words "a circle *is* a regular polygon having an infinite number of infinitesimally small sides" as a part of our doctrine of limits. To take a sensible view of this matter, we can know that there was a time when the best of our scientific ancestors used fuzzy language and whale-oil lamps but we can also know that they worked mightily to produce better products.

29 As was stated in Section 1.1, a number x appearing in this book is a real number unless an explicit statement to the contrary is made. This circumstance does not prohibit recognition of the fact that numbers other than real numbers can appear in mathematics. It is possible, and is sometimes worthwhile, to define and employ a set S^* of numbers which contains each real number x in the set S of real numbers and, in addition, two numbers $-\infty$ and ∞. When the set S^* is employed, each real number x is said to be *finite* and the numbers $-\infty$ and ∞ are said to be *infinite* (not finite). Order relations are introduced in such a way that $-\infty < \infty$ and $-\infty < x < \infty$ whenever x is a real number. While these order relations are simple and attractive, it turns out to be impossible to formulate a useful collection and algebraic laws (or postulates) in such a way that $\infty - \infty$ and $0\cdot\infty$ are numbers in S^*. Persons starting with enthusiasm for $-\infty$ and ∞ usually lose most of their fascination when they learn that the relations $0\cdot\infty = 1$, $\frac{1}{0} = \infty$, and $\frac{1}{0} = -\infty$ are as absurd in the "algebra" of S^* as the symbol $\frac{1}{0}$ is in the algebra of S. We can be momentarily delighted by the "algebraic law" which says that $\infty + x = \infty$ whenever x is a real number, but general usefulness of the unorthodox "algebra" is greatly impaired by the fact that the relation $y + x = y$ does not imply that $x = 0$ because y might be ∞ and x might be 416. For present purposes, we do not need substantial information about these matters, but a little basic information can be very helpful. There are circumstances in which $-\infty$ and ∞ are considered to be numbers, but there are no circumstances in which $-\infty$ and ∞ are real numbers to which we can apply the algebraic rules (or laws or axioms or postulates) that apply to real numbers. Whether or not we consider $-\infty$ and ∞ to be numbers, it is worthwhile to recognize that some of the most convenient terminologies and notations of modern mathematics are relics of times when the "doctrine of limits" was based upon visions of a number x galloping toward infinity and becoming so infinitely great (but still not ∞) that its reciprocal becomes infinitesimally small (but still not 0). These infinitesimals of mathematics, like the aether and phlogiston of physics and chemistry, can now be regarded as mystic absurdities, but they were hardy concepts having tremendous impacts upon present as well as past science and philosophy. We can conclude these remarks with another bit of history. In the good old days when mathematical terminology was incredibly erratic, sane physicists got the habit of saying that a number is "finite" when they wished to emphasize their idea that it is neither zero nor infinite nor infinitely small nor infinitely large. It will be interesting to see how long physicists continue to make modern mathematicians shudder by using the word "finite" to mean "good honest nonzero noninfinite number, with no nonsense." The physicists have good intentions, but mathematicians consider zero to be a finite number, with no nonsense.

3.4 *Continuity* This section contains information about functions and limits that we will need. Our first task is to obtain a full understanding of the following definition.

Definition 3.41 *A function f is continuous at x_0* (or at the point with coordinate x_0, or at the point x_0) *if*

$$\lim_{x\to x_0} f(x) = f(x_0).$$

The assertion that f is continuous at x_0 is nothing more nor less than the assertion that $f(x)$ is near $f(x_0)$ whenever x is near x_0. It means that to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that

$$|f(x) - f(x_0)| < \epsilon \qquad (|x - x_0| < \delta). \tag{3.42}$$

The definition implies that f cannot be continuous at x_0 unless $f(x_0)$ exists, that is, unless x_0 belongs to the domain of f. In case $f(x_0)$ exists, the first inequality in (3.42) automatically holds when $x = x_0$ and we do not need to bother with the restriction $x \neq x_0$ that appears in the definition of limit. With a small change in notation, we can see that f is continuous at x if and only if

$$\lim_{\Delta x \to 0} f(x + \Delta x) = f(x) \tag{3.421}$$

or

$$\lim_{\Delta x \to 0} [f(x + \Delta x) - f(x)] = 0. \tag{3.422}$$

Figure 3.43 shows, for the case in which $f(x) = x^2$ and $\Delta x > 0$, the geometric interpretations that can be given to the numbers appearing in these formulas.

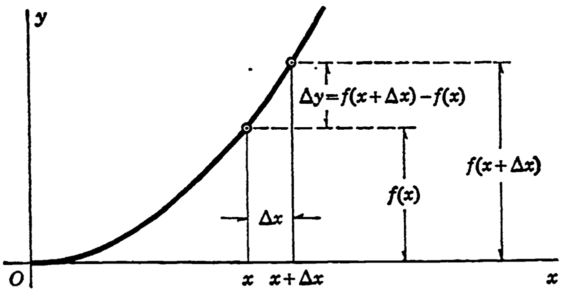


Figure 3.43

Definition 3.44 *A function f is said to have right-hand continuity at a if the first of the assertions*

$$\lim_{x \to a+} f(x) = f(a), \qquad \lim_{x \to b-} f(x) = f(b) \tag{3.441}$$

is valid and to have left-hand continuity at b if the second is valid.

Supposing that $a < b$, we can let f_1 be the function having the graph in Figure 3.442 so that $f_1(x) = 0$ when $x < a$, $f_1(x) = 1$ when $a \leqq x \leqq b$, and $f_1(x) = 0$ when $x > b$. This function is continuous at each x for which $x \neq a$ and $x \neq b$. The function has right-hand continuity at a

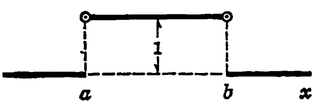


Figure 3.442

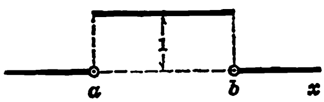


Figure 3.443

and has left-hand continuity at b. It does not have left-hand continuity at a, and it does not have right-hand continuity at b. Let f_2 be the function having the graph in Figure 3.443 so that $f_2(x) = 0$ when $x \leqq a$, $f_2(x) = 1$ when $a < x < b$, and $f_2(x) = 0$ when $x \geqq b$. This function, like f_1, is continuous except when $x = a$ and $x = b$. However, f_2 has left-hand continuity at a and right-hand continuity† at b.

Definition 3.45 *A function f is continuous over an interval $a \leqq x \leqq b$ if it is continuous at each x_0 for which $a < x_0 < b$ and, in addition, has right-hand continuity at a and left-hand continuity at b.*

The definitions of this section are designed to be useful in discussions of examples of functions, and we begin by looking at examples of functions. Let g be the function, defined for $x \neq 0$, for which

$$g(x) = \frac{1}{x} \qquad (x \neq 0). \tag{3.451}$$

This function is continuous at each $x_0 \neq 0$ because, when $x_0 \neq 0$, our theorems on limits imply that

$$\lim_{x \to x_0} g(x) = \lim_{x \to x_0} \frac{1}{x} = \frac{1}{\lim\limits_{x \to x_0} x} = \frac{1}{x_0} = g(x_0). \tag{3.452}$$

However, g cannot be continuous at 0, because $g(0)$ is undefined and there is no possibility of having $\lim\limits_{x \to 0} g(x) = g(0)$. We say that g is discontinuous at 0. Now let h be the function defined over $-\infty < x < \infty$ (this means merely that the domain of h is the entire set of numbers) by $h(0) = 0$ and

$$h(x) = \frac{1}{x} \qquad (x \neq 0). \tag{3.453}$$

This function, like g, is continuous at each $x_0 \neq 0$, but this time $h(0)$ exists and there is no possibility of having $\lim\limits_{x \to 0} h(x) = h(0)$, because $\lim\limits_{x \to 0} h(x)$ does not exist. Let ω (omega) be the peculiar function for which $\omega(0) = 1$ and $\omega(x) = 0$ when $x \neq 0$. For this function both $\omega(0)$ and $\lim\limits_{x \to 0} \omega(x)$ exist, but the function is discontinuous at 0 because

$$\lim_{x \to 0} \omega(x) = 0 \neq 1 = \omega(0). \tag{3.454}$$

† It is to be expected that some readers, particularly those more interested in applied mathematics than in pure mathematics, may feel that matters now being considered are much too theoretical to have practical interest. Some people know, and others can learn, that when a battery has its terminals connected to appropriate electrical hardware, it almost instantly produces an electromotive force (the kind of a force that pushes or pulls electrons around) which we may, for present purposes, suppose to have the constant value 1. When the battery is not connected, the electromotive force produced by it is 0. Thus, batteries which are connected over some time intervals, and disconnected over other time intervals, produce electromotive forces that are, as functions of time, very closely approximated by step functions such as those we have been considering. The discontinuous functions are introduced to simplify problems, not to complicate them. This is one of the reasons why persons interested in applications of science must recognize existence of discontinuous functions.

The graphs of the signum and Heaviside functions shown in Figures 3.191 and 3.192 should indicate that these functions are continuous everywhere except at $x = 0$. One who has seen numerous examples of functions and their graphs should realize that he can enter the construction business to produce more examples. He can start with a clean coordinate system and, as in Figures 3.46 and 3.47, mark points $\pm x_1$, $\pm x_2$, $\pm x_3$, $\cdots$

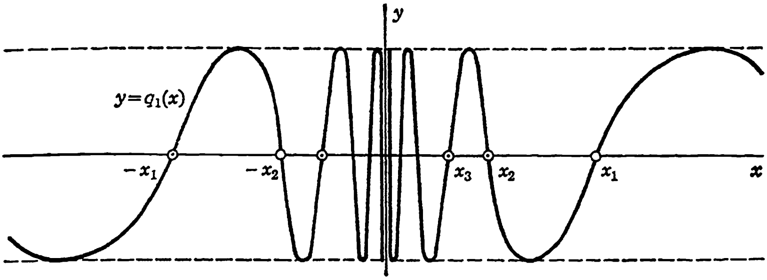


Figure 3.46

on the x axis and then sketch a part of a graph which oscillates through these points in any way he likes. Provided only that the graph contains no two different points having the same x coordinate, the graph will be the graph of a function. In Figure 3.46 the graph is drawn tangent over and over again to the lines having equations $y = 1$ and $y = -1$. In Figure 3.47 the graph is drawn tangent over and over again to the

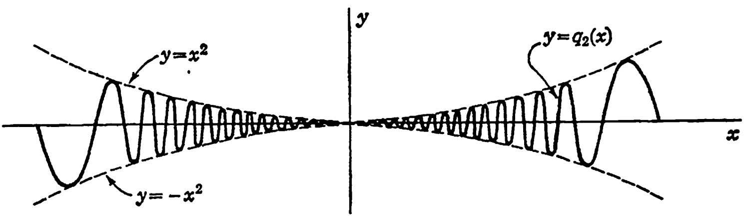


Figure 3.47

parabolas having the equations $y = x^2$ and $y = -x^2$. It can be shown that the graphs of the functions defined by $q_1(x) = \sin(1/x)$ when $x \neq 0$ and $q_2(x) = x^2 \sin(1/x)$ when $x \neq 0$ and $q_2(0) = 0$ look very much like the graphs in Figures 3.46 and 3.47, but we need not worry about this matter now. It should be clear from Figure 3.46 that q_1 cannot be continuous at $x = 0$ because $\lim_{x\to 0} q_1(x)$ does not exist. For the function q_2 the story is different. Since

$$-x^2 \leqq q_2(x) \leqq x^2, \tag{3.471}$$

it follows from the sandwich (or flyswatter) theorem that

$$(3.472) \qquad \lim_{x\to 0} q_2(x) = 0 = q_2(0),$$

so q_2 must be continuous† at $x = 0$.

It is easy to prove fundamental facts about functions formed by combining continuous functions in various ways. With the aid of Theorem 3.285 on limits, we see that if $h(x) = f(x) + g(x)$ over an interval containing x_0, and if f and g are continuous at x_0, then

$$\lim_{x\to x_0} h(x) = \lim_{x\to x_0} f(x) + \lim_{x\to x_0} g(x) = f(x_0) + g(x_0) = h(x_0).$$

This shows that the sum of two continuous functions is continuous whereever the terms being added are both continuous. Very similar arguments show that the product of two continuous functions is continuous wherever the factors are continuous and that the quotient of two continuous functions is continuous whenever the numerator and denominator are continuous and the denominator is not zero.

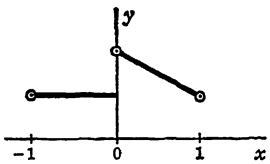


Figure 3.48

We should now see that the function f, which is defined over the interval $-1 \leqq x \leqq 1$ and which has the graph shown in Figure 3.48, is continuous over the interval $0 \leqq x \leqq 1$; it is continuous at each x_0 for which $0 < x_0 < 1$, it has right-hand continuity at 0, and it has left-hand continuity at 1. As a bonus for knowing about limits, unilateral limits, and continuity, we find that we can easily understand and remember some fundamental facts that are frequently used in applied as well as in pure mathematics. A function f has a limit as x approaches a if and only if the two unilateral (right and left) limits exist and are equal. The function is continuous at a if and only if the two unilateral limits exist and are equal to $f(a)$.

Problems 3.49

1 The statement that

$$5x^3 + 2x^2 - 4x + 16$$

is continuous is an abbreviation of the statement that the polynomial function P having values $P(x)$ defined by the formula

$$P(x) = 5x^3 + 2x^2 - 4x + 16$$

† It has sometimes been thought to be meaningful, and perhaps even true or helpful or both, to say that a function f is continuous if and only if "it is possible to draw the graph of f without lifting the pencil from the paper." Enthusiasm for this statement must be chilled when we realize that a continuous function may have an infinite set of oscillations in a finite interval and that feeble mortals never succeed in drawing more than a finite set of them.

is continuous. Prove the statement by filling in the intermediate steps in the formula

$$\lim_{x\to a} P(x) = \cdots = P(a)$$

and tell which theorems on limits are used in the process. *Remark:* The same procedure shows that each polynomial is continuous.

2 Letting

$$Q(x) = \frac{(x-1)(x-3)}{(x-2)(x-4)},$$

show that Q is continuous at each x except 2 and 4.

3 Prove that the quotient of two functions is continuous wherever both functions are continuous and the denominator is not zero. *Remark:* We recall that the quotient of two polynomials is sometimes called a rational function. Our results show that a rational function is continuous wherever the denominator is not zero.

4 Determine the points of discontinuity of the functions f_1, etcetera, for which

(*a*) $f_1(x) = \dfrac{x}{1+x^2}$ (*b*) $f_2(x) = \dfrac{x}{1-x^2}$

(*c*) $f_3(x) = \dfrac{1}{x(1-x)}$ (*d*) $f_4(x) = |x|$

(*e*) $f_5(x) = \dfrac{1}{x^2+2x-3}$ (*f*) $f_6(x) = \dfrac{1}{x^2+2x+3}$

5 Does the assertion

$$\operatorname*{approx}_{\epsilon,|x-a|<\delta} f(x) = f(a)$$

abbreviate the assertion that to each positive number ϵ there corresponds a positive number δ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$? *Ans.:* It can, but it does only if we agree that it does.

6 Taxi fare is 50 cents plus 10 cents for each quarter mile or fraction thereof. Letting $f(x)$ denote the fare for a ride of x miles, sketch a graph of f and tell where f is discontinuous.

7 Assume (as is not quite true) that it takes 0.5 calorie of heat to raise the temperature of 1 gram of ice 1 degree centigrade, that it takes 80 calories to melt the ice at 0°C, and that it takes 1.0 calorie to raise the temperature of 1 gram of water one degree centigrade. Supposing that $-40 \leqq x \leqq 20$, let $Q(x)$ be the number of calories of heat required to raise one gram of H_2O from temperature -40°C to x°C. Sketch a graph of Q. *Ans.:* Figure 3.491.

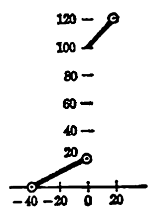


Figure 3.491

8 The magnitude of the gravitational force which the earth exerts upon a particle is called its weight W. Suppose (as would be true in the mechanics of Newton if the earth were a homogeneous spherical ball) that there exist constants k_1 and k_2 such that

$$W = k_1x \qquad (0 \leqq x < R)$$

$$W = \frac{k_2}{x^2} \qquad (x \geqq R),$$

where x is the distance from the center of the earth to the particle and R is the radius of the earth. Supposing that W is a continuous function of x and that $W = 100$ when $x = R$, calculate k_1 and k_2 and sketch a graph of W versus x.

9 Prove that if f is continuous at x_0, then so also is the function g having values defined by $g(x) = |f(x)|$.

10 It is never too soon to start becoming acquainted with the idea that if, during some time interval $t_1 < t < t_2$, a bumblebee or molecule or rocket buzzes around, then at each time t in the interval it is surely someplace and that if we let $f_1(t)$, $f_2(t)$, $f_3(t)$ denote its x, y, z coordinates at time t, then f_1, f_2, f_3 are continuous functions of t. Since wholesome comprehension of mathematics is salubrious, we recognize that we do not quite know how to prove that bumblebees never fly out of our E_3 for a minute or two. Moreover, we do not know how to devise a mathematical proof that a bumblebee cannot gather honey all morning in Pennsylvania, be in Chicago at noon, and hunt clover in Los Angeles all afternoon. The best we can do is make the physical assumption that f_1, f_2, f_3 are continuous and know what the assumption means. What does the assumption mean? *Ans.:* If $t_1 < t < t_2$, then

$$\lim_{\Delta t \to 0} f_k(t + \Delta t) = f_k(t)$$

when $k = 1$, when $k = 2$, and when $k = 3$.

11 Abandoning some of the notation of the preceding problem, we suppose that x, y, z are given functions that are continuous over some interval in which t is supposed to lie. Let $P(t)$ denote the point in E_3 having coordinates x, y, z for which $x = x(t)$, $y = y(t)$, and $z = z(t)$. While the fact will be considered later with more details, we can pause to learn that the ordered set of points $P(t)$, ordered so that $P(t')$ precedes $P(t'')$ when $t' < t''$, is called a *curve C*. The point $P(t)$ is then said to *move along or traverse the curve C as t increases*. Figure 3.492 may be helpful. For each t, let $\mathbf{r}(t)$ be the vector running from the origin O to $P(t)$. This determines a *vector function* $\mathbf{r}$ for which

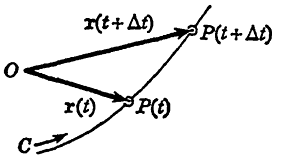


Figure 3.492

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}. \tag{1}$$

Conversely, if $\mathbf{r}$ is a given vector function, then it (and the given coordinate system) determines its scalar components. From (1) and

$$\mathbf{r}(t + \Delta t) = x(t + \Delta t)\mathbf{i} + y(t + \Delta t)\mathbf{j} + z(t + \Delta t)\mathbf{k} \tag{2}$$

we obtain

$$\mathbf{r}(t + \Delta t) - \mathbf{r}(t) = [x(t + \Delta t) - x(t)]\mathbf{i} + [y(t + \Delta t) - y(t)]\mathbf{j} + [z(t + \Delta t) - z(t)]\mathbf{k} \tag{3}$$

and

$$|\mathbf{r}(t + \Delta t) - \mathbf{r}(t)| = [|x(t + \Delta t) - x(t)|^2 + |y(t + \Delta t) - y(t)|^2 + |z(t + \Delta t) - z(t)|^2]^{1/2}. \tag{4}$$

As is easy to guess, the vector function $\mathbf{r}$ is said to be continuous at t if

$$\lim_{\Delta t \to 0} |\mathbf{r}(t + \Delta t) - \mathbf{r}(t)| = 0, \tag{5}$$

and we write

$$\lim_{t \to t_0} \mathbf{r}(t) = \mathbf{w} \tag{6}$$

if $\mathbf{w}$ is a vector for which

$$\lim_{t \to t_0} |\mathbf{r}(t) - \mathbf{w}| = 0. \tag{7}$$

It is a consequence of (4) that *a vector function is continuous if and only if its scalar components are continuous.*

12 Using ideas from the preceding problem, let

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$$
$$\mathbf{w} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

and prove that $\lim_{t \to t_0} \mathbf{r}(t) = \mathbf{w}$ if and only if

$$\lim_{t \to t_0} x(t) = a, \qquad \lim_{t \to t_0} y(t) = b, \qquad \lim_{t \to t_0} z(t) = c.$$

Hint: Write and use a formula for $|\mathbf{r}(t) - \mathbf{w}|$.

13 Once again, let the symbol $[q]$ denote the greatest integer which is less than or equal to q. Let f be the function for which

$$f(x) = \left[\frac{1}{x}\right]$$

when $x > 0$. Draw the graph of f and tell where f is discontinuous.

14 Using the "bracket notation" of the preceding problem, determine whether

$$\lim_{x \to 0+} x\left[\frac{1}{x}\right]$$

exists.

15 Letting D be our old friend, the dizzy dancer function, for which

$$D(x) = 0 \qquad (x \text{ irrational})$$
$$D(x) = 1 \qquad (x \text{ rational}),$$

show that there is no a for which

$$\lim_{x \to a} D(x) = D(a)$$

and hence that this function is everywhere discontinuous.

16 A potential new friend g is defined over the closed interval $0 \leqq x \leqq 1$ in an interesting way. If x is irrational, then $g(x) = 0$. If x is 0, then $g(x) = 1$, and if $x = 1$, then $g(x) = 1$. If x is a rational number for which $0 < x < 1$ and if $x = m/n$, where m and n are positive integers having no common positive integer factor exceeding 1, then $g(x) = 1/n$. Thus $g(\frac{1}{2}) = \frac{1}{2}$, $g(\frac{1}{3}) = \frac{1}{3}$, $g(\frac{2}{3}) = \frac{1}{3}$, $g(\frac{1}{4}) = \frac{1}{4}$, $g(\frac{3}{4}) = \frac{1}{4}$, $g(\frac{1}{5}) = \frac{1}{5}$, $g(\frac{2}{5}) = \frac{1}{5}$, etcetera. Sketch a figure indicating the nature of the graph of g. Show that g is discontinuous at each x for which x is rational and that g is continuous at each x for which x is irrational. *Hint:* If ϵ is a given positive number, then the set of numbers x for which $g(x) > \epsilon$ contains only a finite number of elements. This fact is useful. *Remark:* While interest in the matter should be postponed, this is an example of a bounded function having a countably infinite set of discontinuities. Moreover, each subinterval of the interval $0 \leqq x \leqq 1$ contains an infinite set of these discontinuities, but the set of discontinuities has Lebesgue measure zero. The function g is the famous *corn-popper function.*

17 Some people know very much about the function F for which $F(r)$ is the number of lattice points (points having integer coordinates) lying inside and on the circle of radius r having its center at the origin. Give at least a little precise information about F.

18 Give an example of a function f such that $0 \leqq f(x) \leqq 1$ when $0 \leqq x \leqq 1$ and such that f is continuous at each point of the interval $0 \leqq x \leqq 1$ except at $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \cdots$.

19 Give an example of a function which (i) is defined over the closed interval $0 \leqq x \leqq 1$, (ii) is continuous over the open interval $0 < x < 1$, and (iii) is not continuous over the closed interval $0 \leqq x \leqq 1$.

20 Show that if x_1, x_2, x_3 and A, B, C, D, E are constants for which $x_1 < x_2 < x_3$ and $C \neq 0$, $D \neq 0$, $E \neq 0$, and if

$$f(x) = Ax + B + C|x - x_1| + D|x - x_2| + E|x - x_3|,$$

then f is continuous and the graph of f is a broken line consisting of line segments joined at vertices whose x coordinates are x_1, x_2, x_3.

21 Let

$$\begin{aligned} &f(x) = -x && (x \leqq 0) \\ &f(x) = x && (0 \leqq x \leqq 1) \\ &f(x) = 2 - x && (1 \leqq x \leqq 2) \\ &f(x) = 0 && (x \geqq 2), \end{aligned} \tag{1}$$

so that the graph of f is a broken line having corners at the points (0,0), (1,1), and (2,0). Determine five constants A, B, C, D, E such that

$$f(x) = Ax + B + C|x| + D|x - 1| + E|x - 2|. \tag{2}$$

Hint: For each of the four intervals $x \leqq 0$, $0 \leqq x \leqq 1$, $1 \leqq x \leqq 2$, and $x \geqq 2$, replace the left member of (2) by the appropriate expression and replace the right member of (2) by the appropriate expression not involving absolute-value signs. *Ans.:*

$$f(x) = -\tfrac{1}{2}x + |x| - |x - 1| + \tfrac{1}{2}|x - 2|.$$

3.5 *Difference quotients and derivatives*

Let f be defined over an interval $a \leqq x \leqq b$ and let x be a number for which $a < x < b$. Let Δx be a number, which may be positive or negative but not 0, for which $a \leqq x + \Delta x \leqq b$. We may then set

$$y = f(x), \qquad y + \Delta y = f(x + \Delta x), \tag{3.51}$$

subtract to obtain

$$\Delta y = f(x + \Delta x) - f(x),$$

and then divide by Δx to obtain

$$\frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}. \tag{3.52}$$

This quotient, which is clearly a quotient of differences that are calculated in a special way, is called a *difference quotient.* Difference quotients have already appeared in our problems, and we shall see later that they have important interpretations. Leaving the hosts of applications to be partially revealed later in this textbook, and to be continually revealed to those who pursue further studies in the sciences (including mathematics), we now come to one of the two most important ideas in the calculus. If the difference quotient in (3.52) has a limit as Δx approaches zero, then f is said to be *differentiable* at x and the limit is called the *derivative* of f at x. In case the limit fails to exist, the function is said to be nondifferentiable at x and we say that the derivative of f at x does not exist. There are two very different and very useful notations for derivatives. The first, appearing in the formula

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}, \tag{3.53}$$

is usually read "eff prime of ex," but it can be read "eff prime at x" or "the derivative of f at x." This "prime notation" is called the *Newton* (1642–1727) *notation.*† The second notation, appearing in the formula

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}, \tag{3.54}$$

is read "dee y dee x" or "the derivative of y with respect to x" and was originated by Leibniz‡ (1646–1716). There will be times in the future when we will consider dy/dx to be the quotient of the two numbers dy and dx. Meanwhile, the whole symbol dy/dx is to be regarded as a single symbol, just as the symbol H represents a single letter of the alphabet and not 11 divided by 11. A longer and perhaps dismal discussion of this terminology and notation appears in a remark at the end of the problems of this section; congratulations can be bestowed upon readers wise enough to know that the discussion is semisuperfluous.

According to an old and honorable tradition, the definition of dy/dx and

† The original Newton notation was the "dot notation" or the "flyspeck notation" which employed $\dot{f}$ instead of f', but replacing the dot by the prime is a clerical modification that preserves the original idea of Newton.

‡ Leibniz, like Newton, published his scientific works in Latin. The Latin spelling "Leibnitz" is sometimes seen and sometimes helps people to pronounce the name correctly.

the manner in which it is applied can be (or should be) remembered with the aid of the famous "four-step rule." We may not always get 4 when we count the steps, but the rule is the four-step rule anyway.

Four-step rule 3.55

Definition	*Application*
$y = f(x)$	$y = x^2$
$y + \Delta y = f(x + \Delta x)$	$y + \Delta y = (x + \Delta x)^2 = x^2 + 2x\,\Delta x + \Delta x^2$
$\Delta y = f(x + \Delta x) - f(x)$	$\Delta y = 2x\,\Delta x + \Delta x^2$
$\dfrac{\Delta y}{\Delta x} = \dfrac{f(x + \Delta x) - f(x)}{\Delta x}$	$\dfrac{\Delta y}{\Delta x} = 2x + \Delta x$
$\dfrac{dy}{dx} = \lim\limits_{\Delta x \to 0} \dfrac{\Delta y}{\Delta x}$	$\dfrac{dy}{dx} = 2x$

The steps are as follows: select (or "fix") an x in the domain of f, write $y = f(x)$, introduce Δx, write $y + \Delta y = f(x + \Delta x)$, subtract to get Δy, divide by Δx to get $\Delta y/\Delta x$, and, finally, find the limit as $\Delta x \to 0$ to obtain dy/dx. Whether we consciously use the four-step rule or not, we all need experience in the art of calculating derivatives by finding limits of difference quotients, and problems at the end of this section provide some of it. Meanwhile, we gain experience by proving the following formulas which can be and must be remembered.

Theorem 3.56 *If u and v are differentiable functions of x and if c and n are constants, then*

$$\frac{d}{dx}(u + v) = \frac{du}{dx} + \frac{dv}{dx} \tag{3.561}$$

$$\frac{d}{dx} cu = c\frac{du}{dx} \tag{3.562}$$

$$\frac{d}{dx} x^n = nx^{n-1} \tag{3.563}$$

$$\frac{d}{dx} u^n = nu^{n-1}\frac{du}{dx} \tag{3.564}$$

$$\frac{d}{dx} uv = u\frac{dv}{dx} + v\frac{du}{dx} \tag{3.565}$$

$$\frac{d}{dx}\frac{u}{v} = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2} \tag{3.566}$$

provided that $v \neq 0$ in (3.566) *and that in* (3.563) *and* (3.564) *we have $x \neq 0$ and $u \neq 0$ when n is a negative integer and* (except in some special cases) *$x > 0$ and $u > 0$ when n is not an integer.*

The first three of these formulas enable us to obtain results like

$$\frac{d}{dx}(x^4 - 3x^3 + 5x^2 - 7x + 6) = 4x^3 - 9x^2 + 10x - 7$$

as rapidly as we can write. Thus scientists differentiate polynomials with gusto. Using (3.563) with $n = -\frac{1}{2}$ gives

$$\frac{d}{dx}\frac{1}{\sqrt{x}} = \frac{d}{dx}x^{-1/2} = -\frac{1}{2}x^{-3/2}$$

when $x > 0$, and using it with $n = \frac{3}{2}$ gives

$$\frac{d}{dx}x\sqrt{x} = \frac{d}{dx}x^{3/2} = \frac{3}{2}x^{1/2}$$

when $x > 0$.

The last formula (3.566) can be remembered for years with the aid of a little trick. We remember that the derivative of a quotient is a bigger and better one and begin by drawing a long line to separate the numerator from the denominator. We continue by putting v^2 in the denominator and then, while the v is in mind, begin the numerator by writing v. This starts things right, and the rest can be remembered.

In our proof of the theorem, we fix (or select) an x in the domain of the functions and put $u = u(x)$, $v = v(x)$, $u + \Delta u = u(x + \Delta x)$,

$$v + \Delta v = v(x + \Delta x)$$

so that

$$\frac{du}{dx} = \lim_{\Delta x \to 0}\frac{\Delta u}{\Delta x} = \lim_{\Delta x \to 0}\frac{u(x + \Delta x) - u(x)}{\Delta x}$$
$$\frac{dv}{dx} = \lim_{\Delta x \to 0}\frac{\Delta v}{\Delta x} = \lim_{\Delta x \to 0}\frac{v(x + \Delta x) - v(x)}{\Delta x}.$$

We prove (3.561) and (3.562) together by starting with

$$y = cu + c_1 v,$$

where c and c_1 are constants; we can put $c = c_1 = 1$ to get (3.561) and we can take $c_1 = 0$ to get (3.562). Then

$$y + \Delta y = c(u + \Delta u) + c_1(v + \Delta v)$$

and subtraction gives

$$\Delta y = c\,\Delta u + c_1\,\Delta v.$$

Hence

$$\frac{\Delta y}{\Delta x} = c\frac{\Delta u}{\Delta x} + c_1\frac{\Delta v}{\Delta x}.$$

The hypothesis of Theorem 3.56 implies

$$\lim_{\Delta x \to 0}\frac{\Delta u}{\Delta x} = \frac{du}{dx}, \qquad \lim_{\Delta x \to 0}\frac{\Delta v}{\Delta x} = \frac{dv}{dx}.$$

Therefore, an application of a theorem on limits gives

$$\frac{dy}{dx} = c\frac{du}{dx} + c_1\frac{dv}{dx}.$$

As has been remarked, putting $c = c_1 = 1$ gives (3.561) and putting $c_1 = 0$ gives (3.562).

Postponing (3.563) and (3.564), we start proving the *product formula* (3.565) by setting $y = uv$. Then

$$\begin{aligned} y + \Delta y &= (u + \Delta u)(v + \Delta v) \\ &= uv + u\,\Delta v + v\,\Delta u + \Delta u\,\Delta v, \end{aligned}$$

so $\Delta y = u\,\Delta v + v\,\Delta u + \Delta u\,\Delta v$. Dividing by Δx and inserting an extra factor Δx in the numerator and denominator of the last term give

$$(3.57) \qquad \frac{\Delta y}{\Delta x} = u\frac{\Delta v}{\Delta x} + v\frac{\Delta u}{\Delta x} + \frac{\Delta u}{\Delta x}\frac{\Delta v}{\Delta x}\Delta x.$$

Taking limits as Δx approaches zero gives

$$\frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx} + \frac{du}{dx}\frac{dv}{dx}\cdot 0.$$

The last term is zero, and this proves (3.565). Proof of the *quotient formula* (3.566) is very similar, but the formula is important and we shall prove it. Let $y = u/v$. Then

$$y + \Delta y = \frac{u + \Delta u}{v + \Delta v}$$

$$\Delta y = \frac{u + \Delta u}{v + \Delta v} - \frac{u}{v} = \frac{v\,\Delta u - u\,\Delta v}{v^2 + v\,\Delta v}$$

$$(3.571) \qquad \frac{\Delta y}{\Delta x} = \frac{v\dfrac{\Delta u}{\Delta x} - u\dfrac{\Delta v}{\Delta x}}{v^2 + v\dfrac{\Delta v}{\Delta x}\Delta x}.$$

Taking limits as Δx approaches zero gives

$$\frac{dy}{dx} = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2},$$

and this proves (3.566).

The *power formulas* (3.563) and (3.564) remain to be proved, and we deal with (3.563) first. Let $y = x^n$. In case $n = 0$, we have $y = 1$ and must prove that $dy/dx = 0$. This is true because if $y = 1$ for each x, then $\Delta y = 0$, so $\Delta y/\Delta x = 0$ and hence $dy/dx = 0$. In case $n = 1$, we have $y = x$ and must prove that

$$\frac{dx}{dx} = 1.$$

This is true because if $y = x$, then $\Delta y = \Delta x$, so $\Delta y/\Delta x = 1$ and $dy/dx = 1$. The case $n = 2$ is covered in the application under the four-step rule headline. There are several somewhat different ways to obtain (3.563) for greater integer values of n. Perhaps the most informative method consists of using the product formula (3.565) to obtain

$$\begin{aligned}\frac{d\,u_1u_2u_3}{dx} &= \frac{d(u_1u_2)u_3}{dx} = \frac{d\,u_1u_2}{dx}\,u_3 + u_1u_2\frac{d\,u_3}{dx}\\ &= \frac{du_1}{dx}\,u_2u_3 + u_1\frac{du_2}{dx}\,u_3 + u_1u_2\frac{du_3}{dx}\end{aligned}$$

and then putting $u_1 = u_2 = u_3 = x$ to obtain

$$\frac{dx^n}{dx} = nx^{n-1} \tag{3.572}$$

when $n = 3$. Another application of the same idea, in which the product $u_1u_2u_3u_4$ is written as the product $(u_1u_2u_3)u_4$ of two factors, gives

$$\frac{du_1u_2u_3u_4}{dx} = \frac{du_1}{dx}\,u_2u_3u_4 + u_1\frac{du_2}{dx}\,u_3u_4 + u_1u_2\frac{du_3}{dx}\,u_4 + u_1u_2u_3\frac{du_4}{dx},$$

and putting $u_1 = u_2 = u_3 = u_4 = x$ gives the formula (3.572) for the case in which $n = 4$. The same procedure gives the result for greater integers. Perhaps the simplest proof can be based upon the fact that if (3.572) holds for a given n, then use of the product rule gives

$$\frac{dx^{n+1}}{dx} = \frac{d}{dx}\,x\cdot x^n = xnx^{n-1} + x^n\cdot 1 = (n+1)x^n.$$

Since the formula is valid when $n = 0$, mathematical induction shows that it is valid when n is a nonnegative integer. In case n is a negative integer (so that $-n$ is a positive integer) and $x \neq 0$, the result is proved by the calculation

$$\frac{dx^n}{dx} = \frac{d}{dx}\frac{1}{x^{-n}} = \frac{x^{-n}\cdot 0 - 1\cdot(-n)x^{-n-1}}{x^{-2n}} = nx^{n-1}$$

which involves the formula for the derivative of a quotient. In case n is a constant which is not an integer, (3.563) is still valid at least when $x > 0$. Proof of this appears in Theorem 9.27, and proof of (3.564) then follows from the chain rule of Theorem 3.65.

Very much more about derivatives remains to be learned, and we give a modest but important contribution to theory by proving the following theorem.

Theorem 3.58 *If $f'(x)$ exists, then f must be continuous at x.*

Our hypothesis and a theorem on limits enable us to write

$$\lim_{\Delta x\to 0}[f(x+\Delta x)-f(x)] = \lim_{\Delta x\to 0}\left[\frac{f(x+\Delta x)-f(x)}{\Delta x}\Delta x\right]$$
$$= \left[\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}\right]\left[\lim_{\Delta x\to 0}\Delta x\right] = f'(x)\cdot 0 = 0.$$

Therefore,

$$\lim_{\Delta x\to 0} f(x+\Delta x) = f(x),$$

and it follows from this that f is continuous at x.

The attainment of a technique for differentiating accurately and efficiently is of prime importance in the calculus. When we are called upon to evaluate the left member of the equation

$$\frac{d}{dx}\frac{x}{1+x^2} = \frac{(1+x^2)(1)-x(2x)}{(1+x^2)^2} = \frac{1-x^2}{(1+x^2)^2},$$

we should say to ourselves "the derivative with respect to x of u (meaning x) over v (meaning $1+x^2$) is equal to the quotient with denominator v^2 [write $(1+x^2)^2$] and numerator v [write $(1+x^2)$] times du/dx [write 1] minus u [write x] times dv/dx [write $2x$]." We must learn to talk to ourselves in such a way that we can quickly produce such results as

$$\frac{d}{dx}\frac{1+x^2}{x} = \frac{x(2x)-(1+x^2)}{x^2} = \frac{x^2-1}{x^2}$$
$$\frac{d}{dx}\frac{x}{1-x^2} = \frac{(1-x^2)(1)-x(-2x)}{(1-x^2)^2} = \frac{1+x^2}{(1-x^2)^2}$$
$$\frac{d}{dx}\frac{1}{1+x^2} = \frac{0-1(2x)}{(1+x^2)^2} = -\frac{2x}{(1+x^2)^2}.$$

With the formula for the derivative of a product in mind, we obtain

$$\frac{d}{dx}(x^2-x+1)(x^2+2x+1) = (x^2-x+1)(2x+2)$$
$$+ (x^2+2x+1)(2x-1)$$

by saying "the derivative with respect to x of u (meaning x^2-x+1) times v (meaning x^2+2x+1) is equal to u (write x^2-x+1) times dv/dx (write $2x+2$) plus v (write x^2+2x+1) times du/dx (write $2x-1$)."

Problems 3.59

1 Give the definition of the derivative of f at x. *Ans.:* (3.53) or (3.54).

2 Find dy/dx from the definition of derivatives and then check the answers by use of formulas for differentiation when

(*a*) $y = \sqrt{x}$ (*b*) $y = \dfrac{1}{\sqrt{x}}$

(*c*) $y = \dfrac{1}{x+1}$ (*d*) $y = \dfrac{x}{1+x}$

(*e*) $y = \dfrac{x}{4-x}$ (*f*) $y = \dfrac{1}{1-2x}$

(*g*) $y = \dfrac{1}{1+x^2}$ (*h*) $y = \dfrac{1}{1-x^2}$

3 Sometimes we are given a formula for $f(x)$ and are required to find the derivative of f at a, where a is a number given in decimal form. In some cases it is easiest to find $f'(a)$ directly from the formula

$$f'(a) = \lim_{\Delta x\to 0} \frac{f(a+\Delta x) - f(a)}{\Delta x} \qquad \text{or} \qquad f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}.$$

In some cases it is easiest to work out a formula for $f'(x)$ and put $x = a$ in the result. Work the problem both ways when

(*a*) $f(x) = x^2 - 3x + 1$, $a = 416$
(*b*) $f(x) = x^2 - 3x + 1$, $a = 0$
(*c*) $f(x) = \dfrac{x}{1+x}$, $a = 2$
(*d*) $f(x) = \dfrac{x}{1-x^2}$, $a = 0$

√**4** Formulas for derivatives are often wonderful, but there are times when it is best to use the definition of derivatives to obtain $f'(a)$. Letting $g(x) = x|x|$, find $g'(0)$ or show that $g'(0)$ does not exist. *Ans.:*

$$g'(0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h} = \lim_{h\to 0} \frac{h|h| - 0}{h} = 0.$$

5 Supposing that

$$y = 1 + x + x^2 + x^3 + x^4$$
$$z = 5 + 4x - 3x^2 + 5x^3 + x^4,$$

tell what facts or formulas or both enable us to write

$$\frac{dy}{dx} = 1 + 2x + 3x^2 + 4x^3$$
$$\frac{dz}{dx} = 4 - 6x + 15x^2 + 4x^3.$$

6 Calculate

$$\frac{d}{dx}(x^2 + 3)(x^2 - 2)$$

by use of the product formula. Then multiply the given factors and differentiate the result. Make the answers agree. *Hint:* Look at $(x^2 + 3)(x^2 - 2)$ and read u (meaning $x^2 + 3$) times v (meaning $x^2 - 2$). Then apply the formula for the derivative of uv.

7 This is another lesson on use of formulas. It is expected that persons studying calculus are familiar with the "quadratic formula." When we want to find the values of x for which

$$2x^2 + 3x - 4 = 0,$$

we say "$ax^2 + bx + c = 0$" and, without writing anything, realize that we put $a = 2$ and $b = 3$ and $c = -4$ in the memorized formula

$$x = \frac{-b \pm \sqrt{b^2 - 4\,ac}}{2a}.$$

Then we write only

$$x = \frac{-3 \pm \sqrt{9 + 32}}{4}.$$

When we use differentiation formulas, we should be equally efficient. When we must differentiate

$$y = (3x^2 + 1)^5, \tag{1}$$

we should realize that we must differentiate something of the form u^n (not x^n), where u is a function of x. The formula

$$\frac{d}{dx}u^n = nu^{n-1}\frac{du}{dx} \tag{2}$$

should come into our minds but should not be written. We should look at (1) and read "y equals u to the nth power" and realize without writing anything that $u = 3x^2 + 1$ and $n = 5$. We should then say "dy/dx equals n (write 5) u (write $3x^2 + 1$) to the power $n - 1$ (write 4) times du/dx (write $6x$)." Thus we look at (1) and, after a little chat with ourselves, write

$$\begin{aligned} \frac{dy}{dx} &= 5(3x^2 + 1)^4 \cdot 6x \\ &= 30x(3x^2 + 1)^4. \end{aligned} \tag{3}$$

Minor modifications of this technique can be tolerated, but speed and accuracy must be developed. Write the formula (1) and practice differentiating it as a golfer practices putting; perfection is required.

8 Look at the calculations

$$y = (1 - x^2)^{1/2}, \qquad z = (1 + x^2)^{-1}$$

$$\frac{dy}{dx} = \frac{1}{2}(1 - x^2)^{-1/2}(-2x), \qquad \frac{dz}{dx} = -(1 + x^2)^{-2}(2x)$$

until you see where they come from and understand them thoroughly. Nothing is to be written.

9 Calculate

$$\frac{d}{dx}\frac{1-x^2}{1+x^2}, \qquad \frac{d}{dx}(1-x^2)(1+x^2)^{-1}$$

by the quotient formula and by the product formula. Make the results agree.

10 It can be observed that the sum of the first two of the expressions

$$\frac{x^2}{\sqrt{x^2+1}}, \qquad \frac{1}{\sqrt{x^2+1}}, \qquad \sqrt{1+x^2}$$

is the third. Find the derivatives of these things and check the results by showing that the sum of the first two is the third.

11 The formulas

$$\frac{d}{dx}\sin x = \cos x, \qquad \frac{d}{dx}\cos x = -\sin x$$

will be proved in Section 8.1. Copy them on a nice clean piece of paper, and take a casual look at the calculations

$$\frac{d}{dx}\tan x = \frac{d}{dx}\frac{\sin x}{\cos x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x$$
$$\frac{d}{dx}\cot x = \frac{d}{dx}\frac{\cos x}{\sin x} = \frac{-\sin^2 x - \cos^2 x}{\sin^2 x} = \frac{-1}{\sin^2 x} = -\csc^2 x$$
$$\frac{d}{dx}\sec x = \frac{d}{dx}(\cos x)^{-1} = -(\cos x)^{-2}(-\sin x) = \frac{1}{\cos x}\frac{\sin x}{\cos x} = \sec x \tan x$$
$$\frac{d}{dx}\csc x = \frac{d}{dx}(\sin x)^{-1} = -(\sin x)^{-2}(\cos x) = \frac{-1}{\sin x}\frac{\cos x}{\sin x} = -\csc x \cot x.$$

Then, with the calculations out of sight, try to reproduce them.

12 Show that

$$\frac{d}{dx}\frac{ax+b}{cx+d} = \frac{ad-bc}{(cx+d)^2}.$$

13 Supposing that n is a positive integer and $x \neq 1$, show how the identity

$$1 + x + x^2 + x^3 + \cdots + x^n = \frac{x^{n+1}-1}{x-1} \tag{1}$$

can be used to obtain the less elementary identity

$$1 + 2x + 3x^2 + \cdots + nx^{n-1} = \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2}. \tag{2}$$

Multiply by x and differentiate again to obtain another identity.

14 Calculate the coordinates of the points on the graph of $y = f(x)$ at which $f'(x) = 0$ when

(*a*) $f(x) = x^3 - 3x$ — *Ans.:* $(-1,2)$ and $(1,-2)$
(*b*) $f(x) = x^3 - 3x + 2$ — *Ans.:* $(-1,4)$ and $(1,0)$
(*c*) $f(x) = 2x + 3$ — *Ans.:* None
(*d*) $f(x) = \dfrac{x^2}{1+x^2}$ — *Ans.:* $(0,0)$
(*e*) $f(x) = \dfrac{x}{1+x^2}$ — *Ans.:* $(-1,-\frac{1}{2})$ and $(1,\frac{1}{2})$
(*f*) $f(x) = ax^2 + bx + c$ — *Ans.:* $(-b/2a, -(b^2 - 4ac)/4a)$

15 A long time ago it was discovered that if P is a polynomial, then between each pair of values of x for which $P(x) = 0$ there must be at least one value of x for which $P'(x) = 0$. For the special case in which

$$P(x) = (x - 1)(x - 2)(x - 3),$$

find the values of x for which $P(x) = 0$ and the values of x for which $P'(x) = 0$, and verify the statement about the zeros of P and P'. *Remark:* This matter will become quite unmysterious when we learn about the Rolle theorem.

16 In connection with the definition of the derivative of a function f at a point x, we recognized the possibility that this derivative may fail to exist. To clarify this matter, we should know about the simplest example of a continuous function f which is not everywhere differentiable. To investigate such matters, we should know about the *right-hand derivative* $f'_+(x)$ and the *left-hand derivative* $f'_-(x)$ that are defined by the formulas

$$f'_+(x) = \lim_{\Delta x \to 0+} \frac{f(x + \Delta x) - f(x)}{\Delta x}, \qquad f'_-(x) = \lim_{\Delta x \to 0-} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

when these limits exist. It is easy to guess and almost as easy to prove that f is differentiable at x if and only if

$$f'_+(x) = f'_-(x) = f'(x).$$

For the simplest example in which $f(x) = |x|$, show that $f'_+(0) = 1$ and $f'_-(0) = -1$ and hence that $f'(0)$ does not exist. It is not so easy to construct a continuous function which is everywhere nondifferentiable, but Weierstrass (1815–1897) started this construction business a long time ago.

17 Construct and look at a graph similar to those in Figures 3.46 and 3.47 but having the loops tangent to the lines having equations $y = x$ and $y = -x$. Letting this be the graph of f and, letting $f(0) = 0$, discuss continuity of f at 0 and discuss $f'_+(0)$, $f'_-(0)$, and $f'(0)$.

18 For reasons that may be partially explained by the remark at the end of this set of problems, we give, in terms of the notation of Newton, a complete statement and proof of the part of Theorem 3.56 that involves the product formula (3.565).

(1) **Theorem** *If g and h are functions differentiable at x and if f is the function for which*

$$f(x) = g(x)h(x) \tag{2}$$

when x belongs to the domains of g and h, then f is differentiable at x and

$$f'(x) = g(x)h'(x) + h(x)g'(x). \tag{3}$$

Proof: Since g and h are differentiable at x, there must be an interval I with center at x over which g and h are defined. When $x + \Delta x$ lies in this interval, we have

$$f(x + \Delta x) = g(x + \Delta x)h(x + \Delta x) \tag{4}$$

and hence

(5) $$f(x + \Delta x) - f(x) = g(x + \Delta x)h(x + \Delta x) - g(x)h(x).$$

To make the right side more tractable, we subtract and add the term $g(x + \Delta x)h(x)$ and then divide by Δx to obtain

(6) $$\frac{f(x + \Delta x) - f(x)}{\Delta x} = g(x + \Delta x)\frac{h(x + \Delta x) - h(x)}{\Delta x} + h(x)\frac{g(x + \Delta x) - g(x)}{\Delta x}.$$

But the hypotheses of our theorem, the definition of derivative, and Theorem 3.58 imply that

(7) $$\lim_{\Delta x \to 0} g(x + \Delta x) = g(x), \qquad \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x} = h'(x),$$
$$\lim_{\Delta x \to 0} \frac{g(x + \Delta x) - g(x)}{\Delta x} = g'(x).$$

It follows that the limit, as Δx approaches zero, of the right member of (6) is the right member of the formula

(8) $$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = g(x)h'(x) + h(x)g'(x).$$

The limit of the left member must be the same. Therefore, (8) holds, and (3) then follows from the definition of the derivative of f at x. This proves the theorem. This proof is essentially the same as the proof involving (3.57). If we set $u = g(x)$, $v = h(x)$, $y = f(x)$, $u + \Delta u = g(x + \Delta x)$, $v + \Delta v = h(x + \Delta x)$, and $y + \Delta y = f(x + \Delta x)$, then (6) becomes

(9) $$\frac{\Delta y}{\Delta x} = (u + \Delta u)\frac{\Delta v}{\Delta x} + v\frac{\Delta u}{\Delta x}$$

which is, except for a minor shuffling of terms, the same as (3.57). The version involving (3.57) is usually preferred in elementary courses because the formulas involving Δu, Δv, and Δy flow more smoothly and quickly than those given above.

19 *Remark:* As was said in passing in the text, discussions of names and symbols can be long and perhaps dismal. We call a rose "a rose" because everyone else does, and we do not need another reason. We call the number

(1) $$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

when it exists, "the derivative of f at x" because everyone else does, and we do not need another reason. We can denote this number by $f'(x)$ because everyone else does, and we do not need another reason. If we want to know what $f'(x)$ means, we do not look at $f'(x)$; we look at the definition of $f'(x)$ and see that

(2) $$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

We can observe that the Newton notation in the formula (2) uses functional notation in a thoroughly standard way; if f is a function and x is a number, then the right side of (2), when it exists, determines the value of the function f' at x. If, for example, $f'(x) = 2x$ for each real x, then $f'(0) = 0$, $f'(6) = 12$, $f'(x^2) = 2x^2$, and $f'(\sin x) = 2 \sin x$. It could be presumed that one good symbol for the number in (1) should be enough, but it is not enough. Even if there were no other reason, we would still be required to know another symbol in order to be able to read scientific literature and to converse with scientists. We must know that we can set $y = f(x)$, so that in a particular case we have $y = x^2$, and we can denote the derivative of f at x by the symbol $\dfrac{dy}{dx}$. If we want to know what $\dfrac{dy}{dx}$ means, we do not look at $\dfrac{dy}{dx}$; we look at the definition of $\dfrac{dy}{dx}$ and find that, in the particular case,

$$\frac{dy}{dx} = 2x. \tag{3}$$

According to the definition, $\dfrac{dy}{dx}$ is the derivative of f at x, and the meaning of $\dfrac{dy}{dx}$ is not changed when we read "dee y dee x" or "the derivative of y with respect to x" or even "the derivative of y with respect to x at x." The assertion (3) always means that the derivative of f at x is $2x$, and weird ways of reading the assertion do not change the meaning of the assertion. The meaning of the assertion is not changed when we realize that a silly result is obtained by supposing that the d's and the x and the y in (3) are numbers and canceling the d's to get $y/x = 2x$. The meaning of the assertion is still unchanged when we realize that we never put $x = 6$ in the two members of (3) to obtain

$$\frac{dy}{d6} = 12. \tag{4}$$

We do, however, allow ourselves the liberty of writing

$$\frac{dx^2}{dx} = 2x \qquad \text{or} \qquad \frac{d}{dx}x^2 = 2x \tag{5}$$

to abbreviate the statement that if $y = f(x)$, where f is the function for which $f(x) = x^2$, then the derivative of f at x is $2x$. From a logical point of view, everything we have done can be summarized very simply. If we want to know the meaning of the word "quibble," we do not look at the word "quibble"; we look at a definition. Let us then quit quibbling about the meaning of $\dfrac{dy}{dx}$. We can conclude with a cheerful remark. Whenever we are likely to encounter difficulties with the Leibniz notation, we can discard it and use the Newton notation.

3.6 *The chain rule and differentiation of elementary functions*

To be able to illustrate methods by which fundamental formulas for derivatives are used, we suppose that we know the five fundamental formulas

$$(3.61)\quad \frac{d}{dx}x^n = nx^{n-1}, \qquad \frac{d}{dx}\sin x = \cos x, \qquad \frac{d}{dx}\cos x = -\sin x$$

$$(3.62)\quad \frac{d}{dx}e^x = e^x, \qquad \frac{d}{dx}\log x = \frac{1}{x}$$

only the first of which has been partially proved.† In the last two of these formulas, the base is e, the base of natural exponentials and logarithms, which appears in (3.272) and which will appear later. One of our tasks is to learn a procedure by which we can obtain a correct formula for dy/dx when $y = \sin u$ and u is a differentiable function of x which is not necessarily x itself. The answer is

$$(3.621)\quad \frac{dy}{dx} = \cos u \frac{du}{dx}.$$

To see why this is so, and to see how many similar formulas can be obtained, we consider the general situation in which y is a function of u and u is a function of x, say $y = f(u)$ and $u = g(x)$. Then y is linked to x through the links of a short chain; x determines u and u determines y, so y is a function of x. While the operation may seem somewhat ponderous when

$$(3.622)\quad y = \phi(x) = f(g(x)) = \sin g(x) = \sin u = \sin 2x,$$

we can let $\phi(x) = f(g(x))$ and sketch the schematic Figure 3.63 which catches the functions g, f, and ϕ, respectively, in the act of transforming (or mapping or carrying) x into u, u into y, and x into y. The function ϕ for which $\phi(x) = f(g(x))$ is sometimes called a *composite function*.

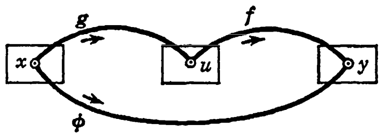


Figure 3.63

The following theorem is the *chain rule*, which sets forth conditions under which y has a derivative with respect to x that can be calculated from the *chain formula*

$$(3.64)\quad \frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

The result is given in terms of the "d" notation of Leibniz and the "prime" notation of Newton, so that we can, in applications, choose the one that seems to be most convenient or informative.

† These formulas will be proved in Chapters 8 and 9. A contention that we can and should learn and use these formulas before they are proved is pedagogically sound. It is as practical as the contention that embryonic electrical engineers should learn that copper wires conduct electric current, and use this information in various ways, before they study solid-state physics and learn mechanisms by which electrons travel along conductors and semiconductors.

Theorem 3.65 (**chain rule**) *If f and g are functions such that g is differentiable at x and f is differentiable at $g(x)$ and if we set $y = f(u)$ and $u = g(x)$ so that $y = f(g(x))$, then the chain formula*

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = f'(u)g'(x) = f'(g(x))g'(x) \tag{3.66}$$

is valid at x.

To prove this theorem, we use the notation of the theorem to obtain

$$\Delta u = g(x + \Delta x) - g(x), \qquad \Delta y = f(u + \Delta u) - f(u) \tag{3.661}$$

and observe that u and y are determined by x alone, while Δu and Δy are determined by x and Δx. Consider first the usual case in which there is a number δ_1 such that $\delta_1 > 0$ and $\Delta u \neq 0$ whenever $0 < |\Delta x| < \delta_1$. Then, when $0 < |\Delta x| < \delta_1$, we can write

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u}\frac{\Delta u}{\Delta x} = \frac{f(u + \Delta u) - f(u)}{\Delta u}\frac{g(x + \Delta x) - g(x)}{\Delta x} \tag{3.662}$$

and, after observing that

$$\lim_{\Delta x \to 0} \Delta u = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x}\Delta x = \frac{du}{dx}\cdot 0 = 0, \tag{3.663}$$

take limits as Δx approaches zero to obtain the required result. Because division by zero is taboo, exceptional cases are more troublesome. We can avoid this difficulty and handle all cases at once by setting

$$\phi(\Delta u) = \frac{\Delta y}{\Delta u} = \frac{f(u + \Delta u) - f(u)}{\Delta u} \qquad (\Delta u \neq 0) \tag{3.664}$$

$$\phi(\Delta u) = \frac{dy}{du} = f'(u) \qquad (\Delta u = 0). \tag{3.665}$$

Then, whether Δu is zero or not, we can write

$$\frac{\Delta y}{\Delta x} = \phi(\Delta u)\frac{\Delta u}{\Delta x} = \phi(\Delta u)\frac{g(x + \Delta x) - g(x)}{\Delta x} \tag{3.666}$$

and take limits as Δx approaches zero to obtain the required result.

The basic elementary functions can be separated into three classes. The first class contains powers and roots of x, that is, functions of the form x^a, where a is a constant. The second class contains the six trigonometric functions and the six inverse trigonometric functions. The third class contains exponential functions of the form b^x and logarithmic functions of the form $\log_b x$, the base b being a constant. Thus there are just 15 types of basic elementary functions. The class of *elementary functions* includes the frightful function ϕ having values

$$\phi(x) = \frac{\log(1 + x^2) + [e^x + (x^4 - 7x^2 + \sin^{-1} 3x^2)^4]^{1/2}}{\sin e^{2x} + e^{\sin 3x} + x\sin^2 4x - \cos x^5} \tag{3.667}$$

and all others obtainable by making "finite combinations" of basic elementary functions together with addition, subtraction, multiplication, and division. This class contains very many important functions. It is therefore important to know that we can work out a formula for the derivative of any given elementary function when we know (i) Theorem 3.56, (ii) 15 basic formulas for derivatives of basic elementary functions, (iii) the chain rule, and, in addition, we possess (iv) a technique which enables us to apply these things.

Of the 15 basic formulas, the most important 5 were listed at the beginning of this section and are relisted in the first column of the following little table.

$$\frac{d}{dx}x^n = nx^{n-1} \qquad \frac{d}{dx}u^n = nu^{n-1}\frac{du}{dx} \tag{3.671}$$

$$\frac{d}{dx}\sin x = \cos x \qquad \frac{d}{dx}\sin u = \cos u\frac{du}{dx} \tag{3.672}$$

$$\frac{d}{dx}\cos x = -\sin x \qquad \frac{d}{dx}\cos u = -\sin u\frac{du}{dx} \tag{3.673}$$

$$\frac{d}{dx}e^x = e^x \qquad \frac{d}{dx}e^u = e^u\frac{du}{dx} \tag{3.674}$$

$$\frac{d}{dx}\log x = \frac{1}{x} \qquad \frac{d}{dx}\log u = \frac{1}{u}\frac{du}{dx} \tag{3.675}$$

If we know the first formula on the left, we can set $y = u^n$ and use the chain formula (3.66) to obtain the chain formula

$$\frac{d}{dx}u^n = \frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = nu^{n-1}\frac{du}{dx}$$

written opposite it. If we know the second formula on the left, we can set $y = \sin u$ and use (3.66) again to obtain the chain formula

$$\frac{d}{dx}\sin u = \frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = (\cos u)\frac{du}{dx}$$

written opposite it. The same procedure shows that each basic formula has a chain extension. Of the ten basic formulas not listed above, four (which appear in Problem 11 of Section 3.5 and have probably been forgotten) give derivatives of the last four trigonometric functions, and the remaining six give derivatives of the inverse trigonometric functions. Proofs of all of the formulas will appear later. Except for three formulas that are rarely used, the formulas are listed on the page opposite the back cover of this book.

Our fund of information about logarithms is quite meager, but we can slowly add to it. We begin with the idea that $\log x$ exists (as a real number) only when $x > 0$. In case $x < 0$, $\log x$ does not exist but $|x| > 0$ and $\log |x|$ does exist. When $x < 0$, we can use the chain formula to obtain

$$\frac{d}{dx}\log|x| = \frac{d}{dx}\log(-x) = \frac{1}{-x}\frac{d(-x)}{dx} = \frac{1}{x}.$$

Thus we can extend the two formulas in (3.675) to obtain the more general formulas

$$\text{(3.676)} \qquad \frac{d}{dx}\log|x| = \frac{1}{x}, \qquad \frac{d}{dx}\log|u| = \frac{1}{u}\frac{du}{dx}$$

in which it is required that $x \neq 0$ and $u \neq 0$ but it is not required that x and u be positive.

Up to the present time, our work with difference quotients and derivatives has involved only fundamental definitions and formulas. Figures and geometric ideas, which might be helpful but which might also be misleading, have been completely absent. Section 5.1 will present our thorough introduction to matters relating to slopes of graphs and tangents to graphs. Meanwhile, we may be helped and may be unmisled by looking at Figure 3.68, which shows the graph C of a differentiable function f. The points P and Q having coordinates (x,y) or $(x, f(x))$ and $(x + \Delta x,\ y + \Delta y)$ or $(x + \Delta x,\ f(x + \Delta x))$ are shown, but the line PQ joining P and Q is not drawn. The first of the two formulas

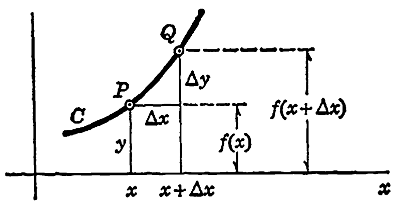


Figure 3.68

$$\text{(3.681)} \qquad \frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x} = \text{slope of line } PQ$$

$$\text{(3.682)} \qquad \frac{dy}{dx} = f'(x) = \text{slope of tangent to } C \text{ at } P$$

$$= \text{slope of } C \text{ at } P$$

is correct because it is obtained by applying the definition of the slope of a line. The second formula is correct by definition; the line through P whose slope is the limit as Δx approaches zero of the slopes in (3.681) is, by definition, the line tangent to C at P, and, moreover, the slope of C at P is, by definition, the slope of the line tangent to C at P. The definition gives precision to the venerable idea that the slope of the line PQ is close to the slope of the tangent at P whenever Q is close to P. The definition (3.682) turns out to be very important. Indeed, there are many situations in which magnitudes play minor roles and it is important to know that the graph C of $y = f(x)$ has a horizontal tangent (tangent of zero slope) at each point (x,y) on C for which $f'(x) = 0$, has a tangent of positive slope at each point (x,y) on C for which $f'(x) > 0$, and has a tangent of negative slope at each point (x,y) on C for which $f'(x) < 0$.

Problems 3.69

1 Calculate $f'(x_0)$ and write the equation of the line tangent to the graph of $y = f(x)$ at the point (x_0,y_0) when

(a) $f(x) = x^2, \quad x_0 = 1$ *Ans.:* $y - 1 = 2(x - 1)$
(b) $f(x) = x(1 - x), \quad x_0 = 0$ *Ans.:* $y = x$
(c) $f(x) = e^x, \quad x_0 = 0$ *Ans.:* $y = x + 1$
(d) $f(x) = \dfrac{x}{1 + x^2}, \quad x_0 = 1$ *Ans.:* $y = \frac{1}{2}$

2 Become thoroughly familiar with the following technique, because it enables us to do many chores quickly and correctly. Suppose we are required to find dy/dx when

$$y = \sin 2x. \tag{1}$$

We must realize that we are not required to differentiate $\sin x$ but are required to differentiate $\sin u$, where u is a function of x. We look at (1) and read "y equals sine u" and realize without making a lot of noise and without writing anything that $u = 2x$. We then write dy/dx and say this is equal to $\cos u$ (write $\cos 2x$) times du/dx (write 2). When we follow orders, we get

$$\frac{dy}{dx} = (\cos 2x)2, \tag{2}$$

but it is always better to put the answer in the neater form

$$\frac{dy}{dx} = 2 \cos 2x \tag{3}$$

which does not require parentheses.

3 Read the equations

(a) $y = (x^2 + 1)^n$ (b) $y = \cos e^x$
(c) $y = \sin (ax + b)$ (d) $y = e^{ax}$
(e) $y = \cos ax$ (f) $y = \log (x^2 + 1)$

the way we read them when we want to find dy/dx. In the first case, we can tolerate "y equals u to the nth" as a contraction of "y equals u with the exponent n" or "y equals u to the nth power." In another case, we can tolerate "y equals e to the u," which looks bad in print but is universally understood. Now, supposing that n, a, and b are constants, concentrate upon the task of learning five basic formulas and applying them to obtain the answers

(a) $\dfrac{dy}{dx} = 2nx(x^2 + 1)^{n-1}$ (b) $\dfrac{dy}{dx} = -e^x \sin e^x$

(c) $\dfrac{dy}{dx} = a \cos (ax + b)$ (d) $\dfrac{dy}{dx} = ae^{ax}$

(e) $\dfrac{dy}{dx} = -a \sin ax$ (f) $\dfrac{dy}{dx} = \dfrac{2x}{x^2 + 1}$

Practice the technique until the answers can be obtained quickly and effortlessly.

4 Each of the formulas

$$y = x^2 \sin x, \qquad y = xe^x$$

can be read "y equals u times v." Do this and obtain the derivatives

$$\frac{dy}{dx} = x^2 \cos x + 2x \sin x, \qquad \frac{dy}{dx} = xe^x + e^x.$$

5 Each of the formulas

$$y = \frac{\sin x}{x}, \qquad y = \frac{\log x}{x}$$

can be read "y equals u over v." Do this and obtain the formulas

$$\frac{dy}{dx} = \frac{x \cos x - \sin x}{x^2}, \qquad \frac{dy}{dx} = \frac{1 - \log x}{x^2}.$$

6 Derivatives of derivatives are called *derivatives of higher order;* the derivative of f at x is $f'(x)$, the derivative of f' at x is $f''(x)$, the derivative of f'' at x is $f'''(x)$ or $f^{(3)}(x)$, and so on. Supposing that z is a number and

$$f(x) = \frac{1}{z + x} = (z + x)^{-1},$$

show that

$$\begin{aligned} f'(x) &= -(z + x)^{-2}, & f''(x) &= 2!(z + x)^{-3}, & f^{(3)}(x) &= -3!(z + x)^{-4}, \\ f^{(4)}(x) &= 4!(z + x)^{-5}, & f^{(5)}(x) &= -5!(z + x)^{-6}, & f^{(6)}(x) &= 6!(z + x)^{-7}, \end{aligned}$$

where $2! = 1 \cdot 2$, $3! = 1 \cdot 2 \cdot 3$, $4! = 1 \cdot 2 \cdot 3 \cdot 4$, etcetera. Supposing as usual that $0! = 1$ and $1! = 1$, observe that

$$f^{(n)}(x) = (-1)^n n!(z + x)^{-n-1} \qquad (n = 0, 1, 2, \cdots)$$

when we agree that the result of differentiating f zero times is f itself.

7 Letting $f(x) = (1 - 2x)^{-1}$, show that

$$f^{(n)}(x) = 2^n n!(1 - 2x)^{-n-1} \qquad (n = 0, 1, 2, \cdots).$$

8 Letting $f(x) = \log(1 + x^2)$, show that

$$f'(x) = \frac{2x}{1 + x^2}, \qquad f''(x) = \frac{2 - 2x^2}{(1 + x^2)^2}$$

and calculate one more derivative.

9 The formulas

(*a*) $\sin(a + b)x = \sin ax \cos bx + \cos ax \sin bx$
(*b*) $\cos(a + b)x = \cos ax \cos bx - \sin ax \sin bx$
(*c*) $e^{(a+b)x} = e^{ax}e^{bx}$
(*d*) $\log ax = \log a + \log x$

are permanently remembered by all good scientists. For each formula, calculate the derivatives with respect to x of the two sides and show that the results are equal.

10 Supposing that a and ω (omega, to keep physicists and engineers happy) are constants and

$$Q = e^{at} \sin \omega t,$$

show how the formula

$$\begin{aligned}\frac{dQ}{dt} &= e^{at}(\omega \cos \omega t) + (\sin \omega t) a e^{at} \\ &= e^{at}(\omega \cos \omega t + a \sin \omega t)\end{aligned}$$

is obtained. Then let $I = dQ/dt$ and show that

$$\frac{dI}{dt} = e^{at}[2a\omega \cos \omega t + (a^2 - \omega^2) \sin \omega t].$$

Remark: It is not necessary for us to know that, if $a < 0$, Q might be the charge on the capacitor of an LRC oscillator, in which case the electric current would be I and the voltage drop across the inductor would be the product of dI/dt and the inductance L of the inductor. It is, however, a good idea to know that the things we are learning are important in applied mathematics.

11 Prove that

$$\frac{d}{dx} \frac{e^x - e^{-x}}{e^x + e^{-x}} = \left(\frac{2}{e^x + e^{-x}}\right)^2.$$

12 If, for a positive integer n,

$$y_n(x) = \frac{\sin x}{1} + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \cdots + \frac{\sin nx}{n},$$

show that

$$y_n'(x) = \cos x + \cos 2x + \cos 3x + \cdots + \cos nx.$$

13 Calculate $f'(x)$ from the first and then from the second of the formulas

$$f(x) = \log \left| \frac{1 - x}{1 + x} \right|, \qquad f(x) = \log |1 - x| - \log |1 + x|.$$

Make the results agree. *Hint:* Do not forget the second formula in (3.676); the derivative with respect to x of $\log |u|$ is $(1/u)\, du/dx$ and the absolute-value signs quietly disappear.

14 Observe that if y is a differentiable function of x, so also is the function F having values

$$F(x) = x^2 + xy(x) + [y(x)]^2.$$

Tell precisely what formulas are used to obtain the formula

$$F'(x) = 2x + xy'(x) + y(x) + 2y(x)y'(x)$$

or

$$\frac{dF}{dx} = 2x + x\frac{dy}{dx} + y + 2y\frac{dy}{dx}.$$

Ans.: The power formula, the formula for the derivative of a product, the chain formula, and the formula for the derivative of a sum.

15 Supposing that y is a differentiable function of x for which

$$x^2 + xy(x) + [y(x)]^2 = 3,$$

apply our fundamental formulas for calculating derivatives to obtain the formula

$$y'(x) = -\frac{2x + y(x)}{x + 2y(x)}.$$

Hint: Equate the derivatives with respect to x of the two members of the given equation. *Remark:* This process, by which we start with an equation involving $y(x)$ and [without obtaining an explicit formula for $y(x)$] obtain an explicit formula for $y'(x)$, is called "implicit differentiation." To gain understanding of this terminology, we can note that the formula $y = x + 1$ says *explicitly* that y is $x + 1$ while the equation $y - x - 1 = 0$ only implies, and hence says *implicitly*, that y is $x + 1$. It is sometimes said that the equation $x^2 + y^2 = 1$ determines y implicitly, but the fact is that the equation does not determine y. Saying that "y is either $\sqrt{1 - x^2}$ or $-\sqrt{1 - x^2}$" does not determine y any more than saying "a blonde did it" determines the culprit in a whodonit.

16 Write the first displayed formula of the preceding problem in the form

$$x^2 + xy + y^2 = 3$$

and use the Leibniz notation for derivatives to obtain the formula

$$\frac{dy}{dx} = -\frac{2x + y}{x + 2y}.$$

Observe, however, that the calculation is illusory unless y is a differentiable function for which the given equation holds.

17 Clarify matters relating to the two preceding problems by showing that y is a differentiable function satisfying the given equation if

$$y(x) = \frac{-x - \sqrt{3(4 - x^2)}}{2} \qquad (-2 < x < 2)$$

and also if

$$y(x) = \frac{-x + \sqrt{3(4 - x^2)}}{2} \qquad (-2 < x < 2).$$

18 A graph of the equation

$$x^2 + xy + y^2 = 3$$

appears in Figure 1.592. Find the equations of the tangents to this graph at the two points for which $x = 0$. Be sure to obtain results that agree with Figure 1.592.

19 It is not a simple matter to "solve" the equation

$$y^3 + y = x \tag{1}$$

for y. If y is a differentiable function of x for which the equation holds, however, we can differentiate with respect to x with the aid of the chain rule to obtain

(2) $$3y^2\frac{dy}{dx}+\frac{dy}{dx}=1 \qquad \text{or} \qquad 3[y(x)]^2y'(x)+y'(x)=1$$

and hence

(3) $$\frac{dy}{dx}=\frac{1}{3y^2+1} \qquad \text{or} \qquad y'(x)=\frac{1}{3[y(x)]^2+1}.$$

From (3) and the assumption that y is a differentiable function of x, we see that the derivative itself is a differentiable function of x. The derivative of the derivative is called the second derivative and is denoted by the symbols in the left members of the formulas

(4) $$\frac{d^2y}{dx^2}=\frac{-6y}{(3y^2+1)^3}, \qquad y''(x)=\frac{-6y}{(3y^2+1)^3}.$$

By differentiating the members of (3), show that the formulas (4) are correct.

20 Supposing that y is a differentiable function of x for which the given relation holds, differentiate with respect to x to find dy/dx when

(*a*) $xy=7$ *Ans.*: $\dfrac{dy}{dx}=-\dfrac{y}{x}$

(*b*) $\sin y=x$ *Ans.*: $\dfrac{dy}{dx}=\dfrac{1}{\cos y}$ or $\pm\dfrac{1}{\sqrt{1-x^2}}$

(*c*) $e^y=x$ *Ans.*: $\dfrac{dy}{dx}=\dfrac{1}{e^y}$ or $\dfrac{1}{x}$

(*d*) $\sin xy=x+y$ *Ans.*: $\dfrac{dy}{dx}=-\dfrac{1-y\cos xy}{1-x\cos xy}$

21 Find $f'(x)$ when

(*a*) $f(x)=\log(x+\sqrt{a^2+x^2})$ *Ans.*: $\dfrac{1}{\sqrt{a^2+x^2}}$

(*b*) $f(x)=\log(\sqrt{a^2+x^2}-x)$ *Ans.*: $\dfrac{-1}{\sqrt{a^2+x^2}}$

(*c*) $f(x)=(\log\sin 2x)^2$ *Ans.*: $\dfrac{4\cos 2x\log\sin 2x}{\sin 2x}$

(*d*) $f(x)=\log(\sin 2x)^2$ *Ans.*: $\dfrac{4\cos 2x}{\sin 2x}$

(*e*) $f(x)=\log\sin(2x)^2$ *Ans.*: $\dfrac{8x\cos 4x^2}{\sin 4x^2}$

(*f*) $f(x)=\left(\dfrac{x}{1+e^{2x}}\right)^n$ *Ans.*: $\dfrac{n[1+(1-2x)e^{2x}]x^{n-1}}{(1+e^{2x})^{n+1}}$

22 The *Hermite polynomials*, depending upon a parameter a that is usually taken to be 1 or 2 in applications, are defined by the formulas $H_0(x)=1$ and

(1) $$H_n(x)=(-1)^ne^{ax^2/2}\frac{d^n}{dx^n}e^{-ax^2/2} \qquad (n\geqq 1).$$

Show that if (1) holds, then

(2) $$H_n(x)e^{-ax^2/2} = (-1)^n \frac{d^n}{dx^n} e^{-ax^2/2}$$

(3) $$[-axH_n(x) + H_n'(x)]e^{-ax^2/2} = (-1)^n \frac{d^{n+1}}{dx^{n+1}} e^{-ax^2/2}$$

(4) $$axH_n(x) - H_n'(x) = (-1)^{n+1}e^{ax^2/2} \frac{d^{n+1}}{dx^{n+1}} e^{-ax^2/2}$$

and

(5) $$H_{n+1}(x) = axH_n(x) - H_n'(x).$$

Use (5) and the fact that $H_0(x) = 1$ to obtain the formulas

$$\begin{aligned}
H_0(x) &= 1\\
H_1(x) &= ax\\
H_2(x) &= a^2x^2 - a\\
H_3(x) &= a^3x^3 - 3a^2x\\
H_4(x) &= a^4x^4 - 6a^3x^2 + 3a^2\\
H_5(x) &= a^5x^5 - 10a^4x^3 + 15a^3x\\
H_6(x) &= a^6x^6 - 15a^5x^4 + 45a^4x^2 - 15a^3\\
H_7(x) &= a^7x^7 - 21a^6x^5 + 105a^5x^3 - 105a^4x.
\end{aligned}$$

23 The *Laguerre polynomials* are defined by the formulas $L_0(x) = 1$ and

$$L_n(x) = e^x \frac{d^n}{dx^n} (x^n e^{-x}) \qquad (n = 1,2,3, \cdots).$$

Show that

$$\begin{aligned}
L_0(x) &= 1\\
L_1(x) &= -x + 1\\
L_2(x) &= x^2 - 4x + 2\\
L_3(x) &= -x^3 + 9x^2 - 18x + 6\\
L_4(x) &= x^4 - 16x^3 + 72x^2 - 96x + 24.
\end{aligned}$$

24 Supposing that $y = e^{\sin t}$ or $h(t) = e^{\sin t}$, use the chain rule and the formula for derivatives of products to obtain the first three derivatives with respect to t of these things. *Ans.:*

(1) $$\frac{dy}{dt} = h'(t) = e^{\sin t} \cos t$$

(2) $$\frac{d^2y}{dt^2} = h''(t) = -e^{\sin t} \sin t + e^{\sin t} \cos^2 t$$

(3) $$\frac{d^3y}{dt^3} = h'''(t) = -e^{\sin t} \cos t - 3e^{\sin t} \cos t \sin t + e^{\sin t} \cos^3 t.$$

25 Assuming existence of all of the derivatives we want to use, show that if $h(t) = f(g(t))$, then

(1) $$h'(t) = f'(g(t))g'(t)$$
(2) $$h''(t) = f'(g(t))g''(t) + f''(g(t))[g'(t)]^2$$

and write a formula for $h'''(t)$. Then show that these formulas reduce to those of Problem 24 when $f(x) = e^x$ and $g(t) = \sin t$.

26 The preceding problem involved three functions and the Newton notation for derivatives. This problem requires use of the Leibniz notation. Supposing that "y is a function of x and x is a function of t so y is a function of t," and that each function has three or more derivatives, write formulas for the first three derivatives of y with respect to t. Finally, check your answers against those of the preceding two problems. *Partial ans.:*

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} \tag{1}$$

$$\frac{d^2y}{dt^2} = \frac{dy}{dx}\frac{d^2x}{dt^2} + \frac{d^2y}{dx^2}\left(\frac{dx}{dt}\right)^2 \tag{2}$$

$$\frac{d^3y}{dt^3} = \frac{dy}{dx}\frac{d^3x}{dt^3} + 3\frac{d^2y}{dx^2}\frac{d^2x}{dt^2}\frac{dx}{dt} + \frac{d^3y}{dx^3}\left(\frac{dx}{dt}\right)^3. \tag{3}$$

27 Suppose we momentarily agree that the first of the formulas

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}, \qquad \frac{d^2y}{dt^2} = \frac{d^2y}{dx^2}\left(\frac{dx}{dt}\right)^2 \ (?)$$

is true "because" we get a correct result by canceling dx's from the right side. Show that we should not apply the same "reasoning" to the second formula.

28 Read Theorem 3.65 and observe that the hypotheses are satisfied if $f(x) = 1 + x + x^2$, $g(x) = 0$, and $u = g(x) = 0$ for each x, while $y = f(u)$ for each u so that $y = f(g(x)) = 1$ for each x. Hence the conclusion of the theorem implies that

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

for each x. Observe that $dy/dx = 0$ and $du/dx = 0$ for each x. Our major question now appears. Is there a reason for uneasiness about the meaning of dy/du when $u = 0$ for each x? *Remark and ans.:* This question was raised by an extremely sane person who happened at the moment to be thinking too much about the manner in which we read dy/du and too little about the meaning of dy/du. According to our basic definition, dy/du is $f'(u)$, the derivative of f at u. Since $f(x) = 1 + x + x^2$ for each x, we find that $f'(x) = 1 + 2x$ for each x, so $f'(u) = 1 + 2u$ for each u. Thus, $dy/du = 1 + 2u$. If it happens that $u = 0$ for each x, then $dy/du = 1$ for each x. We have no reason to be uneasy unless we manufacture trouble by recreating old tales about varying variables that we sometimes call galloping numbers. The notation of Leibniz is often more convenient than that of Newton, but it is also more likely to engender mental aberrations. Nobody expects u to be galloping around while we calculate $f'(u)$.

29 Is the function f for which $f(x) = |x|$ an elementary function? *Remark and ans.:* This is a tricky question. An intelligent person should make an incorrect guess until he discovers or is reminded that $|x| = \sqrt{x^2}$. The function f is elementary, but $f'(0)$ does not exist.

30 Let p and q be positive integers. Let $y(0) = 0$ and let

$$y(x) = x^p \sin\frac{1}{x^q} \qquad (x \neq 0).$$

Show that, when $x \neq 0$,

$$y'(x) = -qx^{p-q-1} \cos \frac{1}{x^q} + px^{p-1} \sin \frac{1}{x^q}.$$

Tell why this formula cannot be valid when $x = 0$. Then show with the aid of the sandwich theorem 3.287 that if $p \geqq 2$, then $y'(0) = 0$. Show that if $p > q + 1$ then $\lim_{x \to 0} y'(x) = 0$. Show that if $p \leqq q + 1$, then $\lim_{x \to 0} y'(x)$ does not exist.

31 Before starting this problem, we make the profound observation that 0 times a number is 0 but that nobody ever tries to define the product of 0 and something that does not exist. With this in mind, show that the first of the formulas

$$(1) \qquad \frac{d}{dx}|x|^2 = 2x, \qquad \frac{d}{dx}|x|^2 = 2|x|\frac{d|x|}{dx}$$

is valid for each x and that the second is valid if, and only if, $x \neq 0$. *Hint:* For the first part, observe that $|x|^2 = x^2$. For the second part, consider separately the cases for which $x > 0$, $x < 0$, and $x = 0$. *Remark:* Putting $f(x) = x^2$ and $g(x) = |x|$ shows that $g'(x)$ can fail to exist even when it is known that $df(g(x))/dx$ exists. The calculations in

$$(2) \qquad \phi(x) = f(g(x)), \qquad \phi'(x) = f'(g(x))g'(x), \qquad \frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

may therefore be incorrect even when ϕ and f are both differentiable. In any case, we are not doing rigorous mathematics when we start with the first of the formulas

$$(3) \qquad \sin y = x, \qquad \cos y \frac{dy}{dx} = 1$$

and obtain the second without giving a thought to the question whether y is a differentiable function of x. Congratulations can therefore be showered upon students who, at this time, have a healthy lack of enthusiasm for problems like Problem 20.

3.7 *Rates, velocities* Let f be defined over some interval $a \leqq x \leqq b$ and let $y = f(x)$. When x and $x + \Delta x$ both lie between a and b and $\Delta x \neq 0$, the difference quotient in

$$(3.71) \qquad \frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

is the *average rate of change of y with respect to x* over the interval from the lesser to the greater of x and $x + \Delta x$. If this average rate (which is the difference quotient) has a limit as Δx approaches zero, then this limit [which is the derivative dy/dx or $f'(x)$] is the *rate of change of y with respect to x* at the given x. These are definitions which can, perhaps without disastrous loss of meaning, be abbreviated to the forms

$$(3.72)\quad \text{Average rate} = \text{difference quotient} = \frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$(3.73)\quad \text{Rate} = \text{derivative} = \frac{dy}{dx} = f'(x).$$

Of course, we are never required to prove definitions, but these are important and we must have or acquire an understanding of them and a feeling that they do (or do not) use words of the English language in a reasonably appropriate way. Shifting the letters from y and x to x and t, we see that the definition involving (3.72) shows that if x is a number of miles and t is a number of hours, then the average rate of change of x with respect to t is a number Δx of miles divided by a number Δt of hours and hence is a number of miles per hour. Some applications of this are very simple and agree with all primitive ideas about rates. When we are thinking about a particular automobile journey in which the automobile moves steadily in one direction along a straight road, we can let x and $f(t)$ denote the distance (number of miles) traveled during the first t hours of the trip. We are all accustomed to calculating the average rate over a given time interval and to calling this average rate an average speed. Suppose now that an untutored (but not necessarily stupid) individual is asked how he might, without looking at a perfect speedometer, determine a number Q which could reasonably be called the speed at a particular time t. His reply might be lengthy and partially intelligible. He should, sooner or later, arrive at the idea that the average speed over a long trip is likely to be a very bad approximation to Q, but that the average speed over the time interval from t to $t + \Delta t$ (or from $t + \Delta t$ to t in case $\Delta t < 0$) should be near Q whenever Δt is near 0 but $\Delta t \neq 0$. We have learned how to make this idea precise. It is done in the definitions we are discussing. Similar stories involving other rates (degrees centigrade per centimeter, coulombs per second, and dollars per year, for examples) show that the definitions are sensible and should have swarms of important applications.

Our simple discussion of the journey of an automobile moving steadily in one direction along a straight road involved the word "speed" but carefully avoided the words "velocity" and "acceleration." To appreciate what is coming, we should know some history. The words "speed," "velocity," and "acceleration" are very old. A long time ago, say before the year 1900, they were all numbers (scalars); velocity and acceleration could be negative but speed never could be. Nowadays, in all enlightened communities, velocities and accelerations are always vectors and we must learn about them. To get started, we consider the path traced by a bumblebee (or molecule or rocket or satellite or what not) as it buzzes through space E_3. While other tactics are both possible and useful, we suppose that we have a rectangular x, y, z coordinate system bearing unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ as in Section 2.2. At each time t, the coordinates of the bumblebee can be denoted by $x(t)$, $y(t)$, $z(t)$. Letting $\mathbf{r}(t)$ denote the vector running from the origin to the bumblebee, we obtain the vector equation

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}. \tag{3.74}$$

This vector $\mathbf{r}(t)$ is called the *displacement* (or *displacement vector*) of the bumblebee at time t. Supposing that $\Delta t \neq 0$, we can write

$$\mathbf{r}(t + \Delta t) = x(t + \Delta t)\mathbf{i} + y(t + \Delta t)\mathbf{j} + z(t + \Delta t)\mathbf{k} \tag{3.75}$$

and form the difference quotient

$$\frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} = \frac{x(t + \Delta t) - x(t)}{\Delta t}\mathbf{i} + \frac{y(t + \Delta t) - y(t)}{\Delta t}\mathbf{j} + \frac{z(t + \Delta t) - z(t)}{\Delta t}\mathbf{k}$$

which can be written in the abbreviated form

$$\frac{\Delta \mathbf{r}}{\Delta t} = \frac{\Delta x}{\Delta t}\mathbf{i} + \frac{\Delta y}{\Delta t}\mathbf{j} + \frac{\Delta z}{\Delta t}\mathbf{k}. \tag{3.751}$$

In accordance with general terminology, this difference quotient is called the average rate of change of the vector $\mathbf{r}(t)$ with respect to t over the interval from the lesser to the greater of t and $t + \Delta t$. It is also called the *average velocity* of the bumblebee over this same interval. In case the above difference quotients have limits as $\Delta t \to 0$, the limit of the average velocity is called the *velocity* at time t and is denoted by $\mathbf{v}(t)$. Thus

$$\mathbf{v}(t) = \mathbf{r}'(t) = x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k} \tag{3.76}$$

or

$$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k} \tag{3.761}$$

provided the derivatives exist. Figures 3.762, 3.763, and 3.764 show how the vectors $\Delta\mathbf{r}$, $\Delta\mathbf{r}/\Delta t$, and $\mathbf{v}(t)$ might appear in a particular example.

Figure 3.762

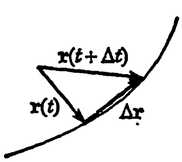


Figure 3.763

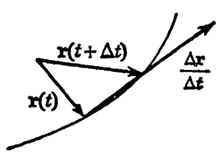
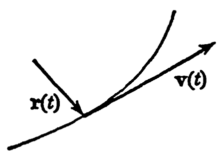


Figure 3.764



The scalar components of the velocity $\mathbf{v}$ or $\mathbf{v}(t)$ are sometimes denoted by the symbols v_x, v_y, v_z, so that

$$(3.765)\qquad v_x = x'(t) = \frac{dx}{dt},\qquad v_y = y'(t) = \frac{dy}{dt},\qquad v_z = z'(t) = \frac{dz}{dt}$$

and

$$(3.766)\qquad \mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k}.$$

The *acceleration* $\mathbf{a}(t)$ is a vector which is defined in terms of velocities in the same way that velocities are defined in terms of displacements. Thus, provided the derivatives exist,

$$(3.77)\qquad \mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t) = x''(t)\mathbf{i} + y''(t)\mathbf{j} + z''(t)\mathbf{k}$$

or

$$(3.771)\qquad \mathbf{a}(t) = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} = \frac{d^2x}{dt^2}\mathbf{i} + \frac{d^2y}{dt^2}\mathbf{j} + \frac{d^2z}{dt^2}\mathbf{k},$$

where the "double prime" in (3.77) and the number 2 appearing in "dee squared x dee t squared" in (3.771) denote second derivatives, that is, derivatives of derivatives. We still have to learn what is meant by the *speed* (a scalar) of the bumblebee. It is defined by

$$(3.78)\qquad \text{Speed} = \text{length of velocity vector},$$

so that, in our notation,

$$(3.781)\qquad \text{Speed} = |\mathbf{v}(t)| = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}.$$

Perhaps it should be explained that the t appearing in the above equations is called a *parameter*, that (3.74) is a *parametric equation* of the path, and that the path is the *graph* of the parametric equation. According to this definition, a parameter is a number. It is an element of the domain of the functions in (3.74), and we need not complicate our lives by harboring impressions that parameters are complicated things. Section 7.1 gives a careful explanation of circumstances in which the graph is called a curve.

In Section 5.1 we shall give a rather detailed discussion of tangents to graphs. Meanwhile, it can be remarked that if the vectors in (3.751) and (3.761) are not 0, then the line through $P(x,y,z)$ and $P(x + \Delta x, y + \Delta y, z + \Delta z)$ is called a *chord* of the curve being considered, and the line through $P(x,y,z)$ having the direction of $\mathbf{v}(t)$ is called the *tangent* to the curve at $P(x,y,z)$. Therefore, we can find the direction of the tangent to a sufficiently decent curve by finding the velocity of a particle which moves along the curve with nonzero velocity. The tangent line and the velocity vector have, by definition, the same direction. To bridge the gap between our work and plebeian terminology used in the prosaic workaday world, we need still another definition. When we say that a moving body is, at time t, "going in the direction of a vector $\mathbf{w}$" we mean that its velocity $\mathbf{v}$ has the direction of $\mathbf{w}$, that is, $\mathbf{v} = k\mathbf{w}$, where k is a positive scalar. In case $\mathbf{w}$ is a unit vector, the scalar k is the speed of the body. It is not so easy to tell what the body is "doing" when $\mathbf{v} = 0$ or $\mathbf{v}$ does not exist. The ancient Greek philosophers tried to make people think about motion, and we never quite know how much they smiled when they insisted that an arrow cannot move where it is and cannot move where it isn't and, hence, cannot move at all. Thoughts about such matters can bring the conviction that definitions are not superfluous.

A few simple observations should be made. In case the bumblebee buzzes around in a plane which we take to be the xy plane, the above story is unchanged but calculations are simplified by the fact that $z(t) = 0$ for each t. In case the bumblebee buzzes around in (or on) a line which we take to be the x axis, we have $y(t) = z(t) = 0$ for each t. In modern terminology, scalars cannot be velocities but can be scalar components of velocities. In case a particle moves on a coordinate axis or on a line parallel to a coordinate axis, its velocity and acceleration are still vectors but their scalar components in the direction of the axis are scalars which we shall call the *scalar velocity* and *scalar acceleration* of the particle. For example, if a particle is moving on an x axis in such a way that its x coordinate at time t is the scalar (or number)

$$x = At^3 + Bt^2 + Ct + D + E \sin \omega t,$$

then the scalar (or number) v (not $\mathbf{v}$) defined by

$$v = \frac{dx}{dt} = 3At^2 + 2Bt + C + E\omega \cos \omega t$$

is its scalar velocity at time t and the scalar (or number) a (not $\mathbf{a}$) defined by

$$a = \frac{d^2x}{dt^2} = 6At + 2B - E\omega^2 \sin \omega t$$

is its scalar acceleration at time t. The speed is $|dx/dt|$. When the positive x lies to the right of the origin, the particle is "moving to the right" at those times for which $dx/dt > 0$ and is "moving to the left" when $dx/dt < 0$. It is not so easy to tell what the particle is "doing" when $dx/dt = 0$ or dx/dt does not exist.

Problems 3.79

1 A stone is thrown downward 10 feet per second from the deck of a bridge. The distance s it will have fallen t seconds later is assumed to be

$$s = 10t + 16t^2.$$

Supposing that $0 < t < t + \Delta t$, work out a formula for the average speed of the stone over the time interval from t to $t + \Delta t$. Then work out a formula for the speed at time t. *Ans.:* $10 + 32t + 16\Delta t$ and $10 + 32t$.

2 A vertical y axis has its positive part above the origin. A particle moves upon this axis in such a way that its coordinate at each time t is

$$y = -At^2 + Bt + C,$$

where A is a positive number. Show that the scalar velocity v is

$$v = -2At + B,$$

that the particle is going up when $t < B/2A$, and that the particle is going down when $t > B/2A$. Show that the scalar acceleration is always $-2A$. Show that the greatest height attained by the particle is $B^2/4A + C$.

3 In the context of the preceding problem, so determine A, B, and C that the scalar acceleration is always -32 and the particle is 3 units below the origin and going upward with speed 8 when $t = 0$.

4 A particle moves along the x axis in such a way that its x coordinate at time t is

$$x = 2t^5 - 5t^4 - 2t^2 - 2t + 1.$$

Find its scalar velocity and scalar acceleration at time t. *Ans.:*

$$10t^4 - 20t^3 - 4t - 2, \qquad 40t^3 - 60t^2 - 4.$$

5 A body moves on a line in such a way that its coordinate x at time t is

$$x = \frac{t^3}{3} - 4t^2 + 15t + 6.$$

Find the time interval over which the scalar velocity is negative, and find the distance the body moves during the interval. *Ans.:* $\frac{4}{3}$.

6 A particle moves along an x axis in such a way that, when $t \geqq 0$, its coordinate is

$$x = \sqrt{k^2t^2 + c^2},$$

where k and c are positive constants. Show that its speed is always less than k and approaches k as t becomes infinite.

7 If an oak tree in Ohio was 20 feet tall when it was 15 years old and was 36 feet tall when it was 25 years old, the average rate of change of height (measured in feet) with respect to time (measured in years) over the 10-year interval is 1.6 feet per year. Tell why it is not reasonable to suppose that the tree grew steadily at the rate of 1.6 feet per year for 10 years. If the tree grew from height 30 feet to height 32 feet in a calendar year from January 1 to December 31, sketch a reasonably realistic graph which shows how the height of the tree might depend upon t during the year.

8 The charge Q (measured in coulombs) on an electrical capacitor at time t is $Q_0 \sin \omega t$, where Q_0 and ω (omega) are constants. The rate of change of Q with respect to t (measured in coulombs per second, that is, in amperes) is the current I in the circuit containing the capacitor. Find a formula which gives I in terms of t.

9 This problem involves *uniform circular motion.* Let a particle P start at the point $(a,0)$ of the plane Figure 3.791 and move around the circle with center at the origin and radius a in such a way that the vector $\overrightarrow{OP}$ rotates at the constant positive rate ω (omega) radians per second. Letting $\mathbf{r} = \overrightarrow{OP}$, show that

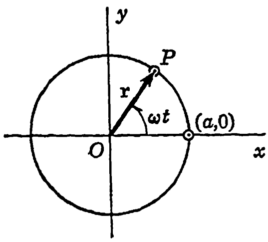


Figure 3.791

$$\begin{aligned}
&(1) \qquad & \mathbf{r} &= a(\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j}) \\
&(2) \qquad & \mathbf{v} &= a\omega(-\sin \omega t\mathbf{i} + \cos \omega t\mathbf{j}) \\
&(3) \qquad & \mathbf{a} &= -a\omega^2(\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j})
\end{aligned}$$

and hence that

$$(4) \qquad \mathbf{a} = -a\omega^2\mathbf{u},$$

where, at each time t, $\mathbf{u}$ is a unit vector running from the origin toward P. Show that $\mathbf{r}\cdot\mathbf{v} = 0$ and interpret this result. Show that $|\mathbf{v}| = a\omega$ and interpret this result. *Remark:* The result (4) is important in physics. It says that, in uniform circular motion, the particle is always accelerated toward the center and that the magnitude of the acceleration is $a\omega^2$. Some additional terminology should be encountered frequently and slowly absorbed. When a particle moves upon a line in such a way that its coordinate at time t is $A + B \sin (\omega t + \phi)$, the motion is said to be *sinusoidal* or (particularly in old books) *harmonic* or *simple harmonic.* The numbers ϕ, $\omega/2\pi$, and B are the *phase*, the *frequency* (cycles per unit time), and the *amplitude* of the motion. Glances at the components of $\mathbf{r}$ and $\mathbf{a}$ in the above formulas show that the projection of P upon a diameter (line, not number) of the circle executes sinusoidal motion. Moreover, the projection is always accelerated toward the center, and the magnitude of the acceleration is proportional to the distance from the center. See also Problem 16.

10 As in the text of this section, let a particle move in E_3 in such a way that its displacement, velocity, and acceleration are

$$\begin{aligned}
&(1) \qquad & \mathbf{r}(t) &= x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} \\
&(2) \qquad & \mathbf{v}(t) &= x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k} \\
&(3) \qquad & \mathbf{a}(t) &= x''(t)\mathbf{i} + y''(t)\mathbf{j} + z''(t)\mathbf{k}
\end{aligned}$$

and the square of its speed is

$$(4) \qquad |\mathbf{v}(t)|^2 = [x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2.$$

Using this information, prove that if the particle moves with constant speed c, then the acceleration is always orthogonal to the velocity. *Hints:* Do not get scared. Equate the right member of (4) to c^2. Equate the derivatives of the members of your equation. Look at your result. *Remark:* One who thinks that this result is mysterious should remember or discover which way he tends to topple when he sits in an automobile which rounds an unbanked curve at constant speed.

11 This problem involves the *uniform circular helical motion* of a particle Q in E_3 which runs up the helix (spiral staircase) of Figure 3.792 in such a way that its projection P upon the xy plane executes the uniform circular motion of Problem 9 whiles its z coordinate increases at the positive rate b units per second. Supposing that Q occupies the position $(a,0,0)$ when $t = 0$ and letting $\mathbf{r} = \overrightarrow{OQ}$, show that

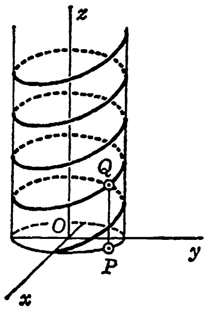


Figure 3.792

$$\begin{aligned} \mathbf{r} &= a(\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j}) + bt\mathbf{k} \\ \mathbf{v} &= a\omega(-\sin \omega t\mathbf{i} + \cos \omega t\mathbf{j}) + b\mathbf{k} \\ \mathbf{a} &= -a\omega^2(\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j}). \end{aligned}$$

Find the speed of Q. *Remark:* One who gets interested in this helix may try to find the length of one turn by two different methods. First, use the speed of Q in an appropriate way. Second, find out what happens when the cylinder upon which the helix lies is cut along a vertical generator and rolled out flat.

12 A projectile P moves in such a way that its displacement vector at time t is

$$\mathbf{r} = (v_0 \cos \alpha)t\mathbf{i} + [(v_0 \sin \alpha)t - \tfrac{1}{2}gt^2]\mathbf{j}, \tag{1}$$

where α, v_0, g are constants for which $0 < \alpha < \pi/2$, $v_0 > 0$, $g > 0$. Show that its velocity at time t is

$$\mathbf{v} = [(v_0 \cos \alpha)\mathbf{i} + (v_0 \sin \alpha)\mathbf{j}] - gt\mathbf{j}. \tag{2}$$

Show that $\mathbf{a} = -g\mathbf{j}$. Show that the coordinates x, y of P at time t are

$$x = (v_0 \cos \alpha)t, \qquad y = (v_0 \sin \alpha)t - \tfrac{1}{2}gt^2. \tag{3}$$

Eliminate t to obtain the equation

$$y = (\tan \alpha)x - \frac{gx^2}{2v_0^2 \cos^2 \alpha} \tag{4}$$

and note that the path of the projectile is a part of a parabola. Show that $y = 0$ when $t = 0$ and that the projectile is then at the origin. Show that $y = 0$ when $t = (2v_0 \sin \alpha)/g$ and that the projectile is then at the point $(R,0)$, where

$$R = \frac{2v_0^2 \sin \alpha \cos \alpha}{g} = \frac{v_0^2 \sin 2\alpha}{g}. \tag{5}$$

This number R is called the *range* of the projectile, and this range is clearly a maximum when $\sin 2\alpha = 1$ and hence when $2\alpha = \pi/2$ and $\alpha = \pi/4$. Show that the initial velocity (velocity at time $t = 0$) is

$$(v_0 \cos \alpha)\mathbf{i} + (v_0 \sin \alpha)\mathbf{j} \tag{6}$$

and that this makes the angle α with the positive x axis. Show that the initial speed is v_0. Tell, in terms of vectors, how the velocity at later times is related to the initial velocity. Find the velocity (not merely speed) of the projectile when it hits the point $(R,0)$.

13 While the matter must remain mysterious until some mathematical secrets have been revealed, the tip P of a cog of a particular hypocyclic gear moves in such a way that its displacement vector at time t is

$$\mathbf{r} = a(\cos^3 \omega t\mathbf{i} + \sin^3 \omega t\mathbf{j}),$$

where a and ω are positive constants. Find and simplify formulas for its velocity, speed, and acceleration. *Ans.:*

$$\mathbf{v} = \frac{3a\omega}{2} \sin 2\omega t(-\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j})$$

$$\text{Speed} = \frac{3a\omega}{2} |\sin 2\omega t|$$

$$\mathbf{a} = \frac{3a\omega^2}{2} \sin 2\omega t(\sin \omega t\mathbf{i} + \cos \omega t\mathbf{j}) + 3a\omega^2 \cos 2\omega t(-\cos \omega t\mathbf{i} + \sin \omega t\mathbf{j}).$$

14 A particle P moves in such a way that its displacement vector at time t is

$$\mathbf{r} = \frac{2t}{t^2+1}\mathbf{i} + \frac{t^2-1}{t^2+1}\mathbf{j}.$$

Show that $|\mathbf{r}| = 1$ at all times and hence that the path of P must lie on the unit circle with center at the origin. If a particle moves on a circle in such a way that it has a nonzero velocity vector $\mathbf{v}$, then $\mathbf{v}$ must be tangent to the circle and hence orthogonal (or perpendicular) to $\mathbf{r}$. Check this story by calculating $\mathbf{v}$ and showing that $\mathbf{v}\cdot\mathbf{r} = 0$. Find the times at which the particle crosses the coordinate axes, and then obtain more information about the motion of P.

15 Prove that if a particle P moves in E_3 in such a way that it has displacement vectors and velocity vectors $\mathbf{r}(t)$ and $\mathbf{v}(t)$ at time t, then

$$\frac{d}{dt}|\mathbf{r}(t)| = \frac{\mathbf{r}(t)\cdot\mathbf{v}(t)}{|\mathbf{r}(t)|}$$

when the particle is not at the origin. Tell why this implies that if the path of P lies on a sphere with center at the origin, then $\mathbf{v}$ must be normal (or orthogonal or perpendicular) to $\mathbf{r}$ and hence $\mathbf{v}$ must be tangent to the sphere. *Remark:* It can be presumed that we do not known much about curves and surfaces in E_3, but we can presume that if a particle P makes a decent trip along a decent curve lying on a decent surface, then at each time the velocity vector having its tail at P must be tangent to the surface as well as to the curve.

16 If a particle moves along the x axis in such a way that, at time t,

$$x = a \sin (\omega t + \phi)$$

where a, ω, ϕ are constants for which $a > 0$ and $\omega > 0$, the particle is said to describe (or execute) sinusoidal (or harmonic) motion. Calculate the first and second derivatives of x with respect to t and show that

$$\frac{d^2x}{dt^2} = -\omega^2 x.$$

Remark: This shows that the scalar acceleration of the particle is proportional to the scalar displacement of the particle from the origin (or equilibrium position) about which it oscillates.

17 A particle P moves in E_3 in such a way that the vector $\mathbf{r}$ running from the origin to P is

$$\mathbf{r} = r[\sin\theta\cos\phi\mathbf{i} + \sin\theta\sin\phi\mathbf{j} + \cos\theta\mathbf{k}],$$

where r, θ, and ϕ are all differentiable functions of t. Find the velocity and speed of the particle at time t. *Ans.:*

$$\mathbf{r}'(t) = r'(t)\mathbf{u} + r(t)\theta'(t)\mathbf{v} + r(t)\phi'(t)\sin\theta\mathbf{w}$$

where
$$\begin{aligned}\mathbf{u} &= \sin\theta\cos\phi\mathbf{i} + \sin\theta\sin\phi\mathbf{j} + \cos\theta\mathbf{k}\\ \mathbf{v} &= \cos\theta\cos\phi\mathbf{i} + \cos\theta\sin\phi\mathbf{j} - \sin\theta\mathbf{k}\\ \mathbf{w} &= -\sin\phi\mathbf{i} + \cos\phi\mathbf{j}\end{aligned}$$

and

$$|\mathbf{r}'(t)| = \{[r'(t)]^2 + [r(t)\theta'(t)]^2 + [r(t)\phi'(t)\sin\theta]^2\}^{1/2}.$$

Remark: Problem 5 of Problems 2.69 shows that the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are orthonormal. The numbers r, ϕ, and θ are spherical coordinates which appear in Figure 10.12 and are studied in Chapter 10. When a is a positive number and $r = a$ for each t, P is always on a sphere and the above formulas become the standard formulas used for study of curves that lie on spheres. Chapter 7 gives solid information about curves.

18 A spherical earth has its center at the origin and has radius a. A particle P moves on the surface S of the earth in such a way that the vector $\mathbf{r}$ running from the origin to P is

$$\mathbf{r} = a[\sin\theta\cos\phi\mathbf{i} + \sin\theta\sin\phi\mathbf{j} + \cos\theta\mathbf{k}] \tag{1}$$

where θ and ϕ are differentiable functions of t. Show that the velocity of P at time t is

$$\mathbf{r}'(t) = a\theta'(t)\mathbf{v} + a\phi'(t)\sin\theta\mathbf{w} \tag{2}$$

where

$$\mathbf{v} = \cos\theta\cos\phi\mathbf{i} + \cos\theta\sin\phi\mathbf{j} - \sin\theta\mathbf{k} \tag{3}$$
$$\mathbf{w} = -\sin\phi\mathbf{i} + \cos\phi\mathbf{j}. \tag{4}$$

Remark: We invest a moment to look at some facts involving compass directions. When $0 < \theta < \pi$ so that P is neither at the north pole nor at the south pole, the vector $\mathbf{v}$ points south from P and the vector $\mathbf{w}$ points east from P. When θ and ϕ are so related that, for some constant q,

$$\phi'(t)\sin\theta = q\theta'(t), \tag{5}$$

the vector $\mathbf{r}'(t)$ and the path of P always make the same constant angle with $\mathbf{w}$ and hence always have the same compass direction. The path of P is then said to be a *rhumb curve*, or *loxodrome*. Such curves are followed by ships that keep sailing northeast. When we have learned more calculus, we will be able to show that (5) holds if and only if there is a constant c for which

$$\phi = q\log\frac{1-\cos\theta}{\sin\theta} + c \quad\text{or}\quad \phi = q\log(\csc\theta - \cot\theta) + c \tag{6}$$

$$\text{or}\quad \phi = q\log\tan\frac{\theta}{2} + c.$$

19 The vector formula

$$\mathbf{r} = (b + a \cos \theta)\cos \phi\mathbf{i} + (b + a \cos \theta)\sin \phi\mathbf{j} + a \sin \theta\mathbf{k}$$

of Problem 22 at the end of Section 2.2 provides the possibility of studying curves on a torus. Supposing that θ and ϕ are differentiable functions of t, find $\mathbf{r}'(t)$. *Ans.:*

$$\mathbf{r}'(t) = a\theta'(t)[-\sin \theta \cos \phi\mathbf{i} - \sin \theta \sin \phi\mathbf{j} + \cos \theta\mathbf{k}] + (b + a \cos \theta)\phi'(t)[-\sin \phi\mathbf{i} + \cos \phi\mathbf{j}].$$

20 The rod OP of the linkage of Figure 3.793 has length a and has one end fixed at the origin O. The rod QP has length b. Its lower end moves to and fro on the x axis in such a way that its x coordinate is $c + \sin \omega t$ at time t. Its upper end is fastened to the first rod at P, and the motion of Q causes the first rod to rotate. Write the formula (the law of cosines) which expresses b^2 in terms of other quantities, and differentiate to obtain a formula for $d\theta/dt$. Then use the formula

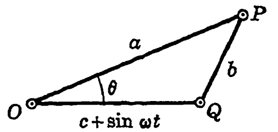


Figure 3.793

$$\mathbf{r} = \overrightarrow{OP} = a\,(\cos \theta\mathbf{i} + \sin \theta\mathbf{j})$$

to obtain a formula for the velocity $\mathbf{v}$ of P. *Ans.:*

$$\mathbf{v} = a\omega \frac{[a \cos \theta - (c + \sin \omega t)] \cos \omega t}{a(c + \sin \omega t) \sin \theta} (-\sin \theta\mathbf{i} + \cos \theta\mathbf{j}).$$

21 A cam can furnish us something to differentiate. A circular disk of radius a is mounted on a cam shaft at O. Supposing that C is the center of the disk and that $0 < b \leqq a$, let $|\overrightarrow{OC}| = b$. The eccentric disk rotates about O with constant angular speed, the angle POC being ωt. The mechanism to the right of the disk in Figure 3.794 keeps the point $P(x,0)$ of a rod pressed against the rotating disk so that $P(x,0)$ moves to and fro on the x axis as the disk rotates. The formulas

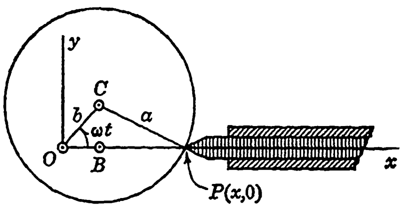


Figure 3.794

(1) $\qquad \overrightarrow{OB} = b \cos \omega t\mathbf{i}, \qquad \overrightarrow{BC} = b \sin \omega t\mathbf{j}, \qquad |\overrightarrow{BP}| = \sqrt{a^2 - |BC|^2}$

show that

(2) $$\mathbf{r} = [b \cos \omega t + \sqrt{a^2 - b^2 \sin^2 \omega t}]\mathbf{i},$$

where $\mathbf{r} = \overrightarrow{OP} = \overrightarrow{OB} + \overrightarrow{BP}$. Find the velocity $\mathbf{v}$ and the acceleration $\mathbf{a}$ of P, and do not spend all day trying to discover the significance of the fact that (2) reduces to

(3) $$\mathbf{r} = a[\cos \omega t + |\cos \omega t|]\mathbf{i}$$

when $b = a$.

22 Let a particle of mass m move in a vertical plane in such a way that its coordinates x, y are differentiable functions of t (time) and the vector $\mathbf{r}(t)$ running from the origin to the particle at time t is

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}.$$

The kinetic energy E_1 of the particle can then be calculated from the formula

$$E_1 = \tfrac{1}{2}m|\mathbf{v}(t)|^2.$$

Assuming that a constant gravitational force $-mg\mathbf{j}$ is exerted upon the body, we can calculate its gravitational potential energy E_2 from the formula

$$E_2 = mgy(t).$$

In appropriate circumstances, the number E defined by

$$E = E_1 + E_2$$

is the total energy of the particle. For present purposes we do not need basic information about these things, but we should know enough calculus to be able to calculate the total energy at time t of a projectile of mass m for which

$$\mathbf{r}(t) = c_0t\mathbf{i} + (c_1t - \tfrac{1}{2}gt^2)\mathbf{j}.$$

Do it. *Ans.:* $E = \frac{1}{2}m(c_0^2 + c_1^2)$. *Remark:* The fact that E has the same value at all times is no surprise to persons who know about "conservation of energy."

23 Supposing that

$$\mathbf{r}(t) = \rho(\cos\phi\mathbf{i} + \sin\phi\mathbf{j}), \tag{1}$$

where ρ and ϕ are functions of t having two derivatives, differentiate to obtain $\mathbf{r}'(t)$ and then differentiate again to obtain $r''(t)$. *Ans.:*

$$\mathbf{r}''(t) = \left[\frac{d^2\rho}{dt^2} - \rho\left(\frac{d\phi}{dt}\right)^2\right](\cos\phi\mathbf{i} + \sin\phi\mathbf{j}) + \left[2\frac{d\rho}{dt}\frac{d\phi}{dt} + \rho\frac{d^2\phi}{dt^2}\right](-\sin\phi\mathbf{i} + \cos\phi\mathbf{j}). \tag{2}$$

24 Give a full statement of reasons why it is true that if ρ and ϕ are functions of t such that ρ has one derivative and ϕ has two derivatives, then the product $\rho^2\dfrac{d\phi}{dt}$ has one derivative and

$$\frac{d}{dt}\rho^2\frac{d\phi}{dt} = 2\rho\frac{d\rho}{dt}\frac{d\phi}{dt} + \rho^2\frac{d^2\phi}{dt^2}.$$

25 We can pick up assorted ideas by thinking about income tax rates. Let $I(x)$ denote the income tax which T, a taxpayer in a particular class, must pay when his net taxable income is x. Of course x and $I(x)$ are to be measured in appropriate units such as dollars or marks or kilobucks. A government may decree that if $x_1 \leqq x \leqq x_2$, then $I(x)$ is y_1 plus k_1 per cent of the excess of x over x_1. We can simplify this by letting $m_1 = k_1/100$ and writing

$$I(x) = y_1 + m_1(x - x_1) \qquad (x_1 \leqq x \leqq x_2).$$

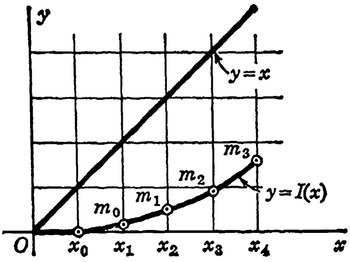


Figure 3.795

Our knowledge of the natures of graphs of lines now shows that the graph of $y = I(x)$ over the interval $x_1 \leqq x \leqq x_2$ is a line segment having slope m_1. Figure 3.795 shows the graph of $y = I(x)$ that can be constructed from information given in a possible tax table. Everybody says that tax rates are different in different tax brackets, being for example m_1 (or $100m_1$ per cent) when $x_1 < x < x_2$. Our fundamental problem is the following. Show that if $x_1 < x < x_2$, then our definition of a rate as a derivative is in agreement with the things that have been said about tax rates. Show that when x is x_1, the tax rate does not exist but that the "right-hand rate" and "left-hand rate" do exist.

3.8 *Related rates* Useful information about derivatives and their applications can be gained by solving problems more or less like the following one. Figure 3.81 represents a ladder which is 10 units (feet or meters) long. The top of the ladder rests against a vertical wall and is 8 units above the horizontal floor upon which the bottom of the ladder rests 6 units from the wall. It is supposed that the bottom of the ladder is moving away from the wall at the rate of 2 units per second, and we are required to find the rate at which the invisible man at the top of the ladder is plunging earthward. To solve this problem, we begin by constructing the more propitious Figure 3.82 in which the ladder still has length 10 but x and y are variables which (unlike 8 and 6) can have different values at different times t when the ladder is skidding. The variables x and y are related by the formula

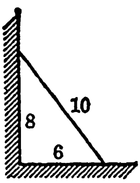


Figure 3.81

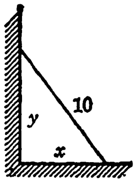


Figure 3.82

$$x^2 + y^2 = 100. \tag{3.83}$$

In order to obtain a relation involving dx/dt (the rate at which the bottom of the ladder is moving away from the wall) and dy/dt (the rate at which our man is moving *upward*), we need the fundamental idea that we should consider x and y to be functions of t and differentiate with respect to t. Equating the derivatives of the members of (3.83) and dividing by 2 gives the formula

$$x\frac{dx}{dt} + y\frac{dy}{dt} = 0 \tag{3.84}$$

which relates the related rates dx/dt and dy/dt. Putting $x = 6$, $y = 8$, and $dx/dt = 2$ shows that $dy/dt = -\frac{3}{2}$. This shows that our poor man is *rising* $-\frac{3}{2}$ units per second and is therefore *falling* $\frac{3}{2}$ units per second, and our problem is solved. It can be insisted that our solution of the problem would have been more easily understood if we had used the more elaborate symbols $x(t)$ or $f_1(t)$ and $y(t)$ or $f_2(t)$ instead of x and y to denote distances. Thus we could have written

$$[x(t)]^2 + [y(t)]^2 = 100, \qquad x(t)x'(t) + y(t)y'(t) = 0$$

or

$$[f_1(t)]^2 + [f_2(t)]^2 = 100, \qquad f_1(t)f_1'(t) + f_2(t)f_2'(t) = 0$$

instead of (3.83) and (3.84). One who wishes to do so may insist that (3.83) and (3.84) abbreviate more meaningful formulas just as the symbols AA and AAA abbreviate Alcoholics Anonymous and American Automobile Association. It is, however, required that we learn the abbreviations to expedite our work and to enable us to understand others who use the abbreviations.

When we are interested in problems involving rates of change of the volume V and the radius r of a sphere, we start with the formula

$$V = \tfrac{4}{3}\pi r^3 \quad \text{or} \quad V(t) = \tfrac{4}{3}\pi[r(t)]^3. \tag{3.85}$$

Supposing that V and r are differentiable functions of time t, we can differentiate to obtain

$$\frac{dV}{dt} = 4\pi r^2 \frac{dr}{dt}. \tag{3.851}$$

When numerical values are assigned to two of the three quantities r, dr/dt, dV/dt, we can solve (3.851) for the remaining quantity.

We need very little information about the external world to appreciate the idea that if an appropriate piston is pushed into a closed cylinder containing a gas, then the volume V of the confined gas will decrease and the pressure p exerted by the confined gas will increase. In appropriate circumstances, calculations can be based upon the formula

$$pV = c, \tag{3.86}$$

where p and V are differentiable functions of t and c is a constant. Differentiation with respect to t gives the formula

$$p\frac{dV}{dt} + V\frac{dp}{dt} = 0, \tag{3.861}$$

which involves four numbers. When three of these numbers are known, we can calculate the remaining one.

If a particle is moving along the graph of the equation $y = x^2$ in such a way that its coordinates x, y are differentiable functions of t, then

$$\frac{dy}{dt} = 2x\frac{dx}{dt}. \tag{3.87}$$

When two of the three numbers in this formula are known, we can calculate the remaining one. Some more or less instructive problems involving such motions require use of the formula

$$|\mathbf{v}| = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \tag{3.871}$$

for the speed of the particle. When motions in E_3 are to be investigated, it may be advantageous to use the vector formulas

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{3.88}$$

$$\mathbf{v} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k} \tag{3.881}$$

in which x, y, z represent the coordinates of the particle at time t. Of course, this motion reduces to motion in the xy plane when $z = 0$ for each t.

Problems 3.89

1 As in Figure 3.891, a rope 13 feet long extends from a boat to a point on a dock 5 feet higher. A man on the dock is pulling rope in at the rate of 72 feet per minute. How fast is the boat moving? *Ans.:* 78 feet per minute.

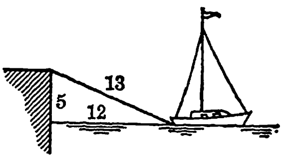


Figure 3.891

2 A light atop a pole is H feet above a level street. A man h feet tall walks steadily, F feet per second, along a line leading away from the base of the pole. At what rate is the tip of his shadow moving when he is x feet from the pole? *Ans.:* $\dfrac{HF}{H-h}$ feet per second.

3 We may be short on information about formation of raindrops in clouds, but we can study the growth of a spherical drop during the part of its development when, for some constant k, the rate, in cubic centimeters per second, at which it collects water is the product of k and the area of its surface. At what rate does the radius increase? *Ans.:* k centimeters per second.

4 It is observed that the radii of volatile mothballs decrease at the rate of 0.5 centimeter per year. Find the rate at which mothballstuff is evaporating from a collection of 100 mothballs of radius 0.6 centimeter. *Ans.:* About 226 cubic centimeters per year or about 0.62 cubic centimeter per day.

5 Sand is falling at the rate of 2 cubic feet per minute upon the tip of a conical sandpile which maintains the form of a right circular cone the height of which is always equal to the radius of the base. Sketch a figure and calculate the rate at which the height is increasing when the height is 6 feet. *Ans.:* $1/18\pi$ feet per minute.

6 Thread is being unwound at the rate of A centimeters per second from an ordinary circular cylindrical spool of radius R centimeters. The unwound part of the thread has length s and is stretched into a line segment TE tangent to the spool at the point T. Find the rate of increase of the distance from the axis of the spool to the end E of the thread. *Hint:* It is not necessary to make an extensive study of the path traversed by the end E; it is sufficient to construct and use an appropriate right triangle *Ans.:* $As/\sqrt{R^2 + s^2}$ centimeters per second.

7 A particle is moving with constant speed k along the graph of $y = \sin x$ in such a way that its x coordinate is always increasing. Derive the formulas

$$\frac{dy}{dt} = \cos x \frac{dx}{dt}, \qquad \frac{dx}{dt} = \frac{k}{\sqrt{1 + \cos^2 x}}, \qquad \frac{dy}{dt} = \frac{k \cos x}{\sqrt{1 + \cos^2 x}}$$

involving the scalar components of the velocity. Show also that

$$\frac{d^2x}{dt^2} = \frac{k^2 \sin x \cos x}{(1 + \cos^2 x)^2}.$$

8 A particle of mass m starts from rest (that is, starts with speed 0) at the point (x_0,y_0) of Figure 3.892 and, with the earth's gravitational field pulling it downward, slides without friction on the graph of the equation $y = x^2$. We grasp an opportunity to see how basic scientific concepts can be employed to obtain information about the motion of the particle. When $0 \leqq y < y_0$ and the particle is at the point (x,y), the loss in potential energy is $mg(y_0 - y)$ and the gain in kinetic energy is $\frac{1}{2}m|\mathbf{v}|^2$, so

$$|\mathbf{v}|^2 = 2g(y_0 - y).$$

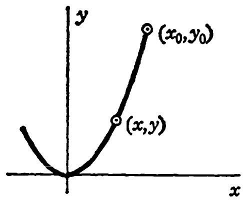


Figure 3.892

With the aid of this information, derive the formulas

$$\left(\frac{dx}{dt}\right)^2 = 2g\frac{y_0 - y}{1 + 4x^2}, \qquad \left(\frac{dy}{dt}\right)^2 = 8g\frac{x^2(y_0 - y)}{1 + 4x^2}$$

which determine (except for algebraic sign) the horizontal and vertical scalar components of the velocity of the particle when it is at the point (x,y). *Remark:* There is a reason why the formulas refuse to tell the signs of dx/dt and dy/dt. As time passes, the particle oscillates to and fro over an arc of the parabola in such a way that the scalar components of the velocity are sometimes positive and sometimes negative.

9 Figure 3.893 shows a connecting rod of length b which earns its name by connecting a piston (which is free to move to and fro in a cylinder) to a point P on a crankshaft which is free to rotate in a circle of radius a having its center at O. We should not be too busy to observe that b exceeds $2a$ in ordinary engines and pumps. Obtain a formula relating dx/dt, the scalar velocity of the piston, to $d\theta/dt$, the angular speed of the crankshaft. *Hint:* Use the law of cosines in the form

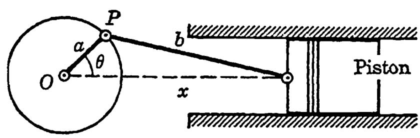


Figure 3.893

$$b^2 = a^2 + x^2 - 2ax \cos \theta.$$

Ans.: $\dfrac{dx}{dt} = -\dfrac{ax \sin \theta}{x - a \cos \theta}\dfrac{d\theta}{dt}.$

10 Let ϕ be the angle between the lines of Figure 3.893 that have lengths b and x. Show that

$$\frac{d\phi}{dt} = \frac{a \cos \theta}{b \cos \phi} \frac{d\theta}{dt}.$$

Hint: At each time the numbers $a \sin \theta$ and $b \sin \phi$ are equal to each other because they are both equal to the distance (or the negative of the distance) from P to the line having length x. Thus we use a slight extension of the trigonometric law of sines.

11 A circle of radius R has its center at the point $(0,R)$ of an xy plane. A motorcycle is racing at night along the circle in the first quadrant toward the origin. When the motorcycle is at distance s (measured along the circle) from the origin, its headlight illuminates a spot at the point $(x,0)$ on the x axis. Show how the rate at which the spot is moving depends upon R, s, and the rate at which the motorcycle is moving. *Outline of solution:* Study of an appropriate figure can lead us to the first and then the second of the formulas

$$\frac{x}{R} = \tan \phi, \qquad x = R \tan \frac{s}{2R}, \tag{1}$$

where ϕ is the angle between the vectors running from the center of the circle to the origin and the point $(x,0)$, and

$$\phi = \frac{\text{length of arc}}{\text{radius}} = \frac{s/2}{R} = \frac{s}{2R}. \tag{2}$$

Since

$$\frac{d}{dt} \tan \phi = \frac{d}{dt} \frac{\sin \phi}{\cos \phi} = \frac{\cos^2 \phi + \sin^2 \phi}{\cos^2 \phi} \frac{d\phi}{dt} = \sec^2 \phi \frac{d\phi}{dt}, \tag{3}$$

differentiating with respect to t gives the answer

$$\frac{dx}{dt} = \frac{1}{2} \sec^2 \frac{s}{2R} \frac{ds}{dt}. \tag{4}$$

Remark: In the context of the motorcycle problem, ds/dt and dx/dt are both negative. The answer will give very interesting information to those who study it. Those who do not use the metric system for measuring distances and speeds may observe that if s and R are numbers of feet and ds/dt is a number of miles per hour (or furlongs per fortnight), then dx/dt will be a number of miles per hour (or furlongs per fortnight).

12 A man or a boy or a particle is, for reasons that are sometimes explained, at the point $P_1(x_1,y_1,z_1)$ and is moving with speed q_1 in the direction of the unit vector $a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$. A second animate or inanimate object is at the point $P_2(x_2,y_2,z_2)$ and is moving or being moved with speed q_2 in the direction of the unit vector $a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}$. We are required to find the rate at which the distance between the two objects is changing. Do it. *Solution:* Let the bodies be at the points $P(x,y,z)$ and $Q(u,v,w)$ at time t. The distance between the bodies at time t is then the positive number s for which

$$s^2 = (u - x)^2 + (v - y)^2 + (w - z)^2. \tag{1}$$

Even though we have written (1) in such a way that t does not appear, we rise to the occasion and differentiate with respect to t to obtain

$$(2)\quad s\frac{ds}{dt} = (u - x)\left(\frac{du}{dt} - \frac{dx}{dt}\right) + (v - y)\left(\frac{dv}{dt} - \frac{dy}{dt}\right) + (w - z)\left(\frac{dw}{dt} - \frac{dz}{dt}\right).$$

To obtain the answer, we determine s from (1) and then ds/dt from (2) when

$$(3)\qquad u = x_2, \quad v = y_2, \quad w = z_2, \quad x = x_1, \quad y = y_1, \quad z = z_1$$

$$(4)\qquad \frac{du}{dt} = q_2a_2, \quad \frac{dv}{dt} = q_2b_2, \quad \frac{dw}{dt} = q_2c_2, \quad \frac{dx}{dt} = q_1a_1, \quad \frac{dy}{dt} = q_1b_1, \quad \frac{dz}{dt} = q_1c_1.$$

Remark: The formula (2) is trying to tell us something. The vector $\overrightarrow{PQ}$ in the formula

$$(5)\qquad \overrightarrow{PQ} = (u - x)\mathbf{i} + (v - y)\mathbf{j} + (w - z)\mathbf{k}$$

is the displacement of Q relative to P. The velocity $\mathbf{v}$ of Q relative to P is obtained by differentiating this with respect to t. Thus

$$(6)\qquad \mathbf{v} = \left(\frac{du}{dt} - \frac{dx}{dt}\right)\mathbf{i} + \left(\frac{dv}{dt} - \frac{dy}{dt}\right)\mathbf{j} + \left(\frac{dw}{dt} - \frac{dz}{dt}\right)\mathbf{k}.$$

Hence (2) tells us that

$$(7)\qquad s\frac{ds}{dt} = \overrightarrow{PQ}\cdot\mathbf{v},$$

that is, the left side is the scalar product of the displacement vector and the velocity of Q relative to P. Since $|\overrightarrow{PQ}| = s$, the vector $\overrightarrow{PQ}/s$ is a unit vector in the direction of $\overrightarrow{PQ}$. The relations (2) and (7) therefore bear a simple message. They tell us that *the rate of change of the distance between two bodies is the scalar component of the relative velocity of the bodies in the direction of the line joining the bodies.* This agrees with and is perhaps even a consequence of the fact that if one body moves in a circle having its center at the other body, then the distance between the bodies is always the radius of the circle.

13 Kitty was riding a horse on a merry-go-round of radius R. When she was south of the center pole and going east with speed s, she exuberantly threw a ball toward the pole. Kitty expected to hit the pole, but unfortunately Chester was riding gallantly ahead of her and the ball hit him on the chin when he was east of the center. Sketch a figure showing the east and north vector components of the velocity (relative to terra firma) of the ball and mark the place where Chester was sitting when Kitty threw the ball.

3.9 *Increments and differentials*

As we become educated, we pick up assorted ideas akin to the idea that we never try to find the weight (1 avoirdupois grain or 0.0648 gram) of a kernel of medieval wheat by finding the weight of a truckload of the stuff and subtracting the weight of the decreased load resulting from removal of the kernel. The difficulty lies in the fact that small relative errors in the large weights can produce huge relative errors in the small difference. The obvious way to find the order of magnitude of the number Δy defined by

$$\Delta y = f(x + \Delta x) - f(x) \tag{3.91}$$

is to calculate the terms on the right side and subtract. However, when the numbers Δx and Δy are very small in comparison to x and $f(x)$ and $f(x + \Delta x)$, this calculation can be thoroughly tedious and impractical. The following alternative way of estimating Δy is very often used. If f and x are such that $f'(x)$ exists, we can define $\phi(x,\Delta x)$ by the formula

$$\frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x) + \phi(x,\Delta x) \tag{3.92}$$

and conclude that

$$\lim_{\Delta x \to 0} \phi(x,\Delta x) = 0. \tag{3.921}$$

Multiplying (3.92) by Δx gives the formula

$$\Delta y = f'(x)\,\Delta x + \phi(x,\Delta x)\,\Delta x, \tag{3.922}$$

which separates Δy into the sum of two "parts." In case $f'(x) \neq 0$ and Δx is near 0, the "part" $\phi(x,\Delta x)\,\Delta x$ is small in comparison to the "part" $f'(x)\,\Delta x$ and the number $f'(x)\,\Delta x$ is the "principal part" of Δy. Therefore, when $f'(x) \neq 0$, we can write the formula

$$\Delta y \sim f'(x)\,\Delta x \tag{3.93}$$

to mean that the numbers Δy and $f'(x)\,\Delta x$ have the same order of magnitude when Δx is near 0, that is,

$$\lim_{\Delta x \to 0} \frac{\Delta y}{f'(x)\,\Delta x} = 1.$$

In any case, it is a common practice to use the number $f'(x)\,\Delta x$ as an approximation to Δy when f, x, and Δx are given and $|\Delta x|$ is judged to be small enough to make the approximation useful. In some cases it is equally useful to use the number $\Delta y/f'(x)$ as an approximation to Δx when f, x, and Δy are given.

We have seen that, in appropriate circumstances, the numbers Δy (an *increment* of y) and Δx (an *increment* of x) are such that Δy and $f'(x)\,\Delta x$ are nearly equal in the sense that their ratio is nearly 1. While it may be difficult to see why we should become excited about the matter, it is worthwhile to think about and even use pairs of numbers dy and dx for which dy is exactly (not merely approximately) equal to $f'(x)\,dx$ so that

$$dy = f'(x)\,dx. \tag{3.94}$$

Such numbers dy and dx are called *differentials*, and some useful observations can be made. When f and x are such that $f'(x)$ exists, we can let dx be any number that pleases us and calculate dy, and, provided $f'(x) \neq 0$, we can also let dy be any number we please and calculate dx. Our interest in differentials can start to develop when we see that, as Figure 3.95 indicates, the point $(x + dx, y + dy)$ must lie on the line tangent to the graph of f at the point (x,y). This is true because $f'(x)$ is the slope of the tangent and (3.94) implies that dy/dx is this slope when $dx \neq 0$. Since increments Δy and Δx are numbers such that the point $(x + \Delta x, y + \Delta y)$ lies on the graph of f and differentials dx and dy are numbers such that the point $(x + dx, y + dy)$ lies on the tangent to the graph at the point (x,y), it is clear that both of the two equalities $dx = \Delta x$ and $dy = \Delta y$ can be true only when the two points $(x + \Delta x, y + \Delta y)$ and $(x + dx, y + dy)$ coincide at a point of intersection of the graph and the tangent.

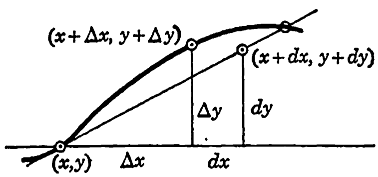


Figure 3.95

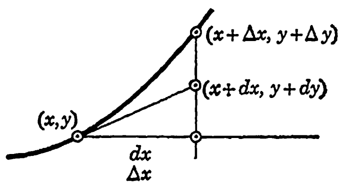


Figure 3.951

It is particularly easy to produce the differential formula (3.94) when we use the Leibniz notation for derivatives. The calculation in the left-hand column

$$\begin{aligned} y &= f(x) & \qquad y &= x^2 \\ \frac{dy}{dx} &= f'(x) & \qquad \frac{dy}{dx} &= 2x \\ dy &= f'(x)\,dx & \qquad dy &= 2x\,dx \end{aligned}$$

produces the formula whenever f and x are such that $f'(x)$ exists, and the calculation in the second column shows how things go when $f(x) = x^2$. This circumstance emphasizes the fact that, when f and x are such that $f'(x)$ exists and $\frac{dy}{dx}$ is as usual the derivative $f'(x)$, the differentials dy and dx are defined in such a way that the quotient $\frac{dy}{dx}$ (dy divided by dx) is the same as the derivative $\frac{dy}{dx}$ when $dx \neq 0$. To find the differential formula relating dy and dx when f and x are given, it is therefore sufficient to set $y = f(x)$, differentiate to obtain the formula

$$\frac{dy}{dx} = f'(x), \tag{3.952}$$

and then multiply by dx. A little experience with these things makes us realize that if $y = \sin x$, we can write the formula $dy = \cos x\, dx$ without bothering to write the intermediate step $dy/dx = \cos x$.

In most situations where increments Δy, Δx and differentials dy, dx simultaneously appear, it is convenient to suppose that $dx = \Delta x$. In such cases, glances at figures more or less like Figure 3.951 can fortify the idea that Δy can easily be twice dy when Δx and dx are equal but not small, but that Δy and dy must have the same order of magnitude when $f'(x) \neq 0$ and the equal numbers Δx and dx are near 0. Thus a useful cookbook *modus operandi* runs as follows.

When f, x, and Δx are given such that $f'(x) \neq 0$ and we want an approximation to the number Δy defined by

$$(3.96) \qquad \Delta y = f(x + \Delta x) - f(x),$$

we put $\Delta x = dx$, calculate the number (or differential) dy defined by

$$(3.961) \qquad dy = f'(x)\, dx,$$

and use dy as an approximation to Δy.

It is instructive to consider a thoroughly simple example in which all of the details are easily understood. Letting x and Δx be numbers which could be 38.27 and 0.05, we can determine the increment Δy in the area of a square when the lengths of its sides are increased from x to $x + \Delta x$. Letting $y = x^2$ and $y + \Delta y = (x + \Delta x)^2$, we find that

$$(3.97) \qquad \Delta y = (x + \Delta x)^2 - x^2 = 2x\,\Delta x + \Delta x^2.$$

When $\Delta x > 0$, the number $2x\,\Delta x$ is the sum of the areas of the two rectangles of Figure 3.971 which have dimensions x and Δx. The number

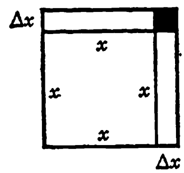


Figure 3.971

Δx^2 is the area of the smaller square in the upper right-hand corner of the figure. The differential dy is, when $dx = \Delta x$,

$$(3.972) \qquad dy = 2x\, dx = 2x\,\Delta x,$$

and it is easily seen that this is a good approximation to Δy when Δx is small in comparison to x.

It is sometimes convenient to solve problems more or less like the following one in order to determine the accuracy of measurements required to produce required accuracy of results computed from the measurements.

If we measure the edges of a cube and decide that, subject to errors in measurement, each side has length x, we conclude that, subject to consequences of errors in measurement, the volume V of the cube is x^3. If the edges have exact lengths $x + \Delta x$, then the exact volume is $(x + \Delta x)^3$, or $V + \Delta V$, and the number ΔV is the error in V produced by the error Δx in x. In quantitative treatments of this matter, we let $V = x^3$ and use the differential dV defined by

$$dV = 3x^2\,dx \tag{3.98}$$

as an approximation to ΔV. In some practical situations, it is reasonable to assume that (for some positive constant p that might be $\frac{1}{2}$ or 2 or 10), the error dx in x has a magnitude not exceeding p per cent of x. This means that

$$|dx| \leqq \frac{p}{100}|x|. \tag{3.981}$$

With this assumption, we find from (3.981) that

$$|dV| \leqq 3x^2 \frac{p}{100} x = \frac{3p}{100} x^3 = \frac{3p}{100} V. \tag{3.982}$$

This leads us to the idea that if we measure the length of the edge of a cube with an error not exceeding p per cent, the resulting error in the computed volume will not exceed $3p$ per cent.

Problems 3.99

1 Find the increment ΔA and the differential dA of area produced when a circular disk of radius r is expanded or contracted to a circular disk of radius $r + h$. *Ans.:* $\Delta A = \pi(2hr + h^2)$ and $dA = 2\pi rh$. *Remark:* We have another opportunity to try to understand a formula. Sketch a figure in which $|h|$ is small in comparison to r and observe that the difference of the two disks is a circular ring of thickness $|h|$. Since the inner (or outer) boundary of this ring has length $2\pi r$, it is not surprising that the area of the ring is approximately $2\pi r|h|$.

2 The area A of a sphere of radius r in E_3 is $4\pi r^2$; this should seem to be reasonable because the area of a hemisphere should be about twice the area of an equatorial disk. The volume V of the spherical ball bounded by this sphere is $\frac{4}{3}\pi r^3$. Find the increment ΔV and the differential dV of volume produced when the radius changes from r to $r + h$. Show that the formula for dV can be put in the form $dV = Ah$ and try to see a geometrical reason why Ah should be a good approximation to ΔV when $|h|$ is small.

3 Use differentials to obtain an approximation to the number of cubic centimeters of chromium plate that must be applied to the lateral surface of a circular cylindrical rod 30 centimeters long to increase its radius from 2.34 centimeters to 2.35 centimeters. *Ans.:* About 4.4 cubic centimeters.

4 Suppose that x and y are differentiable functions of t such that

$$x^2 + y^2 = 1. \tag{1}$$

Show that differentiating with respect to t and multiplying by dt gives the formula

$$x\,dx + y\,dy = 0. \tag{2}$$

Remark: In case $t = x$, we can divide (2) by dx and recover the first of the formulas

$$x + y\frac{dy}{dx} = 0, \qquad x\frac{dx}{dy} + y = 0, \tag{3}$$

which is valid when y is a differentiable function of x for which (1) holds. In case $t = y$, we can divide (2) by dy and recover the second formula in (3), which is valid when x is a differentiable function of y for which (1) holds.

5 Supposing that n is a constant and x is positive, observe that the first of the formulas

$$y = x^n, \qquad \log y = n \log x \tag{1}$$

is equivalent to the second. Use these formulas to obtain

$$dy = nx^{n-1}\,dx \tag{2}$$

and

$$\frac{1}{y}\,dy = n\frac{1}{x}\,dx. \tag{3}$$

Use (2) to show that if $|dx| \leqq (p/100)x$, then $|dy| \leqq (|n|p/100)y$. Then use (3) to obtain the same result.

6 Gradual changes in tensions or compressions or temperatures can produce gradual changes in the lengths of iron rods that form a triangle such as that shown in Figure 3.991. Therefore, because the law of cosines must hold at each time t, it makes sense to suppose that we have four positive differentiable functions of t such that

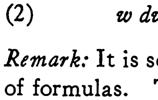


Figure 3.991

$$w^2 = u^2 + v^2 - 2uv\cos\theta. \tag{1}$$

Equate the derivatives with respect to t of the members of (1) and multiply by dt to obtain the differential formula

$$w\,dw = (u - v\cos\theta)\,du + (v - u\cos\theta)\,dv + uv\sin\theta\,d\theta. \tag{2}$$

Remark: It is sometimes both possible and unwise to underestimate potentialities of formulas. The formulas (1) and (2) contain eight numbers u, v, w, θ, du, dv, dw, and $d\theta$. There are many situations in which some information about some of these numbers is known and the two formulas can be used to eke out more information. Some problems are much more recondite than the one solved by finding w from (1) and then finding dw from (2) when the values of the other six numbers are known.

7 Supposing that T, L, and g are positive, observe that the first of the formulas

$$T = 2\pi\sqrt{\frac{L}{g}}, \qquad T^2 = 4\pi^2\frac{L}{g}, \qquad \log T = \log 2\pi + \frac{1}{2}\log L - \frac{1}{2}\log g \tag{1}$$

is equivalent to the other two. Use these formulas to obtain

$$dT = \pi\left(\frac{L}{g}\right)^{-\frac{1}{2}}\frac{g\,dL - L\,dg}{g^2} \tag{2}$$

$$2T\,dT = 4\pi^2\frac{g\,dL - L\,dg}{g^2} \tag{3}$$

$$\frac{1}{T}\,dT = \frac{1}{2}\frac{1}{L}\,dL - \frac{1}{2}\frac{1}{g}\,dg. \tag{4}$$

Making a suitable application of the fact that $|A| \leqq |B| + |C|$ whenever A, B, and C are numbers (not necessarily positive) for which $A = B + C$ or $A = B - C$, use (2) to show that if $|dL| \leqq (p/100)L$ and $|dg| \leqq (q/100)g$, then $|dT| \leqq [\frac{1}{2}(p + q)/100]T$. Repeat the process by use of (3) and (4). *Remark:* The first of the formulas (1) is a standard formula for the period T (a number of seconds) of a pendulum of length L which oscillates in a world where the scalar acceleration of gravity is g. Our result shows that if errors in measurement of L and g do not exceed p and q per cent, respectively, then the error in T will not exceed $\frac{1}{2}(p + q)$ per cent.

8 A pendulum clock gains 3 minutes in 24 hours. By what per cent should the pendulum be lengthened? *Ans.:* 0.42 per cent.

9 Under appropriate conditions the pressure p and the volume V of confined gas satisfy the relation

$$pV^\gamma = C, \tag{*}$$

where γ (gamma) and C are constants that depend upon the gas and the conditions. Obtain the formula

$$\frac{dp}{p} + \frac{\gamma\,dV}{V} = 0 \tag{**}$$

in two different ways. First, differentiate the members of (*) as they stand. Then operate upon the equation obtained by taking logarithms of the members of (*). *Remark:* It is so often desirable to take logarithms before differentiating that the process is named *logarithmic differentiation.* The derivative of the logarithm of a function is called the *logarithmic derivative* of the function.

10 Apply the procedure of the preceding problem to the relation

$$pV = nRT,$$

in which n is the number of gram-moles of a gas, R is a universal proportionality constant known as "the gas constant," and T is the absolute, or Kelvin, temperature. It is now supposed that p, V, and T are all functions of t and the relation

$$\frac{dp}{p} + \frac{dV}{V} = \frac{dT}{T}$$

is to be obtained.

11 For dense projectiles fired short distances over a horizontal plane, the range R is calculated from the formula

$$R = \frac{v_0^2}{g}\sin 2\alpha,$$

where g is the (scalar) acceleration of gravity, v_0 is the initial speed, and α is the angle of elevation of the gun so $0 < \alpha < \pi/2$. Find a formula in which we can put estimates of errors in g, v_0, and α to obtain an estimate of the resulting error in R. *Hint:* Use logarithms. *Ans.:*

$$\left|\frac{dR}{R}\right| \leqq 2\left|\frac{dv_0}{v_0}\right| + \left|\frac{dg}{g}\right| + \left|\frac{2\alpha\cos 2\alpha}{\sin 2\alpha}\right|\left|\frac{d\alpha}{\alpha}\right|.$$

Remark: When α is not too close to $\pi/2$, the factor multiplying $|d\alpha/\alpha|$ has the order of magnitude of $|\cos 2\alpha|$. When α is near $\pi/2$, the factor is very large. We are now able to enlighten rabbit hunters when they ask us why they are unlikely to hit their targets when they shoot almost straight up.

12 Find the maximum possible percentage of error in the computed estimate of the volume of a cone that can be caused by errors not exceeding p per cent and q per cent in measurements of the height and the base radius of the cone. *Ans.:* $p + 2q$.

13 When three resistors having resistances r_1, r_2, r_3 are connected in parallel, the resulting resistance R is determined by the formula

$$\frac{1}{R} = \frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3}.$$

With the aid of the fact that resistances are always positive, prove that if no error in a resistor exceeds p per cent, then the error in R produced by these errors cannot exceed p per cent. *Remark:* This conclusion really means something to those who design the mazes hidden in our television sets. Problems involving silver bands and gold bands and tolerances (percentages of error) cannot be ignored. Engineers do not like to behave like rabbit hunters who shoot almost straight up.

14 Show that the conclusion of the preceding problem is violently false if the numbers r_1, r_2, r_3 are not resistances of resistors but are numbers of which some can be positive and some can be negative.

15 Let $\mathbf{r}$ be a differentiable function of t. For example, we may have

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}, \tag{1}$$

where all of the functions are differentiable and the vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are the unit orthonormal vectors of Section 2.2. By a definition analogous to the one involving scalar differentials, a vector $d\mathbf{r}$ and a scalar dt constitute a pair of differentials if

$$d\mathbf{r} = \mathbf{r}'(t)\,dt. \tag{2}$$

With or without using the rectangular representation (1) and the fact that

$$|\mathbf{r}(t)|^2 = [x(t)]^2 + [y(t)]^2 + [z(t)]^2,$$

prove that

$$d|\mathbf{r}(t)|^2 = 2\mathbf{r}(t)\cdot d\mathbf{r}(t).$$

16 The specific heat of a substance is sometimes said to be the number of calories of heat required to raise the temperature of 1 gram of the stuff 1 degree centigrade. This definition is sometimes useful but, because substances have different specific heats at different temperatures, the following definition is much better. *The specific heat σ (sigma) of a substance at temperature x is $Q'(x)$, where $Q(x)$ is the number of calories of heat required to raise (or lower) the temperature of* 1 *gram of the stuff from* 0°C *to* x°C. For study of this matter, let σ^* be the specific heat calculated from the first definition so that

$$(1) \qquad \sigma^* = \frac{Q(x+1) - Q(x)}{1}, \qquad \sigma = Q'(x) = \lim_{h \to 0} \frac{Q(x+h) - Q(x)}{h}.$$

We can see one of the reasons why knowledge of calculus is needed for study of physical chemistry when we see that σ^* is a difference quotient and σ is a derivative. Since the graph of Q is never (or not ordinarily) a line, σ^* and σ are usually different. The schematic Figure 3.992 illustrates one situation. As is easily

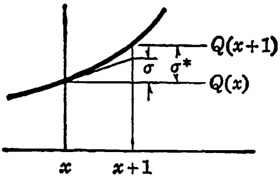


Figure 3.992

imagined, there are situations in which the graph of Q is "almost straight" over the interval from x to $x + 1$ and σ^* is a "good approximation" to σ. On the other hand, σ^* can be dust and ashes when the interval from x to $x + 1$ straddles a temperature at which a substance changes from a solid state to a liquid or from a liquid to a gas.

17 There are reasons why we should conclude with a historical remark. In the good old days when the "doctrine of limits" was based upon visions of galloping numbers and the "infinitely small infinitesimals" were considered to be almost the most wonderful products of human thought, differentials were considered to be the most wonderful. Differentials were the important things, and the things that we now call derivatives were merely the "differential coefficients" appearing in formulas like $dy = f'(x)\,dx$ or $dy = 2x\,dx$. Thus differentials have their origin in old mathematics; it was the fashion to consider them to be "infinitely small" but not quite zero. When at long last the concept of the "infinitely small" was becoming obsolete, attempts were made to salvage differentials by promoting the idea that they really are not ordinary numbers at all but are numbers that are in the process of approaching zero.† So far as this course is concerned, the details of this remark are unimportant. We should, however, know that differentials have a long and checkered history and that we may expect to encounter some quite strange concepts as we get around in the world.

† For those who have not peered into old books and consider this to be too incredible to be true, we quote three passages from W. E. Byerly, "Elements of the Differential Calculus," Ginn and Heath, Boston, 1879. Page 149 tells us that "*An infinitesimal* or *infinitely small* quantity is *a variable which is supposed to decrease indefinitely;* in other words, *it is a variable which approaches the limit zero.*" Page 185 tells us that "It is to be noted that a *differential is an infinitesimal, and that it differs from an infinitesimal increment by an infinitesimal of a higher order.*" Page 186 tells us that "*there is a practical advantage . . . in regarding the differential as the main thing, and looking at the derivative as the quotient of two differentials.*"

4 *Integrals*

4.1 Indefinite integrals There are about as many different types of integrals in mathematics as there are elements in chemistry, but only a few of them occur in first courses in calculus. This chapter introduces basic ideas about two kinds of integrals. These ideas may not be coming too soon to meet the needs of students taking other courses in which mathematics appears. In this section, and in some other places where the deviation from complete linguistic rectitude does not create deceptive statements, we sometimes refer to "the function $f(x)$" or to "the function having values $f(x)$" instead of to the function f which, for each x in some interval, has the value $f(x)$.

It is very often true that we have a given function $f(x)$ and we are interested in those functions $F(x)$ or y or $y(x)$ (if any) for which

$$F'(x) = f(x) \qquad \text{or} \qquad \frac{dy}{dx} = f(x) \tag{4.11}$$

when x lies in some interval. Before discussing this situation, we introduce notation that is universally used. In case $F(x)$ or y is a function for which (4.11) holds, we represent it by the ingenious symbol in the formula

$$F(x) = \int f(x)\,dx \qquad \text{or} \qquad y = \int f(x)\,dx. \tag{4.111}$$

The second equation is read "y equals an integral of eff of x dee x." We should all know that it can be read "y equals an indefinite integral of eff of x dee x," or "y equals a function whose derivative with respect to x is f" or "y equals an antiderivative with respect to x of f," but simplicity always prevails and we read what we see and say what is to be written. The *integral sign* $\int$ is an elongated S, the $f(x)$ is called the *integrand*, and the dx tells us that derivatives with respect to x are involved.† This matter turns out to be so important that we must continually remember the following definition.

Definition 4.12 *The indefinite integral in the formula*

$$y = \int f(x)\,dx = \phi(x) \tag{4.121}$$

is (if it exists) a function of x whose derivative is the integrand $f(x)$; in other words, the formula

$$\frac{dy}{dx} = f(x) = \phi'(x) \tag{4.122}$$

and the formula (4.121) *are both true or both false.*

For an example, let us see what we know and can learn about the functions y for which the equivalent formulas

$$\frac{dy}{dx} = 2x, \qquad y = \int 2x\,dx \tag{4.123}$$

are valid. We may remember that we differentiated x^2 and got $2x$. Hence a function y for which the formulas are valid might be x^2 but it does not have to be because y might be $x^2 + 1$ or $x^2 - 5$ or $x^2 + 416$. It can be proved that a given function y will satisfy the equivalent formulas (4.123) if and only if there is a constant c such that $y = x^2 + c$. Thus

$$\int 2x\,dx = x^2 + c. \tag{4.124}$$

To be precise about this matter, we state the following theorem which will be proved later in a remark following the proof of Theorem 5.57.

† Perhaps it should be emphasized at once that the dx in the symbol is not a number. If we resist temptations to jump to the conclusion that the dx and the crossbar on the f and the integral sign are numbers, we overcome a difficulty that makes some people feel that the good old symbol should be abandoned in favor of another which provides fewer temptations.

Theorem 4.13 *If two functions y and F have the same derivative over an interval, then there is a constant c such that*

$$y(x) = F(x) + c$$

for each x in the interval.

Considerable information is packed into the little formula

$$y = \int f(x)\, dx = F(x) + c \tag{4.14}$$

in which $F(x)$ is any one particular function whose derivative with respect to x is, over some interval, the integrand $f(x)$ of the integral It tells us that, whatever the value of the constant c may be, $F(x) + c$ is a function y whose derivative with respect to x is the integrand $f(x)$. Moreover, it tells us that if y is a function whose derivative is the integrand, then there must be a constant c for which $y = F(x) + c$. The full meaning of the assertion (4.14) has been stated, and this is what is important. Simply because we must converse with our fellow men and must read scientific writings, we must join our fellow men in learning some terminology. The constant c in (4.14) is a "constant of integration" and the poor fellow is sometimes said to be "arbitrary." The integral is called an "indefinite integral" to distinguish it from other types of integrals that are sometimes called "definite integrals." This rather strange terminology will not injure us if we do not allow it to interfere with our understanding of the meaning of (4.14). The assertion "each indefinite integral of f is the sum of a particular indefinite integral and a constant of integration" sounds weird but is true. The "meaning" of the word "indefinite" can be understood if we realize that when c is a constant, say 416, $F(x) + c$ is an "indefinite integral" of $f(x)$ just as the mayor of Chicago is an "indefinite citizen" of Chicago.

In case $F'(x) = f(x)$, $G'(x) = g(x)$, and a, b are constants, differentiation formulas show that

$$\int [af(x) + bg(x)]\, dx = aF(x) + bG(x) + c \tag{4.15}$$

and

$$\int [af(x) + bg(x)]\, dx = a\int f(x)\, dx + b\int g(x)\, dx. \tag{4.16}$$

These formulas tell us that "integrals of sums are sums of integrals" and that "constants can be moved across integral signs." The formulas do not provide justification for moving functions across integral signs; otherwise, we could replace $\neq$ by $=$ in the formula

$$\int f(x)\, dx \neq f(x)\int dx = f(x)(x + c) \tag{4.161}$$

and eliminate all of our troubles.

The following little table gives two versions of each of the five simplest and most useful integration formulas.

$$(4.171) \quad \int x^n\,dx = \frac{x^{n+1}}{n+1} + c \qquad \int u^n \frac{du}{dx}\,dx = \frac{u^{n+1}}{n+1} + c$$

$$(4.172) \quad \int \sin x\,dx = -\cos x + c \qquad \int \sin u \frac{du}{dx}\,dx = -\cos u + c$$

$$(4.173) \quad \int \cos x\,dx = \sin x + c \qquad \int \cos u \frac{du}{dx}\,dx = \sin u + c$$

$$(4.174) \quad \int e^x\,dx = e^x + c \qquad \int e^u \frac{du}{dx}\,dx = e^u + c$$

$$(4.175) \quad \int \frac{1}{x}\,dx = \log|x| + c \qquad \int \frac{1}{u}\frac{du}{dx}\,dx = \log|u| + c$$

In the formulas of the second column, u is supposed to be a differentiable function of x. Subject to the requirement that $n \neq -1$ in (4.171), and that x and u are confined to intervals over which the integrands in (4.171) and (4.175) exist, these formulas are proved by observing that they have the form (4.14) where $F'(x) = f(x)$. We need not learn all of the formulas we see, but the formulas in the above table are used so often that they must be learned.

When the formulas in the column on the right are being used, presence of the factor du/dx must be carefully observed. It is not correct to think of u as being $\sin x$ and to claim that use of (4.171) shows that the members of the formula

$$(4.181) \qquad \int \sin^2 x\,dx \neq \frac{\sin^3 x}{3} + c$$

are equal. We can, however, think of u as being $\sin x$ and read the left member of the formula

$$(4.182) \qquad \int \sin^2 x \cos x\,dx = \frac{\sin^3 x}{3} + c$$

in the form "integral of u to the nth power dee u dee x dee x" and then apply (4.171) to obtain the right member.

It is not correct to claim that the members of the formula

$$(4.183) \qquad \int (5x+7)^2\,dx \neq \frac{(5x+7)^3}{3} + c$$

are equal. We can, however, let I denote the left member, observe that the integrand has the form u^n, where $du/dx = 5$, and write

$$(4.184) \qquad I = \frac{1}{5}\int (5x+7)^2(5)\,dx = \frac{1}{5}\frac{(5x+7)^3}{3} + c.$$

Thorough understanding of this particular example is of utmost importance because it involves an idea that is very often used to overcome a difficulty. In (4.183) we have an integral of the form $\int u^n\,dx$ which does not have the form $\int u^n\,(du/dx)\,dx$. However, du/dx is 5, a constant, so we can insert the factor 5 in the integrand and *compensate* for the deed by inserting the factor $\frac{1}{5}$ before the integral.

To obtain the formula

$$\int \frac{1}{1+x^2}\,2x\,dx = \log\,(1+x^2) + c, \tag{4.185}$$

we read the left side "integral of one over u dee u dee x dee x" and apply (4 175). If the factor 2 had been missing from the integrand in (4.185), it would have been necessary to insert the factor and compensate for the deed. Thus

$$\int \frac{x}{1+x^2}\,dx = \frac{1}{2}\int \frac{1}{1+x^2}\,2x\,dx = \frac{1}{2}\log\,(1+x^2) + c. \tag{4.186}$$

Our very modest table of integrals beginning with (4.171) does not reveal the answer to the question whether there are any functions $F(x)$ for which the formulas

$$F'(x) = \frac{1}{1+x^2}, \qquad \int \frac{1}{1+x^2}\,dx = F(x) + c \tag{4.187}$$

are valid. Many useful purposes are served by this table and the more extensive one appearing opposite the back cover of this book, but one who has solved several of the problems at the end of this section is ready to recognize the fact that there exist much more elaborate tables of integrals. The books of Burington† and Dwight‡ are exceptionally useful examples of books that give hundreds of integration formulas, tables of values of functions, and other mathematical information. It is possible to proceed through our course without using tables other than those on the back cover and facing page of this textbook. However, students who contemplate following educational programs in which mathematics appears are well advised to purchase one of these books (or perhaps another more or less similar one recommended by teachers) and to spend occasional moments (and sometimes hours) inspecting its organization and studying its contents. Ability to understand and use the tables is not inherited but can develop rapidly as more calculus is learned. Experience shows that persons who have completed courses in calculus

† R. S. Burington, "Handbook of Mathematical Tables and Formulas," 3d ed., McGraw-Hill Book Company, Inc., New York, 1948, 296 pages.

‡ Herbert Bristol Dwight, "Tables of Integrals and Other Mathematical Data," 4th ed., The Macmillan Company, New York, 1961, 336 pages.

refer to tables in books of tables in preference to tables in calculus textbooks. Teachers can be particularly helpful when they require that their students purchase identical books of tables and make frequent comments about use of the tables. Sometimes use of a book of tables is permitted in tests and examinations where use of a calculus textbook is forbidden.

Problems 4.19

1 Tell the meaning of $\int f(x)\,dx$. Be prepared to give full information at any time.

2 Show that, when x is properly restricted,

$$(a)\quad \int (2 + 3x + 4x^2)\,dx = 2x + \frac{3x^2}{2} + \frac{4x^3}{3} + c$$

$$(b)\quad \int \frac{1-x^2}{1-x}\,dx = x + \frac{x^2}{2} + c$$

$$(c)\quad \int \left(\sqrt{x} + \frac{1}{\sqrt{x}}\right) dx = \frac{2}{3}(x+3)\sqrt{x} + c$$

$$(d)\quad \int x(1-x)\,dx = \frac{x^2}{2} - \frac{x^3}{3} + c$$

3 Is the formula

$$\int x^3\,dx = x\int x^2\,dx \ (?)$$

true or false?

4 Brevity is sometimes but not always a virtue. It can be claimed that the second formula in (4.171) would be much more easily understood and used if it were written in the form

$$\int [u(x)]^n u'(x)\,dx = \frac{[u(x)]^{n+1}}{n+1} + c.$$

Think about this, and then write the other four formulas in terms of the Newton notation for derivatives. Note that the last formula takes the form

$$\int \frac{u'(x)}{u(x)}\,dx = \log |u(x)| + c.$$

Remark: We can abbreviate (4.171) to the form $\int u^n\,du = u^{n+1}/(n+1) + c$, but for present purposes we can hold the view that further abbreviation of (4.171) is a step in the wrong direction. We need not be in a hurry to join the ranks of gullible people who think that the du appearing in the symbol $\int u\,du$ *is* a number because a correct result is obtained by pretending that du is "the" differential $u'(x)\,dx$ and writing

$$\int u\,du = \int u(x)u'(x)\,dx = \tfrac{1}{2}[u(x)]^2 + c.$$

When we do not use the abbreviation, we do not need to worry about it.

5 Look at the integral

$$\int(1 + 5x)^{1/2}\, dx$$

and tell what must be done to enable us to evaluate the integral by use of the formula involving u^n.

6 Evaluate the integrals

(*a*) $\displaystyle\int (1 - x)^3\, dx$ (*b*) $\displaystyle\int \sin 2x\, dx$ (*c*) $\displaystyle\int \cos 3x\, dx$

(*d*) $\displaystyle\int (1 - 2x)^2\, dx$ (*e*) $\displaystyle\int e^{2x}\, dx$ (*f*) $\displaystyle\int \frac{1}{2x + 3}\, dx$

Ans.:

(*a*) $-(1 - x)^4/4 + c$ (*b*) $-\frac{1}{2}\cos 2x + c$
(*c*) $\frac{1}{3}\sin 3x + c$ (*d*) $-\frac{1}{6}(1 - 2x)^3 + c$
(*e*) $\frac{1}{2}e^{2x} + c$ (*f*) $\frac{1}{2}\log|2x + 3| + c$

7 Sometimes we can make small alterations in the way integrands are written to put the integrands into forms where basic formulas are easily applied. Pay careful attention to the examples

$$\int \tan x\, dx = \int \frac{\sin x}{\cos x}\, dx = -\int \frac{1}{\cos x}(-\sin x)\, dx = -\log|\cos x| + c$$

$$\int \frac{1}{x \log x}\, dx = \int \frac{1}{\log x}\frac{1}{x}\, dx = \log|\log x| + c$$

$$\int xe^{x^2}\, dx = \tfrac{1}{2}\int e^{x^2}(2x)\, dx = \tfrac{1}{2}e^{x^2} + c.$$

Then evaluate

(*a*) $\displaystyle\int x\sqrt{1 + x^2}\, dx$ (*b*) $\displaystyle\int \frac{(\log x)^2}{x}\, dx.$

8 While the terminology plays a minor role in elementary calculus, we can start learning that the equation

$$\frac{dy}{dx} = 2x \tag{1}$$

is an example of an equation that should be called a derivative equation but is called a *differential equation*. Functions y for which (1) holds are called *solutions* of (1), and we know that (1) has many solutions. The particular solution of (1) satisfying the *boundary condition*

$$y = 16 \text{ when } x = 3 \tag{2}$$

is found in a straightforward way. If (1) holds, then

$$y = \int 2x\, dx = x^2 + c, \tag{3}$$

where c is a constant that can be 5 or -3 or 416 but cannot be all of these things at once. The function in (3) will satisfy (2) if $16 = 9 + c$ and hence if $c = 7$. Thus the answer is $y = x^2 + 7$. With clues to methods being provided by this example, find the solutions of the following differential equations satisfying the given boundary conditions.

(*a*) $\dfrac{dy}{dx} = 0$, $y = 1$ when $x = 0$ — *Ans.:* $y = 1$

(*b*) $\dfrac{dy}{dx} = 1$, $y = 2$ when $x = 3$ — *Ans.:* $y = x - 1$

(*c*) $\dfrac{dy}{dx} = \cos 2x$, $y = 0$ when $x = 0$ — *Ans.:* $y = \frac{1}{2}\sin 2x$

(*d*) $\dfrac{dy}{dx} = e^{3x}$, $y = 1$ when $x = 0$ — *Ans.:* $y = \frac{1}{3}e^{3x} + \frac{2}{3}$

9 A body moves to and fro on a line in such a way that its scalar velocity v at time t is given by the formula

$$v = t^2 - 8t + 15.$$

During what interval of time is the scalar velocity negative, and how far does the body move during that time? *Hint:* If s is the coordinate of the body at time t, then $ds/dt = v$ and hence

$$s = \frac{t^3}{3} - 4t^2 + 15t + c,$$

where c is a constant that is 0 if we choose the origin such that $s = 0$ when $t = 0$. *Ans.:* $\frac{4}{3}$ units.

10 If y is a function of x satisfying the differential equation

$$\frac{dy}{dx} = ky \tag{1}$$

and if we know that $y > 0$, then we can divide by y to obtain the first and then the rest of the formulas

$$\frac{1}{y}\frac{dy}{dx} = k, \qquad \log y = kx + c, \qquad y = e^{kx+c}, \qquad y = e^{kx}e^c, \tag{2}$$

where c is a constant that depends upon the particular function y with which we started. But e^c is a constant that we can call A, so

$$y = Ae^{kx}. \tag{3}$$

If we know that y satisfies the boundary condition $y = y_0$ when $x = 0$, then we can put $x = 0$ in (3) to find that $A = y_0$ and hence

$$y = y_0e^{kx}. \tag{4}$$

Remark: Without assuming that $y \neq 0$, we can solve (1) with the aid of a trick. Transposing a term in (1) and multiplying by e^{-kx} give the first and hence the second of the formulas

$$e^{-kx}\left(\frac{dy}{dx} - ky\right) = 0, \qquad \frac{d}{dx}e^{-kx}y = 0.$$

This gives the first and hence the second of the formulas

$$e^{-kx}y = A, \qquad y = Ae^{kx}.$$

More complete treatments of these matters are given in textbooks on differential equations.

11 After having digested the preceding problem (which required that some ideas and methods be absorbed), find the solutions of the following differential equations that satisfy the given boundary conditions

(*a*) $\dfrac{dy}{dx} = y$, $y = 1$ when $x = 0$ *Ans.:* $y = e^x$

(*b*) $\dfrac{dy}{dx} = 2y$, $y = 3$ when $x = 4$ *Ans.:* $y = 3e^{2(x-4)}$

(*c*) $\dfrac{dy}{dx} = y$, $y = 0$ when $x = 0$ *Ans.:* $y = 0$

12 Fill in the right members of the three formulas

$$\frac{d^2x}{dt^2} = ?, \qquad \frac{dx}{dt} = ?, \qquad x = ?$$

when

(*a*) $\dfrac{d^2x}{dt^2} = g$ *Ans.:* g, $gt + c_1$, $\frac{1}{2}gt^2 + c_1t + c_2$

(*b*) $x = \frac{1}{2}gt^2 + c_1t + c_2$ *Ans.:* Same as (*a*)

(*c*) $\dfrac{dx}{dt} = \sin t$ *Ans.:* $\cos t$, $\sin t$, $-\cos t + c$

(*d*) $\dfrac{dx}{dt} = \cos 2t$ *Ans.:* $-2 \sin 2t$, $\cos 2t$, $\frac{1}{2} \sin 2t + c$

(*e*) $\dfrac{dx}{dt} = e^{2t}$ *Ans.:* $2e^{2t}$, e^{2t}, $\frac{1}{2}e^{2t} + c$

13 A particle P moves in the xy plane in such a way that its acceleration $\mathbf{a}$ (a vector) is always $-g\mathbf{j}$. Thus

$$\mathbf{a} = 0\mathbf{i} - g\mathbf{j}.$$

Show that its velocity and displacement vectors must be

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{i} + (-gt + k_1)\mathbf{j} \\ \mathbf{r} &= (c_1t + c_2)\mathbf{i} + (-\tfrac{1}{2}gt^2 + k_1t + k_2)\mathbf{j}, \end{aligned}$$

where the c's and k's are constant. Find the equation of the path in rectangular coordinates when $c_1 \neq 0$ and again when $c_1 = 0$. *Hint:* To solve the last part, put $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$ so that

$$x = c_1t + c_2, \qquad y = -\tfrac{1}{2}gt^2 + k_1t + k_2$$

and eliminate t.

14 This problem requires us to think about indefinite integrals and gives our first glimpse of the famous and important formula for integration by parts.

Let x be confined to an interval I over which two given functions u and v are differentiable. The standard formula

$$(1)\qquad \frac{d}{dx}\,u(x)v(x) = u(x)v'(x) + v(x)u'(x)$$

then (why?) gives the formula

$$(2)\qquad \int[u(x)v'(x) + v(x)u'(x)]\,dx = u(x)v(x) + c.$$

In case the separate integrals are cooperative enough to exist, we can (why?) put (2) in the form

$$(3)\qquad \int u(x)v'(x)\,dx + \int v(x)u'(x)\,dx = u(x)v(x) + c$$

and transpose to obtain the formula

$$(4)\qquad \int u(x)v'(x)\,dx = u(x)v(x) - \int v(x)u'(x)\,dx,$$

which is known as the *formula for integration by parts*. For the particular case in which $u(x) = x$ and $v(x) = e^x$, show that (4) reduces to

$$(5)\qquad \int xe^x\,dx = xe^x - e^x + c.$$

Finally, check (5) by showing that the derivative of the right side actually is the integrand.

15 Read and work the preceding problem again.

16 With Problem 14 out of sight, start with the formula for the derivative of a product and construct the formula for integration by parts and give an application of it.

17 Start with the function f_0 for which $f_0(x) = 1$ when $-1 < x < 1$ and determine the natures of the functions $f_1, f_2, f_3, \cdots$ for which the formulas

$$f_n'(x) = f_{n-1}(x), \qquad f_n(x) = \int f_{n-1}(x)\,dx, \qquad (n = 1,2,3, \cdots)$$

are valid. *Ans.:* There exist constants $c_1, c_2, c_3, \cdots$ such that

$$\begin{aligned}
f_1(x) &= x + c_1\\
f_2(x) &= \tfrac{1}{2}x^2 + c_1x + c_2\\
f_3(x) &= \tfrac{1}{6}x^3 + \tfrac{1}{2}c_1x^2 + c_2x + c_3\\
f_4(x) &= \tfrac{1}{24}x^4 + \tfrac{1}{6}c_1x^3 + \tfrac{1}{2}c_2x^2 + c_3x + c_4
\end{aligned}$$

etcetera. *Remark:* These things will appear later.

18 Prove that the first of the formulas

$$\int \operatorname{sgn} x\,dx = |x| + c_1, \qquad \int \operatorname{sgn} x\,dx = |x| + c_2$$

is correct when $x > 0$ and the second is correct when $x < 0$. Prove that there is no constant c such that the formula

$$\int \operatorname{sgn} x\,dx = |x| + c$$

is correct when $x = 0$.

19 This problem contains hundreds of parts, and there is much to be said for spending several hours or days solving a considerable number of them. To solve one part, pick an integral formula from a (preferably your) book of tables that has the form

$$\int f(x)\,dx = F(x) + c$$

(except that most tables omit the constants) and show that $F'(x) = f(x)$. This promotes an understanding of integral formulas and provides practice in differentiation. It is never too early to start acquaintance with formulas involving X where X is one or another of $a + bx$, $a^2 + x^2$, $a^2 - x^2$, $ax^2 + bx + c$, etcetera. In these situations, modesty and timidity are not virtues. We profit most when we attack the problems that seem most impenetrable and discover that they really are very simple.

4.2 *Riemann sums and integrals* This section introduces the sums and integrals that are named after Riemann (1826–1866) in spite of the fact that Archimedes (287–212 B.C.) knew how special ones could be used in a few special cases. Let f be a function which is defined over an interval $a \leqq x \leqq b$ and has values $f(x)$ such that

$$(4.21)\qquad m \leqq f(x) \leqq M \qquad (a \leqq x \leqq b)$$

where m and M are constants. This amounts to saying that f is *bounded* over the interval; M is an *upper bound* and m is a *lower bound.* Our next few steps are so simple that it may be difficult to see why they are important. As in Figure 4.212, let x be a fixed (or selected) number for which $a < x \leqq b$. Thus x can be b, but it is not necessarily so. Let n be a positive integer. We make a *partition* P of the interval from a to x into n subintervals by inserting points $t_0, t_1, t_2, \cdots, t_{n-1}, t_n$, where

$$(4.211)\qquad a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = x.$$

These points are the circled points of the figure and are the end points of the subintervals.

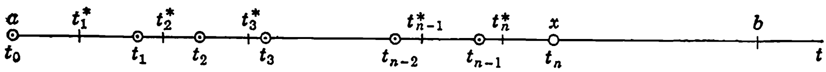


Figure 4.212

Let Δt_1 denote the length of the first subinterval so that $\Delta t_1 = t_1 - t_0$, let Δt_2 denote the length of the second subinterval so that $\Delta t_2 = t_2 - t_1$, and so on so that

$$(4.213)\qquad \Delta t_k = t_k - t_{k-1} \qquad (1 \leqq k \leqq n).$$

It is not required that the points $t_0, t_1, \cdots, t_n$ be equally spaced. The greatest of the numbers $\Delta t_1, \Delta t_2, \cdots, \Delta t_n$ is called the *norm* of the partition P and is denoted by the symbol $|P|$. Thus $|P|$ is the length of the longest of the subintervals in P. Our next act introduces the star characters. Let t_1^* (read tee one star) be a number (or point) in the first subinterval so that $t_0 \leqq t_1^* \leqq t_1$, let t_2^* be in the second subinterval so that $t_1 \leqq t_2^* \leqq t_2$, and so on so that

$$t_{k-1} \leqq t_k^* \leqq t_k \qquad (1 \leqq k \leqq n). \tag{4.214}$$

Our machinery, which is still very much simpler than that in an electrically operated dishwasher, enables us to produce numbers that are called *Riemann sums*. We multiply $f(t_k^*)$, the value of f at t_k^*, by Δt_k, the length of the interval containing t_k^*, and add the results. Thus, denoting the Riemann sum by the symbol RS, we have

$$\text{RS} = f(t_1^*)\,\Delta t_1 + f(t_2^*)\,\Delta t_2 + f(t_3^*)\,\Delta t_3 + \cdots + f(t_n^*)\,\Delta t_n. \tag{4.22}$$

Because it takes too long to write this, we abbreviate it to the form

$$\text{RS} = \sum_{k=1}^{n} f(t_k^*)\,\Delta t_k. \tag{4.221}$$

The right side is read "sigma k running from 1 to n eff of tee kay star delta tee kay" and it denotes the sum of the terms obtained by giving k the values $1, 2, 3, \cdots, n$. The Σ (sigma) is called the summation symbol, and it is very convenient.

Everybody should see that, when the function f and the numbers a and x are given, it is easy to select the partition P in very many different ways and to select the points t_k^* in very many different ways. When an electronic computer is kind enough to do the arithmetical chores, it is even easy to produce very many Riemann sums.

Experience shows that we should avoid future difficulties by allowing the partitions and Riemann sums to slumber peacefully while we invest a moment to think about the names which we have attached to the partition points and the intermediate star points that determine them. The points in Figure 4.212 were called $t_0, t_1, \cdots, t_n$ and $t_1^*, t_2^*, \cdots, t_n^*$. We could, without changing the value of the Riemann sum, have called these same points $\lambda_0, \lambda_1, \cdots, \lambda_n$ and $\lambda_1^*, \lambda_2^*, \cdots, \lambda_n^*$. Thus there is a sense in which the names of these points are "dummy names"; we could have called the points t's or u's or v's or λ's or μ's or ϕ's. When this matter is understood, we must ask and answer two questions. First, why did we avoid the "natural" names $x_0, x_1, \cdots, x_n$ and $x_1^*, x_2^*, \cdots, x_n^*$? The answer is that we already have the interval from a to x on an x axis appearing in our work, and we will have too many x's around the house if we allow any more to enter. Secondly, why did we use the names $t_0, t_1, \cdots, t_n$ and $t_1^*, t_2^*, \cdots, t_n^*$? The only answer we can give is that they are as good as any and better than most alternatives. In situations where we can conveniently use the "natural" names $x_0, x_1, \cdots, x_n$ and $x_1^*, x_2^*, \cdots, x_n^*$, we usually do so. Finally, we do not use the letter i to denote "dummy integers" in (4.214) because the habit of using i leads to awkwardness when we finish study of calculus and enter realms where i is always the imaginary unit whose square is -1. We use k because it is as good as any and better than most.

We now come to the most fundamental remark that appears in the theory of Riemann integration; analogous remarks appear in theories of other integrals. Depending upon the function f and the numbers a and x that have been selected, it may be true (or it may be false) that there is a number I such that to each positive number ϵ there corresponds a positive number δ such that

$$\left| \sum_{k=1}^{n} f(t_k^*)\, \Delta t_k - I \right| < \epsilon \tag{4.23}$$

whenever $|P| < \delta$. This is, of course, just a precise way of saying that there may be a number I such that each Riemann sum with a small norm is near I. If this I exists, then f is said to be *Riemann integrable* over the interval from a to x and I is said to be the *Riemann integral* of f over the interval. This integral is denoted by the symbol in the formula

$$I = \int_a^x f(t)\, dt \tag{4.24}$$

and the symbol is read "the integral from a to x of eff of tee dee tee." The numbers a and x are called the *lower limit* and the *upper limit* of integration, and we always read the lower one first. The symbol t is called a *dummy variable* of integration, the derogatory terminology being applied because the value of the integral would be the same if t were replaced by s or u or α or θ or any other symbol that cannot be confused with a, x, f, and d. It is a convenience (and sometimes also a source of misunderstanding, confusion, and controversy) to drag in the notation of limits and write

$$\lim_{|P| \to 0} \sum_{k=1}^{n} f(t_k^*)\, \Delta t_k = \int_a^x f(t)\, dt. \tag{4.25}$$

A much more substantial convenience results from boiling this down to

$$\lim \sum f(t)\, \Delta t = \int_a^x f(t)\, dt, \tag{4.251}$$

the idea being that we can restore the omitted embellishments whenever there is a reason for doing so.

In case no such number I exists, we say that f is not Riemann integrable over the interval from a to x and that $\int_a^x f(t)\, dt$ does not exist (that is, does not exist as a Riemann integral). To emphasize the fact that a bounded function f and an interval $a \leqq x \leqq b$ can be such that $\int_a^b f(t)\, dt$

does not exist, we look briefly at an example. Let f be the dizzy dancer function D, defined over the interval $0 \leqq x \leqq 1$, for which

$$\begin{cases} D(x) = 0 & (x \text{ irrational}) \\ D(x) = 1 & (x \text{ rational}). \end{cases} \tag{4.252}$$

It is clear that, whatever the partition P of the interval $0 \leqq x \leqq 1$ may be, the Riemann sum

$$\sum_{k=1}^{n} D(t_k^*)\, \Delta t_k \tag{4.253}$$

has the value 0 if the numbers $t_1^*, t_2^*, \cdots, t_n^*$ are all irrational and has the value 1 if the numbers $t_1^*, t_2^*, \cdots, t_n^*$ are all rational. It follows from this that there is no number I such that to each positive number ϵ there corresponds a positive number δ such that (4.23) holds whenever $|P| < \delta$. This shows that the symbol $\int_0^1 D(t)\, dt$ has no meaning or that $\int_0^1 D(t)\, dt$ does not exist.

If we suppose, as above, that f is a function which is bounded over an interval, then the following theorem shows that the answer to the question whether f is integrable over the interval depends only upon the set of discontinuities of f.

Theorem 4.26 *A function f is Riemann integrable over an interval if and only if it is bounded and the set of discontinuities of f which lie in the interval has Lebesgue measure zero.*

This theorem is proved in modern textbooks that fully earn the right to be called textbooks on advanced calculus. The proof is, from our present point of view, both long and difficult, and we do not need to know anything about it. Moreover, we do not need to understand the theorem, but we should not be injured by taking a hasty look at Figure 4.261 and making a modest attempt to understand one of the definitions

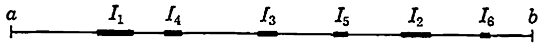


Figure 4.261

which has fundamental importance in more advanced mathematics. A set D of points on a line is said to have *Lebesgue measure* 0 if to each $\epsilon > 0$ there corresponds a collection $I_1, I_2, I_3, \cdots$ of intervals such that each point of D lies in at least one of these intervals and, for each $n = 1, 2, 3, \cdots$, the sum of the lengths of the first n intervals is less than ϵ. Sometimes it is very easy to show that a given set D has Lebesgue measure 0 by showing that if $\epsilon > 0$, then there exist intervals $I_1, I_2, I_3 \cdots$ such that each point of D is in at least one of these intervals and, moreover, the length of I_1 is less than $\epsilon/2$, the length of I_2 is less than $\epsilon/2^2$, the length of I_3 is less than $\epsilon/2^3$, etcetera. The collection of intervals may be a finite collection, that is, it may contain only 1 or 2 or 3 or 416 or 31,690 or some other positive integer number of intervals. The collection of intervals may be a "countably infinite collection," that is, it may contain a first, a second, a third, etcetera, so that to each positive integer k there corresponds an interval I_k. In each of these two cases, the collection of intervals is said to be a "countable collection." Only a most rudimentary understanding of these matters enables us to reach the conclusion that if D contains only 0 or 1 or 2 or 3 or 416 or any other finite number of points, then D must have Lebesgue measure 0. In any case, we should have at least a hazy understanding of the fact that Lebesgue (1875–1941) was a great French mathematician and that Theorem 4.26 implies the much simpler following theorem which we are required to know in this course.

Theorem 4.27 *If f is bounded over the interval $a \leqq x \leqq b$ and if f is continuous over the interval (or is discontinuous but has only a finite set of discontinuities in the interval), then the Riemann integral in*

$$\int_a^x f(t)\,dt = \lim \sum f(t_k^*)\,\Delta t_k \tag{4.271}$$

exists when $a < x \leqq b$.

As we near the end of the text of our introductory section on Riemann sums and integrals, we pause to think about our present state and future development. We have a new symbol, namely, $\int_a^x f(t)\,dt$. If $a < b$, if $a < x \leqq b$, and if f is defined over the interval $a \leqq x \leqq b$, then (depending upon a, x, and f) the symbol may be meaningless or it may be a number. Answers to questions depend upon partitions and Riemann sums. Partitions are so simple that our little sister can understand them completely and be puzzled only by our great interest in them. Riemann sums $\Sigma f(t_k^*)\,\Delta t_k$ are less simple, but we can construct them in great profusion. Matters grow substantially more complex when we ask whether there is a number $\int_a^x f(t)\,dt$ such that to each positive number ϵ there corresponds a positive number δ such that

$$\left| \sum_{k=1}^{n} f(t_k^*)\,\Delta t_k - \int_a^x f(t)\,dt \right| < \epsilon \tag{4.28}$$

whenever P is a partition of the interval $a \leqq t \leqq x$ for which $|P| < \delta$. We should all recognize this and admit that full comprehensions of machinery and its applications is not quickly attained. In fact a substantial part of this textbook is devoted to promotion of understanding of Riemann sums and their applications. We shall have plenty of opportunities to learn.

So far, the integral in (4.271) has been defined only when $x > a$. We now complete the definition by setting

$$(4.281) \qquad \int_a^a f(t)\,dt = 0, \qquad \int_a^x f(t)\,dt = -\int_x^a f(t)\,dt,$$

the second formula being valid when $x < a$ and f is integrable over the interval from x to a.

Problems 4.29

1 Practice the art of telling how the number

$$\int_a^x f(t)\,dt$$

is defined. Be prepared to give the full details, including Riemann sums, at any time.

2 Tell whether you think it wise to abbreviate the statement "To each positive number ϵ there corresponds a positive number δ such that

$$\left| \sum_{k=1}^{n} f(t_k^*)\,\Delta t_k - \int_a^x f(t)\,dt \right| < \epsilon$$

whenever the sum is a Riemann sum formed for a partition P of the interval $a \leqq t \leqq x$ having norm less than δ" to the statement

$$\lim_{|P|\to 0} \sum_{k=1}^{n} f(t_k^*)\,\Delta t_k = \int_a^x f(t)\,dt.$$

Remark: If you do not have an opinion, think about the matter and get one.

3 For better or for worse, the "formula"

$$\lim \sum f(t_k^*)\,\Delta t_k = \int_a^x f(t)\,dt$$

is considered to be an assertion. Tell precisely what it means.

4 Tell whether you would like to learn and use a completely new notation by which the "formula"

$$\operatorname*{approx}_{\epsilon, |P|<\delta} \sum_{k=1}^{n} f(t_k^*)\,\Delta t_k = \int_a^x f(t)\,dt$$

is used to abbreviate the statement that to each positive number ϵ there corresponds a positive number δ such that

$$\left| \sum_{k=1}^{n} f(t_k^*)\,\Delta t_k - \int_a^x f(t)\,dt \right| < \epsilon$$

whenever P is a partition of the interval $a \leqq t \leqq x$ for which $|P| < \delta$. *Remark:* If you do not have an opinion, think about the matter and get one.

5 We often hear about the great scientific progress of our modern era, and we should think about an example. One of the great strides forward is made by abandoning the good old idea that the elementary functions (polynomials, trigonometric functions, etcetera) are always the simplest and most useful functions. There are very many situations in which step functions are the simplest and most useful functions. Our first problem is to follow instructions to prove that

$$(1) \qquad \int_2^5 7\,dt = 21.$$

Draw a figure showing a partition P of the interval $2 \leqq x \leqq 5$ into subintervals having lengths $\Delta t_1, \Delta t_2, \cdots, \Delta t_n$ and observe that

$$(2) \qquad \sum_{k=1}^{n} \Delta t_k = 3.$$

Observe that the integral in (1) involves the function f for which $f(x)$ is always 7. Show that, with the notation of the text,

$$(3) \qquad \begin{aligned} \mathrm{RS} &= \sum_{k=1}^{n} f(t_k^*)\,\Delta t_k = \sum_{k=1}^{n} 7\,\Delta t_k \\ &= 7\sum_{k=1}^{n} \Delta t_k = 7\cdot 3 = 21. \end{aligned}$$

Since each Riemann sum is 21, it is quite apparent that RS is near 21 whenever the norm of P is near 0. This proves the formula (1).

6 Supposing that $a < b$ and k is a constant, prove that

$$(1) \qquad \int_a^b k\,dt = k(b - a).$$

Remark: We are (or soon will be) authorities on areas of rectangular regions. We can observe that if $k > 0$, then the right side of (1) is the area of a rectangular region having base length $(b - a)$ and height k and hence is the area of the region of Figure 4.291 which is bounded by the graphs of the equations $x = a$,

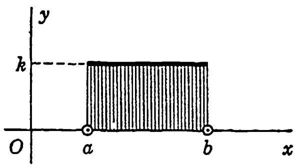


Figure 4.291

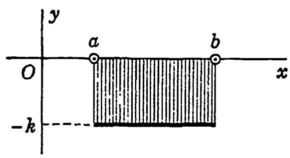


Figure 4.292

$x = b$, $y = 0$, and $y = k$. In case $k < 0$, we can put (1) in the form

$$(2) \qquad \int_a^b k\,dt = -(-k)(b - a),$$

where $-k > 0$, and observe that the right side is the negative of the area of the region in Figure 4.292 bounded by the graphs of the equations $x = a$, $x = b$, $y = 0$, and $y = k$. We must always know that areas of rectangles are positive The idea that rectangles below the x axis have negative areas is as absurd as the idea that cities south of the equator have negative populations.

7 This problem requires us to attain a complete understanding of a more complex situation. Let

$$f(x) = 3 \qquad (2 < x < 4)$$
$$f(x) = 4 \qquad (4 < x < 5)$$

and let $f(2)$, $f(4)$, and $f(5)$ be defined in any way that pleases (or displeases) the fancy. Then Theorem 4.27 implies existence of the integral I in

$$I = \int_2^5 f(t)\,dt.$$

We want to find I, that is, to find the numerical value of I. Draw a figure showing the interval $2 \leqq x \leqq 5$ and, in addition, the graph of f. Make a partition P of the interval $2 \leqq x \leqq 5$ in which 4 is one of the partition points. Choose the points t_k^* in such a way that they are *not* at the ends of the intervals in which they lie. Show that the terms in the Riemann sum RS can be split into two sums RS_1 and RS_2 in such a way that RS_1 contains those terms for which $2 < t_k^* < 4$ and RS_2 contains those terms for which $4 < t_k^* < 5$. Show that $\text{RS}_1 = 6$ and $\text{RS}_2 = 4$ and hence that $\text{RS} = 10$. Our next step is to realize what we are trying to do. We are not trying to prove that f is integrable and are not required to prove that $|\text{RS} - I|$ is small whenever the norm of P is small. We are trying to find I, and we can use the known fact that $|\text{RS} - I|$ must be near 0 whenever the norm of P is near 0. Therefore $|\text{RS} - I|$ must be near 0 whenever P is a partition of the type constructed above and the norm of P is near 0. But $\text{RS} = 10$ for each partition of the type constructed above, and it follows that $I = 10$. Notice that we have, in the course of our work, proved that

$$\int_2^5 f(t)\,dt = \int_2^4 f(t)\,dt + \int_4^5 f(t)\,dt.$$

Interpret the numerical results in terms of areas of rectangular regions.

8 Supposing that $x_1 < x_2 < x_3$ and that k_1 and k_2 are constants, draw a figure showing the interval $x_1 \leqq x \leqq x_3$ and a graph of the function f for which $f(x_1) = f(x_2) = f(x_3) = 0$ and

$$f(x) = k_1 \qquad (x_1 < x < x_2)$$
$$f(x) = k_2 \qquad (x_2 < x < x_3).$$

Show that

$$\begin{aligned}\int_{x_1}^{x_3} f(t)\,dt &= \int_{x_1}^{x_2} f(t)\,dt + \int_{x_2}^{x_3} f(t)\,dt \\ &= k_1(x_2 - x_1) + k_2(x_3 - x_2).\end{aligned}$$

Tell how the result can be interpreted in terms of areas of rectangular regions when (*a*) k_1 and k_2 are both positive, (*b*) $k_1 > 0$ and $k_2 < 0$, and (*c*) k_1 and k_2 are both negative. Explain the manner in which these results can be extended to step functions that have constant values over 3 or 300 intervals instead of just 2.

9 Tell why each of the following Riemann integrals exists or fails to exist.

(*a*) $\int_0^1 (2 + 3t + 4t^2)\,dt$ (*b*) $\int_0^1 \frac{t+1}{t+2}\,dt$

(*c*) $\int_{-1}^1 \frac{1}{t}\,dt$ (*d*) $\int_1^x \frac{1}{t}\,dt$

Hint: Use theorems given in the text.

10 In a campaign to obtain good ideas about Riemann sums and integrals, we can use the discontinuous function ϕ, defined over the interval $0 \leqq x \leqq 1$, for which

$$(1)\qquad \begin{cases} \phi(x) = 0 \text{ when } x \neq 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \cdots \\ \phi(x) = 1 \text{ when } x = \dfrac{1}{m} \text{ and } m \text{ is a positive integer.} \end{cases}$$

Sketch a figure which shows the nature of the graph of ϕ. Then mark the end points $x_0, x_1, \cdots, x_n$ and the intermediate points $x_1^*, x_2^*, \cdots, x_n^*$ of a partition P of the interval $0 \leqq x \leqq 1$ for which $|P| < 0.1$; to mark the end points of a partition for which $|P| < 0.0001$ would be a tedious operation requiring sharper pencils and better microscopes than we normally carry around. Try to see reasons why the Riemann sum

$$(2)\qquad \sum_{k=1}^{n} \phi(x_k^*)\,\Delta x_k$$

must be near 0 whenever $|P|$ is near 0 and hence that

$$(3)\qquad \int_0^1 \phi(x)\,dx = 0.$$

Then start cultivating the art of understanding and originating thoughts about Riemann sums more or less like the following. Let ϵ be a given positive number for which $0 < \epsilon < 1$. Let $h = \epsilon/10$ and suppose at first that $|P| \leqq h$. The terms of the Riemann sum (2) are all nonnegative. Those terms for which x_k^* can be a point of the interval $0 \leqq x \leqq h$ contribute at most $2h$ to the sum. Those terms for which $x_k^* > h$ will be 0 unless $x_k^* = 1/m$, where m is an integer for which $1/m > h$ or $m < 1/h$. Thus there are less than $1/h$ nonzero terms for which $x_k^* > h$ and $\phi(x_k^*) \neq 0$. Since each one of these terms can contribute at most $|P|$ to the sum of these terms, it follows that the sum of all of these terms cannot exceed $(1/h)|P|$. Therefore,

$$(4)\qquad 0 \leqq \sum_{k=1}^{n} \phi(x_k^*)\,\Delta x_k \leqq 2h + \frac{1}{h}|P| < 0.2\epsilon + \frac{10|P|}{\epsilon}.$$

If we let $\delta = 2\epsilon^2/25$, we will have

$$(5)\qquad 0 \leqq \sum_{k=1}^{n} \phi(x_k^*)\,\Delta x_k < \epsilon$$

whenever $|P| < \delta$. This implies the first and hence the second of the formulas

$$(6)\qquad \lim_{|P|\to 0} \sum_{k=1}^{n} \phi(x_k^*)\,\Delta x_k = 0, \qquad \int_0^1 \phi(x)\,dx = 0.$$

11 Supposing that g is the corn-popper function of Problem 16 of Problems 3.49, determine the value (if any) of $\int_0^1 g(x)\,dx$.

12 Our purpose is to discover that some very obvious and superficially useless remarks about Riemann sums lead to the very useful conclusion that the formula

$$\text{(1)} \qquad \int_a^x f(t)\,dt = \int_{(a-q)/p}^{(x-q)/p} f(pu+q)p\,du$$

is correct whenever p and q are constants for which $p \neq 0$ and the integral on the left exists. Let us begin by looking at (1). If we suppose that two variables t and u are related by the formulas

$$\text{(2)} \qquad t = pu + q, \qquad u = \frac{t-q}{p}, \qquad \frac{dt}{du} = p,$$

we can put $t = pu + q$ and $dt = (dt/du)\,du$ in everything after the integral sign in the left side of (1) to obtain everything after the integral sign in the right side of (1). We can observe that the lower limit of integration on the right side is the value attained by u when t is the lower limit of integration on the left side. Similarly, the upper limit of integration on the right side is the value attained by u when t is the upper limit of integration on the left side. Of course, we are entitled to take a dim view of these manipulations until we discover how simple and useful they are. Meanwhile, we forget about (1) and start working with some Riemann sums. Let P be, as in Figure 4.293, a partition of the interval $a \leqq t \leqq x$ having partition points $t_0, t_1, \cdots, t_n$ and intermediate points $t_1^*, t_2^*, \cdots, t_n^*$. Supposing that t and u are related by the formulas (2) and that $p > 0$, we set

$$\text{(3)} \qquad u_k = \frac{t_k - q}{p}, \qquad u_k^* = \frac{t_k^* - q}{p} \qquad (k = 1,2, \cdots, n).$$

To simplify writing, we set

$$\text{(4)} \qquad A = \frac{t_0 - q}{p} = \frac{a-q}{p}, \qquad X = \frac{t_n - q}{p} = \frac{x-q}{p}.$$

The numbers $u_0, u_1, \cdots, u_n$ and $u_1^*, u_2^*, \cdots, u_n^*$ then form the partition

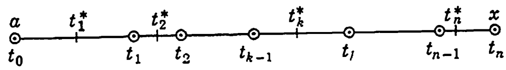


Figure 4.293

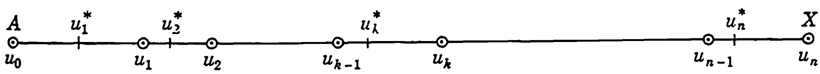


Figure 4.294

points and the intermediate points of partition Q of the interval $A \leqq u \leqq X$ shown in Figure 4.294. Moreover, when

$$\text{(5)} \qquad \Delta t_k = t_k - t_{k-1}, \qquad \Delta u_k = u_k - u_{k-1},$$

we find that

$$\text{(6)} \qquad \Delta t_k = (pu_k + q) - (pu_{k-1} + q) = p\,\Delta u_k$$

and

$$\sum_{k=1}^{n} f(t_k^*)\,\Delta t_k = \sum_{k=1}^{n} f(pu_k^* + q)p\,\Delta u_k. \tag{7}$$

The result (1) follows from this. Let I denote the left member of (1). Let $\epsilon > 0$. There is then a $\delta > 0$ such that

$$\Big| I - \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k \Big| < \epsilon \qquad (|P| < \delta) \tag{8}$$

Since (6) implies that $|P| = p|Q|$, we see from (8) and (7) that

$$\Big| I - \sum_{k=1}^{n} f(pu_k^* + q)p\,\Delta u_k \Big| < \epsilon \qquad (|Q| < \delta/p). \tag{9}$$

But the sum in (9) is a Riemann sum formed for the partition Q of the interval $A \leqq u \leqq X$ and the function F having values

$$F(u) = f(pu + q)p \tag{10}$$

when $A \leqq u \leqq X$. It therefore follows from the definition of Riemann integrals that

$$\int_A^X F(u)\,du = I. \tag{11}$$

But (4) and (10) show that the left member of (11) is the right member of (1). This proves our conclusion (1) for the case in which $p > 0$. In case $p < 0$, some details must be modified because Figure 4.294 must be turned end-for-end, but the result is still correct. The formula (1) which we have proved is called the formula for linear changes of variables in Riemann integrals.

13 By use of formula (1) of Problem 12 show that

(*a*) $\displaystyle\int_a^b (t-c)^2\,dt = \int_{a-c}^{b-c} u^2\,du$ *Hint:* Put $t - c = u$.

(*b*) $\displaystyle\int_0^{\pi/2} \sin 2t\,dt = \frac{1}{2}\int_0^{\pi} \sin u\,du$

(*c*) $\displaystyle\int_0^h \frac{1}{a^2+x^2}\,dx = \frac{1}{a}\int_0^{h/a} \frac{1}{1+t^2}\,dt$ *Hint:* Put $x = at$.

(*d*) $\displaystyle\int_0^h \frac{1}{\sqrt{a^2-x^2}}\,dx = \int_0^{h/a} \frac{1}{\sqrt{1-t^2}}\,dt$

(*e*) $\displaystyle\int_a^b \sin x\,dx = \int_{a-h}^{b-h} \sin(x+h)\,dx$

Hint: Before you start, replace one of the variables of integration by a different variable of integration.

(*f*) $\displaystyle\int_a^b f(x)\,dx = \int_{a-h}^{b-h} f(x+h)\,dx.$

Remark: This last formula shows that we can add a constant to the variable of integration if we subtract it from the limits of integration. This information is sometimes very useful.

14 Assuming that the integrals exist, show that

$$\int_{-h}^{0} f(x)\,dx = \int_{0}^{h} f(-x)\,dx. \tag{1}$$

Remark: This innocent formula and a result of the next section enable us to produce the better formula

$$\begin{aligned}\int_{-h}^{h} f(x)\,dx &= \int_{-h}^{0} f(x)\,dx + \int_{0}^{h} f(x)\,dx \\ &= \int_{0}^{h} [f(-x) + f(x)]\,dx.\end{aligned} \tag{2}$$

This gives the very useful fact that $\int_{-h}^{h} f(x)\,dx = 0$ when f is an odd function, that is, $f(-x) = -f(x)$, and that

$$\int_{-h}^{h} f(x)\,dx = 2\int_{0}^{h} f(x)\,dx$$

when f is an even function, that is, $f(-x) = f(x)$.

15 *Remark:* This remark is designed to indicate that mathematics is a lively subject in which even good ideas can be modified in various ways, and that there are integrals of many different types. We can be irked by the fact that the Riemann integral

$$\int_{0}^{2} f(x)\,dx$$

does not exist when f is the function for which $f(x) = 1$ when $0 \leqq x < 1$ and $f(x) = 2$ when $1 < x \leqq 2$. The difficulty is that $f(1)$ is undefined and that $\Sigma f(t_k^*)\,\Delta t_k$ is undefined when $t_k^* = 1$ for some k. If, however, we extend the definition of f by setting $f(1) = 75$, then the new extended function is Riemann integrable over the interval $0 \leqq x \leqq 2$. We cannot reasonably undertake to remove this irksome situation by changing the definition of the Riemann integral, because changing basic definitions destroys our means for communication of information. We can, however, introduce new types of integrals. We can, for example, use the letter F to make us think of a finite set and produce the following definition. A function f is Riemann-F integrable over $a \leqq x \leqq b$ if there is a finite set F such that f is defined at all points of the interval $a \leqq x \leqq b$ except at the points of F and, moreover, there is a number I such that to each $\epsilon > 0$ there corresponds a number $\delta > 0$ such that $|\text{RS} - I| < \epsilon$ whenever RS is a Riemann sum for which $|P| < \delta$ and the points t_k^* are all different from points of F. This definition does not require f to be defined everywhere over $a \leqq x \leqq b$ and it removes the irritation. Still another definition can be constructed by making similar use of the letter C to make us think of a countable set of points, this being either a finite set or a set whose elements can be placed in one-to-one correspondence with the set $1, 2, 3, \cdots$ of positive integers. A more sophisticated definition makes use of the letter N to make us think of a null set, this being a set having Lebesgue measure 0. As has been remarked, there are many kinds of integrals. Mathematicians who use integrals without knowing which ones they are using are comparable to chemists who use chemicals without knowing which ones they are using.

4.3 Properties of integrals In what follows, all integrals bearing limits of integration are Riemann integrals. They are limits of Riemann sums, and it could be expected that, except for cases in which the integrands are step functions, it must be impossible to obtain their exact values and it must be difficult to obtain reasonably good approximations to them. It turns out, however, that there is a calculus, an invention of Newton and Leibniz, by which exact values of very many of the most important integrals can be calculated very quickly. Dictionaries tell us that a calculus is "a method of computation." The particular calculus that appears at the end of this section was found to be so overwhelmingly important that it came to be known as "the calculus." This calculus enables us, for example, to evaluate the integral in the formula

$$\int_2^3 x^2\,dx = \frac{x^3}{3}\Big]_2^3 = \tfrac{27}{3} - \tfrac{8}{3} = \tfrac{19}{3} \tag{4.31}$$

by writing nothing more than this. Meanings of words have evolved in such a way that we now consider "calculus" or "the calculus" to be a name assigned to a part of mathematics involving derivatives and integrals.†

For making calculations involving integrals, we often need the results set forth in the following theorems. Proofs of these theorems may be omitted; these theorems are rather simple consequences of Theorem 4.26 and properties of Riemann sums and their limits.

Theorem 4.32 *If f is integrable over an interval containing a, b, and c, then*

$$\int_a^c f(t)\,dt + \int_c^b f(t)\,dt = \int_a^b f(t)\,dt.$$

Theorem 4.33 *If f and g are integrable over $a \leqq x \leqq b$ and A and B are constants, then*

$$\int_{x_1}^{x_2} [Af(t) + Bg(t)]\,dt = A\int_{x_1}^{x_2} f(t)\,dt + B\int_{x_1}^{x_2} g(t)\,dt$$

whenever x_1 and x_2 lie in the interval $a \leqq x \leqq b$.

Theorem 4.34 *If $a < b$, if f_1, f_2, f_3 are integrable over $a \leqq x \leqq b$, and if*

$$f_1(x) \leqq f_2(x) \leqq f_3(x) \qquad (a \leqq x \leqq b)$$

then

$$\int_a^b f_1(t)\,dt \leqq \int_a^b f_2(t)\,dt \leqq \int_a^b f_3(t)\,dt.$$

† Historians who claim that Archimedes knew calculus do not always point out that the knowledge was attained posthumously when the meaning of "calculus" changed. Complete misunderstanding of this matter can serve as a basis for the absurd contention that Newton and Leibniz merely rediscovered inventions of Archimedes.

Theorem 4.341 *If k is a constant, then*

$$\int_{x_1}^{x_2} k\,dt = k(x_2 - x_1).$$

Theorem 4.342 *If $a < b$, if f is integrable over $a \leqq x \leqq b$, and if*

$$m \leqq f(x) \leqq M \qquad (a \leqq x \leqq b)$$

then

$$m(b-a) \leqq \int_a^b f(t)\,dt \leqq M(b-a)$$

and

$$m \leqq \frac{1}{b-a}\int_a^b f(t)\,dt \leqq M.$$

Theorem 4.343 *If f is integrable over $a \leqq x \leqq b$, then so also is the function having values $|f(x)|$ and*

$$\left|\int_{x_1}^{x_2} f(x)\,dx\right| \leqq \left|\int_{x_1}^{x_2} |f(x)|\,dx\right|$$

whenever x_1 and x_2 lie between a and b.

The next theorem is not so obvious, and it is so important that we shall discuss it and prove it. Much of the theory and many of the applications of the calculus involve relations between derivatives and integrals. Theorems which give information about derivatives of integrals or integrals of derivatives are called *fundamental theorems of the calculus*. The following theorem is one of the best of these. It has very many applications and shows, among other things, that if f is continuous over $a \leqq x \leqq b$, then there exists a function F for which $F'(x) = f(x)$ when $a \leqq x \leqq b$. In fact, it shows that if f is continuous, then the Riemann integral in (4.351) is an "indefinite integral" of f.

Theorem 4.35 *If f is integrable over $a \leqq x \leqq b$, then the function F defined by*

$$F(x) = \int_a^x f(t)\,dt \tag{4.351}$$

is continuous over $a \leqq x \leqq b$ and

$$F'(x) = f(x) \tag{4.352}$$

for each x for which f is continuous.

To start the proof, we observe that if x and $x + \Delta x$ both lie in the interval, then

$$F(x+\Delta x) - F(x) = \int_a^{x+\Delta x} f(t)\,dt - \int_a^x f(t)\,dt = \int_x^{x+\Delta x} f(t)\,dt. \tag{4.353}$$

To prove continuity of F, we use Theorem 4.26 to see that f must be bounded and hence that there is a constant positive M for which $-M \leqq f(x) \leqq M$ or $|f(x)| \leqq M$. Therefore,

$$|F(x + \Delta x) - F(x)| \leqq \left| \int_x^{x+\Delta x} M\, dt \right| = M\, |\Delta x|.$$

The sandwich theorem then implies that

$$\lim_{\Delta x \to 0} F(x + \Delta x) = F(x) \tag{4.354}$$

and hence that F is continuous at x. It can be observed that we have proved more than was promised; the function F must have bounded difference quotients. To prove (4.352), let x be a point at which f is continuous. From the two formulas

$$\frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{1}{\Delta x} \int_x^{x+\Delta x} f(t)\, dt, \qquad f(x) = \frac{1}{\Delta x} \int_x^{x+\Delta x} f(x)\, dt$$

we obtain

$$\frac{F(x + \Delta x) - F(x)}{\Delta x} - f(x) = \frac{1}{\Delta x} \int_x^{x+\Delta x} [f(t) - f(x)]\, dt. \tag{4.355}$$

Let $\epsilon > 0$. Choose a number $\delta > 0$ such that

$$|f(t) - f(x)| \leqq \epsilon/2 \qquad (|t - x| < \delta).$$

Then when $|\Delta x| < \delta$, we can use Theorems 4.343 and 4.341 to obtain

$$\begin{aligned} \left| \frac{F(x + \Delta x) - F(x)}{\Delta x} - f(x) \right| &\leqq \left| \frac{1}{\Delta x} \int_x^{x+\Delta x} |f(t) - f(x)|\, dt \right| \\ &\leqq \left| \frac{1}{\Delta x} \int_x^{x+\Delta x} \frac{\epsilon}{2}\, dt \right| = \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

Therefore,

$$\lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = f(x) \tag{4.356}$$

and (4.352) follows from the definition of $F'(x)$. This completes the proof of Theorem 4.35.

Supposing now that f is continuous over $a \leqq x \leqq b$, we proceed to show how Theorem 4.35 can be used to obtain the promised method for evaluating Riemann integrals. Putting $x = a$ in (4.351) shows that $F(a) = 0$. Putting $x = b$ in (4.351) and then changing the dummy variable of integration from t to x gives

$$F(b) = \int_a^b f(x)\, dx.$$

Therefore,

$$\int_a^b f(x)\, dx = F(b) - F(a). \tag{4.36}$$

When problems are being solved, it is always convenient to use the *bracket symbol* in the formula

$$F(x)\Big]_a^b = F(b) - F(a). \tag{4.361}$$

This symbol can be read "eff of x bracket a, b." The symbol means exactly what the formula says it does; to obtain its value, we write the value of $F(x)$ when x has the upper value b and subtract the value of $F(x)$ when x has the lower value a. For example, $x^3\Big]_2^3 = 27 - 8 = 19$. It is easy to see that the value of the bracket symbol is unchanged when we add a constant to the function appearing in it. Thus

$$F(x) + c\Big]_a^b = [F(b) + c] - [F(a) + c] = F(b) - F(a).$$

Therefore, we can put (4.36) in the form

$$\int_a^b f(x)\,dx = F(x) + c\Big]_a^b \tag{4.362}$$

where c is 0 or any other constant. Since we have assumed that f is continuous over $a \leqq x \leqq b$, it is a consequence of Theorem 4.35 that $F'(x) = f(x)$ when $a \leqq x \leqq b$. Since each function whose derivative with respect to x is $f(x)$ must have the form $F(x) + c$, the result (4.362) can be put in the following form.

Theorem 4.37 *If f is continuous over $a \leqq x \leqq b$ and if $F'(x) = f(x)$ when $a \leqq x \leqq b$, then*

$$\int_a^b f(x)\,dx = F(x)\Big]_a^b = F(b) - F(a).$$

In substantially all applications of this theorem, the notation of indefinite integrals is used. In such cases the following version of Theorem 4.37 gives precisely the information we actually use to evaluate integrals.

Theorem 4.38 *The formula*

$$\int_a^b f(x)\,dx = F(x)\Big]_a^b = F(b) - F(a) \tag{4.381}$$

is correct if f is continuous over $a \leqq x \leqq b$ and

$$\int f(x)\,dx = F(x) + c \tag{4.382}$$

when $a \leqq x \leqq b$.

When we are able to find a useful expression for the $F(x)$ in (4.382), the integral in (4.381) can be evaluated with remarkable ease. We simply ignore the limits of integration on the first integral until (4.382) has been obtained and then, taking $c = 0$ unless it seems desirable to give c some other value, insert the bracket symbol to obtain (4.381). For example,

$$\int_2^3 x^2\,dx = \frac{x^3}{3}\Big]_2^3 = \tfrac{27}{3} - \tfrac{8}{3} = \tfrac{19}{3}.$$

Problems 4.39

1 Make a small table of integrals by copying formulas from the second column of (4.171) to (4.175). Combine the processes of learning these formulas and using them to show that

(a) $\int_0^1 (x - x^2)\, dx = \frac{1}{6}$ (b) $\int_0^1 x(1 - x)\, dx = \frac{1}{6}$

(c) $\int_0^1 x^2(1 - x^2)\, dx = \frac{2}{15}$ (d) $\int_0^1 x^2(1 - x)^2\, dx = \frac{1}{30}$

(e) $\int_1^x \frac{1}{t}\, dt = \log x$ (f) $\int_0^2 \frac{x}{1 + x^2}\, dx = \frac{1}{2}\log 5$

(g) $\int_0^\pi \sin x\, dx = 2$ (h) $\int_0^\pi \cos x\, dx = 0$

(i) $\int_0^1 e^{-x}\, dx = 1 - e^{-1}$ (j) $\int_1^2 \left(x + \frac{1}{x}\right)^2 dx = \frac{29}{6}$

(k) $\frac{1}{h}\int_0^h \cos (a + bt)\, dt = \frac{\sin (a + bh) - \sin a}{bh}$

2 Verify the formula

$$\int_0^1 x^p(1 - x)^q\, dx = \frac{p!q!}{(p + q + 1)!}$$

for some pairs of small nonnegative integers p and q. *Remark:* Anyone who wishes to augment his corpus of scientific information should be informed that this is a famous and important formula. The integral is the *beta integral.* The formula is correct whenever p and q are real or complex numbers with nonnegative real parts. When Cauchy extensions of Riemann integrals have been defined and are used, it can be proved that the formula is correct when p and q are complex numbers with real parts exceeding -1.

3 While we are not now indulging in proofs of such things, the two integrals

(1) $$\int_0^1 \frac{1}{(1 + x)^s}\, dx, \qquad \int_0^1 \frac{1}{1 + x}\, dx$$

are nearly equal when s is near 1. Nevertheless, we must use different integration formulas to evaluate the integrals. Obey the rules and show that

(2) $$\int_0^1 \frac{1}{(1 + x)^s}\, dx = \frac{2^{1-s} - 1}{1 - s} \qquad (s \neq 1)$$

(3) $$\int_0^1 \frac{1}{1 + x}\, dx = \log 2.$$

Remark: While the details need not be fully understood at the present time, we pause to learn that the right member of (2) really is near log 2 when s is near 1. This means that

(4) $$\lim_{s \to 1} \frac{2^{1-s} - 1}{1 - s} = \log 2.$$

To see that (4) is correct, we can let

(5) $$f(x) = e^{x \log 2}$$

and, after observing that $x \log 2 = \log 2^x$ and hence $f(x) = 2^x$, put (4) in the form

$$\lim_{s\to 1} \frac{f(1-s) - f(0)}{1-s} = \log 2. \tag{6}$$

But it follows from the definition of derivatives that the left member of (6) is $f'(0)$. From (5) we find that $f'(0) = \log 2$. Therefore, (6) and (4) are correct. The conclusion to be drawn from this story is that the function F defined by

$$F(s) = \int_0^1 \frac{1}{(1+x)^s}\,dx \tag{7}$$

is a continuous function and, unless we can find another scapegoat, we must blame the well-known perversity of inanimate matter for the strange fact that $F(s)$ is expressed in terms of exponentials when $s \neq 1$ and is expressed as a logarithm when $s = 1$.

4 This problem, like very many of the fundamental problems of science, requires much more looking and thinking than calculating. Look at Figure 4.391, which shows the graph of a step function f defined over the interval $a \leqq x \leqq b$, and observe that $f(x) \geqq 0$. Remember that, in the problems at the end of Section 4.2, we discovered (or almost discovered) that

$$\int_a^b f(x)\,dx = |S|, \tag{1}$$

where $|S|$ is the area of the set S of points (x,y) for which $a \leqq x \leqq b$ and $0 \leqq y \leqq f(x)$. The next step is the most difficult one. We should realize that, at the present time, our ideas about areas of nonrectangular point sets are at best somewhat vague and nebulous and are at worst nonexistent or even erroneous. The rest of our work is much easier. We look at Figure 4.392, which shows the graph of f over the interval $a \leqq x \leqq b$ for the special case in which $f(x) = x^2$, $a = 0$, and $b = 1$. As above, let S be the set of points (x,y) for which $a \leqq x \leqq b$ and $0 \leqq y \leqq f(x)$. Our next step is to look again at Figure 4.392 and express the

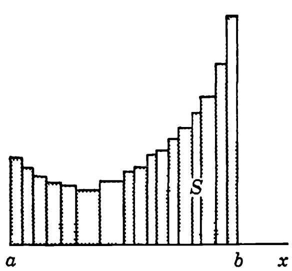


Figure 4.391

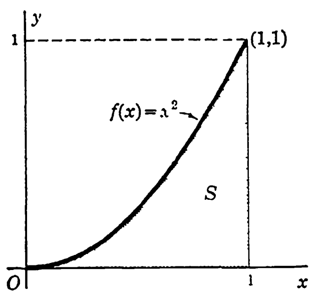


Figure 4.392

cheerful opinion that the set S ought to have an area which we can denote by $|S|$ and that the formula (1), which holds whenever f is a nonnegative step function, ought to hold whenever f is a nonnegative integrable function. Our final step is to seek what a physicist could call experimental verification of this cheerful opinion. Look at Figure 4.392 and note that S seems to fill up about one-third of the square having opposite vertices at the points (0,0) and (1,1), and hence that the area $|S|$ of S should be about one-third. Now comes the calculation. Show that

$$\int_0^1 x^2\,dx = \tfrac{1}{3}.$$

5 Sketch a graph of the equation $y = x + 1$ over the interval $1 \leqq x \leqq 3$ and use elementary geometrical ideas to find the area of the part of the plane bounded by this graph and the graphs of the equations $x = 1$, $x = 3$, and $y = 0$. Then evaluate the integral

$$\int_1^3 (x+1)\,dx$$

and find out whether we obtain more experimental verification of the cheerful opinion of the preceding problem.

6 Figures 4.393 and 4.394 show graphs of $y = \sin x$ and $y = \cos x$ over the interval $0 \leqq x \leqq \pi$. Observe that a particular region of Figure 4.393 seems to

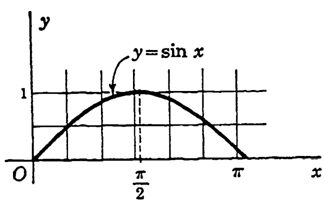


Figure 4.393

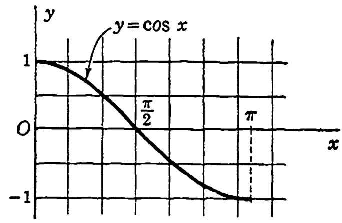


Figure 4.394

fill about two-thirds of the enclosing rectangle and hence that the region ought to have area about $2\pi/3$. Then obtain the first of the formulas

$$\int_0^\pi \sin x\,dx = 2, \qquad \int_0^{\pi/2} \cos x\,dx = 1, \qquad \int_{\pi/2}^{\pi} \cos x\,dx = -1$$

and give an interpretation of the result. Then obtain the second and third formulas and interpret the results in terms of regions in Figure 4.394.

7 Prove that if u and v have continuous derivatives over the interval $a \leqq x \leqq b$, then

$$\int_a^b u(x)v'(x)\,dx = u(x)v(x)\Big]_a^b - \int_a^b v(x)u'(x)\,dx.$$

Hint: Decide how the formula

$$\int_a^b F'(x)\,dx = F(x)\Big]_a^b$$

can be used. *Remark:* The formula to be proved is one of the most useful formulas in the calculus; it is the formula for *integration by parts*.

8 Some of the most important applications of integrals involve inequalities, and we look at an example. Let A be a positive number and start with the fact that

$$1 \leqq e^x \leqq e^A \tag{1}$$

when $0 \leqq x \leqq A$. Replace x by t in (1) and integrate over the interval $0 \leqq t \leqq x$ to obtain, with the aid of Theorem 4.34,

$$x \leqq e^x - 1 \leqq e^A x \qquad (0 \leqq x \leqq A). \tag{2}$$

Replace x by t in (2) and integrate over the interval $0 \leqq t \leqq x$ to obtain

$$\frac{x^2}{2} \leqq e^x - (1 + x) \leqq e^A \frac{x^2}{2} \qquad (0 \leqq x \leqq A). \tag{3}$$

Continue the process to obtain

$$\frac{x^3}{3!} \leqq e^x - \left(1 + x + \frac{x^2}{2!}\right) \leqq e^A \frac{x^3}{3!} \tag{4}$$

$$\frac{x^4}{4!} \leqq e^x - \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}\right) \leqq e^A \frac{x^4}{4!} \tag{5}$$

when $0 \leqq x \leqq A$. *Remark:* Continuation of the process (with the aid of mathematical induction) shows that, for each positive integer n,

$$\frac{x^n}{n!} \leqq e^x - \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}\right) \leqq e^A \frac{x^n}{n!}. \tag{6}$$

While we now have so many other things to do that we shall not look at the details, we can observe that (6) provides a straightforward and foolproof way to obtain decimal approximations to $e^{1/2}$, e, e^2, etcetera, correct to 4 or 40 decimal places. We can discard much of the information in (6) by observing that $x^n/n!$ approaches 0 as $n \to \infty$ and hence that

$$e^x = \lim_{n\to\infty} \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}\right). \tag{7}$$

The formula (7) is the spectacular one, but (6) is often much more useful. We shall learn more about these things later.

9 Applying the idea of the preceding problem to the inequality

$$-1 \leqq \sin x \leqq 1,$$

show that

$$-x \leqq 1 - \cos x \leqq x,$$
$$-\frac{x^2}{2} \leqq x - \sin x \leqq \frac{x^2}{2}$$

when $x > 0$. *Remark:* More extensive information will appear in a problem of Section 8.2.

10 The *Bernoulli functions* $B_0(x)$, $B_1(x)$, $B_2(x)$, $\cdots$ satisfy the conditions

$$B_0(x) = 1 \tag{1}$$

$$B_n'(x) = B_{n-1}(x) \qquad (n = 1,2,3, \cdots) \tag{2}$$

$$\int_0^1 B_n(x)\, dx = 0 \qquad (n = 1,2,3, \cdots) \tag{3}$$

over the interval $-\infty < x < \infty$, except that (2) fails to hold when n is 1 or 2 and x is an integer. They all have period 1, that is, $B_n(x+1) = B_n(x)$ for each n and x. They are all continuous except that $B_1(x)$, the saw-tooth function having the graph shown in Figure 4.395, is discontinuous at the integers. In fact, $B_1(x) = 0$ when x is an integer and

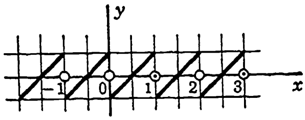


Figure 4.395

(4) $$B_1(x) = x - [x] - \tfrac{1}{2}$$

when x is not an integer. Show that (1) and (2) imply existence of constants $B_0, B_1, B_2, \cdots$ such that, when $0 < x < 1$ and $0! = 1$ as usual,

(5) $$B_0(x) = \frac{B_0}{0!0!}$$

(5.1) $$B_1(x) = \frac{B_0x}{0!1!} + \frac{B_1}{1!0!}$$

(5.2) $$B_2(x) = \frac{B_0x^2}{0!2!} + \frac{B_1x}{1!1!} + \frac{B_2}{2!0!}$$

(5.3) $$B_3(x) = \frac{B_0x^3}{0!3!} + \frac{B_1x^2}{1!2!} + \frac{B_2x}{2!1!} + \frac{B_3}{3!0!}$$

(5.4) $$B_4(x) = \frac{B_0x^4}{0!4!} + \frac{B_1x^3}{1!3!} + \frac{B_2x^2}{2!2!} + \frac{B_3x}{3!1!} + \frac{B_4}{4!0!}$$

and write two more of these formulas. Because of continuity, each of these formulas except (5.1) holds when $0 \leqq x \leqq 1$. The numbers $B_0, B_1, B_2, \cdots$ are the *Bernoulli numbers and*, when $n \geqq 2$,

(6) $$B_n = n!B_n(0).$$

Show that the above formulas can be put in the neater forms

(7) $$0!B_0(x) = B_0$$
(7.1) $$1!B_1(x) = B_0x + B_1$$
(7.2) $$2!B_2(x) = B_0x^2 + 2B_1x + B_2$$
(7.3) $$3!B_3(x) = B_0x^3 + 3B_1x^2 + 3B_2x + B_3$$
(7.4) $$4!B_4(x) = B_0x^4 + 4B_1x^3 + 6B_2x^2 + 4B_3x + B_4$$

involving binomial coefficients and write two more of these equations. Use (3) to show that, when $n \geqq 2$,

$$B_n(1) - B_n(0) = \int_0^1 B_n'(x)\,dx = \int_0^1 B_{n-1}(x)\,dx = 0$$

and hence

(8) $$B_n(1) = B_n(0).$$

Use (1) and (7) to show that $B_0 = 1$. Then use (7.2) and (8) to show that $B_1 = -\frac{1}{2}$. Then use (7.3) and (8) to show that $B_2 = \frac{1}{6}$. Then calculate one or two more Bernoulli numbers. *Remark:* Bernoulli functions and numbers have important applications and some people know very much about them. It can be shown that $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$, $B_3 = 0$, $B_4 = -\frac{1}{30}$, $B_5 = 0$,

$B_6 = \frac{1}{42}$, $B_7 = 0$, $B_8 = -\frac{1}{30}$, and that $|B_{2n}|$ is very large when n is large. Some books, particularly those that give a few formulas involving Bernoulli numbers but do not treat Bernoulli functions, use notation which conflicts with the notation used above.

11 Prove that if f is integrable over the interval $0 \leqq x \leqq 1$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k}{n}\right) = \int_0^1 f(x)\,dx. \tag{1}$$

Solution: To keep all of the bewitching mysticism of mysterious mathematics out of our solution, let ϵ be a given positive number. Choose a positive number δ such that

$$\left| \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k - \int_0^1 f(x)\,dx \right| < \epsilon \tag{2}$$

whenever the sum is a Riemann sum formed for a partition P of the interval $0 \leqq x \leqq 1$ for which $|P| < \delta$. Let N be an integer for which $N > 2$ and $N > 1/\delta$. Let n be an integer greater than N. Let P_n be a partition of the interval $0 \leqq x \leqq 1$ into n *equal* subintervals each having length $1/n$. Then $x_k = k/n$ for each k. Let $x_k^* = x_k$ so that $x_k^* = k/n$ for each k. Since $\Delta x_k = x_k - x_{k-1} = 1/n$ for each k, we see that $|P_n| = 1/n$. Since $n > N$, we have $1/n < 1/N$ and hence $1/n < \delta$. Therefore, $|P_n| < \delta$ and (2) holds when the sum is the Riemann sum formed for the partition P_n. But for the partition P_n we have $x_k^* = k/n$ and $\Delta x_k = 1/n$, so

$$\sum_{k=1}^{n} f(x_k^*)\,\Delta x_k = \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right). \tag{3}$$

It follows that

$$\left| \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right) - \int_0^1 f(x)\,dx \right| < \epsilon \tag{4}$$

when $n > N$, and this gives the desired conclusion (1).

12 The basic formula (1) of Problem 11 has numerous quite astonishing applications. Letting s be a nonnegative number and letting $f(x) = x^s$, use the formula to prove that

$$\lim_{n\to\infty} \frac{1^s + 2^s + 3^s + \cdots + n^s}{n^{s+1}} = \frac{1}{1+s}. \tag{1}$$

Write the formulas to which this reduces when $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2$, and 3. *Remark:* Textbooks that specialize in proofs by mathematical induction give the formulas

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2} \tag{2}$$

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} \tag{3}$$

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}. \tag{4}$$

With the aid of these formulas, it is easy to verify (1) for the cases in which s is 1 and 2 and 3. In fact, Archimedes did it. The formulas of this problem have tremendous importance in the history of science because they stimulated interest in limits of sums that culminated in the invention of "the calculus" by Leibniz and Newton.

13 Letting $f(x) = (1 + x)^s$, where s is a constant for which $s \neq -1$, derive the formula

$$\lim_{n\to\infty} \frac{(n+1)^s + (n+2)^s + \cdots + (n+n)^s}{n^{s+1}} = \frac{2^{s+1} - 1}{s+1}.$$

Write the formulas to which this reduces when s has the values -2, $-\frac{1}{2}$, 0, $\frac{1}{2}$, 1, and 2.

14 Letting $f(x) = (1 + x)^{-1}$, derive the formula

$$\lim_{n\to\infty} \left(\frac{1}{n+1} + \frac{1}{n+2} + \frac{1}{n+3} + \cdots + \frac{1}{n+n}\right) = \log 2.$$

15 Letting $f(x) = 2x/(1 + x^2)^s$, where s is a constant for which $s \neq -1$, derive the formula

$$\lim_{n\to\infty} n^{2s-2}\left[\frac{1}{(n^2+1^2)^s} + \frac{2}{(n^2+2^2)^s} + \frac{3}{(n^2+3^2)^s} + \cdots + \frac{n}{(n^2+n^2)^s}\right] = \frac{1}{2(s-1)}\left[1 - \frac{1}{2^{s-1}}\right].$$

Write the formulas to which this reduces when s has the values -2, -1, $-\frac{1}{2}$, 0, $\frac{1}{2}$, and 1.

16 Letting $f(x) = 2x/(1 + x^2)$, derive the formula

$$\lim_{n\to\infty} n^2\left[\frac{1}{n^2+1^2} + \frac{2}{n^2+2^2} + \frac{3}{n^2+3^2} + \cdots + \frac{n}{n^2+n^2}\right] = \frac{1}{2}\log 2.$$

17 Letting $f(x) = \dfrac{1}{1+x^2}$ and borrowing the fact that $\displaystyle\int_0^1 \frac{1}{1+x^2}\,dx = \frac{\pi}{4}$, derive the implausible formula

$$\lim_{n\to\infty} n\left(\frac{1}{n^2+1^2} + \frac{1}{n^2+2^2} + \frac{1}{n^2+3^2} + \cdots + \frac{1}{n^2+n^2}\right) = \frac{\pi}{4}.$$

18 Persons are sometimes credited with substantial knowledge of calculus when they can simplify

$$(1) \qquad \frac{d}{dx}\int_0^{x^2} e^{-t^2}\,dt.$$

The problem can and should be solved by noticing that putting $f(t) = e^{-t^2}$ enables us to use the fundamental theorem of the calculus (Theorem 4.35) to obtain

$$(2) \qquad \frac{d}{du}\int_0^u f(t)\,dt = f(u).$$

When (2) holds and u is a differentiable function of x, we can use the chain rule to obtain

$$(3) \qquad \frac{d}{dx}\int_0^u f(t)\,dt = \left[\frac{d}{du}\int_0^u f(t)\,dt\right]\frac{du}{dx} = f(u)\,\frac{du}{dx}.$$

Use these ideas to show that (1) is $2xe^{-x^4}$.

19 Letting

$$F(x) = \int_0^{x^2} e^{-t}\,dt,$$

use the ideas of the preceding problem to obtain a simple formula for $F'(x)$. Then find $F(x)$, that is, find a simpler expression for $F(x)$, and differentiate it to obtain $F'(x)$. Make the results agree.

20 Prove that if f is continuous and u and v are differentiable, then

$$\frac{d}{dx}\int_v^u f(t)\,dt = f(u)\frac{du}{dx} - f(v)\frac{dv}{dx}.$$

Hint: Use the formula

$$\int_v^u f(t)\,dt = \int_0^u f(t)\,dt - \int_0^v f(t)\,dt$$

and the ideas of Problem 18.

21 Supposing that λ is a positive constant, $x > 0$, and

$$F(x) = \int_x^{\lambda x} \frac{1}{t}\,dt,$$

show that $F'(x) = 0$ without use of the formula $\int \frac{1}{t}\,dt = \log t + c$.

4.4 *Areas and integrals*

We all know what is meant by a rectangular region R having base length b and height h. When the x axis of a rectangular coordinate system is parallel to the base, R is the set of points (x,y) for which $x_0 \leqq x \leqq x_0 + b$ and $y_0 \leqq y \leqq y_0 + h$ as in Figure 4.41. We are all familiar with the idea that R has an area and that this area is bh, the product of the base length and the height. There is an old-fashioned view that this matter is quite simple, but modern mathematicians, like modern atomic physicists, find that there is much to be learned about things that our ancestors thought were simple. It is quite absurd to presume that it is easy to prove that the area of R is bh; in fact it is quite absurd to presume that it is possible to prove that there is a number (bh or not) which is the area of R unless we have some definitions or postulates or something upon which proofs can be based. We

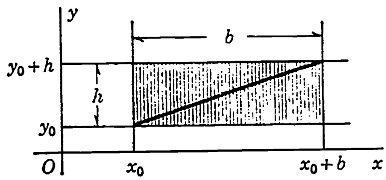


Figure 4.41

escape this awkward situation with the aid of a definition designed for the purpose.

Definition 4.42 *If R is a rectangular region having base length b and height h, then the product bh is called the area of R. This area is denoted*† *by the symbol $|R|$ so that $|R| = bh$.*

This takes care of the matter of areas of rectangular regions, but we are not yet out of trouble. Let T be the triangular set consisting of those points within the rectangular region of Figure 4.41 which lie on and beneath the diagonal drawn there. When we try to decide whether there is a number which is the area of T, we find that we still need definitions or postulates or something before we can do anything. If we try to take care of triangular regions, circular disks, circular sectors, and sets of other special types by hordes of special definitions, we will find ourselves forever wallowing in confusion. While students in elementary calculus courses are normally not expected to know much if anything about the matter, we should at least know that our friend Lebesgue constructed a theory of area which is usually called the theory of Lebesgue two-dimensional measure. This eliminates the confusion and is now very important in applied as well as in pure mathematics. We should not be injured and may possibly be benefited by a brief look at the Lebesgue theory. Let S be a set of points (x,y) which is contained in but does not completely fill a rectangle R. Figure 4.43 may be helpful, but may also be misleading because the set S need not look at all like the one shown in the figure. Let S' be the set of points in R but not in S.

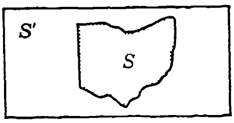


Figure 4.43

Definition 4.44 *The set S is said to have area (or two-dimensional Lebesgue measure) $|S|$ if $|S|$ is a number such that to each $\epsilon > 0$ there correspond* (i) *a countable collection $R_1, R_2, R_3, \cdots$ of rectangular regions such that each point of S lies in at least one of these regions and*

$$(4.45) \qquad |R_1| + |R_2| + \cdots + |R_n| < |S| + \epsilon \quad (n = 1, 2, 3, \cdots)$$

and (ii) *another countable collection $R_1', R_2', R_3', \cdots$ of rectangular regions such that each point of S' lies in at least one of these regions and*

$$(4.46) \qquad |R_1'| + |R_2'| + \cdots + |R_n'| < |R| - |S| + \epsilon \quad (n = 1,2,3, \cdots).$$

† This notation accords with a general principle with which we are slowly becoming acquainted. If Q is a number or a partition or a point set or perhaps even an assertion or a crate of oranges, we expect $|Q|$ to be a real nonnegative number which is associated with Q in some particular way and is, in some sense or other, a measure or a norm or a value of Q. The simplest useful example is that in which Q is a real number and $|Q|$ is its absolute value. When applications of areas are involved, it is often necessary to recognize that h and k are numbers representing lengths measured in particular units (say centimeters) and that the area is a number of appropriate "square units" (say, square centimeters).

It is possible to describe complicated rules for constructing sets S for which no such number $|S|$ exists, and we say that such sets do not possess area† (or are nonmeasurable). However, such sets are much more complicated than those that appear in this book. This discussion of areas will have served a purpose if it provides a reason for acceptance of the fact that the theory of area is much more complicated than the theory of Riemann integrals and that intuitive ideas about areas do not provide a satisfactory basis for proofs of theorems about Riemann integrals. We can, however, be reassured by the facts that many of the results of the theory of area are thoroughly simple and that *they are in complete agreement with all of the results we shall obtain by use of Riemann integrals.* We shall not use Riemann integrals to obtain illusory information about areas of sets that do not possess areas. More information about this matter will appear in Section 5.7.

In what follows, we suppose that M, a, and b are constants and that f is a function, Riemann integrable over $a \leqq x \leqq b$, for which $0 \leqq f(x) \leqq M$ when $a \leqq x \leqq b$. Let S be the set of points (x,y) for which $a \leqq x \leqq b$ and $0 \leqq y \leqq f(x)$. The set S may look more or less like the sets shown in Figures 4.471 and 4.472. In each case we can describe S as the set of

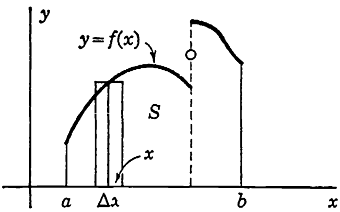


Figure 4.471

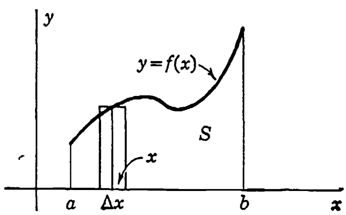


Figure 4.472

points or part of the plane or region bounded on the left and right by the graphs of the equations $x = a$ and $x = b$ and bounded below and above by the graphs of the equations $y = 0$ and $y = f(x)$. In case f is continuous and the graph of $y = f(x)$ looks like that shown in Figure 4.472, we can comfortably describe S as the region bounded by the graphs of the four equations.‡

† Newspapers and magazines keep us permanently aware of the fact that there are inadequacies in old-fashioned intuitive physical theories of matter and that these intuitive theories do not provide an adequate basis for modern physics. Since these newspapers and magazines keep us quite generally uninformed about theories of areas and volumes, it may be necessary to consult Appendix 2 at the end of this book to learn that there are bugs in intuitive theories of areas and volumes.

‡ Of course climatologists who talk about areas of abundant rainfall, and philosophers who talk about areas of scientific thought, could be expected to call S the area bounded by the graphs. But in mathematics and perhaps even in climatology (we never know about philosophy) an area is always a number and scientists do not, in their most brilliant moments, call S an area.

When an enlightened scientist must calculate the area $|S|$ of S, he writes

$$|S| = \lim \sum f(x)\,\Delta x = \int_a^b f(x)\,dx \tag{4.48}$$

and, at least in simple cases, evaluates the integral with the aid of Theorem 4.38. The procedure by which (4.48) is obtained must now be explained. The first step is to sketch an appropriate figure which will look more or less like Figure 4.471 or Figure 4.472. The next step is to make a partition P of the interval $a \leqq x \leqq b$ into subintervals, but we do not bother to draw more than one of the subintervals. Without bothering with subscripts and stars, we let Δx denote the length of the interval and let x be a point of the interval. We then draw the rectangle whose width is Δx and whose height is $f(x)$. The first step in building the formula (4.48) is to write $f(x)\,\Delta x$, because this is the area of the rectangle (or rectangular region). We then tell ourselves that this area is a good approximation to the area of the part of S that lies between the vertical sides of the rectangle, and, while this is no time to get excited about the matter, we could tell ourselves that the two areas might be exactly equal if we choose the x shrewdly enough. The next step is to add the area of the rectangle we have drawn to the areas of the other rectangles which we have not drawn to obtain $\Sigma f(x)\,\Delta x$. Even if we did not know in advance that $\lim \Sigma f(x)\,\Delta x$ exists, we should have a feeling that $\Sigma f(x)\,\Delta x$ should be near $|S|$ whenever the numbers Δx are all small (that is, whenever the norm of P is near zero) and hence we should write

$$|S| = \lim \Sigma f(x)\,\Delta x. \tag{4.481}$$

The final step is to recognize that the right side of this equation is the limit of Riemann sums and hence is the Riemann integral in (4.48).

The ritual involving partitioning (or splitting up), estimating, summing (or adding), and taking a limit to obtain a Riemann integral equal to a number in which we are interested is known as "the process for setting up the integral." The ability to "set up integrals" efficiently and correctly is very valuable, and problems in calculus textbooks that require the finding of areas are designed to promote abilities in this art. Students cannot know, unless they are told, that they are wasting their time if they never bother to set up integrals but only use remembered formulas to calculate areas and volumes and the ubiquitous moments of inertia.

Problems 4.49

1 Figure 4.491 shows graphs of two equations $y = f_1(x)$ and $y = f_2(x)$ which intersect at the points $(-2,-6)$, $(0,0)$, and $(2,6)$. The graphs bound two regions R_1 and R_2. Use partitions and Riemann sums to obtain the formulas

$$|R_1| = \int_{-2}^{0} [f_1(x) - f_2(x)]\, dx, \qquad |R_2| = \int_{0}^{2} [f_2(x) - f_1(x)]\, dx,$$
$$|R_1| + |R_2| = \int_{-2}^{2} |f_1(x) - f_2(x)|\, dx.$$

Remark: The widths and heights of rectangles are always positive, and mistakes in sign are undesirable. When hasty calculations indicate that an area or a population of a city is negative, the calculations should be examined.

2 The graphs in Figure 4.491 are graphs of $y = 3x$ and $y = x^3 - x$. Find $|R_1| + |R_2|$, this being the sum of the areas of the two regions bounded by the graphs. *Ans.:* 8.

3 With Figure 4.492 to provide assistance, make a partition of the interval $0 \leqq x \leqq 2$ to obtain the area $|S_1|$ of the set S_1 bounded by the graphs of $y = 0$, $x = 2$, and $y = x^2$. Try to repair

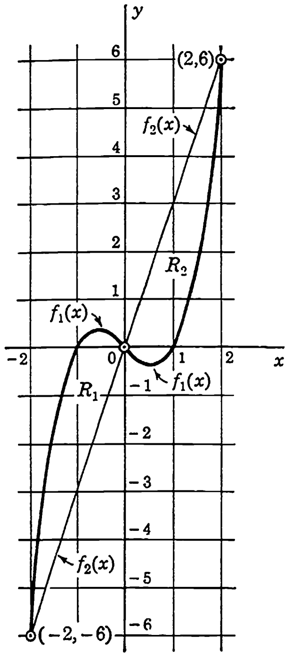


Figure 4.491

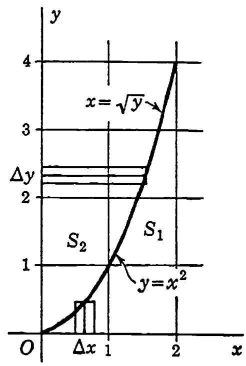


Figure 4.492

the work if the result does not have reasonable agreement with an estimate made by counting squares and partial squares included by S_1. Then interchange the roles of x and y to find the area $|S_2|$ of the set S_2 bounded by the graphs of $x = 0$, $y = x^2$, and $y = 4$. Make a partition of the interval $0 \leqq y \leqq 4$ and be sure that the correct integrand and limits of integration appear in the calculation

$$|S_2| = \lim \sum f(y)\, \Delta y = \int_{y_1}^{y_2} f(y)\, dy.$$

In this case also, try to repair the work if the result clashes with the result of counting squares. Finally, have another look at Figure 4.492 and see what $|S_1| + |S_2|$ should be.

4 Referring again to Figure 4.492, obtain $|S_2|$ by starting with a partition of the interval $0 \leqq x \leqq 2$ and using an estimate of the area of the part of S_2 that stands above the interval of length Δx (or Δx_k) containing the point x (or x_k).

5 Use the technique of the text to find the area of the triangular patch bounded by the lines having the equations $y = 2x$, $y = 0$, and $x = 3$. Check your answer by use of elementary geometry.

6 Let A be the area of the region bounded by the x axis and the graph of the equation $y = x(1 - x)$. Sketch an appropriate graph showing a sample rectangle and fill in the details involving the formula

$$A = \lim \sum x(1 - x)\,\Delta x = \int_0^1 (x - x^2)\,dx = \tfrac{1}{6}.$$

7 Find the area of the region bounded by the coordinate axes and the graph of the equation $y = x^3 - 8$. *Ans.:* 12.

8 Find the area of the part of the plane bounded by the graphs of the equations $y = x^3 - 3x$ and $y = x$. *Ans.:* 8.

9 Find the area of the region bounded by the graphs of the equations $y = x$, $y = 2x$, and $y = x^2$. *Ans.:* $\frac{7}{6}$.

10 Find the area of the region in the first quadrant bounded by the x axis and the graphs of the equations $y = x$ and $y = 2 - x^2$. *Ans.:* $(8\sqrt{2} - 7)/6$.

11 Let A be the area of the part of the plane which lies between the lines having the equations $x = \pi$ and $x = 2\pi$ and is bounded by the x axis and the graph of the equation $y = \sin x$. Sketch an appropriate graph showing a sample rectangle and, observing that the height of the rectangle is the positive number $-\sin x$ (not the negative number $\sin x$), fill in the details involving the formula

$$A = \lim \sum (-\sin x)\,\Delta x = -\int_\pi^{2\pi} \sin x\,dx = 2.$$

12 Someday we will be able to show that the graph of the equation $x^{1/2} + y^{1/2} = a^{1/2}$ is, when a is a positive constant, a part of a parabola. Find the area of the region bounded by the graph and the coordinate axes. *Ans.:* $a^2/6$.

13 Is the area of the region bounded by the graphs of the equations

$$y = x^3 + x^2, \qquad y = x^3 + 1$$

the same as the area of the region bounded by the graphs of the equations

$$y = x^2, \qquad y = 1?$$

14 The graph of each of the following equations contains a loop; determine the nature of the graph and find the area of the region bounded by the loop, it being assumed that a is a positive constant.

(*a*) $y^2 = x(a - x)^2$ *Ans.:* $\frac{8}{15}a^{5/2}$

(*b*) $y^2 = x(x - a)^2$ *Ans.:* $\frac{8}{15}a^{5/2}$

15 The graphs of the equations $y = \frac{1}{2}x^2$ and $y = x + 4$ bound a region R. With the aid of a reasonably good figure, make an estimate of the area $|R|$ of R. Then find $|R|$ by making partitions of an appropriate part of the x axis so that vertical strips appear in the calculation. Then find $|R|$ by a method in which horizontal strips appear. Make the results agree with each other and use your estimate to provide assurance that the two answers are reasonable.

16 Let A be the area of the circular disk of radius a shown in Figure 4.493. Explain the ideas associated with the calculation

$$A = \lim \sum 2\pi x\,\Delta x = 2\pi \int_0^a x\,dx = \pi a^2.$$

Hint: Think of the ring between the two inner circles as being a ribbon of width Δx and length $2\pi x$, the length being (by definition of π) the circumference of a circle of radius x.

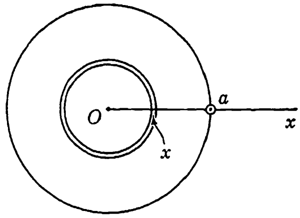


Figure 4.493

17 Sketch graphs of $\sin x$ and $\cos x$ over the interval $0 \leqq x \leqq \pi$ and then, with the aid of this information, sketch graphs of $\sin^2 x$ and $\cos^2 x$. Use these graphs to obtain a reason why it should be true that

$$(1) \qquad \int_0^\pi \sin^2 x\,dx = \int_0^\pi \cos^2 x\,dx.$$

Note also that

$$(2) \qquad \int_0^\pi (\sin^2 x + \cos^2 x)\,dx = \int_0^\pi 1\,dx = \pi.$$

What can we now conclude about the integrals in (1)? Taking a totally different tack, use the formulas

$$(3) \qquad \sin^2 x = \frac{1 - \cos 2x}{2}, \qquad \cos^2 x = \frac{1 + \cos 2x}{2}$$

to evaluate the integrals in (1). Make all of the results agree.

18 Prove the formula

$$\int_0^a \sqrt{a^2 - x^2}\,dx = \frac{\pi a^2}{4}$$

by observing that the integrand is nonnegative and constructing a region of which the integral is the area. *Hint:* Let $y = \sqrt{a^2 - x^2}$ and, after tinkering with this equation, draw an appropriate figure.

19 Let a and b be constants for which $0 < b < a$. Show that if $y \geqq 0$, $0 \leqq x \leqq a$, and

$$(1) \qquad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

then

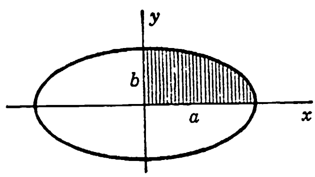
$$y = \frac{b}{a}\sqrt{a^2 - x^2}.$$

Let S be the set of points inside the graph of (1); as we shall learn later, the graph is an ellipse. With the aid of Figure 4.494 show that

$$|S| = \frac{4b}{a}\int_0^a \sqrt{a^2 - x^2}\,dx.$$

With the aid of the preceding problem, show that $|S| = \pi ab$. This is a result that many people remember: the area of a circular disk is πaa and the area of an elliptic disk is πab.

Figure 4.494



20 This problem is interesting because it shows how a basic formula involving areas (a well-known formula which we have not yet proved) can be used to obtain preliminary derivations of formulas involving trigonometric and inverse trigonometric functions. Rigorous derivations will be given in Chapter 8. Supposing that $0 < t < a$, construct and look at an appropriate figure to derive the formula

$$(1) \qquad \int_0^t \sqrt{a^2 - x^2}\,dx = \tfrac{1}{2}t\sqrt{a^2 - t^2} + \tfrac{1}{2}a^2 \sin^{-1}\frac{t}{a}$$

in which the first term on the right is the area of a particular triangle and the last term is the area of a circular sector having radius a and central angle θ, where $\theta = \sin^{-1}(t/a)$ and $0 < \theta < \pi/2$. Anyone who is short on information about areas of circular sectors is reminded that the area of a sector having central angle θ is, as it ought to be, the product of $\theta/2\pi$ and the area πa^2 of the whole circle. We can suddenly become interested in (1) if we realize that we have theorems and rules that enable us to write formulas for the derivatives with respect to t of everything in it except the last term and hence that we can obtain a formula for the derivative of the last term. To capitalize this idea, put (1) in the form

$$(2) \qquad \sin^{-1}\frac{t}{a} = \frac{1}{a^2}\left[2\int_0^t \sqrt{a^2 - x^2}\,dx - t\sqrt{a^2 - t^2}\right]$$

and then differentiate and simplify results to obtain the formula

$$(3) \qquad \frac{d}{dt}\sin^{-1}\frac{t}{a} = \frac{1}{\sqrt{a^2 - t^2}}.$$

Remark: We invest a moment to look at the formula

$$(4) \qquad \frac{d}{dt}\sin^{-1} t = \frac{1}{\sqrt{1 - t^2}}$$

to which (3) reduces when $a = 1$. At least in the case where $0 < t < 1$ and $0 < \theta < \pi/2$, trigonometry books emphasize the fact that the angle θ of Figure 4.495 is "the angle whose sine is t" or "the inverse sine of t," so that $\theta = \sin^{-1} t$

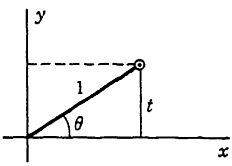


Figure 4.495

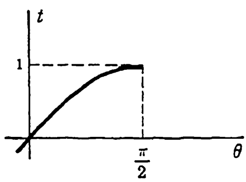


Figure 4.496

and $t = \sin\theta$. The graph in Figure 4.496 shows how θ and t are related. The relation (4) is equivalent to the relation

$$(5) \qquad \lim_{\Delta t \to 0} \frac{\Delta\theta}{\Delta t} = \frac{1}{\sqrt{1 - t^2}}$$

and this is equivalent to the relation

$$\lim_{\Delta\theta \to 0} \frac{\Delta t}{\Delta \theta} = \sqrt{1 - t^2} = \sqrt{1 - \sin^2 \theta} = \cos \theta \tag{6}$$

or to the first of the formulas

$$\frac{d}{d\theta} \sin \theta = \cos \theta, \qquad \frac{d}{d\theta} \cos \theta = -\sin \theta. \tag{7}$$

The second follows from the first and the calculation

$$\frac{d}{dx} \cos \theta = \frac{d}{dx} \sin \left(\frac{\pi}{2} - \theta\right) = -\cos \left(\frac{\pi}{2} - \theta\right) = -\sin \theta. \tag{8}$$

21 Show how formulas of the preceding problem can be used to obtain the integration formulas

$$\int \sqrt{a^2 - x^2}\, dx = \tfrac{1}{2}x \sqrt{a^2 - x^2} + \tfrac{1}{2}a^2 \sin^{-1} \frac{x}{a} + c$$

$$\int \frac{1}{\sqrt{a^2 - x^2}}\, dx = \sin^{-1} \frac{x}{a} + c.$$

Then keep in contact with the external world by finding these formulas in your book of tables.

22 Sketch a few figures which illustrate applications of the following fact. If f is integrable (and hence bounded) over $a \leqq x \leqq b$, we can choose a positive constant B such that $f(x) + B > 0$ when $a \leqq x \leqq b$ and write the formula

$$\int_a^b f(x)\, dx = \int_a^b [f(x) + B]\, dx - \int_a^b B\, dx,$$

which shows that $\int_a^b f(x)\, dx$ is the result of subtracting the area of a rectangle from the area of the set of points (x,y) for which $a \leqq x \leqq b$ and $-B \leqq y \leqq f(x)$. *Remark:* This problem and the next provide ways of reducing questions involving integrals to questions involving integrals with nonnegative integrands.

23 Sketch a few figures which illustrate applications of the following fact. If f is integrable (and hence bounded) over $a \leqq x \leqq b$, so also are the functions g and h defined by

$$g(x) = \frac{|f(x)| + f(x)}{2}, \qquad h(x) = \frac{|f(x)| - f(x)}{2}.$$

Moreover, $g(x) \geqq 0$, $h(x) \geqq 0$, $f(x) = g(x) - h(x)$, $|f(x)| = g(x) + h(x)$, and

$$\int_a^b f(x)\, dx = \int_a^b g(x)\, dx - \int_a^b h(x)\, dx$$

$$\int_a^b |f(x)|\, dx = \int_a^b g(x)\, dx + \int_a^b h(x)\, dx.$$

24 If $f(x) = |x|$, then $f'(x) = \operatorname{sgn} x$ except when $x = 0$. When $a < 0 < b$, Theorem 4.37 does not guarantee correctness of the formula

$$\int_a^b \operatorname{sgn} x\, dx = |x|\Big]_a^b = |b| - |a|,$$

but the formula may be correct anyway. What are the facts? *Ans.:* The formula is correct.

25 This remark is dedicated to a distinguished professor in a distinguished university in New Jeisey. He claimed that it does not make sense to ask a student to evaluate the integral $\int_0^2 x^3\, dx$. The man was right. The integral is a number, the limit of Riemann sums, and the number is 4. Thus, the man was insisting that it does not make sense to ask a student to evaluate 4. What the foxy professor really wanted to do was to emphasize the fact that $\int_0^2 x^3\, dx$ is something more than some black ink on white paper. It is a number. There are times when the thing is called a symbol, but it is not a symbol. The fact that $\left[\int_0^2 x^3\, dx\right]^2 = 16$ would be hard to explain if the thing were considered to be a symbol because we do not square symbols to get 16; we square numbers to get 16. We must agree that we should know what we are doing when we are asked to "evaluate" $\int_0^2 x^3\, dx$ and then go to work to find that the "answer" is 4. A few thoughts about these matters may even pay off sometime.

4.5 *Volumes and integrals* It could hardly be expected that fundamental ideas and definitions involving volumes of sets in E_3 could be simpler than the corresponding ideas and definitions involving areas of sets in E_2. In the best treatments of the subject, the volume of a set is its three-dimensional Lebesgue measure. The theory begins modestly with the definition which asserts that the volume V of a rectangular parallepiped (or brick or three-dimensional interval) having length a, width b, and height c is the product of the dimensions, so that $V = abc$. In the theory of volumes, bricks play the same role that rectangles play in the theory of areas of sets in E_2. It turns out that each bounded set in E_3 that we shall dream of considering has associated with it a number which is the volume of the set. If two of our sets S_1 and S_2 are such that S_1 is a subset of S_2, which means that each point of S_1 is also a point of S_2, we can be sure that the volume $|S_1|$ of S_1 is less than or equal to the volume $|S_2|$ of S_2. If a set S is composed of two parts S_1 and S_2 which have no points in common, we can be sure that $|S| = |S_1| + |S_2|$. If one of our sets S_1 has a volume $|S_1|$ and if S_2 is another set congruent to S_1, then S_2 has a volume and $|S_2| = |S_1|$. Appendix 2 at the end of this book shows that the theory of volumes is (like the theory of "solid" physical matter) not as simple as the naive believe. While a full discussion of volumes lies far beyond the scope of this book, the theory of Lebesgue measure in E_3 justifies all of the methods we shall use for finding volumes.

We now illustrate the "slab method" for finding volumes of three-dimensional sets that are commonly called "solids." With the expectation that the method will be fully understood and applied to find volumes of other solids, we find the volume of the solid cone of Figure 4.51 which consists of the points in E_3 lying between the planes $x = 0$ and $x = h$ and inside or on the conical surface. When we are not required to explain the details of the method, we solve this problem in two lines by writing

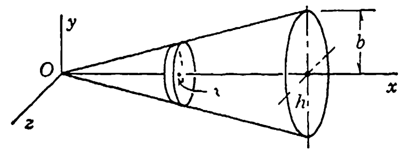


Figure 4.51

$$(4.52)\qquad V = \lim \sum A(x)\,\Delta x = \int_0^h A(x)\,dx$$
$$= \int_0^h \pi\left(\frac{b}{h}x\right)^2 dx = \frac{\pi b^2}{h^2}\frac{x^3}{3}\bigg]_0^h = \tfrac{1}{3}\pi b^2 h.$$

Even when we are not required to give explanations to someone else, we do not write this without talking to ourselves. We make a partition P of the interval $0 \leqq x \leqq h$, but we draw only one subinterval having length Δx and let x be a point of the subinterval. Planes perpendicular to the x axis at the ends of the interval have between them a part of the solid that we can call a slab. Let $A(x)$ be the area of the section in which the solid intersects the plane which contains the point we have selected and is perpendicular to the x axis. In case $|P|$ is small, the number $A(x)\,\Delta x$ is exactly equal to the volume of our slab or is a good approximation to it. We next write

$$(4.53)\qquad \Sigma A(x)\,\Delta x$$

and tell ourselves that this is either exactly or approximately the sum of the volumes of the slabs and hence is exactly or approximately V. Hence it should be true that

$$(4.54)\qquad V = \lim \Sigma A(x)\,\Delta x.$$

But the right side of this formula is the integral in (4.52). Our next step is to observe that $A(x)$ is the area of a circular disk whose radius y is such that $y/x = b/h$ and hence $y = (b/h)x$. Thus

$$(4.55)\qquad A(x) = \pi\left(\frac{bx}{h}\right)^2 = \frac{\pi b^2}{h^2}x^2.$$

With this information, we can quickly complete the work in (4.52).

Observe that it would not be easy to find the volume of the solid cone of Figure 4.51 by employing slabs resulting from a partition of the interval $-b \leqq y \leqq b$ of the y axis. The difficulty resides in the fact that planes perpendicular to the y axis intersect the solid in plane sets the areas of which are not easily found.

Finally, we illustrate the "cylindrical shell method" for finding volumes of solids by finding the volume of a solid cone in another way. We consider the solid cone to be the solid obtained by rotating, about the x axis, the triangular region T in which the solid cone intersects the first quadrant of the xy plane. This region appears in Figure 4.56. This time we make a partition P of the interval $0 \leqq y \leqq b$ of the y axis. When $y_{k-1} \leqq y_k^* \leqq y_k$, the lines in the xy plane having the equations $y = y_{k-1}$ and $y = y_k$ cut from T a strip approximating a rectangular region of length $[h - (h/b)y_k^*]$ and width Δy_k. When this rectangular region is rotated about the x axis, it generates a cylindrical shell resembling a tin tomato can from which both top and bottom have been removed. Different points in this shell have different distances from the x axis, but when $|P|$, the norm of P, is small, these distances are all nearly y_k^*. Taking $2\pi y_k^*$ to be the circumference of the shell, we use the number

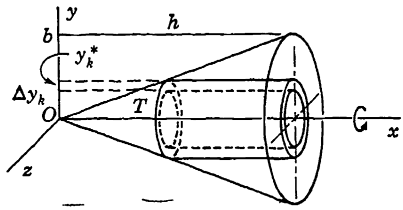


Figure 4.56

$$2\pi y_k^* \left(h - \frac{h}{b} y_k^*\right) \tag{4.57}$$

to approximate the area of the shell. Multiplying this by Δy_k, the thickness of the shell, gives an approximation to the volume of the shell. This leads us to the formulas

$$V = \lim \sum 2\pi y_k^* \left(h - \frac{h}{b} y_k^*\right) \Delta y_k \tag{4.571}$$

and

$$V = 2\pi h \int_0^b y\left(1 - \frac{1}{b} y\right) dy = \tfrac{1}{3}\pi b^2 h. \tag{4.58}$$

For finding volumes of cones, the slab method produces answers much more easily and quickly than the cylindrical shell method. Most of the problems at the end of this section should be solved by the slab method, but the cylindrical shell method sometimes works better than the slab method.

Problems 4.59

1 Find the volume of the spherical solid (or ball) of radius a which has its center at the origin. Find out whether it is easier to partition the whole interval $-a \leqq x \leqq a$ or to take double the result of partitioning the interval $0 \leqq x \leqq a$. *Remark:* Scientists should always remember that the volume is $\frac{4}{3}\pi a^3$.

2 Supposing that $0 \leqq h \leqq 2r$, find the volume V of the segment of the spherical ball with center at the origin and radius r which lies between the planes having the equations $x = r - h$ and $x = r$. *Ans.:*

$$V = \pi \int_{r-h}^{r} (r^2 - x^2)\,dx = \tfrac{1}{3}\pi h^2(3r - h).$$

3 The region in the first quadrant bounded by the graphs of the equations $y = kx^2$, $x = 0$, and $y = A$ is rotated about the y axis to produce a solid S which is a part of a solid paraboloid like the nose of a bullet. Show that $|S|$, the volume of S, is exactly half the volume of a solid circular cylinder having the same base and altitude.

4 The region bounded by the graphs of the equations $y = kx^2$ and $y = A$ is rotated about the line having the equation $y = A$. Find the volume of the resulting solid. *Ans.:*

$$\frac{16\pi A^2}{15}\sqrt{\frac{A}{k}}.$$

5 A region R is bounded by the graphs of the equations $xy = 1$, $y = 0$, $x = a$, and $x = b$ for which $0 < a < b$. Find the volume $|S|$ of the solid S obtained by rotating R about the x axis. *Ans.:* $\pi/a - \pi/b$.

6 The region bounded by the graphs of the equations $x = 1$, $x = 2$, $y = 0$, and $y = \frac{2}{3}\sqrt{9 - x^2}$ is rotated about the x axis. Find the volume of the resulting solid. *Ans.:* $80\pi/27$.

7 The region bounded by the line and hyperbola having the equations $x + y = 5$ and $xy = 4$ is rotated about the y axis. Find the volume V of the solid thus generated. *Ans.:* 9π.

8 Let a cylindrical shell (which resembles the part of a tomato can remaining after the top and bottom have been removed) have length L and have inner and outer radii r_{k-1} and r_k. Supposing as usual that $\Delta r_k = r_k - r_{k-1}$, prove that the volume of the shell is

$$(2\pi r_k^*)L\,\Delta r_k,$$

where r_k^* is the number defined by $r_k^* = \frac{1}{2}(r_{k-1} + r_k)$.

9 Set up two different integrals for the volume of the solid torus (or anchor ring) obtained by rotating the circular disk of Figure 4.591 about the y axis. First make a partition of the interval $0 \leqq y \leqq a$ of the y axis and estimate volumes of washers (things normally associated with nuts and bolts). Then make a partition of the interval $b - a \leqq x \leqq b + a$ and estimate volumes of cylindrical shells (things which, if they had tops and bottoms, would be tin cans). Evaluate one of the integrals. *Remark:* The correct answer agrees with the result of applying a famous old theorem which says that the volume of the solid of revolution is the product of the area of the set rotated and the distance the centroid (in this case, the center) goes. The theorem is the *theorem of Pappus.*

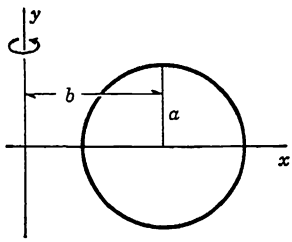


Figure 4.591

10 Find, in two or three different ways, the volume of the solid obtained by replacing the disk of the preceding problem by the square with horizontal and vertical sides tangent to the disk. One of the methods is suggested by the remark at the end of the preceding problem.

11 Find the volume of the solid obtained by rotating, about the y axis, the region bounded by the graphs of the equations $y = 3x^2$ and $y = 12$. *Ans.:* 24π.

12 Find the volume of the solid generated by rotating, about the x axis, the region in the first quadrant bounded by the graphs of the equations $y = x^3$, $x = 0$, and $y = 8$. *Ans.:* $\frac{768}{7}\pi$.

13 Let $a > 0$. Two circular cylinders of radius a have their axes on the x and y axes. With axes so oriented that the z axis is vertical, sketch the part of the first cylinder which lies in the first octant and between the planes $x = 0$ and $x = 5a$. Then sketch the part of the second cylinder which lies in the first octant and between the planes $y = 0$ and $y = 2a$. For three values of z, sketch the lines in which a horizontal plane through $(0,0,z)$ intersects the parts of the cylinders in the figure, and then sketch the curve in which the parts of the cylinders intersect. If the figure is reasonably good, it should be easy to find the volume V of the solid bounded by the three coordinate planes and parts of the two cylinders. Do it. *Ans.:* Figure 4.592 and $V = 2a^3/3$.

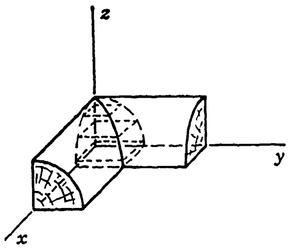


Figure 4.592

14 Find a reason why the answer to the preceding problem must be less than a^3.

15 A cylindrical hole is drilled through the center of a spherical ball. It is observed that the length of the hole is L. Show that the volume of the part of the ball remaining is the same as the volume of a spherical ball of diameter L.

16 A section of a tree trunk is a section of a right circular cylinder of radius a. A wedge is removed by making two cuts to a diameter (line, not number), one cut being in a horizontal plane and the other being in a plane which makes the angle θ with the horizontal plane. Find the volume of the wedge.

Ans.: $\frac{2}{3}a^3 \tan\theta$.

17 It is of interest to know that our methods are powerful enough to enable us to derive the standard formula

$$V = \tfrac{4}{3}\pi abc \tag{1}$$

for the volume V of the solid in E_3 bounded by the ellipsoid having the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{2}$$

in which a, b, c are positive constants. The formula for the volume can be remembered with the aid of the fact that if $a = b = c = r$, then (2) is the (or an) equation of a sphere of radius r and (1) gives the volume of the ball which it bounds. To start to find the volume of the part of our solid containing points

(x,y,z) for which $y \geqq 0$, we make a partition of the interval $0 \leqq y \leqq b$. When $0 < y < b$, we can put (2) in the form

$$(3) \qquad \frac{x^2}{a^2} + \frac{z^2}{c^2} = \frac{1}{b^2}(b^2 - y^2)$$

and hence in the form

$$(4) \qquad \frac{x^2}{\left(\frac{a}{b}\sqrt{b^2 - y^2}\right)^2} + \frac{y^2}{\left(\frac{c}{b}\sqrt{b^2 - y^2}\right)^2} = 1.$$

When y has the constant value y_k^*, (4) has the form

$$(5) \qquad \frac{x^2}{A^2} + \frac{z^2}{B^2} = 1,$$

where

$$(6) \qquad A = \frac{a}{b}\sqrt{b^2 - y_k^{*2}}, \qquad B = \frac{c}{b}\sqrt{b^2 - y_k^{*2}}.$$

This shows that, as Figure 4.593 indicates, the plane having the equation $y = y_k^*$ intersects our solid in an elliptic disk which, according to Problem 19 of Section 4.4, has area πAB or

$$(7) \qquad \frac{\pi ac}{b^2}(b^2 - y_k^{*2}).$$

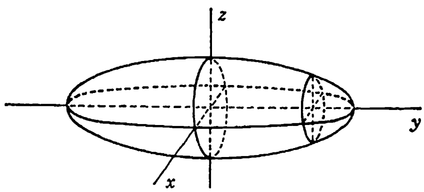


Figure 4.593

The volume of the slab of our solid which lies between the planes having the equations $y = y_{k-1}$ and $y = y_k$ is then exactly or approximately the result of multiplying (7) by Δy_k. Thus

$$(8) \qquad V = 2 \lim \sum \frac{\pi ac}{b^2}(b^2 - y_k^{*2})\,\Delta y_k,$$

the factor 2 being required because we partitioned only the interval $0 \leqq y \leqq b$. The limit of Riemann sums being a Riemann integral, we obtain

$$(9) \qquad V = \frac{2\pi ac}{b^2}\int_0^b (b^2 - y^2)\,dy$$

and hence (1). In case two of the three numbers a, b, c are equal, say $a = c$, the graph of (2) is called a *spheroid*. When finding the volume of the solid bounded by a spheroid, it is possible to simplify matters by using circular disks instead of elliptic disks. Some scientists consider it to be more fun to work out the above formulas than to remember that a spheroid for which $a = c < b$ is called a *prolate spheroid* (like the surface of a cucumber or a watermelon) and that a spheroid for which $a = c > b$ is an *oblate spheroid* (like the surface of a pancake or an unscarred earth that bulges at its equator and is flattened at its poles because of its rotation).

18 From time to time, we recognize the fact that some scientific terminologies and notations have their historical origins in primitive ideas that are fuzzy or incorrect. The number in the right member of the formula

$$\lim \sum f(x)\, \Delta x = \int_a^b f(x)\, dx \tag{1}$$

is, when it exists, defined in terms of Riemann sums in a way which we must now understand. If (1) holds, then to each positive number ϵ there corresponds a positive number δ such that

$$\left| \sum_{k=1}^{n} f(x_k^*)\, \Delta x_k - \int_a^b f(x)\, dx \right| < \epsilon \tag{2}$$

whenever P is a partition of the interval $a \leqq x \leqq b$ for which $|P| < \delta$. For a long time before this precise idea of Riemann revolutionized (or counter-revolutionized) mathematics, it was generally considered to be meaningful to regard the limit of sums as "the sum of infinitely many infinitesimals." Thus $\int_a^b f(x)\, dx$ was considered to be an "infinite sum" of products of "finite" numbers $f(x)$ and "infinitesimal" numbers dx. The "reasoning" involved is quite as flimsy and unrewarding as the "reasoning" which reaches the "conclusion" that "a circle is a polygon having infinitely many infinitesimal sides because it is a limit of polygons." In mathematics, as in other sciences, many of our ancestors were intrigued by ideas which are now considered to be obsolete. Nowadays we accept the idea that the sum of the volumes of many thin slabs can be a good approximation to the volume of a spherical ball, but we reject the fuzzy idea that the volume of the ball is the sum of the volumes of infinitely many infinitely thin slabs. It is not easy for historians to decide which of our great ancestors really had quite correct ideas about approximations and limits and, without swallowing ideas about sums of infinitesimals, merely used the fuzzy terminology because it was the fashion to do so. There can be tenuous connections between ideas and words. If Leonhard Euler wrote in a language in which apples were called "potatoes that grow in the air," historians unaware of the fact have an opportunity to conclude that this intellectual giant did not know the difference between potatoes and apples. Some people believe that the notation for integrals is bad because it makes too many people think that the dx is a number. The author believes that terminologies and notations involving limits are the real sinners because they make too many people think that numbers and partitions and other things are mobile. Perhaps replacing "lim" by "approx" in (1) would cure many of our ills.

4.6 Riemann-Cauchy integrals and work This section introduces integrals that are, in some cases, not Riemann integrals but are constructed from Riemann integrals by use of ideas that were made precise by the French mathematician Cauchy (1789–1857). It may happen that the integral in the right member of the formula

$$\int_a^\infty f(x)\, dx = \lim_{h\to\infty} \int_a^h f(x)\, dx \tag{4.61}$$

exists as a Riemann integral whenever $h \geqq a$ and that this integral, as a function of h, has a limit as h becomes infinite. In such cases, this limit is the *Riemann-Cauchy integral* of f over the semi-infinite interval $x \geqq a$ and is denoted by the symbol in the left member of (4.61). In each other case, we say that the integral in the left member of (4.61) does not exist as a Riemann-Cauchy integral. For example, when $r > 0$,

$$(4.611)\qquad \int_r^\infty \frac{1}{x^2}\,dx = \lim_{h\to\infty} \int_r^h x^{-2}\,dx = \lim_{h\to\infty} \frac{x^{-1}}{-1}\bigg]_r^h = \lim_{h\to\infty}\left[\frac{1}{r} - \frac{1}{h}\right] = \frac{1}{r}.$$

We can bravely start to calculate a Riemann-Cauchy integral by tentatively writing

$$(4.612)\qquad \int_0^\infty \cos x\,dx = \lim_{h\to\infty} \int_0^h \cos x\,dx = \lim_{h\to\infty}\Big[\sin x\Big]_0^h = \lim_{h\to\infty} \sin h$$

with the understanding that we will get an answer if the last limit exists. The last limit does not exist, however, so the integral does not exist.

Riemann-Cauchy integrals of another type are defined by the formula

$$(4.62)\qquad \int_0^a f(x)\,dx = \lim_{h\to 0+} \int_h^a f(x)\,dx$$

when $a > 0$ and the integrals and limit exist. Consider the example for which $f(x) = x^{-1/2}$ when $x > 0$, while $f(x)$ is either undefined or is defined in some other way when $x \leqq 0$. Then f is not bounded over the interval $0 < x \leqq 1$ and hence $\int_0^1 f(x)\,dx$ cannot exist as a Riemann integral. However,

$$(4.621)\qquad \begin{aligned} \int_0^1 x^{-1/2}\,dx &= \lim_{h\to 0+} \int_h^1 x^{-1/2}\,dx \\ &= \lim_{h\to 0+} \Big[2x^{1/2}\Big]_h^1 = \lim_{h\to 0+}\Big[2 - 2\sqrt{h}\Big] = 2, \end{aligned}$$

so the first integral exists as a Riemann-Cauchy integral. Riemann-Cauchy integrals of still other types are defined by the formulas

$$(4.622)\qquad \begin{aligned} \int_{-a}^0 f(x)\,dx &= \lim_{h\to 0+} \int_{-a}^{-h} f(x)\,dx \\ \int_{-\infty}^a f(x)\,dx &= \lim_{h\to -\infty} \int_h^a f(x)\,dx \end{aligned}$$

when the integrals and limits exist. Finally, the Riemann-Cauchy integrals in the left members of the formulas

$$(4.623)\qquad \int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^a f(x)\,dx + \int_a^\infty f(x)\,dx$$

$$(4.624)\qquad \int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

are defined by these formulas whenever the integrals on the right exist as Riemann-Cauchy integrals. Perhaps attention should be called to the fact that some elementary books reserve the term "definite integral" for application to an integral of a particular brand (which is sometimes the Riemann brand and is sometimes not carefully delineated) and apply the term "improper integral" to each integral of another kind.

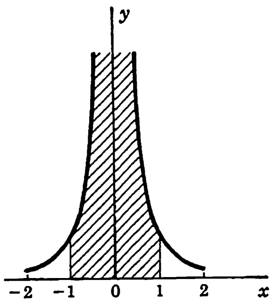


Figure 4.631

It is impossible to have a tranquil scientific career without thorough understanding of matters relating to

$$\int_{-1}^{1} \frac{1}{x^2}\,dx. \tag{4.63}$$

The graph of the integrand is shown in Figure 4.631. The integral cannot exist as a Riemann integral because the integrand x^{-2} is undefined when $x = 0$. Even if we set $f(0) = 0$ and $f(x) = x^{-2}$ when $x \neq 0$, the integral $\int_{-1}^{1} f(x)\,dx$ will still fail to exist as a Riemann integral because f is not bounded over the interval $-1 \leqq x \leqq 1$. According to (4.624), the formula

$$\int_{-1}^{1} \frac{1}{x^2}\,dx = \int_{-1}^{0} \frac{1}{x^2}\,dx + \int_{0}^{1} \frac{1}{x^2}\,dx \tag{4.632}$$

will be valid when the integrals are Riemann-Cauchy integrals provided the two integrals on the right side exist. The calculation

$$\int_{-1}^{0} \frac{1}{x^2}\,dx = \lim_{h\to 0+} \int_{-1}^{-h} \frac{1}{x^2}\,dx = \lim_{h\to 0+} \left[\frac{1}{h} - 1\right] = \infty \tag{4.633}$$

shows that the first integral on the right does not exist, and the calculation

$$\int_{0}^{1} \frac{1}{x^2}\,dx = \lim_{h\to 0+} \int_{h}^{1} \frac{1}{x^2}\,dx = \lim_{h\to 0+} \left[\frac{1}{h} - 1\right] = \infty \tag{4.634}$$

shows that the second does not exist. Hence, the integrals in (4.632) do not even exist as Riemann-Cauchy integrals. The calculations do, however, enable us to convey information by writing

$$\int_{-1}^{1} \frac{1}{x^2}\,dx = \int_{-1}^{0} \frac{1}{x^2}\,dx + \int_{0}^{1} \frac{1}{x^2}\,dx = \infty + \infty = \infty. \tag{4.635}$$

Persons do not lead these tranquil scientific lives when they realize that

$$\frac{d}{dx}\frac{-1}{x} = \frac{1}{x^2} \tag{4.636}$$

except when $x = 0$ and cheerfully make the calculation

$$\int_{-1}^{1} \frac{1}{x^2}\,dx = \frac{-1}{x}\bigg]_{-1}^{1} = -2 \quad (????) \tag{4.637}$$

which would be correct if (4.636) were valid over the whole interval $-1 \leqq x \leqq 1$. Since the wide world contains many definitions of integrals in addition to those of Riemann and Riemann-Cauchy, it is somewhat presumptuous to assert that (4.637) is ridiculous. However, when we confine our attention to Riemann and Riemann-Cauchy integrals, we can observe that (4.637) is incorrect.

Integrals of the types in (4.61) and (4.623) are particularly useful. For example, the formula

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-M)^2}{2\sigma^2}}\,dx = 1 \qquad (\sigma > 0) \tag{4.64}$$

is not easily proved, but it lies at the foundation of very much work in probability and statistics. Proof of this formula will appear later.

We conclude this section with a discussion of work in which the formula

$$\int_{a}^{\infty} \frac{1}{x^2}\,dx = \frac{1}{a} \qquad (a > 0) \tag{4.65}$$

plays a fundamental role. To begin, we study the amount of work done by a force **F** which pulls a particle P from the place on an x axis where $x = a$ to the place where $x = b$. The force **F** may have the direction of the x axis but, as in Figure 4.651, this is not necessarily so. Let $f(x)$ denote the scalar component of the force **F** in the direction of the motion, that is, in the direction of the x axis. In case $f(x)$ is a constant, measured in pounds (or dynes), and the distance $b - a$ is measured in feet (or centimeters), the work W done by the force is measured in foot-pounds (or dyne-centimeters) and is defined by the formula

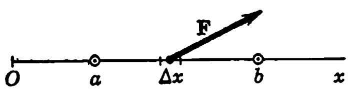


Figure 4.651

$$W = f(x)(b - a). \tag{4.652}$$

Since work and distance are scalars and force is a vector, it is quite incorrect to perpetuate the ancient idea that "work is force times distance"; we must use scalar components of forces. In case the scalar component $f(x)$ is different for different numbers x, the definition (4.652) is inapplicable and we need integration to calculate W. The procedure is almost identical with the procedure used to calculate areas and volumes. We make a partition P of the interval from a to b with a "small" norm $|P|$ and write $f(x_k^*)\,\Delta x_k$ as an approximation to the amount of work done in pulling the particle from the left end to the right end of the kth subinterval. The sum

$$\Sigma f(x_k^*)\,\Delta x_k \tag{4.653}$$

should then be a good approximation to our answer W and hence we should have

$$W = \lim \sum f(x_k^*)\,\Delta x_k = \int_a^b f(x)\,dx. \tag{4.66}$$

Our statements about (4.653) and (4.66) were necessarily vague and optimistic because the quantity W that we are trying to calculate has not yet been defined. We must recognize the fact that we cannot prove correctness of a formula for W when we have no definition or other information that tells us what W is. In the absence of another definition or other information, we must adopt the principle that our work with partitions and Riemann sums provides the motivation for the definition whereby W is *defined* by the formula

$$W = \int_a^b f(x)\,dx \tag{4.661}$$

whenever f is a function for which the integral exists as a Riemann integral.

The above ideas will now be applied to basic problems. The Newton (1642–1727) law of universal gravitation says that if two particles of mass m_1 and m_2 are concentrated at distinct points P_1 and P_2, then these particles attract each other with a force whose magnitude is proportional to the product of the masses and inversely proportional to the square of the distance between them. Suppose we have a particle of mass m_1 concentrated permanently at the origin, Figure 4.67, and that we have a "test particle" of unit mass that we wish to move along the positive x axis. There is then a constant k, which depends only upon the units used to measure mass, force, and distance, such that the force on the test particle has magnitude km_1/x^2 when the particle is at distance x from the origin. The work $W_{a,b}$ required to move the particle from the point a (that is, the point with coordinate a) to the point b is then to be calculated from the formula

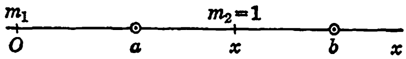


Figure 4.67

$$W_{a,b} = \int_a^b \frac{km_1}{x^2}\,dx. \tag{4.671}$$

From this we find the remarkably simple formula

$$W_{a,b} = \frac{km_1}{a} - \frac{km_1}{b}. \tag{4.672}$$

It follows from these formulas that

$$\lim_{b\to\infty} W_{a,b} = \int_a^\infty \frac{km_1}{x^2}\,dx = \frac{km_1}{a}. \tag{4.673}$$

This formula is responsible for some terminology that scientists often use. The limit in (4.673) is called "the amount of work required to take the test particle from a to infinity" and this amount of work is called the *potential* (or gravitational potential) at the point a due to the particle of mass m_1 at the origin. It is an easy consequence of these definitions and formulas that the potential, say u, at the point $P(x,y,z)$ due to a particle of mass m_1 concentrated at the point $P_0(x_0,y_0,z_0)$ is

$$u = \frac{km_1}{\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2}}. \tag{4.68}$$

The basic importance of the concept of potential u lies in the fact that if a particle of mass m is moved from a point P_1 to a point P_2 with no forces upon it except gravitational forces and a force $\mathbf{F}$, and if the speeds at P_1 and P_2 are equal, then the work done by the force $\mathbf{F}$ is equal to the product of m and the *potential difference*, that is, the potential at the starting point P_1 minus the potential at the destination P_2.

All of the above ideas and formulas apply to electrostatic potentials as well as to gravitational potentials. In the electrical case, we start with two charges q_1 and q_2 and apply the Coulomb (1736–1806) law $|\mathbf{F}| = kq_1q_2/x^2$, which is the electrical analogue of the Newton law of gravitation.

Problems 4.69

1 Suppose somebody writes

$$\int_0^1 \frac{1}{x}\,dx = \infty, \qquad \int_1^\infty \frac{1}{x}\,dx = \infty$$

with the hope that he is conveying information to you. What does he mean? *Ans.:*

$$\lim_{h\to 0+}\int_h^1 \frac{1}{x}\,dx = \infty, \qquad \lim_{h\to\infty}\int_1^h \frac{1}{x}\,dx = \infty.$$

2 Prove that

(*a*) $$\int_1^\infty \frac{1}{x^p}\,dx = \frac{1}{p-1}, \qquad \int_0^1 \frac{1}{x^p}\,dx = \infty \qquad (p>1)$$

(*b*) $$\int_1^\infty \frac{1}{x^p}\,dx = \infty, \qquad \int_0^1 \frac{1}{x^p}\,dx = \frac{1}{1-p} \qquad (p<1).$$

3 Show that, when $k > 0$,

$$\int_0^\infty e^{-kx}\,dx = \frac{1}{k}.$$

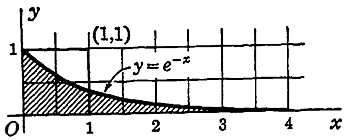


Figure 4.691

4 Remembering that $e = 2.71828$, and remembering or learning that e^3 is about 20, e^6 is about 400, and e^9 is about 8000, make some calculations to indicate that Figure 4.691 shows the nature of the graph of $y = e^{-x}$. Observe that the area of the shaded region seems to be about the same as the area of the unit square. What are the facts?

5 The region bounded by the *cissoid* having the equation

$$y^2 = \frac{x^3}{2a - x}$$

and its asymptote is rotated about the asymptote. Using the cylindrical shell method, set up an integral for the volume V of the solid thus generated. *Clue and ans.:*

$$V = 2 \lim \sum 2\pi(2a - x)y\,\Delta x$$
$$V = 4\pi \int_0^{2a} x^{3/2}(2a - x)^{1/2}\,dx.$$

Remark: With the aid of information about beta integrals, it can be shown very quickly that $V = 2\pi^2 a^3$.

6 Show that putting $M = 0$ and $\sigma = 1/\sqrt{2}$ in (4.64) gives the formula

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

Sketch a graph of $y = e^{-x^2}$ which is good enough to show that this result seems to be correct.

7 Prove that if $f(x) = 4$ when $0 < x < 1$ and $f(x) = 5$ when $1 < x < 2$, then

$$\int_0^2 f(x)\,dx = 9.$$

Note that $f(x)$ is undefined when $x = 0$, when $x = 1$, and when $x = 2$.

8 Prove that

$$\lim_{h\to\infty} \int_{-h}^{h} x\,dx = 0, \qquad \lim_{h\to\infty} \int_0^h x\,dx = \infty.$$

9 Even persons having little contact with the external physical world know that rods and wires and springs stretch when they are pulled and that the amount of stretching depends in some way upon the amount of pulling. Engineers have understanding of elastic limits and of circumstances under which useful results are obtained by applying the law of Robert Hooke. The Hooke law says that the magnitude of the force required to stretch a rod of natural length L to length $L + x$ is

$$\frac{k}{L}x,$$

where k is a constant that depends upon the rod. The number x is the *elongation* of the rod, and the magnitude of the force is proportional to the elongation.

Figure 4.692, which shows the rod before and after stretching, may be helpful. Supposing that $0 < a < b$, find the work done in stretching the rod from length $L + a$ to length $L + b$. *Ans.:*

$$\frac{k}{2L}(b^2 - a^2).$$

10 A conical container (see Figure 4.693) has height a feet and base radius R feet. It is filled with substance (water or wheat, for example) which weighs w pounds per cubic foot and which must be elevated (by a pump or shovel or other elevator) to a level H feet above the vertex. Suppose that $H \geqq a$. Find the work W required to accomplish the task. *Hint:* Start by making a partition of the interval $0 \leqq y \leqq a$ and calculating an approximation to the work required to lift the material which constitutes a horizontal sheet or slab. All calculations are based upon the fundamental idea that gravity pulls things downward, and that the magnitude of the force on a thing is its weight. *Ans.:*

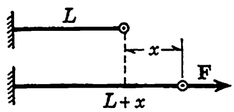


Figure 4.692

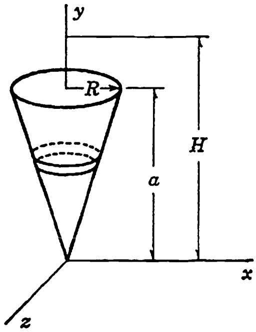


Figure 4.693

$$W = w\pi R^2 a\left(\frac{H}{3} - \frac{a}{4}\right).$$

Note that if V is the volume of the conical solid, then the answer can be put in the form $W = wV(H - \frac{3}{4}a)$.

11 Modify Problem 10 by replacing the conical container by a container such that, for each y^* for which $0 \leqq y^* \leqq a$, the plane having the equation $y = y^*$ intersects the contents of the container in a set having area $A(y^*)$. Then set up an integral for the work W. *Ans.:*

$$W = w\int_0^a (H - y)A(y)\,dy.$$

12 In many problems involving motion of particles, we need the concept of *kinetic energy*, or energy due to motion. This problem requires us to study and learn a method by which we can use calculus to derive an important formula. We suppose that, at time $t = 0$, a particle of mass m starts from rest, with kinetic energy zero, at the origin of an x axis and is pulled in the direction of the positive x axis by a force $\mathbf{F}$ of constant magnitude for which $\mathbf{F} = C\mathbf{i}$ at all times. We suppose that no force other than $\mathbf{F}$ operates on the particle. Letting x denote the coordinate of the particle at time t, we use the Newton law $\mathbf{F} = m\mathbf{a}$ to obtain the vector equation

$$(1) \qquad m\frac{d^2x}{dt^2}\mathbf{i} = m\mathbf{a} = \mathbf{F} = C\mathbf{i}.$$

From this we conclude that there must be a constant vector $\mathbf{c}_1$ such that

(2) $$m\frac{dx}{dt}\mathbf{i} = Ct\mathbf{i} + \mathbf{c}_1.$$

But $(dx/dt)\mathbf{i}$ is the velocity $\mathbf{v}$ at time t, and putting $t = 0$ in (2) shows that $\mathbf{c}_1 = 0$. Therefore,

(3) $$m\mathbf{v} = m\frac{dx}{dt}\mathbf{i} = Ct\mathbf{i}.$$

From this we conclude that there is a constant vector $\mathbf{c}_2$ such that

(4) $$mx\mathbf{i} = \tfrac{1}{2}Ct^2\mathbf{i} + \mathbf{c}_2.$$

But $x = 0$ when $t = 0$, so $\mathbf{c}_2 = 0$. Therefore,

(5) $$mx\mathbf{i} = \tfrac{1}{2}Ct^2\mathbf{i}.$$

The kinetic energy KE of our particle at time t is defined to be the amount of work done by the force $\mathbf{F}$ in bringing the particle from its state of rest at time $t = 0$ to its state of motion at time t. Since $|\mathbf{F}|$ has the constant magnitude C and has the direction of motion of the particle as the particle moves the distance x, the amount of work done is Cx. Thus KE $= Cx$ and, with the aid of (5) and (3), we find that

(6) $$\text{KE} = Cx = \frac{1}{2m}(Ct)^2 = \tfrac{1}{2}m|\mathbf{v}|^2.$$

Therefore,

(7) $$\text{KE} = \tfrac{1}{2}m|\mathbf{v}|^2.$$

The next problem requires that the same result be worked out by a different method without the assumption that $\mathbf{F}$ is a constant.

13 A particle P of mass m is moved around in E_3 by a continuous net force $\mathbf{F}$ which operates over a time interval $a \leqq t \leqq b$. The Newton law $\mathbf{F} = m\mathbf{a}$ then shows that the displacement vector $\mathbf{r}$ (which is $\overrightarrow{OP}$), the velocity vector $\mathbf{v}$, and the acceleration vector $\mathbf{a}$ are continuous functions of t. Make a partition of the interval $a \leqq t \leqq b$ and look at Figure 4.694 which shows, among other things, the positions of P at times t_{k-1} and t_k. Tell why the scalars

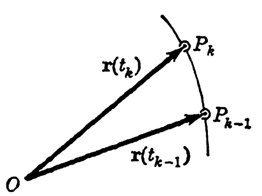


Figure 4.694

(1) $\mathbf{F}(t_k)\cdot[\mathbf{r}(t_k) - \mathbf{r}(t_{k-1})]$ and $\mathbf{F}(t_k)\cdot\mathbf{v}(t_k)\,\Delta t_k$

should, when $|P|$ is small, both be good approximations to the work done by $\mathbf{F}$ in forcing P from P_{k-1} to P_k. Tell why it should be reasonable to adopt either one of the formulas

(2) $$W = \lim_{|P|\to 0}\sum_{k=1}^{n}\mathbf{F}(t_k)\cdot[\mathbf{r}(t_k) - \mathbf{r}(t_{k-1})]$$

(3) $$W = \lim_{|P|\to 0}\sum_{k=1}^{n}\mathbf{F}(t_k)\cdot\mathbf{v}(t_k)\,\Delta t_k$$

as the *definition* of the work done by $\mathbf{F}$ over the time interval $a < t < b$. Show that (3) is equivalent to the definition

$$W = \int_a^b \mathbf{F}(t)\cdot\mathbf{v}(t)\,dt. \tag{4}$$

It remains for us to learn a little trick by means of which information can be gleaned from this formula. Using the Newton formula $\mathbf{F} = m\mathbf{a}$ gives

$$\begin{aligned}\mathbf{F}(t)\cdot\mathbf{v}(t) &= m\mathbf{a}(t)\cdot\mathbf{v}(t) = m\mathbf{v}'(t)\cdot\mathbf{v}(t)\\ &= \frac{1}{2}m\frac{d}{dt}[\mathbf{v}(t)\cdot\mathbf{v}(t)] = \frac{1}{2}m\frac{d}{dt}|\mathbf{v}(t)|^2.\end{aligned} \tag{5}$$

Hence

$$W = \frac{1}{2}m\int_a^b \frac{d}{dt}\Big[|\mathbf{v}(t)|^2\Big]\,dt = \frac{1}{2}m\Big[|\mathbf{v}(t)|^2\Big]_a^b \tag{6}$$

and

$$W = \frac{1}{2}m|\mathbf{v}(b)|^2 - \frac{1}{2}m|\mathbf{v}(a)|^2. \tag{7}$$

In case $\mathbf{v}(a) = 0$, our work gives another derivation of the formula for the kinetic energy of a particle of mass m having speed $|\mathbf{v}(b)|$.

14 The graph of the equation

$$y = \frac{a^2x}{x^2+b^2},$$

which usually appears in the disguised form $x^2y + b^2y - a^2x = 0$, is called a *serpentine*. Find the area (finite or infinite) of the region in the first quadrant between the serpentine and its asymptote.

15 Accumulation of familiarity with Riemann sums may bring a desire to know why $\int_a^b f(x)\,dx$ cannot exist as a Riemann integral when f is defined but unbounded over the interval $a \leqq x \leqq b$. If the integral exists and has the value I, then there must be a partition P of the interval $a \leqq x \leqq b$ such that

$$\Big|\sum_{k=1}^{n} f(x_k^*)\,\Delta x_k - I\Big| < 1 \tag{1}$$

whenever $x_{k-1} \leqq x_k^* \leqq x_k$ for each k. Show that if (1) holds, then

$$|f(x_1^*)| \leqq (\Delta x_1)^{-1}\Big[1 + I + \sum_{k=2}^{n} |f(x_k)|\,\Delta x_k\Big] \tag{2}$$

when $x_0 \leqq x_1^* \leqq x_1$. This shows that f must be bounded over the first subinterval of the partition P. Similar arguments show that f must be bounded over the other subintervals and hence also over the whole interval $a \leqq x \leqq b$.

4.7 *Mass, linear density, and moments* This section involves some ideas that turn out to be important in many ways. Let F be a function which is defined over some finite interval $a \leqq x \leqq b$ and is *monotone increasing* over the interval. This means that $F(x_1) \leqq F(x_2)$ whenever $a \leqq x_1 \leqq x_2 \leqq b$. Such functions F arise in many ways. We can, for example, let P (a number) denote the population of an island, state, or country, let $P(x)$ denote the number of persons having age less than or equal to x, and let

$$F(x) = \frac{P(x)}{P}. \tag{4.71}$$

We can also suppose that the interval $a \leqq x \leqq b$ represents a line segment or a slim beam, as in Figure 4.72, upon which sand and perhaps other things are piled and from which hams and other things are hung, and let $F(x)$ be the total mass which rests upon or hangs from the part of the interval from a to x.

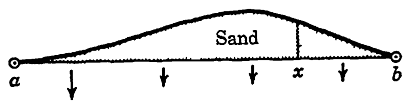


Figure 4.72

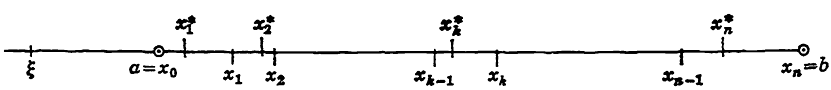
Because of the vividness of the latter interpretation, F is sometimes called a mass function even when $F(x)$ is a number which is important in social sciences and which has nothing whatever to do with such things as pounds and tons and grams and slugs. When x and $x + \Delta x$ both lie in the interval from a to b, the difference quotient

$$\frac{F(x + \Delta x) - F(x)}{\Delta x} \tag{4.73}$$

represents the average mass per unit length or the average linear density over the interval with end points at x and $x + \Delta x$. In case this quotient has, for a given x, a limit as Δx approaches zero, this limit is called the *density* at x. When this density exists, we call it $f(x)$ so that, by our definition of derivatives, $f(x) = F'(x)$. This idea of density has its simplest applications in cases where $F(x)$ has a continuous derivative. In these cases the function f having values $f(x)$ is called the *density function* of the mass function F, and $f(x) = F'(x)$ for each x.

We are now ready to start introducing moments. Without assuming that F is differentiable or even continuous, let ξ(xi) be a number (or point) not necessarily in the interval from a to b, and let p be an integer which is either 0 or positive. Let P be a partition of the interval $a \leqq x \leqq b$ as shown in Figure 4.74. The number $F(x_k) - F(x_{k-1})$ is the number obtained by starting with the total mass in the interval $a \leqq x \leqq x_k$ and

Figure 4.74



subtracting the total mass in the interval $a \leqq x \leqq x_{k-1}$. Thus it is the total mass in the interval $x_{k-1} < x \leqq x_k$. The number

$$(x_k^* - \xi)^p[F(x_k) - F(x_{k-1})] \tag{4.75}$$

represents the pth moment about the point ξ of a single particle of mass $F(x_k) - F(x_{k-1})$ concentrated at the point x_k^* and, when the norm of P is small, this should be a good approximation to the pth moment about ξ of the total mass in the interval $x_{k-1} < x \leqq x_k$. Moreover, the sum

$$\sum_{k=1}^{n} (x_k^* - \xi)^p[F(x_k) - F(x_{k-1})] \tag{4.76}$$

should be a good approximation to the pth moment about ξ of the total mass in the interval $a \leqq x \leqq b$. Our statement about (4.76) was necessarily vague and optimistic because the quantity we are trying to calculate has not yet been defined. It is a fundamental fact, which is proved in the theory of Riemann-Stieltjes integrals, that there is a number $M^{(p)}_{x=\xi}$ such that the sum in (4.76) is near it whenever $|P|$ is small, that is,

$$M^{(p)}_{x=\xi} = \lim_{|P|\to 0} \sum_{k=1}^{n} (x_k^* - \xi)^p[F(x_k) - F(x_{k-1})]. \tag{4.77}$$

This number $M^{(p)}_{x=\xi}$ is called the pth moment about the point ξ of the mass in the interval $a \leqq x \leqq b$. In case $p = 0$, the pth moment is the total mass in the interval $a \leqq x \leqq b$. In mechanics, the second moment is called *moment of inertia*. In statistics and elsewhere, the particular number $\bar{x}$ for which $M^{(1)}_{x=\bar{x}} = 0$ is called the *mean* (or *mean value*) of F over the interval $a \leqq x \leqq b$. In mechanics and elsewhere, the point having coordinate $\bar{x}$ is called the *centroid* of the mass. The number $M^{(2)}_{x=\bar{x}}$, the second moment about the centroid or mean, is particularly important in mechanics and statistics.

The above discussion applies equally well to mass functions F that possess continuous density functions f and to those that do not. When F does possess a continuous density function f, we can solve problems with the aid of only Riemann integrals. In the latter case the number

$$f(x_k^*)\,\Delta x_k \tag{4.78}$$

is taken to be an approximation to the total mass in the interval $x_{k-1} < x \leqq x_k$, and instead of (4.76) we use the Riemann sum

$$\sum_{k=1}^{n} (x_k^* - \xi)^p f(x_k^*)\,\Delta x_k \tag{4.781}$$

as an approximation to $M^{(p)}_{x=\xi}$. Taking limits as the norm of the partition P approaches 0 then gives the formula

$$M^{(p)}_{x=\xi} = \int_a^b (x - \xi)^p f(x)\,dx. \tag{4.782}$$

Problems 4.79

1 As suggested by Figure 4.791, let a rod having constant linear density (mass per unit length) δ be supposed to be concentrated on the interval $a \leqq x \leqq b$ of the x axis. Starting by making a partition of the interval $a \leqq x \leqq b$, calculate $M^{(p)}_{x=\xi}$, the pth moment about ξ of the rod. *Ans.:*

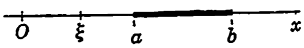


Figure 4.791

$$M^{(p)}_{x=\xi} = \frac{\delta}{p+1}[(b-\xi)^{p+1} - (a-\xi)^{p+1}].$$

2 Using the result of the preceding problem, prove that $M^{(1)}_{x=\xi} = 0$ if and only if $\xi = \frac{1}{2}(a+b)$.

3 Supposing that

$$\int_a^b f(x)\,dx = M > 0,$$

show that the constant $\bar{x}$ satisfies the equation

$$\int_a^b (x-\bar{x})f(x)\,dx = 0$$

if and only if

$$M\bar{x} = \int_a^b xf(x)\,dx.$$

Remark: Always remember that, in statistics and elsewhere, $\bar{x}$ is called the mean (or mean value) of f over the interval $a \leqq x \leqq b$ and that, in mechanics and elsewhere, $\bar{x}$ is the x coordinate of a centroid. Remember (or learn) that a centroid is, as it should be, a point "like a center."

4 Supposing that

$$\int_a^b f(x)\,dx = M > 0$$

and that the mean (or x coordinate of the centroid) is $\bar{x}$, prove that

$$M^{(2)}_{x=\xi} = M^{(2)}_{x=\bar{x}} + (\bar{x}-\xi)^2 M.$$

State the meaning of this formula in words, and use the formula to determine the value of ξ for which $M^{(2)}_{x=\xi}$ has the least possible value. *Hint:* Start by writing

$$M^{(2)}_{x=\xi} = \int_a^b (x-\xi)^2 f(x)\,dx = \int_a^b [(x-\bar{x}) + (\bar{x}-\xi)]^2 f(x)\,dx.$$

5 Let f be the function for which $f(x) = 0$ when $x < 0$ and $f(x) = e^{-x}$ when $x > 0$. Determine and graph the mass function F of which f is the density function.

6 The density function f defined by the first of the formulas

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-M)^2}{2\sigma^2}}, \qquad F(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{(t-M)^2}{2\sigma^2}}\,dt$$

has the mass function (or cumulative function) F defined by the second formula. With the aid of the formula (4.64) make a preliminary attempt to learn the natures of the graphs of $y = f(x)$ and $y = F(x)$ when $M = 0$ and $\sigma = 0.01$.

7 When $M = 7$, the graphs of the functions in the preceding problem are different from the graphs obtained when $M = 0$. What is the difference?

8 When a particle having mass m rotates in a circular path with angular speed ω radians per second at a constant distance r from an axis of rotation, its speed is $r\omega$ and its kinetic energy is $\frac{1}{2}mr^2\omega^2$. With the aid of this information, calculate the kinetic energy of a circular disk of radius a which has mass δ per unit area and which is rotating with speed ω radians per second about an axis through its center perpendicular to its plane. *Hint:* Base the solution on estimates of the area of a ring, the mass of the ring, and then the kinetic energy of the ring. *Ans.:* $\text{KE} = \frac{1}{4}\pi\delta a^4\omega^2$. The answer has the form $\text{KE} = \frac{1}{2}I\omega^2$, where I, the polar moment of inertia of the disk about the axis used, is $\frac{1}{2}\pi\delta a^4$.

9 The cone of Figure 4.51 has mass δ per unit volume and is rotating ω radians per second about its axis. Find its kinetic energy. *Hint:* Use the answer of Problem 8.

10 Figure 4.792 can make us wonder whether we are becoming wise enough to determine the attractive force $\mathbf{F}$ upon a particle of mass m at $P(x,y,z)$ that is produced by a bar or rod concentrated upon an interval $a \leqq x \leqq b$ of the x axis of an x, y, z coordinate system. We suppose that the bar has linear density $\delta(x)$ at the point $(x,0,0)$ and that $\delta(x)$ is integrable but not necessarily constant over the interval $a \leqq x \leqq b$. The first task is to set up an integral for $\mathbf{F}$. The following solution of this problem should be read even by those who can solve the problem without aid and assistance, because it fortifies our understanding of the process by which integrals are set up. We make a partition Q of the interval $a \leqq x \leqq b$, but we call the partition points $t_0, t_1, \cdots, t_n$ because the number x is the x coordinate of P. If the trick helps us, we can consider the x axis to be simultaneously an x axis and a t axis. For each $k = 1, 2, \cdots, n$ let t_k^* be chosen such that $t_{k-1} \leqq t_k^* \leqq t_k$. We then use the number

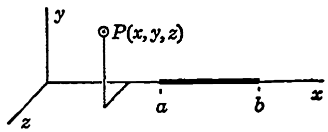


Figure 4.792

$$\delta(t_k^*)\,\Delta t_k \tag{1}$$

as an approximation to the mass of the part of the rod in the interval $t_{k-1} < t \leqq t_k$. Supposing that this mass is concentrated at the point $P_k(t_k,0,0)$, we use the number

$$Gm\frac{\delta(t_k^*)\,\Delta t_k}{|\overrightarrow{PP_k}|^2} \tag{2}$$

as an approximation to the magnitude of the force $\Delta\mathbf{F}_k$ on the particle at P produced by the part of the rod in the interval $t_{k-1} \leqq t \leqq t_k$. Problem 19 of Problems 2.39 discusses this matter and shows how we derive the formula

$$\Delta\mathbf{F}_k = Gm\frac{\delta(t_k^*)\,\Delta t_k\,\overrightarrow{PP_k}}{|\overrightarrow{PP_k}|^3} \tag{3}$$

by use of the fact that a nonzero vector is the product of its magnitude and a unit vector in its direction. Our next step is to write $\overrightarrow{PP_k}$ in terms of the coordinates of P and P_k and to write

$$\sum \Delta \mathbf{F}_k = Gm \, \Sigma \, \delta(t_k^*) \frac{(t_k - x)\mathbf{i} - y\mathbf{j} - z\mathbf{k}}{[(t_k - x)^2 + y^2 + z^2]^{3/2}} \Delta t_k. \tag{4}$$

Everything is now prepared for the crucial steps. When the norm $|Q|$ of the partition Q is small, the sums in (4) should be good approximations for the force $\mathbf{F}$ that we are trying to define. In other words, $\mathbf{F}$ should be the limit of these sums. But these sums are Riemann sums and, provided $P(x,y,z)$ is not a point on the interval $a \leqq x \leqq b$ of the x axis, they have a limit which is the Riemann integral in the formula

$$\mathbf{F} = Gm \int_a^b \delta(t) \frac{(t - x)\mathbf{i} - y\mathbf{j} - z\mathbf{k}}{[(t - x)^2 + y^2 + z^2]^{3/2}} \, dt. \tag{5}$$

Our work motivates the definition whereby $\mathbf{F}$ is defined by (5). While (5) serves as a source of information about $\mathbf{F}$ in other cases, we confine our attention here to the case in which the density is a constant, say $\delta(t) = \delta_0$ for each t, and, moreover, $y = z = 0$ and $x < a < b$. In this case, $\mathbf{F}$ has the direction of $\mathbf{i}$, and if we denote its magnitude by $F_1(a,b,x)$, then

$$\begin{aligned} F_1(a,b,x) &= Gm\delta_0 \int_a^b (t - x)^{-2} \, dt \\ &= Gm\delta_0 \left(\frac{1}{a - x} - \frac{1}{b - x} \right). \end{aligned} \tag{6}$$

It is easy to see that

$$\lim_{b \to \infty} F_1(a,b,x) = \frac{Gm\delta_0}{a - x}, \qquad \lim_{x \to a-} F_1(a,b,x) = \infty. \tag{7}$$

If these formulas agree with our intuitive notions, then at least some of our intuitive notions are good. The second result in (7) gives us a lesson in approximation. Since particles near ends of actual steel rods are not subject to huge attractive forces, we must conclude that very bad approximations to forces on these particles are obtained from calculations based on assumptions that the rods are concentrated on their axes.

11 Modify Figure 4.792 to fit the case in which $h > 0$, $a = -h$, $b = h$, $\delta(x) = \delta_0$, $x = 0$, and $z = 0$. Show that, in this case, formula (5) of the preceding problem becomes

$$\mathbf{F} = Gm\delta_0 \int_{-h}^{h} \frac{t\mathbf{i} - y\mathbf{j}}{(t^2 + y^2)^{3/2}} \, dt.$$

After having a good look at the coefficients of $\mathbf{i}$ and $\mathbf{j}$, show that

$$\mathbf{F} = -2Gm\delta_0 y\mathbf{j} \int_0^h \frac{1}{(t^2 + y^2)^{3/2}} \, dt.$$

12 A thin cylindrical shell S of radius R has its axis on the x axis of an x, y, z coordinate system and has its ends in the planes having the equations $x = a$ and $x = b$. This shell has constant areal density (mass per unit area) δ. Find the gravitational force $\mathbf{F}$ which it exerts upon a particle m^* of mass m which is concentrated at the point $(c,0,0)$. *Hint:* As suggested by Figure 4.793, make a partition P of the interval $a \leqq x \leqq b$. Consider the part of the shell between the planes having the equations $x = x_{k-1}$ and $x = x_k$ to be a circular ring having its mass M_k concentrated in the plane having the equation $x = x_k^*$. Let $\Delta\mathbf{F}_k$ be the force exerted upon m^* by this ring. Because of symmetry, the components of $\Delta\mathbf{F}_k$ orthogonal to $\mathbf{i}$ are zero. Moreover, the $\mathbf{i}$ component of $\Delta\mathbf{F}_k$ (which is $\Delta\mathbf{F}_k$) is the same as the $\mathbf{i}$ component of the force on m^* produced by a single particle of mass M_k concentrated at the point $(x_k^*,R,0)$ in E_3. Therefore,

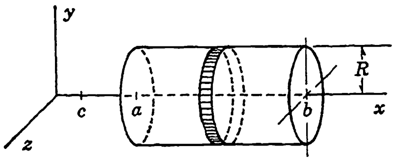


Figure 4.793

$$\Delta\mathbf{F}_k = Gm\frac{M_k(x^* - c)\mathbf{i}}{[(x^* - c)^2 + R^2]^{3/2}}, \tag{1}$$

and we are ready to calculate M_k and get on with the calculus. *Ans.:*

$$\mathbf{F} = 2\pi\delta GmR\left(\frac{1}{\sqrt{(a-c)^2 + R^2}} - \frac{1}{\sqrt{(b-c)^2 + R^2}}\right)\mathbf{i}. \tag{2}$$

13 We can claim that if the density δ and the radius R of the cylindrical shell of Problem 12 are so related that the total mass is M, then the answer to Problem 12 should be nearly the same as one of the answers to Problem 10 when R is near zero. Is it so? *Ans.:* Yes, unless misprints disrupt the harmony.

14 A circular disk of radius H has its center on the x axis of an x, y, z coordinate system and lies in the plane having the equation $x = x_0$. This disk has constant areal density (mass per unit area) δ. Set up an integral for the gravitational force $\mathbf{F}$ which the disk exerts upon a particle m^* of mass m which is concentrated at the point $(c,0,0)$ when $c \neq x_0$. *Hint:* Make a partition with the aid of which the disk is split into a collection of concentric rings so that a representative ring has radius r_k^*. The hint of Problem 12 provides a formula that can be adapted to give the force which the representative ring exerts upon m^*. *Ans.:*

$$\mathbf{F} = 2\pi\delta Gm(x_0 - c)\mathbf{i}\int_0^H \frac{r}{[r^2 + (x_0 - c)^2]^{3/2}}\,dr \tag{1}$$

$$\mathbf{F} = 2\pi\delta Gm(x_0 - c)\left[\frac{1}{|x_0 - c|} - \frac{1}{\sqrt{H^2 + (x_0 - c)^2}}\right]\mathbf{i}. \tag{2}$$

When M is the total mass of the disk so that $M = \pi H^2\delta$, the answer can be put in the form

$$\mathbf{F} = 2\frac{GmM}{H^2}\left[\frac{x_0 - c}{|x_0 - c|} - \frac{x_0 - c}{\sqrt{H^2 + (x_0 - c)^2}}\right]\mathbf{i}. \tag{3}$$

Remark: We really should look at these formulas. For example, (2) gives very interesting information when H is large and our disk is a huge part of a homogeneous plane. One who wishes additional mental elevation should undertake to realize that we can replace gravitational laws and constants by electrostatic ones and obtain information about forces on electrons produced by charges on plates of capacitors.

15 A cylindrical solid of radius R has its axis on the x axis of an x, y, z coordinate system and has its ends in the planes having the equations $x = a$ and $x = b$. This solid has uniform density (mass per unit volume) δ. Find the gravitational force $\mathbf{F}$ which this solid exerts upon a particle m^* of mass m located at the point $(c,0,0)$ of E_3, it being assumed that $c < a$. *Hint:* Make a partition of the interval $a \leqq x \leqq b$. Consider the part of the cylinder between the planes having equations $x = x_{k-1}$ and $x = x_k$ to be a circular disk in the plane having the equation $x = x_k^*$. Let $\Delta\mathbf{F}_k$ be the force exerted upon the particle m^* by this disk. A formula of Problem 14 can then be applied. *Ans.:*

$$\mathbf{F} = 2\pi Gm\delta\mathbf{i}\int_a^b \left[1 - \frac{x - c}{\sqrt{(x - c)^2 + R^2}}\right] dx$$

$$\mathbf{F} = 2\pi Gm\delta[b - a - (\sqrt{(b - c)^2 + R^2} - \sqrt{(a - c)^2 + R^2})]\,\mathbf{i}.$$

This can be put in the form

$$\mathbf{F} = 2\frac{GmM}{R^2}\left[1 - \frac{\sqrt{(b - c)^2 + R^2} - \sqrt{(a - c)^2 + R^2}}{b - a}\right]\mathbf{i},$$

where $M = \pi R^2(b - a)\delta$, the total mass of the cylindrical solid.

16 Let S be a thin spherical shell which is assumed to be concentrated on a sphere (surface, not ball) of radius a having its center at the origin. The shell has constant areal density (mass per unit area) δ. Let m^* be a particle of mass m which is concentrated at a point $(-b,0,0)$ which lies at the origin or at distance b from the origin on the negative x axis. Thus $b \geqq 0$, and we suppose that $b \neq a$ so m^* does not lie on the sphere. The gravitational force $\mathbf{F}$ exerted upon m^* by the shell depends upon the location of m^*. If $0 \leqq b < a$ so that m^* is inside the sphere, then $\mathbf{F} = 0$. If $b > a$ so that m^* is outside the sphere, then

$$(1) \qquad \mathbf{F} = G\frac{mM}{b^2}\,\mathbf{i},$$

where M is the total mass of the shell. Thus when m^* lies outside the shell, the force on it exerted by the shell is the same as the force exerted on it by a particle at the center of the shell whose mass is the total mass of the shell. From our present point of view, proofs of these famous and important results (which are discussed in more general terms in Section 13.8) can be comprehended more easily than they can be originated. To start our proof, we slice the spherical shell into ribbons to which we can apply a basic result given in Problem 12.

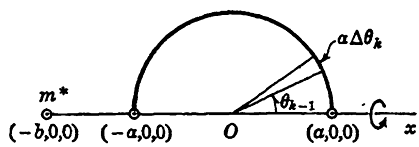


Figure 4.794

The spherical shell is obtained by rotating the semicircle of Figure 4.794 about the x axis. We make a partition P of the interval $0 \leqq \theta \leqq \pi$. With the aid of the basic formula

$$(2) \qquad \text{Angle} = \frac{\text{length of arc}}{\text{radius}}$$

we see that the lines making angle θ_{k-1} and θ_k with the positive x axis have between them an arc of the circle of length $a(\theta_k - \theta_{k-1})$, or $a\,\Delta\theta_k$. When this arc is rotated about the x axis, it produces a part of the spherical shell which can be described roughly as a circular ribbon having radius $a \sin \theta_k^*$, width $a\,\Delta\theta_k$, length $2\pi\, a \sin \theta_k^*$, area $2\pi\, a^2 \sin \theta_k^* \,\Delta\theta_k$, and mass M_k, where

(3) $$M_k = 2\pi\delta a^2 \sin \theta_k^* \,\Delta\theta_k.$$

Considering this ribbon to be a circular ring of mass M_k and radius $a \sin \theta_k^*$ which has its center on the x axis and which lies in the plane having the equation $x = a \cos \theta_k^*$, we use formula (1) of Problem 12 with $c = -b$ to obtain the formula

(4) $$\Delta\mathbf{F}_k = 2\pi Gm\delta a^2 \mathbf{i}\, \frac{(b + a \cos \theta_k^*) \sin \theta_k^*}{[(b + a \cos \theta_k)^2 + (a \sin \theta_k)^2]^{3/2}}\, \Delta\theta_k$$

for an approximation to the force upon m^* produced by one element of the spherical shell. The limit of the sum of these things should be the force $\mathbf{F}$ that we are seeking. But the sum is a Riemann sum and its limit is a Riemann integral. This leads us to the formula

(5) $$\mathbf{F} = 2\pi Gm\delta a\mathbf{i}\,U,$$

where U is the unruly integral defined by

(6) $$U = \int_0^\pi \frac{(b + a \cos \theta)a \sin \theta}{[b^2 + 2ab \cos \theta + a^2]^{3/2}}\, d\theta.$$

The hypothesis that $b \geqq 0$ and $b \neq a$ implies that the denominators in (4) and (6) are never zero and hence that the integrand in (6) is continuous. Before making a serious attack on the integral, we can observe that it is certainly positive when $b > a$ and that it is 0 when $b = 0$. To simplify the integral, we make the substitution (or change of variable)

(5) $$a \cos \theta = x,$$

so that $-a \sin \theta\, d\theta = dx$. Since $x = a$ when $\theta = 0$ and $x = -a$ when $\theta = \pi$, rules which have not yet been adequately treated imply that

(6) $$U = \int_{-a}^{a} \frac{b + x}{[b^2 + 2bx + a^2]^{3/2}}\, dx.$$

To simplify the integral some more when $b \neq 0$, we make the substitution

$$b^2 + 2bx + a^2 = t, \qquad x = \frac{t - b^2 - a^2}{2b}.$$

Since $dx = (1/2b)\, dt$, $t = (b - a)^2$ when $x = -a$, and $t = (b + a)^2$ when $x = a$. substitution in (6) gives

(7) $$U = \frac{1}{2b} \int_{(b-a)^2}^{(b+a)^2} \frac{b + \dfrac{t - b^2 - a^2}{2b}}{t^{3/2}}\, dt.$$

Thus

(8) $$U = \frac{1}{4b^2} \int_{(b-a)^2}^{(b+a)^2} [t^{-1/2} + (b^2 - a^2)t^{-3/2}]\, dt,$$

(9) $$U = \frac{1}{2b^2} \Big[t^{1/2} - (b^2 - a^2)t^{-1/2} \Big]_{(b-a)^2}^{(b+a)^2}$$

(10) $$U = \frac{1}{2b^2} \left[|b + a| - |b - a| - \frac{b^2 - a^2}{|b + a|} - \frac{b^2 - a^2}{|b - a|} \right].$$

In case $0 < b < a$, this gives

$$(11) \qquad U = \frac{1}{2b^2}\left[(b + a) - (a - b) - \frac{b^2 - a^2}{b + a} - \frac{b^2 - a^2}{b - a}\right] = 0$$

and (5) shows that $\mathbf{F} = 0$ as we wish to prove. In case $0 < a < b$, (10) gives

$$(12) \qquad U = \frac{1}{2b^2}\left[(b + a) - (b - a) - \frac{b^2 - a^2}{b + a} + \frac{b^2 - a^2}{b - a}\right] = \frac{2a}{b^2}.$$

Putting this in (5) gives

$$\mathbf{F} = \frac{Gm(4\pi a^2\delta)\mathbf{i}}{b^2} = \frac{GmM}{b^2}\mathbf{i},$$

where $M = 4\pi a^2\delta$, the total mass of the spherical shell. This is the desired result (1) and the fundamental facts about attractions of spherical shells are now established.

17 Use the method of Problem 16 but modify the details in appropriate places to obtain the force $\mathbf{F}_H$ exerted upon m^* by the hemispherical shell that remains after removal of the part of the spherical shell whose points have negative x coordinates.

18 A spherical ball (or solid sphere) is said to be *radially homogeneous* if there is a function δ such that the ball has density (mass per unit volume) $\delta(r)$ at each point having distance r from the center of the sphere. Supposing that $0 < a < b$, find the gravitational force exerted upon a particle m^* of mass m located at the point $(-b,0,0)$ in E_3 by a radially homogeneous spherical ball (like an idealized earth or golf ball) B which has radius a, which has its center at the origin, and which has a density function δ which is not necessarily constant but is integrable. *Solution:* As suggested by Figure 4.795, we make a partition P of the interval $0 \leqq x \leqq a$. When $x_{k-1} \leqq x_k^* \leqq x_k$ as usual, the points of B having distance x_k^* from 0 such that $x_{k-1} < x_k^* \leqq x_k$ form a spherical shell whose volume is approximately the product of the area $4\pi x_k^{*2}$ and the thickness Δr_k. The mass M_k of this shell is therefore approximately

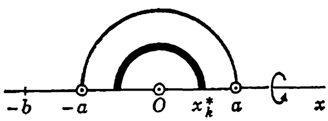


Figure 4.795

$$(1) \qquad M_k = 4\pi x_k^{*2}\delta(x_k^*)\,\Delta x_k.$$

Considering the shell to be concentrated upon the sphere having center at O and having radius x_k^* enables us to use a result of Problem 16 to show that the force $\Delta\mathbf{F}_k$ which the shell exerts upon m^* is approximately

$$(2) \qquad \Delta\mathbf{F}_k = \frac{Gm[4\pi x_k^{*2}\delta(x_k^*)\,\Delta x_k]}{b^2}\mathbf{i}.$$

The limit of the sum of these things should be $\mathbf{F}$. But the sum is a Riemann sum and its limit is the integral in the formula

$$(3) \qquad \mathbf{F} = \frac{Gm\mathbf{i}}{b^2}\int_0^a 4\pi x^2\delta(x)\,dx.$$

This answer can be greatly improved if we notice that methods very similar to those which we have used enable us to show that the integral in (3) is the total mass M of the ball B. Thus we can put (3) in the form

$$\mathbf{F} = \frac{GmM}{b^2}\,\mathbf{i}. \tag{4}$$

This proves that the force exerted upon m^* by a radially homogeneous ball is the same as the force exerted upon m^* by a single particle, at the center of the ball, whose mass is the total mass of the ball. Thus, when computing forces upon particles outside the ball, we may "consider the mass of the ball to be concentrated at its center," the assertion in quotation marks being rather weird because mass is a number and we do not ordinarily squeeze numbers.

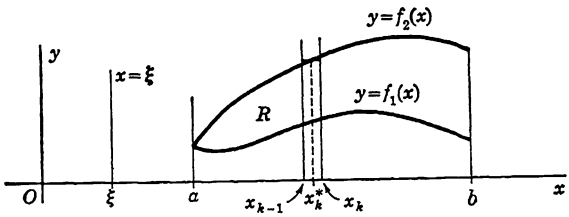
19 Use the method of Problem 18 to show that if S is a radially homogeneous spherical shell having inner and outer radii r_1 and r_2 for which $r_1 < r_2$, then $\mathbf{F} = 0$ when $\mathbf{F}$ is the gravitational force which the shell exerts upon a particle inside the shell.

20 With the aid of arguments involving continuity, the final formulas of preceding problems for gravitational forces upon particles exerted by solid spherical balls and solid unconcentrated spherical shells can be proved to be correct when the particles lie on boundaries of the balls and shells. Using this fact, show how it is possible to split a given radially homogeneous solid ball into an inner solid ball and an outer spherical shell to calculate the force which the given ball exerts upon a particle m^* of mass m concentrated at an inner point of the given ball.

4.8 *Moments and centroids in E_2 and E_3* Section 4.7 introduced us to moments, about a point on a line, of material concentrated upon the line. This section introduces us to two similar ideas. In the first place, we consider moments, about a line, of material concentrated in a plane containing that line. In the second place we consider moments, about a plane, of material in E_3.

To begin, let R be a bounded region in the xy plane which lies between the lines having the equation $x = a$ and $x = b$. It is supposed that when $a \leqq x^* \leqq b$, the line having the equation $x = x^*$ intersects R in an interval (or collection of intervals) having length (or total length) $f(x^*)$. It is not necessary that f be continuous, but we do assume that R has area $|R|$ and that $|R| = \int_a^b f(x)\,dx$. If the region R is, as in Figure 4.81,

Figure 4.81



the set of points (x,y) for which $a \leqq x \leqq b$ and $f_1(x) \leqq y \leqq f_2(x)$, then everything is quite simple and $f(x) = f_2(x) - f_1(x)$. Our next step is to suppose that the points (x,y) of R are points of a material body (much like a sheet of paper or a sheet of copper) that has been compressed into a plane in such a way that, for some positive constant δ, a part of the compressed material body has mass $\delta\,\Delta R$ if that part occupies a part of the region R having area ΔR. The compressed body is called a *lamina*, and it is a *homogeneous lamina* because the areal density (mass per unit area) has the same constant value δ at all places in the lamina.

Letting p be an integer which is either 0 or positive, we proceed to define the number $M_{x=\xi}^{(p)}$, *the* pth *moment of the lamina about the line having the equation* $x = \xi$, by a formula from which it can be calculated. Following the method of Section 4.7, we make a partition P of the interval $a \leqq x \leqq b$ into subintervals. For each k, the lines having the equations $x = x_{k-1}$ and $x = x_k$ have between them a part of the lamina that can be called a strip parallel to the line having the equation $x = \xi$. Supposing as usual that $x_{k-1} \leqq x_k^* \leqq x_k$, we use the number

$$f(x_k^*)\,\Delta x_k \tag{4.82}$$

as an approximation to the area of the strip and accordingly use the number

$$\delta f(x_k^*)\,\Delta x_k \tag{4.821}$$

as an approximation to the mass of the strip. When the norm of P is small, all points of the strip lie at about the same distance $|x_k^* - \xi|$ from the line having the equation $x = \xi$, and multiplying the above mass by $(x_k^* - \xi)^p$ should therefore give a good approximation to the pth moment of the strip about the line having the equation $x = \xi$. The Riemann sum

$$\delta\Sigma(x_k^* - \xi)^p f(x_k^*)\,\Delta x_k \tag{4.822}$$

should then be a good approximation to the moment of the whole lamina. Since the Riemann sums have a limit which is the Riemann integral in the right member of the formula

$$M_{x=\xi}^{(p)} = \delta \int_a^b (x - \xi)^p f(x)\,dx, \tag{4.83}$$

our work motivates the definition by which the required moment is defined by this formula.

The number $M_{y=\eta}^{(p)}$, the pth moment of the lamina about the line having the equation $y = \eta$, is defined by the analogous formula

$$M_{y=\eta}^{(p)} = \delta \int_c^d (y - \eta)^p g(y)\,dy, \tag{4.831}$$

where c and d are numbers such that the lamina lies between the lines having the equation $y = c$ and $y = d$ and $g(y^*)$ is the length of the interval (or the sum of the lengths of the intervals) in which the line having the equation $y = y^*$ intersects the lamina. In case $p = 0$, the pth moment is the mass of the lamina. In mechanics and some other places, the second moment is called the moment of inertia.

While the facts can be established only by considering the different rectangular coordinate systems in the plane of the lamina, the lamina itself determines a point in the plane of the lamina that is called the *centroid* of the lamina. With reference to the particular coordinate system which we have chosen, the x coordinate of this centroid is the number $\bar{x}$ for which $M^{(1)}_{x=\xi} = 0$ when $\xi = \bar{x}$. Thus

$$\delta \int_a^b (x - \bar{x})f(x)\, dx = 0 \tag{4.84}$$

and it follows that

$$M\bar{x} = \delta \int_a^b xf(x)\, dx, \qquad \bar{x} = \frac{\delta \int_a^b xf(x)\, dx}{\delta \int_a^b f(x)\, dx}, \tag{4.841}$$

where M, the denominator in the second formula, is the mass of the lamina. Similar formulas suffice to determine the y coordinate $\bar{y}$ of the centroid. For example,

$$M\bar{y} = \delta \int_c^d yg(y)\, dy. \tag{4.842}$$

The centroid of a lamina has an important physical property. If the lamina is in a plane perpendicular to the direction of the forces in a uniform parallel force field, then the lamina will balance upon each line (or knife-edge) which passes through the centroid and will balance upon a pin placed at the centroid. It follows that if L is a line of symmetry of a lamina, then the centroid lies on L. Moreover, if a point P is a center of symmetry of a lamina, then the centroid is P.

We now turn our attention to the three-dimensional world which contains, in addition to cubes and spherical balls, so many distractions that relatively few of its inhabitants assimilate substantial information about nonmeasurable sets in E_3. To keep these complicated and paradoxical sets out of our gardens, we should† (and therefore do) start with a set S in E_3 which is assumed to possess positive volume V. In order to be able to use Riemann integrals, we assume that wherever we introduce an x, y, z coordinate system in E_3, there will be numbers a and b for which our set S lies between the planes having the equations $x = a$ and $x = b$. We assume that, for each t for which $a \leqq t \leqq b$, the plane having the equation $x = t$ intersects S in a section having area which we denote by $A(t)$. In many applications this area function is continuous. To be

† This is another situation in which we can be kept on the path of rectitude by knowledge of the contents of Appendix 2 at the end of this book.

fully rigorous about the matter, we assume that the Riemann integral in

$$V = \int_a^b A(x)\,dx \tag{4.85}$$

exists and is the volume V of the set S. Our next step is to suppose that the points (x,y,z) of the set S are points of a material body B such that, for some positive constant δ, a part of the body has mass $\delta\,\Delta V$ if that part occupies a part of the set S having volume ΔV. Our body B (which really is somewhat different from the conglomeration of atomic particles that constitute a potato) is said to be *homogeneous* because its density (mass per unit volume) has the same constant value δ at all places in the body. At last we have a body B which might, for example, be what a child thinks a potato is.

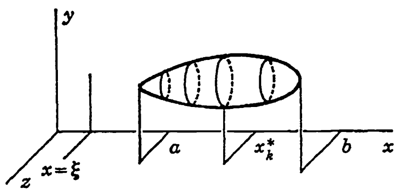


Figure 4.86

Supposing that ξ is a number and that p is 0 or a positive integer, we should now find it easy to construct formulas for calculation of the number $M^{(p)}_{x=\xi}$, *the pth moment of the body B about the plane having the equation* $x = \xi$. Realizing that schematic figures can be helpful even when some wise people consider them to be semisuperfluous, we sketch Figure 4.86. We make a partition P of the interval $a \leqq x \leqq b$ into subintervals. Supposing as usual that $x_{k-1} \leqq x_k^* \leqq x_k$, we use the number

$$A(x_k^*)\,\Delta x_k \tag{4.87}$$

as an approximation to the volume of the slab which lies between the planes having the equations $x = x_{k-1}$ and $x = x_k$. Multiplying by the density δ gives an approximation to the mass of the slab. When the norm of P is small, all points of the slab lie at about the same distance $|x_k^* - \xi|$ from the plane having the equation $x = \xi$ and multiplying the mass by $(x_k^* - \xi)^p$ should therefore give a good approximation to the pth moment of the slab about the plane having the equation $x = \xi$. The Riemann sum

$$\delta\Sigma(x_k^* - \xi)^p A(x_k^*)\,\Delta x_k \tag{4.871}$$

should then be a good approximation to the moment of the whole body. Since the Riemann sums have a limit which is the Riemann integral in the right member of the formula

$$M^{(p)}_{x=\xi} = \delta \int_a^b (x - \xi)^p A(x)\,dx, \tag{4.872}$$

our work motivates the definition by which the required moment is defined by this formula. Analogous formulas define moments about planes parallel to the other coordinate planes.

In case $p = 0$, the pth moment is the mass of the body. We shall not comment upon second moments of solid bodies about planes, but brief comments about first moments may be appropriate. As was the case for laminas, our body B determines a point in E_3 which is called the *centroid* of the body. With reference to the particular coordinate system which we have chosen, the x coordinate of the centroid is the number $\bar{x}$ for which $M^{(1)}_{x=\xi} = 0$ when $\xi = \bar{x}$. Thus

$$\delta \int_a^b (x - \bar{x})A(x)\,dx = 0 \tag{4.88}$$

and it follows that

$$M\bar{x} = \delta \int_a^b xA(x)\,dx, \qquad \bar{x} = \frac{\delta \int_a^b xA(x)\,dx}{\delta \int_a^b A(x)\,dx}, \tag{4.881}$$

where M, the denominator in the second formula, is the mass of the body B. Analogous formulas serve to determine the other coordinates $\bar{y}$ and $\bar{z}$ of the centroid. As was the case for centroids of laminas, the centroid of a body B in E_3 has an important physical property. An ordinary wheel mounted on an axle through its center balances in the gravitational field of the earth which is (so far as an ordinary wheel near the surface is concerned) nearly a uniform parallel force field. Similarly, the body B, when mounted on an axis through its centroid, must balance in a uniform parallel force field. If a plane π is a plane of symmetry of the body B, then the centroid of B is a point in π. If a line L is a line of symmetry of B, then the centroid of B is a point on L. If a point P is a center of symmetry of B, then the centroid of B is P.

All through this section, the moments that have appeared have been "moments of mass," that is, moments of lamina or solid bodies that possess mass. Our methods and formulas are easily modified to produce numbers that are "moments of area," that is, moments of sets in E_2 that possess positive area, and "moments of volumes," that is, moments of sets in E_3 that possess positive volumes. The moments and the centroid of a set S in E_2 which possesses positive area are, by definition, the same as those of the lamina of unit areal density (unit mass per unit area) which coincides with the set. Similarly, the moments and the centroid of a set S in E_3 possessing positive volume are defined to be the moments and the centroid of a body of unit density (unit mass per unit volume) which coincides with S. Thus formulas for moment and centroids of "geometrical" sets are obtained by putting $\delta = 1$ in formulas for "moments and centroids of mass." These concepts are introduced because they are useful. For example, calculations involving forces which bend a beam depend upon a number I which is the moment of inertia of a cross section of the beam about a line through the centroid of the cross section.

Problems 4.89

1 Find the pth moment about the line having the equation $x = \xi$ of the lamina of constant areal density (mass per unit area) δ which occupies the plane region consisting of points (x,y) for which

$$a \leqq x \leqq b, \qquad 0 \leqq y \leqq h. \qquad \textit{Ans.: } \frac{\delta h}{p+1}[(b-\xi)^{p+1} - (a-\xi)^{p+1}].$$

2 A semicircular disk of radius a has its center at the origin and lies in the right half-plane containing points (x,y) for which $x \geqq 0$. Find its centroid.

Ans.: $\bar{x} = \dfrac{4a}{3\pi}$, $\bar{y} = 0$.

3 A homogeneous spherical ball of radius a has its center at the origin. Find the centroid of the hemispherical part of the ball containing points for which $x \geqq 0$. *Ans.:* $\bar{x} = \frac{3}{8}a$, $\bar{y} = 0$, $\bar{z} = 0$.

4 Prove that the centroid of a right circular conical solid of height h has distance $h/4$ from the base of the solid.

5 Find the pth moment about the y axis of the region bounded by the x axis, the line having the equation $x = 1$, and the graph of the equation $y = x^r$, it being supposed that r is a nonnegative constant. *Ans.:* $1/(p + r + 1)$.

6 Copy Figure 1.292 and let T be the triangular region bounded by the triangle having vertices at A, B, C. Set up and evaluate all of the integrals required for evaluation of $|T|$, the area of T, $M^{(1)}_{x=0}$, the first moment of T about the y axis, and $M^{(1)}_{y=0}$, the first moment of T about the x axis. Then use the formulas $|T|\bar{x} = M^{(1)}_{x=0}$ and $|T|\bar{y} = M^{(1)}_{y=0}$ to find $\bar{x}$ and $\bar{y}$. *Remark:* The point $(\bar{x},\bar{y})$ is the point $(2h/3, 0)$. This shows that the centroid of T lies on the median AD. More remarks can be made.

7 This problem involves hydrostatic forces which liquids exert upon surfaces of bodies immersed in them. Before formulating our problem, we digress to eke out some information. If an ordinary rectangular or cylindrical tank has horizontal sections having area A square feet and if the tank is filled to depth d feet with a liquid weighing w pounds per cubic foot, then the total weight of the contents of the tank is wdA. If we divide this total weight wdA by the area A of the base of the tank, we obtain the number wd, which is the weight per square foot that the base supports. This number wd, the product of w and the depth, is called the *pressure* at depth d. This pressure wd is a scalar, the magnitude of the force per unit area which the liquid exerts upon the flat horizontal base of the tank. Our next task is to capture the idea that the jumble of words "pressure in a gas and in a liquid is transmitted (sent across?) equally in all directions" is often presumed to convey. To be very humble about this matter, we can believe or perhaps even know that water will spurt from a hole in the bottom of a tank of water and will spurt almost as vigorously from a hole near the bottom of the tank but in a vertical side of the tank. Fortified by this idea, we can cheerfully accept the ponderous physical principle or law which says that if a plane region having area A is beneath the surface of a liquid, and if d_1 and d_2 are numbers such that each point of the region has a depth d for which $d_1 \leqq d \leqq d_2$, then the force which the liquid exerts upon one side of this region is orthogonal or normal or perpendicular to the region and there is a number d^* such that

$d_1 \leqq d^* \leqq d_2$ and the magnitude of the force is p^*A, where p^* is wd^*, the pressure at depth d^*. Our little lesson in hydrostatics is ended, and we can now formulate our problem. We confine our attention to forces upon one side of a plane region R which lies, as in Figure 4.891, beneath the surface of a liquid and in a vertical plane. The x axis is taken to be horizontal and in the top surface of the liquid. The y axis is taken to be vertical with y positive measured downward; this means that the point (x,y) lies below the x axis when $y > 0$. It is supposed that the region R lies between the lines having the equations $y = a$ and $y = b$ and that, when $a < y^* < b$, the line having the equation $y = y^*$ intersects R in an interval (or collection of intervals) having length (or total length) $f(y^*)$. It is assumed that the region R and the function f are bounded. It is not necessary that f be continuous, but we do assume that R has area $|R|$ and that $|R| = \int_a^b f(x)\,dx$. Our problem is to set up an integral for the magnitude of the force which the liquid exerts upon one side of the region R. The procedure should now be completely familiar. We make a partition of the interval $a \leqq y \leqq b$ into subintervals and choose y_k^* such that $y_{k-1} \leqq y^* \leqq y_k$. The number

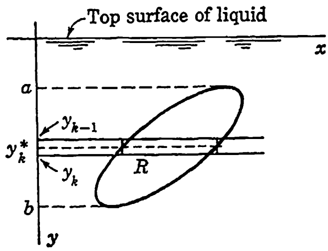


Figure 4.891

(1) $$f(y_k^*)\,\Delta y_k$$

is taken as an approximation to the area of the part of R that lies in a strip parallel to the surface of the liquid. Multiplying this by wy_k^* gives an approximation to the magnitude of the force of the part of R. Since the forces on the parts of R all have the same direction, the sum

(2) $$w\Sigma y_k^* f(y_k^*)\,\Delta y_k$$

gives an approximation to the magnitude $|\mathbf{F}|$ of the force on the whole region R, and the approximation should be good when the norm of the partition is small. Thus it should be true that

(3) $$|\mathbf{F}| = \lim w\Sigma y_k^* f(y_k^*)\,\Delta y_k.$$

Since the right member of (3) exists and is a Riemann integral, our work motivates the definitions where $|\mathbf{F}|$ is defined by the formula

(4) $$|\mathbf{F}| = w\int_a^b yf(y)\,dy.$$

In order to make numerical calculations, we must know or be able to compute a, b, and $f(y)$. The really interesting thing about our result is that it can be put in the form

(5) $$|\mathbf{F}| = w\bar{y}A,$$

where A is the area of the region R and $\bar{y}$ is the depth of its centroid. Thus *the magnitude of the force on the plane region R is the product of the pressure at the centroid and the area of the region.* Many problems can be solved very quickly by use of this fact.

8 Using the fact that the pressure at depth d below the surface of water is wd, but without using more formulas from the preceding problem, find the magnitude of the force exerted upon one face of an isosceles right triangle submerged in water so that one leg is horizontal and 5 feet below the surface while the other leg extends 3 feet upward. *Ans.:* $18w$.

9 According to an examination given at Cornell, the cost, in dollars per mile, of improving the road from Alibab to the Babila oil field 400 miles down the road is 10,000 plus 500 $\sqrt{x}$, where x is the distance from Alibab. Find the total cost of the improvement. *Ans.:* $6 + \frac{2}{3}$ millions.

10 A circle of radius a lies in the xy plane and has its center at the origin. For each positive integer n, points $P_0, P_1, \cdots, P_n$ are equally spaced on the arc of the circle lying in the first quadrant and, for each k for which $1 \leqq k \leqq n$, a vector $\mathbf{r}_k$ is drawn from the origin to a point on the circle between P_{k-1} and P_k. Show that

$$\lim_{n\to\infty} \frac{\mathbf{r}_1 + \mathbf{r}_2 + \cdots + \mathbf{r}_n}{n} = \frac{2a}{\pi}(\mathbf{i} + \mathbf{j}), \tag{1}$$

where, as usual, $\mathbf{i}$ and $\mathbf{j}$ are unit vectors on the x and y axes. *Hint:* Make use of the fact that if

$$\theta_k = \frac{k\pi}{2n}, \qquad \Delta\theta_k = \frac{\pi}{2n}, \tag{2}$$

then

$$\frac{\mathbf{r}_k}{n} = \frac{2a}{\pi}(\cos\theta_k^*\mathbf{i} + \sin\theta_k^*\mathbf{j})\,\Delta\theta_k, \tag{3}$$

where θ_k^* lies between θ_{k-1} and θ_k. The left member of (1) is therefore the limit of a Riemann sum.

11 One way to review Riemann integrals and make them seem simpler is to learn about Riemann-Stieltjes integrals. Let $f(t)$ and $g(t)$ be defined over an interval $a \leqq t \leqq x$, and let P be a partition of the interval $a \leqq t \leqq x$ with partition points t_k and intermediate points t_k^* as in Section 4.2. If there is a number I such that to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that

$$\left| \sum_{k=1}^{n} f(t_k^*)[g(t_k) - g(t_{k-1})] - I \right| < \epsilon \tag{1}$$

whenever $|P| < \delta$, then I is called the *Riemann-Stieltjes integral* of f with respect to g over the interval $a \leqq t \leqq x$ and is denoted by

$$\int_a^x f(t)\,dg(t). \tag{2}$$

These integrals are very important in more advanced mathematics, and some people think that they should at least be mentioned in elementary calculus. Many people have devoted substantial parts of their lives to study of problems for which $f(t) = t^z$. Start picking up ideas by evaluating (2) when $a = -1$, $x = 1$, $f(t) = t^2 + 2t + 3$, and $g(t) = \operatorname{sgn} t$. *Ans.:* 6.

4.9 *Simpson and other approximations to integrals*

When f is a polynomial in x, and in some other cases, we can discover an elementary function F for which $\int f(x)\,dx = F(x)$ or $f(x) = F'(x)$ and can then evaluate $\int_a^b f(x)\,dx$ by the calculation

$$\int_a^b f(x)\,dx = F(x)\Big]_a^b = F(b) - F(a). \tag{4.91}$$

As was pointed out in Section 3.6, derivatives of elementary functions are always elementary functions that can be calculated by use of appropriate rules. It must not be presumed, however, that if f is an elementary function, then there must exist an elementary function F for which $f(x) = F'(x)$. While proofs of such things do not grow in ordinary gardens, it is nevertheless known that if $f(x)$ is one or another of

$$\sqrt{1+x^4},\qquad \frac{\sin x}{x},\qquad \sqrt{1+\sin^2 x},\qquad \frac{1}{\sqrt{4-\sin^2 x}},\qquad \frac{e^x}{x},\qquad e^{-x^2},$$

then there is no elementary function F for which $f(x) = F'(x)$.

This section is devoted to methods by which we can obtain useful decimal approximations to

$$\int_a^b f(x)\,dx \tag{4.92}$$

in cases where it is impossible or difficult to obtain a useful formula for a function F such that (4.91) holds. Some pedestrian methods are worthy of brief mention. When a reasonably accurate graph of f is drawn on graph paper as in Figure 4.93, we can obtain an informative approximation by counting the squares and estimating the partial squares that lie within the appropriate region. Chemists and others who have access to scissors and appropriate scales can cut out the region and weigh the paper. Another method involves use of a *planimeter*, an instrument which will reveal a useful approximation to the area of a region after it has been suitably adjusted and a needle point on a movable arm has traced the boundary of the region. In some situations, the simplest and most direct method is illustrated by Figure 4.931. The interval $a \leqq x \leqq b$ is cut into n equal subintervals of length h, where $h = (b-a)/n$, and a point x_k^* is selected in the kth subinterval. Then

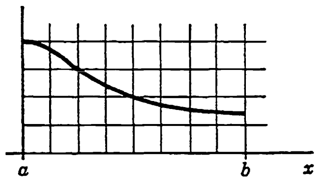


Figure 4.93

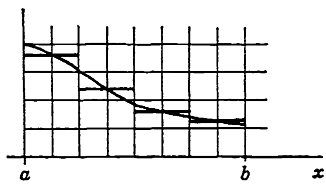


Figure 4.931

$$\int_a^b f(x)\,dx = \epsilon + h\sum_{k=1}^{n} f(x_k^*), \tag{4.932}$$

where ϵ is an error term and the sum is a particular Riemann sum. Of course, we should try to minimize errors by choosing the heights of the rectangles in such a way that, in each strip, the area of the set which lies in the region but outside the rectangle is nearly equal to the area of the set which lies inside the rectangle but outside the region.

This paragraph introduces the *trapezoidal formula*

$$\int_a^b f(x)\,dx = \epsilon + h\left[\frac{y_0}{2} + y_1 + y_2 + y_3 + \cdots + y_{n-1} + \frac{y_n}{2}\right], \tag{4.94}$$

the derivation of which will help us to understand the much better formula (4.95) which will appear in the next paragraph. We separate the interval $a \leqq x \leqq b$ into n equal subintervals of length h, where $h = (b - a)/n$, by points $x_0, x_1, \cdots, x_n$ such that $x_0 = a$, $x_n = b$, and $x_k = x_{k-1} + h$ for each $k = 1, 2, \cdots, n$. As in Figure 4.943, where $n = 4$, we let $y_k = f(x_k)$ for each k. As an approximation to $\int_{x_0}^{x_1} f(x)\,dx$ we use $\int_{x_0}^{x_1} L(x)\,dx$, where $L(x) = Ax + B$ and the constants are chosen such that the graph of $L(x) = Ax + B$ is a line passing through the two points $P_0(x_0,y_0)$ and $P_1(x_1,y_1)$. The details of the calculation

$$\int_{x_0}^{x_1} L(x)\,dx = \int_{x_0}^{x_1}\left[y_0 + \frac{y_1 - y_0}{h}(x - x_0)\right]dx = h\frac{y_0 + y_1}{2} \tag{4.941}$$

are easily supplied; in case y_0 and y_1 are positive, the details are superfluous because the quantities are equal to the area of a trapezoid and elementary geometry shows that the formula is correct. Using (4.941) and analogous formulas, we see that

$$\int_{x_{k-1}}^{x_k} f(x)\,dx = \epsilon_k + h\frac{y_{k-1} + y_k}{2}, \tag{4.942}$$

where the "error term" ϵ_k will be "relatively small" if the graph of f over the interval $x_{k-1} \leqq x \leqq x_k$ is "near" the chord joining P_{k-1} and P_k. Summing the members of (4.942) gives the trapezoidal formula (4.94).

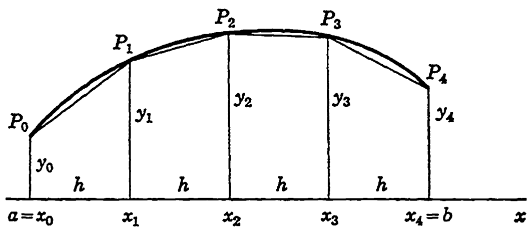


Figure 4.943

To derive the trapezoidal formula (4.94), we began by approximating $f(x)$ over the interval $x_0 \leqq x \leqq x_1$ by the function $L(x)$ whose graph is a line passing through the two points P_0 and P_1. To derive the more useful *Simpson formula*

$$\int_a^b f(x)\,dx = \epsilon + \frac{h}{3}[y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 4y_{n-1} + y_n], \tag{4.95}$$

in which n is always an even positive integer, we begin by approximating $f(x)$ over the interval $x_0 \leqq x \leqq x_2$ by the function $Q(x)$ of the form

$$Q(x) = A(x - x_1)^2 + B(x - x_1) + C \tag{4.951}$$

whose graph passes through the three points $P_0(x_0,y_0)$, $P_1(x_1,y_1)$, and $P_2(x_2,y_2)$. Since (4.951) can be put in the form $Q(x) = Ax^2 + B_1x + C_1$, its graph is a parabola if $A \neq 0$ and is a line if $A = 0$. As is easy to guess, the graph of $Q(x)$ is ordinarily a much better approximation to the arc P_0P_2 than the graph consisting of the two straight chords P_0P_1 and P_1P_2 is, and hence the error term in the Simpson formula is ordinarily much nearer 0 than the error term in the trapezoidal formula. We find that

$$\int_{x_0}^{x_2} Q(x)\,dx = \left[A\frac{(x - x_1)^3}{3} + B\frac{(x - x_1)^2}{2} + C(x - x_1)\right]_{x_1-h}^{x_1+h}$$

so

$$\int_{x_0}^{x_2} Q(x)\,dx = \frac{h}{3}[2Ah^2 + 6C]. \tag{4.952}$$

The three formulas

$$\begin{aligned} y_0 &= Q(x_0) = Q(x_1 - h) = Ah^2 - Bh + C \\ y_1 &= Q(x_1) = C \\ y_2 &= Q(x_2) = Q(x_1 + h) = Ah^2 + Bh + C \end{aligned}$$

enable us to determine A, B, C in terms of y_0, y_1, y_2. It serves our purpose, however, to add the first and last of the formulas to obtain

$$y_0 + y_2 = 2Ah^2 + 2C$$

and to note that $4y_1 = 4C$ so

$$y_0 + 4y_1 + y_2 = 2Ah^2 + 6C.$$

This and (4.952) give the formula

$$\int_{x_0}^{x_2} Q(x)\,dx = \frac{h}{3}[y_0 + 4y_1 + y_2]. \tag{4.953}$$

Using (4.953) and analogous formulas, we see that

$$\begin{aligned} \int_{x_0}^{x_2} f(x)\,dx &= \epsilon_1 + \frac{h}{3}[y_0 + 4y_1 + y_2], \\ \int_{x_2}^{x_4} f(x)\,dx &= \epsilon_2 + \frac{h}{3}[y_2 + 4y_3 + y_4], \quad \cdots, \\ \int_{x_{n-2}}^{x_n} f(x)\,dx &= \epsilon_{n/2} + \frac{h}{3}[y_{n-2} + 4y_{n-1} + y_n]. \end{aligned}$$

Adding these gives the Simpson formula

$$\int_a^b f(x)\,dx = \epsilon + \frac{h}{3}[y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 4y_{n-1} + y_n], \tag{4.96}$$

which appears in (4.95) and is so important that it merits reproduction. We recall that n must be even and that $h = (b - a)/n$. Whenever f is Riemann integrable over the interval $a \leqq x \leqq b$, the error term ϵ is near zero when n is large. When f is continuous, and perhaps in some other cases as well, experienced operators of pencils and slide rules and calculators and electronic computers neglect the ϵ and habitually use the remaining Simpson sum in the right member of (4.96) as an approximation to the integral. A particular sum is often judged to be as accurate as desired when this sum agrees to the desired number of decimal places with the sum obtained by doubling n. In many practical applications, surprisingly small values of n yield the desired accuracy.

Nearly everyone who understands the trapezoidal and Simpson formulas generates the following idea. It should be possible to derive still better formulas by approximating f by polynomials of higher degree having graphs passing through more of the points $P_0, P_1, P_2, P_3, \cdots$. It turns out, however, that these formulas are more complicated than the Simpson formula, and using them for a given h is not as satisfactory as using the Simpson formula with a smaller h.

Problems 4.99

1 Tables give

$$\int_1^2 \frac{1}{x}\,dx = \log 2 = 0.69314\,71806.$$

Show that the trapezoidal formula with $n = 4$ gives $h = \frac{1}{4}$, $y_0 = 1$, $y_1 = \frac{4}{5}$, $y_2 = \frac{4}{6}$, $y_3 = \frac{4}{7}$, $y_4 = \frac{4}{8}$, and

$$\int_1^2 \frac{1}{x}\,dx = \epsilon + \tfrac{1}{4}[\tfrac{1}{2} + \tfrac{4}{5} + \tfrac{4}{6} + \tfrac{4}{7} + \tfrac{1}{4}] = \epsilon + 0.69702\,4$$

and that use of the Simpson formula with $n = 4$ gives $h = \frac{1}{4}$ and

$$\int_1^2 \frac{1}{x}\,dx = \epsilon + \tfrac{1}{12}[1 + \tfrac{16}{5} + \tfrac{8}{6} + \tfrac{16}{7} + \tfrac{1}{2}] = \epsilon + 0.69325\,4.$$

Show that the error terms are respectively -0.003877 and -0.000107. Observe that it is almost equally easy to use the trapezoidal and Simpson formulas. Remember that properly educated persons use the Simpson formula whenever suitable occasions arise, but that they rarely if ever use the trapezoidal formula.

2 Tables give

$$\log 2.5 = 0.91629\,07319.$$

Using the Simpson formula with two subintervals, obtain the approximation

$$\int_2^{2.5} \frac{1}{x}\,dx = \frac{1}{12}\left[\frac{1}{2} + \frac{4}{2.25} + \frac{1}{2.5}\right] = 0.22314\,8.$$

Show how this and the last numerical result of Problem 1 give the approximation

$$\log 2.5 = 0.91640\,2.$$

3 Using the Simpson formula with $n = 2$, obtain the approximations

$$\int_{2.5}^{2.7} \frac{1}{x}\,dx = \frac{0.1}{3}\left[\frac{1}{2.5} + \frac{4}{2.6} + \frac{1}{2.7}\right] = 0.07696\ 106$$

and

$$\int_{2.7}^{2.718} \frac{1}{x}\,dx = \frac{0.009}{3}\left[\frac{1}{2.7} + \frac{4}{2.709} + \frac{1}{2.718}\right] = 0.00664\ 454.$$

Use these formulas and the first formula of Problem 2 to obtain the approximation

$$\log 2.718 = 0.99989\ 633.$$

Remark: With a little skill and a desk calculator that makes divisions, it is not difficult to extend these calculations to obtain good approximations to the number

$$e = 2.71828\ 18284\ 59045$$

for which $\int_1^e \frac{1}{x}\,dx = 1$ and $\log e = 1$. Better ways to approximate logarithms and e will appear later.

4 Someday we will learn the formulas

$$\int \frac{1}{1+x^2}\,dx = \tan^{-1} x + c, \qquad \int_0^1 \frac{1}{1+x^2}\,dx = \frac{\pi}{4} = 0.78539\ 81634.$$

Use the Simpson formula to find approximations to the last of these integrals, and find the errors in the approximations, to obtain the numbers in the first two or three rows of the following table.

n	*Simpson value*	*Error*
2	.78333 332	.00206 484
4	.78539 212	.00000 604
6	.78539 782	.00000 034
8	.78539 802	.00000 014
10	.78539 809	.00000 007
12	.78539 812	.00000 004
14	.78539 812	00000 004
16	78539 809	.00000 007
18	.78539 812	.00000 004
20	.78539 809	.00000 007
40	.78539 789	.00000 027
60	.78539 782	.00000 034
80	.78539 782	.00000 034
100	.78539 769	.00000 047
200	.78539 769	.00000 047
400	.78539 569	.00000 247
600	.78539 465	.00000 351
800	.78539 425	.00000 391
1000	.78539 395	.00000 421
10000	.78535 725	.00004 091
15000	.78535 442	.00004 374
20000	.78535 265	.00004 551
100000	.78499 059	.00040 757

Remark: The last cases show results obtained from an electronic computer that makes 8D calculations. When n is large, rounding errors seriously affect the last digit or digits kept.

5 A loaded freighter is anchored in still water. At water level, the boat is 200 feet long and, for each $k = 0, 1, 2, \cdots, 20$, has breadth y_k at distance $10k$ feet from the prow. Assign semireasonable numerical values to the numbers y_k and do not allow anyone to claim that you have not partially designed a boat. Then use Mr. Simpson's idea to approximate the area of the water-level section of your boat. Finally, recall an exploit of Archimedes and make an estimate of the number of tons of freight that should be removed in order to raise your boat 1 foot.

6 Use the Simpson formula to obtain decimal approximations to the following integrals. Keep two and three decimal places in the calculations, use a slide rule if possible, and use the value of n given in parentheses.

(*a*) $\int_0^4 x^2\,dx$, $(n = 4)$ (*b*) $\int_0^4 x^3\,dx$, $(n = 4)$

(*c*) $\int_{100}^{101} \frac{1}{x}\,dx$, $(n = 2)$ (*d*) $\int_0^\pi \sin x\,dx$, $(n = 6)$

(*e*) $\int_0^\pi \sqrt{\sin x}\,dx$, $(n = 6)$ (*f*) $\int_0^\pi \frac{\sin x}{x}\,dx$, $(n = 6)$

(*g*) $\int_0^1 (1 + x^2)^{1/3}\,dx$, $(n = 4)$ (*h*) $\int_0^1 e^{-x^2}\,dx$, $(n = 10)$

(*i*) $\int_1^2 e^{-x^2}\,dx$, $(n = 10)$ (*j*) $\int_2^3 e^{-x^2}\,dx$, $(n = 4)$

7 Using the fact that $x^2 > 3x$ when $x > 3$, show that

$$\int_3^\infty e^{-x^2}\,dx < \int_3^\infty e^{-3x}\,dx = -\frac{1}{3}e^{-3x}\Big]_3^\infty = \frac{1}{3}e^{-9} \doteq \frac{1}{3}\frac{1}{(20)^3} = \frac{1}{24{,}000}.$$

8 Using the notation and ideas employed to derive (4.953), prove that if the graph of the function f for which

$$f(x) = K(x - x_1)^3 + A(x - x_1)^2 + B(x - x_1) + C \tag{1}$$

contains the three points $P_0(x_0, y_0)$, $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, then

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}[y_0 + 4y_1 + y_2]. \tag{2}$$

Remark: This result shows that the error term is zero and the Simpson formula gives the exact value of the integral when f is a polynomial of degree three or less. Thus we catch the idea that the Simpson formula gives good approximations even when the integrand cannot be closely approximated over the intervals $x_k \leqq x \leqq x_{k+2}$ by quadratic polynomials but can be closely approximated over the intervals by cubic polynomials. Further investigation shows that if we add to the right member of (1) an integrable term $\phi(x)$ for which $|\phi(x)| \leqq M(x - x_1)^4$, then (2) will contain an error term ϵ for which $|\epsilon| \leqq (\frac{2}{5})Mh^5$.

9 An application of (2) of Problem 8 gives the famous old *prismoidal formula*

$$V = \frac{H}{6}[|B_1| + 4|M| + |B_2|] \tag{3}$$

for volumes of solids. To investigate this matter, put $x_0 = a$ and $x_2 = a + H$ in (2) to obtain

$$\int_a^{a+H} f(x)\,dx = \frac{H}{6}\left[f(a) + 4f\left(a + \frac{H}{2}\right) + f(a + H)\right]. \tag{4}$$

If a reasonably decent solid has bases in the planes having the equations $x = a$ and $x = a + H$, and if for each x' for which $a \leqq x' \leqq a + H$ the plane having the equation $x = x'$ intersects the solid in a plane region having area $f(x')$, then the left member of (4) is the volume V of the solid. The quantity in brackets in (4) is the sum of the area $|B_1|$ of one base B_1, the area $|B_2|$ of the other base B_2, and four times the area $|M|$ of the section M midway between the two bases. Thus the formula (3) is correct when the solid has volume V equal to the left member of (4) and $f(x)$ has the form

$$f(x) = K_1x^3 + K_2x^2 + K_3x + K_4.$$

Nearly everyone acquires substantial respect for the prismoidal formula when it is discovered that the formula yields the correct formula for the volume of a spherical ball of radius a. In this case $H = 2a$, the bases are points having area 0, and the midsection M is an equatorial disk having area πa^2.

10 While the matter cannot be fully explored in a course in elementary calculus, we can know that persons who study Lebesgue measure and integration may learn that E_3 contains sets much queerer than those considered in this book. It can happen that each plane section perpendicular to the x axis is a square of unit area so that (in the context of Problem 9) $f(x) = 1$ when $0 \leqq x \leqq 1$, but, nevertheless, the squares are so heterogeneously scattered that the set fails to possess a volume. For such queer sets the prismoidal formula is invalid because the left member of (4) of Problem 9 is not the volume of the set. Experts in the theory of measure can have sympathy for students of solid geometry who are a bit mystified by the "Cavalieri theorem." This "theorem" says that two sets in E_3 have equal volumes if they have parallel bases and equal altitudes, and if each plane parallel to the bases intersects the two sets in two plane regions having equal areas. The queer sets which we have mentioned show that the "theorem" is false. Appendix 2 at the end of this book shows how we can reconcile ourselves to these matters. Some of us will learn more about these things than others, but we can all know that there is much to be learned.

5 *Functions, graphs, and numbers*

5.1 Graphs, slopes, and tangents It is quite possible that we first heard about tangents, or tangent lines, when we were very young. We may have been shown a circle as in Figure 5.11 and have been solemnly told that some lines in the plane of the circle intersect the circle twice, some others do not intersect the circle at all, and some others, the tangents to the circle, intersect the circle just once. When graphs more complicated than circles appear, no such simple story can adequately describe tangents. For example, the line T of Figure 5.12 intersects the graph twice and seems to be tangent to the graph at P_0, while the line L intersects the graph only once and does not seem to be tangent to the graph. To attack this rather delicate matter, we start with a given function f defined over some interval and draw the graph G of $y = f(x)$ as in Figure 5.13. We next select an x within the domain of f and call it

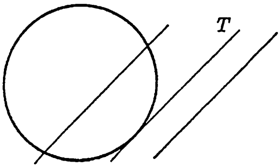


Figure 5.11

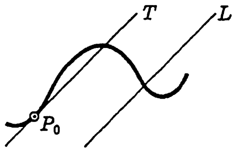


Figure 5.12

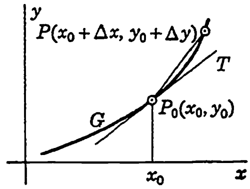


Figure 5.13

x_0 to emphasize the fact that it remains fixed throughout our discussion. Our task is to try to decide what we should mean when we say that a line T is tangent to G at $P(x_0,y_0)$. We gain the possibility of making progress when we choose a number Δx for which $\Delta x \neq 0$, plot the point $P(x_0 + \Delta x, y_0 + \Delta y)$ on G, and draw the chord joining our two points on G. Our first feeble idea can be that T is tangent to G at $P(x_0,y_0)$ if the chord is nearly coincident with T whenever Δx is near zero. We can, so far as nonvertical tangents are concerned, improve this idea to gain the concept that the line T through $P(x_0,y_0)$ having slope m is tangent to G at $P(x_0,y_0)$ if the slope $\Delta y/\Delta x$ of the chord is near m whenever Δx is near 0. We know how to express this concept in terms of limits and derivatives, and we do it in the following definition.

Definition 5.14 *If $f'(x_0)$ exists, then the line T through the point (x_0,y_0) having slope $f'(x_0)$ is said to be tangent to the graph of $y = f(x)$ at the point (x_0,y_0). If $f'(x_0)$ fails to exist, then the graph fails to possess a nonvertical tangent at the point (x_0,y_0).*

From this definition and the point-slope formula for the equation of a line, we obtain the following theorem.

Theorem 5.141 *If $f'(x_0)$ exists, then the equation*

$$y - y_0 = f'(x_0)(x - x_0)$$

is the equation of the tangent to the graph of $y = f(x)$ at the point (x_0,y_0).

To assist in the development and communication of ideas, it turns out to be exceptionally useful to agree that if a graph has a nonvertical tangent at a point (x_0,y_0), then the slope of this tangent will be called the slope of the graph at the point (x_0,y_0). In accordance with this idea, we adopt the following definition.

Definition 5.15 *If $f'(x_0)$ exists, then $f'(x_0)$ is said to be the slope of the graph of $y = f(x)$ at the point $P(x_0,y_0)$.*

In order to obtain a full understanding of tangents to graphs, and for other purposes, it is helpful to know about "lines of support" of graphs and other point sets that lie in a plane. We confine attention here to cases in which f is a continuous function defined over $a \leqq x \leqq b$ and $P_0(x_0,y_0)$ is a point on the graph of $y = f(x)$ for which $a < x_0 < b$. A line L through P_0 is said to be a *line of support* of the graph of $y = f(x)$ if there is a positive number δ such that the part of the graph of $y = f(x)$

for which $x_0 - \delta < x < x_0 + \delta$ lies entirely on or above L or lies entirely on or below L. To emphasize that tangent lines were defined as we defined them because of custom and not because of logical necessity, we can imagine that a man from Mars might come to our earth with a language identical with ours except that his meanings of the terms "line of support" and "tangent line" could be obtained by interchanging ours. This man from Mars might wonder why on earth we study our tangent lines instead of his. The problems at the end of this section may provide reasons.

To be honorable, we must show that the remark made in Section 3.7 about tangents to curves is in agreement with the ideas of this section. Putting $z(t) = 0$ gives the assertion that if $\mathbf{r}(t)$ is the vector $\overrightarrow{OP}$ running from the origin to a particle P which traverses a curve C as t increases, and if

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}, \tag{5.16}$$

where x and y are differentiable functions of t for which $\mathbf{r}'(t) \neq 0$, then, for each t, the vector

$$\mathbf{r}'(t) = x'(t)\mathbf{i} + y'(t)\mathbf{j} \tag{5.161}$$

is tangent to the path. In case the particle P always lies on the graph of the equation $y = f(x)$, we always have $y(t) = f(x(t))$. Therefore,

$$\mathbf{r}(t) = x(t)\mathbf{i} + f(x(t))\mathbf{j}, \tag{5.162}$$

and differentiating with the aid of the chain rule gives the result that, at each time t, the vector

$$\mathbf{r}'(t) = x'(t)[\mathbf{i} + f'(x(t))\mathbf{j}] \tag{5.163}$$

is tangent to the graph. The hypothesis that $\mathbf{r}'(t) \neq 0$ implies that $x'(t) \neq 0$. Since $x'(t)$ is a nonzero scalar, our result is equivalent to the statement that, for each x, the vector

$$\mathbf{i} + f'(x)\mathbf{j} \tag{5.164}$$

having its tail at the point (x,y) on the graph is tangent to the graph at the point. With or without the aid of Figure 5.165, we can see that this vector lies on the line through (x,y) having slope $f'(x)$. Thus the tangent line obtained by use of vectors is the same as the tangent line obtained by use of slopes.

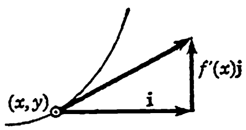


Figure 5.165

The remainder of the text of this section is devoted to a useful theorem which is, from our present point of view, thoroughly difficult. The theorem is important because it gives precise information that is very often used. The proof of the theorem shows that we must learn more mathematics before we can fully comprehend the details. In the worst of circumstances, we are like a person who cannot swim but is thrown into the water and given a chance to fight for his life. Most of us will soon be swimming around in the scientific oceans, and Theorem 5.17 will slowly metamorphose from an ugly demon to a friendly angel. The theorem is closely related to the preceding paragraph and to the chain rule, but it is different from both. Using different notation, it sets forth conditions under which the first of the two equations

$$x = f_1(t), \qquad y = f_2(t)$$

can be "solved" for t and the result substituted in the second equation to obtain y as a function of x. Moreover, the theorem tells how we can find a formula for the derivative of y with respect to x even though we cannot work out a useful formula that gives y in terms of x. The usefulness of the theorem and the difficulty of the proof are both due to the fact that the conclusion of the theorem guarantees existence of various things. If we replace the condition $x'(t) > 0$ by the condition $x'(t) < 0$ in the theorem, the intervening details become somewhat different but the final conclusion (5.171) is valid. We could say that the theorem is a theorem about *elimination of parameters*, but in case $f_2(t) = t$ so $t = y$ it is an inverse-function theorem.

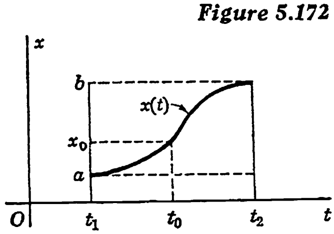
Theorem 5.17 *Let $x(t)$ and $y(t)$ be continuous over the closed interval $t_1 \leqq t \leqq t_2$ and be differentiable over the open interval $t_1 < t < t_2$ and let $x'(t) > 0$ when $t_1 < t < t_2$. Let $x(t_1) = a$ and $x(t_2) = b$. Then $a < b$, and to each x_0 for which $a < x_0 < b$ there corresponds exactly one t_0 for which $t_1 < t_0 < t_2$ and $x(t_0) = x_0$, and t_0 in turn determines exactly one y_0 for which $y_0 = y(t_0)$. This correspondence between numbers x_0 and y_0 determines a function f for which $y_0 = f(x_0)$ when $a < x_0 < b$, and hence $y = f(x)$ when $a < x < b$.* Moreover, *this function f is differentiable and the first of the formulas*

$$f'(x) = \frac{y'(t)}{x'(t)}, \qquad \frac{dy}{dx} = \frac{\dfrac{dy}{dt}}{\dfrac{dx}{dt}} \tag{5.171}$$

is valid when $x = x(t)$ and $t_1 < t < t_2$. The second is also valid when it is understood to mean what the first does.

To help us understand the things we do to prove this theorem, we start sketching Figure 5.172. We mark the points (t_1,a) and (t_2,b) in a tx plane. For a schematic graph of $x(t)$, we sketch a curve headed upward to the right because we think it should be so because $x'(t) > 0$. Theorem

Figure 5.172



5.27, which we do not bother to read now, proves that this idea is correct and that $a < b$. We next mark x_0 such that $a < x_0 < b$. We can easily believe that the rising graph of $x(t)$ must intersect the dotted horizontal line at exactly one point (t_0,x_0). Theorems 5.48 and 5.27 prove that this is correct, and we now have t_0. Our given function y, of which we have not sketched a schematic graph, then determines the number y_0 defined by $y_0 = y(t_0)$, and we put $y_0 = f(x_0)$ in Figure 5.173. This gives one point on the graph of $y = f(x)$, and the same procedure gives each other point on the graph. We now have the formula $y(t) = f(x(t))$. If we had proof that f is differentiable, we could apply the chain rule to obtain $y'(t) = f'(x(t))x'(t)$ and divide by $x'(t)$ to get our answer, but this will not work because we do not yet have the required proof. We therefore start a direct attack upon difference quotients by taking a fixed t for which $t_1 < t < t_2$ and writing

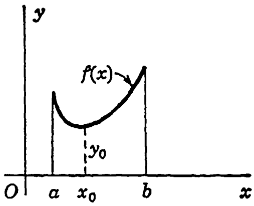


Figure 5.173

$$f(x(t + \Delta t)) - f(x(t)) = y(t + \Delta t) - y(t). \tag{5.174}$$

Dividing by $x(t + \Delta t) - x(t)$ gives the more promising formula

$$\frac{f(x(t + \Delta t)) - f(x(t))}{x(t + \Delta t) - x(t)} = \frac{y(t + \Delta t) - y(t)}{x(t + \Delta t) - x(t)}. \tag{5.175}$$

Since $y'(t)$ and $x'(t)$ both exist and $x'(t) \neq 0$, we can divide the numerator and denominator of the right side by Δt and see that the right side has the limit $y'(t)/x'(t)$ as $\Delta t \to 0$. The left side therefore has the same limit and we obtain

$$\lim_{\Delta t \to 0} \left[\frac{f(x(t + \Delta t)) - f(x(t))}{x(t + \Delta t) - x(t)}\right] = \frac{y'(t)}{x'(t)}. \tag{5.176}$$

This seems to be almost the desired result (5.171), but we must use it to obtain additional information. Let $\epsilon > 0$. Choose a positive number δ_1 such that

$$\left|\frac{f(x(t + \Delta t)) - f(x(t))}{x(t + \Delta t) - x(t)} - \frac{y'(t)}{x'(t)}\right| < \epsilon \tag{5.177}$$

whenever $|\Delta t| < \delta_1$. Another appeal to theorems given later in this chapter shows that there is a positive number δ_2 such that when $|h| < \delta_2$, there is a number Δt for which $|\Delta t| < \delta_1$ and

$$x(t + \Delta t) = x(t) + h. \tag{5.178}$$

It follows that

$$\left|\frac{f(x(t) + h) - f((x(t))}{h} - \frac{y'(t)}{x'(t)}\right| < \epsilon \tag{5.179}$$

whenever $|h| < \delta_2$. This gives (5.171) and completes the proof of Theorem 5.17.

As was remarked, the proof of Theorem 5.17 is difficult because existence of various things must be proved. To help us understand that questions involving existence and differentiability of functions can be significant, we look at an example. Let us assume that y is a differentiable function of x for which

$$x^2 + y^2 + \sin x + \sin y + 46 = 0. \tag{5.18}$$

Differentiating with respect to x with the aid of the chain rule then gives

$$2x + 2y\frac{dy}{dx} + \cos x + \cos y\frac{dy}{dx} = 0 \tag{5.181}$$

or

$$(2y + \cos y)\frac{dy}{dx} = -(2x + \cos x) \tag{5.182}$$

and, when $(2y + \cos y) \neq 0$, dividing by $(2y + \cos y)$ gives a formula for dy/dx. The formula is illusory, however, because the original assumption is incorrect. The inequalities $x^2 \geqq 0$, $y^2 \geqq 0$, $\sin x \geqq -1$, and $\sin y \geqq -1$ imply that, whatever x and y may be,

$$x^2 + y^2 + \sin x + \sin y + 46 \geqq 44. \tag{5.183}$$

Consequently, there are no numbers x and y for which (5.18) is true. The assumption that there is a differentiable function f, defined over some interval $a < x < b$, such that

$$x^2 + [f(x)]^2 + \sin x + \sin f(x) + 46 = 0 \qquad (a < x < b) \tag{5.184}$$

is false. This example can help us understand the nature of Theorem 5.17. The theorem is not a weak one which tells what dy/dx must be if it exists. The theorem sets forth conditions under which dy/dx must exist and gives a formula which must be correct when these conditions are satisfied. Proof of a weak theorem can be obtained by mixing a few words with the calculation

$$\frac{dy}{dx} = \lim_{\Delta x \to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta t \to 0}\frac{\dfrac{\Delta y}{\Delta t}}{\dfrac{\Delta x}{\Delta t}} = \frac{\dfrac{dy}{dt}}{\dfrac{dx}{dt}}, \tag{5.185}$$

but this one line is very far from the equivalent of a theorem which sets forth conditions under which y is a differentiable function of x and the formula is valid. Examples show that matters involving (5.185) are not always completely simple. The distance r from Earth to Mars and the blood pressure p of a particular yogi are both functions of time t, but the unqualified assertion that one of r and p is a differentiable function of the other is quite dubious.

Returning to simpler considerations, we note that if the graph of a function has a tangent line at a point P on the graph, then the line through P perpendicular to the tangent is called the *normal* to the graph at P.

Problems 5.19

1 Find the equation of the line tangent to the graph of the given equation at the given point

(*a*) $y = x^2$, $(1,1)$ *Ans.:* $y - 1 = 2(x - 1)$
(*b*) $y = \sin 2x$, $(0,0)$ *Ans.:* $y = 2x$
(*c*) $y = x \log x$, $(1,0)$ *Ans.:* $y = x - 1$
(*d*) $y = e^{ax}$, $(0,1)$ *Ans.:* $y = ax + 1$
(*e*) $y = \sin x^2$, $(0,0)$ *Ans.:* $y = 0$
(*f*) $y = x \cos x$, $(2\pi,2\pi)$ *Ans.:* $y = x$
(*g*) $y = (x + x^2)^5$, $(1,32)$ *Ans.:* $y = 240x - 208$

2 Find the equation of the tangent to the graph of the equation $y = x^n$ at the point (x_1,x_1^n). *Ans.:*

$$y = nx_1^{n-1}x - (n - 1)x_1^n.$$

3 First find the slopes of the graph of the equation $y = x^3$ at the points for which $x = -1$, $x = -\frac{1}{2}$, $x = 0$, $x = \frac{1}{2}$, and $x = 1$. Use this information to help construct a figure showing the graph and five tangents.

4 Find the area of the region bounded by the graph of $y = x^3$ and the tangent to this graph at the point $(1,1)$. *Ans.:* $\frac{27}{4}$.

5 Even a crude graph suggests that at least one line can be drawn through the point $(-2,-3)$ tangent to the graph of the equation $y = x^2 + 2$. Investigate this matter.

6 Sketch reasonably accurate graphs of $y = \sin x$, $y = x$, and $y = -x$ over the interval $-2\pi \leqq x \leqq 4\pi$. Let

$$f(x) = x \sin x$$

and, after observing that $f(x) = 0$ when $\sin x = 0$, $f(x) = x$ when $\sin x = 1$, and $f(x) = -x$ when $\sin x = -1$, sketch a graph of $f(x)$. It is easy to guess that the graph of $y = f(x)$ is tangent to the graph of $y = x$ wherever $\sin x = 1$ and that the graph of $y = f(x)$ is tangent to the graph of $y = -x$ wherever $\sin x = -1$. Prove that it is so. *Hint:* Calculate $f'(x)$ and observe that $\cos x = 0$ wherever $\sin x$ is 1 or -1.

7 As we know, the part of the graph of the equation

$$y = \sqrt{a^2 - x^2} = (a^2 - x^2)^{1/2}$$

for which $-a < x < a$ is an" upper semicircle" with center at the origin. Let $P_0(x_0,y_0)$ be a point on this graph. Use definitions or theorems of this section to prove that the graph has exactly one tangent at P_0 and that the equation of this tangent is

$$y - y_0 = -\frac{x_0}{y_0}(x - x_0).$$

Prove that this line is perpendicular to the line joining the origin to P_0 and hence that the definition of tangent given in this section is in agreement with ideas of tangents employed in elementary plane geometry.

8 With or without more critical investigation of the matter, sketch a figure which indicates that the graph of the preceding problem has exactly one line of support at P_0.

9 If $x = a \cos t$ and $y = a \sin t$, it is easy to make the calculation

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{a \cos t}{-a \sin t}$$

over each interval of values of t for which $\sin t > 0$ or $\sin t < 0$. Letting $x_0 = a \cos t_0$ and $y_0 = a \sin t_0$, we find that the equation of the tangent to the graph at the point $P_0(x_0,y_0)$ is

$$y - y_0 = -\frac{a \cos t_0}{a \sin t_0}(x - x_0) \qquad \text{or} \qquad y - y_0 = -\frac{x_0}{y_0}(x - x_0).$$

Sketch a graph which shows the geometric interpretations of these things.

10 Find the equation of the tangent to the graph of $y = x^3$ at the origin. Sketch the graph and show that it does not have a line of support at the origin.

11 Draw a graph of the equation $y = |x|$. Show that this graph has no tangent at the origin but does have many lines of support. *Remark:* Our word "tangent" has its root in a Latin verb meaning "to touch," and a mathematician from Mars can defend his contention that our lines of support are "touching lines" and hence should be called tangents. We must, however, stick to our guns and insist that, in languages used on earth, these lines are not tangents.

12 Sketch the graph of $y = \sin x$ and the normal to the graph at the point $(x, \sin x)$. The normal intersects the x axis at the point $(f(x), 0)$. Determine whether $f(x)$ increases as x increases. *Hint:* Borrow, from the next section, the unsurprising fact that $f(x)$ is increasing over an interval if $f'(x) > 0$ over the interval.

13 In connection with Problem 12, we note that problems in applied mathematics sometimes involve extraneous material that may obscure their mathematical aspects. A witch with a broom sweeps the x axis while walking along the graph of $y = \sin x$ in such a way that x is always increasing. She keeps the handle of her broom perpendicular to her path. Is the broom always pushing dust to the right?

14 The two formulas

$$\frac{d}{dx} \sin x = \cos x, \qquad \frac{d}{dx} \cos x = -\sin x$$

say something very specific about the slopes of graphs of $y = \sin x$ and $y = \cos x$. Sketch these graphs and observe that the formulas seem to be correct.

15 Let $x = L \cos^3 \theta$, $y = L \sin^3 \theta$, where L is a given positive constant. Find the equation of the tangent to the graph at the point for which $\theta = \theta_0$ and show that this tangent intersects the x and y axes at the points $A(L \cos \theta_0, 0)$ and $B(0, L \sin \theta_0)$. Show that $|\overrightarrow{AB}| = L$.

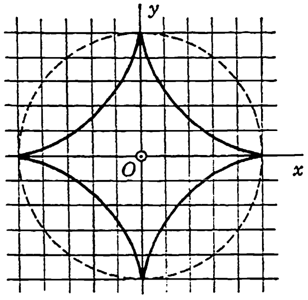


Figure 5.191

16 Let $x = L \cos^3 \theta$, $y = L \sin^3 \theta$ as in the preceding problem. Show that

$$x^{2/3} + y^{2/3} = L^{2/3}$$

and use this to find equations of tangents and to find the final result $|\overrightarrow{AB}| = L$ of the preceding problem. *Remark:* The graph of these equations is, as we shall see later, a hypocycloid of four cusps. It appears in Figure 5.191.

17 We now solve a problem that is similar to Problem 11 of Section 3.6. It is a rather tedious task to draw a graph of the equation

$$x^6 - x^2y - 2x - 7x^3 + y^6 = 721 \tag{1}$$

unless we have an electronic computer to help us do the chores. The graph does contain the point $P_0(2,3)$, the constant 721 having been so determined that this is so. Our problem is to find the equation of the tangent (if any) to the graph at P_0. Without being sure about the facts, we *assume* that there is a function ϕ, defined over some interval $2 - \delta < x < 2 + \delta$, such that the part of the graph near P_0 has the equation $y = \phi(x)$ and, moreover, ϕ is differentiable. Then (1) holds when $y = \phi(x)$ and, with the aid of our formula for differentiating products of differentiable functions of x, we differentiate the members of (1) and equate the results to obtain

$$6x^5 - x^2\frac{dy}{dx} - 2xy - 2 - 21x^2 + 6y^5\frac{dy}{dx} = 0 \tag{2}$$

or

$$\frac{dy}{dx} = -\frac{6x^5 - 2xy - 2 - 21x^2}{6y^5 - x^2}. \tag{3}$$

At the point (2,3) this has the value $\frac{98}{1454}$. The required equation of the tangent line is

$$y - 3 = \tfrac{98}{1454}(x - 2), \tag{4}$$

provided, of course, that our assumption is correct.

18 Apply the method of the preceding problem to find the slope of the graph of the equation

$$x^2 + y^2 = 25$$

at the point (3,4).

19 It is easy to show that the graph of the equation

$$(1)\quad x^4 + 2x^2y^2 + y^4 - 2x^3 - 2x^2y - 2xy^2 - 2y^3 + 5x^2 + 5y^2 - 6x - 6y + 6 = 0$$

contains the point (1,1). What have we learned that could make us sure that the graph contains another point? *Ans.:* Nothing. *Remark:* We do not yet have enough mathematical equipment to enable us to answer basic questions about natures of graphs of complicated equations. One who has or develops interest in such matters must continue study of calculus. Problem 7 and the following problems at the end of Section 11.3 provide reasons why profound study of graphs should follow (not precede) study of calculus. While the operation gives no information about the natures of graphs of other equations, one who cares to do so may show that (1) can be put in the form

$$(2)\qquad [(x-1)^2 + (y-1)^2][x^2 + y^2 + 3] = 0$$

and hence that the point (1,1) is in fact the only point on the graph.

20 Let f be the function for which $f(0) = 0$ and

$$f(x) = x^2 \sin \frac{1}{x^2}$$

when $x \neq 0$. Prove that $f'(0) = 0$ and hence that the x axis is tangent to the graph of f at the origin. Sketch the graph of f and tell why the x axis is not a line of support of the graph. *Hint:* To calculate $f'(0)$, use the fact that

$$\left|\frac{f(x) - f(0)}{x - 0}\right| = \left|x \sin \frac{1}{x^2}\right| \leqq |x| \qquad (x \neq 0)$$

and apply the sandwich theorem.

21 When we study a science, it is sometimes worthwhile to obtain preliminary ideas about machinery that we are not yet prepared to understand fully. This is an example which involves curves and tangents. Let S be a set of points in a plane (in E_2) which is *bounded* (this means that there is a rectangle which contains S), is *convex* (this means that if P_1 and P_2 lie in S, then the whole line segment joining P_1 and P_2 lies in S), and which contains at least one *inner point* (this means that there is a point P in S and a positive number δ such that S contains each point inside the circle with center at P and radius δ). Figure 5.192 shows an example. Let Γ (capital gamma) be the boundary of S; a point Q is a point of Γ if each circular disk with center at Q contains at least one point in S and also at least one point not in S. We can wonder whether Γ should be called a curve. We can observe that there may be points, such as A, B, C, D in the figure, at which Γ has many lines of support but has no tangent. We can observe that there may be points, such as E in the figure, at which Γ has only one line of support and has a tangent. We can say that Γ has a *corner* at a point B if B is on Γ and Γ has more than one line of support at B. We can wonder whether Γ

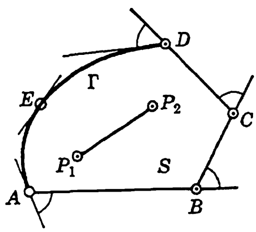


Figure 5.192

has a tangent at each point where it does not have a corner. We can wonder whether we can associate angles with corners in such a way that, whenever we take a finite set of corners, the sum of the angles at these corners must be less than or equal to 2π. We can wonder whether it is possible to construct a set S such that Γ has a corner at each of its points. We can ask questions much faster than we can answer them. We can conclude that we have some very substantial and useful information about tangents, but we do not yet know everything.

5.2 Trends, maxima, and minima Everyone who knows what it means to say "it has been getting hotter all morning" or "the temperature has been increasing all morning" should easily comprehend the following definition. In most of the applications we shall meet, the set S will be a set in E_1 (that is, a set of real numbers) which is either (i) the whole set of real numbers or (ii) the set of numbers in a *closed interval* $a \leqq x \leqq b$ or (iii) the set of numbers in an *open interval* $a < x < b$. However, S can be the set of positive integers or any other set in which we may be interested.

Definition 5.21 *A function f is said to be increasing over a set S in E_1 if $f(x_1) < f(x_2)$ whenever x_1 and x_2 are two numbers in S for which $x_1 < x_2$. The function is said to be decreasing over S if $f(x_1) > f(x_2)$ whenever x_1 and x_2 are two numbers in S for which $x_1 < x_2$.*

The following definition is more subtle. If the temperature was 30° from 10:00 A.M. to 11:00 A.M., an articulate and truthful person would not be expected to say that the temperature was increasing from 10:00 A.M. to 11:00 A.M. However, in accordance with the following definition, the temperature might have been *monotone increasing* all morning.

Definition 5.22 *A function f is said to be monotone increasing over a set S in E_1 if $f(x_1) \leqq f(x_2)$ whenever x_1 and x_2 are two numbers in S for which $x_1 < x_2$. The function is said to be monotone decreasing over S if $f(x_1) \geqq f(x_2)$ whenever x_1 and x_2 are two numbers in S for which $x_1 < x_2$.*

The terminology in this definition is very useful, and it may seem to be less than utterly foolish when we realize that f is called *monotone* (some people have preferred the word monotonous) if it is either monotone increasing or monotone decreasing. For example, the function f having the graph shown in Figure 5.23 is increasing over the interval $a \leqq x \leqq x_1$, is decreasing over the interval $x_2 \leqq x \leqq b$, is monotone increasing over the interval $a \leqq x \leqq x_2$, and is monotone decreasing over the interval $x_1 \leqq x \leqq b$. To appreciate the necessity for the following definitions, it may be sufficient to realize that it is impossible to be quite sure what is meant when someone says that "the temperature at Pike's Peak reached a maximum at noon last Friday."

Figure 5.23

Definition 5.24 *Let f be defined over a nonempty set S in E_1. We say that f has a local (or relative) maximum over S at x_0 and that $f(x_0)$ is a local maximum of f over S, if there is a positive number h such that $f(x) \leqq f(x_0)$ whenever x is in S and $|x - x_0| < h$. We say that f has a global maximum (or absolute maximum) over S at x_0 if $f(x) \leqq f(x_0)$ whenever x is in S. A local minimum and a global minimum are similarly defined, the relation $f(x) \leqq f(x_0)$ being replaced by $f(x) \geqq f(x_0)$.*

Applications of these definitions can be quite diverse. For example, $f(x)$ might be the number of telephones ringing in Chicago x hours after the beginning of the nineteenth century and S might be the whole set of positive integers or any other set of numbers we wish to select. For many purposes it suffices to see how these definitions are applied when S is an interval and f is differentiable over the interval. Let f be the function whose graph is shown in Figure 5.25. Assuming that there is

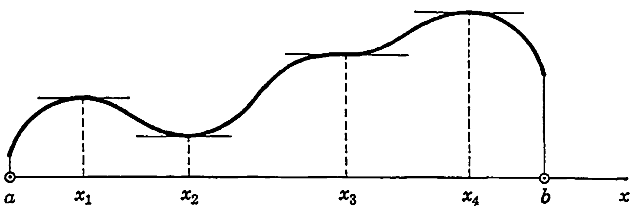
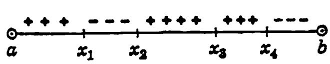


Figure 5.25

nothing deceptive about the graph, we can see that f is increasing over the intervals $a \leqq x \leqq x_1$ and $x_2 \leqq x \leqq x_4$ and that f is decreasing over the intervals $x_1 \leqq x \leqq x_2$ and $x_4 \leqq x \leqq b$. Supposing that $f'(x_3) = 0$ so that the graph has a horizontal tangent at the point $(x_3, f(x_3))$, it appears that f has local maxima at x_1 and x_4, a global maximum at x_4, local minima at a, x_2, and b, and a global minimum at a. There is neither a local minimum nor a local maximum at x_3 even though $f'(x_3) = 0$. We have described the *trends* (the increasings and the decreasings) and the *extrema* (the maxima and minima) of f.

Sometimes we are required to obtain information about a function f when we do not have a graph of f but do have a formula which determines values of $f(x)$ for different numbers x. As the discussion of Figure 5.25 indicates, it is often quite impossible to give precise information about a differentiable function f until we have found the values of x for which $f'(x) = 0$. These are the values of x for which the graph of f has horizontal tangents, and they are called *critical values* of x. After we succeed in finding when $f'(x) = 0$, when $f'(x) > 0$, and when $f'(x) < 0$, we may find it convenient to construct a figure more or less like Figure 5.251 in which we (i) mark the points at which $f'(x) = 0$ and the graph of f has horizontal tangents, (ii)

Figure 5.251



put plus signs above intervals over which $f'(x) > 0$ and the graph of f has positive slope, and (iii) put minus signs above intervals over which $f'(x) < 0$ and the graph of f has negative slope. Information about f can then be obtained with the aid of the two following theorems.

Theorem 5.26 *If $a < x_0 < b$ and if $f'(x_0)$ exists, then f cannot have a maximum or a minimum over the interval $a \leqq x \leqq b$ at x_0 unless $f'(x_0) = 0$.*

Theorem 5.27 *If f is continuous over an interval $a_0 \leqq x \leqq b_0$ and $f'(x) > 0$ when $a_0 < x < b_0$, then f is increasing over the interval*

$$a_0 \leqq x \leqq b_0.$$

If f is continuous over an interval $a_0 \leqq x \leqq b_0$ and $f'(x) < 0$ when

$$a_0 < x < b_0$$

then f is decreasing over the interval $a_0 \leqq x \leqq b_0$.

The second of these theorems is much more forthright and potent than the first. It will be proved in Section 5.5. The first theorem says that if $a < x_0 < b$ and $f'(x_0) > 0$ or $f'(x_0) < 0$, then f cannot have even a local maximum or a local minimum at x_0. To prove this, we suppose first that $a < x_0 < b$ and $f'(x_0) = p$, where p is a positive number. Then

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = p$$

and we can choose a positive number δ such that

$$a < x_0 - \delta < x_0 + \delta < b$$

and

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} > \frac{p}{2}$$

whenever $0 < |\Delta x| < \delta$. If $0 < \Delta x < \delta$, then the denominator of the above quotient is positive, so the numerator must also be positive and $f(x_0) < f(x_0 + \Delta x)$. If $-\delta < \Delta x < 0$, then the denominator of the difference quotient is negative, so the numerator must also be negative and $f(x_0 + \Delta x) < f(x_0)$. Thus if $x_0 - \delta < x_1 < x_0 < x_2 < x_0 + \delta$, then $f(x_1) < f(x_0) < f(x_2)$, so f cannot have either a local maximum or a local minimum at x_0. In case $a < x_0 < b$ and $f'(x_0) < 0$, a similar argument shows existence of a number δ such that if $x_0 - \delta < x_1 < x_0 < x_2 < x_0 + \delta$, then $f(x_1) > f(x_0) > f(x_2)$ and f cannot have a local maximum or a local minimum at x_0. This completes the proof of Theorem 5.26.

It is quite as important to know what Theorem 5.26 does not imply as it is to know what the theorem does imply. It does not imply that f has an extremum (a maximum or a minimum) any place and it does not imply that f has an extremum at x_0. It does imply that if f has an extremum at x_0, then x_0 must be either

(i) one of the end points a and b or
(ii) such that $f'(x_0)$ does not exist or
(iii) such that $f'(x_0) = 0$.

The points x_0 in these three categories are therefore the only ones that need be examined when we are seeking extrema of f over the interval $a \leqq x \leqq b$. This information, meager as it is, is often helpful. Figure 5.271 may help us to understand it. The following theorems, which are

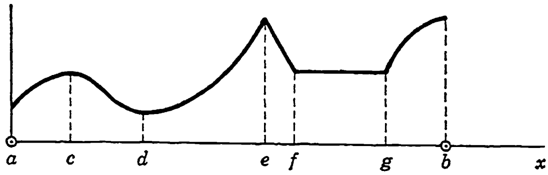


Figure 5.271

easily interpreted in terms of Figure 5.25 and which are easy consequences of Theorem 5.27, give all the information we need to solve many problems.

Theorem 5.28 (maximum) *If f is continuous over $a \leqq x \leqq b$, if $a < x_0 < b$, and if there is a positive number h such that $f'(x) > 0$ when $x_0 - h < x < x_0$ and $f'(x) < 0$ when $x_0 < x < x_0 + h$, then f has a local maximum (which may be a global maximum) at x_0.*

Theorem 5.281 (minimum) *If f is continuous over $a \leqq x \leqq b$, if $a < x_0 < b$, and if there is a positive number h such that $f'(x) < 0$ when $x_0 - h < x < x_0$ and $f'(x) > 0$ when $x_0 < x < x_0 + h$, then f has a local minimum (which may be a global minimum) at x_0.*

Theorem 5.28 says, in slightly different words, that if we travel a smooth road in such a way that we go uphill from 8:58 A.M. to 9:00 A.M. and go downhill from 9:00 A.M. to 9:02 A.M., we are atop a hill (but not necessarily atop the highest mountain) at 9:00 A.M. Theorem 5.281 has a similar interpretation.

Problems 5.29

1 Letting f be defined over E_1 by the formula $f(x) = x^2 - 2x + 3$, show that

$$f'(x) = 2x - 2 = 2(x - 1).$$

Observe that $f'(1) = 0$, and then make the more profound observation that $f'(x) < 0$ and f is decreasing when $x < 1$ and that $f'(x) > 0$ and f is increasing when $x > 1$. Show that $f(1) = 2$ and use the information to sketch a graph of $y = f(x)$. Give all of the facts involving extrema (maxima and minima) of f.

2 Letting $f(x) = ax^2 + bx + c$ where $a > 0$, show that

$$f'(x) = 2a\left(x + \frac{b}{2a}\right).$$

Tell why f is decreasing when $x < -b/2a$ and increasing when $x > -b/2a$. Show that f has a global minimum at the point

$$\left(-\frac{b}{2a}, \quad -\frac{b^2 - 4ac}{4a}\right).$$

3 Letting $f(x) = ax^2 + bx + c$ where $a < 0$, show that f is increasing when $x < -b/2a$, that f is decreasing when $x > -b/2a$, and that f has a global maximum at $-b/2a$.

4 Letting $f(x) = 1/(1 + x^2)$, show that f is increasing when $x < 0$, is decreasing when $x > 0$, and hence has a global maximum when $x = 0$.

5 Show that the function f for which

$$f(x) = \frac{x^3}{1 + x^2}$$

is everywhere increasing and hence has no extrema.

6 Show that

$$\frac{x}{x^2 - 1}$$

is decreasing except when $x = \pm 1$.

7 Find all trends and extrema of the function f for which

$$f(x) = \frac{x}{1 + x^2}$$

and sketch a graph of $y = f(x)$. *Hint:* After calculating $f'(x)$, put the result in the form

$$f'(x) = \frac{(1 + x)(1 - x)}{(1 + x^2)^2}$$

and, after observing that $f'(-1) = 0$ and $f'(1) = 0$, find the sign of $f'(x)$ over each of the intervals $x < -1$, $-1 < x < 1$, and $x > 1$. Then find $f(-1)$ and $f(1)$ and make efficient use of this information. Find $f'(0)$ and make the graph have the correct slope at the origin.

8 Supposing first that $x > 0$, find the trends and extrema of the function f for which

$$f(x) = x + \frac{1}{x}$$

and sketch the graph of $y = f(x)$. Then let $x < 0$ and repeat the process without use of symmetry, but use symmetry to check the results that are obtained.

9 This problem requires us to think about making tanks from rectangular pieces of sheet iron. Starting with a rectangle 15 units wide and 24 units long, we cut equal squares from the four corners and fold up the flaps to form a tank.

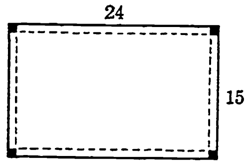


Figure 5.291

Our first step is to draw Figure 5.291 and look at it. Our good sense should tell us that if the squares are small, then the tank will be shallow and the volume will be small. Taking larger squares should yield greater volumes unless we make the squares so large that the area of the base of the tank is small enough to overcome the advantage of making the tank deeper. To become quantitative about this matter, we let x denote the lengths of the sides of the squares and ask how the volume $V(x)$ of the resulting tank depends upon x. In particular, we want to know what x maximizes $V(x)$. Show that

$$\begin{aligned} V(x) &= x(15 - 2x)(24 - 2x) \\ &= 4x^3 - 78x^2 + 360x \end{aligned}$$

and tell why x must be restricted to the interval $0 < x < \frac{15}{2}$. Show that

$$\begin{aligned} V'(x) &= 12x^2 - 156x + 360 \\ &= 12(x - 3)(x - 10). \end{aligned}$$

Tell why $V(x)$ is increasing when $0 < x < 3$ and is decreasing when $3 < x < \frac{15}{2}$. Show that the maximum V attainable is 486 cubic units.

10 A sheet-iron tank without a top is to have volume V. A rectangular sheet h feet high and $2\pi r$ feet long, costing A dollars per square foot, is bent and welded into a circular cylinder to form the lateral surface of the tank. A sheet $2r$ feet square of different material, costing B dollars per square foot, is trimmed to form a circular base which is welded to the cylinder to form the tank. Find the radius and height of the tank for which the total cost T of the material (the total amount purchased, not merely the amount actually used in the tank) is a minimum. *Ans.:*

$$r = \sqrt[3]{\frac{AV}{4B}}, \qquad h = \frac{1}{\pi}\sqrt[3]{\frac{16B^2V}{A^2}}.$$

Hint: Start by showing that

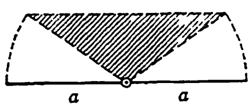
$$T = 2\pi Arh + 4Br^2$$

and then use the relation $V = \pi r^2 h$ to express T in terms of just one of the variables r and h. Standard methods may then be used to minimize T.

11 Referring to Problem 10, find the radius and height of the tank for which the cost of the material actually used is a minimum.

12 Referring to Problems 10 and 11, find the radius and height which minimize the cost of the material actually used in making a tank which has a top exactly like the base.

Figure 5.292



13 A long rectangular sheet of tin is $2a$ inches wide. Find the depth of the V-shaped trough of maximum cross-sectional area (see Figure 5.292) that can be made by bending the plate along its central longitudinal axis. *Ans.:* $a/\sqrt{2}$.

Figure 5.293

14 After referring to Problem 13 and Figure 5.293, formulate and solve a problem involving construction of troughs having rectangular cross sections.

15 Show that of all rectangles having a given area, a square has the least perimeter.

16 Find the radius and height of the cone of greatest volume that can be made from a circular disk of radius a by cutting out (or folding over, as chemists do) a sector and bringing the edges of the remaining part together.

17 An ordinary tomato can is to be constructed to have a given volume V. Determine the height h and radius r of the can for which total surface area is a minimum.

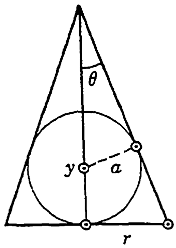


Figure 5.294

18 As in Figure 5.294, the base and lateral surface of a solid right circular cone are tangent to a sphere of radius a. Find the height of the solid having minimum volume.

Outline of solution: The height y and base radius r are related by the formula

$$\frac{a}{y-a} = \frac{r}{\sqrt{r^2+y^2}}$$

which equates two expressions for $\sin\theta$. Squaring, solving for r^2, and using the formula $V = \frac{1}{3}\pi r^2 y$ for the volume of the solid, we find that

$$V = \frac{\pi a^2}{3}\frac{y^2}{y-2a}, \qquad \frac{dV}{dy} = \frac{\pi a^2 y}{3}\frac{y-4a}{(y-2a)^2}.$$

The conditions of the problem require that $y > 2a$. If $2a < y < 4a$, then $dV/dy < 0$ and V is decreasing. If $y > 4a$, then $dV/dy > 0$ and V is increasing. Thus V is minimum when $y = 4a$.

19 Supposing that $x_1, x_2, \cdots, x_n$ are given numbers, find the values of x, if any, for which

$$\sum_{k=1}^{n} (x - x_k)^2$$

attains maximum and minimum values.

20 The elementary theory of probability tells us that the number $p_{n,k}$ defined by

$$p_{n,k} = \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k}$$

is the probability of exactly k successes in n trials when p is the probability of success in each trial. Supposing that n and k are given integers for which $n > 0$ and $0 \leqq k \leqq n$, find the number p which maximizes $p_{n,k}$. *Hint:* Ignore the numerical coefficient and find the p which maximizes $p^k(1-p)^{n-k}$. *Ans.:* $p = k/n$.

21 An observant senator observes that if he hires just one secretary, she will work nearly 30 hours per week but that each additional secretary produces conversations that reduce her effectiveness. In fact, if there are x secretaries, x not exceeding 30, then each one will work only

$$30 - \frac{x^2}{30}$$

hours per week. Find the number of secretaries that will turn out the most work. *Discussion and solution:* If there are x secretaries, the number $y(x)$ of hours of work done per week is

$$(1) \qquad y(x) = x\left(30 - \frac{x^2}{30}\right) = 30x - \frac{x^3}{30}.$$

It is required that x be an integer in the interval $0 \leqq x \leqq 30$, so there are only 31 possibilities. We can calculate the 31 numbers $y(0)$, $y(1)$, $\cdots$, $y(30)$ and select the x which gives the greatest $y(x)$. It is easier and more enlightening, however, to use some calculus. Forgetting momentarily that x is an integer number of secretaries, we observe that (1) defines $y(x)$ for each real x. Differentiating gives

$$(2) \qquad y'(x) = 30 - \frac{x^2}{10} = \frac{1}{10}(300 - x^2).$$

Thus $y'(x) > 0$ and y is increasing when $0 \leqq x < \sqrt{300}$. Moreover, $y'(x) < 0$ and y is decreasing when $x > \sqrt{300}$. Since $\sqrt{300} = 17.32$, we see that $y(x) < y(17)$ when $0 \leqq x < 17$ and that $y(x) < y(18)$ when $18 < x \leqq 30$. Thus the answer is 17 if $y(17) > y(18)$ and is 18 if $y(18) > y(17)$.

22 As in Figure 5.295, a triangle is inscribed in a semicircular region having diameter a. Find the θ which maximizes the area of the triangle. *Ans.:* $\theta = \pi/4$.

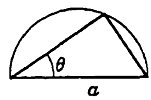


Figure 5.295

23 A printed page is required to contain A square units of printed matter. Side margins of widths a and top and bottom margins of widths b are required. Find the lengths of the printed lines when the page is designed to use the least paper. *Ans.:* $\sqrt{aA/b}$.

24 Sketch a reasonably good graph of $y = x^2$ and then mark the point or points on this graph that seem to be closest to the point $(0,\frac{3}{2})$. Then calculate the coordinates of the closest point or points. *Hint:* Minimize the square of the distance from the point $(0,\frac{3}{2})$ to the point (x,x^2).

25 The strength (ability to resist bending) of a rectangular beam is proportional to the width x and to the square of the height y of a cross section. Find the width and height of the strongest beam that can be sawed from a cylindrical log whose cross sections are circular disks of diameter L. *Ans.:* Width $= L/\sqrt{3}$, height $= \sqrt{2}\,L/\sqrt{3}$.

26 The x axis of Figure 5.296 is the southern shore of a lake containing a little island at the point (a,b), where $a > 0$. A man who is at the origin can run r feet per second along the x axis and can swim s feet per second in the water. He wants to reach the island as quickly as possible. Should he do some running before he starts to swim and, if so, how far? *Partial ans.:* He should run

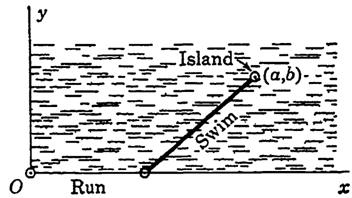


Figure 5.296

$$a - \frac{s}{\sqrt{r^2 - s^2}}\,b$$

feet if $r > s$ and this number is positive. Investigation of the whole matter is not as simple as might be supposed. *Hint:* If $x > 0$ and the man runs from the origin to the point $(x,0)$, we should be able to calculate (in terms of x and the given constants) the distance he runs, the distance he swims, and the total time T required to reach the island.

27 Light travels with speed s_1 in air and with speed s_2 in water. Figure 5.297 can interest us in possible paths by which light might journey from a point A in the air to a point S on the surface of the water and then to a point W in the water. Show that the total time T is a minimum when the point S is so situated that the angle θ_1 of incidence and the angle θ_2 of refraction satisfy the condition

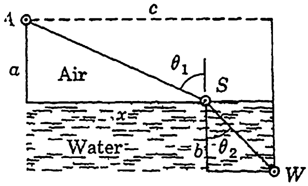


Figure 5.297

$$\frac{\sin\theta_1}{s_1} = \frac{\sin\theta_2}{s_2}.$$

Remark: The above formula is the Snell formula, one of the fundamental formulas of optics. Phenomena such as the one revealed by this problem are of great interest in physics and philosophy.

28 As in Figure 5.298 a heavy object of weight W is to be held by two identical cables. A kind engineer tells us that the tension T in the cables is $W\sqrt{a^2 + x^2}/2x$. A solemn merchant tells us that the cost per foot of his cables is kT dollars, where T is the tension they will safely withstand. We must buy the cables, and we have a problem. *Ans.:* We buy $2\sqrt{2}\,a$ feet of cable costing $Wk/\sqrt{2}$ dollars per foot, so we need $2Wka$ dollars.

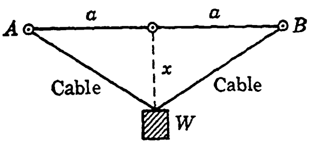


Figure 5.298

29 Modify the preceding problem by supposing that the body must hang below the point which lies between A and B at unequal distances a and b from A and B.

30 The lower free corner of a page of a book is folded up and over until it meets the inner edge of the page and then the folded part is pressed flat to leave a triangular flap and a crease of length L. Supposing that the page has width a, find the distance from the inner edge of the page to the bottom of the crease when L is a minimum and find the minimum L. *Hint:* Minimize L^2. *Ans.:* $a/4$ and $3\sqrt{3}\,a/4$.

31 Sketch the part G_1 of the graph of the equation

$$y = x + \frac{1}{x}$$

that lies in the first quadrant and observe that the y axis and the line having the equation $y = x$ are asymptotes of G_1. Someday we will learn that G_1 is a branch of a hyperbola and that the point V on G_1 closest to the origin is a vertex of the hyperbola. Find the coordinates of V and the distance from the origin to V. *Ans.:* $(2^{-1/4}, (1 + \sqrt{2})2^{-1/4})$ and $\sqrt{2 + 2\sqrt{2}}$.

32 A given circle has radius a. A second circle has its center on the given one, and the arc of the second circle which lies inside the given circle has length L.

Prove that L is maximum when an appropriate angle θ satisfies the equation $\cot \theta = \theta$.

33 A spherical ball of radius r settles slowly into a full conical glass of water and causes an overflow of water. The glass has height a, and the lines on the surface of the glass make the angle θ with the axis of the glass. Find the radius r for which the overflow is a maximum. *Remark:* This problem is famous because it is difficult enough to be remembered and discussed by those who have solved it. *Solution:* With or without careful scrutiny of other cases, we suppose that the ball is neither so small that it can be completely submerged nor so large that it will fail to be tangent to the glass when it is in its lowest position. Letting O, C, and B be the vertex of the conical glass, the center of the ball, and the bottom of the ball, we see that $|\overrightarrow{OC}| = r \csc \theta$ and $|\overrightarrow{OB}| = r(\csc \theta - 1)$. The submerged segment of the ball has thickness h, where

$$h = a - r(\csc \theta - 1). \tag{1}$$

The overflow (measured in cubic units) is equal to the volume V of this segment, and Problem 2 of Problems 4.59 shows that

$$V = \tfrac{1}{3}\pi h^2(3r - h) = \pi[h^2 r - \tfrac{1}{3}h^3]. \tag{2}$$

Differentiation gives

$$\begin{aligned} \frac{dV}{dr} &= \pi\left[h^2 + 2hr\frac{dh}{dr} - h^2\frac{dh}{dr}\right] \\ &= \pi h\left[2r\frac{dh}{dr} + h\left(1 - \frac{dh}{dr}\right)\right]. \end{aligned} \tag{3}$$

Using (1) and the formula for dh/dr calculated from it gives

$$\frac{dV}{dr} = \frac{\pi h}{\sin^2 \theta}[a \sin \theta - (\sin \theta + \cos 2\theta)r]. \tag{4}$$

Since $a \sin \theta$ and $(\sin \theta + \cos 2\theta)$ are positive when $0 < \theta < \pi/2$, it follows that V is a maximum when

$$r = \frac{a \sin \theta}{\sin \theta + \cos 2\theta}. \tag{5}$$

34 When distances are measured in feet, the equation of the path followed by water projected from our fire hose is

$$y = mx - \frac{(1 + m^2)x^2}{100},$$

where m is the slope of the path at the nozzle which is located at the origin. Find the value of m for which the water will reach the greatest height on a wall 40 feet from the nozzle and find the greatest height. *Partial ans.:* One of the two answers is 9.

35 *Remark:* The following big-government problem need not be taken too seriously; its purpose is to neutralize a problem involving a country that allowed its unemployed boomerang repairmen to starve to death. Determine the number of officials that must be supported in a country containing n workers, and use the result to determine the population of a Utopian country that minimizes the burdens which individual workers must bear. *Hint:* Information about officers in efficient productive organizations is not relevant, but Parkinson laws may be used.

36 Find the minimum of the function F for which

$$F(\lambda) = \int_0^1 [x^2 - (x + \lambda)]^2\, dx. \qquad \textit{Ans.: } \tfrac{1}{180}$$

5.3 *Second derivatives, convexity, and flexpoints* In Section 3.6 we called attention to the connection between second derivatives and accelerations. This section shows how second derivatives can be used to obtain information about functions and their graphs. To begin the proceedings, we look at Figure 5.31, which shows the graph of a function for

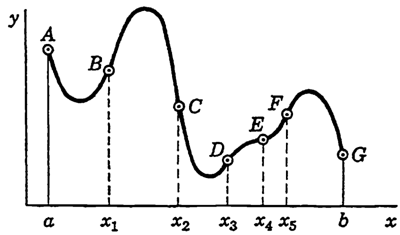


Figure 5.31

which the derivative (or first derivative) $f'(x)$ and the second derivative (the derivative of the derivative) $f''(x)$ exist when $a \leqq x \leqq b$. To get some ideas, we think of the graph as being a road in a vertical plane upon which we can travel from A to G, and we take the x axis to be at sea level. During the whole trip, we are always above sea level. The sign of $f(x)$ gives us this information. At some times during the trip we are going uphill, and at other times we are going downhill. The sign of $f'(x)$ gives us this information. As we travel from A to B, from C to D, and from E to F we are passing over depressions (or pits), and as we travel from B to C, from D to E, and from F to G we are passing over humps (or peaks). As we shall see, $f''(x)$ is our source of information about these things and about *points of inflection* or *flexpoints* B, C, D, E, F at which slopes attain local extrema.

The two following theorems are obtained by replacing f by f' in Theorems 5.26 and 5.27.

Theorem 5.32 *If f' is differentiable over $a < x < b$ and $a < x_0 < b$, then f cannot have a flexpoint at x_0 unless $f''(x_0) = 0$.*

Theorem 5.33 *If f' is continuous over an interval $a_0 \leqq x \leqq b_0$ and $f''(x) > 0$ when $a_0 < x < b_0$, then f' is increasing over the interval $a_0 \leqq x \leqq b_0$. If f' is continuous over an interval $a_0 \leqq x \leqq b_0$ and $f''(x) < 0$ when $a_0 < x < b_0$, then f' is decreasing over the interval $a_0 \leqq x \leqq b_0$.*

The information contained in these theorems is sometimes very helpful when graphs of given functions f are being drawn. For example, Figure 5.34 shows an application of the first part of this theorem; $f''(x)$ is positive

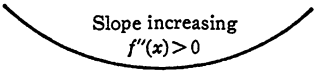


Figure 5.34

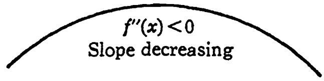


Figure 5.35

over an interval and $f'(x)$, the slope, increases from -1 through 0 to $+1$ as x increases over the interval. Figure 5.35 shows an application of the second part of the theorem; $f''(x)$ is negative over an interval and $f'(x)$, the slope, decreases from 1 through 0 to -1 as x increases over the interval. Sometimes it is helpful to put ideas involving derivatives in the form

$$f''(x) = \frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{dm}{dx} = \frac{d\text{ slope}}{dx}. \tag{5.351}$$

The important thing to remember is that $f''(x)$ is the derivative of $f'(x)$ and that *a positive second derivative implies an increasing first derivative and hence an increasing slope*, and that *a negative second derivative implies a decreasing first derivative and hence a decreasing slope.* It is sometimes useful and even necessary to know about attempts to describe the differences between the arcs of Figures 5.34 and 5.35 in other words. The first runs through a depression and the second runs over a hump. The first bends upward and the second bends downward. The first is convex upward and the second is convex† downward. In the first case, the chord joining two points on the graph lies above the arc joining the two points, and in the second case the chord lies below the arc. In the first case each tangent to the graph lies (at least locally) below the arc, and in the second case each tangent lies (at least locally) above the arc.

The virtue of the following theorem lies in the fact that it is a "local theorem" which we can apply without determining signs of functions over whole intervals and which is therefore sometimes easier to apply than Theorem 5.28.

Theorem 5.36 *If* $f'(x_0) = 0$ *and* $f''(x_0) > 0$, *then* (as Figure 5.34 indicates) f *has a local minimum at* x_0. *If* $f'(x_0) = 0$ *and* $f''(x_0) < 0$, *then* (as Figure 5.35 indicates) f *has a local maximum at* x_0.

† In mathematics and optics a point set (which might in some cases be a lens) is *convex* if it contains the line segment joining P_1 and P_2 whenever it contains P_1 and P_2. The set is sometimes said to be *concave* if it is not convex. When we say that a part of a graph in the xy plane is convex upward, we mean that the set lying above it is convex; we do not mean that the graph is a convex set.

To prove the first part of this theorem, let $f'(x_0) = 0$ and let $f''(x_0) = p$, where p is a positive number. Then

$$\lim_{\Delta x \to 0} \frac{f'(x_0 + \Delta x) - f'(x_0)}{\Delta x} = p.$$

Let $\epsilon = p/2$. Then there is a positive number δ such that $f'(x_0 + \Delta x)$ exists when $|\Delta x| < \delta$ and

$$\frac{f'(x_0 + \Delta x) - f'(x_0)}{\Delta x} > \frac{p}{2}$$

whenever $0 < |\Delta x| < \delta$. But $f'(x_0) = 0$, and hence

$$\frac{f'(x_0 + \Delta x)}{\Delta x} > 0$$

and therefore $f'(x_0 + \Delta x)$ and Δx are both positive or both negative whenever $0 < |\Delta x| < \delta$. When $x_0 < x < x_0 + \delta$, we can set $x = x_0 + \Delta x$ and conclude that $0 < \Delta x < \delta$ and $f'(x) > 0$. When $x_0 - \delta < x < x_0$, we can set $x = x_0 + \Delta x$ and conclude that $-\delta < \Delta x < 0$ and $f'(x) < 0$. It therefore follows from the last part of Theorem 5.28 that f has a local minimum at x_0. In case $f''(x_0) < 0$, everything is the same except that some signs are reversed and f has a local maximum at x_0.

Problems 5.39

1 Sketch a graph of $y = 1/x$. Calculate dy/dx and d^2y/dx^2. If appropriate connections between these things are not immediately clear, there are only three possibilities: (i) the graph needs repairs or (ii) the formulas for derivatives need repairs or (iii) the text of this section must be studied more carefully.

2 The values of

$$f(x) = \frac{x}{1 + x^2}$$

are certainly near 0 when x is near 0 and when $|x|$ is large. Give a full account of the nature of the graph.

3 Supposing that a, b, and c are constants for which $a > 0$ and that

$$f(x) = ax^2 + bx + c,$$

calculate $f'(x)$ and $f''(x)$. Show that the only extremum of f is a minimum which is attained when $x = -b/2a$. Show that the graph of f is everywhere bending upward and that there are no flexpoints.

4 Supposing that a, b, c, d are constants for which $a > 0$ and that

$$f(x) = ax^3 + bx^2 + cx + d,$$

calculate $f'(x)$ and $f''(x)$. Show that the graph of f has exactly one flexpoint for which $x = -b/3a$ and that f' is increasing when $x > -b/3a$.

5 Show that if the x, y coordinate system is chosen in such a way that the graph of

$$y = x^3 + bx^2 + cx + d$$

passes through the origin and has its flexpoint at the origin, then $b = d = 0$ and

$$y = x^3 + cx.$$

Show that the graph of the latter equation has no extrema if $c \geqq 0$ and has two local extrema if $c < 0$.

6 Starting with the first of the relations

$$(1) \qquad y = (a^2 - x^2)^{1/2}, \qquad \frac{d^2y}{dx^2} = -\frac{a^2}{y^3},$$

differentiate twice and obtain the second relation. Then start with the first of the relations

$$(2) \qquad x^2 + y^2 = a^2, \qquad x + y\frac{dy}{dx} = 0$$

$$(3) \qquad 1 + y\frac{d^2y}{dx^2} + \left(\frac{dy}{dx}\right)^2 = 0$$

and show that differentiating with respect to x gives the others. Use (2) and (3) to obtain the second relation in (1). Tell why you should expect the sign of the second derivative to be opposite to the sign of y.

7 Supposing that a and p are given positive numbers and considering positive values of x and y, use the two methods of the preceding problem to find d^2y/dx^2 when

$$x^p + y^p = a^p.$$

Make the results agree with each other and, for the case $p = 2$, with a result of the preceding problem. Tell why the sign of the second derivative should (or should not) depend upon p as it does in your answer.

8 Supposing that $a > 0$ and $b > 0$, show that the graph of

$$f(x) = a \sin (bx + c)$$

has a flexpoint wherever it intersects the x axis.

9 Sketch a reasonably accurate graph of the function f for which

$$f(x) = x \sin x$$

and observe that the graph seems to have flexpoints on or near the x axis. Show that if (x,y) is a flexpoint, then $\tan x = 2/x$ and $y = 2 \cos x$. *Remark:* These results show that if (x,y) is a flexpoint for which $|x|$ is large, then $\tan x$ is near 0, $\sin x$ is near 0, $\cos x$ is near 1 or -1, and y is near 2 or -2.

10 Supposing that n is 10 or 20, sketch the graph of $y = \sin^n x$ over the interval $0 \leqq x \leqq \pi/2$ and mark a point which seems to be a flexpoint. Then, supposing that $n > 1$, show that the graph has a flexpoint at the point (x_n, y_n) for which

$$\cos x_n = \frac{1}{\sqrt{n}} \quad \text{or} \quad \sin x_n = \left(1 - \frac{1}{n}\right)^{1/2} \quad \text{and} \quad y_n = \left(1 - \frac{1}{n}\right)^{n/2}.$$

Remark: Unless we know about the famous number e, it is still not easy to estimate y_n when n is large. When we have learned the first of the formulas

$$\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = e^x, \qquad \lim_{n\to\infty} y_n = e^{-1/2} = 0.606531,$$

we will be able to put $x = -1$ and take square roots to obtain the second one.

11 This problem, like the preceding one, involves ideas. Supposing that $n > 1$ and

$$(1) \qquad y = x^n(2 - x)^n = (2x - x^2)^n,$$

show that

$$(2) \qquad \frac{d^2y}{dx^2} = 2n(2n - 1)(2x - x^2)^{n-2}\left[x^2 - 2x + \frac{2n - 2}{2n - 1}\right]$$

and hence that the graph of (1) has a flexpoint at the point (x_n, y_n) where

$$x_n = 1 - \sqrt{\frac{1}{2n - 1}}, \qquad y_n = \left[1 - \frac{1}{2n - 1}\right]^n.$$

12 Determination of the natures of the graphs of equations like

$$(1) \qquad y^4 = x^2(1 + x^2)$$

is an ancient and honorable pastime. Observe that if the point (x,y) lies on the graph, then so also do the points $(x,-y)$, $(-x,y)$, and $(-x,-y)$. If we find the part G of the graph in the first quadrant, we can therefore use symmetry to obtain the rest of the graph. Henceforth we consider only points on the graph for which $x \geqq 0$ and $y \geqq 0$. For these points,

$$(2) \qquad y = (x^4 + x^2)^{1/4}.$$

To each x there corresponds exactly one y for which the point (x,y) is on G Moreover, $y \geqq (x^4)^{1/4} = x$, so G lies on or above the line having the equation $y = x$. Show that, when $x > 0$,

$$(3) \qquad \frac{dy}{dx} = \frac{2x^3 + x}{2(x^4 + x^2)^{3/4}}, \qquad \frac{d^2y}{dx^2} = \frac{x^2(2x^2 - 1)}{4(x^4 + x^2)^{7/4}}.$$

Show that the slope is decreasing over the interval $0 < x < 1/\sqrt{2}$, increasing over the interval $x > 1/\sqrt{2}$, and attains the minimum value $\sqrt[4]{625/1728}$ at the flexpoint having the coordinates $1/\sqrt{2}$ and $\sqrt[4]{3/4}$. Show that, when $x > 0$,

$$(4) \quad 0 < \sqrt[4]{x^4 + x^2} - x = \frac{x^2}{(\sqrt[4]{x^4 + x^2} + x)(\sqrt{x^4 + x^2} + x^2)} < \frac{x^2}{(2x)(2x^2)} = \frac{1}{4x}$$

and hence that the line having the equation $y = x$ is an asymptote of G. The graph of (1) is shown on a small scale in Figure 5.391.

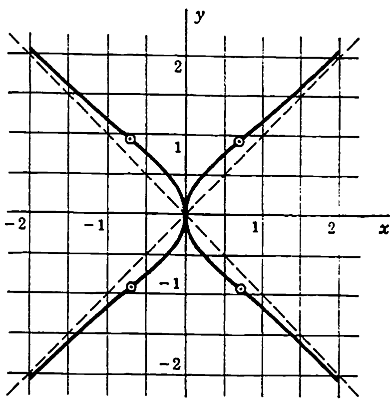


Figure 5.391

13 Determine the natures of the graphs of the equations

(a) $y^4 = x(1 + x^2)$ (b) $y^4 = x(x^2 - 1)$

(c) $y^4 = x^2(x^2 - 1)$ (d) $y^4 = x(1 - x)$

(e) $y^2 = \dfrac{x}{1 + x^2}$ (f) $y^2 = \dfrac{x^2}{1 + x^2}$

(g) $y = \dfrac{1}{(x + 3)(x - 4)}$ (h) $y^2 = \dfrac{1}{(x + 3)(x - 4)}$

14 Sketch graphs of $y = x$, $y = \sin x$, and then $y = x + \sin x$. Then make repairs in the last graph that may be necessary to make it agree with formulas for the slope and the derivative of the slope.

15 Persons interested in themselves and the surrounding world should not neglect opportunities to learn about the honorable *Gauss* (or *normal*) probability density function Φ defined by the formula

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-M)^2}{2\sigma^2}} \tag{1}$$

in which σ (sigma) and M are constants for which $\sigma > 0$. We should know that $e^0 = 1$, and we can cheerfully accept the facts that

$$e = 2.71828 \cdots, \qquad e^{-\frac{1}{2}} = 0.60653 \cdots. \tag{2}$$

We want to determine the manner in which the graph of $y = \Phi(x)$ depends upon the constants M (which is called the *mean* of Φ) and σ (which is called the standard deviation of Φ). Show that

$$\Phi'(x) = \frac{-1}{\sqrt{2\pi}\,\sigma^3}(x - M)e^{-\frac{(x-M)^2}{2\sigma^2}}. \tag{3}$$

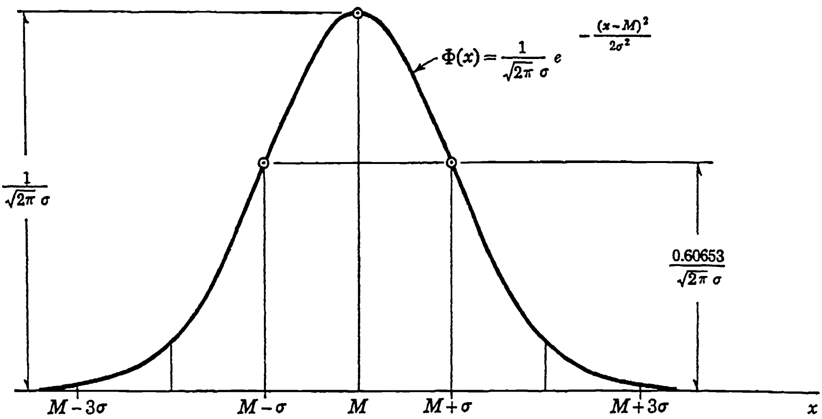


Figure 5.392

Tell why it is true that, as Figure 5.392 indicates, Φ is increasing over the interval $x \leqq M$, Φ is decreasing over the interval $x \geqq M$, Φ has a maximum at M, and $\Phi(M) = 1/\sqrt{2\pi}\,\sigma$. Show that

(4) $$\Phi''(x) = \frac{1}{\sqrt{2\pi}\,\sigma^5}[(x - M)^2 - \sigma^2]e^{-\frac{(x-M)^2}{2\sigma^2}}.$$

Tell why it is true that, as the figure indicates, $\Phi'(x)$ is increasing and the graph is bending upward when $x < M - \sigma$ and when $x > M + \sigma$, whereas $\Phi'(x)$ is decreasing and the graph is bending downward when $M - \sigma < x < M + \sigma$. Finally, show that, as the figure indicates,

(5) $$\Phi(M - \sigma) = \Phi(M + \sigma) = \frac{0.60653}{\sqrt{2\pi}\,\sigma} = 0.60653\Phi(M).$$

Remark: The index will tell where this and other bits of information about Gauss probability functions are concealed. We shall learn that

(6) $$\int_{-\infty}^{\infty} \Phi(x)\,dx = 1,$$

and budding scientists are never too young to start hearing that, in appropriate circumstances, the number

(7) $$\int_a^b \Phi(x)\,dx$$

is taken to be the probability that a number x lies between a and b.

16 Sketch rough graphs of $y = \cos x$, $y = 2\cos x$, $y = \cos 2x$, and then

(1) $$f(x) = 2\cos x - \cos 2x$$

over the interval $0 \leqq x \leqq \pi$. Find the maxima and minima of f and the flexpoints of its graph. Make the results agree.

17 Make a sketch showing the points (x,y) on the graph of the equations

$$x = a(\phi - \sin\phi), \qquad y = a(1 - \cos\phi)$$

for which $\phi = 0, \pi/2, \pi, 3\pi/2$, and 2π. Show that the graph is convex downward at each point for which $y \neq 0$. *Remark:* The graph is a cycloid. Answers can be simplified by use of trigonometric identities. Thus

$$\frac{dx}{d\phi} = a(1 - \cos\phi) = 2a\sin^2\frac{\phi}{2}$$

$$\frac{dy}{d\phi} = a\sin\phi \qquad = 2a\sin\frac{\phi}{2}\cos\frac{\phi}{2}$$

$$\frac{dy}{dx} = \frac{\sin\phi}{1 - \cos\phi} \qquad = \frac{\cos\dfrac{\phi}{2}}{\sin\dfrac{\phi}{2}} = \cot\frac{\phi}{2},$$

so $dy/dx > 0$ when $0 < \phi < \pi$ and $dy/dx < 0$ when $\pi < \phi < 2\pi$. Moreover, since

$$\frac{d}{d\phi}\cot\frac{\phi}{2} = \frac{d}{dt}\frac{\cos\dfrac{\phi}{2}}{\sin\dfrac{\phi}{2}} = -\frac{1}{2\sin^2\dfrac{\phi}{2}}$$

when $\sin\dfrac{\phi}{2} \neq 0$, we find that

$$\frac{d^2y}{dx^2} = \frac{\dfrac{d}{d\phi}\dfrac{dy}{dx}}{\dfrac{dx}{d\phi}} = \frac{\dfrac{d}{d\phi}\cot\dfrac{\phi}{2}}{2a\sin^2\dfrac{\phi}{2}} = -\frac{1}{4a\sin^4\dfrac{\phi}{2}}$$

when $\sin\dfrac{\phi}{2} \neq 0$ and hence when $y \neq 0$. The slope is therefore decreasing when $y \neq 0$.

18 Verify that the hypotheses and conclusion of the following theorem are satisfied when $f(x) = \sin x$, $g(x) = (\sin x)/x$, and $a = \pi$.

Theorem *If $a > 0$, if f has two derivatives over the interval $0 \leqq x \leqq a$, if $f(0) \geqq 0$, if $f''(x) < 0$ when $0 < x < a$, and if $g(x) = f(x)/x$, then g is decreasing over the interval $0 < x < a$.* *Remark:* This theorem has a very interesting geometric interpretation. The hypotheses imply that, as in Figure 5.393, $f(0) \geqq 0$ and the graph of f is convex downward. The graph can make us feel that, as x increases, the angle θ must decrease and hence $f(x)/x$ must decrease because $f(x)/x$ is tan θ. It is, however, necessary to recognize that feelings and impressions obtained by looking at one or a dozen figures do not constitute a proof of the theorem. To prove the theorem, we begin by observing that, when $0 < x < a$,

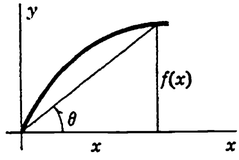


Figure 5.393

$$g'(x) = \frac{xf'(x) - f(x)}{x^2} = \frac{h(x)}{x^2},$$

where $h(x) = xf'(x) - f(x)$. We will know that g is decreasing over the interval $0 < x < a$ if we can show that $h(x) < 0$ when $0 < x < a$. Since

$$h'(x) = xf''(x) + f'(x) - f'(x) = xf''(x),$$

our hypothesis that $f''(x) < 0$ when $0 < x < a$ implies that $h'(x) < 0$ when $0 < x < a$. Thus $h(x)$ is decreasing over the interval $0 < x < a$. Since h is continuous and $h(0) = -f(0) \leqq 0$, it follows that $h(x) < 0$ when $0 < x < a$. This gives the conclusion of the theorem.

19 Ideas of this section and the preceding one can be used to obtain information about the graphs of the Bernoulli functions $B_0(x)$, $B_1(x)$, $B_2(x)$, $\cdots$ that appeared in Section 4.3, Problem 10. We recall that $B_0(x) = 1$, that

$$B_n'(x) = B_{n-1}(x) \tag{1}$$

$$\int_0^1 B_n(x)\,dx = 0 \tag{2}$$

when $n = 1, 2, 3, \cdots$ except that (1) fails to hold when n is 1 or 2 and x is an integer, and that all of the functions except $B_1(x)$ are continuous. To keep our task within reasonable bounds, we suppose we know the fundamental fact that $B_n(0) = 0$, $B_n(\frac{1}{2}) = 0$, and $B_n(1) = 0$ when n is odd, that is, when $n = 1, 3, 5, 7, \cdots$. We want to show, without making tedious calculations, that the miniature graphs of $B_1(x)$, $B_2(x)$, $\cdots$, $B_8(x)$ over the interval $0 \leqq x \leqq 1$ appearing in Figure 5.394 give correct information about the trends and the

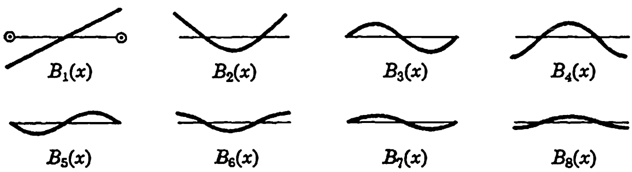


Figure 5.394

zeros of these functions. When (12.384) and related formulas have been studied, we will be able to see that scales on the vertical axes have been adjusted to make the graphs visible; it can be shown that $|B_n(x)| < 4/(2\pi)^n$ when $n > 1$ and hence that numerical values cannot be estimated from the graphs in Figure 5.394. Supposing that $0 < x < 1$, show that the formulas $B_1'(x) = 1$ and $\int_0^1 B_1(x)\,dx = 0$ imply that $B_1(x) = x - \frac{1}{2}$ and hence that the graph of $B_1(x)$ is correct. Show that the formula $B_2'(x) = B_1(x)$ implies that $B_2(x)$ is decreasing over the interval $0 < x < \frac{1}{2}$ and is increasing over the interval $\frac{1}{2} < x < 1$. Show that this fact and the formula $\int_0^1 B_2(x)\,dx = 0$ imply that $B_2(0) > 0$ and $B_2(\frac{1}{2}) < 0$ and hence that $B_2(x)$ has exactly two zeros between 0 and 1. Show that the formula $B_3''(x) = B_1(x)$ implies that the graph of $B_3(x)$ is bending downward over the interval $0 < x < \frac{1}{2}$ and is bending upward over the interval $\frac{1}{2} < x < 1$. Show that the formula $B_3'(x) = B_2(x)$ implies that $B_3(x)$ is increasing over the first part of the interval $0 < x < 1$, is decreasing over a central part, and is increasing over the remaining part. Tell why $B_4(x)$ is increasing over the interval $0 < x <$

$\frac{1}{2}$ and is decreasing over the interval $\frac{1}{2} < x < 1$, and why $B_4(x)$ has exactly two zeros between 0 and 1. Continue this investigation until general conclusions about the functions $B_n(x)$ and their graphs have been reached.

5.4 Theorems about continuous and differentiable functions

It is possible to look at Figure 5.42 and others more or less like it and claim that these figures provide experimental evidence supporting the following theorem of which we shall give a stronger version in Theorem 5.52.

Theorem 5.41 *If L is a chord joining two points on the graph of a differentiable function, then there must be at least one point on the graph at which the tangent is parallel to L.*

There are at least two reasons why this theorem is surprising. It is thoroughly important, and it is impossible to prove it without making use of some substantial mathematical machinery that has not yet appeared in this course. The source of the difficulty can be stated very crudely by saying that Theorem 5.41 would be false if there were "holes" in the set of real numbers so that the graph of Figure 5.42 contains no points having x coordinates x_1 and x_2. To prove Theorem 5.41, and for many other purposes, we need a property or postulate or axiom which guarantees that the set of real numbers is *complete*. While several different equivalent axioms can be given, the following one involving a fundamental idea of Dedekind (1831–1916) is in some respects the most natural one to adopt.

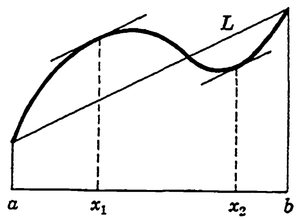


Figure 5.42

Axiom 5.43 **(Dedekind)** *Let the set of real numbers be partitioned into two subsets A and B* (see Figure 5.44) *in such a way that* (i) *each real number is put into either A or B,* (ii) *each of A and B contains at least one real number, and* (iii) *if x_1 is in A and x_2 is in B, then $x_1 < x_2$. Then there is a real number* ξ (the partition number *xi*) *which is either the greatest number in A or the least number in B.*

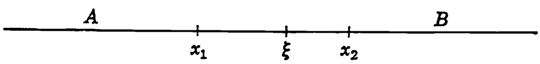


Figure 5.44

Once again we are in a position where we should know something about our present state and prospects for future development. To attain full comprehension of the Dedekind axiom, and the manner in which it is used to prove basic theorems of mathematical analysis, is not a short task. Experience shows that, except in unusual special circumstances, it is quite unreasonable to suppose that enough time is available in a first course in calculus to explore these matters thoroughly. In order to understand the proofs, it is necessary to sketch and study illustrations of various kinds, and progress is very slow. Until considerable experience has been obtained, it is not easy to reproduce the proofs even after they have been completely understood. Students who peek at the proofs can be compared with children of jewelers who peek at the innards of watches. They start accumulation of knowledge of reasons why things tick, and the overwhelming importance of getting started must be recognized by everyone who knows that we toddle and walk before we run. So far as this course is concerned, it is of primary importance to understand the axiom and theorems of this section and to cultivate the habit of formulating and using precise mathematical statements.

We begin a campaign to learn something about continuous functions and differentiable functions and their graphs by proving the following theorem.

Theorem 5.45 *If f is continuous over an interval $a \leqq x \leqq b$, then f is bounded over the interval, that is, there is a constant M such that*

$$|f(x)| \leqq M \qquad (a \leqq x \leqq b).$$

Our proof will use the Dedekind axiom. Let x_1 be put in A if $x_1 \leqq a$. Moreover, let x_1 be put in A if $a < x_1 \leqq b$ and there is a constant M_1 such that $|f(x)| \leqq M_1$ when $a \leqq x \leqq x_1$. Let B contain all other numbers. This determines a Dedekind partition, and we can let ξ be the partition number. It is easy to see that $a \leqq \xi \leqq b$, but the remainder of the proof is more delicate. Since f has right-hand continuity at a, we can let $\epsilon = 1$ and choose a positive number δ such that $f(a) - 1 \leqq f(x) \leqq f(a) + 1$ and hence $|f(x)| \leqq |f(a)| + 1$ whenever $a \leqq x \leqq a + \delta$. Hence f is bounded over the interval $a \leqq x \leqq a + \delta$, so $a + \delta$ belongs to A and $\xi \geqq a + \delta > a$. Our next step is to show that $\xi = b$. If $\xi < b$, then we have $a < \xi < b$ as in Figure 5.451. Since f is continuous at ξ,

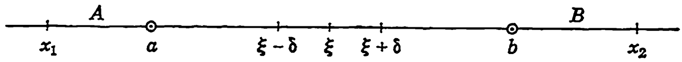


Figure 5.451

we can let $\epsilon = 1$ and choose a positive number δ such that $a < \xi - \delta < \xi + \delta < b$ as in Figure 5.451 and $|f(x)| < |f(\xi)| + 1$ when $\xi - \delta \leqq x \leqq \xi + \delta$. But $\xi - \delta$ belongs to A, so there must be a constant M_1 such that $|f(x)| \leqq M_1$ when $a \leqq x \leqq \xi - \delta$. If we let M_2 be the greater of M_1 and $|f(\xi)| + 1$, then $|f(x)| \leqq M_2$ when $a \leqq x \leqq \xi + \delta$. Therefore, $\xi + \delta$ must belong to A and we have a contradiction of the fact that the partition number ξ must be either the greatest number in A or the least number in B. All this shows that $\xi = b$, and we are almost finished. Let $\epsilon = 1$ and choose a positive number δ such that $a < b - \delta < b$ and

$|f(x)| < f(b) + 1$ when $b - \delta \leqq x \leqq b$. Since $b - \delta$ belongs to A, there must be a constant M_3 such that $|f(x)| \leqq M_3$ when $a \leqq x \leqq b - \delta$. If we let M be the greater of M_3 and $|f(b)| + 1$, then $|f(x)| \leqq M$ when $a \leqq x \leqq b$. This completes the proof of Theorem 5.45.

To strengthen Theorem 5.45, and for other purposes, we need information about upper and lower bounds. A set S of numbers is said to have an *upper bound* M_1 if $x \leqq M_1$ whenever x is in S and is said to have a *least upper bound* (l.u.b.) or supremum (sup) M if M is an upper bound and there is no upper bound M_1 for which $M_1 < M$. Analogously, S is said to have a lower bound m_1 if $x \geqq m_1$ whenever x is in S and is said to have a *greatest lower bound* (g.l.b.) or *infimum* (inf) m if m is a lower bound and there is no lower bound m_1 for which $m_1 > m$. The inequality

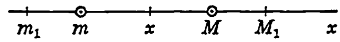


Figure 5.453

$$m_1 \leqq m \leqq x \leqq M \leqq M_1 \tag{5.452}$$

shows how these numbers must be related when x is in S, and Figure 5.453 shows a way in which they are sometimes related.

Theorem 5.46 *If a nonempty set S of numbers has an upper bound M_1, then it has a least upper bound. Similarly, if a nonempty set S of numbers has a lower bound m_1, then it has a greatest lower bound.*

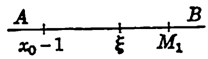


Figure 5.461

As Figure 5.461 indicates, we make a partition of numbers by putting a number in B if it is an upper bound of S and putting a number in A if it is not an upper bound of S. The set B contains M_1, and if x_0 is a number in S, then A contains the number $x_0 - 1$. Let ξ be the partition number. Let x be a number in S. Then, for each positive number ϵ, $x \leqq \xi + \epsilon$. Hence $x \leqq \xi$ and it follows that ξ is an upper bound of S. If $x' < \xi$, then x' is in A and hence x' is not an upper bound of S. Therefore, ξ is the least upper bound of S. This completes the proof of the first part of the theorem. The second part is proved similarly.

Theorem 5.47 *If f is continuous over an interval $a \leqq x \leqq b$, then $f(x)$ attains minimum and maximum values over the interval at points of the interval, that is, there exist numbers m, M, x_1, and x_2 such that $a \leqq x_1 \leqq b$, $a \leqq x_2 \leqq b$, and*

$$m = f(x_1) \leqq f(x) \leqq f(x_2) = M$$

whenever $a \leqq x \leqq b$.

To prove the part of the theorem involving M, we use Theorem 5.45 to conclude that f must have an upper bound M_1. It follows from Theorem 5.46 that the set of numbers $f(x)$ for which $a \leqq x \leqq b$ has a least upper bound which we now denote by M. Then $f(x) \leqq M$ when $a \leqq x \leqq b$, and it remains to be shown that there is a number x_2 for which $f(x_2) = M$.

If $f(a) = M$ or $f(b) = M$, we set $x_2 = a$ or $x_2 = b$. Otherwise, assuming that $f(a) < M$ and $f(b) < M$, we determine the required number x_2 by means of a Dedekind partition. Put a number x' in A if $x' \leqq a$ and also if $a < x' \leqq b$ and for some $\epsilon > 0$ the interval $a < x \leqq x'$ contains no point for which $f(x) > M - \epsilon$. Let B contain all other numbers, and observe that b is in B. Let x_2 be the partition number of this partition. Clearly, $f(x_2) \leqq M$. If we assume that $f(x_2) < M$, say $f(x_2) = M - \epsilon_0$, where $\epsilon_0 > 0$, we can choose a positive number δ such that $a < x_0 - \delta < x_2 + \delta < b$ and $f(x) < M - \epsilon_0/2$ whenever $x_0 - \delta \leqq x \leqq x_2 + \delta$. The fact that $x_2 - \delta$ is in A will then enable us to draw the erroneous conclusion that $x_2 + \delta$ is in A. Therefore, $f(x_2) = M$. This completes the proof of the part of Theorem 5.47 involving M. To prove the part of the theorem involving m, we can use an analogous argument. We can, alternatively and more easily, use the fact that $-f(x)$ must have a maximum $-m$ attained when $x = x_1$ and hence that $f(x)$ must have a minimum m attained when $x = x_1$.

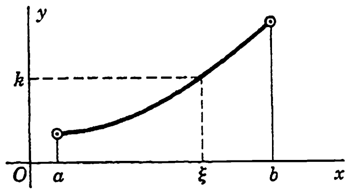


Figure 5.481

The following theorem is known as the *intermediate-value theorem;* Figure 5.481 provides experimental evidence.

Theorem 5.48 *If f is continuous over an interval $a \leqq x \leqq b$ and if k is a constant for which $f(a) < k < f(b)$ or $f(a) > k > f(b)$, then there exists at least one number ξ for which $a < \xi < b$ and $f(\xi) = k$.*

Taking first the case in which $f(a) < k < f(b)$, we prove the result with the aid of a Dedekind partition. Let x_1 be put in A if $x_1 \leqq a$ and also if $a \leqq x_1 \leqq b$ and $f(x) < k$ whenever $a \leqq x \leqq x_1$. Let x_2 be put in B if $x_2 \geqq b$ and also if $a \leqq x_2 \leqq b$ and the interval $a \leqq x \leqq x_2$ contains a number x for which $f(x) \geqq k$. Let ξ be the partition number. Since $f(a) < k$ and f has right-hand continuity at a, we can let $\epsilon = k - f(a)$ and choose a positive number δ such that $|f(x) - f(a)| < \epsilon$ and hence $f(x) < k$ when $a \leqq x \leqq a + \delta$. Hence $a + \delta$ belongs to A and $\xi \geqq a + \delta > a$. A similar argument shows that $\xi < b$. Therefore $a < \xi < b$. If we suppose that $f(\xi) > k$, then we can choose a positive number δ such that $a < \xi - \delta < \xi + \delta < b$ and $f(x) > k$ when $\xi - \delta \leqq x \leqq \xi + \delta$. This contradicts the fact that $f(x) < k$ when $a < x < \xi - \delta$ and $f(x) > k$ for some x in the interval $a < x < \xi + \delta$. If we suppose that $f(\xi) \geqq k$, a similar argument leads to a contradiction. Therefore $f(\xi) = k$. A very similar proof covers the case in which $f(a) > k > f(b)$ and Theorem 5.48 is proved.

In ordinary circumstances we try to be too efficient to clutter our books and our memories with obvious corollaries and applications of our theorems, but one corollary of the intermediate-value theorem is so important that we relax to look at it. *If a function f is negative at x_1*

and is positive at x_2, and if f is continuous over the closed interval having end points at x_1 and x_2, then there must be at least one x_3 between x_1 and x_2 for which $f(x_3) = 0$. This implies that the graph of a continuous function f cannot run from a point (x_1,y_1) below the x axis to a point (x_2,y_2) above the x axis without intersecting the x axis at a point $(x_3,0)$ for which x_3 lies between x_1 and x_2. Instead of asking whether this result is "obvious," we can ask whether it is obvious that a man cannot walk from the Capitol of South Dakota to the Capitol of North Dakota without stepping upon the common boundary of the two Dakotas.

Problems 5.49

1 With the text of this section out of sight, try to produce adequate responses to the following orders; if unsuccessful, study the text again and try again.

(*a*) Write a full statement of the Dedekind axiom.
(*b*) Write a theorem which gives precise information about boundedness of continuous functions.
(*c*) Write a theorem which gives precise information about extrema of continuous functions.
(*d*) Write a full statement of the intermediate-value theorem.

2 Using known properties of the function f for which $f(x) = x^2$, show how the intermediate-value theorem (Theorem 5.48) can be used to prove that there is a positive number ξ for which $\xi^2 = 2$. Give all of the details, recognizing that Theorem 5.48 cannot be applied until appropriate values of a and b have been captured.

3 Modify the work of the preceding problem to prove that there must be at least one x for which $f(x) = 0$ when

(*a*) $f(x) = x^3 - 7$ (*b*) $f(x) = x^3 - x - 7$

(*c*) $f(x) = \dfrac{x^3}{1 + x^2} - 40$ (*d*) $f(x) = x - \cos x$

4 Letting

$$f(x) = 1 + x + x^2 + x^3 + x^4, \tag{1}$$

determine whether there are any numbers x for which $f(x) < 0$. *Hint:* Use the fact that $f(1) = 5$ and

$$f(x) = \frac{x^5 - 1}{x - 1} \qquad (x \neq 1). \tag{2}$$

Show that $x^5 = 1$ only when $x = 1$, so $f(x)$ is never zero and hence $f(x)$ is never negative.

5 Let f be defined over the closed interval $-1 \leqq x \leqq 1$ by the formulas $f(0) = 0$ and $f(x) = 1/x^2$ when $-1 \leqq x \leqq 1$ and $x \neq 0$. Show that there is no constant M such that $|f(x)| \leqq M$ when $-1 \leqq x \leqq 1$. *Solution:* Suppose, intending to establish a contradiction of the supposition, that there is a number M

for which $|f(x)| \leqq M$ when $-1 \leqq x \leqq 1$. Let $x = 1/\sqrt{|M| + 2}$. Then $x \neq 0$ and $-1 \leqq x \leqq 1$, but $f(x) = 1/x^2 = |M| + 2$, so $|f(x)| > |M| \geqq M$.

6 Read Theorem 5.45. Then construct a figure which illustrates the meaning of the following remark. If $a < b$ and $M > 0$, then condition

$$|f(x)| < M \qquad (a \leqq x \leqq b) \tag{1}$$

is satisfied if and only if f is defined over the interval $a \leqq x \leqq b$ and the graph of $y = f(x)$ over the interval $a \leqq x \leqq b$ lies between the graphs of the lines having the equations $y = -M$ and $y = M$. Note that this gives a "geometrical meaning" to Theorem 5.45. Note that the inequality (1) holds if and only if $M > 0$ and $-M < f(x) < M$ when $a \leqq x \leqq b$.

7 Sketch a graph of the function f for which $f(0) = 0$ and $f(x) = 1/x$ when $x \neq 0$ and $-1 \leqq x \leqq 1$. Show that there is no M such that the graph of f over the interval $-1 \leqq x \leqq 1$ lies entirely above the line having the equation $y = -M$.

8 Give an example of a function which has an upper bound over the interval $-1 \leqq x \leqq 1$ but has no lower bound.

9 Show that the function f for which $f(x) = x$ has upper and lower bounds over the open interval $0 < x < 1$ but possesses neither a maximum nor a minimum over this interval.

10 Show that the function f defined over the closed interval $0 \leqq x \leqq 2$ by the formula

$$f(x) = x - [x],$$

in which $[x]$ denotes the greatest integer less than or equal to x, has an upper bound but does not have a maximum.

11 Prove that there is a number x^* for which $a < x < b$ and $f'(x^*) = [f(b) - f(a)]/(b - a)$ when

(*a*) $f(x) = x^2$, $a = 0$, $b = 1$
(*b*) $f(x) = x^6 - 7x^2 + 3x + 40$, $a = -1$, $b = 1$

12 Without undertaking extensive calculations that are easily made when appropriate computing equipment is available, we call attention to the Newton (1642–1727) method by which zeros of reasonable functions are approximated in decimal form. The method is based upon the elementary observation that, in many cases more or less like the one shown in Figure 5.491, if x_n (where n may initially be 1) is a reasonably good approximation to a number z for which $f(z) = 0$, then the tangent to the graph of $y = f(x)$ at the point $(x_n, f(x_n))$ will intersect the x axis at a point $(x_{n+1}, 0)$ for which x_{n+1} is a much better approximation to z.

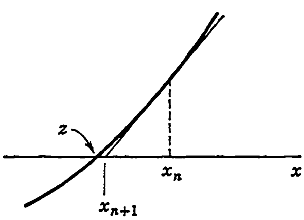


Figure 5.491

The Newton method is normally applied in cases where f has many continuous derivatives and $f'(x) \neq 0$ when x is near z but $x \neq z$. In such cases the equation of the tangent at $(x_n, f(x_n))$ is

$$y - f(x_n) = f'(x_n)(x - x_n)$$

and setting $y = 0$ gives

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

When the method is applied, we start with a first approximation x_1, set $n = 1$ to calculate a second approximation x_2, set $n = 2$ to calculate a third approximation x_3, and so on. To test the Newton method and understanding of it in situations where computations are not too onerous to be done with a pencil, calculate x_2 when

(*a*) $f(x) = x^2 - 2$, $x_1 = 1.4$
(*b*) $f(x) = x^3 - 20$, $x_1 = 3$
(*c*) $f(x) = x^4 + x - 20$, $x_1 = 2$
(*d*) $f(x) = x^4 + 5x - 50$, $x_1 = 2$

13 It is not always easy to tell what is obvious and what is not, the fundamental difficulty being that some things that have been thought to be "obviously true" are false. Consider, for example, the "obvious" statement that "each finite set of numbers contains a greatest element." If, as is usual, the empty set is considered to be a finite set, the statement is false. Consider, then, the revised statement "each nonempty finite set S of numbers contains a greatest element." Is this obviously true? Let n be a positive integer and let the numbers be $x_1, x_2, x_3, \cdots, x_n$. The only thing we know about these things is that they are numbers. One possible method of attacking the problem starts with a comparison of x_1 and x_2. If $x_1 \leqq x_2$, we discard x_1 and consider the remaining set, but if $x_2 < x_1$, we discard x_2. Instead of employing this "finite mathematics," we introduce some "infinite mathematics" that will make us think about least upper bounds. The fact that $x_k \leqq |x_k|$ for each k implies that

$$x_k \leqq |x_1| + |x_2| + \cdots + |x_n|$$

for each k. Hence the set S has an upper bound, and it follows from Theorem 5.46 that the set has a least upper bound M. If there is a k for which $x_k = M$, then this x_k must be a (or perhaps the) greatest element of S. If we suppose that there is no k for which $x_k = M$, and hence that $x_k < M$ for each k, we can obtain a contradiction of the hypothesis that S contains only n elements. To do this, let y_1 be an element of S. Then $y_1 < M$ and there must be an element y_2 of S for which $y_1 < y_2 < M$. The same argument shows that there must be an element y_3 of S for which $y_2 < y_3 < M$, and so on. We run into a contradiction of the assumption after we have used n elements of S. This proves that the set S does contain a greatest element and provides the possibility that schemes for finding "it" might even work.

13a *Remark:* To put the following problems and their consequences upon a rigorous base, we should have a definition of the set S_1 of positive integers. This set S_1 can be defined to be the subset of the set S of positive real numbers for which the number 1 is the least element in S_1; $a + b$ is in S_1 whenever a and b are in S_1; $b - a$ is in S_1 whenever a and b are in S_1 and $a < b$. It follows from this that if a and λ are numbers for which $0 < \lambda < 1$, then the interval $a \leqq x \leqq a + \lambda$ can contain at most one integer.

14 Prove the Archimedes property of numbers: if $\epsilon > 0$ and $a > 0$, then there is an integer n for which $n\epsilon > a$. *Solution:* Suppose $n\epsilon \leqq a$ for each n. Then the set S of numbers $\epsilon, 2\epsilon, 3\epsilon, \cdots$ is nonempty and has an upper bound. Hence S must have a least upper bound M. There must be an integer m for which $m\epsilon > M - \epsilon$. Then $(m+1)\epsilon > M$, and hence M is not an upper bound of S. This contradiction proves that there is an n for which $n\epsilon > a$.

15 Prove that each nonempty set of positive integers contains a least element. *Solution:* Let S be a set of positive integers. Since S is nonempty and has the lower bound 1, S must have a greatest lower bound m. Let $0 < \epsilon < \frac{1}{2}$. The interval $m \leqq x < m + \epsilon$ must contain an integer n in S, since otherwise $m + \epsilon$ would be a lower bound greater than m. Since the interval has length less than $\frac{1}{2}$ and cannot contain two integers, it follows that n is the one and only integer in S which is less than $m + \epsilon$. Therefore, n is the least integer in S. *Remark:* The fact that each nonempty set of positive integers contains a least element will now be used to prove the following *principle of mathematical induction. If a particular assertion involving a positive integer k is true when $k = 1$, and if the assertion is true when $k = n + 1$ provided $n \geqq 1$ and it is true when $k = n$, then the assertion is true for each positive integer.* Let T be the set of positive integers for which the assertion is true, and let F be the set for which the assertion is false. If F is nonempty, then F must have a least element m which is a positive integer greater than 1. Then $m - 1$ must be in T and our hypothesis gives the conclusion that m is in T. Thus m is in both F and T and (on the basis of the tacit assumption that we are dealing with statements that are either true or false but not both) we have a contradiction that proves that F is empty. Therefore, T contains each positive integer and the assertion is therefore true for each positive integer.

16 Prove that if x is a number, then there is an integer n for which $n \leqq x < n + 1$. *Remark:* This property of numbers was mentioned in Section 1.1, and the integer n is $[x]$, the greatest integer in x. *Solution:* Suppose first that $x > 2$. Using the Archimedes property of real numbers (Problem 14) with $\epsilon = 1$ and $a = x$ shows that the set S of integers greater than x is a nonempty set of positive integers. Hence, as Problem 15 shows, S must have a least element m. Then $x < m$, but the inequality $x < m - 1$ must be false because otherwise m would not be the least element of S. Therefore, $m - 1 \leqq x < m$ and we obtain our result by setting $n = m - 1$. In case $x \leqq 2$, we can choose an integer k such that $x + k > 2$. Letting m be an integer for which $m \leqq x + k < m + 1$, we find that $m - k$ is an integer n for which $n \leqq x < n + 1$.

17 Prove that if x is a number and n is a positive integer, then there is an integer k for which

$$\frac{k}{n} \leqq x < \frac{k+1}{n}.$$

Remark: In case m is a nonnegative integer and $n = 10^m$, the result shows how x is related to "finite decimals." *Proof:* Problem 16 shows that there is an integer k for which

$$k \leqq nx < k + 1,$$

and the result is obtained by dividing by n.

18 Prove that if x is a number, then there exist an integer N and a sequence $d_1, d_2, d_3, \cdots$ of digits (a digit being one of the integers $0, 1, 2, \cdots, 9$) such that

$$(1)\quad N + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_{n-1}}{10^{n-1}} + \frac{d_n}{10^n} \leqq x < N + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_{n-1}}{10^{n-1}} + \frac{d_n + 1}{10^n}$$

for each $n = 1, 2, 3, \cdots$. *Solution:* Let $N = [x]$, so that N is the greatest integer in x, and let $\theta = x - N$. Then $0 \leqq \theta < 1$ and the required result will follow if we prove that there exist digits $d_1, d_2, d_3, \cdots$ such that

$$(2)\quad \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_{n-1}}{10^{n-1}} + \frac{d_n}{10^n} \leqq \theta < \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_{n-1}}{10^{n-1}} + \frac{d_n + 1}{10^n}$$

for each $n = 1, 2, 3, \cdots$. To prove (2), it is sufficient to prove that

$$(3)\quad 0 \leqq \theta - \frac{d_1}{10} - \frac{d_2}{10^2} - \cdots - \frac{d^n}{10^n} < \frac{1}{10^n}.$$

While it is of interest to take time to use (2) to determine what $d_1, d_2, d_3, \cdots$ must be if they exist, we save time by defining integers $d_1, d_2, d_3, \cdots$ by the formulas

$$(4)\quad d_1 = [10\theta]$$

$$(5)\quad d_2 = \left[10^2\left(\theta - \frac{d_1}{10}\right)\right]$$

$$(6)\quad d_3 = \left[10^3\left(\theta - \frac{d_1}{10} - \frac{d_2}{10^2}\right)\right]$$

and, in general, for each $n = 1, 2, 3, \cdots$

$$(7)\quad d_{n+1} = \left[10^{n+1}\left(\theta - \frac{d_1}{10} - \frac{d_2}{10^2} - \cdots - \frac{d_n}{10^n}\right)\right].$$

Since $0 \leqq \theta < 1$, we find that $0 \leqq 10\theta < 10$ and hence that d_1 is a digit. Moreover,

$$(8)\quad d_1 \leqq 10\theta < d_1 + 1$$

and hence

$$(9)\quad 0 \leqq \theta - \frac{d_1}{10} < \frac{1}{10},$$

so (3) holds when $n = 1$. Multiplying (9) by 10^2 gives

$$(10)\quad 0 \leqq 10^2\left(\theta - \frac{d_1}{10}\right) < 10.$$

Hence (5) shows that d_2 is a digit and

$$d_2 \leqq 10^2\left(\theta - \frac{d_1}{10}\right) < d_2 + 1. \tag{11}$$

Dividing by 10^2 and transposing give

$$0 < \theta - \frac{d_1}{10} - \frac{d_2}{10^2} < \frac{1}{10^2}, \tag{12}$$

so (3) holds when $n = 2$. This procedure enables us to prove (3) by induction. If $d_1, d_2, \cdots d_n$ are digits and (3) holds, then

$$0 < 10^{n+1}\left(\theta - \frac{d_1}{10} - \frac{d_2}{10^2} - \cdots - \frac{d^n}{10^n}\right) < 10 \tag{13}$$

and (7) shows that d_{n+1} is a digit and

$$d_{n+1} \leqq 10^{n+1}\left(\theta - \frac{d_1}{10} - \frac{d_2}{10^2} - \cdots - \frac{d_n}{10^n}\right) < d_{n+1} + 1. \tag{14}$$

Dividing by 10^{n+1} and transposing give the result of replacing n by $n + 1$ in (3). This proves (3) by induction, that is, by use of the principle of mathematical induction of Problem 15.

19 Let $F(\theta)$ be the temperature or pressure at the place P where a circle having its center at the origin of an x,y coordinate system is intersected by the ray from the origin which makes the angle θ (as in trigonometry) with the positive x axis. It is supposed that F is continuous and $F(\theta + 2\pi) = F(\theta)$ for each θ. Prove that there are two diametrically opposite points of the circle at which F has equal values. *Hint:* Apply the intermediate-value theorem to the function f for which $f(\theta) = F(\theta) - F(\theta + \pi)$. Observe that if $f(\theta_0) > 0$, then $f(\theta_0 + \pi) = -f(\theta_0) < 0$. *Remark:* While we do not yet have equipment required for proof, we can learn an interesting property of continuous functions defined over surfaces like spheres. There are two antipodal (or diametrically opposite) places on the surface of the earth having both equal temperatures and equal atmospheric pressures.

20 Suppose that a world has existed so long and so favorably to fish that an infinite number of fish have existed but that only a finite number of fish have existed at any one time because the world contains only a finite number of atoms. Prove that there is a least number m_0 such that the mass m (measured in some standard system) of each past and present and future fish is less than or equal to m_0.

21 It is easy to presume that if f is differentiable over the interval $-1 \leqq x \leqq 1$ and if $f'(0) = 1$, then there must be a positive number h such that f is increasing over the interval $-h \leqq x \leqq h$. Use the function f for which $f(0) = 0$ and

$$f(x) = x + x^2 \sin \frac{1}{x^2} \qquad (x \neq 0)$$

to show that the presumption is false. *Hint:* Show that $f'(0) = 1$ and that, when $x \neq 0$,

$$f'(x) = 1 + 2x \sin \frac{1}{x^2} - \frac{2}{x} \cos \frac{1}{x^2}.$$

Observe that if n is a positive integer and $x_n = 1/\sqrt{2n\pi}$, then

$$f'(x_n) = 1 - 2\sqrt{2n\pi}.$$

It follows that each interval $0 < x < h$ contains subintervals over which $f'(x) < 0$ and f is decreasing.

22 Boom-and-bust processes occur (or seem to occur) in economic and political life. Persons who get their political information from clever press secretaries of astute chiefs of state discover that the fortunes of their countries are at low ebbs when new chiefs are installed and that these fortunes steadily improve during the tenure of each chief. Such processes occur in electrical engineering when a charge on a capacitor steadily increases until a spark jumps and the charge disappears. This problem involves a particular boom-and-bust process in which a and q are positive constants. It is supposed that, for each integer n, the quantity y is 0 when $t = na$ and that y increases at a constant rate over the interval $na \leqq t < (n+1)a$ in such a way that y approaches q as t approaches $(n+1)a$ from the left. Sketch a graph of y versus t and find a formula giving y in terms of t. *Partial ans:*

$$y = q\left(\frac{t}{a} - \left[\frac{t}{a}\right]\right),$$

where $[x]$ denotes the greatest integer in x.

23 While persons confining their mathematical contacts to modern mathematics books need not worry about the matter, others may need a warning. In the good old days, the word "finite" was used in place of the word "bounded." In order to understand assertions involving the word finite, it is sometimes not sufficient to understand modern mathematics. Sometimes we need substantial information about history, and sometimes we need conscious recognition of the fact that assertions involving the word "finite" have different meanings at different places and at different times. For example, the assertion that "f is finite at x_0" can mean that there is an interval with center at x_0 such that f is bounded over the interval. The assertion can, however, have other meanings, and this is the reason why we should shudder when we hear it.

24 A function f is said to have a *generalized first derivative* $Gf'(x)$ at x if

$$\text{(1)} \qquad Gf'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x-h)}{2h}$$

and is said to have a *generalized second derivative* $Gf''(x)$ at x if

$$\text{(2)} \qquad Gf''(x) = \lim_{h\to 0} \frac{f(x-h) - 2f(x) + f(x+h)}{h^2}.$$

Prove that $Gf'(x) = f'(x)$ when $f'(x)$ exists. *Hint:* Use the fact that

$$\text{(3)} \qquad \frac{f(x+h) - f(x-h)}{2h} = \frac{1}{2}\left[\frac{f(x+h) - f(x)}{h} + \frac{f(x-h) - f(x)}{-h}\right].$$

Remark: The wide world contains several persons who have sharpened their wits by trying to answer two questions which are not guaranteed to be easily answered. Does the hypothesis that $Gf'(x) = 0$ when $a < x < b$ imply that there is a constant c for which $f(x) = c$ when $a < x < b$? Does the hypothesis that $Gf''(x) = 0$ when $a < x < b$ imply that there are constants c_1 and c_2 for which $f(x) = c_1x + c_2$ when $a < x < b$?

5.5 *The Rolle theorem and the mean-value theorem* In this section we prove some fundamental theorems and use them to review and prove some theorems that have been previously given without proof. The following theorem must be permanently remembered and known as the *Rolle* (Michel, 1652–1719) *theorem.* It is not to be presumed that Rolle proved or even knew this theorem, but he did discover some of its applications to polynomials.

Theorem 5.51 (**Rolle theorem**) *If $a < b$, if f is continuous over $a \leqq x \leqq b$, if f is differentiable over $a < x < b$, and if $f(a) = f(b) = 0$, then there is at least one number x^* for which $a < x^* < b$ and $f'(x^*) = 0$.*

The proof of this theorem is mildly tricky because it seems to be necessary to consider three different cases. Suppose first that $f(x) = 0$ over the whole interval $a \leqq x \leqq b$. Then $f'(x) = 0$ when $a < x < b$ and we can choose x^* to be any number between a and b. Suppose next that there is a number x_1 for which $a < x_1 < b$ and $f(x_1) > 0$. Then with the aid of Theorem 5.47 we see that f must attain a positive maximum $f(x^*)$ at some point x^* for which $a \leqq x^* \leqq b$, and we can be sure that $a < x^* < b$ because $f(a) = f(b) = 0$. Since $f'(x^*)$ exists, it follows from Theorem 5.26 that $f'(x^*) = 0$. Suppose finally that there is a number x_2 for which $a < x_2 < b$ and $f(x_2) < 0$. Arguments similar to those used above then show that f must have a negative minimum at some point x^* for which $a < x^* < b$ and that $f'(x^*) = 0$.

The following theorem is known as the law of the mean or the mean-value theorem of the derivative calculus. It is a strengthened version of Theorem 5.41, which we have discussed briefly, the right member of (5.53) being the slope of the chord joining the points $(a, f(a))$ and $(b, f(b))$. It, like the Rolle theorem, must be permanently remembered.

Theorem 5.52 (**mean-value theorem**) *If $a < b$, if f is continuous over $a \leqq x \leqq b$, and if f is differentiable over $a < x < b$, then there is at least one number x^* for which $a < x^* < b$ and*

$$f'(x^*) = \frac{f(b) - f(a)}{b - a} \tag{5.53}$$

or

$$f(b) - f(a) = f'(x^*)(b - a). \tag{5.54}$$

This theorem differs from the Rolle theorem because it is not assumed that $f(a) = f(b) = 0$. It happens, however, that the theorem can be proved by applying the Rolle theorem to the function ϕ defined by $\phi(x) = f(x) - g(x)$, where g is the function whose graph is the chord

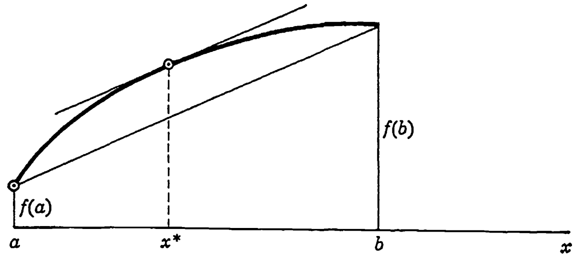


Figure 5.55

joining the points $(a, f(a))$ and $(b, f(b))$ of Figure 5.55. The point-slope form of the equation of a line gives the formula

$$g(x) - f(a) = \frac{f(b) - f(a)}{b - a}(x - a) \tag{5.561}$$

in which $g(x)$ appears instead of the more familiar y. Hence,

$$\phi(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a) \tag{5.562}$$

when $a \leqq x \leqq b$ and

$$\phi'(x) = f'(x) - \frac{f(b) - f(a)}{b - a} \tag{5.563}$$

when $a < x < b$. It is easily seen that ϕ satisfies the hypotheses of the Rolle theorem. Choosing x^* such that $a < x^* < b$ and $\phi'(x^*) = 0$ gives the required conclusion (5.53). Multiplying by $(b - a)$ then gives (5.54), and Theorem 5.52 is proved.

Theorem 5.57 *If f is continuous over $a \leqq x \leqq b$ and $f'(x) = 0$ when $a < x < b$, then $f(x) = f(a)$ when $a \leqq x \leqq b$.*

To prove this theorem, we note first that $f(x) = f(a)$ when $x = a$. If $a < x_1 \leqq b$, we can apply the mean-value theorem to the interval $a \leqq x \leqq x_1$ to conclude that there is a number x^* for which $a < x^* < x_1$ and

$$f(x_1) - f(a) = f'(x^*)(x_1 - a).$$

But $f'(x^*) = 0$ and hence $f(x_1) = f(a)$. Therefore, $f(x) = f(a)$ when $a \leqq x \leqq b$ and Theorem 5.57 is proved. It follows from this theorem that if two functions F_1 and F_2 have the same derivative over an interval, say $F_1'(x) = F_2'(x) = g(x)$ when $a \leqq x \leqq b$, and we put $f(x) = F_2(x) - F_1(x)$, then $f'(x) = 0$ when $a \leqq x \leqq b$ so $f(x)$ must be a constant c and

$$F_2(x) = F_1(x) + c$$

when $a \leqq x \leqq b$. This proves Theorem 4.13.

We now prove Theorem 5.27 the first part of which says that if f is continuous over an interval $a_0 \leqq x \leqq b_0$ and $f'(x) > 0$ when $a_0 < x < b_0$, then f is increasing over the interval $a_0 \leqq x \leqq b_0$. Let $a_0 \leqq x_1 < x_2 \leqq b_0$. The mean-value theorem guarantees existence of a number x^* such that $x_1 < x^* < x_2$ and

$$f(x_2) - f(x_1) = f'(x^*)(x_2 - x_1).$$

But $f'(x^*) > 0$ and $(x_2 - x_1) > 0$, so $f(x_2) - f(x_1) > 0$. Thus $f(x_2) > f(x_1)$ and f is increasing. For the second part of Theorem 5.27, everything is the same except that $f'(x^*) < 0$ and f is decreasing.

The following theorem expresses the fact that if a function f is continuous over a closed interval $a \leqq x \leqq b$, then it is *uniformly continuous* over the interval.

Theorem 5.58 *If f is continuous over a closed interval $a \leqq x \leqq b$, then to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that*

$$|f(x_1) - f(x_2)| < \epsilon$$

whenever $a \leqq x_1 \leqq b$, $a \leqq x_2 \leqq b$, and $|x_2 - x_1| < \delta$.

While neater proofs of this theorem can be given in advanced calculus after more mathematics has been digested, there is virtue in knowing that it is possible to base a proof upon a straightforward application of the Dedekind axiom 5.43. The bookkeeping by which we inch along toward the answer is really very elementary, and students who have the patience to see that this is so are very likely to become the leading scientists of the future. Let ϵ be a given positive number. Let a number x be placed in the set A if $x \leqq a$ and also if $a < x \leqq b$ and there is a positive number δ such that $|f(x_2) - f(x_1)| < \epsilon$ whenever $a \leqq x_1 \leqq x$, $a \leqq x_2 \leqq x$, and $|x_2 - x_1| \leqq \delta$. Let B contain each number x not placed in A. Let ξ be the partition number. Clearly, $a \leqq \xi \leqq b$. Since f is continuous at a, we can choose a positive number δ_1 such that $|f(x) - f(a)| < \epsilon/2$ whenever $a \leqq x \leqq a + \delta_1$. Then

$$\begin{aligned} |f(x_2) - f(x_1)| &= |[f(x_2) - f(a)] - [f(x_1) - f(a)]| \\ &\leqq |f(x_2) - f(a)| + |f(x_1) - f(a)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} < \epsilon \end{aligned}$$

whenever $a \leqq x_1 \leqq a + \delta_1$, $a \leqq x_2 \leqq \delta_1$, and $|x_2 - x_1| \leqq \delta_1$. Therefore, $a + \delta_1$ belongs to A and it follows that $\xi > a$. Our next step is to prove that $\xi = b$ by obtaining a contradiction from the assumption that $a < \xi < b$. Suppose, then, that $a < \xi < b$. Since f is continuous at ξ, we can choose a positive number δ_2 such that $a < \xi - \delta_2 < \xi + \delta_2 < b$ and $|f(x) - f(\xi)| < \epsilon/2$ whenever $|x - \xi| \leqq \delta_2$. Moreover, since $\xi - \delta_2$ must belong to A, we can choose a positive number δ_3 such that $\delta_3 < \delta_2$ and $|f(x_2) - f(x_1)| < \epsilon/2$ whenever $a \leqq x_1 \leqq \xi - \delta_2$, $a \leqq x_2 \leqq \xi - \delta_2$, and $|x_2 - x_1| < \delta_3$. Now suppose that $a \leqq x_1 \leqq \xi + \delta_2$, $a \leqq x_2 \leqq \xi +$

δ_2, and $|x_2 - x_1| < \delta_3$. We may suppose that $x_1 \leqq x_2$. If x_1 and x_2 both lie in the interval $|x - \xi| \leqq \delta_2$, then $|f(x_2) - f(x_1)| < \epsilon$. If x_1 and x_2 both lie in the interval $a \leqq x \leqq \xi - \delta_2$, then again $|f(x_2) - f(x_1)| < \epsilon$. If $a \leqq x_1 \leqq \xi - \delta_2$ and $\xi - \delta_2 < x_2 \leqq \xi + \delta_2$, then

$$|f(x_2) - f(x_1)| \leqq |f(x_2) - f(\xi - \delta_2)| + |f(\xi - \delta_2) - f(x_1)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, in each case, $|f(x_2) - f(x_1)| < \epsilon$, so $\xi + \delta_2$ is in A and ξ cannot be the partition number. This contradiction shows that $\xi = b$, but to complete the proof of the theorem, we must show that b is in the set A. Since f has left-hand continuity at b, we can choose $\delta' > 0$ such that $a < b - \delta' < b$ and $|f(x) - f(b)| < \epsilon/2$ whenever $b - \delta' \leqq x \leqq b$. Since $b - \delta'$ must belong to A, we can choose a positive number δ such that $\delta < \delta'$ and $|f(x_2) - f(x_1)| < \epsilon/2$ whenever $a \leqq x_1 \leqq b - \delta'$, $a \leqq x_2 \leqq b - \delta'$, and $|x_2 - x_1| \leqq \delta$. Arguments used above show that $|f(x_2) - f(x_1)| < \epsilon$ whenever $a \leqq x_1 \leqq b$, $a \leqq x_2 \leqq b$, and $|x_2 - x_1| < \delta$. This completes the proof of Theorem 5.58.

Problems 5.59

1 With the text out of sight, write completely accurate statements of (*a*) the Rolle theorem, (*b*) the mean-value theorem.

2 Sketch several graphs that seem to be graphs of functions satisfying the hypotheses of the Rolle theorem, and see whether it seems to be true that the star points exist.

3 Sketch several graphs that seem to be graphs of functions which, for one reason or another, do *not* satisfy the hypotheses of the Rolle theorem but nevertheless it seems to be true that star points exist anyway.

4 Sketch several graphs that seem to be graphs of functions which, for one reason or another, do *not* satisfy the hypotheses of the Rolle theorem and star points do not exist.

5 Tell why there is no request for construction of graphs of functions that satisfy the hypotheses of the Rolle theorem and star points do not exist.

6 Prove that if $F'(x) > 0$ when $a < x < b$, then there is at most one x for which $a < x < b$ and $F(x) = 0$. *Solution:* If we suppose that $a < x_1 < x_2 < b$ and $F(x_1) = F(x_2) = 0$, an application of the Rolle theorem yields the conclusion that there is a number ξ for which $x_1 < \xi < x_2$ and $F'(\xi) = 0$. This contradicts the hypotheses and the result follows.

7 This problem requires us to review fundamental processes of the calculus whose validity depends upon Theorem 5.57. Supposing that f and g are functions defined and continuous over an interval containing the point $x = a$ and that

$$(1) \qquad f'(x) = g(x), \qquad f(a) = A,$$

we can then write

$$(2) \qquad f(x) = \int g(x)\, dx + c,$$

where $\int g(x)\,dx$ stands for some particular function whose derivative is $g(x)$, and then so determine c that $f(a) = A$. We can also determine f from the formula

$$f(x) = f(a) + \int_a^x f'(t)\,dt \tag{3}$$

in which the integral is a Riemann integral. Determine f in two different (or superficially different) ways by using (2) and by using (3), and make the results agree, when

(*a*) $f'(x) = 2x$, $f(2) = 3$
(*b*) $f'(x) = \sin ax$, $f(0) = 0$
(*c*) $f'(x) = \cos ax$, $f(0) = 0$
(*d*) $f'(x) = e^{ax}$, $f(0) = 1$
(*e*) $f'(x) = \dfrac{1}{x}$, $f(2) = 3$
(*f*) $f'(x) = \sqrt{x}$, $f(4) = 0$

8 Prove that if u and v are functions that have continuous derivatives over an interval I containing a and x, then

$$\int_a^x u(t)v'(t)\,dt = u(t)v(t)\Big]_a^x - \int_a^x v(t)u'(t)\,dt.$$

Hint: Let $F_1(x)$ and $F_2(x)$ denote the left and right sides of the formula. Then show that $F_1(a) = F_2(a)$ and that $F_1'(x) = F_2'(x)$ when x is in I.

9 If $f(0) = 0$ and

$$f'(x) = \frac{x^{27}}{27 + x^{27}},$$

the result of writing formula (2) of Problem 7 as an aid to finding $f(1)$ is rather (or more) futile, but we can write a version of (3) and undertake to estimate $f(1)$. Do it.

10 Sketch a graph over the interval $0 \leqq x \leqq 1$ of the function f for which $f(x) = x^2$. Let $\epsilon = \frac{1}{8}$. Use your eyes to select a $\delta > 0$ such that $|f(x_2) - f(x_1)| < \epsilon$ whenever $0 \leqq x_1 \leqq 1$, $0 \leqq x_2 \leqq 1$, and $|x_2 - x_1| < \delta$. Note that if this f were the only continuous function, we would not need to work so long to prove Theorem 5.58.

11 Supposing that f has a continuous derivative over the interval $a \leqq x \leqq b$, show that the functions F and G for which

$$F(x) = f(a) + \tfrac{1}{2}\int_a^x [1 + f'(t) + |f'(t)|]\,dt$$
$$G(x) = \tfrac{1}{2}\int_a^x [1 - f'(t) + |f'(t)|]\,dt$$

are both increasing over the interval $a \leqq x \leqq b$ and

$$f(x) = F(x) - G(x).$$

Remark: This problem contains an important idea. It is sometimes useful to know about the possibility of representing a given function as the difference of two increasing functions.

12 Prove the following theorem, which is known as an extended (not generalized) mean-value theorem or as a Taylor theorem.

(1) **Theorem** *If f is such that f'' exists over an interval containing a and x, then there is at least one number x^* between a and x such that*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(x^*)}{2!}(x - a)^2. \tag{2}$$

Solution: Let

$$\phi(t) = f(x) - f(t) - \frac{f'(t)}{1!}(x - t) - \frac{C}{2!}(x - t)^2, \tag{3}$$

where C is a constant chosen such that $\phi(a) = 0$. Then $\phi(a) = \phi(x) = 0$ and the other hypotheses of the Rolle theorem are satisfied. The Rolle theorem therefore furnishes a number x^* between a and x for which $\phi'(x^*) = 0$. Thus,

$$\phi'(x^*) = -f'(x^*) + f'(x^*) - f''(x^*)(x - x^*) + C(x - x^*) = 0. \tag{4}$$

Therefore, $C = f''(x^*)$. Since $\phi(a) = 0$, we can put $t = a$ in (3), equate the result to 0, and transpose to obtain the required formula (2). *Remark:* With the additional hypothesis that the second derivative f'' is continuous, we shall use integration by parts in Section 12.5 to obtain more straightforward derivations of (2) and related formulas.

13 This problem requires attainment of understanding of matters relating to the following *generalized mean-value theorem* which involves two functions.

(1) **Theorem** *Let f and g be continuous over the closed interval from a to x, let f and g be differentiable over the open interval from a to x, and let the derivative g' be different from zero over the open interval from a to x. Then there exists an x^* in the open interval from a to x such that*

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(x^*)}{g'(x^*)}. \tag{2}$$

We assume that f and g are given functions satisfying the hypotheses of the theorem. Two applications of the mean-value theorem then show that there exist numbers x_1^* and x_2^* between a and x such that

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{\dfrac{f(x) - f(a)}{x - a}}{\dfrac{g(x) - g(a)}{x - a}} = \frac{f'(x_1^*)}{g'(x_2^*)}. \tag{3}$$

While this result can be useful, it is crude because x_1^* and x_2^* are not necessarily equal. We obtain the more useful and elegant result (2) by arranging our work to make a single application of the Rolle theorem. The trick is to define a new function ϕ by the formula

$$\phi(t) = [f(x) - f(a)][g(t) - g(a)] - [f(t) - f(a)][g(x) - g(a)]. \tag{4}$$

It is easy to see that $\phi(a) = \phi(x) = 0$ and that the other hypotheses of the Rolle theorem are satisfied. Hence the Rolle theorem implies that there is a number x^* between a and x for which

$$\phi'(t^*) = [f(x) - f(a)]g'(x^*) - f'(x^*)[g(x) - g(a)] = 0, \tag{5}$$

and the desired formula (2) follows from this. The generalized mean-value theorem has numerous applications, including the following one. Since x^* lies between a and x, (2) implies that

$$\lim_{x\to a}\frac{f(x)-f(a)}{g(x)-g(a)}=\lim_{x\to a}\frac{f'(x)}{g'(x)} \tag{6}$$

provided the limit on the right exists.

For the special case in which $f(a) = g(a) = 0$, (6) reduces to the famous and useful *L'Hôpital formula*

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)} \tag{7}$$

which, like (6), is correct when the limit on the right exists.† In this case the quotient $f(x)/g(x)$ is said to be "an indeterminate form of the form 0/0 when $x = a$." The formula (7), which gives a method for finding limits of "indeterminate forms," is called "the L'Hôpital rule for evaluation of indeterminate forms." Stories about "evaluation of indeterminate forms" will not injure us if we resolutely remember that we sometimes find limits but that we never evaluate 0/0. When we apply the L'Hôpital formula (7), we must not fall asleep at the switch and write the derivative of the quotient $f(x)/g(x)$; we write the quotient of the derivatives. The following rather simple examples show how the formula is applied:

$$\lim_{x\to 1}\frac{x^2-1}{x-1}=\lim_{x\to 1}\frac{2x}{1}=2 \tag{8}$$

$$\lim_{x\to 0}\frac{\sin x}{x}=\lim_{x\to 0}\frac{\cos x}{1}=1 \tag{9}$$

$$\lim_{x\to 0}\frac{1-\cos x}{x}=\lim_{x\to 0}\frac{\sin x}{1}=0 \tag{10}$$

$$\lim_{x\to 0}\frac{1-\cos x}{x^2}=\lim_{x\to 0}\frac{\sin x}{2x}=\lim_{x\to 0}\frac{\cos x}{2}=\frac{1}{2} \tag{11}$$

$$\lim_{x\to 0}\frac{e^x-1}{x}=\lim_{x\to 0}\frac{e^x}{1}=1 \tag{12}$$

$$\lim_{x\to 0}\frac{e^x-1-x}{x^2}=\lim_{x\to 0}\frac{e^x-1}{2x}=\lim_{x\to 0}\frac{e^x}{2}=\frac{1}{2} \tag{13}$$

$$\begin{aligned}\lim_{x\to 0}\left(\frac{1}{\sin x}-\frac{1}{x}\right)&=\lim_{x\to 0}\frac{x-\sin x}{x\sin x}\\&=\lim_{x\to 0}\frac{1-\cos x}{x\cos x+\sin x}=\lim_{x\to 0}\frac{\sin x}{2\cos x-x\sin x}=\frac{0}{2}=0.\end{aligned} \tag{14}$$

14 Show that

$$\lim_{x\to 1}\frac{nx^{n+1}-(n+1)x^n+1}{(x-1)^2}=\frac{n(n+1)}{2}.$$

† Guillaume François de l'Hospital (1661–1704) introduced this formula in a book *Analyse des infiniment petits pour intelligence des lignes courbes* (Paris, 1696, 182 pp.) which enjoyed widespread popularity. While there can be objections to tinkering with names of people, most authorities insist that we must accept evolution from l'Hospital to L'Hôpital quite as cheerfully as we accept evolution from hostel to hotel. Even counterrevolutionists must recognize that the name is spelled in different ways.

15 Supposing that n is a positive integer and $x \neq 1$, differentiate

$$1 + x + x^2 + x^3 + \cdots + x^n = \frac{x^{n+1} - 1}{x - 1} \tag{1}$$

to obtain

$$1 + 2x + 3x^2 + \cdots + nx^{n-1} = \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2}. \tag{2}$$

Multiply by x and differentiate again to obtain

$$1^2 + 2^2x + 3^2x^2 + \cdots + n^2x^{n-1} = \frac{n^2x^{n+2} - (2n^2 + 2n - 1)x^{n+1} + (n+1)^2x^n - x - 1}{(x-1)^3}. \tag{3}$$

Finally, take limits as $x \to 1$ to obtain the formula

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}. \tag{4}$$

Remark: We could multiply (3) by x and differentiate and continue the process to obtain formulas for sums of cubes and higher powers. The details mushroom rapidly as we proceed.

16 Another L'Hôpital rule is embodied in the following theorem.

(1) **Theorem** *If f and g are differentiable over the infinite interval $x \geqq x_0$ and if*

$$\lim_{x\to\infty} f(x) = \infty, \qquad \lim_{x\to\infty} g(x) = \infty, \tag{2}$$

then

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)} \tag{3}$$

provided the limit on the right exists.

To prove this theorem, suppose that the right member of (3) exists and is L. Let

$$\phi(x) = f(x) - Lg(x). \tag{4}$$

Then

$$\lim_{x\to\infty} \frac{\phi'(x)}{g'(x)} = \lim_{x\to\infty} \frac{f'(x) - Lg'(x)}{g'(x)} = 0. \tag{5}$$

Let $\epsilon > 0$. Choose x_1 such that $x_1 > x_0$, $g(x^*) > 0$ when $x^* \geqq x_1$, and

$$\left|\frac{\phi'(x^*)}{g'(x^*)}\right| < \frac{\epsilon}{2} \qquad (x^* \geqq x_1). \tag{6}$$

It then follows from the generalized mean-value theorem of Problem 13 that

$$\left|\frac{\phi(x) - \phi(x_1)}{g(x) - g(x_1)}\right| < \frac{\epsilon}{2} \qquad (x > x_1). \tag{7}$$

Choose x_2 such that $x_2 > x_1$ and $g(x) > g(x_1)$ when $x > x_2$. It then follows from (7) that

$$\left| \frac{\phi(x) - \phi(x_1)}{g(x)} \right| < \frac{\epsilon}{2} \qquad (x > x_2). \tag{8}$$

Choose x_3 such that $x_3 > x_2$ and $|\phi(x_1)/g(x)| < \epsilon/2$ when $x > x_3$. Then

$$\left| \frac{\phi(x)}{g(x)} \right| = \left| \frac{\phi(x) - \phi(x_1)}{g(x)} + \frac{\phi(x_1)}{g(x)} \right| \leqq \left| \frac{\phi(x) - \phi(x_1)}{g(x)} \right| + \left| \frac{\phi(x_1)}{g(x)} \right| < \epsilon \tag{9}$$

when $x > x_3$. Therefore,

$$\lim_{x\to\infty} \left[\frac{f(x)}{g(x)} - L \right] = \lim_{x\to\infty} \frac{\phi(x)}{g(x)} = 0, \tag{10}$$

and this gives the required conclusion (3) which involves "indeterminate forms of the form ∞/∞." The following rather simple examples show how the theorem is applied.

$$\lim_{x\to\infty} \frac{x}{x^2+1} = \lim_{x\to\infty} \frac{1}{2x} = 0 \tag{11}$$

$$\lim_{x\to\infty} \frac{x^2}{x^3+x+1} = \lim_{x\to\infty} \frac{2x}{3x^2+1} = \lim_{x\to\infty} \frac{2}{6x} = 0 \tag{12}$$

$$\lim_{x\to\infty} \frac{x+\sqrt{x}}{x+1} = \lim_{x\to\infty} \frac{1+\frac{1}{2}x^{-1/2}}{1} = 1 \tag{13}$$

$$\lim_{x\to\infty} \frac{\log x}{x^p} = \lim_{x\to\infty} \frac{x^{-1}}{px^{p-1}} = \lim_{x\to\infty} \frac{1}{px^p} = 0 \qquad (p > 0) \tag{14}$$

$$\lim_{t\to 0+} t \log t = \lim_{x\to\infty} \frac{1}{x} \log \frac{1}{x} = \lim_{x\to\infty} \frac{-\log x}{x} = \lim_{x\to\infty} \frac{-1/x}{1} = 0. \tag{15}$$

Remark: Limits of functions of other types can be found by using the above formulas. For example, to find $\lim\limits_{x\to 0+} x^x$, we put $y = x^x$ and find that $\log y = x \log x$, so $\lim\limits_{x\to 0+} \log y = 0$, $\lim\limits_{x\to 0+} y = 1$, and $\lim\limits_{x\to 0+} x^x = 1$. Similar arguments show that $\lim\limits_{x\to\infty} x^{1/x} = 1$.

17 Supposing that f, g, h are three functions satisfying the hypotheses of the mean-value theorem, show that the function F defined by the first of the formulas

$$F(x) = \begin{vmatrix} f(x) & g(x) & h(x) \\ f(a) & g(a) & h(a) \\ f(b) & g(b) & h(b) \end{vmatrix}, \qquad \begin{vmatrix} f'(x^*) & g'(x^*) & h'(x^*) \\ f(a) & g(a) & h(a) \\ f(b) & g(b) & h(b) \end{vmatrix} = 0$$

satisfies the hypotheses of the Rolle theorem and hence that there is a number x^* for which $a < x < b$ and the second formula holds. Examine the case in which $h(x) = 1$ for each x.

18 This problem provides an opportunity to learn some very interesting mathematics but, like a bicycle rider who lacks appreciation of basic principles of physics and engineering, we can pedal along without it. The following theorem is a fundamental theorem of the calculus which is stronger than Theorem 4.37 because it does not require that f be continuous.

Theorem *If f is integrable (Riemann) over $a \leqq x \leqq b$ and if $F'(x) = f(x)$ when $a \leqq x \leqq b$, then*

$$\int_a^b f(x)\,dx = F(x)\Big]_a^b = F(b) - F(a). \tag{1}$$

Our proof of this theorem uses the mean-value theorem. Supposing that P is a partition of the interval $a \leqq x \leqq b$ having partition points $x_0, x_1, \cdots, x_n$ for which

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b, \tag{2}$$

we find that

$$F(b) - F(a) = \sum_{k=1}^{n} [F(x_k) - F(x_{k-1})] = \sum_{k=1}^{n} F'(x_k^*)(x_k - x_{k-1}) \tag{3}$$

when, for each k, x_k^* is an appropriately chosen point in the interval $x_{k-1} < x < x_k$. Since $F'(x_k^*) = f(x_k^*)$, the last sum is a Riemann sum formed for the function f over the interval $a \leqq x \leqq b$. Since each sum is equal to $F(b) - F(a)$, it follows that the limit (which exists by hypothesis) of these sums must be $F(b) - F(a)$. The limit is the integral in (1) and the theorem is proved. *Remark:* The hypothesis that F is differentiable and $F'(x) = f(x)$ over $a \leqq x \leqq b$ does not imply that f is continuous over $a \leqq x \leqq b$; in fact the last of Problems 3.69 gives examples of functions which are differentiable over the interval $-1 \leqq x \leqq 1$ but have derivatives that are unbounded over this interval. Thus, some discontinuous functions can be derivatives, but the following theorem shows that a function cannot be a derivative unless it (like continuous functions) possesses the intermediate-value property.

Theorem *If F is differentiable over $a \leqq x \leqq b$ and if $F'(a) < q < F'(b)$ or $F'(a) > q > F'(b)$, then there is a number ξ for which $a < \xi < b$ and $F'(\xi) = q$.*

To prove this theorem, let $g(x) = F(x) - q(x - a)$. Then g, like F, must be continuous. Hence $g(x)$ must attain a minimum value at some point ξ for which $a \leqq \xi \leqq b$. Consider the case in which $F'(a) < q < F'(b)$. Since $g'(a) = F'(a) - q < 0$, we see that ξ cannot be a. Since $g'(b) = F'(b) - q > 0$, we see that ξ cannot be b. Hence $a < \xi < b$ and therefore $g'(\xi) = 0$ and $F'(\xi) = q$. In case $F'(a) > q > F'(b)$, $g(x)$ must attain a maximum at a point ξ for which $a < \xi < b$ and $F'(\xi) = q$. This proves the theorem.

19 Prove that if f is continuous over $-\infty < x < \infty$, if $f(x) \to 0$ as $x \to \infty$, and if $f(x) \to 0$ as $x \to -\infty$, then f must have a global maximum or a global minimum but not necessarily both. *Hints:* As in the proof of the Rolle theorem, consider three cases. In case $f(x) > 0$ for at least one x, choose x_0 such that $f(x_0) > 0$. Choose a number H such that $|f(x)| < \frac{1}{2}f(x_0)$ when $|x| \geqq H$. The maximum of $f(x)$ over the interval $|x| \leqq H$ must then be the maximum of $f(x)$ over the infinite interval.

20 Persons who manufacture peanut butter and typewriters and electronic organs have an abiding interest in demand curves. It is supposed that x units of a commodity can be sold when the price is $p(x)$ dollars per unit. The graph of p versus x is the *demand curve.* The nature of the demand curve depends upon the commodity, being relatively flat (or *inelastic*) for false teeth, since people are prone to purchase only those that are required no matter what they cost, and being relatively steep (or *elastic*) for water, since people require only enough to drink but like to wash everything and water their gardens when water is cheap. Economists and others construct and study hypothetical demand curves for pleasure and for business. It is usually supposd that p is a positive decreasing differentiable function of x. When x units are sold at $p(x)$ dollars per unit, the total revenue $R(x)$ is the product of x and $p(x)$. Thus $R(x) = xp(x)$. When x units are sold, the profit $P(x)$ is $R(x) - C(x)$, where $C(x)$ is the total cost of producing and selling the x units. Thus

$$P(x) = R(x) - C(x), \tag{1}$$

this being one way of saying that profit is obtained by subtracting expenses from income. For better or for worse, economists use special terminology in studies of price, revenue, cost, and profit. The numbers $p'(x)$, $R'(x)$, $C'(x)$, and $P'(x)$, these being derivatives at x, are called the *marginal price*, the *marginal revenue*, the *marginal cost*, and the *marginal profit*. This terminology is (or is thought to be) appropriate because if we are producing and selling x units and we know $p(x)$, $R(x)$, $C(x)$, and $P(x)$, then a shift to $x + \Delta x$, $p(x + \Delta x)$, $R(x + \Delta x)$, $C(x + \Delta x)$, and $P(x + \Delta x)$ is "marginal" when Δx is near zero and, for example, the number which $[P(x + \Delta x) - P(x)]/\Delta x$ approximates for marginal shifts is a marginal profit. As is easily imagined, knowledge of functions, limits, and derivatives is helpful when these things are being studied. Differentiating (1) gives the formula

$$P'(x) = R'(x) - C'(x), \tag{2}$$

which says that the marginal profit is equal to the marginal revenue minus the marginal cost. When, as frequently happens, $P(x)$ is a maximum when $P'(x) = 0$ and there is just one x for which $P'(x) = 0$, we obtain the following rule for maximizing profits: choose the x for which the marginal cost is equal to the marginal revenue. When equations of demand curves and cost curves are given, we can determine the x that maximizes profits. Our course in analytic geometry and calculus is considered to be a prerequisite for extensive study of economics because it prepares us to understand definitions, work out formulas, solve problems, and attain over-all comprehension of the subject. In fact, knowledge of the mean-value theorem is not superfluous. The formula

$$P(x + 1) - P(x) = \frac{P(x + 1) - P(x)}{1} = P'(x^*), \tag{3}$$

in which x^* is an appropriate number between x and $x + 1$, can help us understand the antics of elementary books that alternately use $P(x + 1) - P(x)$ and the slope of the graph of P for the marginal profit.

5.6 *Sequences, series, and decimals* Our mathematical foundations always remain quite shaky until we obtain precise information about the possibility of approximating and "representing" numbers by decimals. Moreover, we should have some solid information about this "representing" business. We know what we mean when we say that lawyers represent felons in courts of law, but nevertheless our precious corpus of scientific information is not appreciably augmented when a solemn tutor makes the unexplained statement that "decimals represent numbers."

To attack this and other matters, we must learn about some things that have many applications. A *sequence* $s_1, s_2, s_3, \cdots$ of numbers is an ordered collection of numbers in which there is a first, a second, a third, etcetera. The individual numbers are called *elements* of the sequence; they are not called terms because terms are things that are added, and they are not called factors because factors are things that are multiplied. When $s_1, s_2, s_3, \cdots$ is a given sequence, it may be true (or it may be false) that there is a number L such that s_n is near L whenever n is large. This statement is meaningful. It means that when $s_1, s_2, s_3, \cdots$ is a given sequence, it may be true (or it may be false) that there is a number L such that to each $\epsilon > 0$ there corresponds an integer N such that $|s_n - L| < \epsilon$ whenever $n > N$. In case L exists, we write

$$\lim_{n\to\infty} s_n = L,$$

as in Section 3.3, and we say that the sequence *converges* to L. In case the limit does not exist, we say that the sequence is *nonconvergent* or *divergent.*

As we shall see, the elementary theories of sequences and series are closely related. However, a series is very different from a sequence. A *series* (or simple infinite series) is an array of numbers and plus signs of the form

$$u_1 + u_2 + u_3 + \cdots. \tag{5.61}$$

Because the notion of addition is involved, the numbers $u_1, u_2, u_3, \cdots$ are called *terms* of the series. The terms are not necessarily nonnegative, and it is standard practice to write the series

$$1 + (-\tfrac{1}{2}) + \tfrac{1}{3} + (-\tfrac{1}{4}) + \tfrac{1}{5} + (-\tfrac{1}{6}) + \cdots$$

in the form

$$1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \cdots.$$

Our series $u_1 + u_2 + u_3 + \cdots$ contains so many terms that not even a high-speed electronic computer could "add them all up" during its lifetime. In order to determine a number that can reasonably be called the value of the series, we need a procedure involving more than brute-force addition. While other procedures exist and are useful, the following is the most elementary and best known useful procedure. Let the *sequence* $s_1, s_2, s_3, \cdots$ *of partial sums* be defined by the formulas $s_1 = u_1$, $s_2 = u_1 + u_2$, $s_3 = u_1 + u_2 + u_3$, etcetera, so that

$$s_n = u_1 + u_2 + \cdots + u_n = \sum_{k=1}^{n} u_k \quad (n = 1, 2, 3, \cdots). \tag{5.62}$$

If it happens that this sequence of partial sums converges to s, so that

(5.621) $$\lim_{n\to\infty} s_n = s,$$

then we say that the series *converges* to s, and, leaving the significance of the horrendous operations to be revealed in Problem 6 and Chapter 12, we say that the series has the *sum* s and we write

(5.622) $$s = u_1 + u_2 + u_3 + \cdots \quad \text{or} \quad s = \sum_{k=1}^{\infty} u_k.$$

In case a given series is not convergent, we say that it is divergent. The series

$$1 - 1 + 1 - 1 + 1 - 1 + \cdots$$

is a classic example of a divergent series.

We are now ready to attack decimals. Let $d_1, d_2, d_3, \cdots$ be a sequence each element d_n of which is one of the 10 digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The array

(5.63) $$0.d_1d_2d_3 \cdots,$$

in which the first dot is a decimal point, is then an *infinite decimal.* We confine our attention to decimals of this form; presence of a positive integer before the decimal point causes no difficulties. Just as the left side of the equation

$$0.31690 = \frac{3}{10} + \frac{1}{10^2} + \frac{6}{10^3} + \frac{9}{10^4} + \frac{0}{10^5}$$

is a remarkably efficient way of abbreviating the right side, so also (5.63) is a remarkably efficient way of abbreviating the infinite series

(5.631) $$\frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3} + \cdots.$$

Thus the infinite decimal is an infinite series in disguise.

Theorem 5.64 *Each infinite decimal* $0.d_1d_2d_3 \cdots$ *converges to a real number* s.

If we think it will serve a useful purpose, we can say that the decimal "represents" the number to which it converges. In any case, we write

(5.641) $$s = 0.d_1d_2d_3 \cdots$$

when the decimal converges to s. To prove the theorem, let s_n denote the sum of the first n terms of the series (5.631) so that

$$s_n = \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n}{10^n}$$

and

$$s_n = 0.d_1d_2 \cdots d_n.$$

The set S consisting of the numbers $s_1, s_2, s_3, \cdots$ is then nonempty and has the upper bound 1, since $s_n < 1$ for each n. Therefore, Theorem 5.46 implies that S has a least upper bound which we denote by s. Then $s_n \leqq s$ for each n. To each $\epsilon > 0$ there corresponds an index N such that $s_N > s - \epsilon$, since otherwise $s - \epsilon$ would be an upper bound of S less than s. But the numbers $d_1, d_2, d_3, \cdots$ are all nonnegative, and hence $s - \epsilon < s_n \leqq s$ when $n > N$. Therefore, $\lim_{n\to\infty} s_n = s$ or $s = 0.d_1d_2d_3 \cdots$ and Theorem 5.64 is proved. For future reference, we note that very minor modifications of this proof yield proofs of the following two theorems.

Theorem 5.65 *If the terms of the series* $u_1 + u_2 + u_3 + \cdots$ *are nonnegative and if the sequence of partial sums has an upper bound, then the series is convergent.*

Theorem 5.651 *If a sequence* $s_1, s_2, s_3, \cdots$ *is monotone increasing (that is,* $s_m \leqq s_n$ *when* $m < n$*) and bounded above (that is,* $s_n \leqq M$ *for each* n*) then the sequence is convergent. Similarly, each monotone-decreasing sequence which is bounded below must be convergent.*

In connection with Theorem 5.64, it is often necessary to recognize the awkward fact that two different infinite decimals can converge to the same number. For example,

$$\tfrac{1}{4} = 0.250000 \cdots, \qquad \tfrac{1}{4} = 0.249999 \cdots.$$

This situation occurs, however, only when one of the decimals has only nines from some place onward. In Theorem 5.64 we started with a decimal and found that it converges to a number. The next theorem is different; we start with a number and find a decimal which converges to it.

Theorem 5.66 *If* s *is a number for which* $0 < s < 1$, *then there is a decimal* $0.d_1d_2d_3 \cdots$ *which converges to it.*

Our proof of this theorem involves manipulation similar to the manipulations of Problem 18 of Problems 5.49, where more details are given. Let d_1 be the greatest integer for which $0.d_1 \leqq s$. Then $s - 0.1 < 0.d_1 \leqq s$. Let d_2 be the greatest integer for which $0.d_1d_2 \leqq s$. Then $s - 0.1^2 < 0.d_1d_2 \leqq s$. Let d_3 be the greatest integer for which $0.d_1d_2d_3 \leqq s$. Then $s - 0.1^3 < 0.d_1d_2d_3 \leqq s$. Continuation of this procedure yields a decimal $0.d_1d_2d_3 \cdots$ that converges to s so that

$$s = 0.d_1d_2d_3 \cdots.$$

We conclude this section with a study of geometric series and repeating decimals. When $x \neq 1$, the identity

$$\frac{1 - x^n}{1 - x} = 1 + x + x^2 + \cdots + x^{n-1} \tag{5.67}$$

can be proved either by long division or by multiplying by $1 - x$. When $|x| < 1$, the sequence $|x|, |x|^2, |x|^3, \cdots$ is monotone decreasing and bounded below by 0 and hence must have a limit. If we let L denote this limit, then

$$L = \lim_{n\to\infty} |x|^{n+1} = |x| \lim_{n\to\infty} |x|^n = |x|L,$$

so $(1 - |x|)L = 0$ and hence $L = 0$. Therefore, as we have previously proved in another way,

$$\lim_{n\to\infty} x^n = 0 \qquad (|x| < 1). \tag{5.671}$$

But the right member of (5.67) is the sum of the first n terms of the series in the right member of the formula

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \qquad (|x| < 1). \tag{5.672}$$

Hence, when $|x| < 1$, taking the limits as n becomes infinite of the members of (5.67) gives (5.672). Multiplying the members of (5.67) by a constant a gives the very important formula

$$\frac{a}{1-x} = a + ax + ax^2 + ax^3 + \cdots \qquad (|x| < 1) \tag{5.673}$$

which must be permanently remembered. The series is a *geometric series* with ratio x, the ratio being the factor by which we multiply one term to get the next. The easy way to remember the formula is to remember that, when the absolute value of the ratio is less than 1, *a geometric series converges to the first term divided by* 1 *minus the ratio.*

A *repeating decimal* is one, like

$$3.16952\ 952\ 952 \cdots,$$

in which, from some place onward, the digits involve only periodic repetitions of a collection containing one or more digits. With the aid of (5.673) we can show that each repeating decimal converges to (or is) a rational number, that is, a quotient of two integers. For example, if s is the number to which the decimal displayed above converges, then

$$s = \frac{316}{100} + \frac{1}{100}\left(.952 + \frac{.952}{1000} + \frac{.952}{(1000)^2} + \cdots\right)$$
$$= \frac{316}{100} + \frac{1}{100}\frac{.952}{1 - \frac{1}{1000}} = \frac{316}{100} + \frac{1}{100}\frac{952}{999}.$$

It is presumed that we can add fractions when there is a reason for doing so, and we can see that s is a quotient of integers with denominator 99900.

The most important fact concerning repeating decimals is set forth in the following theorem.

Theorem 5.68 *The (terminating or nonterminating) decimal expansion of each rational number is a repeating decimal.*

Proof of this fact can be based on the ordinary process by which "long division" is used to divide one positive integer, say P, by another, say Q. At each sufficiently advanced stage of the process, we obtain a representation of P/Q of the form

$$\frac{P}{Q} = N + 0.d_1d_2 \cdots d_n + \frac{P_n}{10^nQ}, \tag{5.681}$$

where N is an integer, the d's are digits, and P_n is an integer remainder for which $0 \leqq P_n < Q$. After the long division has progressed past the place where no digits other than zero are "brought down," the remainders and hence also the d's run through cycles that produce the repeating decimal. A cycle begins when a remainder becomes equal to a previous remainder, and this must happen because $0, 1, 2, \cdots, Q-1$ are the only values that remainders can have. Dividing 365 by 7 shows an application of the ideas. The long division process never produces decimal expansions which, from some place onward, consist exclusively of nines, but these expansions are clearly repeating decimals.

The elementary arithmetical consequences of Theorem 5.68 are enormous. We can easily write nonrepeating decimals, examples being

$$0.1234567891011121314151617 \cdots$$

where the positive integers are written in order, and

$$0.1010010001000010000001 \cdots .$$

These decimals converge to real numbers that are not rational and are called irrational (not ratio-nal). This proves existence of irrational numbers. Moreover, we can easily generate the idea that if the digits in

$$0.d_1d_2d_3d_4 \cdots$$

are selected in some random way, then it is highly unlikely (or even almost impossible) that the decimal would be a repeating decimal. This leads us to the idea that "almost all" real numbers are irrational, and there are different ways in which this idea can be made precise.

Problems 5.69

1 Show that if a, b, and c are digits, then

(*a*) $0.aaaa \cdots = \dfrac{a}{9}$ (*b*) $0.ababab \cdots = \dfrac{10a+b}{99}$

(*c*) $b.aaaa \cdots = \dfrac{9b+a}{9}$ (*d*) $c.ababab \cdots = \dfrac{99c+10a+b}{99}$

(*e*) $0.baaa \cdots = \dfrac{9b+a}{90}$ (*f*) $0.cabab \cdots = \dfrac{99c+10a+b}{990}$

2 Write an infinite decimal which converges to an irrational number between 0.43211 and 0.43212.

3 Supposing that 0.31690416 · · · and 0.31690444 · · · converge to irrational numbers, write a rational number that lies between them.

4 Supposing that a and b are different positive numbers, give a procedure by which we can find a rational number x_1 and an irrational number x_2 that lie between a and b.

5 For a long time before the advent of electronic computers, the base 10 of the decimal system reigned supreme and most people thought that other bases had only theoretical interest. Nowadays the base 2, which employs the two *binary bits* 0 and 1 instead of the ten decimal digits 0, 1, 2, · · · , 9, is very important. In the binary system, the left member of the formula

$$(1)\qquad b_mb_{m-1}\cdots b_2b_1b_0 = b_m2^m + b_{m-1}2^{m-1} + \cdots + b_22^2 + b_12 + b_0,$$

in which each bit b_k is 0 or 1, abbreviates the right member. Thus the binary representations of the first few positive integers are

$$(2)\qquad 1, 10, 11, 100, 101, 110, 111, 1000, 1001, \cdots .$$

Similarly,

$$(3)\qquad (.b_{-1}b_{-2}b_{-3}\cdots)_2 = \frac{b_{-1}}{2} + \frac{b_{-2}}{2^2} + \frac{b_{-3}}{2^3} + \frac{b_{-4}}{2^4} + \cdots,$$

where the subscript 2 in the left member informs us that the "point" is not a "decimal point" but is a "binary point" and that each b is a binary bit. One reason for importance of binary bits lies in the fact that one "state" such as "light on" or "switch closed" or "true" can be represented by 1, while the opposite "state" such as "light off" or "switch open" or "false" can be represented by 0. Perhaps without knowing why, we can pick up useful ideas by solving a few simple problems. Show that

(*a*) $(29)_{10} = (11101)_2$ (*b*) $(100)_2 = (4)_{10}$
(*c*) $(100)_{10} = (1100100)_2$ (*d*) $(416)_{10} = (110100000)_2$
(*e*) $(10011)_2 + (10110)_2 = (101001)_2$ (*f*) $(\frac{7}{32})_{10} = (0.00111)_2$
(*g*) $(\frac{1}{3})_{10} = (0.01010101\cdots)_2$

Remark: Many persons with substantial lacks of enthusiasm for adding, subtracting, multiplying, and dividing with decimal digits can find genuine amusement in learning to make these manipulations with binary bits. Scientists need never be bored because of lack of interesting things to do.

6 Inquisitive students may ask why we write

$$(1)\qquad s = u_1 + u_2 + u_3 + \cdots$$

when the series converges to s. The answer lies partly in the fact that it is much easier to write (1) than to write the statement that "s is the number to which the series $u_1 + u_2 + u_3 + \cdots$ converges" and partly in the fact that the *method of convergence* which we have described is the simplest useful method for assigning values to series. There are other methods that are both venerable and useful. One of these is the method which is called the method of Abel (1802–1829) even though it was extensively used by Euler (1707–1783) and was used by Leibniz

(1646–1716) and others before Euler. A given series $u_1 + u_2 + u_3 + \cdots$ is assigned the value V by this method if the series

$$u_1 + u_2 r + u_3 r^2 + \cdots \tag{2}$$

converges when $0 < r < 1$ to the values $f(r)$ of a function for which

$$\lim_{r \to 1-} f(r) = V. \tag{3}$$

Use these ideas to find the Abel value of the series $1 - 1 + 1 - 1 + \cdots$. *Hint:* The series $1 - r + r^2 - r^3 + \cdots$ is a geometric series whose ratio is $-r$, and the series converges to $1/(1 + r)$ when $|r| < 1$ *Remark:* Our mathematical notations would be more sensible but less brief if we were accustomed to writing

$$s = C\{u_1 + u_2 + u_3 + \cdots\} \tag{4}$$

to abbreviate the statement that the series in braces is assigned the value s by the method of convergence and to writing

$$\tfrac{1}{2} = A\{1 - 1 + 1 - 1 + \cdots\} \tag{5}$$

to abbreviate the statement that the series in braces is assigned the value $\frac{1}{2}$ by the method of Abel. This more elaborate notation can show just what we are doing when we adopt the convenient but absurd old idea that a conglomeration of numbers and plus signs "is" a number or "represents" a number if and only if it *converges* to the number. An intelligible theory of series requires a suitable mixture of broad ideas of Euler and narrow ideas usually promoted by elementary books of the nineteenth and twentieth centuries.

7 Each sequence $s_1, s_2, s_3, \cdots$ of numbers determines its sequence $M_1, M_2, M_3, \cdots$ of arithmetic means defined by the formulas

$$M_n = \frac{s_1 + s_2 + \cdots + s_n}{n} = \frac{1}{n} \sum_{k=1}^{n} s_k \quad (n = 1,2,3 \cdots). \tag{1}$$

If

$$\lim_{n \to \infty} s_n = s, \tag{2}$$

so that s_n is near s whenever n is large, we can feel that M_n should also be near s whenever n is large and hence that

$$\lim_{n \to \infty} M_n = s. \tag{3}$$

Prove that (2) implies (3). *Solution:* Let $\epsilon > 0$. Choose an integer N such that $|s_n - s| < \epsilon/2$ whenever $n > N$. Then, when $n > N$,

$$M_n - s = \frac{(s_1 - s) + (s_2 - s) + \cdots + (s_n - s)}{n} = \frac{1}{n} \sum_{k=1}^{n} (s_k - s) \tag{4}$$

and hence

$$|M_n - s| \leqq \frac{1}{n} \sum_{k=1}^{N} |s_k - s| + \frac{1}{n} \sum_{k=N+1}^{n} |s_k - s| \leqq \frac{C}{n} + \frac{1}{n} \sum_{k=N+1}^{n} \frac{\epsilon}{2} < \frac{C}{n} + \frac{\epsilon}{2}, \tag{5}$$

where $C = \sum_{k=1}^{N} |s_k - s|$. If we choose N_1 such that $N_1 > N$ and $C/n < \epsilon/2$ when $n > N$, then we will have

$$(6) \qquad |M_n - s| < \epsilon$$

when $n > N_1$. This proves (3). *Remark:* It often happens that the limit in (3) exists when the limit in (2) does not exist. In case $u_1 + u_2 + u_3 + \cdots$ is a series having partial sums $s_1, s_2, \cdots$ and arithmetic means $M_1, M_2, M_3, \cdots$ such that (3) holds, we can write

$$(7) \qquad s = C_1\{u_1 + u_2 + u_3 + \cdots\}$$

and say that the series is evaluable to s by the *method of arithmetic means* or by the *Cesàro method of order* 1.

8 Supposing that n is a positive integer, sketch a graph of the function f_n for which $f_n(x) = n^2x$ when $|x| \leqq 1/n$ and $f_n(x) = 1/x$ when $|x| > 1/n$. Show that f_n is continuous over E_1. Show that

$$\lim_{n\to\infty} f_n(x) = g(x),$$

where $g(0) = 0$ and $g(x) = 1/x$ when $x \neq 0$. *Hint:* Consider separately the cases in which $x = 0$, $x > 0$, and $x < 0$.

9 Using the notation of the preceding problem, let

$$\begin{aligned} u_1(x) &= f_1(x) \\ u_2(x) &= f_2(x) - f_1(x) \\ u_3(x) &= f_3(x) - f_2(x) \\ u_4(x) &= f_4(x) - f_3(x), \end{aligned}$$

etcetera, so that $u_k(x) = f_k(x) - f_{k-1}(x)$ when $k = 2, 3, 4, \cdots$. Show that each function u_n is continuous over E_1 and that

$$\sum_{k=1}^{\infty} u_k(x) = g(x).$$

Remark: It is sometimes necessary to be sophisticated enough to know that a series of continuous functions may converge to a discontinuous function. Moreover, we should be tall enough to peer over the wall of our garden and observe that a series $u_1(x) + u_2(x) + \cdots$ of functions having partial sums $f_1(x)$, $f_2(x)$, $\cdots$ is said to *converge uniformly* over a set E to $f(x)$ if to each positive number ϵ there corresponds an integer N such that $|f_n(x) - f(x)| < \epsilon$ whenever $n \geqq N$ and x is in E. The following theorem is proved in advanced calculus. *If a series $u_1(x) + u_2(x) + \cdots$ of continuous functions converges uniformly over E to $f(x)$, then f must be continuous over E.*

10 Starting with positive numbers a_1 and b_1 for which $a_1 < b_1$, let sequences $a_1, a_2, a_3, \cdots$ and $b_1, b_2, b_3, \cdots$ be defined recursively by the formulas

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

Show that, for each $n = 1, 2, 3, \cdots$,

$$a_n < a_{n+1} < b_{n+1} < b_n. \tag{2}$$

Tell why there must exist numbers L_1 and L_2 such that

$$\lim_{n\to\infty} a_n = L_1, \qquad \lim_{n\to\infty} b_n = L_2. \tag{3}$$

Show also that

$$0 < b_{n+1} - a_{n+1} < \frac{a_n + b_n}{2} - a_n = \frac{b_n - a_n}{2} \tag{4}$$

and hence

$$0 \leqq L_2 - L_1 \leqq \tfrac{1}{2}(L_2 - L_1), \tag{5}$$

so $L_2 = L_1$. *Remark:* The common value of the two limits in (3) has an impressive name; it is the arithmetico-geometric mean of the two given numbers a_1 and b_1.

11 This problem, which is in some respects the most significant problem in this chapter, would be much too difficult if it were not prefaced by a rather elaborate story. We make the reasonable assumption that Mr. C., a particular carpenter, never heard of the Dedekind axiom 5.43, and that his ideas about the real-number system are incomplete. Next we make the reasonable assumption that the class R^* (read R star) of numbers that Mr. C. knows about is the class of rational numbers which he may call "whole numbers and fractions." This class R^* is, for many purposes, a thoroughly useful class of numbers. If x and y belong to R^*, so do $x + y$, $x - y$, xy, and also x/y, provided $y \neq 0$. While we may be somewhat surprised by the fact, it is nevertheless true that Mr. C. could define graphs, limits, derivatives, indefinite integrals, Riemann integrals, and many other things exactly as we defined them. Mr. C. could show, exactly as we did, that if $f(x) = x^2$, then $f'(x) = 2x$. There would be many respects in which his analytic geometry and calculus would be thoroughly satisfactory. He would say, exactly as we did, that f is continuous at x_0 if to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that $|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_0| < \delta$, but of course only rational numbers appear in his work. Mr. C. would be totally unaware of the existence of irrational numbers, but we could nevertheless select an irrational number ξ for which $0 < \xi < 1$ and put Mr. C. to work studying the function f for which

$$\begin{cases} f(x) = -1 & (0 \leqq x < \xi) \\ f(x) = 1 & (\xi < x \leqq 1). \end{cases} \tag{1}$$

Mr. C. would discover that f is defined for each x in R^* for which $0 \leqq x \leqq 1$, and hence he would say that it is defined over the interval $0 \leqq x \leqq 1$. He could prove that f is continuous at each x in R^* for which $0 \leqq x \leqq 1$. He would therefore say that it is continuous over the interval $0 \leqq x \leqq 1$. He could prove that $f'(x) = 0$ for each x in R^* for which $0 < x < 1$. So far there is nothing wrong, but there will be something wrong if Mr. C. tries to tell us that $f'(x) = 0$ when $0 < x < 1$ and *hence* "it is obvious" or "it can be shown" that there must be a constant k such that $f(x) = k$ when $0 < x < 1$. In fact, a look at the formulas (1) defining f shows that there is no constant k such that $f(x) = k$ when

$0 < x < 1$. It is now time to seek the moral of this story. If we are not sure whether the set of numbers we use in our analytic geometry and calculus is the complete set of real numbers for which the Dedekind postulate is valid, then we cannot be sure about the validity of the ideas that we need to enable us to do our chores. It is, therefore, not enough to know the axioms usually given in one way or another in elementary arithmetic and algebra and "finite mathematics." We need, in addition, the Dedekind axiom or an equivalent axiom which guarantees that we are using the complete class of real numbers in our work. We now come to the problem. Tell whether it is necessary to use the Dedekind axiom (or, what amounts to the same thing, to use consequences of the Dedekind axiom or an equivalent axiom) in order to (*a*) prove the Rolle theorem 5.51, (*b*) prove the intermediate-value theorem 5.48, (*c*) define the area of a rectangle to be the product of its dimensions, (*d*) define the derivative of a given function f, (*e*) prove existence of $\int_1^2 (1/x)\,dx$.

5.7 *Darboux sums and Riemann integrals* This section can be omitted from this course without damaging understanding of the rest of the book. There can, however, be no doubt that students with serious interest in pure mathematics should master it and that everyone else should read it. The section gives substantial information about a standard way of attacking matters relating to existence of Riemann integrals. Let f be defined and bounded over an interval $a \leqq x \leqq b$ so that, for some constants m and M, we have

$$m \leqq f(x) \leqq M \qquad (a \leqq x \leqq b). \tag{5.71}$$

As in our definition of Riemann sums, let P be a partition such as the one shown in Figure 5.711 and, for each k, let x_k^* be selected such that $x_{k-1} \leqq$

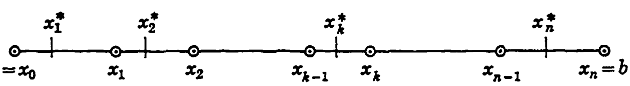


Figure 5.711

$x_k^* \leqq x_k$. Let $\Delta x_k = x_k - x_{k-1}$. For each $k = 1, 2, \cdots, n$, let

$$m_k = \underset{x_{k-1} \leqq x \leqq x_k}{\text{g.l.b.}}\ f(x), \qquad M_k = \underset{x_{k-1} \leqq x \leqq x_k}{\text{l.u.b.}}\ f(x) \tag{5.712}$$

so that m_k and M_k are respectively the greatest lower bound and the least upper bound of f over the interval $x_{k-1} \leqq x \leqq x_k$. The numbers UDS($P$) and LDS($P$) defined by

$$\text{LDS}(P) = \sum_{k=1}^{n} m_k\,\Delta x_k, \qquad \text{UDS}(P) = \sum_{k=1}^{n} M_k\,\Delta x_k \tag{5.72}$$

are called the *lower* and *upper Darboux* (1842–1917) *sums* determined by

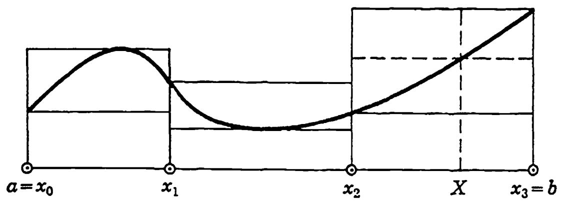


Figure 5.721

P. Figure 5.721 is available for inspection. For each choice of the points x_k^* we have $m_k \leqq f(x_k^*) \leqq M_k$ and hence

$$\text{LDS}(P) \leqq \sum_{k=1}^{n} f(x_k^*)\, \Delta x_k \leqq \text{UDS}(P). \tag{5.722}$$

Therefore, for a given partition P, the different Riemann sums that can be formed by making different choices of the points x_k^* are all sandwiched between the lower and upper Darboux sums. Information about Riemann sums can therefore be gleaned from information about Darboux sums.

The first step in our study of Darboux sums may seem to be a very modest one. Let P be a given partition, and let P' be a simple extension of P. By this we mean that P' is exactly the same as P except that P' contains one additional partition point, say X, which lies between two of the partition points of P, say $x_2 < X < x_3$. The inequality

$$\begin{aligned}&[\operatorname*{l.u.b.}_{x_2 \leqq x \leqq X} f(x)](X - x_2) + [\operatorname*{l.u.b.}_{X \leqq x \leqq x_3} f(x)](x_3 - X)\\ &\qquad \leqq [\operatorname*{l.u.b.}_{x_2 \leqq x \leqq x_3} f(x)](X - x_2) + [\operatorname*{l.u.b.}_{x_2 \leqq x \leqq x_3} f(x)](x_3 - X) \leqq M_3\, \Delta x_3\end{aligned}$$

implies that $\text{UDS}(P') \leqq \text{UDS}(P)$. Figure 5.721 is not needed in the proof of the inequality but may nevertheless be helpful. Consideration of simple extensions of simple extensions of P leads to the conclusion that if P' is any extension of P (so that P' contains all of the partition points of P and perhaps also some additional ones), then $\text{UDS}(P') \leqq \text{UDS}(P)$. An analogous argument, in which greatest lower bounds appear and the inequality signs are reversed, shows that if P' is an extension of P, then $\text{LDS}(P') \geqq \text{LDS}(P)$. Suppose now that P_1 and P_2 are two given partitions, and let P_3 be an extension of both P_1 and P_2. Then

$$\text{LDS}(P_1) \leqq \text{LDS}(P_3) \leqq \text{UDS}(P_3) \leqq \text{UDS}(P_2) \tag{5.723}$$

and hence

$$\text{LDS}(P_1) \leqq \text{UDS}(P_2). \tag{5.724}$$

This is a key result of the theory. Let the symbols in

$$L = \underline{\int_a^b} f(x)\, dx, \qquad U = \overline{\int_a^b} f(x)\, dx \tag{5.73}$$

denote, respectively, the least upper bound of all lower Darboux sums and the greatest lower bound of all upper Darboux sums. These numbers are, respectively, the *lower* and *upper Darboux integrals* of f over the interval $a \leqq x \leqq b$. It follows from (5.73) that, for each partition P_2, $L \leqq \text{UDS}(P_2)$, and it follows in turn from this that $L \leqq U$. Thus

$$\text{LDS}(P) \leqq L \leqq U \leqq \text{UDS}(P) \tag{5.731}$$

for each partition P.

There are bounded functions f for which $L < U$. For example, let $a = 0$, let $b = 1$, and let $f(x) = 0$ when x is irrational and $f(x) = 1$ when x is rational. The $\text{LDS}(P) = 0$ and $\text{UDS}(P) = 1$ for each P and therefore $L = 0$ and $U = 1$.

It can be proved that $\text{LDS}(P)$ is near L and $\text{UDS}(P)$ is near U whenever $|P|$ (the norm of P) is small. This result, which is sometimes called the Darboux theorem, means that to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that

$$|\text{LDS}(P) - L| < \epsilon, \qquad |\text{UDS}(P) - U| < \epsilon \tag{5.74}$$

whenever $|P| < \delta$. This and (5.731) imply that, when $|P| < \delta$, the numbers $\text{LDS}(P)$ and $\text{UDS}(P)$ are respectively located in the left and right intervals of Figure 5.741 when $L < U$ and of Figure 5.742 when

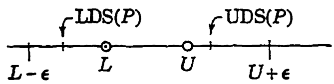


Figure 5.741

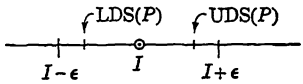


Figure 5.742

$L = U = I$. Consider first the case in which $L < U$. Since each Darboux sum can be approximated as closely as we please by a Riemann sum having the same partition points, it follows that there exist Riemann sums with norm $|P| < \delta$ which differ from L by less than ϵ and that there also exist Riemann sums with norm $|P| < \delta$ which differ from U by less than ϵ. It follows that if $L < U$, then f cannot be Riemann integrable over $a \leqq x \leqq b$.

Consider next the case in which $L = U = I$. In this case

$$I - \epsilon < \text{LDS}(P) \leqq \text{RS}(P) \leqq \text{UDS}(P) < I + \epsilon \tag{5.743}$$

whenever $\text{RS}(P)$ is a Riemann sum formed for a partition P for which $|P| < \delta$. Therefore,

$$\int_a^b f(x)\,dx = I, \tag{5.744}$$

the integral being a Riemann integral.

All this gives the following theorem which involves the numbers L and U defined in the sentence containing (5.73).

Theorem 5.75 *If f is defined and bounded over $a \leqq x \leqq b$, then f is Riemann integrable over $a \leqq x \leqq b$ if and only if $L = U$. Moreover,* (5.744) *holds when $L = U = I$.*

This theorem and (5.731) imply the following useful theorem.

Theorem 5.751 *A function f is Riemann integrable over $a \leqq x \leqq b$ if and only if to each $\epsilon > 0$ there corresponds a partition P such that*

$$\text{UDS}(P) - \text{LDS}(P) \leqq \epsilon. \tag{5.752}$$

The above story provides ideas and results that are used in proofs of the fundamental theorem (Theorem 4.26) on existence of Riemann integrals. We shall use Theorem 5.751 to prove some less pretentious theorems.

Theorem 5.76 *If f is defined and monotone increasing (or monotone decreasing) over $a \leqq x \leqq b$, then the Riemann integral $\int_a^b f(x)\,dx$ exists.*

Let $\epsilon > 0$. Suppose first that f is monotone increasing so that $f(x') \leqq f(x'')$ when $a \leqq x' < x'' < b$. Let P be a partition of $a \leqq x \leqq b$ with partition points $x_0, x_1, \cdots, x_n$ as in Figure 5.711. Then

$$\begin{aligned}\text{UDS}(P) - \text{LDS}(P) &= \sum_{k=1}^{n} [\underset{x_{k-1} \leqq x \leqq x_k}{\text{l.u.b.}} f(x) - \underset{x_{k-1} \leqq x \leqq x_k}{\text{g.l.b.}} f(x)]\,\Delta x_k \\ &= \sum_{k=1}^{n} [f(x_k) - f(x_{k-1})]\,\Delta x_k \leqq \sum_{k=1}^{n} [f(x_k) - f(x_{k-1})]|P| \\ &= [f(b) - f(a)]|P| < \epsilon\end{aligned} \tag{5.761}$$

provided $|P|$ is sufficiently small. This and Theorem 5.751 establish the result for the case in which f is monotone increasing. In case f is monotone decreasing, the proof is exactly the same except that $[f(x_k) - f(x_{k-1})]$ is replaced by $[f(x_{k-1}) - f(x_k)]$ and $[f(b) - f(a)]$ is replaced by $[f(a) - f(b)]$.

It is easy to extend Theorem 5.76 to obtain a better theorem. A function f is said to be bounded and *piecewise monotone* over the closed interval $a \leqq x \leqq b$ if there is a constant M for which $|f(x)| \leqq M$ when $a \leqq x \leqq b$ and if there is a partition P of the interval $a \leqq x \leqq b$ such that, whenever x_{k-1} and x_k are two consecutive partition points, f is monotone (maybe monotone increasing, maybe monotone decreasing) over the open interval $x_{k-1} < x < x_k$.

Theorem 5.762 *If f is bounded and piecewise monotone over $a \leqq x \leqq b$, then the Riemann integral $\int_a^b f(x)\,dx$ exists.*†

† It is sometimes said that this theorem is a poor-man's version of a stronger theorem which says that f is integrable over $a \leqq x \leqq b$ if f has bounded variation over $a \leqq x \leqq b$. Problem 10 at the end of this section provides opportunities to rise above poverty.

Similarly, we apply the fundamental Theorem 5.731 to prove Riemann integrability of continuous functions and piecewise continuous functions.

***Theorem* 5.77** *If f is continuous over $a \leqq x \leqq b$, then the Riemann integral $\int_a^b f(x)\,dx$ exists.*

To prove this, let $\epsilon > 0$. Theorem 5.58 then enables us to choose a positive number δ such that $|f(x_2) - f(x_1)| < \epsilon/(b-a)$ whenever $a \leqq x_1 \leqq b$, $a \leqq x_2 \leqq b$, and $|x_2 - x_1| < \delta$. Let P be a partition of the interval $a \leqq x \leqq b$ for which $|P| < \delta$. Then, with the notation of (5.712) and (5.72), we have $M_k - m_k \leqq \epsilon/(b-a)$ and hence

$$\text{(5.771)} \qquad \mathrm{UDS}(P) - \mathrm{LDS}(P) \leqq \sum_{k=1}^{n} \frac{\epsilon}{b-a}\,\Delta x_k = \epsilon.$$

The required conclusion then follows from Theorem 5.751.

A function f is said to be *piecewise continuous* over the closed interval $a \leqq x \leqq b$ if it is defined over $a \leqq x \leqq b$ and if there is a partition P of the interval $a \leqq x \leqq b$ such that, whenever x_{k-1} and x_k are two consecutive partition points, f is continuous over the open interval $x_{k-1} < x < x_k$ and, in addition, the unilateral limits

$$\lim_{x \to x_{k-1}^+} f(x), \qquad \lim_{x \to x_k^-} f(x)$$

both exist. On account of the fact that functions that are piecewise continuous over a closed interval must be bounded, it is not difficult to use Theorem 5.77 to prove the following more general theorem.

***Theorem* 5.772** *If f is piecewise continuous over $a \leqq x \leqq b$, then the Riemann integral exists.*

Finally, we use ideas and notation of this section to prove the following theorem.

***Theorem* 5.78** *If f is Riemann integrable over the interval $a \leqq x \leqq b$ and $f(x) \geqq 0$ when $a \leqq x \leqq b$, then the set S of points (x,y) for which $a \leqq x \leqq b$, $0 \leqq y \leqq f(x)$ possesses an area $|S|$ and $|S| = \int_a^b f(x)\,dx$.*

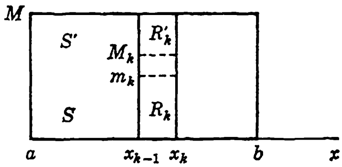


Figure 5.781

The proof depends upon the fundamental definition of area given in Definition 4.44. Choose a constant M such that $f(x) \leqq M - 1$ and observe that S is a subset of the large rectangle R of Figure 5.781. Let $\epsilon > 0$ and let $0 < \epsilon' < \epsilon$. Let the number $|S|$ be defined by the formula $|S| = \int_a^b f(x)\,dx$. To prove the theorem, we shall show that $|S|$ is in fact the area of S. Let P be a partition for which

$$\text{(5.782)} \qquad \sum_{k=1}^{n} f(^*_k)\,\Delta x_k \leqq |S| + \epsilon', \qquad \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k \geqq S - \epsilon'$$

whenever $x_{k-1} \leqq x_k^* \leqq x_k$ for each k. Defining M_k and m_k by (5.712), we conclude that

$$\sum_{k=1}^{n} M_k \, \Delta x_k \leqq |S| + \epsilon', \qquad \sum_{k=1}^{n} m_k \, \Delta x_k \geqq |S| - \epsilon'. \tag{5.783}$$

Let R_k and R_k' be, for each k, the rectangles (meaning rectangular regions) consisting of points (x,y) for which $x_{k-1} \leqq x \leqq x_k$, $0 \leqq y \leqq M_k$ and $x_{k-1} \leqq x < x_k$, $m_k \leqq y \leqq M$. The two formulas (5.783) then give

$$\sum_{k=1}^{n} |R_k| < |S| + \epsilon, \qquad \sum_{k=1}^{n} |R_k'| < |R| - |S| + \epsilon. \tag{5.784}$$

If $P(x,y)$ lies in S, then there is at least one k for which $x_{k-1} \leqq x \leqq x_k$ and hence $0 \leqq y \leqq f(x) \leqq M_k$, so P is a point of at least one rectangle R_k. Similarly, if $P(x,y)$ is a point of the set S' consisting of the points in R but not in S, then there is at least one k for which $x_{k-1} \leqq x \leqq x_k$ and hence $m_k \leqq f(x) \leqq M$, so P is a point of at least one rectangle R_k'. It is therefore a consequence of Definition 4.4 that the set S does possess an area and that its area is $|S|$. This completes the proof of Theorem 5.78.

Problems 5.79

1 Sketch a dozen graphs that look like graphs of functions f that are bounded and piecewise monotone over the interval $0 \leqq x \leqq 1$. Be sure to include graphs of some discontinuous functions and of some nonmonotone functions.

2 Sketch a figure which is like Figure 5.721 except that the partition P contains 10 or 20 partition points that are roughly equally spaced. Then look at your figure and see how LDS(P) and UDS(P) are related.

3 Sketch a figure which shows the geometric meanings in the statement and proof of Theorem 5.76.

4 As was remarked, Archimedes (287–212 B.C.) knew about some special Riemann sums, and this matter may be worthy of brief consideration here. When f is defined over *rational* values of x in the interval $0 \leqq x \leqq 1$, we can make a partition of the interval $0 \leqq x \leqq 1$ into n equal subintervals of length $1/n$ by partition points x_k for which $x_k = k/n$ and form the special Riemann sum

$$A_n = \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\frac{1}{n}$$

which we can call an Archimedes sum. Without implying that Archimedes used modern terminology involving sums, limits, and integrals, we can recognize that there is historical evidence that we are merely putting ideas of Archimedes into modern terminology when we say that f is Archimedes integrable over the interval $0 \leqq x \leqq 1$ and that f has the Archimedes integral I if $A_n \to I$ as $n \to \infty$. Now comes the problem. Supposing that $f(x) = 0$ when x is irrational and $f(x) = 1$ when x is rational, show that if the symbol

$$\int_0^1 f(x)\, dx$$

represents an Archimedes integral, then the integral exists and has the value 1, but that if the symbol represents a Riemann integral, then the integral does not exist.

5 For each $n = 3, 4, 5, \cdots$ the broken line joining in order the points $(0,0)$ $(1/n,n)$, $(2/n,0)$, $(1,0)$ is the graph of a function f_n defined over the interval $0 \leqq x \leqq 1$. Prove that

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 1, \qquad \int_0^1 [\lim_{n\to\infty} f_n(x)]\,dx = 0. \tag{1}$$

Hint: Observe that $f_n(0) = 0$ for each $n \geqq 3$ and hence $\lim_{n\to\infty} f_n(0) = 0$. If $0 < x \leqq 1$, then $f_n(x) = 0$ when $2/n < x$, and hence when $n > 2/x$, so again $\lim_{n\to\infty} f_n(x) = 0$. *Remark:* Persons who push very far into the theory of Fourier series learn that if

$$F_n(x) = \frac{2}{n\pi}\left(\frac{\sin nx}{\sin x}\right)^2 \qquad (n = 1, 2, 3, \cdots) \tag{2}$$

then

$$\lim_{n\to\infty} \int_0^{\pi/2} F_n(x)\,dx = 1, \qquad \int_0^{\pi/2} [\lim_{n\to\infty} F_n(x)]\,dx = 0. \tag{3}$$

While consideration of the matter can be postponed, our course in analytic geometry and calculus should be leading us toward abilities to sketch a graph of F_n and to appreciate the fact that the two formulas in (3) can be valid.

6 This problem invites investment of time in a speculative venture. It was proved in Section 4.3 that the formula

$$\int_a^b F'(x)\,dx = F(x)\Big]_a^b = F(b) - F(a) \tag{1}$$

is valid whenever F has a continuous derivative over the interval $a \leqq x \leqq b$. It was proved in Problem 18 of Problems 5.59 that (1) is valid whenever F' exists and is Riemann integrable over the interval $a \leqq x \leqq b$. Even though nobody requires us to learn everything, we may sometime be benefited by knowledge that there is a function F for which (i) $F'(x)$ exists when $-1 \leqq x \leqq 1$ and (ii) F' is not continuous but is Riemann integrable over $-1 \leqq x \leqq 1$. Let F be defined by the formulas $F(0) = 0$ and

$$F(x) = -x^2 \cos\frac{1}{x} + \int_0^x 2t \cos\frac{1}{t}\,dt \qquad (x \neq 0). \tag{2}$$

That $F'(0) = 0$ can be proved by using the inequality

$$\left|\frac{F(x) - F(0)}{x - 0}\right| \leqq \left|x \cos\frac{1}{x}\right| + \left|\frac{1}{x}\int_0^x |2t|\,dt\right| \leqq 2|x| \tag{3}$$

and the sandwich theorem. When $x \neq 0$, differentiating (2) gives

$$F'(x) = \sin\frac{1}{x} \qquad (x \neq 0). \tag{4}$$

Thus $F'(x)$ exists when $-1 \leqq x \leqq 1$. As x approaches 0, $F'(x)$ oscillates between -1 and 1 and does not have a limit, so F' is not continuous at the place where $x = 0$. However, F' is Riemann integrable over the interval $-1 \leqq x \leqq 1$ because $F'(x)$ exists and is bounded over the interval and is continuous except at one place. Thus our function F has the required properties. We could have used the formula

$$F(x) = \int_0^x \sin\frac{1}{t}\,dt \tag{5}$$

instead of (2) to define F. It would then have been slightly easier to obtain (4) but would not have been so easy to show that $F'(0) = 0$. There is a reason why no simpler example can be given. Derivatives must have the intermediate-value property, and no discontinuous function having the intermediate-value property is simpler than the function ϕ for which $\phi(0) = 0$ and $\phi(x) = \sin(1/x)$ when $x \neq 0$.

7 If the unique individual that some textbooks like to call "the student" is unable to prove that each polynomial is bounded and piecewise monotone over each interval $a \leqq x \leqq b$, there are only three possible places to place the blame. Is it the student? Is it the textbook? Is it the problem?

8 We have, at one time and another, seen examples of faulty applications of the noble but frequently invalid premise that a thing T must be an element of a set S if T is the limit of a sequence of elements of S. One old example involves the "idea" that a circle must be a polygon because it is the limit of polygons. Another old example involves the "idea" that a Riemann integral must be the sum of infinitely many things because it is the limit of sums. Should we swallow the "idea" that an irrational number must be a rational number because it is the limit of rational numbers? *Ans.:* No.

9 We have the possibility of extending our intellectual horizons by investing a few minutes or a few years in study of algebras which differ from the algebra of real numbers. The algebra of rational functions invites us to consolidate old ideas and capture new ones. When $a_0, a_1, \cdots, a_m$ and $b_0, b_1, \cdots, b_n$ are constants for which the b's are not all zero, the two polynomials P and Q for which

$$P(x) = a_0 + a_1x + \cdots + a_mx^m, \qquad Q(x) = b_0 + b_1x + \cdots + b_nx^n$$

determine the rational function f for which $f(x) = P(x)/Q(x)$ for those values of x for which $Q(x) \neq 0$. The sum $f + g$ of two rational functions is the rational function h for which $h(x) = f(x) + g(x)$ for each x for which the sum is defined. If c is a constant and f is a rational function, then cf is the rational function having values $cf(x)$. If f and g are rational functions, then fg is the rational function having values $f(x)g(x)$ and [unless $g(x) = 0$ for each x] f/g is the rational function having values $f(x)/g(x)$ when $g(x) \neq 0$. Textbooks in modern algebra call attention to many respects in which the algebra of rational functions is like the algebra of real numbers. Terminologies involving rings, fields, and groups facilitate discussions of these matters. Nontrivial interest in the algebra of rational functions starts to develop when order relations are introduced in a particular special way. We say that $f < g$ and $g > f$ if there is a number x_0 such that $f(x) < g(x)$ and

$g(x) > f(x)$ for each x for which $x > x_0$. Progress with the theory depends upon the basic fact that if f and g are rational functions, then $f(x) = g(x)$ for each x or there is a number x_0 such that $f(x) < g(x)$ when $x > x_0$ or there is a number x_0 such that $f(x) > g(x)$ when $x > x_0$. This basic fact depends upon the fact that if h is a rational function, then either $h(x) = 0$ for each x or there is a number x_0 such that $h(x)$ is continuous and positive or continuous and negative when $x > x_0$. It follows that if f and g are rational functions, then one and only one of the three relations $f < g$, $f = g$, $f > g$ is valid. Thus the set of rational functions is, like the set of real numbers, now an ordered field. Let f_0, the *zero function*, be the rational function for which $f_0(x) = 0$ for each x. Our algebra of rational functions is said to have the *Archimedes property* (or to be Archimedian) if to each pair of functions f and g for which $f > f_0$ and $g > f_0$ there corresponds an integer n for which $nf > g$. This is of interest to us because we have proved, with the aid of the Dedekind axiom, that the algebra of real numbers is Archimedian. It could be presumed that the algebra of rational functions is so much like the algebra of real numbers that the algebra of rational functions must be Archimedian. However, the presumption is false, the algebra of rational functions is not Archimedian. To prove this, let f and g be the rational functions for which $f(x) = x$ and $g(x) = x^2$. Careful applications of our definitions then imply that $f > f_0$, $g > f_0$, and $nf < g$ for each integer n. Thus our algebra of rational functions is not Archimedian. When all matters which we have discussed are thoroughly understood, it becomes clear that the Archimedian property of the algebra of real numbers is not a consequence of those properties of real numbers that are ordinarily stated in elementary arithmetic and algebra. Algebra books that give adequate treatments of matters relating to order relations, bounds, limits, Dedekind partitions, and Archimedes properties are said to be *modern*. We have seen some of the reasons why knowledge of modern algebra is considered to be an essential part of a mathematical education.

10 While consideration of the matter is usually reserved for more advanced courses, we have enough equipment to understand, and perhaps even prove, basic facts involving functions of bounded variation. Let $f(t)$ be defined over $a \leqq t \leqq b$ and let $a < x \leqq b$. Supposing f such that $T(x)$ exists (is finite) we define numbers $T(x)$, $P(x)$, and $N(x)$ by the formulas

$$\text{l.u.b.} \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| = T(x) \tag{1}$$

$$\text{l.u.b.} \frac{1}{2} \sum_{k=1}^{n} \{|f(t_k) - f(t_{k-1})| + [f(t_k) - f(t_{k-1})]\} = P(x) \tag{2}$$

$$\text{l.u.b.} \frac{1}{2} \sum_{k=1}^{n} \{|f(t_k) - f(t_{k-1})| - [f(t_k) - f(t_{k-1})]\} = N(x). \tag{3}$$

In each case, the least upper bound is the least upper bound of sums obtained for partitions P of the interval $a \leqq t \leqq x$. The function f is said to have *bounded variation* (the term finite variation would be better) over the interval $a \leqq t \leqq b$. Let $T(a) = P(a) = N(a) = 0$. The numbers $T(x)$, $P(x)$, and $N(x)$ are, respectively, the *total variation*, the *positive variation*, and the *negative variation* of f over the interval $a \leqq t \leqq x$. Prove that $T(x)$, $P(x)$, $N(x)$ are all nonnegative and monotone increasing over $a \leqq x \leqq b$. Prove that

$$T(x) = 2P(x) - [f(x) - f(a)] \tag{4}$$
$$T(x) = 2N(x) + [f(x) - f(a)] \tag{5}$$
$$T(x) = P(x) + N(x) \tag{6}$$
$$f(x) = f(a) + P(x) - N(x) \tag{7}$$
$$f(x) = [x + f(a) + P(x)] - [x + N(x)] \tag{8}$$

when $a \leqq x \leqq b$. Use our information to prove that if $f(x)$ has bounded variation over $a \leqq x \leqq b$ [which means that $T(b)$ is finite], then $f(x)$ is the difference of two increasing functions. Prove that if $f(x)$ is the difference of two increasing functions over $a \leqq x \leqq b$, then $f(x)$ has bounded variation over $a \leqq x \leqq b$. *Hint:* To obtain (4), note that (2) contains a telescopic sum and put (2) in the form

$$\text{l.u.b.} \left\{\frac{1}{2} \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| + \frac{f(x) - f(a)}{2}\right\} = P(x). \tag{9}$$

Remark: Our results and Theorem 5.76 imply that the Riemann integral $\int_a^b f(x)\,dx$ exists if f has bounded variation over $a \leqq x \leqq b$. Moreover, we now have enough information to appreciate the most important theorem in the theory of Riemann-Stieltjes integrals; see Problem 11 of Problems 4.89. The theorem says that

$$\int_a^b f(x)\,dg(x) \tag{10}$$

exists if f is continuous and g has bounded variation over $a \leqq x \leqq b$. Methods of this section provide proof for the case in which f is continuous and g is increasing, and the general result is then obtained by expressing g as the difference of increasing functions. A much more difficult theorem says that if g is such that (10) exists whenever f is continuous over $a \leqq x \leqq b$, then g must have bounded variation over $a \leqq x \leqq b$.

6 Cones and conics

6.1 Parabolas Before plunging into the general aspects of this chapter, we obtain more information about the parabolas that were introduced in Section 1.4. Being realistic, we face some facts. We remember that, for some strange reason, the graph of $y = kx^2$ is, when $k > 0$, a parabola, but details involving the focus and directrix of this parabola may have been quite thoroughly forgotten. We try to recall, and henceforth remember, that the parabola has a focus F and a directrix as in Figure 6.11 and that the parabola is the set of points $P(x,y)$ for which $|\overrightarrow{FP}| = |\overrightarrow{DP}|$. We have forgotten how the coordinates of F and the equation of the directrix are related to k, and we may forget again, so we should know how to discover the facts. To put a little variety into our lives, we use the symbol "?" to represent the unknown distance from the origin to F and from the origin to the directrix. Now we make the key observation. The points on the horizontal line through F all lie at distance (2?) from the directrix. Hence the point (2?, ?) which lies (2?) units to the right of F must lie on the parabola. The coordinates of this point must therefore satisfy the equation of the parabola, so $? = k(2?)^2$ and $? = 1/4k$. The coordinates of F are therefore $(0, 1/4k)$, and the equation of the directrix is $y = -1/4k$. The square of Figure 6 11 having a vertex

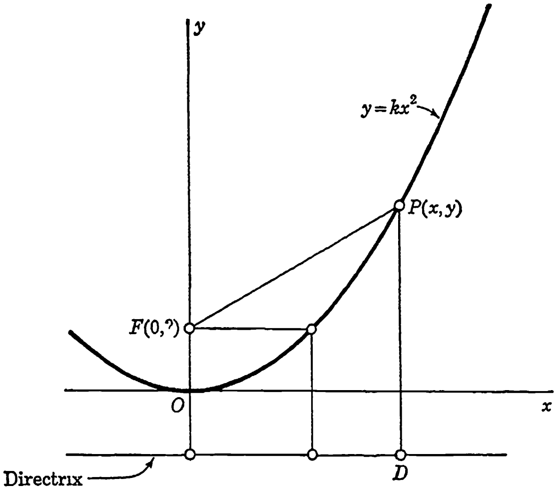


Figure 6.11

at F and two vertices on the directrix is called a *focal square* of the parabola. Another focal square lies to the left of the one in the figure. A figure which shows a parabola together with its focus and directrix is imperfect unless the parabola contains a vertex of each focal square.

The y axis, being an axis of symmetry and the only one, is called the *axis* of the parabola. The point in which the parabola intersects its axis is called the *vertex* of the parabola. More definitions will appear in the problems. While parabolas have important applications in which foci (plural of focus) and directrices (plural of directrix) never appear, most of the problems involve situations in which they do appear.

Since preliminary ideas can be very valuable, we look briefly at Figure 6.12. The figure gives six views of the intersection of a cone and a plane

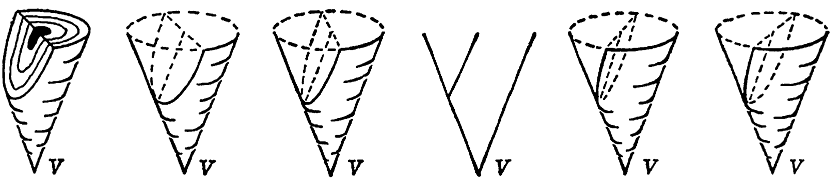


Figure 6.12

The cone is a right circular conical surface a part of which resembles a conical paper cup or ice-cream cone. The vertex V and the axis of the cone are in the plane of the paper. The intersecting plane is parallel to a line on the cone. The intersection is a curve of which a part (the solid part) lies on the front half of the cone and a part (the dotted part) lies on the back half of the cone. Our present requirement is exceedingly modest. All we are required to do is grasp the idea that the curve looks like a parabola. Our solid information about this matter will come in the next section. Meanwhile, persons with artistic flairs can find useful entertainment in sketching sections of cones made by planes not parallel to lines on the cones.

Problems 6.19

1 Sketch a graph showing the parabola whose equation is $y = x^2$ together with the focus and directrix of the parabola. Draw the focal squares and make any repairs that may be necessary to make the parabola contain corners of the focal squares. Prove that the tangents to the parabola at the latter corners are diagonals of the focal squares, and make any additional repairs that may be necessary.

2 Problems of this section deal quite exclusively with parabolas placed upon coordinate systems in such a way that their equations have the standard form $y = kx^2$, where k is a positive constant. We can, however, pause briefly to note that the equation

$$y - y_0 = k(x - x_0)^2 \tag{1}$$

is the equation of a parabola having its vertex at the point (x_0,y_0). Supposing that a, b, c are constants for which $a \neq 0$, show that the graph of the equation

$$y = ax^2 + bx + c \tag{2}$$

is a parabola and find the coordinates of its vertex. *Solution:* From (2) we obtain

$$\begin{aligned} y &= a\left(x^2 + \frac{b}{a}x \qquad\right) + c \\ &= a\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}\right) + c - \frac{b^2}{4a} \\ &= a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a} \end{aligned} \tag{3}$$

and hence

$$y + \frac{b^2 - 4ac}{4a} = a\left(x + \frac{b}{2a}\right)^2. \tag{4}$$

Thus the graph of (2) is a parabola having its vertex at the point $\left(-\frac{b}{2a}, -\frac{b^2 - 4ac}{4a}\right)$. *Remark:* In case $a > 0$, the equation (4) has the form (1), where $k > 0$ and the graph "opens upward" like the graph of $y = kx^2$. In case $a < 0$, (4) has the form

$$y - y_0 = -k(x - x_0)^2, \tag{5}$$

where $k > 0$ and the graph "opens downward" like the graph of $y = -kx^2$.

These matters are important because equations of the form (2) appear very often, but for basic studies of parabolas we use the standard form $y = kx^2$, where $k > 0$.

3 Show that the tangent to the graph of the equation $y = kx^2$ at the point $P_1(x_1,kx_1^2)$ has the equation

$$y - kx_1^2 = 2kx_1(x - x_1) \qquad \text{or} \qquad y = kx_1(2x - x_1).$$

Show that this tangent intersects the y axis, the x axis, and the directrix at the points

$$A(0, -kx_1^2),\ B\left(\frac{x_1}{2}, 0\right),\ C\left(\frac{x_1}{2} - \frac{1}{8k^2x_1}, -\frac{1}{4k}\right)$$

provided, for the last point, $x_1 \neq 0$. Sketch a figure in which the parabola, the tangent, and the points A, B, C all appear.

4 Show that the normal to the graph of the equation $y = kx^2$ at the point $P_1(x_1,kx_1^2)$ has, when $x_1 \neq 0$, the equation

$$y - kx_1^2 = -\frac{1}{2kx_1}(x - x_1).$$

Show that this normal intersects the y axis, the x axis, and the directrix at the points

$$A^*\left(0, \frac{1}{2k} + kx_1^2\right), \qquad B^*(x_1 + 2k^2x_1^3, 0), \qquad C^*\left(\frac{3x_1}{2} + 2k^2x_1^3, -\frac{1}{4k}\right).$$

Sketch a figure showing all of these things.

5 Two points (x_1,y_1) and (x_2,y_2) lie on the parabola having the equation $y = kx^2$. Prove that the coordinates of the intersection R of the y axis and the line through these points can be put in the form $(0, -kx_1x_2)$. Figure 6.191 illustrates results of this and the next two problems, but the figures look quite different when x_1 and x_2 have opposite signs.

6 Two points (x_1,y_1) and (x_2,y_2) lie on the parabola having the equation $y = kx^2$. Prove that the coordinates of the intersection of the tangents to the parabola at these points can be put in the form

$$\left(\frac{x_1 + x_2}{2}, kx_1x_2\right).$$

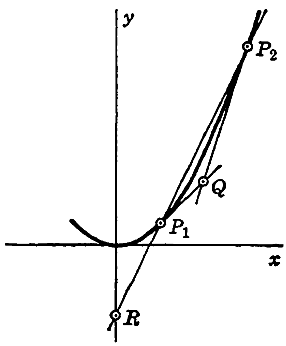


Figure 6.191

7 Show that the results of the two preceding problems yield the following theorem. Let P_1 and P_2 be two points on a parabola. Let Q be the intersection of the tangents to the parabola at P_1 and P_2. Let R be the intersection of the line P_1P_2 and the axis of the parabola. Then the mid-point of the segment QR lies on the line tangent to the parabola at the vertex. *Solution:* The results of the preceding problems show that the mid-point is $\left(\frac{x_1 + x_2}{4}, 0\right)$, and this point is on the tangent to the parabola at the vertex.

8 Let P_1 be a point on a parabola which is not the vertex V. Let W be the intersection of the tangents to the parabola at P_1 and V. Show that the line from the focus F to W is perpendicular to the line WP_1. *Hint:* Let the parabola have the equation $y = kx^2$ and use the fact that F has coordinates $(0, 1/4k)$. See Figure 6.192.

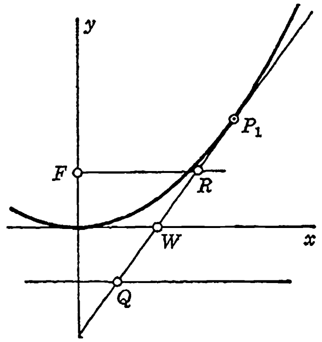


Figure 6.192

9 Let P_1 be a point on a parabola which is not the vertex V. Prove that the tangent to the parabola at P_1 meets the directrix and the line through the focus F parallel to the directrix at two points Q and R that are equidistant from F. See Figure 6.192.

10 A particle P moves on the parabola having the equation $y = kx^2$ in such a way that, at each time t, its x and y coordinates are t and kt^2 and the vector $\mathbf{r}$ running from the origin to P is

$$\mathbf{r} = t\mathbf{i} + kt^2\mathbf{j}.$$

Show that the velocity vector $\mathbf{v}$ is

$$\mathbf{v} = \mathbf{i} + 2kt\mathbf{j},$$

and note that this vector is also a "forward tangent" to the parabola at P. Letting F be the focus of the parabola, show that

$$\overrightarrow{FP} = t\mathbf{i} + \left(kt^2 - \frac{1}{4k}\right)\mathbf{j} \qquad \text{and} \qquad |\overrightarrow{FP}| = kt^2 + \frac{1}{4k}.$$

Letting ϕ_1 be the angle between the vector $\overrightarrow{FP}$ and the tangent vector $\mathbf{v}$, show that

$$\cos\phi_1 = \frac{\overrightarrow{FP}\cdot\mathbf{v}}{|\overrightarrow{FP}|\cdot|\mathbf{v}|} = \frac{2kt}{\sqrt{1 + 4k^2t^2}}.$$

Letting ϕ_2 be the angle between the tangent vector $\mathbf{v}$ and the vertical vector $\mathbf{j}$, show that

$$\cos\phi_2 = \frac{\mathbf{v}\cdot\mathbf{j}}{|\mathbf{v}|\,|\mathbf{j}|} = \frac{2kt}{\sqrt{1 + 4k^2t^2}}$$

and hence that $\phi_2 = \phi_1$. *Remark:* These formulas yield the famous reflection property of parabolas. They imply that the line FP and the line extending upward from P make equal angles with the normal to the parabola at P. This implies that if light or something else goes in a line from F and is reflected from the parabola in such a way that the angle θ_r of reflection is equal to the angle θ_i of incidence, then its path after reflection is parallel to the axis of the parabola.

11 Modify the work of the preceding problem to make a direct attack upon the angles which FP and $\mathbf{j}$ make with the normal to the parabola at P. *Remark:* This should be done when our primary interest lies in angles of incidence and reflection.

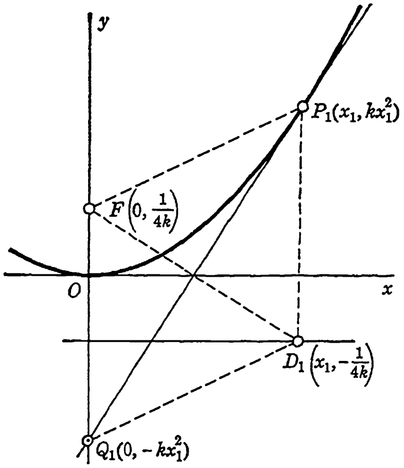


Figure 6.193

12 This problem and Figure 6.193 delve a bit deeper into the geometry of parabolas. The figure shows the parabola having the equation $y = kx^2$, where $k > 0$. The focus F and the directrix have the coordinates $(0, 1/4k)$ and the equation $y = -1/4k$. Let $x_1 > 0$. Show that the line through $P_1(x_1, kx_1^2)$ parallel to the axis of the parabola intersects the directrix at the point $D_1(x_1, -1/4k)$. Show that the tangent to the parabola at P_1 intersects the axis of the parabola at the point $Q_1(0, -kx_1^2)$. Show that the quadrilateral $Q_1D_1P_1F$ is a rhombus, that is, an equilateral parallelogram. This rhombus could be called a focal rhombus; in any case it is a focal square when $x_1 = 1/2k$ and the rhombus is a square. Show that the diagonals of this rhombus are perpendicular to each other and that they intersect at the point $(x_1/2, 0)$. Finally, show how these results and elementary geometry can be used to prove the reflection property of the parabola, namely, that the line from the focus to P_1 and the line through P_1 parallel to the axis of the parabola make equal angles with the tangent to the parabola at P_1.

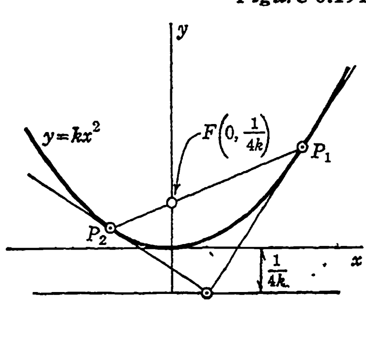
13 Write the equation of the tangent to the graph of the equation $y = x^2$ at the point (x_1,x_1^2) and then try to determine x_1 so the tangent will contain (or pass through) the point

(*a*) (1,1) (*b*) (1,0) (*c*) (0,1) (*d*) (−1,−1).

14 Find the equation of the normal to the graph of $y = x^2$ at the point (x_1,x_1^2) on the graph. Show that if $y_0 \leqq \frac{1}{2}$, then there is only one value of x_1 for which the normal passes through the point $(0,y_0)$, but that if $y_0 > \frac{1}{2}$, then there are three values of x_1 for which the normal passes through the point $(0,y_0)$. Sketch a figure or figures which show that the results seem to be reasonable.

15 As is the case for circles, a line segment joining two points on a parabola is a *chord* of the parabola, and the set of mid-points of the chords parallel to a given chord is called a *diameter* of the parabola. Thus a diameter is a point set, not a number. Letting the parabola have the equation $y = kx^2$, where $k > 0$, prove that for each m the diameter determined by chords having slope m is the line segment containing points (x,y) for which $x = m/2k$ and $y \geqq m^2/4k$.

Figure 6.194



16 A *focal chord* of a parabola is a line segment which contains the focus and has its ends at points on the parabola. Supposing that $x_2 < 0 < x_1$ as in Figure 6.194, show that the two points $P_1(x_1,kx_1^2)$, $P_2(x_2,kx_2^2)$ on the graph of $y = kx^2$ are end points of a focal chord if and only if $(2kx_1)(2kx_2) = -1$ and hence if and only if the tangents to the parabola at P_1 and P_2 are perpendicular.

17 With or without the aid of the results of the preceding problems, show that two different tangents to a parabola intersect on the directrix if and only if the tangents are perpendicular and hence if and only if the points of tangency are ends of a focal chord.

18 Prove that the center of a focal chord of a parabola is equidistant from the directrix and the ends of the chord.

19 Two equilateral triangles in E_2 are similar in the sense that one can be transformed into the other by a translation, a rotation, and a change of scale Show that the same is true of two parabolas in E_2. *Hint:* Suppose that the two given parabolas are translated and rotated so that their equations become $y = k_1x^2$ and $y = k_2x^2$, where k_1 and k_2 are positive constants. Show that if in the first equation we change scale by replacing x and y by λx and λy, we obtain $y = (\lambda k_1)x^2$.

20 Let k be a positive constant. For each positive number a, let $F(a)$ be the y coordinate of the center of the circle tangent to the graph of $y = kx^2$ at the points for which $x = a$ and $x = -a$. Find $F(a)$ and $\lim_{a\to 0} F(a)$. *Ans.:* $ka^2 + 1/2k$ and $1/2k$.

21 Let k be a positive constant. For each positive number a, let $(G(a), H(a))$ be the center of the circle which is tangent to the graph of $y = kx^2$ at the point (a,ka^2) and which contains (or passes through) the origin. Show that

$$G(a) = \frac{3}{2}ka^2 + \frac{1}{2k}, \qquad H(a) = -k^2a^3$$

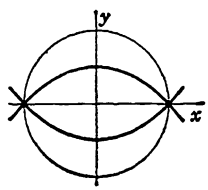
and

$$\lim_{a\to 0} G(a) = 1/2k, \qquad \lim_{a\to 0} H(a) = 0.$$

22 Sketch a graph of the parabola having the equation $y = x^2$ and then sketch several circles which have centers on the positive y axis and are tangent to the x axis at the origin. Observe that sufficiently big circles in this family intersect the parabola at points different from the origin and that small circles leave us in doubt. Supposing that $k > 0$, investigate this matter for the parabola having the equation $y = kx^2$. *Ans.:* The circle with center at $(0,a)$ and radius a intersects the parabola only at the origin (and is elsewhere above or "inside" the parabola) if and only if $a \leqq 1/2k$. Thus the biggest one of these circles which lies completely on or inside the parabola has radius equal to the distance from the focus to the directrix of the parabola.

23 Study the set S which contains a point $P(x,y)$ if and only if the point is equidistant from the x axis and the circle with center at the origin and radius a. *Solution:* This problem is interesting because S contains some points inside the circle as well as some points on and some points outside the circle; see Figure 6.195. Whether a point $P(x,y)$ lies inside or on or outside the circle, it will be in the set S iff (if and only if)

Figure 6.195



$$(1) \qquad |\sqrt{x^2 + y^2} - a| = |y|$$

and hence if

(2) $$\sqrt{x^2 + y^2} = a \pm y.$$

If (2) holds, then

(3) $$x^2 + y^2 = a^2 \pm 2ay + y^2$$

and hence either

(4) $$y = -\frac{a}{2} + \frac{x^2}{2a}$$

or

(5) $$y = \frac{a}{2} - \frac{x^2}{2a}.$$

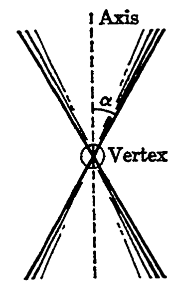
It can be shown that if (4) or (5) holds, then (1) holds. It follows that S is the sum (or union) of two parabolas of which one has the equation (4) and the other has the equation (5). Each parabola has its focus at the origin, and the directrices are the tangents to the circle that are parallel to the x axis. The parabolas intersect the x axis where the circle does.

24 With or without the aid of results of preceding problems, let P be a given parabola and verify the following facts which show students of mechanical drawing how to locate the axis, vertex, focus, and directrix of P. The mid-points M_1 and M_2 of two parallel chords C_1 and C_2 of P determine a line L_1 parallel to the axis of P. The axis L of P is the perpendicular bisector of the line segment joining points where a line perpendicular to L_1 intersects P. In case L_1 is not the axis of P, the mid-points M_1' and M_2' of chords C_1' and C_2' perpendicular to C_1 and C_2 determine another line L_1' parallel to L. Let L_1 and L_1' intersect the parabola at P_1 and P_1'. Then the line T_1 (or T_1') through P_1 (or P_1') parallel to C_1 (or C_1') is tangent to P at P_1 (or P_1'). Moreover T_1 and T_1' are perpendicular, and their intersection is on the directrix of the parabola. Finally, the line segment joining P_1 and P_1' is a focal chord of the parabola so this segment intersects L at the focus. *Remark:* Anyone who spends a substantial part of his lifetime working with parabolas can learn very much about them.

6.2 *Geometry of cones and conics*

Throughout this chapter, a *cone* is always a complete right circular conical surface consisting of two parts, or *nappes*, as in Figure 6.21. We assume that $0 < \alpha < \pi/2$ and that the lines on the cone all make the same angle α with the axis of the cone. For the present, we simplify our discussion by supposing that the axis of the cone is vertical, and hence that each plane perpendicular to the axis is horizontal. A conic (or conic section) is the set of points in which a plane π intersects the cone. In case π contains the vertex V, the resulting conic is either a single point or a single line or a pair of intersecting lines. In case π is perpendicular to the axis of the cone and does not contain the vertex, the conic is a circle. Our interest

Figure 6.21



in this chapter lies in conics of less simple natures. These turn out to be parabolas and ellipses, each of which intersects only one nappe of the cone, and hyperbolas, each of which intersects both nappes of the cone.

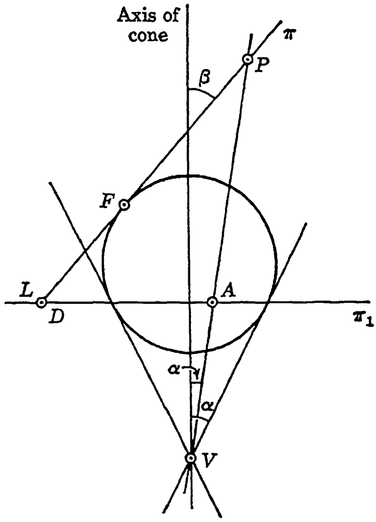


Figure 6.22

Without yet knowing what will be learned, we look at Figure 6.22, which is one of the most remarkable figures of elementary geometry. This is a flat nonperspective figure which must be constructed and studied rather carefully before it can be fully understood. The vertical line in the plane of the paper is the axis of a cone with vertex V and vertex angle α. The line making the acute angle β with the axis represents more than a line. It is supposed to lie in the plane of the paper, and it represents a plane π which makes the angle β with the axis of the cone and which intersects the cone in a conic K. We can, if we wish to do so, think of the plane π as being an xy plane in which the x axis is pointed toward our eyes, and everything in this plane seems to lie on one line. The graph of each point P on the conic K, whether P lies on the part of the cone in front of the plane of the paper or on the part behind the paper, is on the line. Our first significant step is to fit a sphere into the cone, the sphere being just big enough to be tangent to the plane π. The circle in the figure represents this sphere, which is tangent to the cone at the points of a circle which lies in the horizontal plane π_1 and which is tangent to the plane π at the point F.

It can now be revealed that discoveries will be made; in fact we shall show that F is a focus of the conic K. Because π_1 is horizontal and π is not, these planes intersect in a line L which is represented by a single point in the flat figure. We shall show that L is a directrix of the conic K. To start learning something about the conic K, let P be a point on K and draw the line segment PD which lies in π and is perpendicular to L at the point D on L. The line PV lies on the cone and is tangent to the sphere at a point A in π_1. The line PF lies in π and is tangent to the sphere at F. Therefore $|\overrightarrow{PF}| = |\overrightarrow{PA}|$ because the two vectors have their tails at the same point P and are tangent to the sphere at their tips. If we let d be the distance from P to the plane π_1, then $d = |\overrightarrow{AP}| \cos \alpha$ because the vector $\overrightarrow{AP}$ makes the angle α with vertical lines. Also $d = |\overrightarrow{DP}| \cos \beta$

because the vector $\overrightarrow{DP}$ makes the angle β with vertical lines. Equating the two expressions for d gives the formula

$$|\overrightarrow{AP}| \cos \alpha = |\overrightarrow{DP}| \cos \beta.$$

This and the fact that $|\overrightarrow{PF}| = |\overrightarrow{PA}|$ give the fundamental formula

$$|\overrightarrow{PF}| = \frac{\cos \beta}{\cos \alpha} |\overrightarrow{PD}| = e|\overrightarrow{PD}| \tag{6.23}$$

where the constant e defined by the formula

$$e = \frac{\cos \beta}{\cos \alpha}, \tag{6.231}$$

is called the *eccentricity* of the conic. The point F and the line L are respectively a *focus* and a *directrix* of the conic. It is clear from Figure 6.22 that, whenever α and β are given acute angles, we can make the distance from F to L have any positive value p we please by taking π at the appropriate distance from V. The equation (6.23) is called an *intrinsic equation* of the conic, that is, an equation that depends only upon the conic itself and not upon the coordinates of a particular "external" coordinate system.

In case $\beta = \alpha$, the conic is called a *parabola*. In this case, (6.231) shows that $e = 1$ and that the formula (6.23) reduces to the simpler formula $|\overrightarrow{PF}| = |\overrightarrow{PD}|$. Thus we have, as was promised in Section 1.4, proved that a parabola is the set of points P (in a plane) equidistant from a fixed point (the focus F) and a fixed line (the directrix L).

In case $\alpha < \beta < \pi/2$, the conic is called an *ellipse*. In this case

$$|\overrightarrow{PF}| = e|\overrightarrow{PD}|, \tag{6.24}$$

where the eccentricity is a constant e for which $0 < e < 1$. As we can see from Figure 6.22, an ellipse is an oval (or oval curve) that lies entirely on one nappe of the cone. Ellipses will be studied in greater detail in Section 6.3.

In case $0 < \beta < \alpha$, the conic is called a *hyperbola*. In this case

$$|\overrightarrow{PF}| = e|\overrightarrow{PD}|, \tag{6.241}$$

where the eccentricity is a constant e for which $e > 1$. As we can see from Figure 6.22, a hyperbola consists of two branches (or parts) one of which is contained in each nappe of the cone. Hyperbolas will be studied in greater detail in Section 6.4.

The above information enables us to find the equation of a nontrivial conic K (parabola, ellipse, or hyperbola) which lies in an xy plane when we

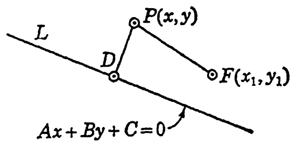


Figure 6.25

know the eccentricity e, the coordinates (x_1,y_1) of a focus F, and the equation $Ax + By + C = 0$ of a directrix L. While different values of e yield conics of different shapes, the schematic Figure 6.25 may be helpful. The formula $|\overrightarrow{PF}| = e|\overrightarrow{PD}|$ is equivalent to the formula $|\overrightarrow{FP}|^2 = e^2|\overrightarrow{PD}|^2$, and use of formulas for distances from points to points and from points to lines (see Theorem 1.48) enables us to put this in the form

$$(6.251)\qquad (x - x_1)^2 + (y - y_1)^2 = \frac{e^2}{A^2 + B^2}(Ax + By + C)^2.$$

While the equation (6.251) has its virtues, we can obtain a more informative equation by choosing the x,y coordinate system in such a way that the focus F lies on the x axis and the directrix is perpendicular to the x axis. With the intention of so determining x_1 that the resulting equation will have its simplest form, we suppose that the focus F has coordinates $(x_1,0)$ and that the directrix lies a given positive distance p to the left of the focus (as in Figure 6.26) so that the equation of the directrix is $x = x_1 - p$. The intrinsic equation $|\overrightarrow{FP}|^2 = e^2|\overrightarrow{PD}|^2$ then gives the coordinate equation

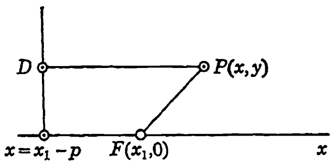


Figure 6.26

$$(x - x_1)^2 + y^2 = e^2[x - (x_1 - p)]^2$$

or

$$(6.261)\qquad (1 - e^2)x^2 + 2[e^2(x_1 - p) - x_1]x + y^2 = e^2(x_1 - p)^2 - x_1^2.$$

We can now begin to see how the nature of the equation depends upon the eccentricity e. In case $e = 1$, so that the conic is a parabola, (6.261) reduces to

$$(6.262)\qquad x = \frac{1}{2p}y^2 + \left(x_1 - \frac{p}{2}\right).$$

This equation has its simplest form when $x_1 = p/2$. The simplest equation has the form $x = ky^2$ which (except that the roles of x and y were interchanged to simplify matters) was studied in Section 6.1.

We now face the task of simplifying (6.261) for cases in which $0 < e < 1$ or $e > 1$ and the conic is an ellipse or a hyperbola. As is easy to guess, the greatest simplification results from so choosing x_1 that the coefficient of x is 0. Therefore, we let x_1 be determined by the equivalent equations

$$(6.27)\qquad e^2(x_1 - p) - x_1 = 0,\qquad x_1 = \frac{-e^2p}{1 - e^2},\qquad x_1 - p = \frac{-p}{1 - e^2}.$$

Putting this value of x_1 in (6.261) gives the equation

$$(1 - e^2)x^2 + y^2 = \frac{e^2p^2}{1 - e^2}. \tag{6.271}$$

This equation of the conic is a good source of information, but we obtain a better source by putting the equation in the "standard form." To do this neatly and correctly, we sacrifice some paper to put (6.271) in the form

$$\frac{(1 - e^2)x^2}{1} + \frac{y^2}{1} = \frac{e^2p^2}{1 - e^2},$$

divide the numerator and denominator of the first term by $(1 - e^2)$ to obtain

$$\frac{x^2}{\dfrac{1}{1 - e^2}} + \frac{y^2}{1} = \frac{e^2p^2}{1 - e^2},$$

and then divide both members of this equation by the right member to obtain an equation which is put in the form

$$\frac{x^2}{\dfrac{e^2p^2}{(1 - e^2)^2}} + \frac{y^2}{\dfrac{e^2p^2}{1 - e^2}} = 1, \qquad \text{(ellipse)} \tag{6.272}$$

when $0 < e < 1$, and in the form

$$\frac{x^2}{\dfrac{e^2p^2}{(e^2 - 1)^2}} - \frac{y^2}{\dfrac{e^2p^2}{e^2 - 1}} = 1, \qquad \text{(hyperbola)} \tag{6.273}$$

when $e > 1$. Everything is so arranged that the denominators in (6.272) and (6.273) are positive.

Unless we think a bit about sections of cones, we cannot fully appreciate the significance of these formulas. It is sometimes said that any reasonably sane person should feel quite sure that an ellipse is an egg-shaped oval which has a "small end" at the part of the ellipse nearest the vertex of the cone and which has a "big end" at the part of the ellipse farthest from the vertex of the cone. However, (6.272) shows very clearly that the x and y axes are axes of symmetry of the ellipse and that the origin is a center (center of symmetry) of the ellipse. Thus (6.272) reveals the astonishing fact that the ellipse has a center and that the two "ends" of the ellipse are alike (or congruent). Similarly, suppose a particular hyperbola H intersects one nappe of a cone at some points near the vertex of the cone but intersects the other nappe only at points very far from the vertex of the cone. It would seem to be incredible that the two branches of this hyperbola should be alike, but they are alike.

The formula (6.273) shows that the x and y axes are axes of symmetry of the hyperbola and that the origin is a center (center of symmetry) of the hyperbola. This proves that the two branches of the hyperbola are congruent.

There are two reasons for putting the above equations in the simpler standard forms

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{6.28}$$

where a and b are positive constants. In the first place the denominators in (6.272) and (6.273) are clumsy things to write, and in the second place equations of the form (6.28) often arise in problems where a and b are not children of eccentricities. Comparing (6.272) and (6.273) with (6.28) shows that a and b are determined in terms of e and p by the formulas

$$a = \frac{ep}{|1 - e^2|}, \qquad b = \frac{ep}{\sqrt{|1 - e^2|}}. \tag{6.281}$$

On the other hand, the formulas

$$|x_1| = \frac{e^2p}{|1 - e^2|} = ae = \text{distance from center to focus} \tag{6.282}$$

$$|x_1 - p| = \frac{p}{|1 - e^2|} = \frac{a}{e} = \text{distance from center to directrix} \tag{6.283}$$

$$ae = \sqrt{a^2 - b^2} \quad \text{for ellipse,} \qquad ae = \sqrt{a^2 + b^2} \quad \text{for hyperbola} \tag{6.284}$$

serve to determine other quantities in terms of a and b. The first two of these formulas are obtained very quickly by comparing the formulas for x_1 and $x_1 - p$ in (6.27) with the formula for a in (6.281). To obtain (6.284), we can square the members of (6.281) and combine the results to obtain $b^2/a^2 = |1 - e^2|$ and then treat separately the cases in which $0 < e < 1$ and $e > 1$. Observe that $0 < b < a$ when $0 < e < 1$ and also $1 < e < \sqrt{2}$ but that $b > a$ when $e > \sqrt{2}$. Graphs of ellipses and hyperbolas, and schemes for remembering essential parts of the above formulas, will appear in later sections.

Problems 6.29

1 Copy Figure 6.25 and the equation (6.251) of the conic K and look at them. Then show that the equation of K can be put in the form

$$\begin{aligned}(1)\quad &\left(1 - \frac{e^2A^2}{A^2 + B^2}\right)x^2 - \frac{2e^2AB}{A^2 + B^2}xy + \left(1 - \frac{e^2B^2}{A^2 + B^2}\right)y^2 \\ &- 2\left(x_1 + \frac{e^2AC}{A^2 + B^2}\right)x - 2\left(y_1 + \frac{e^2BC}{A^2 + B^2}\right)y = \frac{e^2C^2}{A^2 + B^2} - (x_1^2 + y_1^2).\end{aligned}$$

This equation is ponderous and nobody should ever dream of remembering it, but it is useful.

2 With the aid of the result of Problem 1, show that the coefficient of xy in the equation of a conic K is zero if and only if each directrix of K is parallel to one of the coordinate axes.

3 With the aid of the result of Problem 1, show that if the coefficients of x^2 and y^2 in the equation of a conic K are both 0, then $e = \sqrt{2}$ and $A^2 = B^2$. Show that when $e = \sqrt{2}$ and $B = A$, the equation reduces to

$$xy + \left(x_1 + \frac{C}{A}\right)x + \left(y_1 + \frac{C}{A}\right)y = \frac{x_1^2 + y_1^2}{2} - \frac{C^2}{2A^2}. \tag{2}$$

Show that when $e = \sqrt{2}$ and $B = -A$, the equation reduces to

$$xy - \left(x_1 + \frac{C}{A}\right)x - \left(y_1 - \frac{C}{A}\right)y = \frac{C^2}{2A^2} - \frac{x_1^2 + y_1^2}{2}. \tag{3}$$

The graphs of these equations are hyperbolas because $e = \sqrt{2} > 1$.

4 With the aid of Problem 3, show that if $e = \sqrt{2}$, $B = A$, $x_1 = -C/A$, and $y_1 = -C/A$, then the equation of the conic K is

$$xy = \frac{C^2}{2A^2}. \tag{4}$$

Show also that if $e = \sqrt{2}$, $B = -A$, $x_1 = -C/A$, and $y_1 = C/A$, then the equation of the conic K is

$$xy = -\frac{C^2}{2A^2}. \tag{5}$$

Our equations are now much simpler.

5 Let $k > 0$. Observe that formula (4) of Problem 4 reduces to $xy = k$ when $A = 1$ and $C = -\sqrt{2k}$. This shows that $xy = k$ is the equation of the conic having a focus at the point $(\sqrt{2k}, \sqrt{2k})$, having a directrix with the equation

$$x + y - \sqrt{2k} = 0, \tag{6}$$

and having eccentricity $e = \sqrt{2}$. *Remark:* Relatively few persons have enough courage to undertake to sketch or otherwise describe a cone in E_3 which intersects the xy plane in the graph of the equation $xy = k$ and then use methods of synthetic geometry (geometry which, unlike analytic geometry, never uses algebra and other brands of mathematical analysis) to obtain information about the foci and directrices of the conic. However, the results of this problem enable us to put this information in very simple terms. Supposing that $k > 0$, we start with a good clean x, y coordinate system and show how to locate the foci and directrices of the hyperbola having the equation $xy = k$. The point $V(\sqrt{k}, \sqrt{k})$ clearly

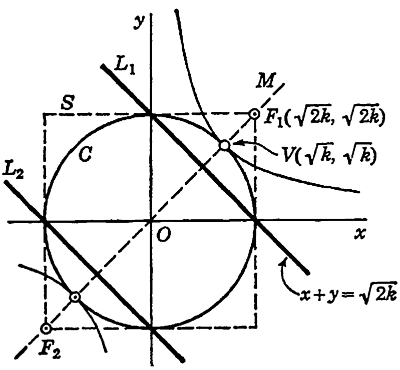


Figure 6.291

lies on the hyperbola, and we start by locating it in Figure 6.291. We then enter the wholesale sketching business and sketch the line M containing O and V, the circle C with center at O which contains V, and the square S whose sides are tangent to C at the points where C intersects the coordinate axes. The points $F_1(\sqrt{2k}, \sqrt{2k})$ and $F_2(-\sqrt{2k}, -\sqrt{2k})$ where the square S meets the line M are the foci of the hyperbola. The line L_1 joining the points where the circle and square meet the positive x and y axes is one directrix of the hyperbola. The line L_2 joining the points where the circle and square meet the negative x and y axes is the other directrix of the hyperbola. Finally, we complete Figure 6.291 by sketching the hyperbola. As we know, the coordinate axes are asymptotes of the hyperbola The origin is the center (the center of symmetry) of the hyperbola. The line M is a line of symmetry; it is called the *transverse axis* of the hyperbola. The line through O perpendicular to M is another line of symmetry; it is called the *conjugate axis* of the hyperbola. The particular hyperbola we have been studying is called a rectangular hyperbola because its asymptotes are at right angles to each other.

6 Information theory teaches that results like those of Problem 5, which depend upon considerable calculation, should be checked when it is relatively easy to do so. Letting F be the point $(\sqrt{2k}, \sqrt{2k})$, letting L be the line having the equation $x + y - \sqrt{2k} = 0$, letting D be the foot of the perpendicular from P to L, and letting $e = \sqrt{2}$, simplify the intrinsic equation $|\overrightarrow{FP}|^2 = e^2|\overrightarrow{DP}|^2$ to obtain the coordinate equation $xy = k$. *Remark:* We are seldom required to find the distance from a point to a line which is not parallel to a coordinate axis, and it is helpful to be able to find Theorem 1.48, which enables us to obtain the result very quickly.

7 Find the set of numbers k such that there exists at least one point (x,y) whose coordinates satisfy the equation $y = kx$ and the equation

(*a*) $x^2 + y^2 = 1$ (*b*) $xy = 1$ (*c*) $xy = -1$

(*d*) $\dfrac{x^2}{9} + \dfrac{y^2}{4} = 1$ (*e*) $\dfrac{x^2}{9} - \dfrac{y^2}{4} = 1$

In each case, sketch a figure which shows the geometric significance of the result.

8 Tell why a circle is not an ellipse. *Remark:* The answer must be based upon definitions and not upon intuitions of the untutored.

9 When $A \neq 0$ and $B > 0$, the graph of the equation

$$y^2 = Ax^2 + B$$

is a *central conic* (circle or ellipse or hyperbola but not a parabola or a degenerate conic) having its center at the origin. Show that if m and b are constants such that the line having the equation

$$y = mx + b$$

intersects the conic at two points, then the mid-point of the chord of the conic joining these points has coordinates (x,y), where

$$x = b\frac{m}{2(A - m^2)}, \qquad y = b\frac{m^2 + 2(A - m^2)}{2(A - m^2)}.$$

Use this result to show that, for each fixed m, the set D of mid-points of chords having slope m lies on a line through the center of the conic. *Remark:* This set D, which is always a line segment when the conic is a circle or an ellipse and is sometimes a whole line when the conic is a hyperbola, is called a *diameter* of the conic.

10 Let m be a positive constant. A surface S contains the origin, and when $y \neq 0$, it contains the circle which lies in a plane parallel to the xz plane and has a diameter coinciding with the line segment joining the two points $(0,y,0)$ and $(0,y,2my)$ in the yz plane. Sketch a figure showing S and show that the equation of S is

$$x^2 + (z - my)^2 = m^2y^2.$$

Remark: One who wishes to rise above minimum requirements may show that (i) S is a quadric surface, (ii) S is a cone and hence is a quadric cone, (iii) S is not a right circular cone. One who wishes to rise to still greater heights may try to decide whether we know enough to determine whether the cone has an axis and, if so, whether sections made by planes perpendicular to this axis are ellipses.

6.3 *Ellipses* Remarkable geometric properties of ellipses can be extracted from Figure 6.31. This figure, like Figure 6.22, shows a cone having a vertical axis. The axis lies in the plane of the paper. The plane π intersects the cone in an ellipse E of which the two points (vertices, in fact) V_1 and V_2 lie in the plane of the paper. The smaller circle represents a sphere which is tangent to the cone at the points of a circle which determines the plane π_1 and is tangent to π at F_1. As we saw in Section 6.2, π_1 and π intersect in a line L_1 and, moreover, F_1 is a focus and L_1 is a directrix of the ellipse E. The larger circle represents a sphere which is tangent to the cone at the points of a circle which determines the plane π_2 and is tangent to π at F_2. The planes π_2 and π intersect in a line L_2, and the same procedure which was applied to F_1 and L_1 shows that F_2 is another

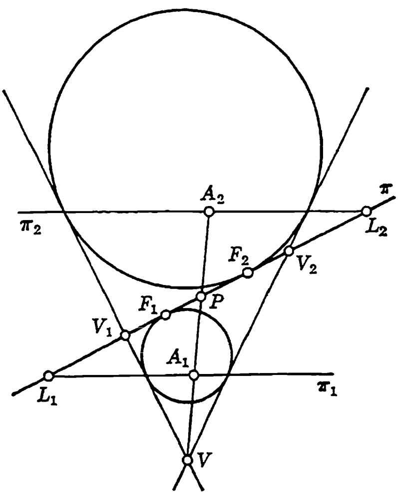


Figure 6.31

focus and L_2 is another directrix of the ellipse E. Thus E has two foci and two directrices.

We can look at Figure 6.31 and make some informal observations that we shall not (and perhaps cannot) make precise. If V_1 and V_2 are nearly equidistant from V, then the ellipse is nearly circular, the foci are close together and nearly midway between V_1 and V_2, and the eccentricity e is, as (6.231) shows, nearly zero. If we keep V_1 where it is and replace V_2 by a point many miles up on the cone, then the eccentricity will be near 1, the ellipse will be relatively flat, and the part of the ellipse within a few miles of V_1 would look so much like a part of a parabola that very careful inspection of this part would be required to enable us to tell whether the conic is an ellipse or a parabola or a hyperbola.

Our next step is to use Figure 6.31 to obtain the famous *string property*

$$|\overrightarrow{F_1P}| + |\overrightarrow{F_2P}| = |\overrightarrow{V_1V_2}| \tag{6.32}$$

of the ellipse E, which shows that the sum of the distances from the foci of an ellipse to a point P on the ellipse has the same constant value for all points P on the ellipse. Let P be V_1 or V_2 or any other point on the ellipse. The line VP lies on the cone and is tangent to the lower and upper spheres at points A_1 and A_2. Then, as was pointed out in Section 6.2, $|\overrightarrow{PF_1}| = |\overrightarrow{PA_1}|$ because the two vectors have their tails at the same point and are tangent to a sphere at their tips. Also, $|\overrightarrow{PF_2}| = |\overrightarrow{PA_2}|$ for the same reason, the upper sphere now being involved. Therefore,

$$|\overrightarrow{F_1P}| + |\overrightarrow{F_2P}| = |\overrightarrow{A_1P}| + |\overrightarrow{PA_2}| = |\overrightarrow{A_1A_2}|. \tag{6.33}$$

Wherever the point P may be on the cone, the number $|\overrightarrow{A_1A_2}|$ is the constant slant height of the segment of the cone that lies between the parallel planes π_1 and π_2; in fact if d is the distance between π_1 and π_2, then $|\overrightarrow{A_1A_2}| = d/\cos\alpha$, where α is the angle at the vertex of the cone. The points V_1, F_1, F_2, and V_2 all lie on the line in which π intersects the plane of the paper. The results of setting $P = V_1$, and then $P = V_2$, in (6.33) give

$$|\overrightarrow{F_1V_1}| + |\overrightarrow{F_2V_1}| = \frac{d}{\cos\alpha}, \qquad |\overrightarrow{F_1V_2}| + |\overrightarrow{F_2V_2}| = \frac{d}{\cos\alpha}, \tag{6.331}$$

and with the aid of Figure 6.31 we can put this in the form

$$2|\overrightarrow{V_1F_1}| + |\overrightarrow{F_1F_2}| = |\overrightarrow{F_1F_2}| + 2|\overrightarrow{F_2V_2}| = \frac{d}{\cos\alpha}. \tag{6.332}$$

This gives the remarkable fact that

$$|\overrightarrow{V_1F_1}| = |\overrightarrow{F_2V_2}|. \tag{6.333}$$

With the aid of Figure 6.31 and these formulas, we find that

$$(6.334)\qquad \begin{aligned}|\overrightarrow{V_1V_2}| &= |\overrightarrow{V_1F_1}| + |\overrightarrow{F_1F_2}| + |\overrightarrow{F_2V_2}|\\ &= 2|\overrightarrow{V_1F_1}| + |\overrightarrow{F_1F_2}| = \frac{d}{\cos\alpha} = |\overrightarrow{A_1A_2}|.\end{aligned}$$

From this and (6.33) we obtain the string property (6.32). One reason for interest in the string property of ellipses lies in the fact that it provides a mechanical method for drawing ellipses. Let a string of length $2a$ have its ends pinned to two points F_1 and F_2 on a sheet of paper. Using a pencil point to stretch the string into two straight segments, we can move the pencil so that its point draws an ellipse having foci at F_1 and F_2 as the string slides over the pencil point.

Figure 6.34 shows an ellipse which was drawn with the aid of the string property, and it also shows some numerical dimensions which display

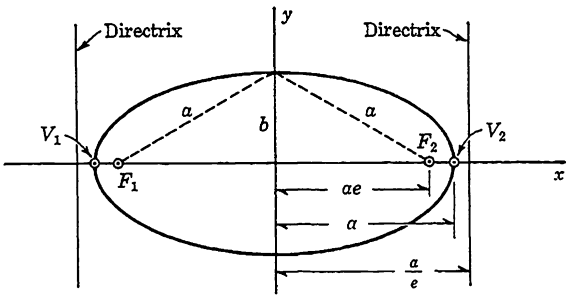


Figure 6.34

information from the paragraph containing (6.281). Even though the equation of the ellipse of the figure has already been derived, it is worthwhile to know about the operations involved in using the intrinsic string property $|\overrightarrow{F_1P}| + |\overrightarrow{F_2P}| = 2a$ to derive the equation. Letting $F_1(-ae,0)$ and $F_2(ae,0)$ be located on the x axis with the origin midway between them as in Figure 6.34, we use the string property to obtain the uninformative equation

$$(6.35)\qquad \sqrt{(x+ae)^2+y^2} + \sqrt{(x-ae)^2+y^2} = 2a,$$

which should be simplified. If we square the members of this equation, the product of the two square roots will complicate our calculations. It is better to transpose one of the square roots (we select the second) and square and simplify the result to obtain

$$\sqrt{(x-ae)^2+y^2} = a - ex.$$

Squaring and simplifying this gives

$$(1-e^2)x^2 + y^2 = a^2(1-e^2),$$

and dividing by the right side gives the standard form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{6.36}$$

where $b^2 = a^2(1 - e^2)$ or

$$ae = \sqrt{a^2 - b^2}. \tag{6.37}$$

It is of interest to see that memorization of only a few details enables us to find the graph, foci, and directrices of the graph of (6.36) when a and b are given constants, say $a = 5$ and $b = 2$. Putting $y = 0$ shows that points $(a,0)$ and $(-a,0)$ lie on the graph. The line segment joining these points is the *major axis* of the graph. Putting $x = 0$ shows that the points $(0,b)$ and $(0,-b)$ lie on the graph. The line joining these points is the *minor axis* of the graph. The graph is an ellipse through these four points. The foci always lie on the major axis. With or without the aid of the string property of the ellipse, we can remember that the foci lie on the major axis and on the circle of radius a having its center at an end of the minor axis. Then Figure 6.38, which is a simpler version of Figure 6.34, shows that the distance from the center of the ellipse to the foci can be calculated by the Pythagoras theorem. The distance is $\sqrt{a^2 - b^2}$, and if we will remember that this is ae, then we can calculate e. Finally, we can calculate the distance from the center to the directrices if we remember that this distance is a/e. The numbers ae and a/e are the key numbers.

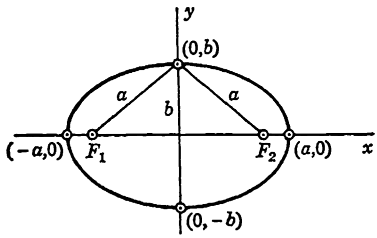


Figure 6.38

Problems 6.39

1 For each of the following pairs of values of a and b, sketch the ellipse having the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

find the eccentricity, find the foci (give coordinates), and find the directrices (give equations). Try to cultivate the ability to use the Pythagoras theorem and key numbers without use of books or notes. Check the numerical results by use of the fact that the distance p from a focus to its directrix must satisfy the equation $e^2p^2 = b^2(1 - e^2)$.

(*a*) $a = 5$, $b = 2$ (*b*) $a = 3$, $b = 1$ (*c*) $a = 5$, $b = 4$

2 The equation

$$\frac{x^2}{2^2} + \frac{y^2}{3^2} = 1$$

differs from equations of ellipses having their foci on the x axis because the denominator under x^2 is not greater than the denominator under y^2. Nevertheless, plot the four points on the graph obtained by setting $x = 0$ and then $y = 0$, and then sketch the graph. Observe that everything is like the preceding problem except that the roles of x and y are interchanged. Then proceed to find the eccentricity, foci, and directrices. Repeat the process when 2 and 3 are respectively replaced by

(*a*) 2 and 5 (*b*) 1 and 5 (*c*) 3 and 5

3 We have known for a long time that the graph of the equation

$$(x - h)^2 + (y - k)^2 = a^2$$

is a circle having its center at the point (h,k). With this hint, sketch graphs of the equations

(*a*) $\frac{(x-1)^2}{5^2} + \frac{(y-2)^2}{2^2} = 1$ (*b*) $\frac{(x-1)^2}{2^2} + \frac{(y-2)^2}{5^2} = 1$

Observe that, in these cases, distances from centers to foci are *not* coordinates of foci; suitable adjustments must be made. *Remark:* A good clean start is made by setting $y = 2$ and calculating $x - 1$ and then x.

4 Find the equations of the ellipses (if any) which have foci at the points $(-2,0)$ and $(2,0)$ and which pass through the point $(1,1)$.

5 Find the equation of the ellipse which has its center at the point $(2,3)$, which has axes parallel to the coordinate axes, and which is tangent to the coordinate axes. Sketch a reasonably good figure.

6 The foci of a particular ellipse lie midway between the center and vertices. Find the eccentricity. Supposing that the major axis has length $2a$, find the length of the minor axis and the distance from the center to the directrices. Sketch a reasonably good figure.

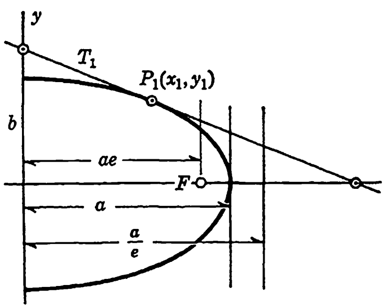
7 Except for minor perturbations, the orbit of the earth is an ellipse having the sun at a focus. The least and the greatest distances from the earth to the sun have the ratio $\frac{29}{30}$. Find the eccentricity of the approximate orbit. *Ans.:* $\frac{1}{59}$.

8 As in Figure 6.391, let $P_1(x_1,y_1)$ be a point on the ellipse having the standard equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Supposing that P_1 is not one of the points where the ellipse intersects the

Figure 6.391



coordinate axes, find the equation of the line T_1 tangent to the ellipse at P_1. *Ans.:*

$$\frac{x_1x}{a^2} + \frac{y_1y}{b^2} = 1.$$

9 Find the coordinates of the points where the tangent T_1 of Problem 8 intersects the coordinate axes. *Ans.:*

$$\left(\frac{a^2}{x_1}, 0\right), \quad \left(0, \frac{b^2}{y_1}\right).$$

10 Find the coordinates of the points where the tangent T_1 of Problem 8 intersects the lines through the foci perpendicular to the major axis. *Ans.:*

$$\left(-ae, \frac{b^2}{y_1}\left(1 + \frac{ex_1}{a}\right)\right), \quad \left(ae, \frac{b^2}{y_1}\left(1 - \frac{ex_1}{a}\right)\right).$$

11 Find the coordinates of the points where the tangent T_1 of Problem 8 intersects the directrices of the ellipse. *Ans.:*

$$\left(-\frac{a}{e}, \frac{b^2}{y_1}\left(1 + \frac{x_1}{ae}\right)\right), \quad \left(\frac{a}{e}, \frac{b^2}{y_1}\left(1 - \frac{x_1}{ae}\right)\right).$$

12 Let the line T_1 tangent to an ellipse at P_1 intersect a directrix at Q_1 and let F be the focus corresponding to the directrix. With the aid of Problem 11 and the fact that $b^2 = a^2(1 - e^2)$, prove that the line FQ_1 is perpendicular to the line FP_1. *Remark:* This result has some quite surprising consequences. If the focus F, directrix D, and one single point $P_1(x_1,y_1)$ of an ellipse are marked in a plane, we can give a simple rule for drawing the line T which is tangent to the undrawn ellipse at P_1. In case P_1 is on the line through F perpendicular to D, the tangent T_1 is the line through P_1 parallel to D. Otherwise, the tangent T_1 is the line containing P_1 and the point Q_1 where the line through F perpendicular to the line FP_1 intersects the directrix. This result implies that *if P_1 and P_2 are points at the end of a focal chord* (a chord containing a focus), *then the tangents at P_1 and P_2 intersect at the point Q_1 on the directrix where the line through the focus perpendicular to the line P_1P_2 meets the directrix.* Figure 6.392, in which a part of the ellipse is drawn, illustrates this elegant geometric fact.

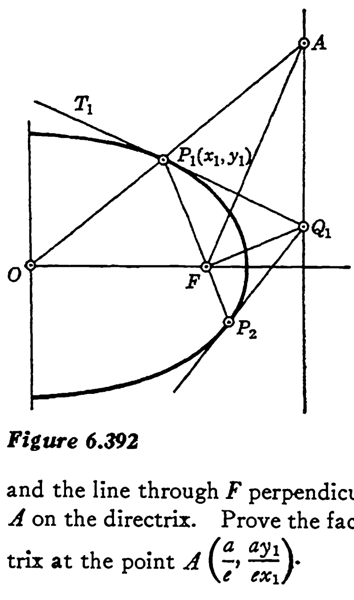


Figure 6.392

13 Figure 6.392 illustrates another interesting geometric fact. Then line OP_1 and the line through F perpendicular to the tangent T_1 at P_1 intersect at a point A on the directrix. Prove the fact by proving that each line intersects the directrix at the point $A\left(\frac{a}{e}, \frac{ay_1}{ex_1}\right)$.

14 Figure 6.393 illustrates the fact that the line V_1P_1 from a vertex V_1 to a point P_1 on an ellipse is parallel to the line OE from the center of the ellipse to the point E where the tangent at P_1 intersects the tangent at the other vertex V_2. Show that the coordinates of E are $\left(a, \frac{b^2}{y_1}\left(1 - \frac{x_1}{a}\right)\right)$ and prove the fact.

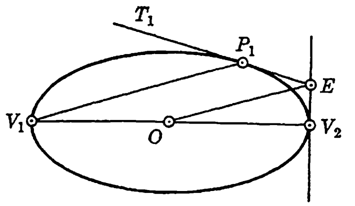


Figure 6.393

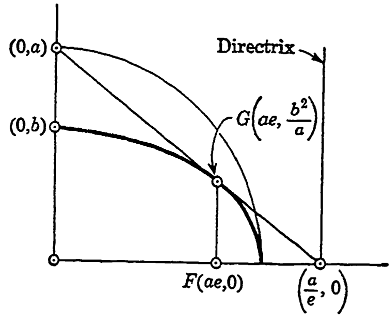


Figure 6.394

15 Figure 6.394 shows a part of the ellipse having, as usual, the standard equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{1}$$

in which $0 < b < a$. Let G be the point in the first quadrant where the ellipse is intersected by the line through the focus F parallel to the directrix. Prove that, as the figure shows, the coordinates of G are $(ae, b^2/a)$. *Remark:* The result can be obtained by use of the fact that the y coordinate of G is the product of e and the distance $(a/e - ae)$ from G to the directrix. It can also be obtained by putting $x = ae$ in (1). In each case, it is necessary to use the relation $b^2 = a^2(1 - e^2)$.

16 Using the notation and results of Problem 15, show that the equation of the tangent to the ellipse at G is $ex + y = a$. *Remark:* This shows that the tangent intersects the y axis at the point $(a,0)$ and intersects the x axis where the directrix does. These results are illustrated in Figure 6.394. The circle which has its center at the center of the ellipse and contains the vertices is called the *major circle* of the ellipse. Thus *the tangent at G intersects the major axis where the directrix does and intersects the minor axis where the major circle does.* This is one of many elegant geometric theorems that have fascinated men for centuries.

17 Supposing that $P(x,y)$ is a point on the ellipse having the standard equation (6.36) which we should now know, use the information in Figure 6.38 (which we should remember) and the distance formula to show that

$$|\overrightarrow{F_1P}| = \frac{a^2 + \sqrt{a^2 - b^2}\,x}{a}, \qquad |\overrightarrow{F_2P}| = \frac{a^2 - \sqrt{a^2 - b^2}\,x}{a}.$$

Then seek a way in which formulas involving eccentricity can be used to obtain the same result. Note that $|\overrightarrow{F_1P}| + |\overrightarrow{F_2P}|$ is what it should be.

18 Study Figure 6.395 and discover the procedure by which the encircled points are determined, and then use the procedure to obtain another encircled point. Prove that the set of points obtained by this procedure lie on an ellipse. *Solution:* If the inner and outer circles have radiuses (this time we use English; the Latin is radii) a and b, and if a line is drawn through the origin making the angle θ with the positive x axis, then the coordinates (x,y) of the resulting encircled point are

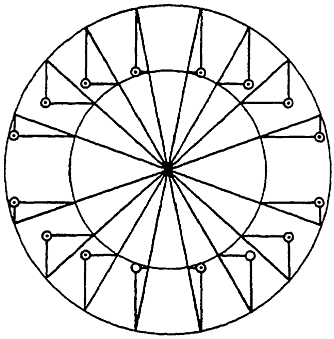


Figure 6.395

$$(1) \qquad x = a\cos\theta, \qquad y = b\sin\theta.$$

From these equations we obtain

$$(2) \qquad \frac{x^2}{a^2} + \frac{y^2}{b^2} = \cos^2\theta + \sin^2\theta = 1,$$

so the point (x,y) lies on an ellipse.

19 Let $0 < b < a$ and let $\omega > 0$. Let a particle move in a plane in such a way that, at each time t, the vector running from the origin of an x, y coordinate system to it is

$$\mathbf{r} = (a\cos\omega t)\mathbf{i} + (b\sin\omega t)\mathbf{j}.$$

Show that its path is an ellipse. Show that it is always accelerated toward the origin and that the magnitude of the acceleration is

$$\omega^2\sqrt{b^2 + (a^2 - b^2)\cos^2\omega t}.$$

Hint: To get started, let $x = a\cos\omega t$, $y = b\sin\omega t$, and observe the result of some dividing and squaring and adding.

20 A point P moves around an ellipse having foci

$$(1) \qquad F_1(-\sqrt{a^2 - b^2}, 0), \qquad F_2(\sqrt{a^2 - b^2}, 0)$$

in such a way that the vector $\mathbf{r}$ running from the origin to P at time t is

$$(2) \qquad \mathbf{r} = (a\cos t)\mathbf{i} + (b\sin t)\mathbf{j}.$$

Show that the vector

$$(3) \qquad \mathbf{v} = -(a\sin t)\mathbf{i} + (b\cos t)\mathbf{j}$$

is a forward tangent to the ellipse at P. Show that

$$(4) \qquad \overrightarrow{F_kP} = (a\cos t + \lambda\sqrt{a^2 - b^2})\,\mathbf{i} + b\sin t\mathbf{j},$$

where $\lambda = 1$ when $k = 1$ and $\lambda = -1$ when $k = 2$, and then show that

$$(5) \qquad |\overrightarrow{F_kP}|^2 = a^2 + 2a\lambda\sqrt{a^2 - b^2}\cos t + \lambda^2(a^2 - b^2)\cos^2 t$$

and hence

$$(6) \qquad |\overrightarrow{F_kP}| = a + \lambda\sqrt{a^2 - b^2}\cos t.$$

Show that

(7) $$|\mathbf{v}| = \sqrt{b^2 + (a^2 - b^2)\sin^2 t}.$$

Show that

(8) $$\overrightarrow{F_kP}\cdot\mathbf{v} = -\lambda\sqrt{a^2 - b^2}\sin t[a + \lambda\sqrt{a^2 - b^2}\cos t].$$

Letting ϕ_k be the angle which the vector $\overrightarrow{F_kP}$ makes with the forward tangent $\mathbf{v}$ at P, show with the aid of (6), (7), and (8) that

(9) $$\cos\phi_k = \frac{\overrightarrow{F_kP}\cdot\mathbf{v}}{|\overrightarrow{F_kP}||\mathbf{v}|} = \frac{-\lambda\sqrt{a^2 - b^2}\sin t}{\sqrt{b^2 + (a^2 - b^2)\sin^2 t}}$$

Show that $y = b\sin t$ and that multiplying the numerator and denominator of the last member of (9) by b gives the formula

(10) $$\cos\phi_k = \frac{-\lambda\sqrt{a^2 - b^2}\,y}{\sqrt{b^4 + (a^2 - b^2)y^2}}.$$

Remark: These remarkable formulas yield the famous reflection property of ellipses. Since $\lambda = 1$ when $k = 1$ and $\lambda = -1$ when $k = 2$, the numbers $\cos\phi_1$ and $\cos\phi_2$ in (10) differ only in sign. This implies that the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$ make supplementary angles with the forward tangent $\mathbf{v}$ to the ellipse at P and hence that the lines F_1P and F_2P make equal angles with the normal to the ellipse at P. This implies that if light or something else goes in a line from F_1 and is reflected from the ellipse in such a way that the angle θ_r of reflection is equal to the angle θ_i of incidence, then its path leads to F_2. Moreover, because of the string property of the ellipse, radiation leaving F_1 at the same time but in different directions will arrive simultaneously (or in phase) at F_2.

21 The reflection property of ellipses is a consequence of another interesting geometric property of ellipses. Let the line T of Figure 6.396 be tangent at P to the ellipse having foci at F_1 and F_2. Let H_1 and H_2 be the reflections in T of F_1 and F_2; this means that T is the perpendicular bisector of the line segments F_1H_1 and F_2H_2. Then, as indicated by the figure, the line segments F_1H_2 and F_2H_1 intersect at P. To prove this fact, let A be any point on T different from P and let B be the point at which the line segment F_1A intersects the ellipse. Then (why?)

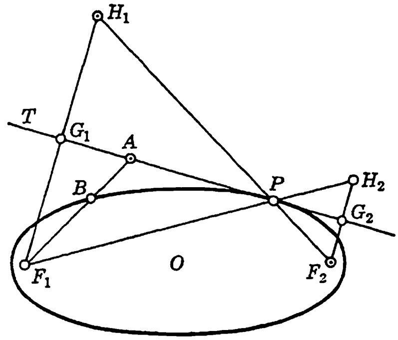


Figure 6.396

(1) $$|\overrightarrow{F_1B}| + |\overrightarrow{BF_2}| < |\overrightarrow{F_1A}| + |\overrightarrow{AF_2}|$$

so (why?)

(2) $$|\overrightarrow{F_1P}| + |\overrightarrow{PF_2}| < |\overrightarrow{F_1A}| + |\overrightarrow{AF_2}|$$

and (why?)

(3) $$|\overrightarrow{F_1P}| + |\overrightarrow{PH_2}| < |\overrightarrow{F_1A}| + |\overrightarrow{AH_2}|.$$

Hence (why?) P must lie on the line F_1H_2. Similarly (why?) P must lie on the line F_2H_1. This proves that the lines F_1H_2 and F_2H_1 intersect at P. Therefore (why?) the angles F_1PG_1 and F_2PG_2 are equal.

22 This problem involves a nested family of curves containing many ellipses and one circle. Let a be a given positive number. For each positive number b, the graph of the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

is an ellipse (or a circle) which intersects the x axis at the points $(-a,0)$ and $(a,0)$. It is easy to generate interest in these graphs by sketching some of them. When $0 < x_1 < a$, the line having the equation $x = x_1$ intersects each of these graphs at two points. Prove that the tangents to these graphs at these points all intersect at a point on the x axis. *Solution:* Using the result of Problem 8 (or working out the result again) shows that each tangent intersects the x axis at the point $(a^2/x_1, 0)$ which does not depend upon b.

23 For each θ for which $0 < \theta < \pi/2$ and $\theta \neq \pi/4$, the graph of the equation

$$\frac{x^2}{\sin^2\theta} + \frac{y^2}{\cos^2\theta} = 1$$

is an ellipse. Sketch several of these graphs.

24 The members of a family of confocal ellipses have foci at the points $(-1,0)$ and $(1,0)$. Sketch good approximations to six of them. *Suggestion:* Do not work too long on an easy problem. Select a point $(0,b)$ and make Figure 6.38 tell you what the a in $(a,0)$ must be.

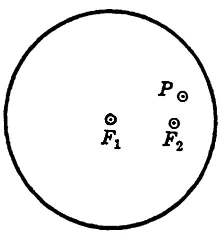
25 Supposing that an ellipse E is given, tell how our little sister can use her new drawing equipment to locate the center, the axes, and the foci of E. *Hints:* The mid-points of parallel chords of E lie on a line through the center C of E. It is easy to choose r such that the circle of radius r having center at C intersects the ellipse in four points.

26 A rod of length L has a red end, a blue end, and a pink dot P at distance q from the red end. Suppose that $0 < q < L$. Show that if the red end is on the x axis and the blue end is on the y axis, then (except when $q = \frac{1}{2}$) P must lie on an ellipse. *Ans.:* The equation of the ellipse is

$$\frac{x^2}{(L-q)^2} + \frac{y^2}{q^2} = 1.$$

27 Let F be a point which is inside a circle C but is not the center of C. A little preliminary sketching shows that the set S of points equidistant from C and F looks much like an ellipse having a focus at F. What are the facts? *Solution:* As in Figure 6.397, let F_1 be the center of the circle and let F_2 be the point F. The condition that P be equidistant from C and F_2 can be put in the form

Figure 6.397



$$(1) \qquad r - |\overrightarrow{F_1P}| = |\overrightarrow{F_2P}|$$

where r is the radius of the circle. Before undertaking to express this result in terms of coordinates, we can suddenly realize that it is almost familiar. If we put it in the form

$$|\overrightarrow{F_1P}| + |\overrightarrow{F_2P}| = r,$$

we see an expression of the string property of an ellipse. Therefore, S is an ellipse having foci at the center of the circle and the given point F.

28 Sketch a figure like Figure 6.31 in which the distance from V to V_2 is about 10 or 20 times the distance from V to V_1. Try to decide whether the center of the ellipse is on the axis of the cone.

29 *Remark:* Figure 6.31 presents an interesting problem in plane geometry. When two lines through V and a line π are given, we can use a ruler and compass to construct the circles of the figure. We can then wonder whether we can give a simple proof that $|\overrightarrow{V_1F_1}| = |\overrightarrow{F_2V_2}|$ without use of the cone and planes and spheres that were employed in the proof in the text. Perhaps consideration of this problem will increase our respect for the methods that were employed.

6.4 Hyperbolas

Geometric properties of hyperbolas can be extracted from Figure 6.41, which, like some preceding ones, shows a cone having a vertical axis. The axis lies in the plane of the paper. The plane π intersects the cone in a hyperbola H of which the two points V_1 and V_2 (vertices, in fact) lie in the plane of the paper. The upper circle represents a sphere, in the upper nappe of the cone, which is tangent to the cone at the points of a circle which determines the plane π_1 and is tangent to π at F_1. As we saw in Section 6.2, π_1 and π intersect in a line L_1 and, moreover, F_1 is a focus and L_1 is a directrix of the hyperbola H. The lower circle represents a sphere, in the lower nappe of the cone, which is tangent to the cone at the points of a circle which determines the plane π_2 and is tangent to π at F_2. The planes π_2 and π intersect in a line L_2, and the same procedure which was applied to F_1 and L_1 shows that F_2 is another focus and L_2 is another directrix of the hyperbola H. Thus H has two foci and two directrices.

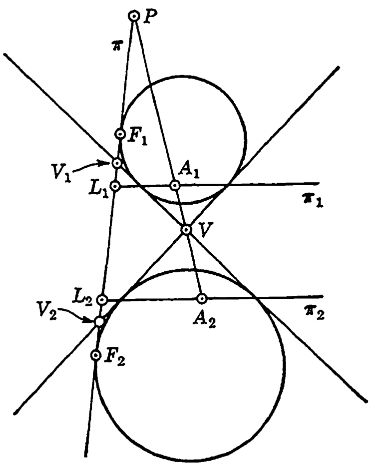


Figure 6.41

Hyperbolas have a string property which involves the *difference* (not sum) of the distances $|\overrightarrow{F_1P}|$ and $|\overrightarrow{F_2P}|$ from foci to points on the hyperbolas. Let P be V_1 or any other point on the upper branch of H. The line VP lies on the cone and is tangent to the lower and upper spheres at points

A_2 and A_1. With the aid of Figure 6.41 and familiar facts about vectors tangent to spheres at their tips, we see that

$$(6.42)\qquad |\overrightarrow{F_2P}| - |\overrightarrow{F_1P}| = |\overrightarrow{A_2P}| - |\overrightarrow{A_1P}| = |\overrightarrow{A_2A_1}|.$$

Wherever the point P may be on the upper branch of H, the number $|\overrightarrow{A_2A_1}|$ is the constant sum of the slant heights of two conical segments having their vertex at V and their bases in the planes π_1 and π_2; in fact if d is the distance between the planes π_1 and π_2, then $|\overrightarrow{A_2A_1}| = d/\cos\alpha$, where α is the angle at the vertex of the cone. In particular, letting $P = V_1$ shows that

$$|\overrightarrow{F_2V_1}| - |\overrightarrow{V_1F_1}| = |\overrightarrow{A_2A_1}|$$

and hence

$$(6.421)\qquad |\overrightarrow{F_2V_2}| + |\overrightarrow{V_2V_1}| - |\overrightarrow{V_1F_1}| = |\overrightarrow{A_1A_2}|.$$

In case P lies on the lower branch of H, we can reverse the roles of the subscripts 1 and 2 to obtain the formulas

$$|\overrightarrow{F_1P}| - |\overrightarrow{F_2P}| = |\overrightarrow{A_1A_2}|,\quad |\overrightarrow{F_1V_2}| - |\overrightarrow{F_2V_2}| = |\overrightarrow{A_1A_2}|$$

and

$$(6.422)\qquad |\overrightarrow{F_1V_1}| + |\overrightarrow{V_2V_1}| - |\overrightarrow{V_2F_2}| = |\overrightarrow{A_1A_2}|.$$

Adding the formulas (6.421) and (6.422) shows that $2|\overrightarrow{V_2V_1}| = 2|\overrightarrow{A_2A_1}|$ and hence that the first of the formulas

$$(6.423)\qquad |\overrightarrow{A_2A_1}| = |\overrightarrow{V_2V_1}|,\quad |\overrightarrow{F_2V_2}| = |\overrightarrow{F_1V_1}|$$

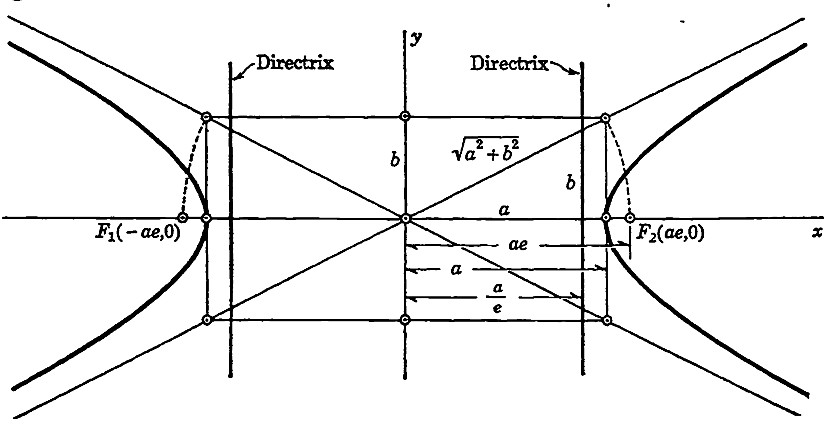
is valid. The second formula is a consequence of the first and (6.421). All this implies the string property of the hyperbola having foci at F_1 and F_2 and vertices at V_1 and V_2. If P is on the hyperbola, then

$$(6.43)\qquad |\overrightarrow{F_1P}| - |\overrightarrow{F_2P}| = \pm|\overrightarrow{V_1V_2}|,$$

the plus sign being required when P is on one branch and the minus sign being required when P is on the other branch.

Figure 6.44 shows a hyperbola and also some numerical dimensions

Figure 6.44



which display information from the paragraph containing (6.281). Even though the equation of the hyperbola of the figure has already been derived, we become acquainted with useful ideas by using the intrinsic string property

$$(6.45) \qquad |\overrightarrow{F_1P}| - |\overrightarrow{F_2P}| = \pm 2a$$

to derive the equation. Letting $F_1(-ae,0)$ and $F_2(ae,0)$ be located on the x axis with the origin midway between them as in Figure 6.44, we use the string property to obtain the equation

$$\sqrt{(x+ae)^2+y^2} - \sqrt{(x-ae)^2+y^2} = \pm 2a.$$

Transposing the second term, squaring, and simplifying give

$$\pm\sqrt{(x-ae)^2+y^2} = a - ex.$$

Squaring and simplifying this give

$$(e^2-1)x^2 - y^2 = a^2(e^2-1),$$

and dividing by the right side gives the standard form

$$(6.46) \qquad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

where $b^2 = a^2(e^2-1)$, or

$$(6.47) \qquad ae = \sqrt{a^2+b^2}.$$

As was the case for ellipses, it is of interest to see that memorization of only a few details enables us to find the graph, foci, directrices, and asymptotes of (6.46) when a and b are given constants. Putting $y = 0$ shows that the points $V_1(-a,0)$ and $V_2(a,0)$ lie on the graph. The line through these points is the *principal axis* of the hyperbola, and the foci lie on it. Putting $x = 0$ shows that the hyperbola contains no points on the y axis, but we mark the points $(0,b)$ and $(0,-b)$ anyway. The line through these points is the *conjugate axis* of the hyperbola. The next steps are to draw the box having horizontal and vertical sides containing the four points and then draw the diagonals of the box. These diagonals are, as we showed in Problem 7 of Section 3.3 and shall quickly show again, asymptotes of the hyperbola. A reasonable approximation to the hyperbola can now be sketched very quickly; it is tangent to the box at the two vertices we have found, and a pencil point which traces a part of the hyperbola becomes steadily closer to the asymptote as it recedes from a vertex. For the part of the hyperbola in the first quadrant, this result comes from the fact that

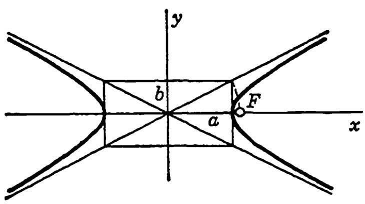


Figure 6.48

$$(6.481) \qquad \frac{b}{a}x - \frac{b}{a}\sqrt{x^2-a^2} = \frac{b}{a}\frac{a^2}{x+\sqrt{x^2-a^2}},$$

and for other quadrants we use symmetry. We can use the Pythagoras theorem to calculate the distance from the center to a focus if we remember Figure 6.48, a simplified version of Figure 6.44, which shows that if we fix one point of a compass at the origin, then the circle through the corners of the box meets the principal axis at the foci. If we remember that this distance is ae (as it also is for ellipses), then we can calculate the eccentricity e. Finally, if we remember that the distance from the center to a directrix is a/e (as it also is for ellipses), we can calculate this distance. The numbers ae and a/e are still the key numbers.

Problems 6.49

1 For each of the following pairs of values of a and b, sketch the hyperbola having the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

find the eccentricity, find the foci (give coordinates), find the directrices (give equations), and find the asymptotes (give equations). Try to cultivate the ability to use the Pythagoras theorem and key numbers without use of books or notes. Check numerical results by use of the fact that the distance p from a focus to its directrix must satisfy the equation $e^2p^2 = b^2(e^2 - 1)$.

(*a*) $a = 5, b = 2$ (*b*) $a = 2, b = 5$ (*c*) $a = b = 1$

2 The equation

$$\frac{y^2}{2^2} - \frac{x^2}{3^2} = 1$$

differs from equations of hyperbolas having their principal axes and foci on the x axis because the roles of x and y are interchanged. Nevertheless, plot the points on the graph obtained by setting $x = 0$ and then $y = 0$ and then draw the helpful box and sketch the hyperbola. Then proceed to find the eccentricity, foci, directrices, and asymptotes. Repeat the process when 2 and 3 are respectively replaced by

(*a*) 2 and 5 (*b*) 5 and 2 (*c*) 1 and 1

Remark: The graphs of the equations

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad \frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$$

are hyperbolas having the same helpful box and the same asymptotes. Each hyperbola is said to be the *conjugate* of the other.

3 Sketch graphs of the equations

(*a*) $\dfrac{(x-1)^2}{5^2} - \dfrac{(y-2)^2}{2^2} = 1$ (*b*) $\dfrac{(y-2)^2}{2^2} - \dfrac{(x-1)^2}{5^2} = 1$

Remark: Good clean starts are made by setting $y = 2$ or $x = 1$ and remembering that squares of our numbers (which are always real numbers) are never negative.

4 The two vertices and one focus of a hyperbola are given. Describe a procedure by which it is possible to draw the asymptotes without using equations.

5 For basic studies of hyperbolas, we place the hyperbolas upon coordinate systems in such a way that their equations have the standard form

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

Supposing $P_1(x_1,y_1)$ is, as in Figure 6.491, a point on the hyperbola that is not a

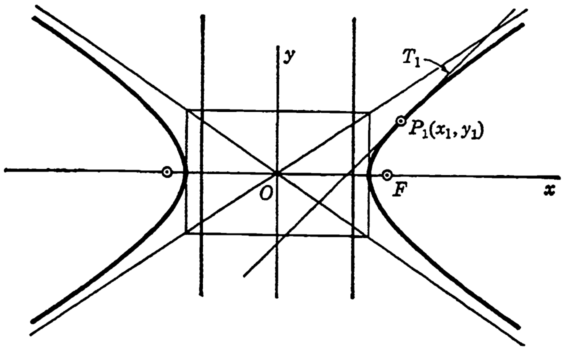


Figure 6.491

vertex, find the equation of the line T_1 tangent to the hyperbola at P_1. *Ans.:*

$$\frac{x_1x}{a^2} - \frac{y_1y}{b^2} = 1.$$

6 Find the coordinates of the points at which the tangent T_1 of Problem 5 intersects the coordinate axes. *Ans.:*

$$\left(\frac{a^2}{x_1}, 0\right), \quad \left(0, -\frac{b^2}{y_1}\right).$$

7 Find the coordinates of the points where the tangent T_1 of Problem 5 intersects the lines through the foci perpendicular to the x axis. *Ans.:*

$$\left(-ae, -\frac{b^2}{y_1}\left(1 + \frac{ex_1}{a}\right)\right), \quad \left(ae, -\frac{b^2}{y_1}\left(1 - \frac{ex_1}{a}\right)\right).$$

8 Find the coordinates of the points where the tangent T_1 of Problem 5 intersects the directrices of the hyperbola. *Ans.:*

$$\left(-\frac{a}{e}, -\frac{b^2}{y_1}\left(1 + \frac{x_1}{ae}\right)\right), \quad \left(\frac{a}{e}, -\frac{b^2}{y_1}\left(1 - \frac{x_1}{ae}\right)\right)$$

9 Let the line T_1 tangent to a hyperbola at P_1 intersect the directrix at Q_1 and let F be the focus corresponding to the directrix. With the aid of Problem 8 and the fact that $b^2 = a^2(e^2 - 1)$, prove that the line FQ_1 is perpendicular to the line FP_1.

10 Sketch a figure somewhat like Figure 6.491 and observe that the line L_1 from a focus F perpendicular to the line T_1 tangent to the hyperbola at P_1 and the line OP_1 from the center of the world to P_1 seem to intersect at a point on the directrix. Prove that each of these lines does intersect the directrix at the point $(a/e, ay_1/ex_1)$.

11 Figure 6.492 illustrates the fact that the line V_1P_1 from a vertex of a hyperbola to a point P_1 on the hyperbola is parallel to the line OE from the center

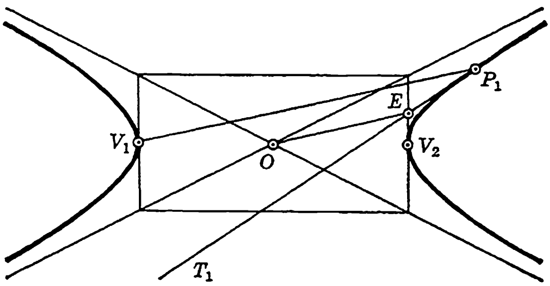


Figure 6.492

of the hyperbola to the point E where the tangent T_1 at P_1 intersects the tangent at the other vertex. Show that the coordinates of E are $\left(a, \frac{b^2}{y_1}\left(\frac{x_1}{a} - 1\right)\right)$ and prove the fact.

12 Figure 6.493 shows a part of the hyperbola having, as usual, the standard equation

$$\text{(1)} \qquad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

Let G be the point in the first quadrant where the hyperbola is intersected by the line through the focus F parallel to the directrix. Prove that, as the figure shows, the coordinates of G are $(ae, b^2/a)$. *Remark:* A relation among a, b, and e is needed.

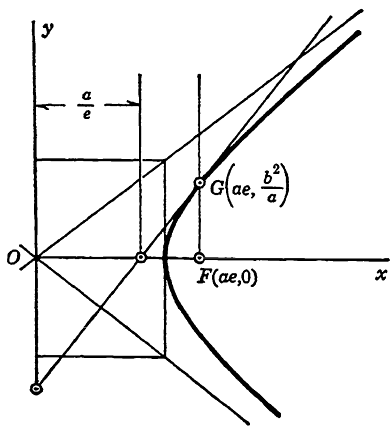


Figure 6.493

13 Using the notation and results of Problem 12, show that the equation of the tangent to the hyperbola at G is $ex - y = a$. *Remark:* This shows that the tangent to the hyperbola at G intersects the x axis (the transverse axis of the hyperbola) where a directrix does and intersects the y axis (the conjugate axis of the hyperbola) at a point on the circle which has its center at the center of the hyperbola and includes the vertices of the hyperbola.

14 Let $P_1(x_1,y_1)$ be a point on a hyperbola having the standard equation $x^2/a^2 - y^2/b^2 = 1$. Find the coordinates of the points at which the tangent T_1 to the hyperbola at P_1 intersects the asymptotes of the hyperbola. *Ans.:* The intersections with the asymptotes $y = (b/a)x$ and $y = -(b/a)x$ are respectively

$$\left(\frac{a^2b}{bx_1 - ay_1}, \frac{ab^2}{bx_1 - ay_1}\right), \qquad \left(\frac{a^2b}{bx_1 + ay_1}, \frac{-ab^2}{bx_1 + ay_1}\right).$$

Remark: Since $(bx_1 - ay_1)(bx_1 + ay_1) = b^2x_1^2 - a^2y_1^2 = a^2b^2$, the denominators are all positive or all negative.

15 Prove that if a line T_1 is tangent to a hyperbola at P_1, then P_1 lies midway between the points at which T_1 intersects the asymptotes of the hyperbola.

16 Find the area of the triangular region bounded by the tangent T_1 and the asymptotes of the hyperbola of Problem 5. *Ans.: ab.*

17 Substantial information about the geometry of the hyperbola having the standard equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{1}$$

is concealed in Figure 6.494. The figure shows the auxiliary rectangle whose vertices lie on the asymptotes, and also the asymptotes. As we know, the circle

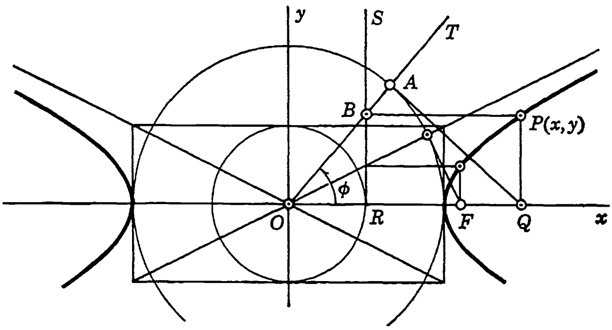


Figure 6.494

with center at O and radius $\sqrt{a^2 + b^2}$ intersects the x axis at the foci. The smaller circles have radii a and b. The line RS is tangent to the circle of radius b at the point $(b,0)$ where this circle intersects the positive x axis. Upon the basic part of the figure which has been described, we can heap more construction. Let OT be a ray (or half-line) from the origin which makes with the positive x axis an angle ϕ for which $0 < \phi < \pi/2$. Let A be the point where the ray OT intersects the circle having radius a, and let Q be the point where the tangent to the circle at A intersects the x axis. Let B be the point where the ray OT intersects the line RS. *The point $P(x,y)$ where the horizontal line through B intersects the vertical line through Q lies on the hyperbola* because

$$\frac{x}{a} = \frac{|\overrightarrow{OQ}|}{|\overrightarrow{OA}|} = \sec\phi = \frac{1}{\cos\phi}, \qquad \frac{y}{b} = \frac{|\overrightarrow{RB}|}{|\overrightarrow{OR}|} = \tan\phi = \frac{\sin\phi}{\cos\phi}$$

and hence

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \sec^2\phi - \tan^2\phi = \frac{1 - \sin^2\phi}{\cos^2\phi} = 1.$$

Aristotle told Alexander the Great that there is no royal road to geometry, but it is quite easy to see what we have done and it is even quite easy to see some of the implications of what we have done. We have proved the theorem given above in italics. We have described a simple ruler-and-compass procedure for construction of points P on the hyperbola. We can use the construction to produce 4 or even 40 points on the hyperbola, but we cannot thus produce all of the points on the hyperbola and "construct the hyperbola." Our result enables us to write the vector $\mathbf{r}$ from the origin O to a point P on the hyperbola in the form

$$\mathbf{r} = a \sec \phi \mathbf{i} + b \tan \phi \mathbf{j},$$

but ϕ is an "eccentric angle" which is not the angle between $\mathbf{r}$ and $\mathbf{i}$. Persons having inherent interest in geometry may, as an extramural excursion, consider the special case in which the ray OT is the part of the asymptote in the first quadrant. Considerable geometry is associated with the fact that, in this case, the point Q is the focus F. To emphasize the fact that consideration of these things need not be tedious, we observe that if $\tan \phi = b/a$, then $\cos \phi = a/\sqrt{a^2 + b^2}$, so $|\overrightarrow{OQ}| = \sqrt{a^2 + b^2}$ and hence Q must be F.

18 A glance at Figure 6.494 suggests that the directrices may be the lines determined by the points at which the asymptotes intersect the circle containing the vertices. Prove that it is so.

19 Find the distance from a focus to a directrix of a hyperbola having the standard equation $x^2/a^2 - y^2/b_2 = 1$. *Ans.:* $b^2/\sqrt{a^2 + b^2}$.

20 Let a hyperbola have the standard equation $x^2/a^2 - y^2/b^2 = 1$. Let F be the focus and let D be the directrix in the right half-plane. Let $P_1(x_1,y_1)$ be a point on the hyperbola in the right half-plane. Show that

$$|\overrightarrow{FP_1}| = ex_1 - a.$$

Let A be the point in which the directrix D is intersected by the line through P_1 parallel to an asymptote. Show that

$$|\overrightarrow{AP_1}| = ex_1 - a.$$

21 The functions in the right members of the formulas

$$\sinh t = \frac{e^t - e^{-t}}{2}, \qquad \cosh t = \frac{e^t + e^{-t}}{2}$$

are used often enough to justify introductions of special names and symbols for them. They are called hyperbolic functions, and the h in sinh t and cosh t tells us that the functions are hyperbolic sines and hyperbolic cosines. Show that if $x = a \cosh t$ and $y = b \sinh t$, then

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

This shows that if a particle P moves in the xy plane in such a way that, for each t, the vector $\mathbf{r}$ running from the origin to P is

$$\mathbf{r} = (a \cosh kt)\mathbf{i} + (b \sinh kt)\mathbf{j},$$

then P traverses a branch of a hyperbola. Calculate the acceleration $\mathbf{a}$ and show that $\mathbf{a} = k^2\mathbf{r}$.

22 Copy Figure 6.48 and let a particle P move along the right branch of the hyperbola in such a way that the vector $\mathbf{r}$ running from the origin to P is, at each time t,

$$\mathbf{r} = (a \cosh t)\mathbf{i} + (b \sinh t)\mathbf{j}. \tag{1}$$

Show that the vector

$$\mathbf{v} = (a \sinh t)\mathbf{i} + (b \cosh t)\mathbf{j} \tag{2}$$

is a forward tangent to the hyperbola at P. Use the figure to show that, where $\lambda_k = +1$ when $k = 1$ and $\lambda_k = -1$ when $k = 2$,

$$\overrightarrow{F_kP} = (a \cosh t + \lambda \sqrt{a^2 + b^2})\, \mathbf{i} + (b \sinh t)\mathbf{j} \tag{3}$$

and then, with the aid of the simple fact that $1 + \sinh^2 t = \cosh^2 t$, show that

$$|\overrightarrow{F_kP}| = \sqrt{a^2 + b^2} \cosh t + \lambda a. \tag{4}$$

Letting ϕ_k be the angle which the vector $\overrightarrow{F_kP}$ makes with the forward tangent $\mathbf{v}$ at P, show that

$$|\mathbf{v}| \cos \phi_k = \frac{|\overrightarrow{F_kP}| \cdot \mathbf{v}}{|\overrightarrow{F_kP}|} = \sqrt{a^2 + b^2} \sinh t \tag{5}$$

and hence that

$$\cos \phi_k = \frac{\sqrt{a^2 + b^2}\, y}{\sqrt{b^4 + (a^2 + b^2)y^2}}. \tag{6}$$

Remark: The formula (6) shows that $\phi_1 = \phi_2$. This means that the lines drawn from the foci to a point P on a hyperbola make equal angles with the tangent to the hyperbola at P.

23 An ellipse and a hyperbola are confocal, that is, have the same foci. Prove that they are orthogonal where they intersect. *Remark:* One way to prove the result is to use the fact that if F_1 and F_2 are the foci and P is a point of intersection of the ellipse and hyperbola, then the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$ make equal angles with the tangent to the hyperbola at P and also make equal angles with the normal to the ellipse at P. The miniature Figure 6.495 can help us remember the fact.

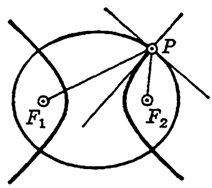


Figure 6.495

24 The members of a family of confocal hyperbolas have their foci at the points $F_1(-1,0)$ and $F_2(1,0)$. Sketch good approximations to graphs of six of them. *Suggestion:* Do not work too long on an easy problem. Make a figure which tells you where the corners of the handy boxes must be, and then sketch hyperbolas at the rate of two per minute. *Remark:* There is a reason for knowing about these things. If the x axis is an impenetrable barrier except for an opening on the interval $-1 \leqq x \leqq 1$, your curves are paths followed by particles of a liquid or gas (or perhaps by football spectators) that stream through the gap.

25 Let $0 < b < a$. Prove that when $\lambda < b^2$, the ellipse having the equation

$$\frac{x^2}{a^2-\lambda} + \frac{y^2}{b^2-\lambda} = 1$$

has its foci at the points $(\pm\sqrt{a^2-b^2}, 0)$. Prove that, when $b^2 < \lambda < a^2$, the hyperbola having the same equation has the same foci.

26 Let H be a given hyperbola. Describe an elementary geometric procedure for locating the center, the axes, and the asymptotes of H. *Hints:* Take clues from the hints of Problem 25 of Problems 6.39. Then use the fact that if the equation of H has the standard form, then $y = \pm b$ when x is the length of a diagonal of a square of which the sides have length a.

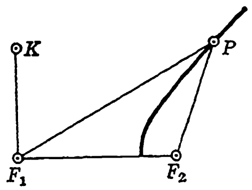


Figure 6.496

27 We have seen string mechanisms for construction of parabolas and ellipses, and we need not slight hyperbolas. Let a red string run from a knot K so that it passes below a tack F_1 to an end tied to a pencil point at P as in Figure 6.496. Let a white string run from the same knot K so that it passes below the tack F_1 and then under a tack F_2 to an end tied at P. Show that if P is held in such a way that both strings are kept taut, then P will trace a part of a branch of a hyperbola as the knot K is pulled away from F_1. *Hint:* Justify and use the fact that if the red and white strings have lengths R and W, then

$$|\overrightarrow{F_1K}| = R - |\overrightarrow{F_1P}| = W - |\overrightarrow{F_1F_2}| - |\overrightarrow{F_2P}|$$

at all times.

28 Sound travels with speed s, and a bullet travels with speed b from a gun at $(-h,0)$ to a target at $(h,0)$ in an xy plane. Where in the plane can the boom of the gun and the ping of the target be heard simultaneously? *Ans.:* At points (x,y) for which

$$\frac{\sqrt{(x+h)^2+y^2}}{s} = \frac{\sqrt{(x-h)^2+y^2}}{s} + \frac{2h}{b}$$

or

$$\sqrt{(x+h)^2+y^2} = \sqrt{(x-h)^2+y^2} + 2qh$$

where $q = s/b$ and hence $q = 1/M$ where M is the *Mach number* of the bullet. In case $M < 1$ and $q > 1$, there are no such points because the length of one side of a triangle cannot exceed the sum of the lengths of the other two sides. In case $M = 1$ and $q = 1$, the required points are those for which $y = 0$ and $x \geqq h$. In case $M > 1$ (so the speed of the bullet is supersonic) and $q < 1$, the required points lie on the right-hand branch of the hyperbola having the equation

$$\frac{x^2}{q^2h^2} - \frac{y^2}{h^2(1-q^2)} = 1.$$

29 We have seen that if S_1 is the set of points equidistant from a line and a circle having its center on the line, then S_1 is the sum or union of *two* parabolas. We have also seen that if S_2 is the set of points equidistant from a circle and a point inside the circle which is not the center of the circle, then S_2 is *one* ellipse. Complete the story by showing that if S_3 is the set of points equidistant from a circle and a point outside the circle, then S_3 is *half* (one branch) of a hyperbola.

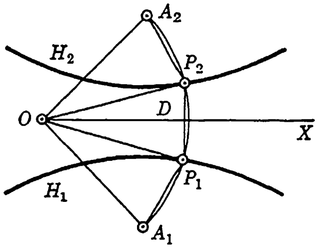


Figure 6.497

30 Let the angle A_1OA_2 of Figure 6.497 be a given angle between 0 and π. Everyone starting study of calculus must know that it is very easy to give a ruler-and-compass construction of the arc A_1A_2 of a circle having its center at O and of the line OX which bisects the angle A_1OA_2. We may be too busy with other affairs to learn how the result can be proved, but we should nevertheless know that it has been proved to be impossible to give a ruler-and-compass construction of points P_1 and P_2 such that the lines OP_1 and OP_2 trisect the given angle. Pappus of Alexandria, who flourished about A.D. 300, trisected the angle by means of hyperbolas. Let H_1 be the branch near A_1 of the hyperbola having eccentricity 2 for which A_1 is a focus and OX is a directrix, and let P_1 be the point at which H_1 intersects the circular arc A_1A_2. A branch H_2 of a similar hyperbola having a focus at A_2 intersects the circular arc at P_2. The definition of eccentricity and the symmetry imply that

$$|\overrightarrow{A_1P_1}| = 2|\overrightarrow{P_1D}| = |\overrightarrow{P_1P_2}| = 2|\overrightarrow{DP_2}| = |\overrightarrow{P_2A_2}|.$$

Thus the three chords A_1P_1, P_1P_2, P_2A_2 have equal lengths, and it follows that the two lines OP_1 and OP_2 trisect the given angle A_1OA_2. In modern mathematics, it is important to know why the Pappus construction does *not* provide a ruler-and-compass construction of P_1 and P_2. It is possible to produce hordes of points on the hyperbolas with rulers and compasses, but it is impossible to produce all of them. In particular, it is impossible to prescribe rules for ruler-and-compass construction of the particular points where the hyperbolas intersect the circular arc. This matter will again be brought to our attention in Problems 10.19.

31 Let $P_1(x_1,y_1)$ be a point on the rectangular hyperbola having the equation

$$x^2 - y^2 = a^2. \tag{1}$$

In terms of x_1 and y_1, find the coordinates of the point (x,y) at which the tangent to the hyperbola at P_1 intersects the line through the origin perpendicular to this tangent. Finally, find an equation which x and y must satisfy by eliminating x_1 and y_1 from your equations and the equation $x_1^2 - y_1^2 = a^2$. *Ans.:* The first two of the required equations are

$$x = \frac{a^2x_1}{x_1^2 + y_1^2}, \qquad y = -\frac{a^2y_1}{x_1^2 + y_1^2}. \tag{2}$$

We can square and add, and then square and subtract, to obtain

$$x^2 + y^2 = \frac{a^4}{x_1^2 + y_1^2}, \qquad x^2 - y^2 = \frac{a^6}{(x_1^2 + y_1^2)^2}. \tag{3}$$

It is then easy to obtain the final answer

$$(x^2 + y^2)^2 = a^2(x^2 - y^2). \tag{4}$$

Remark: The graph of (4) is called the *lemniscate* of Bernoulli. In Problem 5 of Section 10.1 we shall find that its polar coordinate equation is $\rho^2 = a^2 \cos 2\phi$. Its graph appears in Figure 10.171.

32 This long problem should be very easy. The problem is to read about graphs of equations of the form

$$(1) \qquad Ax^2 + By^2 + Cx + Dy + E = 0$$

and to verify correctness of each assertion that is not completely obvious. In case $AB > 0$, completion of squares enables us to put the equation in the form

$$(2) \qquad A(x - x_0)^2 + B(y - y_0)^2 = K,$$

where A and B have the same sign, and the graph is an ellipse or a circle or a single point or the empty set. In case $AB < 0$, the equation (1) can be put in the form (2), where A and B have opposite signs, and the graph is either a hyperbola (when $K \neq 0$) or a pair of intersecting lines (when $K = 0$). In case $A = B = 0$, the graph is a line or the empty set (in case $C = D = 0$ and $E \neq 0$) or the whole plane (in case $C = D = E = 0$). So far we have covered all situations except those in which one of A and B is zero and the other is not. The case in which $A = 0$ and $B \neq 0$ being analogous, we suppose henceforth that $A \neq 0$ and $B = 0$ and that (1) has been reduced to

$$x^2 + C_1x + D_1y + E_1 = 0.$$

In case $D_1 \neq 0$, completion of a square gives the equation $y - y_0 = K(x - x_0)^2$ and the graph is a parabola. In case $D_1 = 0$, the graph consists of two vertical lines or a single vertical line or the empty set. The results may be summarized. Depending upon the values of A, B, C, D, E, the graph of (1) may be a conic (an ellipse, a hyperbola, a parabola, a circle, two intersecting lines, a single line, or a single point) and it may be a set which is not a conic (two distinct parallel lines, a plane, or the empty set). Still more information is available. The equation (1) is said to have *elliptic type* when $AB > 0$ even though there are cases in which the graphs are circles or points or empty sets. Similarly, (1) is said to have *hyperbolic type* when $AB < 0$ and to have *parabolic type* when A and B are not both 0 but $AB = 0$.

33 Let A and B be nonzero constants not both negative. The graph of the equation

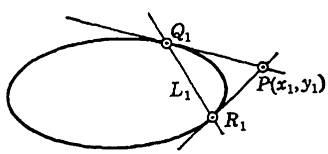
$$(1) \qquad Ax^2 + By^2 = 1$$

is then a central conic K (a circle or ellipse or hyperbola) having its center at the origin O. When $P_1(x_1,y_1)$ is a point different from the origin, the equation

$$(2) \qquad Ax_1x + By_1y = 1$$

is the equation of a line L_1 which is called the *polar line* (with respect to the conic K) of the point P_1. Moreover, the point P_1 is called the *polar point* of the line L_1. These polar points and lines are very important in some parts of mathematics, and we can start learning about them. If the point P_1 lies on the conic K, then L_1, the polar line of P_1, is the line tangent to K at P_1. When P_1 is not on K, matters are much more interesting. Suppose first that P_1 lies "outside" the conic K so that there exist two points Q_1 and R_1 on K such that the tangents to K at Q_1 and R_1 contain P_1. Then, as Figure 6.498

Figure 6.498



can help us remember, *the polar line of* P_1 *is the line* L_1 *containing the two points* Q_1 *and* R_1. A standard simple proof of this result is as amazing as the result. If the coordinates of the point Q_1 on K are (x_2,y_2), then the equation of the tangent to K at Q_1 is

(3) $$Ax_2x + By_2y = 1.$$

The fact that $P_1(x_1,y_1)$ lies on this line implies that

(4) $$Ax_2x_1 + By_2y_1 = 1.$$

This implies that (2) holds when $x = x_2$ and $y = y_2$ and hence that Q_1 lies on the line L_1. When we have scrutinized this sleight of hand closely enough to understand it, we can see that the same argument proves that R_1 is on L_1. Thus we have given geometric interpretations to the polar line L_1 of P_1 for the case in which P_1 lies on K and for the case in which P lies outside K. Our next geometric interpretation of the polar line L_1 of P_1 is equally applicable to the case in which P_1 lies outside the conic (Figure 6.499), the case in which P_1 lies on the conic (Figure 6.4991), and the case in which P_1 lies inside the conic (Figure 6.4992). Let L be a line through P_1 which intersects the conic at two points $P_2(x_2,y_2)$ and $P_3(x_3,y_3)$. The tangents to the conic at P_2 and P_3 have the equations

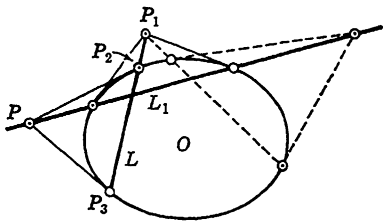


Figure 6.499

(5) $$Ax_2x + By_2y = 1$$
(6) $$Ax_3x + By_3y = 1.$$

Except for the special line L which contains the center O, these tangents intersect

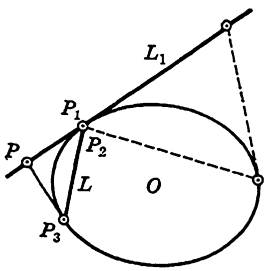


Figure 6.4991

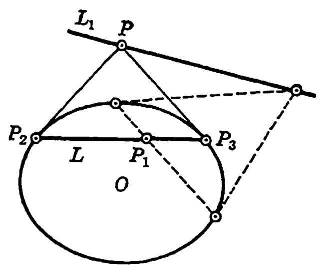


Figure 6.4992

at the point $P(x,y)$ for which the two equations (5) and (6) are both satisfied. Since P_1 lies on the line L containing P_2 and P_3, there exists a constant λ such that

(7) $$x_1 = x_2 + \lambda(x_3 - x_2) = (1 - \lambda)x_2 + \lambda x_3$$
(8) $$y_1 = y_2 + \lambda(y_3 - y_2) = (1 - \lambda)y_2 + \lambda y_3.$$

Multiplying (5) and (6) by $(1 - \lambda)$ and λ, respectively, and adding give the relation

$$(9) \qquad Ax_1x + By_1y = 1,$$

which shows that the intersection of the tangents lies on the polar line L_1 having the equation (2). This is a remarkable geometric fact. Different lines through P_1 yield hordes of pairs of tangents, and all of the intersections lie on the same polar line L_1. We shall not undertake to prove the fact, illustrated by Figure 6.4993, that pairs of chords as well as pairs of tangents intersect on the polar line L_1. When the lines L and L' through P_1 intersect the conic at four known points P_2, P_3, P_2', P_3' as in the figure, the chords P_2P_2' and P_3P_3' intersect at one point on L_1, the chords P_2P_3' and P_3P_2' intersect at another point on L_1, and, moreover, these two intersections determine L_1. Thus (unless parallelism causes trouble) we can start with just four points on a conic and use them to determine a point P_1 and its polar line L_1. This "four-point construction" is remarkable because the four points on the conic do not determine the conic. If we are given a fifth point E on the conic (such that no three of the five are collinear) then the conic is determined and we can produce the dotted lines of the figure to construct a sixth point E'. Appropriate use of polar points and polar lines provides methods for construction of more points. We are not expected to learn much about these matters in our elementary course, but we can be aware of the fact that many persons continue study of geometry to learn more

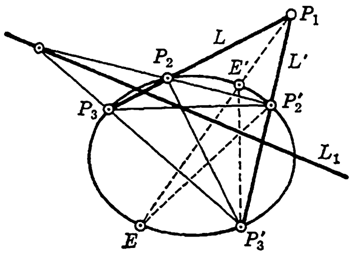


Figure 6.4993

6.5 *Translation and rotation of axes* To start learning about advantages gained by introducing supplementary coordinate systems, we look at an example so simple that the supplementary coordinate system is not needed. When we want to learn about the graph of the equation

$$x^2 + y^2 - 10x - 8y + 32 = 0,$$

we complete squares and write the equation in the form

$$(6.51) \qquad (x-5)^2 + (y-4)^2 = 9.$$

It is then easily seen that the graph is a circle with center at (5,4) and radius 3, and we can sketch the graph in Figure 6.52 without onerous calculations. If in the equation (6.51) we set

$$(6.511) \qquad x' = x - 5, \qquad y' = y - 4$$

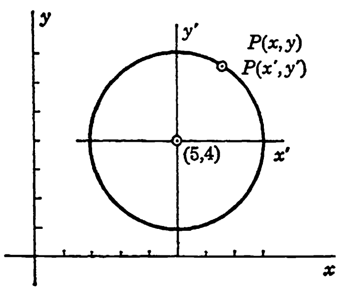


Figure 6.52

or, equivalently,

$$x = x' + 5, \qquad y = y' + 4, \tag{6.512}$$

the equation takes the simpler form

$$x'^2 + y'^2 = 9. \tag{6.513}$$

If we think of x' and y' as being coordinates (the prime or primed coordinates), then (6.513) looks like the equation of a circle having its center at the origin of the new prime coordinate system of Figure 6.52. It is customary to say that we "translate axes" when, as in Figure 6.52, we introduce a supplementary coordinate system that can be obtained from the original one by translations, that is, by slidings free from rotations. It sometimes happens that, in more complicated situations, introduction of a supplementary coordinate system helps us to determine the nature and position of the graph of a given equation.

We now begin consideration of the graph of the equation $Q = 0$, where

$$Q = Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F, \tag{6.53}$$

it being supposed that A, B, C, D, E, F are given constants for which A, B, C are not all 0. In case $B = 0$, the following results can be obtained more quickly by completing squares. In any case, we undertake to determine constants h and k such that the substitution $x = x' + h$, $y = y' + k$ will yield a simpler expression for Q. Substitution gives

$$Q = Ax'^2 + 2Bx'y' + Cy'^2 + 2D'x' + 2E'y' + F' \tag{6.531}$$

where
$$\begin{aligned} D' &= Ah + Bk + D \\ E' &= Bh + Ck + E \\ F' &= Ah^2 + 2Bhk + Ck^2 + 2Dh + 2Ek + F. \end{aligned}$$
In case $B^2 - AC \neq 0$, we can simplify the expression for Q in (6.531) by determining h and k so that $D' = E' = 0$. Except in the special case in which $F' = 0$ when h and k are so determined that $D' = E' = 0$, the equation $Q = 0$ resulting from making $D' = E' = 0$ is not easily graphed unless $A = C = 0$ or $B = 0$. Hence, at least in the study of Q when $B \neq 0$, the results that lie ahead are much more important than those obtained by translation of axes. For those who may become interested in such things, it may be remarked that the quadratic form Q in

$$\begin{aligned} Q = {} & Axx + Bxy + Dxz \\ & + Byx + Cyy + Eyz \\ & + Dzx + Ezy + Fzz \end{aligned} \tag{6.54}$$

reduces to (6.53) when $z = 1$ and that there are places in pure and applied mathematics where these things are important. To attack

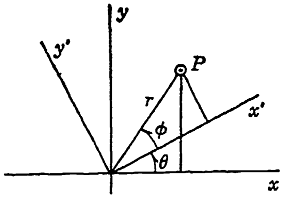


Figure 6.551

(6.54), we would use the formulas (2.67) (with $x_0 = y_0 = z_0 = 0$) from Section 2.6 instead of the simpler formulas of the next paragraph.

It is customary to say that we "rotate axes" when, as in Figure 6.551, we introduce a supplementary coordinate system that can be obtained from the original one by rotation about a line through the origin perpendicular to the plane of the given coordinate system. Without going back to review formulas from Section 2.6, we shall use Figure 6.551 to derive the formulas

$$\text{(6.55)}\qquad \begin{aligned} x &= x'\cos\theta - y'\sin\theta, & x' &= x\cos\theta + y\sin\theta \\ y &= x'\sin\theta + y'\cos\theta, & y' &= -x\sin\theta + y\cos\theta \end{aligned}$$

that relate the original and prime coordinates of a point P when the prime coordinate system is obtained by rotating the original axes through the angle θ. The primitive formulas

$$\frac{x'}{r} = \cos\phi, \qquad \frac{y'}{r} = \sin\phi, \qquad \frac{x}{r} = \cos(\phi+\theta), \qquad \frac{y}{r} = \sin(\phi+\theta)$$

and the formulas for cosines and sines of sums give

$$\begin{aligned} x &= r\cos(\phi+\theta) = r\cos\phi\cos\theta - r\sin\phi\sin\theta = x'\cos\theta - y'\sin\theta \\ y &= r\sin(\phi+\theta) = r\cos\phi\sin\theta + r\sin\phi\cos\theta = x'\sin\theta + y'\cos\theta. \end{aligned}$$

This gives the first set of formulas (6.55). The second set can be obtained by solving the first set for x' and y', or by making appropriate observations about the result of replacing θ by $-\theta$.

Before attacking more ponderous expressions, we observe that the first formulas in (6.55) show that if $Q = xy$, then

$$\text{(6.552)}\qquad \begin{aligned} Q &= (x'\cos\theta - y'\sin\theta)(x'\sin\theta + y'\cos\theta) \\ &= (x'^2 - y'^2)\sin\theta\cos\theta + x'y'(\cos^2\theta - \sin^2\theta) \\ &= \tfrac{1}{2}(x'^2 - y'^2)\sin 2\theta + x'y'\cos 2\theta. \end{aligned}$$

We can eliminate the $x'y'$ term by making $\cos 2\theta = 0$ and hence by setting $2\theta = \pi/2$ and $\theta = \pi/4$ (or 45°). This gives $Q = \frac{1}{2}(x'^2 - y'^2)$. The graph in the xy plane of the equation $xy = 1$ is the same as the graph in the prime coordinate system of the equation

$$\text{(6.553)}\qquad \frac{x'^2}{(\sqrt{2})^2} - \frac{y'^2}{(\sqrt{2})^2} = 1.$$

Thus the graph of the equation $y = 1/x$ is now thoroughly inducted into the hyperbolic fraternity.

As in (6.53), let

$$\text{(6.56)}\qquad Q = Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F$$

and suppose that A, B, C are not all zero. Rotating the x, y axes through the angle θ gives new coordinates x', y', and using (6.55) shows that, in the new coordinates,

$$Q = A'x'^2 + 2B'x'y' + C'y'^2 + 2D'x' + 2E'y' + F \tag{6.561}$$

where
$$\begin{aligned} A' &= A\cos^2\theta + 2B\sin\theta\cos\theta + C\sin^2\theta \\ B' &= (C - A)\sin\theta\cos\theta + B(\cos^2\theta - \sin^2\theta) \\ C' &= A\sin^2\theta - 2B\sin\theta\cos\theta + C\cos^2\theta \\ D' &= D\cos\theta + E\sin\theta \\ E' &= -D\sin\theta + E\cos\theta. \end{aligned}$$

Considerable information can be extracted from the above formulas without the aid of noncranial electronics. For each θ, we obtain the first of the formulas

$$A' + C' = A + C, \qquad B'^2 - A'C' = B^2 - AC \tag{6.562}$$

by adding the expressions for A' and C'. To prove the second formula, we can write an expression for the left side, cancel the terms that cancel, and obtain the result with the aid of the identity

$$\cos^4\theta + 2\cos^2\theta\sin^2\theta + \sin^4\theta = (\cos^2\theta + \sin^2\theta)^2 = 1.$$

Because of (6.562), the quantities $A + C$ and $B^2 - AC$ are said to be *invariant under rotation of axes.*

The basic importance of the formulas for A', B', C', D', E' lies in the fact that they show us how to determine θ and the coefficients in (6.561) so that B' is zero and the objectionable term is missing. When we write the formula for B' in the form

$$B' = \tfrac{1}{2}(C - A)\sin 2\theta + B\cos 2\theta,$$

we can see that it is easy to find a unique angle θ for which $0 \leqq \theta \leqq \pi/2$ and $B' = 0$. In case $C = A$, we set $\theta = \pi/2$. In case $C \neq A$, we choose 2θ such that $0 \leqq 2\theta \leqq \pi$ and

$$\tan 2\theta = \frac{2B}{A - C}. \tag{6.563}$$

Without repeating the formulas for the prime coefficients and the method by which θ is found, we put our main result in the following theorem.

Theorem 6.57 *Each Q of the form*

$$Q = Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F \tag{6.571}$$

can be put in the form

$$Q = A'x'^2 + C'y'^2 + 2D'x + 2E'y + F \tag{6.572}$$

by suitable choice of an angle θ for which $0 \leqq \theta \leqq \pi/2$.

When we are asked about the graph of $Q = 0$, we can calculate the coefficients in (6.572) and then (completing squares when necessary) use

(6.572) to get the information. It is of interest to observe that we can get information about Q without making these calculations. Even when we do not know the numerical values of A' and C', we know that

$$B^2 - AC = -A'C' \tag{6.573}$$

when $B' = 0$ because $B^2 - AC$ is invariant under rotation of axes. In case $B^2 - AC < 0$, we must have $A'C' > 0$, and (6.572) shows that the equation $Q = 0$ has elliptic type. In case $B^2 - AC = 0$, we must have $A'C' = 0$ and the equation $Q = 0$ has parabolic type. In case $B^2 - AC > 0$, we must have $A'C' < 0$ and the equation $Q = 0$ has hyperbolic type. Sometimes it is necessary to know the fact, which was given in Problem 32 of Section 6.4, that graphs of equations of elliptic type are not always ellipses; they may be circles or points or empty sets. The number $(2B)^2 - 4AC$ or $4(B^2 - AC)$ is called the *discriminant* of Q, and its sign is the same as the sign of $B^2 - AC$. Hence we can put the above results in a form that is easier to remember and use.

Theorem 6.58 *The equation $Q = 0$ (and, in fact, Q itself) is elliptic or parabolic or hyperbolic according as its discriminant is negative or zero or positive.*

Finally, we should know something about ways of using (6.563) to calculate the coefficients A', B', C', D', E'. It is possible to write formulas giving $\sin\theta$ and $\cos\theta$ in terms of A, B, C, but these formulas are so unattractive that we shun them. Applied mathematicians do not need lessons from this book to obtain approximations to answers by finding 2θ from a table or slide rule and then finding θ by gaily dividing by 2. When exact results must be obtained, we draw a figure more or less similar to Figure 6.581 and use elementary ideas to obtain $\cos 2\theta$. When $0 \leqq \theta < \pi/2$, $\cos\theta$ and $\sin\theta$ must be nonnegative and their values can be calculated from the formulas

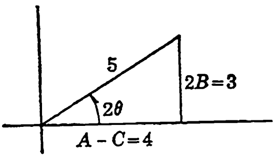


Figure 6.581

$$\cos\theta = \sqrt{\frac{1+\cos 2\theta}{2}}, \qquad \sin\theta = \sqrt{\frac{1-\cos 2\theta}{2}}. \tag{6.582}$$

This gives the information needed for locating the prime axes and calculating the prime coefficients.

Problems 6.59

1 Sketch x, y axes having origin O and, in the same figure, sketch primed (or new) parallel axes with origin O' having unprimed coordinates (2,3). Using the primed axes, sketch the parabola having the equation $y' = x'^2$. Find the primed coordinates of the points A and B where the parabola intersects the lines having the primed equations $x' = -2$ and $x' = 2$. Find the primed coordinates of the focus F and the primed equation of the directrix D of the parabola. Finally, find unprimed equations and coordinates of the parabola and of A, B, F, and D.

2 Write the equation

$$\frac{(x+2)^2}{1^2} + \frac{(y-3)^2}{2^2} = 1 \tag{1}$$

in the form

$$Ax^2 + By^2 + Cx + Dy + E = 0. \tag{2}$$

Then substitute

$$x = x' + h, \qquad y = y' + k \tag{3}$$

into the result and so determine h and k that the coefficients of x' and y' will be zero. Show that the new equation can then be put in the form

$$\frac{x'^2}{1^2} + \frac{y'^2}{2^2} = 1. \tag{4}$$

Finally, sketch properly related unprimed and primed axes and sketch the graph of (4).

3 Show how a primed coordinate system can be introduced to simplify the process of obtaining basic information about the hyperbola having the equation

$$\frac{(x-3)^2}{3^2} - \frac{(y-1)^2}{1^2} = 1.$$

Sketch a figure showing both sets of axes and the hyperbola.

4 Translate axes to remove the first-degree terms from the equation

$$xy - 2x + 3y - 4 = 0.$$

Remark: This gibberish means something. Let $x = x' + h$, $y = y' + k$ and determine h and k in such a way that the coefficients of x' and y' in the new equation will be 0. *Ans.:* $h = -3$, $k = 2$, $x'y' = -2$.

5 Translate axes to simplify the equation

$$x^2 + xy + y = 3.$$

Ans.: Putting $x = x' + h$ and $y = y' + k$, we find that the first-degree terms disappear (have zero coefficients) when $h = -1$, $k = 2$, and that the simplified equation is $x'^2 + x'y' = 2$.

6 Construct and solve more problems similar to the preceding two, but keep the equations simple. We do not have time to do chores that computers should do.

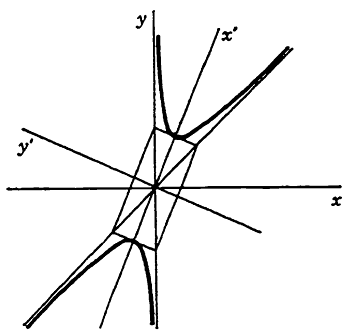
7 A rough graph of the equation

$$y = x + \frac{1}{x} \tag{1}$$

is easily drawn. A miniature version appears in Figure 6.591. Since (1) can be put in the form

$$x^2 - xy + 1 = 0, \tag{2}$$

Figure 6.591



the graph must be a conic. While it is possible to foresee some of the results of putting (2) into standard form by "rotation of axes," we examine the details of the process. Show that substituting

$$(3) \qquad x = x' \cos \theta - y' \sin \theta, \quad y = x' \sin \theta + y' \cos \theta$$

into (2) gives an equation in x' and y' for which the coefficient of $x'y'$ will be zero when $\tan 2\theta = -1$ and hence

$$(4) \qquad 2\theta = \frac{3\pi}{4}, \quad \theta = \frac{3\pi}{8}, \quad \cos^2 \theta = \frac{1 + \cos 2\theta}{2} = \frac{\sqrt{2} - 1}{2\sqrt{2}},$$

$$\sin^2 \theta = \frac{1 - \cos 2\theta}{2} = \frac{\sqrt{2} + 1}{2\sqrt{2}}.$$

Continue the work to show that the equation in x' and y' can be put in the form

$$(5) \qquad \frac{x'^2}{2\sqrt{2} + 2} - \frac{y'^2}{2\sqrt{2} - 2} = 1.$$

Note that the result agrees very well with Figure 6.591. The distance from the center of the hyperbola to its vertices is $\sqrt{2 + 2\sqrt{2}}$. It is possible to use (5) to obtain more information in primed coordinates and to express this information in terms of the original coordinates, but all of these operations consume considerable time.

8 Obtain preliminary information about the graph of the equation

$$x(x + y) = 1$$

by solving for y (or for x) and making a rough sketch. Then obtain more precise information by rotation of axes.

9 Apply the procedure of Problem 8 to the equation

$$x^2 + xy + y^2 = 1.$$

10 Make and solve more problems similar to Problems 8 and 9, but keep the equations simple. We still do not have time to do chores that computers should do.

11 When a and b are positive, the equation

$$a^2x^2 + b^2y^2 - a^2b^2 = 0$$

has elliptic type. For what values of k does the equation

$$a^2x^2 + kxy + b^2y^2 - a^2b^2 = 0$$

have elliptic type? Parabolic type? Hyperbolic type? *Hint:* The discriminant is $k^2 - 4a^2b^2$, and we are put on the right track when we notice that this is negative and our equation is elliptic when $k = 0$.

12 When a and b are positive, the equation

$$a^2x^2 - b^2y^2 - a^2b^2 = 0$$

has hyperbolic type. For what values of k does the equation

$$a^2x^2 + kxy - b^2y^2 - a^2b^2 = 0$$

have hyperbolic type? Parabolic type? Elliptic type? *Hint:* Get on the right track.

13 We should know something about the possibility of obtaining information about the graph of the equation

$$Ax^2 + 2Bxy + Cy^2 + F = 0$$

by solving for x or y by use of the quadratic formula. Other cases being similar or much easier, suppose that $C > 0$, $B \neq 0$, and $F \neq 0$. Show that a point P lies on the graph iff (iff means if and only if) its coordinates satisfy the equation

$$y = \frac{-Bx \pm \sqrt{(B^2 - AC)x^2 - CF}}{C}.$$

Show that if $B^2 - AC > 0$, then the points (x,y) for which $|x|$ is sufficiently great and

$$y = \frac{x}{C}\left[-B \pm \sqrt{(B^2 - AC) - \frac{CF}{x^2}}\right]$$

must lie on the graph, and tell why the graph cannot be an ellipse. Show that if $B^2 - AC < 0$, the graph cannot be a hyperbola.

14 Sketch a graph of the equation

$$y = x + \sqrt{1 - x^2}$$

by sketching graphs of $y_1 = x$ and $y_2 = \sqrt{1 - x^2}$ and adding ordinates (that is, values of y). Give some precise information about the graph.

15 It is sometimes said that the graph of the equation

$$x^{1/2} + y^{1/2} = a^{1/2},$$

where a is a positive constant, is a parabola. Is this true? *Ans.:* No, but the graph is a part of a parabola.

16 A wheel of radius a rolls, without slipping and with angular speed ω, on the top side of the x axis of an xy plane. At time $t = 0$, the center is above the origin, and a pink spot P which rotates with the wheel lies b units below the center of the wheel. Show that if an x', y' coordinate system, with origin at O', travels with the wheel but keeps its unit vectors $\mathbf{i}'$ and $\mathbf{j}'$ in the directions of $\mathbf{i}$ and $\mathbf{j}$, then

$$\overrightarrow{O'P} = -b \sin \omega t\mathbf{i} - b \cos \omega t\mathbf{j}.$$

Show that, because the wheel rolls without slipping,

$$\overrightarrow{OO'} = a\omega t\mathbf{i} + a\mathbf{j}.$$

Show that the vector $\mathbf{r}$ running from O to P is

$$\mathbf{r} = (a\omega t - b \sin \omega t)\mathbf{i} + (a - b \cos \omega t)\mathbf{j}.$$

Show that

$$\mathbf{v} = (a\omega - b\omega \cos \omega t)\mathbf{i} + (b\omega \sin \omega t)\mathbf{j}$$
$$\mathbf{a} = b\omega^2(\sin \omega t\mathbf{i} + \cos \omega t\mathbf{j}).$$

Remark: Moving coordinate systems and vectors provide the simplest way of obtaining neat and correct answers to problems more or less like this one. The paths traced by the pink spots P are *cycloids*. For future reference, we note that if $b = a$, so that P is on the rolling circle and the cycloid is an *ordinary cycloid*, and if we let θ be the angle ωt through which the wheel has rotated at time t, then the coordinates of P are

$$x = a(\theta - \sin\theta) \qquad y = a(1 - \cos\theta).$$

Figure 6.592 exhibits the ordinary cycloid having these equations. The unusual

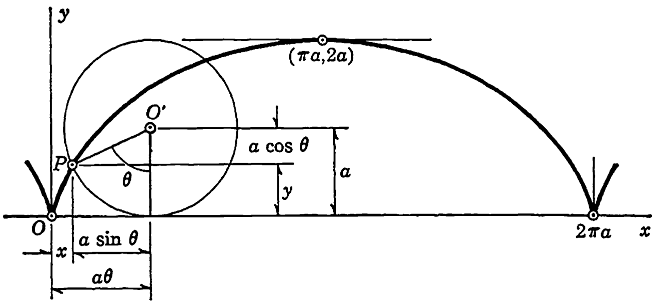


Figure 6.592

cycloids are sometimes called *curtate cycloids* or *prolate cycloids*, and are sometimes called *trochoids*.

17 Find equations of epicycloids, that is, paths traced by points on spokes (or extended spokes) of circular wheels which roll, without slipping, outside a fixed circular wheel. *Outline of solution:* As in Figure 6.593, let the fixed circle have radius a and have its center at the origin of an x, y coordinate system. Let the rolling circle have radius b and have center O' which travels with it, and let ωt (not restricted to the interval $0 \leqq \omega t \leqq 2\pi$) be the angle which $\overrightarrow{OO'}$ makes with the x axis at time t so that

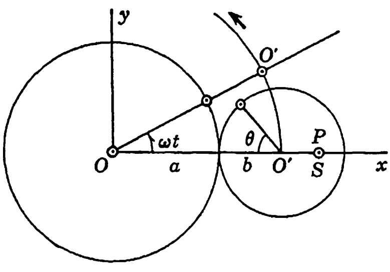


Figure 6.593

(1) $\overrightarrow{OO'} = (a + b)(\cos\omega t\mathbf{i} + \sin\omega t\mathbf{j}).$

We suppose that $\omega > 0$, so the spokes in the moving wheel rotate in a positive direction. Let the point P which traces the epicycloid be the point for which $\overrightarrow{O'P} = c\mathbf{i}$ when $t = 0$, and let S be the spoke of the moving wheel which (extended if necessary) contains P. The angle ψ (psi) through which the spoke S has turned at time t is the sum of two angles: (i) the angle θ through which it would turn if its wheel rolled the distance $a\omega t$ along a straight line and (ii) the angle ωt through which it would turn if it slid without rolling on the fixed circle. We find that $a\omega t = b\theta$, so $\theta = (a/b)\omega t$,

(2) $$\psi = \frac{a + b}{b}\,\omega t,$$

and

$$(3) \qquad \overrightarrow{O'P} = c\left(\cos\frac{a+b}{b}\omega t\mathbf{i} + \sin\frac{a+b}{b}\omega t\mathbf{j}\right)$$

at time t. From (1) and (3) we obtain

$$(4) \quad \mathbf{r} = \left[(a+b)\cos\omega t + c\cos\frac{a+b}{b}\omega t\right]\mathbf{i} + \left[(a+b)\sin\omega t + c\sin\frac{a+b}{b}\omega t\right]\mathbf{j}$$

for the displacement vector of the point P on the epicycloid at time t. An *ordinary epicycloid* is obtained by setting $c = -b$ so P starts at the initial point of tangency of the two circles. *Remark:* For the case in which $c = -b$ and $b = a$, we obtain the epicycloid of one cusp having the equation

$$(5) \qquad \mathbf{r} = a[2\cos\omega t - \cos 2\omega t]\mathbf{i} + a[2\sin\omega t - \sin 2\omega t]\mathbf{j}.$$

Using the trigonometric identities

$$(6) \qquad \cos 2\theta = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1, \qquad \sin 2\theta = 2\sin\theta\cos\theta$$

enables us to put (5) in the form

$$(7) \qquad \mathbf{r} - a\mathbf{i} = 2a(1 - \cos\omega t)(\cos\omega t\mathbf{i} + \sin\omega t\mathbf{j})$$

from which we see that

$$(8) \qquad |\mathbf{r} - a\mathbf{i}| = 2a(1 - \cos\omega t).$$

Letting ρ denote the distance from the point $(a,0)$ to the point P on our epicycloid and setting $\phi = \omega t$ puts (8) in the form

$$(9) \qquad \rho = 2a(1 - \cos\phi).$$

The polar coordinate graph of (9) is a cardioid. Thus we have discovered that an epicycloid with one cusp is a cardioid, and we have started learning about epicyclic gears. That the Greek prefixes *epi* and *hypo* mean outside (or above) and inside (or below) can be remembered by those who have hypodermics put under their skins.

18 Find equations of hypocycloids, that is, paths traced by points on spokes (or extended spokes) of circular wheels which roll, without slipping, inside a fixed larger circular wheel. *Outline of solution:* This problem is much like Problem 17. To get the answer from (4) of Problem 17, replace b by $-b$ because $|\overrightarrow{OO'}| = a - b$. For hypocycloids the spokes of the inner wheels run backwards. The equation giving displacement vectors of points on hypocycloids is

$$(1) \quad \mathbf{r} = \left[(a-b)\cos\omega t + c\cos\frac{a-b}{b}\omega t\right]\mathbf{i} + \left[(a-b)\sin\omega t - c\sin\frac{a-b}{b}\omega t\right]\mathbf{j}.$$

An *ordinary hypocycloid* is obtained by setting $c = b$ so P starts at the initial point of tangency of the two circles. A graph appears in Figure 7.291. *Remark:* For the case in which $c = b$ and $b = a/4$, we obtain the hypocycloid of four cusps having the equation

$$(2) \qquad \mathbf{r} = \frac{a}{4}[3 \cos \phi + \cos 3\phi]\mathbf{i} + \frac{a}{4}[3 \sin \phi - \sin 3\phi]\mathbf{j}$$

or

$$(3) \qquad \mathbf{r} = a[\cos^3 \omega t\mathbf{i} + \sin^3 \omega t\mathbf{j}]$$

or

$$(4) \qquad x^{2/3} + y^{2/3} = a^{2/3}.$$

19 Persons interested in machinery should determine the path traced by a cog on the inner wheel of a hypocyclic gear when the radius of the inner wheel is just half the radius of the outer wheel.

20 A rod of length $2a$, always in the xy plane, is whirling about its center with angular speed ω and, at the same time, is so thrown that its center has coordinates $(At, Bt - Ct^2)$ at time t. Supposing that a red spot on the stick started at the point $(0,a)$ at time $t = 0$, find its position and velocity and acceleration at later times. Try to solve the problem without use of the following outline. If unsuccessful, look hastily at the outline to get some ideas and then try to solve the problem with the outline out of sight. *Outline of solution:* Use vectors and a moving coordinate system. Let an x, y coordinate system with unit vectors $\mathbf{i}$ and $\mathbf{j}$ be drawn and remain fixed (that is, always in the same place). Let an x', y' coordinate system have origin O' at $(At, Bt - Ct^2)$ at time t, and let its unit vectors $\mathbf{i}'$ and $\mathbf{j}'$ have the directions of $\mathbf{i}$ and $\mathbf{j}$ so that, in the world of the moving coordinate system, the stick is doing nothing but rotate about O'. Letting r be the vector running from O to P, we have

$$\begin{aligned}\mathbf{r} &= \overrightarrow{OO'} + \overrightarrow{O'P}\\ &= (At + a \cos \omega t)\mathbf{i} + (Bt - Ct^2 + a \sin \omega t)\mathbf{j}\end{aligned}$$

so

$$\mathbf{v} = (A - a\omega \sin \omega t)\mathbf{i} + (B - 2Ct + a\omega \cos \omega t)\mathbf{j}$$

and

$$\mathbf{a} = (-a\omega^2 \cos \omega t)\mathbf{i} + (-2C - a\omega^2 \sin \omega t)\mathbf{j}.$$

21 Let a and b be positive constants, and let G be the graph in the x', y' plane of the first of the equations

$$y' = ae^{-bx'^2}, \qquad y = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2}.$$

Find two positive constants λ and σ such that G will be the graph of the second equation in the x, y plane, the primed and unprimed coordinates being related by the formulas $x' = \lambda x$ and $y' = \lambda y$. *Ans.:* $\lambda = (\pi a^2/b)^{1/4}$, $\sigma = (1/4\pi a^2 b)^{1/4}$.

6.6 *Quadric surfaces* This brief section, which contains no terminal list of problems, can be omitted, but it can be read even by students who are not required to read it. As is easy to see by considering such special examples as

$$(6.61)\quad x^2 + y^2 + z^2 + 1 = 0, \qquad x^2 + y^2 + z^2 = 0, \qquad x^2 + y^2 = 0,$$

the graph in E_3 of an equation of the form

$$(6.62)\quad Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fzx + Gx + Hy + Iz + J = 0$$

can be the empty set or a point or a line. In case A, B, C, D, E, F are not all zero and the graph is a surface, the surface is called a *quadric surface*. Examples of the form

$$(6.621)\qquad Ax^2 + By^2 + Cxy + D = 0,$$

where A, B, C are not all 0, show that some quadric surfaces are cylinders which may be planes or pairs of planes. Our interest in this section lies in quadric surfaces that are not cylinders.

As we study quadric surfaces, it will be helpful to have some results of Section 2.6 in mind. A line which does not lie completely on a quadric surface can intersect the quadric surface in at most two points. Each plane section of a quadric surface must be the empty set or a single point or a single line or two lines or a circle or a parabola or an ellipse or a hyperbola. From this we can conclude that each nonempty bounded plane section of a quadric surface must be either a point or a circle or an ellipse.

When the equation (6.62) of a quadric surface is given, it is possible to introduce a new coordinate system in such a way that the equation in the new coordinates has one or another of several standard forms. There are places, even in applied mathematics, where it is necessary to know procedures for putting given equations into standard forms, but the procedures are much too complicated for coverage in elementary courses. In this course we can be content with a little information about the standard forms and their graphs. In what follows, it is always assumed that a, b, c are given positive constants. The standard forms are selected in such a way that, insofar as is possible, the coordinate planes are planes of symmetry and the curves in which the surface intersects planes parallel to the coordinate planes are easily described and sketched

The graph of the equation

$$(6.63)\qquad \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

is an *ellipsoid* except when $a = b = c$ and the graph is a sphere. Putting $z = k$ shows that the ellipsoid intersects the plane having the equation

$z = k$ in a set which is the empty set if $|k| > c$, a point if $|k| = c$, and an ellipse (or circle if $a = b$) if $|k| < c$. Very similar remarks apply to the planes having the equations $x = k$ and $y = k$. It is easiest to sketch the intersections (or sections) for which $k = 0$, but full information about other sections parallel to the coordinate planes is easily obtained. For example, when $|k| < c$, we can put $z = k$ in (6.63), transpose the term involving c^2, and then divide by the new right side to obtain

$$\frac{x^2}{a^2\left(1 - \frac{k^2}{c^2}\right)} + \frac{y^2}{b^2\left(1 - \frac{k^2}{c^2}\right)} = 1.$$

In case $a \neq b$, this shows that the graph intersects the plane having the equation $z = k$ in an ellipse which has its center on the z axis and which intersects the planes having the equations $x = 0$ and $y = 0$ at the points

$$\left(0, \pm b\sqrt{1 - \frac{k^2}{c^2}}, k\right), \qquad \left(\pm a\sqrt{1 - \frac{k^2}{c^2}}, 0, k\right).$$

In case $a = b$, the section is a circle. Unshaded and shaded graphs appear in Figures 6.631 and 6.632. In case $a = c$ and b is greater than a and c,

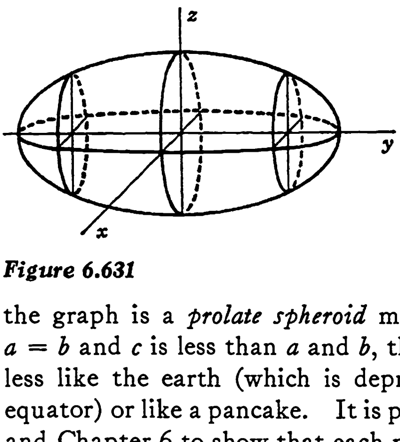


Figure 6.631

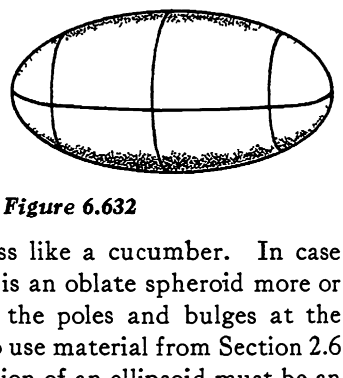


Figure 6.632

the graph is a *prolate spheroid* more or less like a cucumber. In case $a = b$ and c is less than a and b, the graph is an oblate spheroid more or less like the earth (which is depressed at the poles and bulges at the equator) or like a pancake. It is possible to use material from Section 2.6 and Chapter 6 to show that each plane section of an ellipsoid must be an empty set or a single point or a circle or an ellipse.

The graph of the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \tag{6.64}$$

is a *hyperboloid of one sheet*. It intersects the plane having the equation $z = k$ in an ellipse (or circle if $a = b$). It intersects the plane having the equation $y = k$ in a hyperbola when $|k| \neq b$ and in a pair of lines when $|k| = b$. It intersects the plane having the equation $x = k$ in a hyperbola when $|k| \neq a$ and in a pair of lines when $|k| = a$. Unshaded and shaded graphs appear in Figures 6.641 and 6.642.

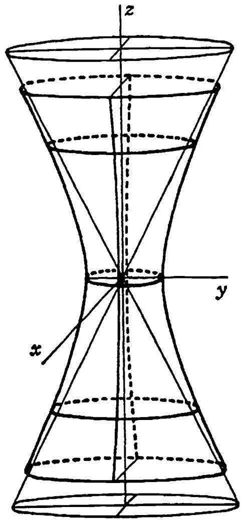


Figure 6.641

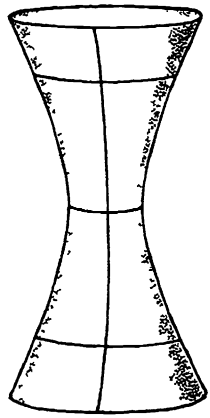


Figure 6.642

The graph of the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1 \qquad \text{or} \qquad -\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{6.65}$$

is a *hyperboloid of two sheets*. It intersects the plane having the equation $z = k$ in the empty set when $|k| < c$, in a point when $|k| = c$, and in an ellipse (or circle) when $|k| > c$. It intersects the planes having the equations $y = k$ and $x = k$ in hyperbolas. Unshaded and shaded graphs appear in Figures 6.651 and 6.652.

Figure 6.651

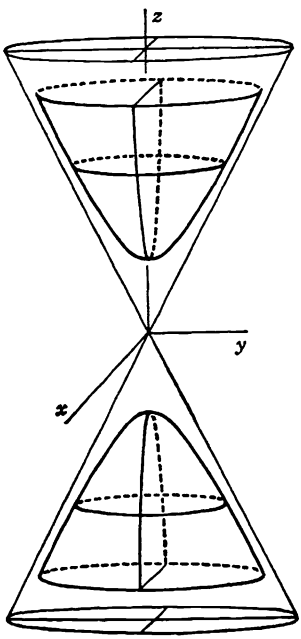
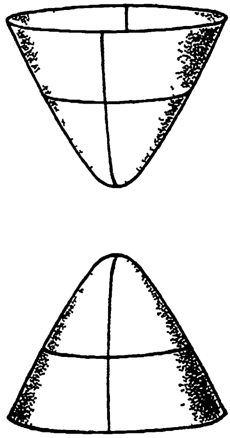


Figure 6.652



The graph of the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = z \tag{6.66}$$

is an *elliptic paraboloid.* It intersects the plane having the equation $z = k$ in the empty set when $k < 0$, in a point when $k = 0$, and in an ellipse (or circle) when $k > 0$. It intersects the planes having the equations $x = k$ and $y = k$ in parabolas. Unshaded and shaded graphs appear in Figures 6.661 and 6.662.

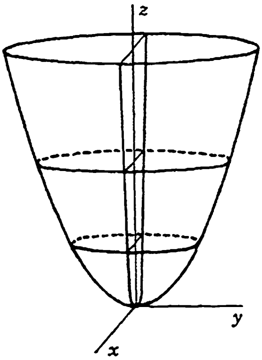


Figure 6.661

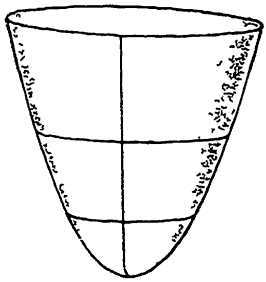


Figure 6.662

The graph of the equation

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = z \tag{6.67}$$

is a *hyperbolic paraboloid.* It intersects the plane having the equation $z = k$ in a hyperbola when $k < 0$, in a pair of lines when $k = 0$, and in a hyperbola when $k > 0$. It intersects the planes having the equations $x = k$ and $y = k$ in parabolas. Unshaded and shaded graphs appear in Figures 6.671 and 6.672. Hyperbolic paraboloids are the simplest

Figure 6.671

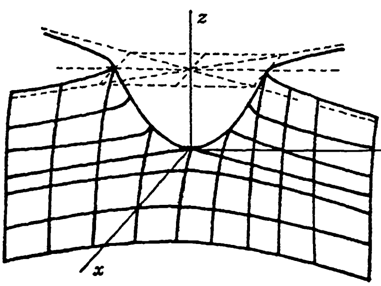
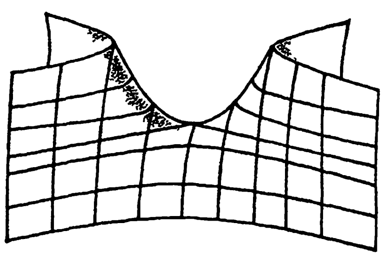


Figure 6.672



examples of surfaces known as *saddle surfaces*. The origin is a *saddle point*, and a huge surprise awaits embryonic mathematical physicists who suppose that only cowboys are interested in them.

Finally, we should not forget the cones. When A and B are nonzero constants not both negative, the graph of the equation

$$Ax^2 + By^2 = z^2 \tag{6.68}$$

is a nondegenerate *quadric cone* It is a cone because the point $(\lambda x,\lambda y,\lambda z)$ lies on the graph whenever λ is a number and the point (x,y,z) lies on the graph. It intersects the plane having the equation $z = k$ in a point (the vertex of the cone) if $k = 0$ and in a central conic (circle. ellipse, or hyperbola) having its center on the z axis if $k \neq 0$. Instead of exhibiting graphs and photographs of quadric cones, we conclude with a remark. Since we know that hyperbolas have asymptotes, we need not be surprised to learn that hyperboloids can have asymptotic cones. The cone having the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2} \tag{6.69}$$

is the common asymptotic cone of the two hyperboloids having the equations (6.64) and (6.65).

While some of us will meet ellipsoids and other quadric surfaces after completing this course, we need not invest our time in consideration of more or less routine problems involving special quadric surfaces. We have earned the right to think about a little (or big?) problem that can be of interest to those who slice onions and other things. Let S be a set in E_3 which contains more than one point. Suppose that, for each plane π, the intersection of S and π is a circle or a set consisting of just one point or the empty set. Do our hypotheses imply that S must be a sphere? If it is easy to make an incorrect guess and if the problem is difficult, so be it. If there is just one easy way to solve the problem and only one person in a million can discover the way, so be it. An anecdote illustrates the fact that we are sometimes provided opportunities to do our own thinking and investigating. In a lecture to advanced students at Cambridge, G. H. Hardy stated that the answer to a particular question is obvious. Becoming dubious about his assertion, Hardy asked his students to excuse him while he sat down to think about the matter. Five minutes later, Hardy arose to report that the answer is obvious and to continue his lecture.

7 *Curves, lengths, and curvatures*

7.1 Curves and lengths One of the main purposes of this chapter may seem at first sight to be quite modest. We want to show that the length L of the circular arc of Figure 7.11, in which $ABCD$ is a rectangle and the arc is tangent to CD at D, satisfies the inequality

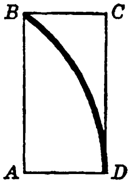


Figure 7.11

$$|\overrightarrow{AB}| \leqq L \leqq |\overrightarrow{BC}| + |\overrightarrow{CD}|. \tag{7.12}$$

Before we ask our little sister to solve the problem, we should ask ourselves a question that shows us that the problem is not completely simple. What is a circular arc, and how do we know that it determines a number that can be called its length? It is clear that we need some definitions before we can do anything.

For an accurate treatment of matters relating to lengths of curves, we need more information about things that are sometimes called curves and are sometimes called oriented curves. A bumblebee can give us preliminary ideas by starting at A, flying to a rose at B, flying on to another rose at C, flying back over the same route to B, and then flying to D and on to E as in Figure 7.13. His total path is a curve, and we get the idea that a curve is not determined by a point set. We cannot know what the curve is until we know the order in which points on the curve were visited. For example, if the bumblebee flies from E back to A by traversing his route in reverse, his path is a new curve which can be called the negative of the original one. While the matter has psychological rather than logical importance to us now, we can feel quite confident that if the bumblebee makes a flight over the interval $t_1 \leqq t \leqq t_2$, and if we introduce an x, y, z coordinate system, then at each time in the interval he is surely someplace and that if we denote his coordinates by $x(t)$, $y(t)$, $z(t)$, then these coordinate functions are continuous functions of t. Of course, the curve does not uniquely determine the coordinate functions because the bumblebee can fly his course at different speeds, but any one set of appropriate coordinate functions does determine the intrinsic curve. The above brief discussion of curves leaves many unanswered questions. It will have served its purpose if it provides a hazy feeling that the following definition uses words in a reasonable way.

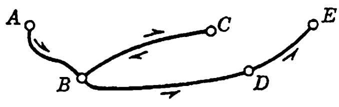


Figure 7.13

Definition 7.14 *If x, y, z are continuous functions of t over an interval, then the ordered set of points*

$$P(t) = P(x(t), y(t), z(t)) \tag{7.141}$$

for which t lies in the interval, and for which $P(t')$ is said to precede $P(t'')$ if $t' < t''$, is a curve (or oriented curve) C.

Professional creators of complicated curves can give an example of a curve in E_3 that is clever enough to "pass through" each point in a given cube or even in the whole E_3. Such curves are space-filling curves. It must not be presumed that all curves are complicated things, however. For example, when $z(t) = 0$ for each t, the curve lies in the xy plane and we omit the $z(t)$ when no confusion can result. We always have the possibility of setting $x(t) = t$ and replacing t by x. This shows that if f is continuous over $a \leqq x \leqq b$, then the set of points $(x, f(x))$ on the graph of $y = f(x)$ becomes a curve when we so order the points that $(x_1, f(x_1))$ precedes $(x_2, f(x_2))$ when $a \leqq x_1 < x_2 \leqq b$. Unless an explicit statement to the contrary is made, it is presumed that the points are ordered in this way whenever the graph of $y = f(x)$ is called a curve. Finally, such things as circles, ellipses, rectangles, and triangles become curves as soon as we think of them as being traversed once in the positive (counterclockwise) or negative (clockwise) direction.

Supposing that C is a given curve for which the interval appearing in Definition 7.14 is a closed interval $a \leqq t \leqq b$, we proceed to define (when it exists) a number $|C|$ which is called the length of C. Let P be a partition of the interval by partition points $t_0, t_1, \cdots, t_n$ for which

$$a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = b. \tag{7.15}$$

For each $k = 0, 1, 2, \cdots, n$, let P_k be the point on C for which $t = t_k$. Figure 7.151 is a schematic figure that may be helpful, but it is much too simple to show how much care is needed to guarantee that the points are not scrambled in an inappropriate way. The polygon (or broken line) running from P_0 to P_1 to P_2 and so on to P_n is said to be *inscribed* in C. The number S_P defined by

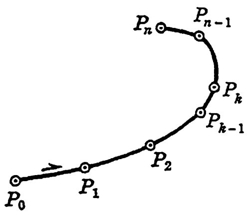


Figure 7.151

$$S_P = \sum_{k=1}^{n} |\overrightarrow{P_{k-1}P_k}|$$

is the sum of the lengths of the sides of the inscribed polygon. Let S be the set of numbers S_P obtained by making partitions P of the interval $a \leqq t \leqq b$. In case there is no number M such that $S_P \leqq M$ whenever S_P is in S, it is said that C does not have length or that C does not have finite length or that C has infinite length or that $|C| = \infty$. In case there is a number M such that $S_P \leqq M$ whenever S_P is in S, then Theorem 5.46 guarantees existence of a least number $|C|$ such that $S_P \leqq |C|$ whenever S_P is in S. This number $|C|$ is then, by definition, the *length*† of C.

For future reference we state without proof the following theorem which involves notation we have been using.

Theorem 7.16 *The curve C has finite length $|C|$ if and only if*

$$|C| = \lim_{|P|\to 0} \sum_{k=1}^{n} |\overrightarrow{P_{k-1}P_k}|.$$

We now use the definition of length to prove the following theorem which gives, as a corollary, the desired result involving Figure 7.11.

Theorem 7.17 *Let f be continuous and monotone increasing (or monotone decreasing) over $a \leqq x \leqq b$. Let C be the graph of $y = f(x)$ so oriented*

† In old books particularly, a curve C having finite length is sometimes said to be *rectifiable;* the ancient idea is that if C is a string we could pull on the ends and straighten it out to get a straight string of finite length. We can all know enough about logic to know that if C is not a string, then little credence is placed upon consequences of the false assumption that C is a string.

that $(x_1, f(x_1))$ *precedes* $(x_2, f(x_2))$ *when* $x_1 < x_2$. *Then C has finite length* $|C|$ *and*

$$|f(b) - f(a)| \leqq |C| \leqq |b - a| + |f(b) - f(a)|. \tag{7.171}$$

To prove this, let P be a partition of the interval $a \leqq x \leqq b$, and, for each k, let $P_k = (x_k, f(x_k))$. Let $\Delta x_k = x_k - x_{k-1}$ and let

$$\Delta y_k = f(x_k) - f(x_{k-1}).$$

Then

$$|\overrightarrow{P_{k-1}P_k}| = \sqrt{\Delta x_k^2 + \Delta y_k^2}. \tag{7.172}$$

But

$$\begin{aligned} \Delta y_k^2 \leqq \Delta x_k^2 + \Delta y_k^2 &\leqq |\Delta x_k|^2 + 2|\Delta x_k\, \Delta y_k| + |\Delta y_k|^2 \\ &= (|\Delta x_k| + |\Delta y_k|)^2. \end{aligned} \tag{7.173}$$

Therefore,

$$|\Delta y_k| \leqq |\overrightarrow{P_{k-1}P_k}| \leqq |\Delta x_k| + |\Delta y_k|. \tag{7.174}$$

But the numbers Δx_k are all positive and the numbers Δy_k are all negative (or all positive), and hence (why?) addition gives

$$|f(b) - f(a)| \leqq \sum_{k=1}^{n} |\overrightarrow{P_{k-1}P_k}| \leqq |b - a| + |f(b) - f(a)|. \tag{7.175}$$

Since this inequality holds for each partition P, the curve C has finite length $|C|$. Moreover, the least upper bound of the central sums cannot exceed the particular upper bound in (7.175), and no upper bound can be less than $|f(b) - f(a)|$. This proves Theorem 7.17.

We have been studying lengths of *curves*, that is, lengths of ordered sets of points of special types. There are other (and different) theories involving lengths of unordered sets S in E_3 and E_2 and E_1. We need not be authorities on these matters, but we should have at least a vague idea that, in elementary mathematics, the part of the set S of Figure 7.18 that lies "between" A and B has length L if there is a "natural ordering" of the points of the part that yields a curve C having length L.

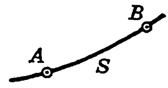


Figure 7.18

Problems 7.19

1 Tell how the length $|C|$ of a curve C is defined.

2 Show that if a curve C having finite length runs from A_1 through A_2 to A_3, then A_2 separates the curve into two parts each having finite length. It is not quite so easy to show that "the whole is equal to the sum of its parts," but it can be done. In particular, the length of a simple polygon is the sum of the lengths of its straight segments.

3 Show that the projection on the xy plane of a curve in E_3 is a curve in E_2.

4 Discover continuous functions x and y such that as t increases over the interval $0 \leqq t \leqq 4$, the point $(x(t), y(t))$ goes once in the positive direction around the square having vertices at the points (0,0), (0,1), (1,1), and (0,1). *Hint:* Write the equations which tell where a pencil point will be at time t if it traverses the square with unit speed. *One ans.:* $x(t) = t$ and $y(t) = 0$ when $0 \leqq t \leqq 1$; $x(t) = 1$ and $y(t) = t - 1$ when $1 \leqq t \leqq 2$; $x(t) = 3 - t$ and $y(t) = 1$ when $2 \leqq t \leqq 3$; $x(t) = 0$ and $y(t) = 4 - t$ when $3 \leqq t \leqq 4$.

5 Let L_1 be the length of the central path from A to B in Figure 7.191 and let L_2 be the length of the upper path. Find $L_2 - L_1$. *Ans.:* $(\pi - 2)h$.

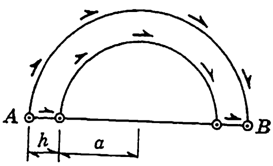


Figure 7.191

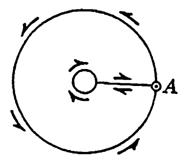


Figure 7.192

6 Supposing that the inner and outer circles of Figure 7.192 have radii ϵ and R, find the length of the curve which begins and ends at A.

7 Draw a square and from each corner draw a line segment toward the center but reaching only halfway to the center. Then insert arrows to specify a curve which forms the boundary of the inner region. Find the length of the curve.

8 This problem requires us to think about and calculate the lengths L_1 and L_2 of two of the paths by which an insect might crawl from the bottom B to the top T of the three-dimensional ring or anchor ring or torus of Figure 7.193. The first path, of length L_1, is a semicircle of radius $b + a$ which lies on the outer circumference of the torus. The second path, of length L_2, consists of a semicircle of radius a, and then a semicircle of radius $b - a$ which lies on the inner circumference of the torus, and finally another semicircle of a radius a. Try to guess which of the paths is shorter and then calculate L_1 and L_2. *Ans.:*

$$L_1 = \pi(a + b), \qquad L_2 = \pi a + \pi(b - a) + \pi a,$$

so the two paths have equal lengths. *Remark:* If a curve G lies in a set S and joins two points P_1 and P_2 of S, and if the length of G is less than or equal to the length of each other curve C in S which joins P_1 and P_2 then G is called a *geodesic* in S. As is easily imagined, the study of geodesics on a torus is an honorable part of an interesting subject.

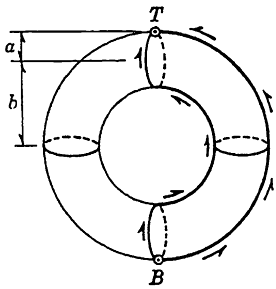


Figure 7.193

9 Creators of interesting tales for children say that, before the time when Columbus sailed across the ocean wet, everybody thought that the earth was flat. It has in fact been widely known for more than two thousand years that the earth is much like a spherical ball, but is not a spherical ball because mountains and valleys are quite noticeable around the Mediterranean sea and some other places. To show that a little information about lengths of arcs can have quite astonishing consequences, we look at the method said to have been used by the industrious Eratosthenes (*c.* 275 B.C. to *c.* 195 B.C.) to find what the radius (a number) of the earth would be if the earth were sandpapered to the shape of a spherical ball. Figure 7.194 shows O, the center of the earth, and a circular arc AB on the surface of the polished earth. The dotted vertical lines represent rays of light coming from the sun, and these rays are so nearly parallel that approximations can be based on a figure in which the rays are parallel to the line OA. Thus an observer at A finds that the sun is at his zenith. At B a pole BC of height h is erected in such a way that an observer at B thinks it is vertical, that is, the points O, B, C lie on the same line. In addition to h, two other lengths are obtained. In the first place, we find the length a of the circular arc DC which has its center at B (the base of the pole BC) and which has a shadow that just covers the pole. In the second place, the length b of the arc AB is obtained by more or less reliable surveyors. Let r be the radius of the earth. Since parallelism of the light rays implies that the angles DBC and AOB are equal, say to θ, we obtain the equation

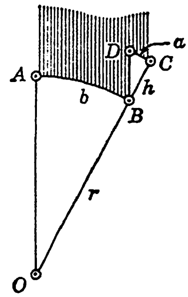


Figure 7.194

$$\frac{a}{h} = \frac{b}{r}$$

from the fact that each of the ratios is equal to θ. From this equation, r is easily calculated. It is not to be presumed that Eratosthenes spoke English and knew about miles, meters, and radians, but it is said that his calculations produced estimates of the radius and circumference of the earth that are almost as good as the estimates (radius 4000 miles, circumference 25,000 miles) that are ordinarily used for rough calculations. The fact that reliability of computed results depends upon accurate surveying is precisely the reason why two men named Mason and Dixon were commissioned to do some accurate surveying.

10 This problem, like some others, does not have a number for an answer. It requires us to think about connections between graphs and curves, and to learn some geometric terminology. The graph of a polynomial of degree n, that is, the graph of an equation of the form

$$(1) \qquad y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n,$$

in which $a_n \neq 0$, used to be (and occasionally still is) called a *parabola of degree n*. For example, a line is a parabola of degree 1 and a cubic is a parabola of degree 3. While this terminology is almost extinct, the graph of the equation

$$(2) \qquad y^2 = x^3$$

still is called a *semicubical parabola*, the old idea being that, when we solve for y, we get an exponent which is not an integer but is half of 3. According to this

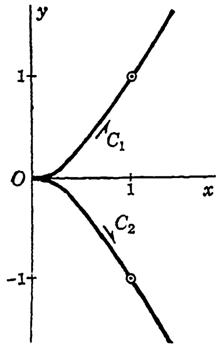


Figure 7.195

terminology, the semicubical parabola is a graph (not necessarily a curve) which is the point set shown in Figure 7.195. The graph has two *branches*, one being the upper branch consisting of points for which $y = x^{3/2}$, and the other being the lower branch consisting of points for which $y = -x^{3/2}$. Each of these branches contains the origin.

When the points on the upper branch are ordered in such a way that $P(x_1,y_1)$ precedes $P(x_2,y_2)$ when $x_1 < x_2$, the branch becomes a curve C_1 in the first quadrant having an end (or end point) at the origin. Similarly, when the points on the lower branch are ordered in such a way that $P(x_1,y_1)$ precedes $P(x_2,y_2)$ when $x_1 < x_2$. this branch becomes another curve C_2 having an end at the origin. With these orderings, the graph consists of the points on the two curves C_1 and C_2, the orderings being indicated by the arrows of Figure 7.195 which lie between the curves and the positive x axis. When $x > 0$, the curve C_1 has a tangent at the point $(x,x^{3/2})$ which has slope $\frac{3}{2}x^{1/2}$, and the curve C_2 has a tangent at the point $(x,-x^{3/2})$ which has slope $-\frac{3}{2}x^{1/2}$. We can observe that C_1 has a forward tangent at the origin whose slope is 0 and that C_2 has a forward tangent at the origin whose slope is 0. Although we have definitions involving tangents to graphs of equations of the form $y = f(x)$ and have definitions involving tangents to curves, we have no definitions which we can apply to decide whether the x axis is tangent to the semicubical parabola at the origin. In accordance with terminology other applications of which are not explained here, we say that the semicubical parabola has a *cusp* at the origin.

11 Prove that the curve C consisting of points P for which

$$\overrightarrow{OP} = t^2\mathbf{i} + t^3\mathbf{j} \tag{1}$$

coincides with the semicubical parabola having the equation $y^2 = x^3$. Sketch a new graph of the semicubical parabola and insert arrows which show how the points on the semicubical parabola are ordered to produce the single curve C. Put (1) in the equivalent form

$$\mathbf{r} = t^2\mathbf{i} + t^3\mathbf{j} \tag{2}$$

and, supposing that a particle moves along C in such a way that its displacement vector at time t is $\mathbf{r}$, find its velocity and acceleration at time t. See what happens when you try to write the equation of the line tangent to C at the point occupied by the particle at time $t = 0$.

12 The *folium of Descartes* is the graph of the parametric equations

$$x = \frac{3at}{1+t^3}, \qquad y = \frac{3at^2}{1+t^3}. \tag{1}$$

The vector $\mathbf{r}$ running from the origin to the point $P(x,y)$ on the folium is

$$\mathbf{r} = 3a\left[\frac{t}{1+t^3}\mathbf{i} + \frac{t^2}{1+t^3}\mathbf{j}\right], \tag{2}$$

there being one such point for each t for which $t \neq -1$. One who has considerable time to study the functions defined by (1) and to make some calculations can discover the manner in which P wanders as t increases. As t increases over the interval $-\infty < t < -1$, P runs over the curve C_1 extending from the origin (but not including the origin) down into the fourth quadrant. As t increases over the interval $-1 < t < \infty$, P traverses the curve C_2 which comes from the third quadrant to the origin and then runs around the loop toward the origin but does not again contain the origin. This problem requires that we obtain some more information without working so hard. Show that if (1) holds, then $y = tx$ and

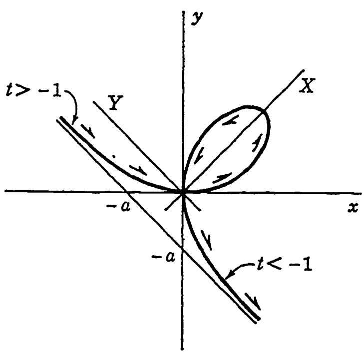


Figure 7.196

(3) $$x^3 + y^3 = 3axy.$$

Show that if (3) holds and $y = tx$, then (1) holds. Thus (3) is an equation of the folium. If $P(x,y)$ is a point on the folium, then (3) shows that $P(y,x)$ is also on the folium. Therefore, the line having the equation $y = x$ is an axis of symmetry of the folium. This suggests that we introduce the X, Y axes of Figure 7.196 which bear unit vectors $\mathbf{I}$ and $\mathbf{J}$. Use the formulas

(4) $$\mathbf{i} = \frac{1}{\sqrt{2}}(\mathbf{I} - \mathbf{J}), \qquad \mathbf{j} = \frac{1}{\sqrt{2}}(\mathbf{I} + \mathbf{J})$$

with (2) to obtain the new equation

(5) $$\mathbf{r} = \frac{3a}{\sqrt{2}}\left[\frac{t}{1 - t + t^2}\mathbf{I} + \frac{t(1-t)}{1+t^3}\mathbf{J}\right]$$

of the folium. Treating (5) as an equation of the form $\mathbf{r} = X\mathbf{I} + Y\mathbf{J}$, show that

$$X = \frac{3a}{\sqrt{2}}\frac{t}{1 - t + t^2}, \qquad \frac{dX}{dt} = \frac{3a}{\sqrt{2}}\frac{1 - t^2}{(1 - t + t^2)^2}$$

and use the result to find the minimum and maximum values of X and obtain more information about the folium.

13 Let f be continuous over the interval $0 \leqq x \leqq 1$ and let C be the set of points $(x, f(x))$ so ordered that $(x_1, f(x_1))$ precedes $(x_2, f(x_2))$ if $x_1 < x_2$. We may feel that C must have finite length, and we investigate the matter. Theorem 7.17 shows that C must have finite length if f is piecewise monotone, but we cannot be so sure when C is not piecewise monotone. When p and q are positive integers, the function $\sin^p x^q$ [or the function g for which $g(x) = \sin^p x^q$] is not piecewise monotone over the infinite interval $x > 1$ and the function $\sin^p (1/x^q)$ is not piecewise monotone over the interval $0 < x < 1$. Functions f for which $f(0) = 0$ and

$$f(x) = x^r \sin^p \frac{1}{x^q}$$

are, when $r > 0$, continuous but not piecewise monotone over $0 \leqq x \leqq 1$. Supposing that $p = q = r = 2$, show that $0 \leqq f(x) \leqq 1$ and f is differentiable as well as continuous over $0 \leqq x \leqq 1$. It can be proved that C does not have finite length.

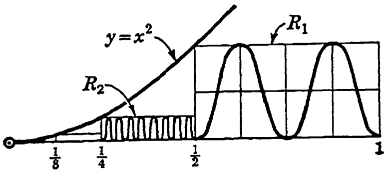


Figure 7.197

14 It is worthwhile to cultivate the ability to understand and even to construct simple examples of curves that have properties of various sorts. Figure 7.197 shows a part of the graph C of a function f, defined over $0 \leqq x \leqq 1$, for which $f'(x)$ exists when $0 \leqq x \leqq 1$ and, moreover, $f(x) = f'(x) = 0$ when $x = 0, 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots$. We do not bother to give formulas expressing $f(x)$ in terms of x, being content to describe the graph C. The rectangles have their upper left corners on the graph of the equation $y = x^2$. The part C_1 of the graph in the rectangle R_1 over the interval $\frac{1}{2} \leqq x \leqq 1$ consists of four monotone parts. Show that $1 \leqq |C_1| \leqq 2$. We now turn our attention to the rectangle R_2 which lies above the interval $\frac{1}{4} < x < \frac{1}{2}$ and has height $(\frac{1}{4})^2$. In this rectangle we construct another part C_2 of C, this part consisting of 4^2 monotone parts so that $1 \leqq |C_2| \leqq 2$. Tell how to continue the construction so that the total resulting curve C will not have finite length.

15 Let $x(t)$ and $y(t)$ be continuous over $a \leqq t \leqq b$ so that the set of points $P(t)$ for which $P(t) = (x(t), y(t))$ and $a \leqq t \leqq b$, so ordered that $P(t_1)$ precedes $P(t_2)$ when $t_1 < t_2$, constitutes a curve C in the xy plane. The curve C is said to be a *closed curve* if $P(a) = P(b)$, that is, if the curve ends where it starts. Give some examples of closed curves and some examples of curves that are not closed.

16 A set A in the xy plane is said to be *connected* if to each pair P_1 and P_2 of points in A there corresponds a curve C of the type described in Problem 15 such that $P(a) = P_1$, $P(b) = P_2$, and each point of C is a point of A. Let Γ (capital gamma) be the circle with center at P_0 and radius r. Prove that the set A_1 of points P for which $|\overrightarrow{P_0P}| < r$ (this set being the *interior* of Γ) is a connected set. Prove that the set A_2 of points P for which $|\overrightarrow{P_0P}| > r$ (this being the *exterior* of Γ) is a connected set. Prove finally that if a set A_3 contains a point P_1 for which $|\overrightarrow{P_0P_1}| < r$ and a point P_2 for which $|\overrightarrow{P_0P_2}| > r$ but contains no point of the circle Γ, then A_3 is not a connected set. *Proof of last part:* Suppose, intending to establish a contradiction, that the set A_3 is connected. Then there is a curve C, determined by continuous functions $x(t)$ and $y(t)$ as in Problem 15, such that $P(a) = P_1$, $P(b) = P_2$, and each point of C is a point of A_3. Let

$$f(t) = |\overrightarrow{P_0P(t)}| = \sqrt{(x_0 - x(t))^2 + (y_0 - y(t))^2}$$

so that $f(t)$ is the distance from P_0 to $P(t)$. Then $f(a) < r$ and $f(b) > r$. Since f is a continuous function, it follows from the intermediate-value theorem (Theorem 5.48) that there is a number t^* for which $a < t^* < b$ and $f(t^*) = r$. The point $P(t^*)$ is a point of the circle Γ, so $P(t^*)$ is not in A_3 and we have the required contradiction. *Remark:* The geometrical nature of a circle in a plane is so simple that basic facts involving its interior and exterior are easily described and easily proved. The remaining problems of this list involve more general situations.

17 Sketch some figures to obtain preliminary ideas about matters relating to the following definition. Let S be a given nonempty set in the xy plane and let P' be a point of S. Let S' be the set of points P'' in S that can be connected to P' by curves lying in S. This means that P'' is a point of S' if and only if there exist functions $x(t)$ and $y(t)$ continuous over an interval $a \leqq t \leqq b$ such that $P(x(a), y(a))$ is P', $P(x(b), y(b))$ is P'', and $P(x(t), y(t))$ is a point in S whenever $a \leqq t \leqq b$. The set S' is a connected set because if P_1 and P_2 lie in S', then P_1 and P_2 can be connected to P' by curves C_1 and C_2 lying in S' and these two curves can be combined to give a single curve C connecting P_1 to P_2 and lying in S'. This set S' is the *maximal connected subset* of S that contains P' or, briefly, the *component* of S that contains P'.

18 The curve C of Problem 15 is said to have a *multiple point* at the point Q if there exist two numbers t_1 and t_2 such that $a \leqq t_1 < t_2 < b$ (or $a < t_1 < t_2 \leqq b$) and the two points $P(t_1)$ and $P(t_2)$ coincide with Q. The curve is said to be *simple* (free from multiple points) if it has no multiple points. Give some examples of closed curves that have multiple points and some examples of simple closed curves.

19 The French mathematician Camille Jordan (1838–1922) was the first person to give serious attention to the difficult question whether each simple closed plane curve "separates the plane" into exactly two nonempty components one of which constitutes the outside of the curve and the other of which constitutes the inside. Such curves are called *Jordan curves.* The following theorem is known as the Jordan curve theorem.

Theorem *If C is a simple closed plane curve (or Jordan curve), then the set S of points of the plane which are not points of C contains exactly two (no more and no fewer) nonempty components S_1 and S_2.*

The author now owes his readers an explanation. It never has been and need not be expected that students of elementary calculus know anything about the Jordan curve theorem. Nevertheless, the author insists that each student should have opportunities to pick up ideas about mathematics. We can know that the Jordan curve theorem is so difficult that Jordan never succeeded in proving it and that we must learn more mathematics before we can undertake to understand proofs that have been given or to construct new proofs. We can know that many persons who never generated much interest in additions of fractions become intensely interested in problems involving sets and curves.

7.2 Lengths and integrals

Let x, y, and z be given functions having continuous derivatives over the closed interval $a \leqq t \leqq b$. Let

$$P(t) = P(x(t), y(t), z(t)) \tag{7.21}$$

so that, for each t, $P(t)$ is the point having coordinates displayed in (7.21). Let $\mathbf{r}(t)$ denote the vector running from the origin to $P(t)$ so that, for each t,

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}. \tag{7.22}$$

We propose to set up an integral for the length $|C|$ of the curve C traversed by $P(t)$, or by the tip of the vector $\mathbf{r}(t)$, as t increases from a to b. It could be supposed that our main interest is in lengths of curves and that we should get our result as quickly as possible, but this is not true. Integrals are very important things, and we proceed slowly to learn as much as we can about them.

Our first step is to make a partition P of the interval $a \leqq t \leqq b$ having partition points $t_0, t_1, \cdots, t_n$ and to sketch schematic figures showing the partition and points $P_{k-1} = P(t_{k-1})$ and $P_k = P(t_k)$ on C. We then tell ourselves that if $|P|$, the norm of the partition, is small then $|\overrightarrow{P_{k-1}P_k}|$ should be a good approximation to the length of the part of C between P_{k-1} and P_k, so $\Sigma|\overrightarrow{P_{k-1}P_k}|$ should be a good approximation to $|C|$ and hence it should be true that

$$|C| = \lim \Sigma|\overrightarrow{P_{k-1}P_k}|. \tag{7.23}$$

We have employed the fundamental ideas about length given in Section 7.1. The next step is to put the right member into a more useful form. While introduction of the abbreviations may be unwise when acres of paper and blackboard space are available, we simplify our formulas by setting

$$\Delta t_k = t_k - t_{k-1}, \qquad \Delta x_k = x(t_k) - x(t_{k-1}), \qquad \Delta y_k = y(t_k) - y(t_{k-1}),$$
$$\Delta z_k = z(t_k) - z(t_{k-1}).$$

Then (7.23) gives

$$|C| = \lim \Sigma \sqrt{\Delta x_k^2 + \Delta y_k^2 + \Delta z_k^2}. \tag{7.231}$$

We can make this look much more like a limit of Riemann sums by introducing factors Δt_k in numerators and denominators to obtain

$$|C| = \lim \sum \sqrt{\left(\frac{\Delta x_k}{\Delta t_k}\right)^2 + \left(\frac{\Delta y_k}{\Delta t_k}\right)^2 + \left(\frac{\Delta z_k}{\Delta t_k}\right)^2}\, \Delta t_k. \tag{7.232}$$

The possibility of making further progress is provided by the hypothesis that x, y, and z have continuous derivatives.

We should be realistic and realize that there are different ways to proceed. We can cheerfully adopt the view that, however we choose t_k^* in the kth subinterval of our partition, the kth term in (7.232) should be closely approximated by $f(t_k^*)\,\Delta t_k$, where

$$f(t) = \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}. \tag{7.233}$$

Thus we can expect that

$$|C| = \lim \Sigma f(t_k)\,\Delta t_k. \tag{7.234}$$

Since our hypothesis guarantees that f is continuous and hence integrable, the Riemann sums do have a limit and we are led to the formula

$$|C| = \int_a^b \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}\, dt. \tag{7.24}$$

This is, in fact, a correct and very satisfying formula which treats all coordinates alike and enables us to calculate lengths of curves that wind around through E_3. In case C lies in the xy plane, everything is the same except that $z(t) = z'(t) = 0$ and the formula is thereby simplified. In applications, it often happens that $x(t) = t$ and the formula is written in one or another of the forms

$$(7.241)\quad |C| = \int_a^b \sqrt{1 + [y'(t)]^2 + [z'(t)]^2}\,dt = \int_a^b \sqrt{1 + [y'(x)]^2 + [z'(x)]^2}\,dx = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2}\,dx$$

and, for plane curves, we have $z = 0$. It is not easy to be sure that the "cheerful derivation" of (7.24) is slovenly. A person can claim that he will not use the definition of length given in Section 7.1, that he is interested only in curves C for which $x'(t)$, $y'(t)$, and $z'(t)$ are continuous, and that all of the work that he has done merely motivates his definition whereby $|C|$ is *defined* by (7.24). His position is tenable and, so long as he stays in his own garden, it is even reasonable. In fact, many quantities in pure and applied mathematics are defined by integrals, and sometimes the definitions are so well motivated that readers (and even writers) fail to recognize the fundamental fact that it is utterly impossible to *prove* correctness of a formula for something that has not been defined.

We start again with the formula

$$(7.25)\qquad |C| = \lim \sum \sqrt{\left(\frac{\Delta x_k}{\Delta t_k}\right)^2 + \left(\frac{\Delta y_k}{\Delta t_k}\right)^2 + \left(\frac{\Delta z_k}{\Delta t_k}\right)^2}\,\Delta t_k,$$

which is (7.232) brought up where we can see it, but this time our situation is different. We suppose that $|C|$ is *defined* by (7.25), and we want to *prove* that the right side is an integral. For present purposes, the work of the preceding paragraph would be slovenly because nothing was proved. We could try to fix everything by trying to prove that there is a t_k^* such that, with the notation of the preceding paragraph, the kth term of (7.232) is exactly equal to $f(t_k^*)\,\Delta t_k$. We do not try this, however, because of the fear that the process would involve only uninformative hard work. Instead, we start by applying the mean-value theorem (Theorem 5.52) to (7.25). Since x' is continuous, there must be a t_k^* for which $t_{k-1} < t_k^* < t_k$ and

$$(7.251)\qquad \frac{\Delta x_k}{\Delta t_k} = \frac{x(t_k) - x(t_{k-1})}{t_k - t_{k-1}} = x'(t_k^*).$$

Similar formulas involving y and z end with $y'(t_k^{**})$ and $z'(t_k^{***})$, there being no reason for hope that the three numbers t_k^*, t_k^{**}, and t_k^{***} are the same. Thus

$$(7.252) \qquad |C| = \lim \Sigma \sqrt{[x'(t_k^*)]^2 + [y'(t_k^{**})]^2 + [z'(t_k^{***})]^2}\,\Delta t_k.$$

This looks much like a Riemann sum formed for the function f having values

$$(7.253) \qquad f(t) = \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2},$$

and there is no difficulty in the important special case in which $x(t) = t$ and $z(t) = 0$ for each t. Except in special cases, we must consider consequences of the fact that t_k^*, t_k^{**}, and t_k^{***} may be different. Since difficulties of this nature (and they were real difficulties in the old days) are usually associated with the name of Duhamel (1797–1872), we can call this a Duhamel difficulty. To treat all coordinates alike and to be precise about this matter, we let T_k be the center of the interval $t_{k-1} \leqq t \leqq t_k$. We can then put (7.252) in the form

$$(7.254) \qquad |C| = \lim_{|P|\to 0} \sum_{k=1}^{n} [f(T_k) + \delta_k]\,\Delta t_k,$$

where

$$(7.255) \quad \delta_k = \sqrt{[x'(t_k^*)]^2 + [y'(t_k^{**})]^2 + [z'(t_k^{***})]^2} \\ - \sqrt{[x'(T_k)]^2 + [y'(T_k)]^2 + [z'(T_k)]^2}.$$

To prove the desired result

$$(7.26) \quad |C| = \int_a^b f(t)\,dt = \int_a^b \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}\,dt,$$

it is therefore sufficient (and also necessary) to prove that

$$(7.261) \qquad \lim_{|P|\to 0} \sum_{k=1}^{n} \delta_k\,\Delta t_k = 0.$$

It is easy to originate the correct idea that the numbers t_k^*, t_k^{**}, and t_k^{***} are all near T_k, that continuity of the functions x', y', z' implies that the two terms on the right side of (7.255) are nearly equal, that the numbers δ_k are all near 0, and hence that the sum in (7.261) is near 0, whenever the norm $|P|$ of P is small. The easy way to make this precise is to use the fact that the first term in the right member of (7.255) is a continuous function of three variables. Using only functions of one variable, we can let $0 < \epsilon' < \epsilon$ and choose a number $\delta > 0$ such that the numbers α, β, and γ defined by

$$(7.262) \quad [x'(u)]^2 - [x'(v)]^2 = \alpha, \qquad [y'(u)]^2 - [y'(v)]^2 = \beta, \\ [z'(u)]^2 - [z'(v)]^2 = \gamma$$

all have absolute values less than ϵ' when $|u - v| < \delta$. When $|P| < \delta$, the right member of (7.255) then has the form

$$\delta_k = \sqrt{A_k \pm h} - \sqrt{A_k}, \tag{7.263}$$

where the quantities under the radicals are nonnegative and $0 \leqq h \leqq 3\epsilon'$. It follows that $|\delta_k| \leqq \sqrt{3\epsilon'}$ and hence

$$\left| \sum_{k=1}^{n} \delta_k \, \Delta t_k \right| \leqq \sum_{k=1}^{n} \sqrt{3\epsilon'} \, \Delta t_k = \sqrt{3\epsilon'} \, (b - a) < \epsilon \tag{7.264}$$

provided ϵ' is so chosen that the last inequality holds. This proves (7.261) and hence (7.26).

In order to introduce coordinates on curves, we suppose that x, y, and z are functions having continuous derivatives over an open interval $a < t < b$ and that C is the curve traversed by the point $P(t)$ having coordinates $(x(t), y(t), z(t))$ as t increases. To avoid difficulties, we suppose that the curve is *simple;* this means that $P(t_1) \neq P(t_2)$ when $t_1 \neq t_2$ so that $P(t)$ cannot be in the same place at two different times. Let t_0 be fixed such that $a < t_0 < b$. Then, as in Figure 7.27, we can assign coordinates to points on C by letting s be the length of the curve running from $P(t_0)$ to $P(t)$ when $t_0 \leqq t < b$ and letting $-s$ be the length of the curve running from $P(t)$ to $P(t_0)$ when $a < t \leqq t_0$. Then, whenever $a < t < b$, the coordinate $s(t)$ of P at time t is

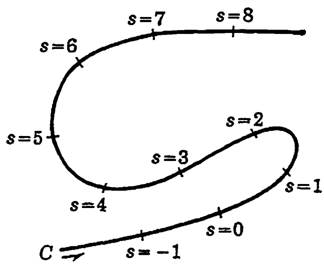


Figure 7.27

$$s(t) = \int_0^t \sqrt{[x'(\tau)]^2 + [y'(\tau)]^2 + [z'(\tau)]^2} \, d\tau. \tag{7.28}$$

In this formula τ (tau, the Greek t) is used as a dummy variable of integration. Since our hypothesis implies that the integrand is continuous, Theorem 4.35 enables us to differentiate with respect to t to obtain the formula

$$s'(t) = \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}, \tag{7.281}$$

which is often written in the form

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}. \tag{7.282}$$

This is related to the formulas we get when we set

$$\mathbf{r} = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} \tag{7.283}$$

and calculate the velocity $\mathbf{v}$ and the speed $|\mathbf{v}|$ from the formulas

$$\mathbf{v} = x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k} \tag{7.284}$$

and

$$|\mathbf{v}| = \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}. \tag{7.285}$$

Since the right members of (7.282) and (7.285) are the same, this shows that, in appropriate circumstances, the speed determined so quickly by taking the absolute value of the velocity is the same as the "speed of the particle in its path" determined by use of coordinates on the path.

Problems 7.29

1 Find the length of the part of the helix

$$x = a\cos t, \qquad y = a\sin t, \qquad z = kt$$

traversed by a particle at $P(x,y,z)$ as t increases from 0 to 2π.

2 Using the first standard equation in

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad x = a\cos\theta, \qquad y = b\sin\theta,$$

where $0 < b < a$, set up an integral for the length of the part of the ellipse lying in the first quadrant. Then use the parametric equations to set up another integral for the same length

Ans.: $$\int_0^a \sqrt{\frac{a^2 - k^2x^2}{a^2 - x^2}}\,dx, \qquad a\int_0^{\pi/2} \sqrt{1 - k^2\cos^2\theta}\,d\theta,$$

where $k^2 = (a^2 - b^2)/a^2$. These are *elliptic integrals*, and others more or less like them are also called elliptic integrals. There is no easy way to obtain their exact numerical values.

3 Set up an integral for the length L of the curve traced by the point P having coordinates

$$x = a\theta - b\sin\theta, \qquad y = a - b\cos\theta$$

as θ increases from 0 to 2π, and show that the result can be put in the form

$$L = (a + b)\int_0^{2\pi} \sqrt{1 - \frac{4ab}{(a+b)^2}\cos^2\frac{\theta}{2}}\,d\theta.$$

When $b \neq a$, this is an elliptic integral. When $b = a$, the graph is, as Problem 16 of Section 6.5 shows, an ordinary cycloid. Sketch a figure and explain a simple geometric argument which shows that, when $b = a$ and the approximations $\pi = 3$ and $\pi^2 = 10$ are used, $\sqrt{52}\,a < L < 10a$. Finally, show that $L = 8a$.

4 Find the length of the part of the curve having the equations

$$x = 2t^3, \qquad y = \frac{t^2 - 3}{2}, \qquad z = \frac{2 - t^2}{2}$$

which lies between the points $(0, -\frac{3}{2}, 1)$ and $(2, -1, \frac{1}{2})$. *Ans.:* $L = \dfrac{\sqrt{2}}{54}[19^{3/2} - 1]$.

5 When $0 < b < a$ and a circle (or cog wheel) of radius b rolls without slipping inside a circle of radius a as in Figure 7.291, the path traced by a point on the small rolling circle is a hypocycloid (inside-cycloid) that appeared in one of Problems 6.59 and that even small dictionaries describe. It can be shown that when the center of the small circle has passed over an angle θ (that could be ωt), the point initially at $(a,0)$ has coordinates

$$x = (a - b)\cos\theta + b\cos\frac{a-b}{b}\theta$$
$$y = (a - b)\sin\theta - b\sin\frac{a-b}{b}\theta.$$

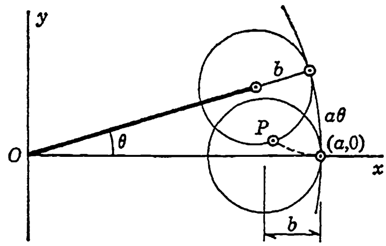


Figure 7.291

With the aid of Figure 7.291 find the amount by which θ must increase as P traverses one arch of the cycloid from the big circle back to the big circle. Then find the length of one arch of the cycloid. *Ans.:* $8b(a - b)/a$.

6 It is a remarkable fact that when $b = a/4$, so that the hypocycloid of the preceding problem has four cusps, the equations can be put in the form

$$x = \frac{a}{4}(3\cos\theta + \cos 3\theta) = a\cos^3\theta$$
$$y = \frac{a}{4}(3\sin\theta - \sin 3\theta) = a\sin^3\theta$$

so that

$$x^{2/3} + y^{2/3} = a^{2/3}.$$

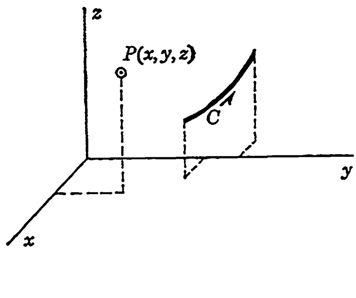
Using this formula, find the length of the part of the hypocycloid in the first quadrant. *Ans.:* $3a/2$.

7 Let f and g have continuous derivatives over $a \leqq u \leqq b$. Let C be the curve which the point $P(f(u), g(u))$ traverses in an xy plane as u increases from a to b. Let C be regarded as a wire having linear density δ so that a piece of C of length L has mass δL. Let p be a nonnegative integer and let x_0 be a constant. Set up an integral for the pth moment of the wire about the line having the equation $x = x_0$.

8 With or without assistance from the preceding problem, set up an integral for the pth moment about a diameter of a circular wire having linear density δ.

9 Centroids of triangles, rectangles, and regular polygons coincide with the centroids of the regions that they bound. Draw some polygons for which there is violent departure from coincidence.

Figure 7.292



10 Figure 7.292 is intended to steer our thoughts toward a wire or cord concentrated on a curve C and to make us realize that we have not yet calculated the gravitational force $\mathbf{F}$ upon a particle P^* of mass m at $P(x,y,z)$ that is produced by the wire. Set up an integral for $\mathbf{F}$. *Outline of solution:* Since x, y, z have been preempted, we take the curve C to be the ordered set of points $P(t) = P(u,v,w)$ for which u, v, w are functions of t having continuous derivatives and

$$(1) \qquad \overrightarrow{OP} = u(t)\mathbf{i} + v(t)\mathbf{j} + w(t)\mathbf{k} \qquad (a \leqq t \leqq b),$$

and suppose that the linear density (mass per unit length) of C at $P(t)$ is $\delta(t)$, where δ is a continuous function of t. Letting Q be a partition of the interval $a \leqq t \leqq b$ with partition points t_k and supposing that $t_{k-1} \leqq t_k^* \leqq t_k$, we take the number $\delta(t_k^*)\,|\overrightarrow{P_{k-1}P_k}|$ as an approximation to the mass of an element of the wire. The force $\Delta\mathbf{F}_k$ which this element produces upon the particle P^* of mass m at $P(x,y,z)$ should have approximate magnitude

$$(2) \qquad \frac{Gm\delta(t_k^*)|\overrightarrow{P_{k-1}P_k}|}{|\overrightarrow{PP}(t_k^*)|^2}.$$

Introducing vectors and coordinates gives the approximation

$$(3) \quad \Delta\mathbf{F}_k = Gm\delta(t_k^*)\sqrt{\left(\frac{\Delta u_k}{\Delta t}\right)^2 + \left(\frac{\Delta v_k}{\Delta t}\right)^2 + \left(\frac{\Delta w_k}{\Delta t}\right)^2}\; \frac{[u(t_k^*) - x]\mathbf{i} + [v(t_k^*) - y]\mathbf{j} + [w(t_k^*) - z)]\mathbf{k}}{\{[u(t_k^*) - x]^2 + [v(t_k^*) - y]^2 + [w(t_k^*) - z]^2\}^{3/2}}\,\Delta t.$$

The limit of the sum of these things should be $\mathbf{F}$, and we are led to the definition

$$(4) \quad \mathbf{F} = Gm\int_a^b \delta(t)\sqrt{[u'(t)]^2 + [v'(t)]^2 + [w'(t)]^2}\; \frac{[u(t) - x]\mathbf{i} + [v(t) - y]\mathbf{j} + [w(t) - z]\mathbf{k}}{\{[u(t) - x]^2 + [v(t) - y]^2 + [w(t) - z]^2\}^{3/2}}\,dt.$$

This is our result. In case the wire coincides with a circle of radius a in the yz plane having its center at the origin, we can set $u(t) = 0$, $v(t) = a\cos t$, $w(t) = a\sin t$ and put (4) in the form

$$(5) \qquad \mathbf{F} = Gma\int_0^{2\pi} \delta(t)\,\frac{-x\mathbf{i} + [a\cos t - y]\mathbf{j} + [a\sin t - z]\mathbf{k}}{\{x^2 + [a\cos t - y]^2 + [a\sin t - z]^2\}^{3/2}}\,dt.$$

In case P^* is on the axis of the wire so that $y = z = 0$, this takes the much simpler form

$$(6) \qquad \mathbf{F} = Gma\int_0^{2\pi} \delta(t)\,\frac{-x\mathbf{i} + a\cos t\mathbf{j} + a\sin t\mathbf{k}}{(x^2 + a^2)^{3/2}}\,dt.$$

In case the wire is a uniform wire so that, for some constant δ_0 we have $\delta(t) = \delta_0$ for each t, the coefficients of $\mathbf{j}$ and $\mathbf{k}$ are zero because

$$(7) \qquad \int_0^{2\pi} \cos t\,dt = \int_0^{2\pi} \sin t\,dt = 0$$

(or, as is usually said, "because of symmetry") and (6) reduces to

$$(8) \qquad \mathbf{F} = -\frac{GmMx}{(x^2 + a^2)^{3/2}}\,\mathbf{i},$$

where $M = 2\pi a\delta_0$, the total mass of the wire. *Remark:* With slight differences in notation, (8) was used extensively in some of the Problems 4.79.

11 Let units of measurement be so employed that the potential at $P_0(x_0,y_0,z_0)$ due to a particle of mass m concentrated at a point $P(x,y,z)$ is $m/|\overline{P_0P}|^2$. Let x, y, z have continuous derivatives over $a \leqq u \leqq b$, and let $P(x,y,z)$ traverse a curve C as u increases from a to b. Let C have linear density δ. Let $P_0(x_0,y_0,z_0)$ be a point not on C. Remembering that potentials are scalars, and remembering or learning quickly that the potential due to a collection of particles is equal to the sum of the potentials due to the individual particles, set up an integral which is equal to the potential at P_0 due to C.

12 Set up an integral equal to the moment of inertia (second moment) about the z axis of the curve C of the preceding problem.

13 When a particle has fallen (or slid without friction) from height $G(b)$ to a lower height $G(u)$, its speed is $\sqrt{2g[G(b) - G(u)]}$, where g is the acceleration of gravity. We all know that if we travel a short distance with a constant speed, we get the time required for the trip by dividing the distance by the speed. With this basic information, we are prepared to attack a problem. Let F and G be functions having positive continuous derivatives over $0 \leqq u \leqq b$ and let the point $P(F(u), G(u))$ traverse a curve C from the origin to a point B in the first quadrant of an xy plane as u increases from 0 to b. Sketch a schematic figure showing the curve C. Set up an integral which gives the time T required for a particle starting from rest at B to slide without friction down the curve C to the origin.

14 The reflecting surface of a headlight is a part of a paraboloid, of depth 4 inches and diameter 12 inches, obtained by rotating a part of a parabola about its axis. Find its area. *Solution:* This is not an easy problem, since a basic difficulty lurks in the fact that areas of such surfaces have not been defined. We start by so determining k that the parabola having the equation $y = kx^2$ contains (or passes through) the point (6,4). This gives $y = \frac{1}{9}x^2$ for the equation of the generating parabola. With the possibility of setting $a = 0$, $b = 6$, and $f(x) = \frac{1}{9}x^2$, we suppose that f has a continuous derivative over an interval $a \leqq x \leqq b$. Let S be the surface generated by rotating, about the y axis, the part of the graph G of $y = f(x)$ for which $a \leqq x \leqq b$. Expecting to use some intuitive ideas about areas, we make a partition P of the interval $a \leqq x \leqq b$ and consider one particular subinterval $x_{k-1} \leqq x \leqq x_k$. As an approximation to the length L_k of the segment G_k of the graph G containing points x for which $x_{k-1} \leqq x \leqq x_k$, we may use the number

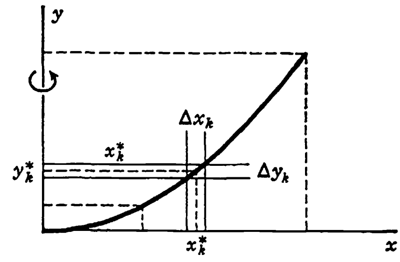


Figure 7.293

$$(1) \qquad \sqrt{\Delta x_k^2 + \Delta y_k^2} = \sqrt{1 + \left(\frac{\Delta y_k}{\Delta x_k}\right)^2}\,\Delta x_k = \sqrt{1 + [f'(x_k^*)]^2}\,\Delta x_k,$$

where $\Delta x_k = x_k - x_{k-1}$, $\Delta y_k = f(x_k) - f(x_{k-1})$, and x_k^* is a number for which $x_{k-1} < x_k^* < x_k$. Figure 7.293 is helpful. Even though areas are not yet defined. we can have a feeling that when the segment G_k is rotated about the y axis, its points all travel approximately the same distance $2\pi x_k$ and that the segment generates a surface essentially like a ribbon having width L_k and length $2\pi x_k$. This leads us to the intuitive idea that one term in the sum

$$(2) \qquad |S| = \lim \Sigma 2\pi \sqrt{1 + [f'(x_k)]^2}\, x_k\, \Delta x_k$$

should, when the norm $|P|$ of P is small, be a good approximation to the area of the part of S generated by one segment of the graph G. The next step is to adopt the tentative intuitive conclusion that the sum is a good approximation to the total area $|S|$ of S when the norm of P is small or, in other words, that (2) should be valid. Our theory of Riemann sums and integrals assures us that if (2) is true, then

$$(3) \qquad |S| = 2\pi \int_a^b \sqrt{1 + [f'(x)]^2}\, x\, dx.$$

If we have sufficient confidence in our calculations (it would, of course, be fatal to use an incorrect formula for the circumference of a circle of radius x_k) and in our intuitive ideas, we can install the formula (3) as one (not *the*) definition of area of surfaces of revolution. It does not make sense to claim that this definition is "correct," but experience shows that it is useful and this is all that we can expect from definitions. For the case in which $f(x) = kx^2$, $a = 0$, and $b = r$, (3) reduces to

$$|S| = 2\pi \int_0^r (1 + 4k^2x^2)^{1/2} x\, dx = \frac{\pi}{6k^2}[(1 + 4k^2r^2)^{3/2} - 1].$$

When $k = \frac{1}{8}$ and $r = 6$, this reduces to $|S| = 49\pi$.

15 We invest a little time to look at some *curve integrals* that are called *line integrals* by those who adhere to the notion that curves are lines. Let functions $x(t)$, $y(t)$, $z(t)$, the point $P(t)$, the vector $\mathbf{r}(t)$, the curve C, the partition P of $a \leqq t \leqq b$, and the numbers Δt_k, Δx_k, Δy_k, Δz_k be defined as in the part of this section preceding (7.231). Let a vector function $\mathbf{F}$ having scalar components f_1, f_2, f_3 be defined and continuous over a part of E_3 that contains the curve C. We consider $\mathbf{F}$ to be a force which operates upon a particle at $P(t)$ as the particle moves from $P(a)$ to $P(b)$.

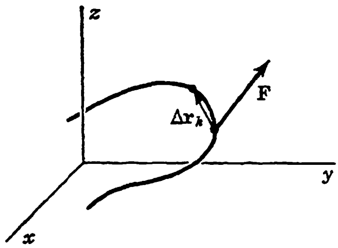


Figure 7.294

The schematic Figure 7.294 may be helpful. With the idea that the work done by $\mathbf{F}$ as the particle moves from $P(t_{k-1})$ to $P(t_k)$ is closely approximated by

$$(1) \qquad \mathbf{F}(x_k,y_k,z_k)\cdot\Delta\mathbf{r}_k$$

when the norm of the partition P is small, we can define the total work W done by $\mathbf{F}$ by the formula

$$(2) \qquad W = \lim \Sigma \mathbf{F}(x_k,y_k,z_k)\cdot\Delta\mathbf{r}_k.$$

The right member of (2) is an example of a *curve integral*, and it is denoted by the symbol in

(3) $$W = \int_C \mathbf{F}(\mathbf{r})\cdot d\mathbf{r},$$

where $\mathbf{F}(\mathbf{r})$ is the vector function of the vector $\mathbf{r}$ defined by

(4) $$\mathbf{F}(\mathbf{r}) = \mathbf{F}(x,y,z)$$

in which x, y, z are the numbers for which $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. The right member of (3) is read "the integral over C of $\mathbf{F}(\mathbf{r})$ dot $d\mathbf{r}$," the fundamental idea being that it is the curve C that is partitioned to produce the approximating sums. To learn something about this curve integral, we can use the formulas

(5) $$\mathbf{F}(x,y,z) = f_1(x,y,z)\mathbf{i} + f_2(x,y,z)\mathbf{j} + f_3(x,y,z)\mathbf{k}$$
(6) $$\Delta\mathbf{r}_k = \Delta x_k\mathbf{i} + \Delta y_k\mathbf{j} + \Delta z_k\mathbf{k}$$

to put (2) in the form

(7) $$W = \lim \sum \left[f_1(x(t_k), y(t_k), z(t_k)) \frac{\Delta x_k}{\Delta t_k} + f_2(x(t_k), y(t_k), z(t_k)) \frac{\Delta y_k}{\Delta t_k} + f_3(x(t_k), y(t_k), z(t_k)) \frac{\Delta z_k}{\Delta t_k} \right] \Delta t_k.$$

A proof very similar to the one centering around (7.255) enables us to show that

(8) $$W = \int_a^b f_1(x(t), y(t), z(t))x'(t)\,dt + \int_a^b f_2(x(t), y(t), z(t))y'(t)\,dt + \int_a^b f_3(x(t), y(t), z(t))z'(t)\,dt.$$

This is a formula from which W can be calculated or approximated. Since we are not electronic computers, we do not (at least at the present time) make calculations, but we do point out that the integrals in (8) are abbreviated to those in the formula

(9) $$W = \int_C [f_1(x,y,z)\,dx + f_2(x,y,z)\,dy + f_3(x,y,z)\,dz].$$

The integrals in (9) are scalar curve integrals. Whenever we want to know what these things mean and how they can be evaluated, we should have the wits to check back to see what they abbreviate. While it may be possible to overemphasize the importance of the matter, we can observe that if we set

(10) $$\mathbf{F}(\mathbf{r}) = f_1(x,y,z)\mathbf{i} + f_2(x,y,z)\mathbf{j} + f_3(x,y,z)\mathbf{k}$$

and make the pretense that $d\mathbf{r}$ is a vector for which

(11) $$d\mathbf{r} = dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}$$

and

(12) $$\mathbf{F}(\mathbf{r})\cdot d\mathbf{r} = f_1(x,y,z)\,dx + f_2(x,y,z)\,dy + f_3(x,y,z)\,dz,$$

then we can substitute (12) into (3) to obtain (9). Of course, it should be thoroughly understood that we have done nothing but explain symbols. This would be a stupid waste of time if it were not for the fact the symbols are useful. They appear even in quite elementary physics and engineering, and mathematicians have a responsibility to tell what they mean.

7.3 *Center and radius of curvature* Courses and textbooks in "differential geometry" provide information about matters relating to curvature of curves that lie in E_3. In this section, we give most of our attention to curves that lie in an xy plane. Our approach to the subject sacrifices brevity to place emphasis upon elementary geometric ideas that can be of interest to everyone and are needed by engineers and others who study the bending of beams.

Let x and y, or $x(t)$ and $y(t)$, have continuous second derivatives over some open interval $a < t < b$ in which t and $t + \Delta t$ are always supposed to lie. Let $P(t)$ denote the point with coordinates $(x(t), y(t))$ and let $\mathbf{r}(t)$ be the vector running from the origin to $P(t)$ so that

$$\text{(7.311)} \qquad \mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$$
$$\text{(7.312)} \qquad \mathbf{v}(t) = x'(t)\mathbf{i} + y'(t)\mathbf{j}$$
$$\text{(7.313)} \qquad \mathbf{a}(t) = x''(t)\mathbf{i} + y''(t)\mathbf{j}.$$

Let C be the curve traversed by $P(t)$ and the tip of $\mathbf{r}(t)$ as t increases, so that $\mathbf{v}(t)$ is tangent to C at $P(t)$ when $\mathbf{v}(t) \neq 0$. Henceforth we consider only values of t for which

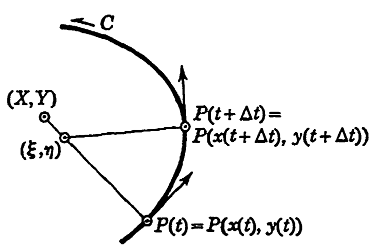
$$\text{(7.32)} \qquad \mathbf{v}(t) \times \mathbf{a}(t) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x'(t) & y'(t) & 0 \\ x''(t) & y''(t) & 0 \end{vmatrix} = [x'(t)y''(t) - x''(t)y'(t)]\mathbf{k} \neq 0.$$

Since $|\mathbf{v} \times \mathbf{a}| = |\mathbf{v}|\,|\mathbf{a}|\,|\sin\theta|$, this means that $\mathbf{v}$ and $\mathbf{a}$ are nonzero vectors which do not lie on parallel lines. This and the first of the formulas

$$\text{(7.321)} \qquad \lim_{\Delta t \to 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t} = \mathbf{a}(t), \qquad \mathbf{v}(t) \times \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t} \neq 0$$

implies existence of a $\delta > 0$ such that the second holds when $0 < |\Delta t| < \delta$. Since $\mathbf{v}(t) \times \mathbf{v}(t) = 0$, we conclude that $\mathbf{v}(t) \times \mathbf{v}(t + \Delta t) \neq 0$ and hence that the tangents $\mathbf{v}(t)$ and $\mathbf{v}(t + \Delta t)$ do not lie on parallel lines when $0 < |\Delta t| < \delta$.

Figure 7.33



Supposing that $0 < |\Delta t| < \delta$, we construct Figure 7.33 and look at it. Since the tangents $\mathbf{v}(t)$ and $\mathbf{v}(t + \Delta t)$ to C at $P(t)$ and $P(t + \Delta t)$ do not lie on parallel lines, the normals to C at these points must intersect at a point (ξ,η) having coordinates ξ (xi) and η (eta). If C is a circle, then the most elementary of elementary geometry tells us that the points (ξ,η) obtained for different values of Δt would all lie at the center (X,Y) of the circle. This is no time to be devoid of ideas. We can have at least a vague feeling that, even when our curve C is not a circle, a small section of C in a neighborhood of $P(t)$ should be so much like a circle that there should be a point (X,Y) such that (ξ,η) is near (X,Y) whenever Δt is small. If all this happens, we should give the point (X,Y) a name and find formulas for X and Y. All this does happen. The point is called the *center of curvature* of C at $P(t)$. We shall find formu'as for X and Y and for the distance from (X,Y) to $P(t)$. This distance is called the *radius of curvature* ρ (rho) of C at $P(t)$. The circle with center at (X,Y) and radius ρ is called the *circle of curvature* of C at $P(t)$.

The coordinates ξ and η are determined by the system of equations

$$\begin{aligned}[]
[\xi - x(t)]x'(t) \quad &+ \quad [\eta - y(t)]y'(t) = 0\\
[\xi - x(t+\Delta t)]x'(t+\Delta t) &+ [\eta - y(t+\Delta t)]y'(t+\Delta t) = 0.
\end{aligned}$$

The left member of the first equation is the scalar product of the vector $\mathbf{v}(t)$ tangent to C at $P(t)$ and the vector running from $P(t)$ to (ξ,η), and the equation expresses the fact that the two vectors are orthogonal. Similarly, the second equation expresses the fact that (ξ,η) lies on the normal to C at $P(t+\Delta t)$. Replacing the quantities in brackets in the second equation by

$$[\xi - x(t) - x(t+\Delta t) + x(t)] \qquad \text{and} \qquad [\eta - y(t) - y(t+\Delta t) + y(t)]$$

and transposing a part of the result enables us to put the two equations in the form

$$[\xi - x(t)]x'(t) \qquad + [\eta - y(t)]y'(t) \qquad = 0 \tag{7.341}$$

$$[\xi - x(t)]x'(t+\Delta t) + [\eta - y(t)]y'(t+\Delta t) = Q, \tag{7.342}$$

where we have simplified matters by letting Q denote the quantity

$$Q = [x(t+\Delta t) - x(t)]x'(t+\Delta t) + [y(t+\Delta t) - y(t)]y'(t+\Delta t). \tag{7.343}$$

To eliminate η from the two equations (7.341) and (7.342), we multiply the first by $y'(t+\Delta t)$ and the second by $-y'(t)$ and add to obtain the first of the formulas

$$D[\xi - x(t)] = -Qy'(t), \qquad D[\eta - y(t)] = Qx'(t), \tag{7.35}$$

where D is the determinant

$$D = x'(t)y'(t+\Delta t) - x'(t+\Delta t)y'(t)$$

which can be put in the form

$$D = x'(t)[y'(t+\Delta t) - y'(t)] - [x'(t+\Delta t) - x'(t)]y'(t). \tag{7.351}$$

A similar procedure, in which the multipliers $-x'(t + \Delta t)$ and $x'(t)$ are used, gives the second of the formulas (7.35). Formulas for ξ and η are now obtained from (7.35) by dividing and transposing. As we shall see, formulas for X and Y are obtained by dividing (7.35) by Δt and taking limits as $\Delta t \to 0$. Since (7.343) and (7.351) show that

$$\lim_{\Delta t \to 0} \frac{Q}{\Delta t} = [x'(t)]^2 + [y'(t)]^2$$
$$\lim_{\Delta t \to 0} \frac{D}{\Delta t} = x'(t)y''(t) - x''(t)y'(t),$$

we find from (7.35) that

$$(7.361) \quad X - x(t) = \lim_{\Delta t \to 0} [\xi - x(t)] = -\frac{[x'(t)]^2 + [y'(t)]^2}{x'(t)y''(t) - x''(t)y'(t)}\, y'(t)$$

$$(7.362) \quad Y - y(t) = \lim_{\Delta t \to 0} [\eta - y(t)] = \frac{[x'(t)]^2 + [y'(t)]^2}{x'(t)y''(t) - x''(t)y'(t)}\, x'(t).$$

Transposing the terms $x(t)$ and $y(t)$ gives formulas

$$(7.363) \qquad X = x(t) - \frac{[x'(t)]^2 + [y'(t)]^2}{x'(t)y''(t) - x''(t)y'(t)}\, y'(t)$$

$$(7.364) \qquad Y = y(t) + \frac{[x'(t)]^2 + [y'(t)]^2}{x'(t)y''(t) - x''(t)y'(t)}\, x'(t)$$

for the coordinates (X,Y) of the center of curvature of C at the point $(x(t), y(t))$, but these new formulas are sometimes less useful than their parents. The definition of radius of curvature ρ implies that

$$\rho = \sqrt{[X - x(t)]^2 + [Y - y(t)]^2}$$

and hence

$$(7.37) \qquad \rho = \frac{\{[x'(t)]^2 + [y'(t)]^2\}^{3/2}}{|x'(t)y''(t) - x''(t)y'(t)|}.$$

In case $x(t) = t$ so $x'(t) = 1$ and $x''(t) = 0$, it is customary to replace t by x and write

$$(7.371) \qquad \rho = \frac{\{1 + [y'(x)]^2\}^{3/2}}{|y''(x)|} = \frac{\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{3/2}}{\left|\frac{d^2y}{dx^2}\right|}.$$

To steer our thoughts toward another problem involving curvature, we consider the rate of change of direction of an automobile which traverses a level road that winds around swamps and between mountains. At each time t the rate depends upon the speed of the automobile and upon another number which is called the curvature of the road at the position of the automobile. To be more precise about this matter, let

x and y be functions of t having two derivatives each over some interval and let C be the curve traversed by the point $P(t)$, having coordinates $x(t)$, $y(t)$, as t increases over the interval. We suppose that there is no t for which $x'(t)$ and $y'(t)$ are both 0. Our results will be obtained with the aid of the three familiar vectors

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} \tag{7.381}$$
$$\mathbf{v}(t) = x'(t)\mathbf{i} + y'(t)\mathbf{j} \tag{7.382}$$
$$\mathbf{a}(t) = x''(t)\mathbf{i} + y''(t)\mathbf{j}. \tag{7.383}$$

The vector $\mathbf{v}(t)$ having its tail at $P(t)$ is the forward tangent to C at $P(t)$. With each t we wish to associate an angle $\phi(t)$ that determines the direction of $\mathbf{v}(t)$, and this is a matter that must not be treated carelessly. To restrict $\phi(t)$ to an interval like $-\pi < \phi \leqq \pi$, so that ϕ would be discontinuous when the vector rotates from northwest to southwest, would defeat our purpose. To formulate general principles by which $\phi(t)$ can be calculated may be beyond our capabilities. Let us then avoid possible unforeseen topological difficulties by restricting attention to curves C for which it is clearly possible to determine $\phi(t)$ by the following procedure. Let t_0 be a particular t in the interval considered and let $\phi(t_0)$ be the angle for which $-\pi < \phi(t_0) \leqq \pi$ and $\phi(t_0)$ is the ordinary trigonometric angle having its initial side on the positive x axis and its terminal side on the vector through the origin parallel to $\mathbf{v}(t_0)$. Then, as t increases (or decreases) from t_0, let $\phi(t)$ vary with the vector $\mathbf{v}(t)$ in such a way that ϕ is continuous. Some curves are less complicated than that shown in Figure 7.384, and some are more complicated. Supposing that $\phi(t)$ is

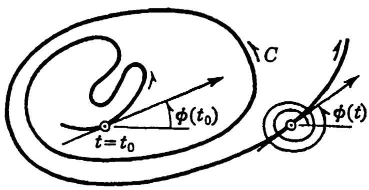


Figure 7.384

satisfactorily determined, we can put the formula (7.382) for $\mathbf{v}(t)$ in the form

$$|\mathbf{v}(t)|[\cos\phi(t)\mathbf{i} + \sin\phi(t)\mathbf{j}] = \mathbf{v}(t), \tag{7.385}$$

where

$$x'(t) = |\mathbf{v}(t)|\cos\phi(t), \qquad y'(t) = |\mathbf{v}(t)|\sin\phi(t) \tag{7.3851}$$

and

$$|\mathbf{v}(t)| = \{[x'(t)]^2 + [y'(t)]^2\}^{1/2} \neq 0. \tag{7.3852}$$

Since we want a formula for $\phi'(t)$, it is reasonable to differentiate (7.385) with respect to t and to try to use the result. We divide by $|\mathbf{v}(t)|$ and then differentiate to obtain

$$(7.386)\quad [-\sin\phi(t)\mathbf{i} + \cos\phi(t)\mathbf{j}]\phi'(t) = \frac{|\mathbf{v}(t)|\mathbf{a}(t) - \mathbf{v}(t)\dfrac{d}{dt}|\mathbf{v}(t)|}{|\mathbf{v}(t)|^2}.$$

The coefficient of $\phi'(t)$ is the unit vector $\mathbf{n}$ for which

$$(7.3861)\quad \mathbf{n} = [-\sin\phi(t)\mathbf{i} + \cos\phi(t)\mathbf{j}] = \frac{-y'(t)\mathbf{i} + x'(t)\mathbf{j}}{|\mathbf{v}(t)|}.$$

Since $\mathbf{n}\cdot\mathbf{v}(t) = 0$, we can equate the scalar products of $\mathbf{n}$ and the members of (7.386) to obtain

$$\phi'(t) = \frac{|\mathbf{v}(t)|\mathbf{a}(t)\cdot\mathbf{n}(t)}{|\mathbf{v}(t)|^2} = \frac{[x''(t)\mathbf{i} + y''(t)\mathbf{j}]\cdot[-y'(t)\mathbf{i} + x'(t)\mathbf{j}]}{|\mathbf{v}(t)|^2}$$

and hence

$$(7.387)\quad \phi'(t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{[x'(t)]^2 + [y'(t)]^2}.$$

For some applications of (7.387), we can set $x(t) = t$ so that $x'(t) = 1$ and $x''(t) = 0$. In such cases, we can replace t by x to obtain

$$(7.3871)\quad \phi'(x) = \frac{y''(x)}{1 + [y\,(x)]^2}.$$

In this, and in any other case in which $x'(t) \neq 0$, we can eliminate practically all of the work of this section by writing

$$(7.3872)\quad \tan\phi = \frac{y'(t)}{x'(t)}$$

with or without the aid of (7.3851) and then differentiating with respect to t to obtain (7.387).

Returning to (7.387), we note that $\phi'(t)$, which might be measured in radians per minute, gives the time rate of change of ϕ with respect to t. In case the point $P(t)$ traverses C with unit speed, which might be 1 kilometer per minute, $\phi'(t)$ becomes also a number of radians per unit distance measured along C, and this is called the *curvature* K or $K(t)$ of C at $P(t)$. To give curvature an additional leg upon which to stand, we introduce upon C a coordinate system like that in Figure 7.27 with the stipulation that the coordinate s of $P(t)$ increases as t increases. The curvature of C at $P(t)$ can then be defined by the formula

$$(7.388)\quad K = K(t) = \frac{d\phi}{ds}.$$

Since $d\phi/ds = \phi'(t)/s'(t) = \phi'(t)/|v(t)|$, the formulas (7.387) and (7.3852) yield the formula

$$K = \frac{x'(t)y''(t) - x''(t)y'(t)}{\{[x'(t)]^2 + [y'(t)]^2\}^{3/2}} \tag{7.389}$$

from which the curvature of C at $P(t)$ can be calculated without reference to other formulas. Perhaps we should take notice of the fact that a curve C is an ordered set of points, and that we can misunderstand (7.389) when we forget this fact. In particular, the sign in (7.389) will be wrong if we put the coordinate system on C backwards so that s decreases as t increases.

We conclude with a fundamental observation. As formulas (7.37) and (7.389) show, *the radius of curvature ρ and the absolute value $|K|$ of the curvature K* are reciprocals wherever we have defined both of them.

Problems 7.39

1 By use of the equations

$$x = a \cos t, \qquad y = a \sin t,$$

show that the curvature of the curve C consisting of a circle of radius a traced in the *positive* direction is identically $1/a$. Then by use of the equations

$$x = a \cos t, \qquad y = -a \sin t,$$

show that the curvature of the curve Γ (capital gamma) consisting of a circle of radius a traced in the *negative* direction is identically $-1/a$.

2 Hindsight can be very good. Look at (7.388). Then run with constant speed and in the positive direction around a circle of radius a and observe that ϕ increases at a constant rate. Then reverse the direction of the run and observe that ϕ decreases at the same constant rate.

3 Determine the radius of curvature of (that is, at points of) the parabola having the equation $y = kx^2$. Find the minimum radius of curvature. Sketch a graph for the case in which $k = 1$ and determine whether the answer seems to be correct.

4 Show that the normals to the graph of $y = x^2$ at the points (0,0) and (0.01, 0.0001) intersect at the point (0, 0.5001). Show that this intersection is at distance 0.0001 from the center of curvature of the graph at the point (0,0).

5 A glance at the graph of $y = \log x$ suggests that the absolute value of the curvature is greatest and that the radius of curvature is least when x is somewhere between $\frac{1}{2}$ and 2. Find the x for which the radius of curvature attains its minimum value. *Ans.:* $\sqrt{2}/2 = 0.707$.

6 When a point $P(x,y)$ moves along an arc or curve C having equations $x = x(t)$, $y = y(t)$ that satisfy appropriate conditions, the center (X,Y) of curvature moves along an arc that is called the *evolute* of C. The formulas (7.363) and (7.364), which we should be able to use but need not remember, show how X

and Y depend upon t. Supposing P moves along the graph of $y = kx^2$ in such a way that $x = t$ and $y = kt^2$, show that the equations of the evolute are

$$X = -4k^2t^3, \qquad Y = \frac{1}{2k} + 3kt^2.$$

7 This problem involves a little story. A string is wound in a clockwise direction around a circular spool of radius a with an end at the point $(a,0)$. When the string is unwound, being kept stretched during the process, the end of the string traces the spiral curve C shown in Figure 7.391. When $a\theta$ units

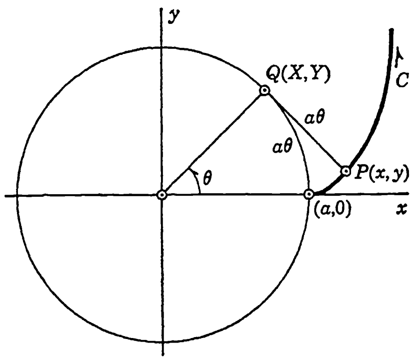


Figure 7.391

of string have been unwound, this part of the string is tangent to the spool at the point $Q(X,Y)$ for which

$$X = a\cos\theta, \qquad Y = a\sin\theta$$

and the end of the string is at the point $P(x,y)$ for which

$$\begin{aligned} x &= x(\theta) = a\cos\theta + a\theta\sin\theta \\ y &= y(\theta) = a\sin\theta - a\theta\cos\theta. \end{aligned}$$

It can be observed that Q moves around the circle just as rapidly as the distance from Q to P increases. It is not unreasonable to guess (or at least to consider the possibility) that the center and radius of curvature of C at P are Q and $a\theta$. On the other hand, a skeptic can be uncertain whether QP is perpendicular to the tangent to C at P. The situation demands clarification. Start with the equations of C and find the center and radius of curvature of C at P. *Ans.:* (X,Y) and $a\theta$. *Remark:* The curve C is called the *involute* of the spool. The spool is the evolute of its involute.

8 Determine the radius of curvature of the "standard ellipse"

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad \text{or} \qquad x = a\cos\theta, \qquad y = b\sin\theta$$

by use of the first "standard equation" and then by use of the latter parametric equations.

9 Using the parametric equations of Problem 8, find the evolute of the standard ellipse.

10 As an alternative to (or in addition to) Problem 8, find the radius of curvature of the "standard hyperbola."

11 A glance at the hypocycloid having the equations

$$x = a \cos^3 t, \qquad y = a \sin^3 t$$

indicates that the radius of curvature ρ should be a maximum at points (x,y) for which $|x| = |y|$ and a minimum at points for which $xy = 0$. Can we believe our eyes? *Ans.:* No, because ρ has no minimum. When $t \neq n\pi/2$, calculations give $\rho = \frac{3a}{2}|\sin 2t|$. Thus ρ is a maximum when $2t = \left(n + \frac{1}{2}\right)\pi$ or $t = \left(\frac{n}{2} + \frac{1}{4}\right)\pi$ and hence when $|\sin t| = |\cos t|$ or $|x| = |y|$. We see that $\rho \to 0$ as $t \to 0$, but the curvature at the cusps is undefined.

12 When a flexible cord or chain (the Latin word for chain is *catena*) is suspended from its ends in a parallel force field, it hangs in a curve (or point set) called a *catenary*. Differential equations textbooks show that a rectangular coordinate system can be chosen in such a way that the equation of the catenary is

$$y = \frac{a}{2}(e^{x/a} + e^{-x/a}).$$

Find the radius of curvature of this catenary at the point (x,y). *Ans.:* y^2/a.

13 Find the radius of curvature of the cycloid having the equation

$$x = a(\theta - \cos\theta), \qquad y = a(1 - \cos\theta).$$

14 Find the evolute of the cycloid of Problem 13.

15 Persons who picnic beside lakes several miles long can wonder whether poor visibility instead of curvature of the earth is responsible for invisibility of distant boats and shores. This problem involving curvature can be solved very simply. Figure 7.392 shows the center of a spherical earth at C on the

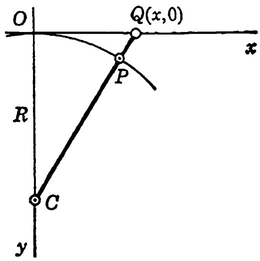


Figure 7.392

positive y axis, the x axis being tangent to the surface of the earth at O. The line from C to $Q(x,0)$ intersects the surface of the earth at P. The number $|\overrightarrow{PQ}|$ is the height h of an object which is visible from the point O on the earth x miles away. The simple calculation

$$(1) \qquad h = \sqrt{R^2 + x^2} - R = \frac{x^2}{\sqrt{R^2 + x^2} + R}$$

shows that the approximate formula

$$h = \frac{x^2}{2R} \tag{2}$$

gives good results when x is small in comparison to R. It is a quite remarkable fact that if h and R are measured in miles and if the height H of the object is measured in feet so that $H = 5280h$, then we can put $R = 3960$ in (2) and multiply by 5280 to obtain

$$H = \tfrac{2}{3}x^2.$$

Putting $H = 6$ shows that only the hair on the top of the head of a man 6 feet tall is visible from a point on the earth 3 miles away. Putting $x = 30$ gives $H = 600$ and shows that less than the top half of the Empire State Building can be seen by persons on a ship 30 miles away.

16 It is sometimes useful to have formulas obtainable from (7.361) and (7.362). Letting **N** [where the N can make us think of a normal to C at $P(t)$] be the vector running from $P(t)$ to the center (X,Y) of curvature, we see that

$$\mathbf{N} = \frac{[x'(t)]^2 + [y'(t)]^2}{x'(t)y''(t) - x''(t)y'(t)}[-y'(t)\mathbf{i} + x'(t)\mathbf{j}]. \tag{1}$$

Let **b** (where the b can make us think of binormal or "second normal") be the unit vector in the direction of $\mathbf{v}(t) \times \mathbf{a}(t)$ so that, as (7.32) shows, **b** is the coordinate vector **k** in our work. Then, with the aid of (7.32) and the fact that $\mathbf{b}\cdot\mathbf{b} = 1$, we can put (1) in the intrinsic form

$$\mathbf{N} = \frac{\mathbf{v}(t)\cdot\mathbf{v}(t)}{[\mathbf{v}(t) \times \mathbf{a}(t)]\cdot\mathbf{b}(t)}\mathbf{b}(t) \times \mathbf{v}(t) \tag{2}$$

in which coordinates do not appear. The formula (2) is valid when C is a curve in E_3 for which x,y,z are functions having continuous second derivatives and

$$\begin{aligned}
\mathbf{r}(t) &= x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}\\
\mathbf{v}(t) &= x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k}\\
\mathbf{a}(t) &= x''(t)\mathbf{i} + y''(t)\mathbf{j} + z''(t)\mathbf{k}
\end{aligned}$$

whenever t is such that $\mathbf{v}(t) \times \mathbf{a}(t) \neq 0$. We shall not prove this, but remark that the point (ξ,η,ζ) analogous to the point (ξ,η) of Figure 7.33 is the intersection of three planes and that (X,Y,Z) is the limit as $\Delta t \to 0$ of (ξ,η,ζ).

17 It is possible to obtain very informative formulas by considering the motion of a particle which moves along a plane curve C, endowed with coordinates as in Figure 7.27, in such a way that s increases as t increases. We assume existence and continuity of all the derivatives we want to use, and we assume that $dx/dt > 0$. Let **t** be the unit forward tangent vector to C at time t so that

$$\mathbf{t} = \cos\phi\mathbf{i} + \sin\phi\mathbf{j}, \tag{1}$$

where ϕ is, as in the discussion of Figure 7.384, an angle giving the direction of **t** at time t. Show that differentiating (1) gives

$$\frac{d\mathbf{t}}{dt} = [-\sin\phi\mathbf{i} + \cos\phi\mathbf{j}]\frac{d\phi}{dt} = \mathbf{n}\frac{d\phi}{ds}\frac{ds}{dt} = \frac{ds}{dt}K\mathbf{n}, \tag{2}$$

where $\mathbf{n}$ is the unit normal obtained by rotating $\mathbf{t}$ clockwise through the angle $\pi/2$, ds/dt is the speed of the particle, and K is the curvature $d\phi/ds$ of C. Show that writing the formula $\mathbf{v} = |\mathbf{v}|\mathbf{t}$ in the form

(3) $$\mathbf{v} = \frac{ds}{dt}\mathbf{t}$$

and differentiating give

(4) $$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2s}{dt^2}\mathbf{t} + \frac{ds}{dt}\frac{d\mathbf{t}}{dt}$$

and that use of (2) then gives

(5) $$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{t} + \left(\frac{ds}{dt}\right)^2 K\mathbf{n}.$$

Remark: This elegant formula gives the normal (or transverse) and tangential scalar components of the acceleration in terms of the speed and rate of change of the speed of the particle. The simplest applications involve the case in which the particle moves along the curve with constant speed σ so that $ds/dt = \sigma$ and $d^2s/dt^2 = 0$ at each time. In this case, (5) reduces to the simple but important formula

(6) $$\mathbf{a} = \sigma^2 K\mathbf{n}.$$

If $K = 0$, then $\mathbf{a} = 0$. If $K \neq 0$, then $\mathbf{a}$ has magnitude $\sigma^2|K|$ or σ^2/ρ (where ρ is the radius of curvature) and is directed toward the center of curvature. Use of (6) and the formula $\mathbf{F} = m\mathbf{a}$ gives the force required to propel a particle of mass m along a curve C with constant speed σ.

18 Let C be a simple closed convex curve which could, for example, be a famous "triangular roller" composed of the three vertices of an equilateral triangle together with three circular arcs of which each has its center at one vertex and contains the other two vertices. Let R be a rod (or line segment) whose length L is small enough to permit the rod to be "slid around" C in such a way that its two ends both remain on C. It is easy to presume, as has sometimes been done in "proofs" of the "Holditch theorem," that each point P on R must traverse a simple closed convex curve C_P as R slides around C. Persons having a compass, a straightedge like the edge of a sheet of paper upon which marks can be placed, and some spare time at their disposal can sketch interesting figures. Surprises await those who let C be a triangular roller, let the length L of the rod R be equal to or only a bit less than the distance between two vertices of the equilateral triangle, and let P be the mid-point of L.

8 Trigonometric functions

8.1 Trigonometric functions and their derivatives We are studying and perhaps even learning mathematics, and it may be amusing and perhaps even useful to see how an unscientific but highly critical Justice of a Supreme Court might be introduced to angles and trigonometry. We would teach him enough about numbers, E_2, continuous functions, and curves to make him realize that if we start with positive numbers a and h for which $0 < h < a$, then the ordered set of points $(x, \sqrt{a^2 - x^2})$ for which $h \leqq x \leqq a$, the point for which $x = x_2$ preceding the point for which $x = x_1$ when $a \leqq x_1 < x_2 \leqq h$, is a curve. Our curve lies on the circle with center at the origin and radius a, and we simplify (or complicate) matters by calling our curve "the arc, of the circle with center at the origin and radius a, which runs in the positive direction from the point $(a,0)$ to the point Q having coordinates $(h, \sqrt{a^2 - h^2})$" or, for temporary purposes, simply "the arc." To help us remember the meanings of our symbols, we can start constructing Figure 8.11 and amplify it as we proceed. Next we must teach our pupil a theory of length and prove to him that the arc has a length which we may denote by s. Knowing that our quantities depend upon h and a, we can define a number θ by the formula

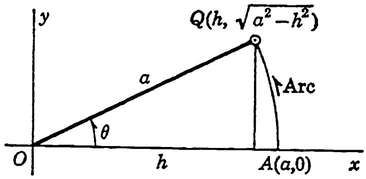


Figure 8.11

$$(8.12) \quad \theta = \frac{s}{a} = \frac{\text{length of arc}}{\text{radius}}.$$

If we multiply each of a and h by the same positive constant λ (lambda), we get a new radius and a new arc, but the ratio s/a remains the same. To prove this, we must prove that the length of the new arc is the product of λ and the length of the original arc. Our theory of length enables us to do this because, by drawing radial lines from the origin, we can see that each polygon inscribed in the new arc determines (and is determined) by a polygon inscribed in the original arc and that the lengths of the straight segments of the polygons all differ by the same factor λ. All this shows that we get the same θ when we take another pair of values of a and h for which the point $(h, \sqrt{a^2 - h^2})$ lies on the same half-line extending from O through Q. Now we can again simplify (or complicate) matters by considering the number θ to be "a measure of the amount of rotation required to bring a line from the position OA to the position OQ" or "a measure of the opening between the lines OA and OQ." Perhaps to remind us where the number θ came from, or perhaps to indicate something of which θ is to be considered a measure, we complete Figure 8.11 by inserting the θ together with the curved arrow which shows the direction of our arc. It is the fashion to call θ an angle, but it should be permanently remembered that θ, like 5, is a number. The facts that we sometimes use θ to measure an amount of rotation and use 5 to measure a number of fingers do not imply that θ *is* a rotation and that 5 *is* a fistful of fingers, but we can nevertheless understand and even use the more or less convenient terminologies involving "angles" that have become a part of nonscientific as well as scientific attempts to convey information. If all this indicates that trigonometry is a subject much too difficult for inclusion in trigonometry textbooks, we do have one consolation. The hard work is done and the rest is easy.

Our theory of curves is sufficiently general to allow us to extend the above account of positive (that is, counterclockwise) arcs and angles to cover situations in which the arc is longer and Q lies in the second or third or fourth quadrant. Moreover, the arc can be so long that we must encircle the origin more than once to traverse it, and the number (or angle) θ is still defined by the same formula (8.12). In case the arc starts at A and is oriented in the negative (or clockwise) direction, everything is the same except that the directions of the arrows are reversed, a negative sign is prefixed to the middle and last members of

(8.12), and the number θ is negative. Even though we could pretend to be appalled by the idea that numbers have sides, we bow to conventions and agree that the way to find the six fundamental trigonometric functions of a given angle θ is to construct "the terminal side of θ" as in Figure 8.13, pick a point $P(x,y)$ on this terminal side, let r be the distance (positive, of course) from O to P, and use the numbers x, y, r in the usual way.

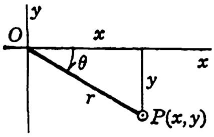


Figure 8.13

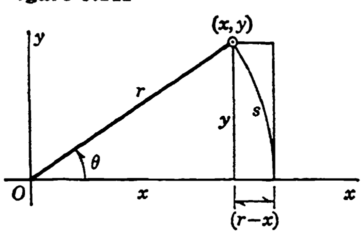
It is our purpose to present formulas for derivatives of trigonometric functions with derivations to which no logical objections can be raised. We could undertake to use the formula $A = \frac{1}{2}\theta r^2$ for the area of a circular sector of radius r which has central angle θ, but it would be immediately recognized that this formula has not been proved. Even if it be conceded that we know that the area of a whole circular disk of radius r is πr^2, it is only the docile acceptance of a crafty fraud that would allow us to "see at once" that the sector "obviously" has area $\frac{1}{2}\theta r^2$ because the area of the sector divided by the area of the circle is "obviously" $\theta/2\pi$. Nothing is obvious. Even a hazy understanding of the theory of area is enough to show that nothing can be proved without making substantial use of precise definitions and completeness of the real-number system or, what amounts to the same thing, consequences of these things. Since we do not now wish to be responsible for furnishing a complete treatment of areas of sectors, we shall base our work on the inequality (7.12). This inequality, the truth of which should seem thoroughly reasonable to a person at B who has the choice of walking two paths to D, has a virtue. It has been proved.

It was indicated in Problem 12 at the end of Section 3.2 that we will be able to derive the formulas for derivatives of trigonometric functions when we have proved the two formulas

$$\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = 1, \qquad \lim_{\theta\to 0}\frac{1-\cos\theta}{\theta} = 0. \tag{8.14}$$

Like a man who has drilled a hole and filled it with blasting powder to split a rock, we are ready to produce results. Figure 7.11 can be put into a coordinate system as in Figure 8.141, and we suppose that $0 < \theta < \pi/2$. In terms of the notation of this figure, the inequality (7.12) becomes

Figure 8.141



$$y \leqq s \leqq y + (r - x).$$

Dividing by r gives

$$\sin\theta \leqq \theta \leqq \sin\theta + (1-\cos\theta). \tag{8.142}$$

Since $1 - \cos\theta > 0$ and $1 + \cos\theta > 1$ and $\sin\theta < \theta$, the last term can be overestimated by the inequality

$$(8.143)\qquad \begin{aligned} 0 < (1 - \cos\theta) &< (1 - \cos\theta)(1 + \cos\theta) \\ &= 1 - \cos^2\theta = \sin^2\theta < \theta\sin\theta. \end{aligned}$$

Hence we can replace the last term in (8.142) by $\theta\sin\theta$ and divide by $\sin\theta$ to obtain the first and hence the second of the significant inequalities

$$(8.144)\qquad 1 \leqq \frac{\theta}{\sin\theta} < 1 + \theta, \qquad \frac{1}{1+\theta} < \frac{\sin\theta}{\theta} \leqq 1.$$

Dividing (8.143) by θ and using (8.142) gives

$$(8.145)\qquad 0 < \frac{1 - \cos\theta}{\theta} < \sin\theta < \theta.$$

The above inequalities have been proved to hold when $0 < \theta < \pi/2$. Since $|\theta| = \theta$ when $\theta > 0$, (8.144) and (8.145) imply that the inequalities

$$(8.146)\qquad \frac{1}{1+|\theta|} < \frac{\sin\theta}{\theta} \leqq 1, \qquad 0 < \frac{1-\cos\theta}{|\theta|} < |\theta|$$

hold when $0 < \theta < \pi/2$. Since the members of these inequalities are not changed when we replace θ by $-\theta$, we conclude that they hold when $|\theta| < \pi/2$ and $\theta \neq 0$. The desired formulas (8.14) follow from this and the sandwich theorem.

To derive the formulas for derivatives of sines and cosines, we use the formulas

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

to obtain

$$\begin{aligned} \sin(x + \Delta x) - \sin x &= -\sin x(1 - \cos\Delta x) + \cos x\sin\Delta x \\ \cos(x + \Delta x) - \cos x &= -\cos x\,(1 - \cos\Delta x) - \sin x\sin\Delta x. \end{aligned}$$

After dividing by Δx, taking limits as $\Delta x \to 0$ of the resulting difference quotients gives, with the aid of (8.14), the fundamental formulas

$$(8.15)\qquad \frac{d}{dx}\sin x = \cos x, \qquad \frac{d}{dx}\cos x = -\sin x.$$

These formulas, which we have used many times, have at long last been proved. For values of x for which the functions are defined, we obtain

$$(8.151)\quad \frac{d}{dx}\tan x = \frac{d}{dx}\frac{\sin x}{\cos x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \sec^2 x$$

$$(8.152)\quad \frac{d}{dx}\cot x = \frac{d}{dx}\frac{\cos x}{\sin x} = \frac{-\cos^2 x - \sin^2 x}{\sin^2 x} = -\csc^2 x$$

$$(8.153)\quad \frac{d}{dx}\sec x = \frac{d}{dx}\frac{1}{\cos x} = \frac{\sin x}{\cos^2 x} = \sec x \tan x$$

$$(8.154)\quad \frac{d}{dx}\csc x = \frac{d}{dx}\frac{1}{\sin x} = \frac{-\cos x}{\sin^2 x} = -\csc x \cot x.$$

The graphs of $\sin x$ and $\cos x$ are so important that we reproduce Figure 1.58 in Figure 8.16 and give further attention to the procedure by

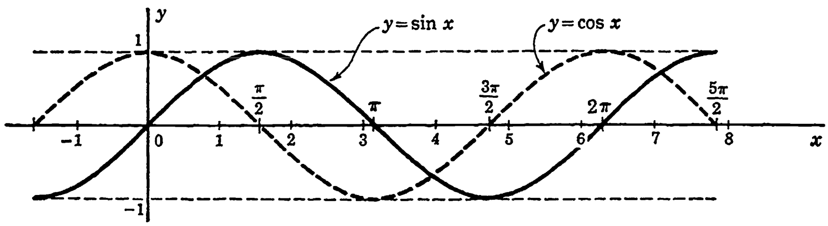


Figure 8.16

which reasonably accurate graphs are quickly sketched with or without use of graph paper. The trick is to sketch guide lines 1 unit above and 1 unit below the x axis, to hop 3 units and a bit more to the right of the origin to mark π, to make another such hop to mark 2π, and then sketch reasonably good copies of the figure. Each graph has slope 1 or -1 where it crosses an axis, and noticeable contradictions of this fact should not appear. In the problems at the end of the next section, we shall obtain formulas from which $\sin\theta$ and $\cos\theta$ can, for a given θ, be calculated as accurately as we wish. Meanwhile, we can be interested in Figure 8.17, which enables us to obtain reasonable estimates of $\sin\theta$ and $\cos\theta$ when $0 < \theta < \pi/2$. The circle has radius 1, and the radial lines make angles 0.1, 0.2, $\cdots$, 1.5 with the positive x axis. For example, the rough approximations

$$\sin 0.35 = \frac{0.34}{1} = 0.34, \qquad \cos 0.35 = \frac{0.94}{1} = 0.94,$$

$$\tan 0.35 = \frac{0.34}{0.94} = 0.36$$

can be read from the figure.

With the aid of information about derivatives, it is easy to see that the graph of the function t for which $t(x) = \tan x$ has the form shown in

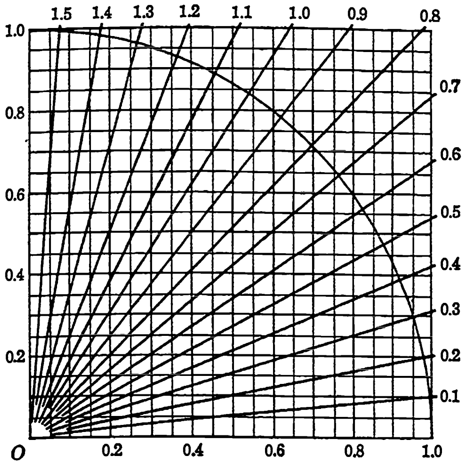


Figure 8.17

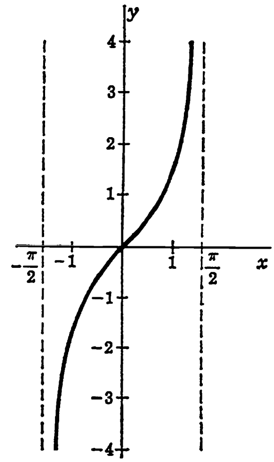


Figure 8.171

Figure 8.171. For each integer n, the interval $|x - n\pi| < \pi/2$ contains an exact copy of the graph in the interval $|x| < \pi/2$. The lines

$$x = n\pi \pm \frac{\pi}{2}$$

are all vertical asymptotes of the graph, and $\tan(n\pi \pm \pi/2)$ is undefined.

We should now be quite familiar with the fact that each formula for a derivative has a chain extension. The following list contains the chain extensions of formulas for derivatives of the six trigonometric functions and three additional formulas. All of these must be learned.

$$\frac{d}{dx}\sin u = \cos u\,\frac{du}{dx} \qquad \frac{d}{dx}\cot u = -\csc^2 u\,\frac{du}{dx}$$

$$\frac{d}{dx}\cos u = -\sin u\,\frac{du}{dx} \qquad \frac{d}{dx}\sec u = \sec u\tan u\,\frac{du}{dx}$$

$$\frac{d}{dx}\tan u = \sec^2 u\,\frac{du}{dx} \qquad \frac{d}{dx}\csc u = -\csc u\cot u\,\frac{du}{dx}$$

$$\frac{d}{dx}u^n = nu^{n-1}\frac{du}{dx}, \qquad \frac{d}{dx}e^u = e^u\frac{du}{dx}, \qquad \frac{d}{dx}\log u = \frac{1}{u}\frac{du}{dx}$$

Problems 8.19

1 With all graphs and tables out of sight, make the pretense that it has been forgotten whether the derivative with respect to x of $\sin x$ is $\sin x$ or $-\sin x$ or $\cos x$ or $-\cos x$. Sketch graphs of $\sin x$ and $\cos x$ and make these graphs give the correct answer.

2 Explain how you should modify the procedure for sketching the graph of

$y = \sin x$ to obtain a procedure for sketching the graph of $y = 3 \sin x$, and sketch the graph.

3 Equate the derivatives of the two members of the identity in the left column and show that obvious simplifications give the identity in the right column when

(*a*) $\sin 2x = 2 \sin x \cos x$	$\cos 2x = \cos^2 x - \sin^2 x$
(*b*) $\cos 2x = \cos^2 x - \sin^2 x$	$\sin 2x = 2 \sin x \cos x$
(*c*) $\sin^2 x + \cos^2 x = 1$	$0 = 0$
(*d*) $\sin (x + \phi)$ $= \sin x \cos \phi + \cos x \sin \phi$	$\cos (x + \phi)$ $= \cos x \cos \phi - \sin x \sin \phi$
(*e*) $\cos (x + \phi)$ $= \cos x \cos \phi - \sin x \sin \phi$	$\sin (x + \phi)$ $= \sin x \cos \phi + \cos x \sin \phi$
(*f*) $\cos^2 x = \dfrac{1 + \cos 2x}{2}$	$2 \sin x \cos x = \sin 2x$
(*g*) $\sin^2 x = \dfrac{1 - \cos 2x}{2}$	$2 \sin x \cos x = \sin 2x$
(*h*) $\sin \left(x + \dfrac{\pi}{2}\right) = \cos x$	$\cos \left(x + \dfrac{\pi}{2}\right) = -\sin x$
(*i*) $e^x e^x = e^{2x}$	$e^x e^x = e^{2x}$
(*j*) $\log x^2 = 2 \log x$	$\dfrac{2x}{x^2} = \dfrac{2}{x}$

4 Supposing that $0 \leqq x < \pi/2$, let $f(x) = \tan x$ and show that

$$\begin{aligned} f'(x) &= \sec^2 x \\ f''(x) &= 2 \sec^2 x \tan x \\ f'''(x) &= 2 \sec^4 x + 4 \sec^2 x \tan^2 x, \end{aligned}$$

and obtain the next derivative. Show that each derivative of higher order will also be a sum of terms of the form $A \sec^p x \tan^q x$, where A, p, q are nonnegative integers, and hence that $f^{(n)}(x) \geqq 0$ for each n and x.

5 Supposing that $0 < x < \pi/2$, calculate the first three derivatives of $\sec x$. Show that these and all derivatives of higher order are nonnegative.

6 Again supposing that $0 < x < \pi/2$, calculate some derivatives of $\sqrt{\tan x}$ and try to decide whether they are all nonnegative.

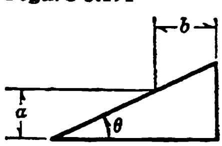
7 Prove the formula

$$\int \frac{\sin \theta}{1 + \sin \theta}\, d\theta = \sec \theta - \tan \theta + \theta + c$$

by differentiating the right side.

8 The length L of the longest beam that can be taken in a horizontal position around the corner of Figure 8.191 is the length L of the shortest line segment placed like the one in the figure. Find L. *Ans.:* $L = (a^{2/3} + b^{2/3})^{3/2}$. *Remark:* Putting the equation in the form $a^{2/3} + b^{2/3} = L^{2/3}$ can make us wonder why (a,b) should, for a fixed L, be a point on a hypocycloid.

Figure 8.191



9 Find the length L of the shortest ladder that can reach from level ground to a high wall when it must go over a fence which is a feet high and b feet from the wall. *Ans.:* $L = (a^{2/3} + b^{2/3})^{3/2}$.

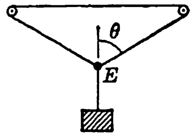


Figure 8.192

10 A heavy body is suspended from a rope, as in Figure 8.192, that runs up from the body to a pulley at E and thence (a wonderful word) over two stationary pulleys at A and B on the same horizontal level and back to E, where the pulley is tied to the other end of the rope. How should old man gravity select the θ of the figure in order to gratify his desire to bring the heavy body to its lowest possible position? *Ans.:* $\cos\theta = \frac{1}{2}$ and $\theta = \pi/3$ or $\theta = 60°$.

11 If $|P_n|$ is the length of a regular polygon of n sides which is inscribed in a circle of radius a, prove that $|P_n| = 2an \sin(\pi/n)$ and hence that $\lim_{n\to\infty} |P_n| = 2\pi a$.

12 Supposing that A and B are constants not both 0 and that

$$y(x) = A \sin x + B \cos x,$$

calculate $y'(x) = y''(x)$ and show that

$$y''(x) = -y(x).$$

In terms of the graph of $y(x)$, tell precisely what it means to say that $y''(x) < 0$ when $y(x) > 0$ and $y''(x) > 0$ when $y(x) < 0$. Sketch graphs and verify your conclusion when $A = 0$ and when $B = 0$. It is now required that we learn a little trick that happens to be very important. Plot the point having coordinates (A,B) and, as in Figure 8.193, let ϕ be an angle having its terminal side on the line running from the origin to the point. Let $C = \sqrt{A^2 + B^2}$ and observe that

$$A = C \cos\phi \qquad \text{and} \qquad B = C \sin\phi$$

so

$$y(x) = C(\sin x \cos\phi + \cos x \sin\phi)$$

and hence

$$y(x) = C \sin(x + \phi).$$

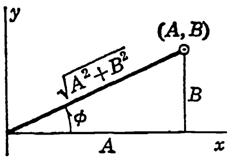


Figure 8.193

Finally, it is required that we learn some technical terminology by which this result can be remembered. Functions of the form $E \sin(\omega t + \phi)$ and $E \cos(\omega t + \phi)$ are called *sinusoids* (things like sines) of *angular frequency* ω. Since x can be ωt, we have proved that *the sum of two sinusoids having the same frequency* ω *is also a sinusoid having frequency* ω.

13 It is very easy to show that if k, A, B are constants for which $k > 0$ and if

$$y = A \sin kt + B \cos kt \tag{1}$$

then

$$\frac{d^2y}{dt^2} + k^2y = 0. \tag{2}$$

All we need to do is differentiate (1) twice and look at the result. The theory of differential equations contains theorems which imply that *if y is a function* *for which* (2) *holds, then there must be constants A and B for which* (1) *holds.* The latter and more difficult result is often needed by students who have not yet made reliable contacts with differential equations. Our solution of the more difficult problem will be postponed until we have solved the easy problem by a new method which involves only reversible steps. Starting with the formulas

$$(3)\qquad \begin{cases} y = A\sin kt + B\cos kt \\ \dfrac{dy}{dt} = Ak\cos kt - Bk\sin kt, \end{cases}$$

we eliminate B, and then eliminate A, from these equations to obtain

$$(4)\qquad \begin{cases} \cos kt\,\dfrac{dy}{dt} + (k\sin kt)y = Ak \\ \sin kt\,\dfrac{dy}{dt} - (k\cos kt)y = -Bk. \end{cases}$$

Therefore,

$$(5)\qquad \begin{cases} \dfrac{d}{dt}\left[\cos kt\,\dfrac{dy}{dt} + (k\sin kt)y\right] = 0 \\ \dfrac{d}{dt}\left[\sin kt\,\dfrac{dy}{dt} - (k\cos kt)y\right] = 0 \end{cases}$$

or

$$(6)\qquad \begin{cases} \cos kt\,\dfrac{d^2y}{dt^2} - k\sin kt\,\dfrac{dy}{dt} + k\sin kt\,\dfrac{dy}{dt} + (k^2\cos kt)y = 0 \\ \sin kt\,\dfrac{d^2y}{dt^2} + k\cos kt\,\dfrac{dy}{dt} - k\cos kt\,\dfrac{dy}{dt} + (k^2\sin kt)y = 0 \end{cases}$$

or

$$(7)\qquad \begin{cases} \cos kt\left[\dfrac{d^2y}{dt^2} + k^2y\right] = 0 \\ \sin kt\left[\dfrac{d^2y}{dt^2} + k^2y\right] = 0. \end{cases}$$

Since there is no t for which $\cos kt$ and $\sin kt$ are both 0, we conclude that

$$(8)\qquad \frac{d^2y}{dt^2} + k^2y = 0.$$

We are now ready to prove the more difficult result. Suppose that y is a given function for which (8) holds. Then we can multiply by $\cos kt$ and $\sin kt$ to obtain (7) and hence (6) and hence (5). Since the derivatives of the quantities in brackets are zero, these quantities must be constants which we can call Ak and $-Bk$ to obtain (4). Solving the equations (4) for y and dy/dt gives (3) and our problem is solved.

14 It is very easy to show that if k, A, B are constants for which $k > 0$ and if

$$(1)\qquad y = Ae^{kx} + Be^{-kx},$$

then

$$(2)\qquad \frac{d^2y}{dt^2} - k^2y = 0.$$

Do this. Then adapt the method of the preceding problem to show that if y is a function for which (2) holds, then there must be constants A and B for which (1) holds.

15 Prove that if y and u are functions of t having second derivatives and if

$$y = e^{-ht}u,$$

then

$$\frac{d^2y}{dt^2} + 2h\frac{dy}{dt} + k^2y = e^{-ht}\left[\frac{d^2u}{dt^2} + (k^2 - h^2)u\right].$$

16 A right circular cylinder (like a tomato can from which the top and bottom have been removed) has height h and base radius a. We examine the idea that good approximations to the area of this surface are obtained by triangulating it and calculating the sum of the areas of the triangles. The surface is partitioned into m strips each having height h/m like the one shown in Figure 8.194. On the top and bottom circular boundaries of each strip, we place the vertices of $2n$ isosceles triangles congruent to the triangle ABC of the figure. The side AB of the triangle subtends the angle $2\pi/n$ at the center O of the top circle. Of the $2n$ triangles in the top strip, n have two vertices on the top circle and the other n have two vertices on the lower circle. The total number of triangles congruent to the triangle ABC is $2mn$. The first factor in the product

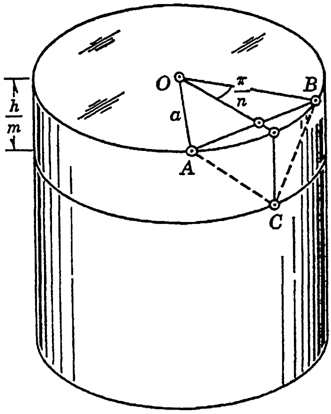


Figure 8.194

$$(1) \qquad \left(a \sin\frac{\pi}{n}\right)\sqrt{\left(\frac{h}{m}\right)^2 + \left(a - a\cos\frac{\pi}{n}\right)^2}$$

is half the length of the base AB of the triangle. The last factor is the altitude, being the length of the hypotenuse of a triangle one leg of which runs from C to the right angle on the upper circle and the other leg of which runs (in the direction of O) from the right angle to the base AB. The sum S_{mn} of the areas of all of the triangles is the product of (1) and $2mn$. Therefore,

$$(2) \qquad S_{mn} = 2\pi ah\frac{\sin\frac{\pi}{n}}{\frac{\pi}{n}}\sqrt{1 + \left(\frac{\pi^2 a}{2h}\right)^2\left(\frac{\sin\frac{\pi}{2n}}{\frac{\pi}{2n}}\right)^4\left(\frac{m}{n^2}\right)^2}.$$

When m and n are both large, the quotients having $\sin(\pi/n)$ in their numerators are near 1 and we have the approximation

$$(3) \qquad S_{mn} \sim 2\pi ah\sqrt{1 + \left(\frac{2\pi^2 a}{h}\right)^2\left(\frac{m}{n^2}\right)^2}.$$

In case $m = n$, and in other cases where

(4) $$\lim_{m,n\to\infty} \frac{m}{n^2} = 0,$$

we have

(5) $$\lim_{m,n\to\infty} S_{mn} = 2\pi ah,$$

the right member being the number usually considered to be the correct area of the cylinder. In case $m = n^2$, we obtain

(6) $$\lim_{m,n\to\infty} S_{mn} = 2\pi ah\sqrt{1 + \left(\frac{2\pi^2 a}{h}\right)^2},$$

the right member being a number which is not usually considered to be the correct area of the cylinder. Other remarks can be made. The above calculations were made by a German mathematician Schwarz, and they constitute the *Schwarz paradox*. The paradox shows that the triangulation idea provides a precarious basis for definitions of area of curved surfaces. The theory of these areas is extremely complicated.

17 It is sometimes useful as well as interesting to have information about the things we see. A *polynomial* P in x and y is the sum of a finite set of terms of the form cx^jy^k, where c is a constant and j and k are nonnegative integers. The polynomial has *degree* n if $j + k = n$ for at least one term in the sum having a nonzero coefficient and $j + k \leqq n$ for each term in the sum having a nonzero coefficient. A polynomial in which the coefficients are all zero is said to be *trivial;* it does not have a degree. Thus the polynomials having values

(1) $$x^4 + xy + y^4 - 34, \qquad (x^2 - y^2 - 1)(x^2 + y^2 - 4)$$

both have degree 4. A nontrivial polynomial is *irreducible* if it is not the product of two polynomials of lower degree. An *algebraic equation* of degree n is an equation of the form $P(x,y) = 0$, where P is a nontrivial polynomial of degree n in x and y. The graph of an algebraic equation is an *algebraic graph*. A function f of one variable x is said to be an *algebraic function* if there is a nontrivial polynomial $P(x,y)$ such that

(2) $$P(x, f(x)) = 0$$

for each x in the domain of f. Since each nontrivial polynomial in x and y can be put in the form

(3) $$Q_0(x) + Q_1(x)y + Q_2(x)y^2 + \cdots + Q_n(x)y^n,$$

where $Q_0, Q_1, \cdots, Q_n$ are polynomials in x at least one of which is nontrivial, it follows that a function f is an algebraic function if and only if there exist polynomials $Q_0, Q_1, \cdots, Q_n$ in x such that Q_n is nontrivial and

(4) $$Q_0(x) + Q_1(x)f(x) + Q_2(x)[f(x)]^2 + \cdots + Q_n(x)[f(x)]^n = 0$$

for each x in the domain of f. Functions that are not algebraic functions are said to be *transcendental functions*, the old idea being that polynomials lie at the foundation of human experience and that things not closely related to polynomials are more ethereal. This matter can be of interest to us now because, even in quite elementary mathematics, the trigonometric functions are called transcendental functions. For example, the assertion that $\sin x$ is transcendental means that there do not exist polynomials $Q_0, Q_1, \cdots, Q_n$ in x such that Q_n is nontrivial and (4) is true for each x when $f(x) = \sin x$. Let us now think briefly about real numbers. A number x is said to be an *algebraic number* if there exist integers $k_0, k_1, k_2, \cdots, k_n$ not all zero such that

$$k_0 + k_1x + k_2x^2 + \cdots + k_nx^n = 0.$$

Thus a number x is algebraic if it is a zero of a nontrivial polynomial in x having integer coefficients. A number x which is not an algebraic number is said to be a *transcendental number*. The numbers π and e are transcendental. Nobody can prove these things unless he devotes very much time and energy to the operation, but nobody requires us to be so busy digging ditches that we never look at the stars. We can know that there exist theories of algebraic functions and algebraic numbers and that these theories invite the attention of persons who have completed studies of analytic geometry and calculus.

8.2 *Trigonometric integrands* Before introducing chain extensions and other modifications of the formulas, we systematically work out formulas for integrals of the six trigonometric functions. Even though the last three or four have minor importance, we must learn about them to be respectable. The first two are

$$\int \sin x\,dx = -\cos x + c, \qquad \int \cos x\,dx = \sin x + c. \tag{8.21}$$

They are immediate consequences of the formulas for derivatives of sines and cosines, since the formula

$$\int f(x)\,dx = F(x) + c$$

is valid over an interval if and only if $F'(x) = f(x)$ over the interval. It is necessary to keep negative signs in their places when we differentiate and integrate sines and cosines, and it happens that a foolish little trick enables us to permanently remember how the signs go. We can mentally write

$$\left\{\begin{array}{ll} \text{derivative} & \text{integral} \\ \text{sine} & \text{cosine} \end{array}\right. \tag{8.22}$$

in their natural orders and remember that "like things give plus," so differentiating sines and integrating cosines give plus signs, but that "unlike things give minus," so differentiating cosines and integrating sines give minus signs.

If an interval contains one of the points $n\pi \pm \pi/2$, there can be no F such that $\int \tan x\,dx = F(x) + c$ over that interval. When x is confined to an interval containing none of these points, we can use the fundamental formula

$$\int [u(x)]^{-1}u'(x)\,dx = \log|u(x)| + c$$

to obtain

$$(8.23)\quad \int \tan x\,dx = -\int \frac{1}{\cos x}(-\sin x)\,dx = -\log|\cos x| + c$$
$$= \log|\sec x| + c.$$

In most applications of this formula, $|x| < \pi/2$, so $\cos x > 0$ and the absolute-value signs can be omitted. For $\cot x$, a miniature graph of which appears in Figure 8.232, we suppose that x is confined to an interval containing none of the points $x = n\pi$ and obtain

$$(8.231)\qquad \int \cot x\,dx = \int \frac{1}{\sin x}\cos x\,dx = \log|\sin x| + c.$$

In most applications of this formula, $0 < x < \pi$, so $\sin x > 0$ and the absolute-value signs can be omitted.

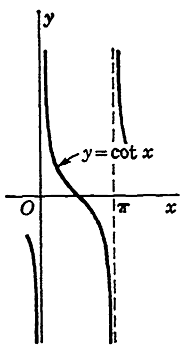


Figure 8.232

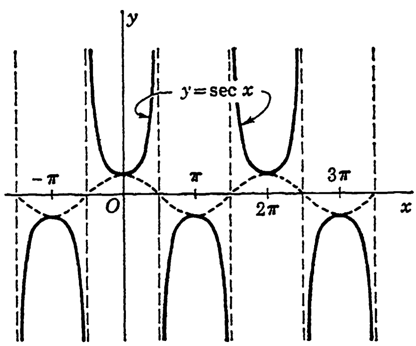


Figure 8.233

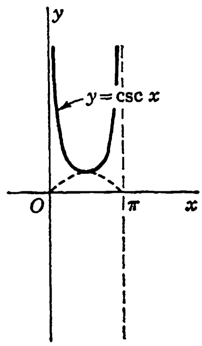


Figure 8.234

A sketch of the graph of $y = \sec x$ is most easily obtained by sketching a graph of $y = \cos x$ and estimating reciprocals. As the graph in Figure 8.233 indicates, $\sec x$ is undefined when x is an odd integer multiple of $\pi/2$. When we integrate $\sec x$ over an interval, we must suppose that the interval contains none of these points. We can then obtain the formula

$$(8.235)\quad \int \frac{\sec x}{1}\,dx = \int \frac{1}{\sec x + \tan x}(\sec^2 x + \sec x \tan x)\,dx$$
$$= \log|\sec x + \tan x| + c$$

provided we happen to know that the result is obtained by multiplying the numerator and denominator of (sec x)/1 by the implausible factor sec x + tan x. We dispose of csc x as rapidly as possible by writing

$$(8.236)\quad \int \frac{\csc x}{1}\,dx = -\int \frac{1}{\csc x + \cot x}(-\csc^2 x - \csc x \cot x)\,dx$$
$$= -\log|\csc x + \cot x| + c.$$

When $0 < x < \pi$ as in Figure 8.234, we can omit the absolute-value signs. The chain extensions of the six formulas are placed in a table at the end of this section where they are most available for reference. The formulas

$$(8.24)\quad \int \sin ax\,dx = \frac{1}{a}\int \sin ax\; a\,dx = -\frac{1}{a}\cos ax + c$$

$$(8.241)\quad \int \cos ax\,dx = \frac{1}{a}\int \cos ax\; a\,dx = \frac{1}{a}\sin ax + c,$$

which hold when $a \neq 0$, are by far the most important applications of the chain extensions.

We now consider some of the more or less important integrals that can be evaluated in terms of elementary functions. Adding and subtracting the two elementary formulas

$$(8.25)\quad \cos^2\theta + \sin^2\theta = 1$$
$$(8.251)\quad \cos^2\theta - \sin^2\theta = \cos 2\theta$$

gives the two formulas

$$(8.252)\quad \cos^2\theta = \frac{1+\cos 2\theta}{2}, \qquad \sin^2\theta = \frac{1-\cos 2\theta}{2}$$

that are so useful that we should either learn them or be able to work them out very quickly. Moreover, these formulas should come to mind when we are asked to evaluate the first integrals in the formulas

$$(8.253)\quad \int \cos^2 ax\,dx = \int \frac{1+\cos 2ax}{2}\,dx$$
$$= \frac{1}{2}\left[x + \frac{1}{2a}\int \cos 2ax(2a)\,dx\right]$$
$$= \frac{1}{2}\left[x + \frac{1}{2a}\sin 2ax\right] + c$$

and

$$(8.254)\quad \int \sin^2 ax\,dx = \int \frac{1-\cos 2ax}{2}\,dx$$
$$= \frac{1}{2}\left[x - \frac{1}{2a}\int \cos 2ax(2a)\,dx\right]$$
$$= \frac{1}{2}\left[x - \frac{1}{2a}\sin 2ax\right] + c.$$

Students are frequently called upon to realize that the formula

$$\cos^4 x = (\cos^2 x)^2 = \left(\frac{1 + \cos 2x}{2}\right)^2 = \frac{1 + 2\cos 2x + \cos^2 2x}{4}$$

is useful when the left member must be integrated, and to know what to do to complete the problem.

The integral in

$$I = \int \sin mx \cos nx \, dx \tag{8.26}$$

looks forbidding until it is discovered or remembered that a useful formula for the integrand can be obtained by adding the formulas

$$\begin{aligned} \sin(\theta + \phi) &= \sin\theta\cos\phi + \cos\theta\sin\phi \\ \sin(\theta - \phi) &= \sin\theta\cos\phi - \cos\theta\sin\phi. \end{aligned}$$

Thus

$$\tfrac{1}{2}[\sin(mx + nx) + \sin(mx - nx)] = \sin mx \cos nx$$

and

$$\begin{aligned} I &= \frac{1}{2(m+n)} \int [\sin(m+n)x](m+n)\, dx \\ &\qquad + \frac{1}{2(m-n)} \int [\sin(m-n)x](m-n)\, dx \\ &= \frac{1}{2(m+n)} \cos(m+n)x - \frac{1}{2(m-n)} \cos(m-n)x + c \end{aligned} \tag{8.261}$$

except when $m = n$ or $m = -n$. Similarly, the integrals

$$\int \sin mx \sin nx\, dx, \qquad \int \cos mx \cos nx\, dx \tag{8.262}$$

can be evaluated by use of formulas obtained by adding and subtracting the formulas

$$\begin{aligned} \cos(\theta - \phi) &= \cos\theta\cos\phi + \sin\theta\sin\phi \\ \cos(\theta + \phi) &= \cos\theta\cos\phi - \sin\theta\sin\phi. \end{aligned}$$

Sometimes integrals can be evaluated by making quite direct use of the power formula and other integration formulas. For example,

$$\begin{aligned} &\int \sin^p ax \cancel{\cos ax\, dx} \\ &\qquad = \frac{1}{a}\int (\sin ax)^p (a \cos ax)\, dx = \frac{1}{a}\frac{\sin^{p+1} ax}{p+1} + c \end{aligned} \tag{8.263}$$

when $a \neq 0$ and $p \neq -1$. Sometimes we need some ingenuity. For example, the integral

$$\int \sin^2 x \cos^3 x\, dx \tag{8.264}$$

is rendered manageable by replacing the factor $\cos^3 x$ by the last member of the formula

$$\cos^3 x = \cos^2 x \cos x = (1 - \sin^2 x) \cos x.$$

The integral in

$$(8.265) \qquad \int e^{\cos x} \sin x \, dx = -\int e^{\cos x}(-\sin x) \, dx = -e^{\cos x} + c$$

is evaluated by making the adjustment and compensation necessary to put it in the form $\int e^{u(x)}u'(x) \, dx$, and opportunities to make such adjustments and compensations should always be observed.

There are reasons why some of the integrals we have evaluated may be said to be so important that everyone should know how to evaluate them. There is an old and perhaps honorable tradition that requires students of the calculus to spend huge amounts of time cultivating "the technique" of "formal integration." The fact that we live in an age of electronic computers makes it much more important to learn fundamental theory than to acquire skill in formal integration. For this reason, the author requests that teachers join him in avoiding all but the simplest formal integration problems that are not likely to be encountered by undergraduates in courses other than calculus courses. "The student" who has not read dozens of calculus books and does not know what we are talking about is invited to look at the shiny example

$$(8.27) \qquad I = \int \sqrt{\sin x \tan x (1 + \sqrt{\cos x})} \, dx$$

of an integral that we shall not expect him to evaluate quickly. A person who has constructed this problem can easily feel very sure that the only sensible attack upon the problem lies in setting

$$(8.271) \qquad u(x) = 1 + (\cos x)^{1/2};$$

so, when $0 < x < \pi/2$,

$$(8.272) \qquad \begin{aligned} u'(x) &= \tfrac{1}{2}(\cos x)^{-1/2}(-\sin x) \\ &= -\tfrac{1}{2}\sqrt{\frac{\sin^2 x}{\cos x}} = -\tfrac{1}{2}\sqrt{\sin x \tan x} \end{aligned}$$

and hence

$$(8.273) \qquad \begin{aligned} I &= -2 \int [u(x)]^{1/2}[u'(x)] \, dx \\ &= -2 \cdot \frac{[u(x)]^{3/2}}{\frac{3}{2}} + c = -\tfrac{4}{3}[1 + \sqrt{\cos x}]^{3/2} + c. \end{aligned}$$

Instead of inviting attention to problems of this nature, we present problems more likely to promote scientific competence.

Table 8.28

$$\int \sin u \frac{du}{dx}\,dx = -\cos u + c \qquad \int \cos u \frac{du}{dx}\,dx = \sin u + c$$

$$\int \tan u \frac{du}{dx}\,dx = \log|\sec u| + c \qquad \int \cot u \frac{du}{dx}\,dx = \log|\sin u| + c$$

$$\int \sec u \frac{du}{dx}\,dx = \log|\sec u + \tan u| + c \qquad \int \csc u \frac{du}{dx}\,dx = \log|\csc u - \cot u| + c$$

$$\int u^n \frac{du}{dx}\,dx = \frac{u^{n+1}}{n+1} + c, \int \frac{1}{u}\frac{du}{dx}\,dx = |\log u| + c, \int e^u \frac{du}{dx}\,dx = e^u + c.$$

Problems 8.29

1 Make all of the calculations necessary to show that

(*a*) $\int_0^{\pi} \sin x\,dx = 2$ (*b*) $\int_0^{\pi/2} \cos x\,dx = 1$

(*c*) $\lim_{\omega\to\infty} \int_a^b \sin \omega x\,dx = 0$ (*d*) $\lim_{\omega\to\infty} \int_a^b \cos \omega x\,dx = 0$

(*e*) $\lim_{x\to\infty} \frac{1}{x}\int_0^x \sin^2 \omega t\,dt = \frac{1}{2}$ (*f*) $\lim_{x\to\infty} \frac{1}{x}\int_0^x \cos^2 \omega t\,dt = \frac{1}{2}$

(*g*) $\int_0^{\pi/4} \tan x\,dx = \frac{1}{2}\log 2$ (*h*) $\int_0^{\pi/2} \tan x\,dx = \infty$

2 Recall that, when u is a differentiable function of x and $u \neq 0$,

$$\frac{d}{dx}\log|u| = \frac{1}{u}\frac{du}{dx}; \tag{1}$$

the absolute-value signs, which are superfluous when $u > 0$, need not bother us. Supposing that x is not an odd multiple of $\pi/2$ and that

$$f(x) = \log|\sec x + \tan x|, \qquad g(x) = \log\left|\tan\left(\frac{x}{2} + \frac{\pi}{4}\right)\right|, \tag{2}$$

show that $f'(x) = \sec x$ and $g'(x) = \sec x$.

Remark: This proves that the two formulas

$$\int \sec x\,dx = \log|\sec x + \tan x| + c_1 \tag{3}$$

$$\int \sec x\,dx = \log\left|\tan\left(\frac{x}{2} + \frac{\pi}{4}\right)\right| + c_2 \tag{4}$$

are both correct. Some integral tables contain both of them. These things imply that, over each interval containing no odd multiple of $\pi/2$, $f(x) - g(x)$

must be a constant, but they do not imply that this constant must be zero. The identity

$$\tan\left(\frac{x}{2}+\frac{\pi}{4}\right) = \frac{1+\tan\frac{x}{2}}{1-\tan\frac{x}{2}} = \frac{\cos\frac{x}{2}+\sin\frac{x}{2}}{\cos\frac{x}{2}-\sin\frac{x}{2}}\,\frac{\cos\frac{x}{2}+\sin\frac{x}{2}}{\cos\frac{x}{2}+\sin\frac{x}{2}}$$

$$= \frac{1+2\sin\frac{x}{2}\cos\frac{x}{2}}{\cos^2\frac{x}{2}-\sin^2\frac{x}{2}} = \frac{1+\sin x}{\cos x} = \sec x + \tan x$$

clarifies the matter.

3 Prove that if x is not an even multiple of $\pi/2$ and

$$f(x) = \log|\csc x - \cot x|, \qquad g(x) = \log\left|\tan\frac{x}{2}\right|,$$

then $f'(x) = \csc x$ and $g'(x) = \csc x$. *Remark:* This proves that the two formulas

$$\int \csc x\,dx = \log|\csc x - \cot x| + c, \qquad \int \csc x\,dx = \log\left|\tan\frac{x}{2}\right| + c$$

are both correct.

4 When a steady (or constant) current I (measured in amperes) flows through a wire (or resistor) having resistance R (measured in ohms), the quantity Q (measured in watthours, or thousandths of kilowatthours) of energy converted into heat in Δt hours is calculated from the formula

(1) $$Q = I^2R\,\Delta t.$$

With this basic information, show that if $I(t)$ is integrable (Riemann) over the time interval $t_1 \leqq t \leqq t_2$, then the quantity $Q(t_1,t_2)$ of energy converged into heat between times t_1 and t_2 is (or should be)

(2) $$Q(t_1,t_2) = R\int_{t_1}^{t_2}[I(t)]^2\,dt.$$

Suppose now that $I(t)$ is the sinusoidal (or alternating) current determined by

(3) $$I(t) = I_0 \sin(\omega t + \phi),$$

where I_0 is a constant "maximum current," ω is a constant "angular frequency," and ϕ is a "phase angle." Show that, in this case,

(4) $$Q(t_1,t_2) = \frac{t_2-t_1}{2}I_0^2R - \frac{I_0^2R}{4\omega}[\sin(2\omega t_2 + 2\phi) - \sin(2\omega t_1 + 2\phi)].$$

The last term is 0 whenever $2\omega t_2 - 2\omega t_1$ is an integer multiple of 2π, and in every case the absolute value of the last term cannot exceed $RI_0^2/2\omega$. In all ordinary applications of this formula, $t_2 - t_1$ is so large in comparison to $1/\omega$ that the last term in (4) is insignificant. In such cases, (4) is always replaced by

(5) $$Q(t_1,t_2) = \tfrac{1}{2}I_0^2R(t_2 - t_1).$$

Putting (5) in the form

$$Q(t_1,t_2) = \left(\frac{I_0}{\sqrt{2}}\right)^2 R(t_2 - t_1) \tag{6}$$

shows that a sinusoidal current having maximum value I_0 produces heat (or dissipates energy) just as rapidly as a steady current of magnitude $I_0/\sqrt{2}$. For this reason, the number $I_0/\sqrt{2}$ is called the *effective value* of the sinusoidal current. For ordinary house current having "effective voltage" 120 volts, the maximum (or peak) voltage is not 120 volts but is 120 $\sqrt{2}$ volts.

5 Let, when L is a positive number and $n = 1, 2, 3, \cdots$,

$$\phi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L}, \qquad \psi_0(x) = \frac{1}{\sqrt{L}}, \qquad \psi_n(x) = \sqrt{\frac{2}{L}} \cos \frac{n\pi x}{L}.$$

Show that the first of the formulas

$$\int_0^L \phi_m(x)\phi_n(x)\,dx = 0, \qquad \int_0^L [\phi_n(x)]^2\,dx = 1$$

holds when m and n are different positive integers and that the second holds when $n = 1, 2, 3, \cdots$. Show that the first of the formulas

$$\int_0^L \psi_m(x)\psi_n(x)\,dx = 0, \qquad \int_0^L [\psi_n(x)]^2\,dx = 1$$

holds when m and n are different nonnegative integers and that the second holds when $n = 0, 1, 2, \cdots$. Show that

$$\int_0^L \psi_0(x)\phi_n(x)\,dx = \sqrt{2}\,\frac{1 - \cos n\pi}{n\pi} \qquad (n = 1, 2, 3, \cdots).$$

Show that, when $m, n = 1, 2, 3, \cdots$,

$$\int_0^L \psi_m(x)\phi_n(x)\,dx = \frac{1 - \cos(m+n)\pi}{(m+n)\pi} + \frac{1 - \cos(m-n)\pi}{(m-n)\pi},$$

where the last term is to be omitted when $m = n$. *Remark:* While students in calculus courses have not yet heard about the matter, the above formulas are of great interest in the theory of orthonormal sets and Fourier series of the trigonometric variety. Our calculations and a little theory produce the interesting formulas

$$\frac{\pi}{4} = \sin\frac{\pi x}{L} + \frac{1}{3}\sin\frac{3\pi x}{L} + \frac{1}{5}\sin\frac{5\pi x}{L} + \cdots \qquad (0 < x < L)$$

$$\frac{\pi^2}{8} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots, \qquad \frac{\pi^2}{6} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots$$

and many others.

6 A problem in Section 8.4 will tell us about the formula

$$\int_0^{\pi/2} \sin^p\theta \cos^q\theta\,d\theta = \frac{\frac{p-1}{2}!\,\frac{q-1}{2}!}{2\,\frac{p+q}{2}!}$$

which is correct whenever $p > -1$ and $q > -1$. As usual, $0! = 1$, $1! = 1$, $2! = 1\cdot 2$, $3! = 1\cdot 2\cdot 3$, $4! = 1\cdot 2\cdot 3\cdot 4$, etcetera. The values of $x!$ when x is not an integer are more esoteric, but $x! > 0$ when $x > -1$. Put $p = q = 0$ and discover that $(-\frac{1}{2})! = \sqrt{\pi}$. Show that the formula is correct when $p = q = 1$. Put $p = q = 2$ and discover that $(\frac{1}{2})! = \sqrt{\pi}/2$. Put $p = 2x - 1$ and $q = 1$ and use the result to prove that $x!$ is the product of x and $(x - 1)!$.

7 With an assist from (8.144), which shows that

$$\int_0^1 \frac{1}{(1+\theta)^2}\,d\theta \leqq \int_0^1 \left(\frac{\sin\theta}{\theta}\right)^2 d\theta \leqq \int_0^1 1\,d\theta,$$

the middle integral being a Riemann-Cauchy integral because the integrand is undefined when $\theta = 0$, show that

$$\tfrac{1}{2} \leqq \int_0^1 \left(\frac{\sin\theta}{\theta}\right)^2 d\theta \leqq 1.$$

Then prove the first of the inequalities

$$0 \leqq \int_1^\infty \left(\frac{\sin\theta}{\theta}\right)^2 d\theta \leqq 1, \qquad \tfrac{1}{2} \leqq \int_0^\infty \left(\frac{\sin\theta}{\theta}\right)^2 d\theta \leqq 2$$

and use it to obtain the second.

8 Let a and b be constants for which $a = 0$ and $b = 1$ or $a = 1$ and $b = 0$ and let $x > 0$. Show that

$$-1 \leqq a\cos x + b\sin x \leqq 1.$$

Replace x by t and integrate over the interval $0 \leqq t \leqq x$ to obtain

$$-x \leqq a\sin x - b(\cos x - 1) \leqq x.$$

Replace x by t and integrate over the interval $0 \leqq t \leqq x$ to obtain

$$-\frac{x^2}{2} \leqq -a(\cos x - 1) - b(\sin x - x) \leqq \frac{x^2}{2}.$$

Repeat the process to obtain

$$-\frac{x^3}{3!} \leqq -a(\sin x - x) + b\left(\cos x - 1 + \frac{x^2}{2}\right) \leqq \frac{x^3}{3!}$$

$$-\frac{x^4}{4!} \leqq a\left(\cos x - 1 + \frac{x^2}{2!}\right) + b\left(\sin x - x + \frac{x^3}{3!}\right) \leqq \frac{x^4}{4!}$$

$$-\frac{x^5}{5!} \leqq a\left(\sin x - x + \frac{x^3}{3!}\right) - b\left(\cos x - 1 + \frac{x^2}{2!} - \frac{x^4}{4!}\right) \leqq \frac{x^5}{5!}.$$

Repeat the process two more times. With or without more attention to details, jump to the conclusion that

$$\left|\sin x - \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \pm \frac{x^{2n+1}}{(2n+1)!}\right)\right| \leqq \frac{|x|^{2n+3}}{(2n+3)!}$$

$$\left|\cos x - \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \pm \frac{x^{2n}}{(2n)!}\right)\right| \leqq \frac{|x|^{2n+2}}{(2n+2)!}$$

for each $n = 1, 2, 3, \cdots$. It is true (and is easy to surmise) that, for each x, $|x|^n/n! \to 0$ as $n \to \infty$. Therefore,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$
$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots.$$

9 The formulas at the end of the preceding problem have been proved to be correct when $x > 0$. Use this fact to prove that the formulas are also correct when $x \leqq 0$. *Hint:* Look first at the case in which $x = 0$. Then use the facts that $\sin(-x) = -\sin x$ and $\cos(-x) = \cos x$.

10 Have a good look at the formulas at the end of Problem 8 and start learning them by obtaining some of the following results correct to four or more decimal places.

sin 0.01 = 0.00999 9833	cos 0.01 = 0.99995 0000
sin 0.02 = 0.01999 8667	cos 0.02 = 0.99980 0007
sin 0.10 = 0.09983 3417	cos 0.10 = 0.99500 4165
sin 0.20 = 0.19866 9331	cos 0.20 = 0.98006 6587
sin 1.00 = 0.84147 0985	cos 1.00 = 0.54030 2306
sin 1.10 = 0.89120 7360	cos 1.10 = 0.45359 6121

11 Digest the following idea. We have a desk calculator and National Bureau of Standards Tables giving the values

$$\sin 0.2345 = 0.23235\,6699$$
$$\cos 0.2345 = 0.97263\,0641$$

If we want to find sin 0.23456 789 correct to eight decimal places, we can use the identity

$$\sin(x + 0.2345) = \sin x \cos 0.2345 + \cos x \sin 0.2345,$$

where $x = 0.00006789$. The values of $\sin x$ and $\cos x$ can easily be found correct to 10 decimal places from the formulas at the end of Problem 8.

8.3 *Inverse trigonometric functions*

Before coming to the announced subject of this section, we think about a general situation applications of which appear in many branches of mathematics. Suppose we have an operator or transformer or mapper or function f that carries or transforms or maps each element x of a set D (the domain of f) into an element y of a set R (the range of f). In some cases the transformer transforms two or more different elements of D into the same element of R. For example, if f is one of the six trigonometric functions and $f(x)$ exists, then $f(x + 2\pi) = f(x)$ and hence more than one element of the domain is carried into the same element $f(x)$ of the range. When problems involving domains and ranges are involved, it is very often possible to eliminate confusion by singling out for special attention a subset D_1 of D such that to each y in R there corresponds exactly one x

in D_1 for which $f(x) = y$. Then we can introduce a function f_1, which is different from f because it has a different domain, for which $f_1(x) = f(x)$ when x is in D_1 and for which $f_1(x)$ is undefined when x is not in D_1. This function f_1 is called *the restriction of f to D_1*. We now have a general situation that may be easier to visualize than its applications to trigonometric functions. The schematic Figure 8.31 shows the domain and

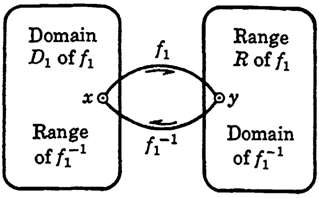


Figure 8.31

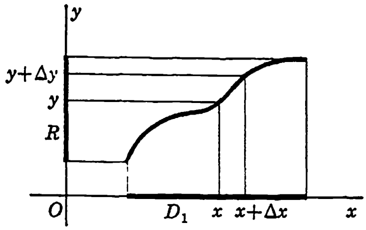


Figure 8.32

range of f_1, and the upper arc from x to y can make us think of a path along which f_1 might carry x to the y which is $f_1(x)$. The function f_1 sets up a one-to-one (or *schlicht*) correspondence between D_1 and R, that is, to each x in D_1 corresponds exactly one y in R for which $f(x) = y$ and to each y in R corresponds exactly one x in D_1 for which $f(x) = y$. The function with domain R and range D_1 which carries each y in R into the x in D_1 for which $f_1(x) = y$ is called the *inverse* of f_1 and is denoted by f_1^{-1}. As indicated by the figure, f_1^{-1} undoes what f_1 does. If $f_1(x) = y$, then $x = f_1^{-1}(y)$. Moreover, $f_1^{-1}(f_1(x)) = x$ when x is in D and

$$f_1(f_1^{-1}(y)) = y$$

when y is in R.

Figure 8.32 shows the graph of a particular function f to which the following theorem applies.

Theorem 8.33 *If f is continuous over an interval D_1 in E_1 and if f has a derivative for which $f'(x) > 0$ at each inner point x of D_1 [or $f'(x) < 0$ at each inner point x of D_1] then the restriction f_1 of f to D_1 has a range R which is an interval and has also an inverse f_1^{-1} which is differentiable at each inner point y of R. Moreover, the first of the formulas*

$$f_1^{-1\prime}(y) = \frac{1}{f_1'(f_1^{-1}(y))}, \qquad \frac{dx}{dy} = \frac{1}{\dfrac{dy}{dx}} \tag{8.331}$$

is valid when y is an inner point of R.

It seems to be possible to put the first formula of (8.331) in the second form without sacrifice of meaning. To prove the theorem, we use the hypotheses to conclude that f_1 is increasing (or decreasing) over D_1 and that (because of the intermediate-value theorem) R is an interval. Hence

f_1 has an inverse defined over R. Since we must *prove* that f_1^{-1} is differentiable, we introduce difference quotients for f_1^{-1}. Supposing that y and $y + \Delta y$ belong to the domain R of f_1^{-1}, we define x and $x + \Delta x$ by the formulas $x = f_1^{-1}(y)$ and $x + \Delta x = f_1^{-1}(y + \Delta y)$ and obtain $y = f_1(x)$, $y + \Delta y = f_1(x + \Delta x)$. Since f_1 has a positive (or negative) derivative, we conclude that $\Delta x \neq 0$ when $\Delta y \neq 0$ and that $\Delta x \to 0$ as $\Delta y \to 0$. Therefore,

$$(8.332)\quad \lim_{\Delta y \to 0} \frac{f_1^{-1}(y + \Delta y) - f_1^{-1}(y)}{\Delta y} = \lim_{\Delta y \to 0} \frac{\Delta x}{f_1(x + \Delta x) - f_1(x)}$$

$$= \lim_{\Delta y \to 0} \frac{1}{\dfrac{f_1(x_1 + \Delta x) - f_1(x)}{\Delta x}} = \frac{1}{f_1'(x)} = \frac{1}{f_1'(f_1^{-1}(y))}.$$

This proves Theorem 8.33. Of course it is possible to abbreviate the calculations by writing

$$(8.333)\qquad \frac{dx}{dy} = \lim_{\Delta y \to 0} \frac{\Delta x}{\Delta y} = \frac{1}{\lim\limits_{\Delta x \to 0} \dfrac{\Delta y}{\Delta x}} = \frac{1}{\dfrac{dy}{dx}},$$

but this one line is not, by itself, the equivalent of a theorem and proof which present conditions under which the formula is correct.

Our general discussion of inverse functions and Theorem 8.33 will now be used to guide us to six functions that are called inverse trigonometric functions even though they are in fact inverses of restrictions of trigonometric functions. It will turn out that the six functions will all have values between 0 and $\pi/2$, and that all of the information implied by Figures 8.341, 8.342, and 8.343 will be correct, provided $0 < x < 1$ in

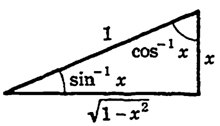


Figure 8.341

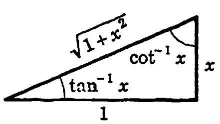


Figure 8.342

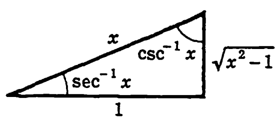


Figure 8.343

Figure 8.341, $x > 0$ in Figure 8.342, and $x > 1$ in Figure 8.343. The first figure shows, for example, that

$$\sin(\sin^{-1} x) = x, \qquad \tan(\sin^{-1} x) = \frac{x}{\sqrt{1 - x^2}}$$

when $0 < x < 1$. These triangles can sometimes provide helpful information even when the functions appear in problems involving the misfortunes of gamblers who bet on horse races.

To begin, let f be the trigonometric function for which $f(x) = \sin x$, the domain D being the infinite interval $-\infty < x < \infty$ and the range

R being the closed interval $-1 \leqq y \leqq 1$. When we are called upon to select a domain D_1 to which we can apply Theorem 8.33, the graph of $y = \sin x$ can remind us of known properties of its derivative which show us that if we are going to be able to apply Theorem 8.33, we must let D_1 be the interval $-\pi/2 < x < \pi/2$ or some other interval lying $n\pi$ (n an integer) units to the right or left of it. Making the simplest choice, we let f_1 be the function, defined only when $-\pi/2 \leqq x \leqq \pi/2$, for which $f_1(x) = \sin x$. The inverse of f_1 is called the inverse sine. If $y = f_1(x)$, so that $y = \sin x$ and $-\pi/2 \leqq x \leqq \pi/2$ and $-1 \leqq y \leqq 1$, we write $x = \sin^{-1} y$. The graph of $x = \sin^{-1} y$ coincides with the part of the graph of $y = \sin x$ for which $-\pi/2 \leqq x \leqq \pi/2$, and the graph is shown in Figure 8.351. When we interchange x and y, which amounts to replacing

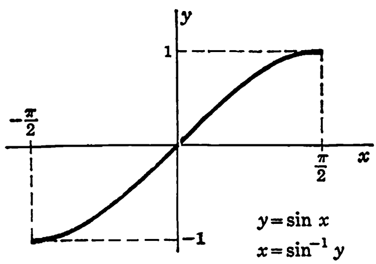


Figure 8.351

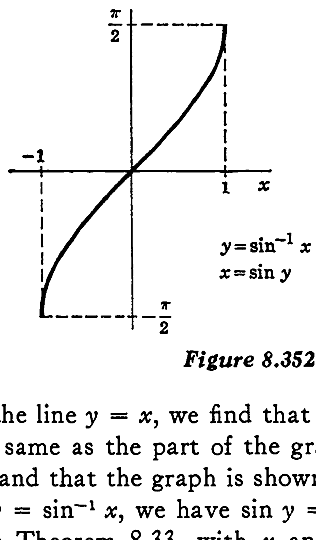


Figure 8.352

each point (x,y) by its "image" (y,x) in the line $y = x$, we find that the graph of the equation $y = \sin^{-1} x$ is the same as the part of the graph of $x = \sin y$ for which $-\pi/2 \leqq y \leqq \pi/2$ and that the graph is shown in Figure 8.352. When $-1 < x < 1$ and $y = \sin^{-1} x$, we have $\sin y = x$, $-\pi/2 < y < \pi/2$, and $\cos y > 0$. Since Theorem 8.33, with x and y interchanged, implies that dy/dx exists, we can use the chain rule to obtain $\cos y(dy/dx) = 1$ and hence

$$(8.353) \qquad \frac{d}{dx}\sin^{-1} x = \frac{dy}{dx} = \frac{1}{\cos y} = \frac{1}{\sqrt{1-\sin^2 y}} = \frac{1}{\sqrt{1-x^2}}.$$

While we could not be expected to guess the result in this formula, we can look at Figure 8.352 and see that it seems to be a very reasonable answer. One who feels that inverse sines and their derivatives could never be useful should be informed at once that (8.353) gives the formula

$$(8.354) \qquad \int \frac{1}{\sqrt{1-x^2}}\,dx = \sin^{-1} x + c \qquad (|x| < 1)$$

which is useful because the integral appears in important problems.

The inverse cosine is the inverse of the function f_1, with domain $0 \leqq x \leqq \pi$ and range $-1 \leqq x \leqq 1$, for which $f_1(x) = \cos x$. Graphs of $x = \cos^{-1} y$, of $y = \cos^{-1} x$, and of the relevant parts of $y = \cos x$ and $x = \cos y$ appear in Figures 8.355 and 8.356. When $-1 < x < 1$ and

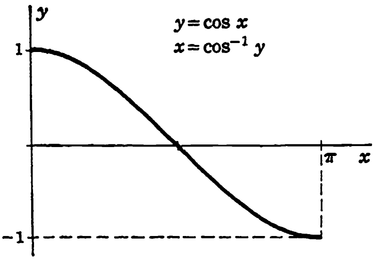


Figure 8.355

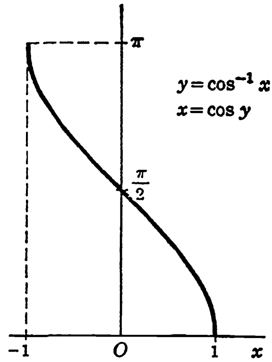


Figure 8.356

$y = \cos^{-1} x$, then $\cos y = x$, $0 < y < \pi$, and $\sin y > 0$. Again Theorem 8.33 implies that dy/dx exists, so $(-\sin y)\,(dy/dx) = 1$ and

$$(8.357) \qquad \frac{d}{dx}\cos^{-1} x = \frac{dy}{dx} = \frac{-1}{\sin y} = \frac{-1}{\sqrt{1-\cos^2 y}} = \frac{-1}{\sqrt{1-x^2}}.$$

This gives the integration formula

$$(8.358) \qquad \int \frac{1}{\sqrt{1-x^2}}\,dx = -\cos^{-1} x + c \qquad (|x| < 1)$$

which, because $\sin^{-1} x + \cos^{-1} x = \pi/2$, does not contradict (8.354).

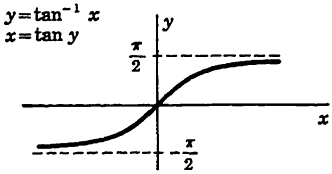


Figure 8.36

As Figure 8.171 may suggest, the inverse tangent is the inverse of the function f_1, with domain $-\pi/2 < x < \pi/2$ and range $-\infty < x < \infty$, for which $f_1(x) = \tan x$. The graph of $y = \tan^{-1} x$, which is the same as a part (sometimes called the principal part) of the graph of $x = \tan y$, is shown in Figure 8.36. The domain of the inverse tangent is the entire set of numbers, and if $y = \tan^{-1} x$, then $\tan y = x$ and $\sec^2 y(dy/dx) = 1$, so

$$(8.361) \qquad \frac{d}{dx}\tan^{-1} x = \frac{dy}{dx} = \frac{1}{\sec^2 y} = \frac{1}{1+\tan^2 y} = \frac{1}{1+x^2}.$$

This gives the important formula

$$(8.362) \qquad \int \frac{1}{1+x^2}\,dx = \tan^{-1} x + c.$$

The inverse cotangent is the inverse of the function f_1, with domain $0 < x < \pi$ and range $-\infty < y < \infty$. A graph of $y = \cot^{-1} x$ is shown in Figure 8.37. The domain of the inverse cotangent is the entire set of

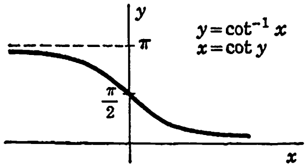


Figure 8.37

real numbers and if $y = \cot^{-1} x$, then $\cot y = x$ and $-\csc^2 y(dy/dx) = 1$, so

$$(8.371) \qquad \frac{d}{dx}\cot^{-1} x = \frac{dy}{dx} = \frac{-1}{\csc^2 y} = \frac{-1}{1 + \cot^2 y} = \frac{-1}{1 + x^2}.$$

As Figure 8.233 indicates, the graph of $y = \sec x$ presents a difficulty that has not previously appeared. It is impossible to select an interval D_1 of values of x such that $\sec x$ is defined and continuous over D_1 and each y in the range of $\sec x$ is attained for some x in D_1. The best we can do is let D_1 be the interval $0 \leqq x \leqq \pi$ with the center point $x = \pi/2$ omitted. The inverse secant is then the inverse of the function f_1 with domain D_1 for which $f_1(x) = \sec x$. The graph of $y = \sec^{-1} x$, which coincides with a part of the graph of $x = \sec y$, is shown in Figure 8.38. If $x > 1$ or $x < -1$ and $y = \sec^{-1} x$, then $\sec y = x$, $0 < y < \pi$, $y \neq \pi/2$, and $\sec y \tan y > 0$. Again Theorem 8.33 implies that dy/dx exists, so

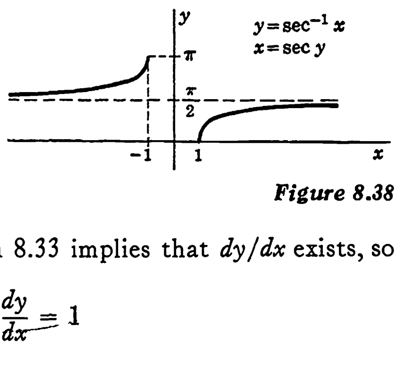


Figure 8.38

$$\sec y \tan y \frac{dy}{dx} = 1$$

and

$$(8.381) \qquad \frac{d}{dx}\sec^{-1} x = \frac{dy}{dx} = \frac{1}{\sec y \tan y} = \frac{1}{|\sec y|\,|\tan y|}$$
$$= \frac{1}{|\sec y|\sqrt{\sec^2 y - 1}} = \frac{1}{|x|\sqrt{x^2 - 1}}.$$

This gives the formula

$$(8.382) \qquad \int \frac{1}{|x|\sqrt{x^2 - 1}}\,dx = \sec^{-1} x + c,$$

which is valid when $x > 1$ and also when $x < -1$.

The graph of $y = \csc x$ presents the same difficulty that the graph of $y = \sec x$ presented. The best we can do is let D_1 be the interval $-\pi/2 \leqq x \leqq \pi/2$ with the center point $x = 0$ omitted. The inverse cosecant is then the inverse of the function f_1 with domain D_1 for which $f_1(x) = \csc x$. The graph of $y = \csc^{-1} x$, which coincides with a part of the graph of $x = \csc y$, is shown in Figure 8.383.

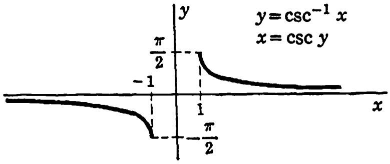


Figure 8.383

If $x > 1$ or $x < -1$ and $y = \csc^{-1} x$, then $\csc y = x$, $-\pi/2 < y < \pi/2$, $y \neq 0$, and hence $\csc y \cot y > 0$. Again Theorem 8.33 implies that dy/dx exists, so

$$-\csc y \cot y \frac{dy}{dx} = 1$$

and

$$(8.384)\quad \frac{d}{dx}\csc^{-1} x = \frac{dy}{dx} = \frac{-1}{\csc y \cot y} = \frac{-1}{|\csc y|\,|\cot y|}$$
$$= \frac{-1}{|\csc y|\sqrt{\csc^2 y - 1}} = \frac{-1}{|x|\sqrt{x^2 - 1}}.$$

Problems 8.39

1 Supposing that $a > 0$ and $b^2 < 4ac$, show how the first of the formulas

$$(1)\quad \int \frac{1}{1+u^2}\,du = \tan^{-1} u + c, \qquad \int \frac{1}{a^2+u^2}\,du = \frac{1}{a}\tan^{-1}\frac{u}{a} + c$$

$$(2)\quad \int \frac{1}{ax^2+bx+c}\,dx = \frac{2}{\sqrt{4ac-b^2}}\tan^{-1}\frac{2ax+b}{\sqrt{4ac-b^2}} + c$$

can be used to obtain the other two. *Remark:* Everyone should know how the last of these formulas and many similar ones are presented in standard integral tables. Let $a > 0$ and let

$$X = ax^2 + bx + c, \qquad q = b^2 - 4ac.$$

Then (2) takes the form

$$(3)\quad \int \frac{1}{X}\,dx = \frac{2}{\sqrt{-q}}\tan^{-1}\frac{2ax+b}{\sqrt{-q}} + c$$

when $q < 0$ and hence $-q > 0$. The derivation of (3) is based upon the identity

$$(4)\quad X = a\left[x^2 + \frac{b}{a}x + \frac{c}{a}\right] = a\left[\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}\right) + \left(\frac{c}{a} - \frac{b^2}{4a^2}\right)\right]$$
$$= a\left[\left(x + \frac{b}{2a}\right)^2 + \left(\frac{\sqrt{4ac-b^2}}{2a}\right)^2\right] = a\left[\left(x + \frac{b}{2a}\right)^2 + \left(\frac{\sqrt{-q}}{2a}\right)^2\right]$$

which involves completing a square. It may be time to renew the idea that normal persons who have forgotten (or never remembered) needed integral formulas either work them out to satisfy their vanities and preserve their mathematical powers or stoop to hunting them in a book containing a table of integrals.

2 Show that

$$\int_0^y \frac{1}{y^2 + x^2}\,dx = \frac{\pi}{4y}$$

when $y > 0$ and that

$$\int_1^a \left\{\int_0^y \frac{1}{y^2 + x^2}\,dx\right\} dy = \frac{\pi \log a}{4}$$

when $a > 0$.

3 Supposing that $-1 < x < 1$ and

$$T_n(x) = 2^{1-n} \cos\,(n \cos^{-1} x), \tag{1}$$

find formulas for $T_n'(x)$ and $T_n''(x)$. Then multiply $T_n(x)$ by n^2, multiply $T_n'(x)$ by $-x$, multiply $T_n''(x)$ by $(1 - x^2)$, and add the results to discover that

$$(1 - x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + n^2y = 0 \tag{2}$$

when $y = T_n(x)$. *Remark:* This problem involves much more than meets the eye. When n is a positive integer, T_n is the Tchebycheff polynomial of degree n and (2) is the Tchebycheff differential equation of order n. One who wishes a simple and brief discussion of these astonishing things can find it in the author's textbook "Differential Equations," 2d ed , McGraw-Hill Book Company, Inc., New York, 1960.

4 Supposing that $-1 < x < 1$ and $u = (\sin^{-1} x)^2$, calculate the first two derivatives of u and show that

$$(1 - x^2)\frac{d^2u}{dx^2} - x\frac{du}{dx} = 2.$$

5 The eye of a man is at a point E that is, as in Figure 8.391, x feet from a vertical wall bearing a picture the bottom and top of which are a and b feet above

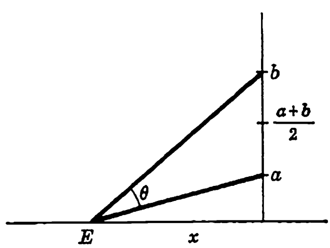


Figure 8.391

eye level. The angle θ which the picture subtends at the eye is surely small when x is small and when x is large. Derive the formulas

$$\theta = \tan^{-1}\frac{b}{x} - \tan^{-1}\frac{a}{x}, \qquad \frac{d\theta}{dx} = \frac{(b-a)(ab - x^2)}{(a^2 + x^2)(b^2 + x^2)}$$

and show that θ is a maximum when $x = \sqrt{ab}$. *Remark:* Addicts of elementary plane geometry can be pleased to see that the x which maximizes θ can be found in another way. As in Figure 8.392 let C be a circle which passes through the

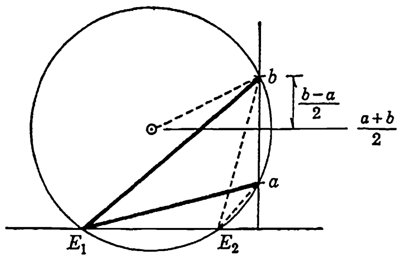


Figure 8.392

top and bottom of the picture and intersects the eye-level line L at two points E_1 and E_2. The angles which the picture subtends at E_1 and E_2 are equal. Of all such circles C that intersect L, the smallest one that produces the greatest θ is the circle C_0 for which E_1 and E_2 coincide, so C_0 is tangent to L. The radius of C_0 is $(b + a)/2$, and use of an appropriate right triangle shows that the distance from the wall to the center and to the critical point of tangency of C_0 is

$$\sqrt{\left(\frac{b+a}{2}\right)^2 - \left(\frac{b-a}{2}\right)^2} \qquad \text{or} \qquad \sqrt{ab}.$$

6 Supposing that $a > 0$ and $|ax| < 1$, prove the formula

$$\tan^{-1} x + \tan^{-1} a = \tan^{-1} \frac{a+x}{1-ax}$$

by showing that the two members of the equation are equal when $x = 0$ and have equal derivatives when $|ax| < 1$.

7 With or without undertaking to prove the fundamental fact that $\cot^{-1} x = \tan^{-1}(1/x)$ when $x > 0$, suppose that $x \neq 0$ and fill in the omitted steps in the calculation

$$\frac{d}{dx}\cot^{-1} x = \frac{d}{dx}\tan^{-1}\frac{1}{x} = \cdots = \frac{-1}{1+x^2}.$$

8 With or without undertaking to prove the fundamental fact that $\sec^{-1} x = \cos^{-1}(1/x)$ when $|x| \geqq 1$, suppose that $|x| > 1$ and fill in the omitted steps in the calculation

$$\frac{d}{dx}\sec^{-1} x = \frac{d}{dx}\cos^{-1}\frac{1}{x} = \cdots = \frac{1}{|x|\sqrt{x^2-1}}.$$

Hint: Do not forget the chain rule and the fact that $x^2 = |x|\,|x|$.

9 With the aid of Figure 8.343 we can see that the formula

$$\sec^{-1} x = \sin^{-1}\frac{\sqrt{x^2-1}}{x} \tag{1}$$

is valid at least when $x > 1$. Differentiate the two members of (1) and determine the set of numbers x for which the derivatives are equal.

10 Letting N be a positive integer (which we shall not call "arbitrary"), manufacture and solve N problems of the nature of the preceding one.

11 Prove that

$$\int_0^1 \frac{1}{1+x^2}\,dx = \frac{\pi}{4}, \qquad \int_0^\infty \frac{1}{1+x^2}\,dx = \frac{\pi}{2}.$$

12 *Remark:* We should not be so busy that we never have time to look at our formulas and see that we can learn things from them. Supposing for simplicity that $x > 0$, we can look at the formula

$$\tan^{-1} x = \int_0^x \frac{1}{1+t^2}\,dt \tag{1}$$

and realize that each member is equal to the area of the shaded region in Figure 8.393. Any information we can obtain about $\tan^{-1} x$ or the area or the integral

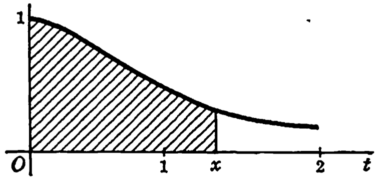


Figure 8.393

therefore provides information about all three. While there are much more complicated ways of getting the information, we can quickly learn some significant facts by starting with the formula

$$\frac{1}{1+t^2} = 1 - t^2 + t^4 - t^6 + \cdots + t^{4n} - \frac{t^{4n+2}}{1+t^2} \tag{2}$$

which can be obtained by long division and can be checked by multiplication by $(1 + t^2)$. Integrating this over the interval $0 \leqq t \leqq x$ gives the formula

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + \frac{x^{4n+1}}{4n+1} - R_n(x), \tag{3}$$

where

$$R_n(x) = \int_0^x \frac{t^{4n+2}}{1+t^2}\,dt. \tag{4}$$

Suppose now that $0 \leqq x \leqq 1$. Then $1 \leqq 1 + t^2 \leqq 2$ and therefore

$$\int_0^x \frac{t^{4n+2}}{2}\,dt \leqq R_n(x) \leqq \int_0^x t^{4n+2}\,dt \tag{5}$$

so

$$\frac{x^{4n+3}}{2(4n+3)} \leqq R_n(x) \leqq \frac{x^{4n+3}}{4n+3}. \tag{6}$$

This gives us an excellent estimate of $R_n(x)$ which implies the more crude estimate

$$0 \leqq R_n(x) \leqq \frac{1}{4n+3}. \tag{7}$$

This and the sandwich theorem imply that $R_n(x) \to 0$ as $n \to \infty$, and the definition involving (5.622) shows that

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \frac{x^{11}}{11} + \cdots \tag{8}$$

when $0 < x \leqq 1$. Changing the sign of x changes the signs of both sides and shows that the formula is correct when $-1 \leqq x \leqq 1$. Putting $x = 1$ gives the famous Leibniz formula

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots, \tag{9}$$

which is of particular interest to those who have not previously seen a formula from which π could be calculated. Putting $x = 1$ in (6) shows that we must take many terms of the series (8) to obtain a sum agreeing with $\pi/4$ to three or four decimal places, and we say that the series "converges slowly." The series in (8) converges more rapidly when x is nearer 0. We are not now in the computing business, but it is easy to verify that

$$\frac{\pi}{4} = \tan^{-1}\frac{1}{2} + \tan^{-1}\frac{1}{3}, \tag{10}$$

by taking tangents of both sides, and to use (8) to show that π is roughly 3.14. The formula of Machin (1680–1751)

$$\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}, \tag{11}$$

which is not so easily proved, is used by professional computers.

13 Find whether the function f for which $f(x) = x + \sin x$ has an inverse and, if so, whether the inverse is differentiable.

14 Let E be an interval of values of x; it could be the interval $x > 0$, and it could be the interval $-1 \leqq x \leqq 1$. Let D be the operator or transformer or differentiator which, when applied to a function f which is defined and differentiable over E, produces the function ϕ for which $\phi(x) = f'(x)$ when x is in E. The domain Δ of D is then the set of functions f that are defined and differentiable over E, and the range R of D is the set of functions ϕ that are defined over E and are derivatives of functions differentiable over E. Show that D does not possess an inverse.

15 Let E be the interval $-1 < x < 1$. Let D_1 be the restriction of the operator D of the preceding problem to the domain Δ_1 consisting of functions f defined over E for which $f(0) = 0$ and f' exists and is continuous over E. Show that D_1 has an inverse.

16 Let E be the interval $-1 < x < 1$. Let I_1 be the operator or transformer or integrator which, when applied to a function g which is continuous over E, produces the function G for which

$$G(x) = \int_0^x g(t)\,dt$$

when x is in E. Describe the domain and range of I_1 and show that I_1 has an inverse.

17 Try to understand and even prove the statement that each of the operators D_1 and I_1 (of the two preceding problems) is the inverse of the other.

8.4 *Integration by trigonometric and other substitutions* We begin with a statement of the fact that there is an elementary function F whose derivative with respect to x is $\sqrt{a^2 - x^2}$. Letting F denote one such function, we try to learn about F by writing

$$F(x) = \int \sqrt{a^2 - x^2}\, dx \tag{8.41}$$

and searching for an idea. Once upon a time somebody discovered that if we substitute $x = a \sin \theta$, then the integrand will become $\sqrt{a^2 - a^2 \sin^2 \theta}$ or $\sqrt{a^2 \cos^2 \theta}$, and this is $a \cos \theta$ if $a \cos \theta > 0$. Thus a trigonometric substitution removes the radical and leaves $\pm a \cos \theta$, but we can still be unsure of the meaning of $\int \cos \theta\, dx$. Hence we must pause to make an observation.

The chain rule of Theorem 3.65 tells us that if

$$F'(x) = f(x) \tag{8.42}$$

and if u is a differentiable function whose range lies in the domain of F, then

$$\begin{aligned} \frac{d}{dt} F(u(t)) &= F'(u(t))u'(t) \\ &= f(u(t))u'(t). \end{aligned} \tag{8.421}$$

This gives the following important substitution theorem which shows how to replace x by $u(t)$ in integrals.

Theorem 8.43 *If*

$$F(x) = \int f(x)\, dx \tag{8.431}$$

and if u is a differentiable function whose range lies in the domain of f, then

$$F(u(t)) = \int f(u(t))u'(t)\, dt. \tag{8.432}$$

This theorem is used so often that it is worthwhile to be able to get from the right side of (8.431) to the right side of (8.432) in a purely formal way without thinking about the way in which the theorem was proved and the meanings of the formulas. We replace the old integrand $f(x)$ by the new integrand $f(u(t))u'(t)$ and the old dx by the new dt. This seems like a simple ritual, but the factor $u'(t)$ might be forgotten when problems are being solved. It seems to be safer and easier to think of $f(x)$ being replaced by $f(u(t))$ and dx being replaced by $u'(t)\, dt$. If we follow historical precedents, we become carried away by our own enthusiasm and try to eliminate all possibility of overlooking the factor $u'(t)$ by creating the fiction that dx is a number (a bunch of bananas would serve the same purpose) and $u'(t)\, dt$ is the same thing. There is a sense in which this whole business is utterly silly, but it really is a convenience to imagine that dx, $u'(t)$, and dt are three numbers for which $dx = u'(t)\,dt$. When $x = u(t)$, we can differentiate to get $\dfrac{dx}{dt} = u'(t)$, and the pretense that $\dfrac{dx}{dt}$ is the quotient of two numbers then enables us to multiply by dt to get the formula $dx = u'(t)\,dt$ that tells us to replace dx by $u'(t)\,dt$. It is possible to try to say more about these matters but, for present purposes, the important thing is that we do not forget the factor $u'(t)$.

Since it is not always convenient to introduce a new symbol, such as u, for the function which carries t into x, we restate the theorem in the following form which can be ignored by those who prefer the first version.

Theorem 8.433 *If*

$$F(x) = \int f(x)\,dx \tag{8.434}$$

and if x is a differentiable function whose range lies in the domain of f, then

$$F(x(t)) = \int f(x(t))x'(t)\,dt \tag{8.435}$$

or

$$F(x(t)) = \int f(x(t))\,\frac{dx}{dt}\,dt \tag{8.436}$$

or

$$F(x) = \int f(x)\,dx. \tag{8.437}$$

We give further attention to changes of variables in integrals by proving the following very useful variant of Theorem 8.433 which involves Riemann integrals.

Theorem 8.44 *If f is continuous over the interval $a \leqq x \leqq b$ and if u is a function which has a continuous derivative and is such that $u(t)$ increases from a to b as t increases (or decreases) from α to β, then*

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f(u(t))u'(t)\,dt. \tag{8.441}$$

To prove this theorem, let F be a function whose derivative is f so that $\int f(x)\,dx = F(x) + c$ and

$$\int_a^b f(x)\,dx = F(x)\Big]_a^b = F(b) - F(a). \tag{8.442}$$

The chain formula for derivatives then implies that

$$\int_\alpha^\beta f(u(t))u'(t)\,dt = F(u(t))\Big]_\alpha^\beta = F(u(\beta)) - F(u(\alpha)). \tag{8.443}$$

The required conclusion (8.441) then follows from the fact that α and β are chosen such that $u(\alpha) = a$ and $u(\beta) = b$. Careful attention must be given to the manner in which the new limits of integration are determined. When we set $x = u(t)$, we must determine α such that x is a when t is α and we must determine β such that x is b when $t = \beta$.

We are now ready to attack (8.41) and, since this is the first time we have made a substitution (or changed the variable) in an integral, we proceed with great caution. Supposing that $a > 0$ and that $-a \leqq x \leqq a$, we set

$$F(x) = \int \sqrt{a^2 - x^2}\, dx \tag{8.45}$$

and let $x = a \sin \theta$, where $-\pi/2 \leqq \theta \leqq \pi/2$. Figure 8.451 always helps

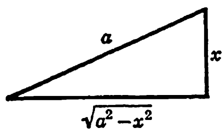


Figure 8.451

us to see what we are doing on occasions like this. Since $dx/d\theta = a \cos \theta$, we obtain

$$F(a \sin \theta) = \int \sqrt{a^2 \cos^2 \theta}\, a \cos \theta\, d\theta. \tag{8.452}$$

But $\cos \theta \geqq 0$ when $-\pi/2 \leqq \theta \leqq \pi/2$, so

$$\begin{aligned} F(a \sin \theta) &= a^2 \int \cos^2 \theta\, d\theta = \frac{a^2}{2} \int (1 + \cos 2\theta)\, d\theta \\ &= \frac{a^2}{2} [\theta + \tfrac{1}{2} \sin 2\theta] + c \\ &= \tfrac{1}{2}[a^2\theta + a^2 \sin \theta \cos \theta] + c. \end{aligned} \tag{8.453}$$

Since each x in the range $-1 \leqq x \leqq 1$ is obtained for a θ in the interval $-\pi/2 < \theta < \pi/2$, we can use Figure 8.451 and the fact that $\cos \theta \geqq 0$ to return to the original variable x so that

$$F(x) = \tfrac{1}{2}\left[a^2 \sin^{-1} \frac{x}{a} + x\sqrt{a^2 - x^2}\right] + c. \tag{8.454}$$

This gives the formula

$$\int \sqrt{a^2 - x^2}\, dx = \tfrac{1}{2}\left[a^2 \sin^{-1} \frac{x}{a} + x\sqrt{a^2 - x^2}\right] + c. \tag{8.455}$$

We are not always so careful about all of the details, and we do not always get correct answers either. There is a reason why we can sometimes be careless about quadrants in which angles lie. There is a "theory of analytic functions" that guarantees that, in many situations, a formula which is correct when angles lie in the first quadrant will be correct wherever we want to use it.

When we wish to make a substitution to evaluate the integral in

$$I = \int \frac{1}{\sqrt{a^2 - x^2}}\, dx, \tag{8.46}$$

we can clarify our work and save writing the integral by denoting it by I, or by I_1 or J if we wish to distinguish it from other integrals. We can then put $x = a \sin \theta$ and allow the variables to shift for themselves while we write

$$\begin{aligned} I &= \int \frac{1}{\sqrt{a^2 \cos^2 \theta}}\, a \cos \theta\, d\theta = \int 1\, d\theta \\ &= \theta + c = \sin^{-1} \frac{x}{a} + c. \end{aligned} \tag{8.461}$$

We have evaluated an integral that previously appeared in (8.354).

The identity $1 - \cos^2 \theta = \sin^2 \theta$ provides the reason why the substitution $x = a \sin \theta$ eliminated the square roots from the integrands of the above integrals. The identities

$$\sec^2 \theta - 1 = \tan^2 \theta, \qquad \tan^2 \theta + 1 = \sec^2 \theta, \tag{8.47}$$

which are obtainable from the one involving sines and cosines by dividing by $\cos^2 \theta$, are less familiar but are nevertheless important when we want to use them. Their uses are illustrated below.

To evaluate the integral in

$$J = \int \frac{1}{(x^2 + a^2)^{3/2}}\, dx, \tag{8.471}$$

we look at it and generate the idea that we should try setting $x = a \tan \theta$. Then $x' = a \sec^2 \theta$, so

$$\begin{aligned} J &= \int \frac{1}{(a^2 \sec^2 \theta)^{3/2}}\, a \sec^2 \theta\, d\theta = \frac{1}{a^2} \int \frac{1}{\sec \theta}\, d\theta \\ &= \frac{1}{a^2} \int \cos \theta\, d\theta = \frac{1}{a^2} \sin \theta + c = \frac{x}{a^2 \sqrt{x^2 + a^2}} + c, \end{aligned} \tag{8.472}$$

the last step being assisted by a figure showing θ, a, and x in the right way.

To evaluate the integral in

$$J_1 = \int \frac{1}{(x^2 - a^2)^{3/2}}\, dx, \tag{8.473}$$

we look at it and generate the idea that we should try setting $x = a \sec \theta$. Then $x' = a \sec \theta \tan \theta$, so

$$(8.474) \qquad \begin{aligned} J_1 &= \int \frac{1}{(a^2 \tan^2 \theta)^{3/2}} a \sec \theta \tan \theta \, d\theta \\ &= \frac{1}{a^2} \int \frac{\sec \theta}{\tan^2 \theta} \, d\theta = \frac{1}{a^2} \int \frac{\cos \theta}{\sin^2 \theta} \, d\theta \\ &= \frac{1}{a^2} \int (\sin \theta)^{-2} \cos \theta \, d\theta = \frac{1}{a^2} \frac{(\sin \theta)^{-1}}{-1} + c \\ &= -\frac{1}{a^2} \frac{x}{\sqrt{x^2 - a^2}} + c. \end{aligned}$$

It is not implied that we know in advance that the integrand should be written in terms of sines and cosines, but trying this possibility and seeing that the power formula can be applied is all a part of the game.

Our last problem has a lengthy solution. We can bravely start to evaluate the integral in

$$(8.48) \qquad J_2 = \int \sqrt{x^2 - a^2} \, dx$$

by setting $x = a \sec \theta$. Then $x' = a \sec \theta \tan \theta$ and

$$J_2 = \int \sqrt{a^2 \tan^2 \theta} \, a \sec \theta \tan \theta \, d\theta$$

so

$$(8.481) \qquad J_2 = a^2 \int \tan^2 \theta \sec \theta \, d\theta.$$

It happens that this integral is an elementary function, but there is no simple direct way of discovering the facts. Authors and teachers, as well as students, can work for hours on this problem unless they remember how to solve it or (and this much better) have a guiding principle. The principle is the following. If we want to learn something about an integral and other methods fail to be helpful, we try integration by parts. By this we mean that we try to determine two functions u and v of the variable of integration in such a way that the integral we are studying will be the left member of the formula

$$(8.482) \qquad \int u(\theta) v'(\theta) \, d\theta = u(\theta) v(\theta) - \int v(\theta) u'(\theta) \, d\theta$$

and, moreover, the "parts" on the right side enable us to make progress with our problem. Assuming that u and v have continuous derivatives over intervals where we use the formula, differentiation shows that the formula is correct. The trick is to use it effectively. Naturally, we want the functions u, v, u', v' to be as simple as possible, and it is quite easy to discover that the best way to convert J_2 into the left member of (8.482) is to set

$$(8.483) \qquad \begin{cases} u(\theta) = a^2 \tan \theta, & v'(\theta) = \sec \theta \tan \theta \\ u'(\theta) = a^2 \sec^2 \theta, & v(\theta) = \sec \theta. \end{cases}$$

It is always a good idea to display u and v' in one line and u' and v in a lower line and to know that the formula (8.482) for integration by parts says that the integral of the product of the top two is equal to the product of u and v minus the integral of the product of the bottom two.† Thus

$$J_2 = a^2 \sec\theta \tan\theta - a^2 \int \sec^3\theta\, d\theta. \tag{8.484}$$

The last integral seems to be quite as recalcitrant as J_2, and this really is true because

$$\begin{aligned} a^2\int \sec^3\theta\, d\theta &= a^2\int \sec^2\theta \sec\theta\, d\theta = a^2\int [1 + \tan^2\theta] \sec\theta\, d\theta \\ &= a^2\int \sec\theta\, d\theta + a^2\int \tan^2\theta \sec\theta\, d\theta \\ &= a^2 \log(\sec\theta + \tan\theta) + c + J_2. \end{aligned}$$

Instead of developing faint hearts by throwing everything into a wastebasket, we substitute our last result into (8.484) and get

$$J_2 = a^2 \sec\theta \tan\theta - a^2 \log(\sec\theta + \tan\theta) - c - J_2. \tag{8.485}$$

If we can now suddenly remember that we are trying to find J_2, we can transpose the term $-J_2$ (or add J_2 to both sides of our equation if transposing is onerous) and divide by 2 to obtain

$$J_2 = \tfrac{1}{2}a^2 \sec\theta \tan\theta - \tfrac{1}{2}a^2 \log(\sec\theta + \tan\theta) + c, \tag{8.486}$$

where the new constant c, like the old $-c/2$, can be any constant. With the aid of a figure showing how θ, a, and x are related, we obtain

$$\begin{aligned} J_2 = \int \sqrt{x^2 - a^2}\, dx &= \tfrac{1}{2}x\sqrt{x^2 - a^2} \\ &\quad - \tfrac{1}{2}a^2 \log \frac{x + \sqrt{x^2 - a^2}}{a} + c. \end{aligned} \tag{8.487}$$

If we wish, we can write the logarithm of the quotient as a difference of logarithms and combine the constant part with the c to obtain the formula

$$\begin{aligned} J_2 = \int \sqrt{x^2 - a^2}\, dx &= \tfrac{1}{2}x\sqrt{x^2 - a^2} \\ &\quad - \tfrac{1}{2}a^2 \log(x + \sqrt{x^2 - a^2}) + c, \end{aligned} \tag{8.488}$$

in which c has another value. The answers in (8.487) and (8.488) look different, but they are both correct. The last one may seem simpler, but the first one gives $c = 0$ if $J_2 = 0$ when $x = a$.

We should recognize the fact that all or nearly all of the integration formulas in the text and problems of this section appear in books of tables. We should hear more than once that books of tables are often used as labor-saving devices, but that many persons like to derive the formulas they use to preserve and improve their mathematical acumen.

† These matters are very important, and it is required that we think about them more than once. Our concentrated attack upon integration by parts appears in Section 9.5.

Problems 8.49

1 By making an appropriate figure and trigonometric substitution, fill in all of the missing details which show that

(a) $\displaystyle\int \frac{1}{\sqrt{x^2 - a^2}}\,dx = \int \sec\theta\,d\theta = \log\,(x + \sqrt{x^2 - a^2}) + c$

(b) $\displaystyle\int \frac{1}{a^2 - x^2}\,dx = \frac{1}{a}\int \sec\theta\,d\theta$

$$= \frac{1}{a}\log\frac{a + x}{\sqrt{a^2 - x^2}} + c = \frac{1}{2a}\log\frac{a + x}{a - x} + c \qquad (|x| < a)$$

(c) $\displaystyle\int \frac{1}{x^2 - a^2}\,dx = \frac{1}{a}\int \csc\theta\,d\theta$

$$= -\frac{1}{a}\log\frac{x + a}{\sqrt{x^2 - a^2}} + c = \frac{1}{2a}\log\frac{x - a}{x + a} + c \qquad (x > a)$$

(d) $\displaystyle\int (a^2 - x^2)^{3/2}\,dx = a^4\int \cos^4\theta\,d\theta$

2 When $x > 0$, the identity

$$(1) \qquad \sqrt{1 + \frac{1}{x^2}} = \frac{\sqrt{x^2 + 1}}{x} = \frac{x^2 + 1}{x\sqrt{1 + x^2}}$$

shows that

$$(2) \qquad \int \sqrt{1 + \frac{1}{x^2}}\,dx = \int (1 + x^2)^{-1/2}x\,dx + \int \frac{1}{x\sqrt{1 + x^2}}\,dx.$$

The first integral on the right is easily evaluated, and the second can be simplified by a trigonometric substitution. Use these ideas to derive the formula

$$(3) \qquad \int \sqrt{1 + \frac{1}{x^2}}\,dx = \sqrt{1 + x^2} + \log x - \log\,(1 + \sqrt{1 + x^2}) + c.$$

3 Sketch graphs of $y = e^x$ and $y = \log x$ in a single figure and note that each is the mirror image of the other in the line $y = x$. Observe that the arc C_1 on the graph of $y = e^x$ which joins two points (p_1,q_1) and (p_2,q_2) on the graph of $y = e^x$ is congruent to the arc C_2 on the graph of $y = \log x$ which joins the two points (q_1,p_1) and (q_2,p_2) on the graph of $y = \log x$. Letting L denote the length of C_1 and C_2, show that

$$(1) \qquad L = \int_{p_1}^{p_2} \sqrt{1 + e^{2x}}\,dx = \int_{e^{p_1}}^{e^{p_2}} \sqrt{1 + \frac{1}{x^2}}\,dx.$$

This gives the formula

$$(2) \qquad \int_{p_1}^{p_2} \sqrt{1 + e^{2t}}\,dt = \int_{e^{p_1}}^{e^{p_2}} \sqrt{1 + \frac{1}{x^2}}\,dx.$$

To show that errors and misprints have not led us to an incorrect formula, show that the substitution $t = \log x$ converts the first integral in (2) into the second integral in (2).

4 In Problem 2 of Problems 4.39, we called attention to the important nonelementary beta integral formula

$$(1) \qquad \int_0^1 x^p(1-x)^q\,dx \, \frac{p!q!}{(p+q+1)!}$$

which is correct when the integral is a Riemann integral and p and q are nonnegative; we can now report that it is correct when the integral is a Riemann-Cauchy integral and p and q exceed -1. Show that putting $x = \sin^2\theta$, where $0 \leqq \theta \leqq \pi/2$ when $0 \leqq x \leqq 1$, yields the formula

$$(2) \qquad \int_0^{\pi/2} \sin^{2p+1}\theta \cos^{2q+1}\theta\,d\theta = \frac{1}{2}\frac{p!q!}{(p+q+1)!}.$$

Be sure to observe the fact that the limits of integration are correctly determined; setting $x = 0$ in formulas involving x is equivalent to setting $\theta = 0$ in formulas involving θ, and setting $x = 1$ in formulas involving x is equivalent to setting $\theta = \pi/2$ in formulas involving θ. Finally, replace θ by x and let $2p + 1 = \alpha$, $2q + 1 = \beta$ to obtain the formula

$$(3) \qquad \int_0^{\pi/2} \sin^\alpha x \cos^\beta x\,dx = \frac{1}{2}\,\frac{\left(\dfrac{\alpha-1}{2}\right)!\left(\dfrac{\beta-1}{2}\right)!}{\left(\dfrac{\alpha+\beta}{2}\right)!},$$

which is valid when α and β exceed -1. *Remark:* If we happen to know that $(-\frac{1}{2})! = \sqrt{\pi}$, we can conclude from (3) that

$$(4) \qquad \int_0^{\pi/2} \sin^\alpha x\,dx = \sqrt{\frac{\pi}{2}}\,\frac{\left(\dfrac{\alpha-1}{2}\right)!}{\left(\dfrac{\alpha}{2}\right)!}.$$

5 Replace x by y^2 in the first formula of the preceding problem and then replace y by x to obtain the formula

$$\int_0^1 x^{2p+1}(1-x^2)^q\,dx = \frac{1}{2}\frac{p!q!}{(p+q+1)!}.$$

Tell why the limits of integration are correct.

6 When we proved the formulas for derivatives of trigonometric functions, we were unwilling to base the proofs upon the formula $A = \frac{1}{2}r^2\theta$ for the area of a sector of a circle having central angle θ and radius r. Supposing that $0 < \theta < \pi/2$, this problem requires that the formula be proved. With the aid of an appropriate figure, show that

$$A = \int_0^r f(x)\,dx$$

where

$$f(x) = x\tan\theta \qquad (0 \leqq x \leqq r\cos\theta)$$
$$f(x) = \sqrt{r^2 - x^2} \qquad (r\cos\theta \leqq x \leqq r).$$

Then find A with the aid of the formula (8.455) which has been proved.

7 It is easy to evaluate the integral

$$K_1 = \int x \sin x \, dx$$

with the aid of the formula

$$\int u(x)v'(x)\,dx = u(x)v(x) - \int v(x)u'(x)\,dx$$

for integration by parts. Do it by filling in the first and then the second row of the formulas

$$\begin{array}{ll} u(x) = & \qquad v'(x) = \\ u'(x) = & \qquad v(x) = \end{array}$$

in the most useful way. Check the result by differentiation.

8 Do not forget our guiding principle. If you have an idea other than integration by parts for evaluating the integral

$$K_2 = \int x^2 \sin x \, dx,$$

write the author a letter. Otherwise, develop a strong heart by solving the problem by integration by parts.

9 When we get good basic ideas we can understand and sometimes even originate modifications of them. If Mr. Watson asks us to evaluate the integral

$$L = \int \frac{1}{\sqrt{x+1}+1}\,dx,$$

we could eliminate the radical by setting $x = \tan^2 \theta$, but the result is unlovely. We can try to simplify matters by making a substitution $x = u(t)$ so cleverly devised that $\sqrt{x+1} = t$. The denominator will then be simply $t + 1$ and the other details may be of such a nature that we can find L. Show that

$$L = 2\sqrt{x+1} - 2\log(\sqrt{x+1}+1) + c.$$

Hint: The simple identity

$$\frac{t}{t+1} = \frac{t+1-1}{t+1} = 1 - \frac{1}{t+1}$$

enables us to integrate the first member in a hurry.

8.5 Integration by substituting $z = \tan\frac{\theta}{2}$

This section provides useful experience in changing variables by treating a particular substitution which is sometimes useful, but the section can be omitted without damaging understanding of the remainder of the book.

We begin by deriving formulas which enable us to use the substitution (or change of variable) $z = \tan\frac{\theta}{2}$ to convert integrals of quotients of polynomials in $\sin\theta$, $\cos\theta$, $\tan\theta$, $\cot\theta$, $\sec\theta$, and $\csc\theta$ into integrals of quotients of polynomials in z. We suppose that $-\pi < \theta < \pi$, so that $-\pi/2 < \theta/2 < \pi/2$, and define z in terms of θ by the formula

$$z = \tan\frac{\theta}{2}. \tag{8.51}$$

Differentiation gives

$$\frac{dz}{d\theta} = \frac{1}{2}\sec^2\frac{\theta}{2} = \frac{1}{2}\left(1 + \tan^2\frac{\theta}{2}\right) = \frac{1+z^2}{2},$$

and hence

$$d\theta = \frac{2}{1+z^2}\,dz. \tag{8.52}$$

Basic trigonometric identities and (8.51) give

$$\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} = \frac{2\tan\frac{\theta}{2}}{\sec^2\frac{\theta}{2}} = \frac{2z}{1+z^2}, \tag{8.531}$$

$$\begin{aligned}\cos\theta &= \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2} = 2\cos^2\frac{\theta}{2} - 1 \\ &= \frac{2}{\sec^2\frac{\theta}{2}} - 1 = \frac{2}{1+z^2} - 1 = \frac{1-z^2}{1+z^2}\end{aligned} \tag{8.532}$$

and, except when θ is $\pi/2$ or $-\pi/2$,

$$\tan\theta = \frac{2\tan\frac{\theta}{2}}{1-\tan^2\frac{\theta}{2}} = \frac{2z}{1-z^2}. \tag{8.533}$$

In fact, Figures 8.541 and 8.542 enable us to write the six trigonometric functions of $\theta/2$ and θ in terms of z. Thus each of $\sin\theta$, $\cos\theta$, $\tan\theta$,

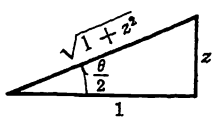


Figure 8.541

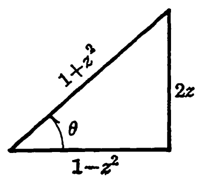


Figure 8.542

$\cot\theta$, $\sec\theta$, $\csc\theta$, and $d\theta/dz$ is a quotient of polynomials in z. It follows from this that if P and Q are polynomials in the six trigonometric functions of x, then

$$\int \frac{P}{Q}\,d\theta = \int \frac{P_1}{Q_1}\,dz,$$

where P_1 and Q_1 are polynomials in z. Several examples appear in the problems at the end of this section.

Except in special cases, evaluating integrals of the form

$$\int \frac{a_0z^m + a_1z^{m-1} + \cdots + a_{m-1}z + a_m}{b_0z^n + b_1z^{n-1} + \cdots + b_{n-1}z + b_n}\, dz$$

is very tedious business. When $n \leqq 2$, and in a few special additional cases, answers can be found in tables of integrals. Some information about the matter will appear in Section 9.4 of this book.

Problems 8.59

1 By use of the substitution $z = \tan \frac{\theta}{2}$, show that

(*a*) $\displaystyle\int \frac{1}{\sin \theta}\, d\theta = \int \frac{1}{z}\, dz = \log |z| + c = \log \left| \tan \frac{\theta}{2} \right| + c$

(*b*) $\displaystyle\int \frac{1}{\cos \theta}\, d\theta = \int \frac{2}{1 - z^2}\, dz = \int \frac{1 - z + 1 + z}{(1 - z)(1 + z)}\, dz$

$$= \int \left(\frac{1}{1 + z} + \frac{1}{1 - z} \right) dz = \log \left| \frac{1 + \tan \frac{\theta}{2}}{1 - \tan \frac{\theta}{2}} \right| + c$$

(*c*) $\displaystyle\int \frac{1}{a + b \sin \theta}\, d\theta = \int \frac{2}{az^2 + 2bz + a}\, dz$

(*d*) $\displaystyle\int \frac{1}{a + b \sin \theta + c \cos \theta}\, d\theta = \int \frac{2}{(a - c)z^2 + 2bz + (a + c)}\, dz$

(*e*) $\displaystyle\int \frac{\sin \theta + \cos \theta}{\tan \theta + \cot \theta}\, d\theta = 4 \int \frac{z(1 - z^2)(1 + 2z - z^2)}{(1 + z^2)^4}\, dz$

(*f*) $\displaystyle\frac{2z}{1 + z^2} + c = \sin \theta + c = \int \cos \theta\, d\theta = 2 \int \frac{1 - z^2}{(1 + z^2)^2}\, dz$

(*g*) $\displaystyle\log \frac{1 + z^2}{1 - z^2} + c = \log \sec \theta + c = \int \tan \theta\, d\theta = 4 \int \frac{z}{1 - z^4}\, dz$

(*h*) $\displaystyle\log \left(a + \frac{2bz}{1 + z^2} \right) + c = \log (a + b \sin \theta) + c$

$$= \int \frac{b \cos \theta}{a + b \sin \theta}\, d\theta$$

$$= 2b \int \frac{1 - z^2}{(az^2 + 2bz + a)(1 + z^2)}\, dz$$

(*i*) $\displaystyle\int \frac{\sin \theta}{1 + \sin \theta}\, d\theta = \int \frac{4z}{(1 + z^2)(1 + z)^2}\, dz$

(*j*) $\displaystyle\int \frac{\cos \theta}{1 + \cos \theta}\, d\theta = \int \frac{1 - z^2}{1 + z^2}\, dz$

$$= \int \frac{2 - 1 - z^2}{1 + z^2}\, dz = 2 \tan^{-1} \left(\tan \frac{\theta}{2} \right) - \tan \frac{\theta}{2} + c$$

9 Exponential and logarithmic functions

9.1 Exponentials and logarithms At least a modicum of basic information about exponentials is possessed by everyone who looks at Chapter 9 of a calculus book. Everyone knows that $3^4 = 3 \cdot 3 \cdot 3 \cdot 3$ and $3^{1/2} = \sqrt{3}$. Such formulas as

$$a^{5/2} = \sqrt{a^5} = (\sqrt{a})^5,$$

in which it is supposed that $a > 0$, are familiar. Remark 20 which appears among the problems at the end of this section provides basic theory of a^x for the case in which $a > 0$ and the exponent x is a rational number, that is, $x = P/Q$, where P and Q are integers for which $Q \neq 0$. There is a reason why this theory is often neglected. It is too difficult for elementary algebra books because it requires use of the intermediate value theorem and hence requires knowledge of completeness of the real-number system. It is too simple for calculus books because it is mostly elementary algebra. The mathematical theory of exponents, like the physical theory of hydrogen, is neither so brief nor unimportant that it is unworthy of several hours of study.

We start here with a given number a, the *base* of our exponential function a^x, for which $a > 1$. We suppose that a^x has been defined for rational numbers x and that, if x and y are rational numbers for which $x = P/Q$, then

$$a^x a^y = a^{x+y}, \qquad (a^x)^y = a^{xy}, \qquad a^x = (a^P)^{1/Q} = (a^{1/Q})^P. \tag{9.11}$$

For the purpose of learning about mathematics in general and about the exponential function a^x in particular, we make a direct frontal attack upon the problem of learning more about a^x when x is rational. We then use this information as a basis for definition and study of a^x when x is real. Our first step is to calculate a^x for several values of x and to start making a graph of $y = a^x$ as in Figure 9.12. We mark only the points whose coordinates we have calculated and overcome the primitive urge to draw a "smooth curve" through these points. We have experimental evidence that $0 < a^x < a^{x+h}$ when x and h are rational and $h > 0$, and this matter should be investigated. When $x = P/Q$, $a^{1/Q}$ is the positive number r for which $r^Q = a$ and $a^x = (a^{1/Q})^P = r^P > 0$, so $a^x > 0$. It turns out that we can obtain a huge amount of information from the identity

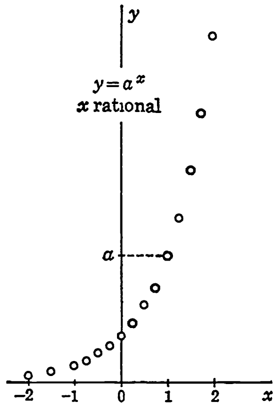


Figure 9.12

$$a^{x+h} - a^x = (a^h - 1)a^x. \tag{9.13}$$

For a modest beginning we suppose that $h = p/q$, where p and q are positive integers, and show that $a^h > 1$. When $f(x) = x^q$, we see that $f(1) = 1$ and that $f(x)$ is continuous and increasing when $x > 0$. Since $a > 1$, the one and only number x_1 for which $f(x_1) = a$ must exceed 1. Thus $x_1^q = a$ and $a^{1/q} = x_1$, where $x_1 > 1$, and hence $a^h = x_1^p > 1$. Thus, when $h > 0$, the right member of (9.13) is the product of two positive numbers and hence is positive. This implies the following theorem.

Theorem 9.14 *If x and h are rational and $h > 0$, then*

$$0 < a^x < a^{x+h}. \tag{9.141}$$

This means that, as a function of the "rational variable" x, the function having values a^x is positive and increasing. Proof of this fundamental fact represents a triumph of mind over matter, but one triumph is not enough and we must continue. Tables and slide rules and brute-force calculations with pencils produce experimental evidence in support of the following theorem.

Theorem 9.15 *As a function of the "rational variable" x, the function having values a^x is continuous over $-\infty < x < \infty$.*

We shall prove this theorem with the aid of (9.13) and an estimate of the troublesome factor $(a^h - 1)$ that is, from our present point of view and from some others, quite amazing. Let h be a positive rational number for which $h \leqq 1$ and let n be the greatest integer for which $nh \leqq 1$. If we wish, we can draw Figure 9.151 and put a part of the graph of $y = a^x$ over it to help us see what we are doing. The equality

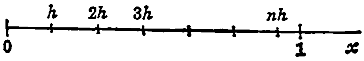


Figure 9.151

$$(9.152) \qquad (a^h - 1)a^{(k-1)h} = a^{kh} - a^{(k-1)h}$$

holds when $k \doteq 1, 2, 3, \cdots, n$ and summing over these values of k gives the formula

$$\begin{aligned}(9.153) \quad (a^h - 1)\Big[\sum_{k=1}^{n} a^{(k-1)h}\Big] &= \sum_{k=1}^{n} (a^{kh} - a^{(k-1)h}) \\ &= (a^h - a^0) + (a^{2h} - a^h) + (a^{3h} - a^{2h}) + \cdots + (a^{nh} - a^{(n-1)h}) \\ &= a^{nh} - a^0 = a^{nh} - 1 \leqq a - 1\end{aligned}$$

which has, in its middle line, a sum which is called a *telescopic sum* because it "telescopes" to $a^{nh} - a^0$. More critical examination of the sum in brackets will appear later. Meanwhile, we observe that if we denote the sum by S, then each term in the sum is 1 or more, so $S \geqq n$. But $nh > \frac{1}{2}$, so $n > 1/2h$ and $S > 1/2h$. Replacing S in (9.153) by the smaller number $1/2h$ gives the inequality

$$(a^h - 1)\frac{1}{2h} < a - 1.$$

Therefore,

$$(9.154) \qquad a^h - 1 < 2(a - 1)h \qquad (0 < h \leqq 1).$$

This and (9.13) show that the formula

$$(9.155) \qquad a^{x+h} - a^x \leqq 2(a - 1)ha^x$$

is valid when x and h are rational and $0 < h \leqq 1$. This result and the fact that a^x is increasing give the following theorem.

Theorem 9.16 *If r_1 and r_2 are rational numbers for which*

$$|r_2 - r_1| = h \leqq 1$$

and if M is a rational number for which $r_1 \leqq M$ and $r_2 \leqq M$, then

$$|a^{r_2} - a^{r_1}| \leqq 2(a - 1)|r_2 - r_1|a^M.$$

This is a stronger version of Theorem 9.15 which implies Theorem 9.15. Our information about the values of a^r for rational numbers r is now very substantial, and we proceed to use it to define and study a^x when x is real. Let x be a real number which is not necessarily rational. In case x is rational, the fact that a^r is an increasing function of the rational variable r implies that

$$a^x = \operatorname*{l.u.b.}_{r \leqq x} a^r, \tag{9.17}$$

where the right side is the least upper bound of the set of numbers a^r for which r is rational and $r \leqq x$. In case x is irrational, *we define the number a^x by this same formula* (9.17). To use this definition, we need only a very little information about least upper bounds of sets of numbers. We need only the elementary fact that if each number in a set S_1 is also a number in a set S_2, then the least upper bound of the first set must be less than or equal to the least upper bound of the second set. Suppose x_1 and x_2 are given real numbers for which $x_1 < x_2$ and we choose two rational numbers r_1 and r_2 such that

$$x_1 < r_1 < r_2 < x_2.$$

Then, since $a^{r_1} < a^{r_2}$,

$$a^{x_1} = \operatorname*{l.u.b.}_{r \leqq x_1} a^r \leqq \operatorname*{l.u.b.}_{r \leqq r_1} a^r = a^{r_1} < a^{r_2} = \operatorname*{l.u.b.}_{r \leqq r_2} a^r \leqq \operatorname*{l.u.b.}_{r \leqq x_2} a^r = a^{x_2},$$

and this proves that a^x is an increasing function of x. If $0 < \delta < M$, we can choose a positive integer n such that $a^n > M$ and $a^{-n} < \delta$. This and the fact that a^x is positive and increasing imply that $a^x \to 0$ as $x \to -\infty$ and $a^x \to \infty$ as $x \to \infty$. To prove that a^x is continuous (and in fact uniformly continuous) over the interval $x < M$, we can employ Theorem 9.16. Let $\epsilon > 0$. We can then choose a $\delta > 0$ such that $|a^{r_2} - a^{r_1}| < \epsilon$ whenever $r_1 < M$, $r_2 < M$, and $|r_2 - r_1| < \delta$. Whenever x_1 and x_2 are real numbers for which $x_1 < M$, $x_2 < M$, and $|x_2 - x_1| < \delta$, we can sandwich x_1 and x_2 between rational numbers r_1 and r_2 for which $r_1 < M$, $r_2 < M$, and $|r_2 - r_1| < \delta$. The fact that a^x is increasing then implies that $|a^{x_2} - a^{x_1}| < |a^{r_1} - a^{r_2}| < \epsilon$.

To complete our study of the basic theory of a^x when x is real, we must prove the formulas

$$a^x a^y = a^{x+y}, \qquad (a^x)^y = a^{xy}. \tag{9.18}$$

Since a^x is continuous, we can let $r_1, r_2, \cdots$ and $s_1, s_2, \cdots$ be sequences of rational numbers converging to x and y so that

$$\begin{aligned} a^x a^y &= (\lim_{n\to\infty} a^{r_n})(\lim_{n\to\infty} a^{s_n}) = \lim_{n\to\infty} (a^{r_n} a^{s_n}) \\ &= \lim_{n\to\infty} a^{r_n+s_n} = a^{x+y} \end{aligned}$$

and the first formula is proved. To prove the second formula, we vary the procedure by letting r, s, t, and u be rational numbers for which $r \leqq x \leqq s$ and $t \leqq y \leqq u$. Then

$$a^r \leqq a^x \leqq a^s \qquad \text{and} \qquad (a^r)^t \leqq (a^x)^y \leqq (a^s)^u,$$

so

$$a^{rt} \leqq (a^x)^y \leqq a^{su}.$$

Taking limits as $rt \to xy$ and $su \to xy$ gives $a^{xy} \leqq (a^x)^y \leqq a^{xy}$ and shows that $(a^x)^y = a^{xy}$.

Our basic theory of exponents enables us to give the basic theory of logarithms in a few lines. With the aid of the intermediate-value theorem, we see that to each positive number x there corresponds exactly one number y such that $a^y = x$. This number y, an exponent, is called *the logarithm with base a of x* and is denoted by $\log_a x$, so that the formulas

$$(9.181) \qquad a^y = x, \qquad y = \log_a x, \qquad a^{\log_a x} = x$$

are equivalent. The logarithmic function is the inverse of the expo-

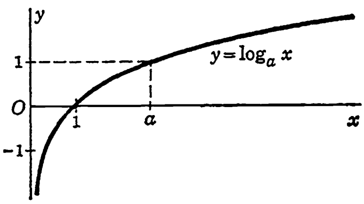


Figure 9.182

nential function, and its graph is shown in Figure 9.182. The fundamental formulas

$$(9.183) \qquad \log_a xy = \log_a x + \log_a y, \qquad \log_a x^y = y \log_a x$$

are proved by setting $u = \log_a x$, $v = \log_a y$, $x = a^u$, and $y = a^v$, so that

$$xy = a^u a^v = a^{u+v}$$

and

$$\log_a xy = u + v = \log_a x + \log_a y.$$

Moreover, $x^y = (a^u)^y = a^{uy}$, so

$$\log_a x^y = uy = y \log_a x.$$

When a, b, and x are positive numbers for which $a > 1$ and $b > 1$, we can equate the logarithms with base b of the members of the identity $x = a^{\log_a x}$ to obtain the first of the formulas

$$(9.184) \qquad \log_b x = (\log_a x)(\log_b a), \qquad 1 = (\log_a b)(\log_b a),$$

and putting $x = b$ in the first gives the second. Putting $b = e$ and $a = 10$ gives the formula

$$\log x = (\log_{10} x)(\log 10) = \frac{\log_{10} x}{\log_{10} e} \tag{9.185}$$

which, together with the estimates

$$\log 10 = 2.30258\ 50929\ 94045\ 68402 \tag{9.186}$$
$$\log_{10} e = 0.43429\ 44819\ 03251\ 82765, \tag{9.187}$$

enables us to find $\log x$ with the aid of a table of values of $\log_{10} x$. These formulas are shunned when a log-log slide rule is available and gives satisfactory accuracy and when a satisfactory table of values of $\log x$ is available. When decimal representations of numbers are used in calculations, it is often necessary to know that each positive number y is representable in the form

$$y = 10^n x, \tag{9.188}$$

where n is an integer and $1 \leqq x < 10$, and that

$$\log_a y = n \log_a 10 + \log_a x. \tag{9.189}$$

This formula enables us to find $\log_a y$ with the aid of a table giving values of $\log_a x$ when $1 \leqq x \leqq 10$. The formula works whenever $a > 1$. It is simplest when $a = 10$ and $\log_a 10 = 1$; in this case $n \log_a 10$ is an integer (the *characteristic* of the logarithm) and rules for finding it are sometimes peddled without revelation of the fact that they are identical with the rules for finding the exponents in the representations

$$\begin{aligned} 416.3 &= 10^2\, 4.16 \\ 0.00004163 &= 10^{-5}\, 4.16 \\ 4.163 &= 10^0\, 4.163. \end{aligned}$$

Our derivations of formulas for derivatives and integrals of logarithms and exponentials will come in the next section. Meanwhile, we close this introductory section with some historical remarks. The first published table of logarithms, the "Mirifici Logarithmorum Canonis · · · ," by John Napier (1550–1617), appeared in 1614. The rare-book collection of the University of Illinois contains this book and an astonishing collection of old tables. To indicate that these books and their titles are nontrivial, we cite the full title of the 1631 edition of the book of Napier.†

† Napier, John, "Logarithmicall Arithmetike, or tables of logarithmes for absolute numbers from an unite to 100000; as also for sines, tangentes and secantes for every minute of a quadrant: with a plaine description of their use in arithmetike, geometrie, geographie, astronomie, navigation, etc. These numbers were first invented by the most excellent Iohn Neper, Baron of Marchiston, and the same were transformed, and the foundation and use of them illustrated with his approbation by Henry Briggs, Sir Henry Savils Professor of geometrie in the Universite of Oxford. The uses whereof were written in Latin by the author himselfe, and since his death published in English by diverse of his friends according to his mind, for the benefit of such as understand not the Latin tongue," London, 1631, 819 pp.

For more than 300 years, logarithms with base 10 were systematically and extensively used to make arithmetical calculations. Respectable scientists of the present and future must know about them, but substantially all of the chores formerly done with the aid of tables of $\log_{10} x$ are now done with slide rules and mechanical and electronic calculators and computers of assorted shapes and sizes. It seems that the first published table of logarithms with base e appeared in a 1618 edition of the Napier tables, and that John Speidell used e as a base of exponentials in a book published in 1620. Much of the present usefulness of e is based upon work of Euler (1707–1783). Except for collectors of rare books, tables of e^x, e^{-x}, and $\log x$ are now vastly more valuable than tables of $\log_{10} x$.

Problems 9.19

1 Find the values of a, x, and y that satisfy the equations

(*a*) $2^x = 32,\ a^5 = 32,\ 2^5 = y$
(*b*) $x = \log_2 32,\ 5 = \log_a 32,\ 5 = \log_2 y$
(*c*) $7^4 = y,\ 7^x = 2401,\ a^4 = 2401$
(*d*) $4 = \log_7 y,\ x = \log_7 2401,\ 4 = \log_a 2401$
(*e*) $a^3 = 1000,\ 10^3 = y,\ 10^x = 1000$
(*f*) $3 = \log_a 1000,\ 3 = \log_{10} y,\ x = \log_{10} 1000$

2 Practice the art of starting with the first of the equations

$$y = a^x, \qquad \log y = x \log a = (\log a)x, \qquad y = e^{kx},$$

taking logarithms (with base e) to obtain the second equation, and then using the definition of logarithms to obtain the third equation where $k = \log a$.

3 Practice the art of starting with the formula $y = a^x$ and writing

$$y = a^x = (e^k)^x = e^{kx},$$

where k is the exponent which we must put upon e to get a and hence $k = \log a$.

4 Using the method of Problem 2 or the method of Problem 3, show that

(*a*) $x^x = e^{x \log x} \qquad (x > 0)$
(*b*) $(1 + x)^{1/x} = e^{\frac{1}{x}\log(1+x)} = e^{\frac{\log(1+x) - \log 1}{x}} \qquad (x > -1)$
(*c*) $f(x)^{g(x)} = e^{g(x) \log f(x)} \qquad (f(x) > 0)$

5 Show that $a^{k_1 x} = e^{k_2 x}$ when $k_2 = k_1 \log a$.

6 Sketch graphs of the equations $y = 2^x$ and $y = \log_2 x$ on the same sheet of graph paper. Tell why the line having the equation $y = x$ is (or is not) a line of symmetry of the set consisting of the two graphs. Sketch a line which appears to be tangent to the graph of $y = 2^x$ at the point (0,1), estimate the coordinates of the point where this line meets the line having the equation $y = x$, and use the results to obtain an estimate of the slope of the tangent. Finally, modify the procedure to obtain an estimate of the slope of the line tangent to the graph of $y = \log_2 x$ at the point (1,0). *Remark:* This is not a dull problem, because moderately careful work produces very good results. To check one of the results, let $f(x) = 2^x$ so $f(x) = e^{x \log 2}$ so $f'(x) = e^{x \log 2}(\log 2)$ and $f'(0) = \log 2 = 0.693 \cdots$.

7 Apply the procedure of Problem 6 to the equations $y = 3^x$ and $y = \log_3 x$. Try to find a way to use a table or slide rule to check the numerical results.

8 Many persons with scientific interests have (and should have) log-log slide rules and should know or learn how to set them to obtain values of a^x when $2 \leqq a \leqq 3$ and $-1 \leqq x \leqq 1$. These persons can quickly produce a good graph of $y = (2.5)^x$ over the interval $-1 \leqq x \leqq 1$ and use it to obtain a good estimate of the slope at the point (0,1). Repeating the process for $y = (2.75)^x$ and using ideas involving interpolation lead to quite good approximations to a base e for which the graph of $y = e^x$ has slope 1 at the point (0,1). This base e is the famous e.

9 Tell why $\int_0^1 2^x\,dx$ and $\int_0^1 3^x\,dx$ exist. Then sketch appropriate graphs and use them to obtain rough estimates of the values of these integrals. Finally, check the estimates with the aid of the formula

$$\begin{aligned}\int_0^1 a^x\,dx &= \int_0^1 e^{x \log a}\,dx = \frac{1}{\log a}\,e^{x \log a}\Big]_0^1 \\ &= \frac{1}{\log a}\,(e^{\log a} - 1) = \frac{a-1}{\log a}.\end{aligned}$$

10 Tell why $\int_0^1 100^x\,dx$ could not be 0.05 and could not be 500.

11 Discover reasons why we could suspect that

$$\int_0^x a^t\,dt < \frac{1 + a^x}{2}$$

when $a > 1$ and $x > 0$.

12 Sketch graphs of the equations

(*a*) $y = 2^{-1/x^2}$ (*b*) $y = 2^{-1/x}$

Remark: More detailed information about the graph of the equation $y = e^{-1/x^2}$ will appear later.

13 In Problem 4 of Section 12.6, we shall discover that if $z > 0$, then

$$\log z! = \tfrac{1}{2}\log 2\pi + (z + \tfrac{1}{2})\log z - z + E(z), \tag{1}$$

where $E(z)$ is a number for which

$$\frac{1}{12z} - \frac{1}{360z^3} < E(z) < \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1260z^5}. \tag{2}$$

Show that (1) can be put in the form

$$\log z! = \log\sqrt{2z\pi} + \log z^z - z + E(z) \tag{3}$$

and hence that

$$z! = \sqrt{2z\pi}\, z^z e^{-z} e^{E(z)}. \tag{4}$$

Remark: The formulas (1) and (3) and (4) are examples of *Stirling formulas* for $\log z!$ and $z!$.

14 This problem requires that we capture the ideas used to obtain the number

$$y = e^{0.41631\ 690216}$$

correct to 10 decimal places, it being assumed that a table giving $e^{0.4163}$ correct to 15 places and a desk calculator are available. The first step is to recognize that $y = AB$, where $A = e^x$, $B = e^h$, $x = 0.4163$, and $h = 0.00001\ 690216$. The number A is copied from the table correct to 12 or 15 places. The number B is readily calculated to 12 or 15 places with the aid of the formula

$$e^h = 1 + h + \frac{h^2}{2} + \frac{h^3}{6} + \frac{h^4}{24} + \cdots$$

which we shall learn about in a problem at the end of the next section. After a few buttons have been pressed, the calculator will give the product $y = AB$ in a hurry.

15 This problem requires that we capture the ideas used to obtain the number

$$z = \log 4.16316\ 90216$$

correct to 10 decimal places, it being assumed that a table giving $\log 4.1631$ correct to 15 places and a desk calculator are available. The first step is to recognize that

$$z = \log 4.1631 + \log \frac{4.16316\ 90216}{4.1631}$$

and hence that $z = A + B$, where $A = \log x$, $B = \log (1 + h)$, $x = 4.1613$, and

$$h = \frac{0.00006\ 90216}{4.1631}.$$

The number A is copied from the table correct to 12 or 15 places. The number B is readily calculated to 12 or 15 places with the aid of the formula

$$\log (1 + h) = h - \frac{h^2}{2} + \frac{h^3}{3} - \frac{h^4}{4} + \cdots .$$

The sum $A + B$ can be obtained with a pencil, but it is safer to use the calculator.

16 Supposing that a desk calculator is available to perform additions, subtractions, multiplications, and divisions, describe the steps by which the numbers

(*a*) 2^π, (*b*) π^π, (*c*) e^π, (*d*) π^e, (*e*) $\sqrt[7]{\pi}$

can be calculated correct to 12 decimal places. *Hint:* Take logarithms with base e so that tables of values of e^x and $\log x$ can be used. *Partial ans.:* Let $w = \pi^\pi$ and obtain the formula $\log w = C$, where $C = AB$, $A = \pi$, and $B = \log \pi$. Find a table giving π correct to 15D (that is, 15 decimal places). Then use the method of Problem 15 to find B correct to 15D. Then multiply to find C. Then $w = e^C$. Let $C = n + x$, where n is an integer and $0 \leqq x < 1$, so $w = e^n e^x$. Find e^n by multiplication or from a table. Find e^x by the method of Problem 14. Finally, multiply to get w.

17 Persons who start fires with matches can be interested in the fact that fires can be started by rubbing sticks together. We shall soon learn modern ways of calculating approximations to exponentials and logarithms, but we can be interested in seeing that more primitive methods will work. Supposing that a and x are positive numbers given in decimal form, and that $0 < x < 1$, we outline an old, old procedure for finding decimal approximations to a^x that was used when all computations were made by hand and when letters $A, B, C, \cdots$ of the alphabet were used in place of $x_0, x_1, x_2, \cdots$. To get started, let $y_0 = 0$, $z_0 = 1$, and observe that the inequalities

$$(1) \qquad y_n \leqq x \leqq z_n, \qquad a^{y_n} \leqq a^x \leqq a^{z_n}$$

hold when $n = 0$. Supposing that $n = 1$, calculate x_1 and then a^{x_1} from the formulas

$$(2) \qquad x_n = \frac{y_{n-1} + z_{n-1}}{2}, \qquad a^{x_n} = \sqrt{a^{y_{n-1}}a^{z_{n-1}}},$$

and let y_n and z_n be two of the three numbers y_{n-1}, x_n, z_{n-1} chosen in such a way that the inequalities (1) hold when $n = 1$. Repeating the process with $n = 2$ gives (2) and then (1) when $n = 2$, and the process can be continued. Since $z_n = y_n + 1/2^n$, we can be sure that our estimates of a^x will be good when n is large. To find

$$(\sqrt{2} + \sqrt{3})^{\pi-3}$$

correct to 32 decimal places by this method (or by any other method) would be quite a chore for an inexperienced person, but persons who run modern computers can enjoy making such computations. Primitive methods often work better than fancier methods of limited applicability.

18 Persons possessing slide rules may wish to study them while reading something that tells why they work. The C and D scales contain numbers from 1 to 10 and, when it is supposed that these scales have unit length (this unit is usually about 10 inches), the number x lies $\log_{10} x$ units from the left end. To find the product X of two positive numbers A and B, we put them in the forms $A = 10^m a$ and $B = 10^n b$ where m and n are integers and $1 \leqq a < 10, 1 \leqq b < 10$ and use the fact that $X = 10^{m+n}x$ where $x = ab$. Noting that

$$(1) \qquad \log_{10} x = \log_{10} a + \log_{10} b,$$

we run along the D scale a distance $\log_{10} a$ to the number a and then, after setting 1 on the C scale opposite a, run along the C scale a distance $\log_{10} b$ to the number b. We have then gone the distance $\log_{10} x$ from the left end of the D scale and hence can read x on the D scale.

19 Persons possessing log-log slide rules may wish to see why and how the esoteric LL scales are made. Suppose we wish to find the number y for which $y = b^q$ when b and q are given. We note that

$$(1) \qquad \log_e y = q \log_e b$$

and

$$(2) \qquad \log_{10} \log_e y = \log_{10} \log_e b + \log_{10} q.$$

The $LL3$ scale contains numbers from e to about 22,000. The distance from the left end of the $LL3$ scale to a number y is $\log_{10}\log_e y$. If we run a distance $\log_{10}\log_e b$ on the $LL3$ scale to hit b and then, with the aid of the C scale, run an additional distance $\log_{10} q$, we go a total distance $\log_{10}\log_e y$ and hence can read y on the $LL3$ scale. We can be comforted by discovery that log-log slide rules give the simple formulas $3^4 = 81$ and $4^{2.5} = 32$ as well as others that are not so easily verified. It is often particularly useful to know that if points b and f lie opposite each other on the $LL3$ and D scales, then

$$(3) \qquad \log_{10}\log_e b = \log_{10} f$$

so $\log_e b = f$ and $b = e^f$. Thus natural exponentials and logarithms are easily read, and the good approximations $e^3 = 20$, $\log 20 = 3$ are always available to show us which way to read the scales. With the aid of this and similar information about other scales, it is possible to both understand and use slide rules.

20 This is the promised development of the theory of a^x for the case in which x is rational. Let a be a positive number. It is the *base* of the exponential function a^x which we are about to define when x is an integer, then when x is the reciprocal of a nonzero integer, and then when x is a rational number which is not necessarily an integer. If n is a nonnegative integer, we define a^n and a^{-n} by the formulas

$$(1) \qquad a^n = 1\cdot a\cdot a\cdot a\cdot \cdots a, \qquad a^{-n} = \frac{1}{1\cdot a\cdot a \cdots a},$$

where in each case 1 is multiplied by a exactly n times. If $n = 0$, no multiplications are involved and $a^0 = 1$. Clearly, $a^1 = a$, $a^{-1} = 1/a$, $a^2 = a\cdot a$, $a^{-2} = 1/a\cdot a$, etcetera, and $a^{-n} = 1/a^n$ or $a^{-n}a^n = a^0 = 1$. Counting numbers of times by which 1 is multiplied or divided by a enables us to show that the *laws of exponents*

$$(2) \qquad a^x a^y = a^{x+y}, \qquad (a^x)^y = a^{xy}$$

hold whenever x and y are nonnegative integers and then whenever x and y are integers. Let n be a nonzero integer and let $f(x) = x^n$ when $x > 0$. Then $f'(x) = nx^{n-1}$, so f is continuous and increasing when $n > 0$ and is continuous and decreasing when $n < 0$. Moreover, $f(x) \to 0$ as $x \to 0$ and $f(x) \to \infty$ as $x \to \infty$ when $n > 0$, and $f(x) \to \infty$ as $x \to 0$ and $f(x) \to 0$ as $x \to \infty$ when $n < 0$. In each case these facts and the intermediate-value theorem 5.48 imply that there is exactly one positive number h for which $h^n = a$. We then define $a^{1/n}$ by the first of the formulas

$$(3) \qquad a^{1/n} = h, \qquad a = h^n, \qquad (a^{1/n})^n = a, \qquad (a^n)^{1/n} = a,$$

remember that the first formula is equivalent to the second, and observe that the last two formulas are correct. Supposing that λ, p, and q are integers for which $\lambda q \neq 0$, we acquire ability to manipulate these things by proving the formula

$$(4) \qquad (a^{\lambda p})^{1/\lambda q} = (a^p)^{1/q} = (a^{1/q})^p = (a^{1/\lambda q})^{\lambda p}.$$

To prove this, let H be defined by $a^{1/\lambda q} = H$, so that $a = H^{\lambda q} = (H^\lambda)^q$ and $a^{1/q} = H^\lambda$. Thus (4) will be true if

$$(5) \qquad ((H^{\lambda q})^{\lambda p})^{1/\lambda q} = ((H^{\lambda q})^p)^{1/q} = (H^\lambda)^p = H^{\lambda p}.$$

But $(H^{\lambda q})^{\lambda p} = (H^{\lambda p})^{\lambda q}$ and $(H^{\lambda q})^{p} = (H^{\lambda p})^{q}$, and it follows that each member of (5) is $H^{\lambda p}$. Thus (5) is true and this proves (4). Just as the rational number r defined by $r = -\frac{63}{28}$ is representable in the forms

$$r = \frac{63}{-28} = \frac{(7)(-9)}{(7)(4)} = \frac{-9}{4},$$

so also each rational number r given in the form $r = P/Q$, where P and Q are integers, is uniquely representable in the form p/q, where $P = \lambda p$, $Q = \lambda q$, and λ, p, and q are integers for which $q > 0$ and the integers $|p|$ and q are relatively prime, that is, have no common positive integer factor different from 1. This fact and (4) show that when P and Q are integers for which $Q \neq 0$, we can define $a^{P/Q}$ by the first of the formulas

$$a^{P/Q} = (a^P)^{1/Q} = (a^{1/Q})^P \tag{6}$$

with assurance that the whole formula is correct and that the formula remains correct when we multiply or divide both P and Q by the same integer λ provided $\lambda \neq 0$ when we divide. This completes the definition of a^x when x is rational. To prove that the laws (2) hold when x and y are rational, we can let $x = p/n$ and $y = q/n$, where p, q, n are integers and $n > 0$, to obtain

$$a^x a^y = (a^{p/n})(a^{q/n}) = (a^{1/n})^p (a^{1/n})^q = (a^{1/n})^{p+q} = a^{(p+q)/n} = a^{x+y}$$

and

$$((a^x)^y)^{n^2} = (((a^{p/n})^q)^{1/n})^{n^2} = (((a^p)^{1/n})^q)^n = (((a^p)^q)^{1/n})^n = a^{pq}$$

so

$$(a^x)^y = (a^{pq})^{1/n^2} = a^{pq/n^2} = a^{xy}.$$

9.2 *Derivatives and integrals of exponentials and logarithms*

Let $a > 1$. Looking forward to a derivation of a formula for the derivative of the function f for which $f(x) = a^x$, we use h instead of Δx, write

$$a^{x+h} - a^x = (a^h - 1)a^x, \tag{9.21}$$

and observe that

$$\frac{d}{dx} a^x = \lim_{h \to 0} \frac{a^{x+h} - a^x}{h} = \left[\lim_{h \to 0} \frac{a^h - 1}{h}\right] a^x \tag{9.22}$$

provided the limits exist. If we succeed in proving that, for some constant A which may depend upon a, the first of the formulas

$$\lim_{h \to 0+} \frac{a^h - 1}{h} = A, \qquad \lim_{h \to 0-} \frac{a^h - 1}{h} = A \tag{9.221}$$

is valid, then the second will also be valid because

$$\lim_{h \to 0-} \frac{a^h - 1}{h} = \lim_{h \to 0+} \frac{a^{-h} - 1}{-h} = \lim_{h \to 0+} \frac{1 - a^{-h}}{h} = \lim_{h \to 0+} a^{-h} \frac{a^h - 1}{h} = A$$

and we will be able to conclude that

$$\frac{d}{dx} a^x = Aa^x. \tag{9.222}$$

To investigate the factor $a^h - 1$, we employ the method which we used to prove Theorem 9.15 when we knew practically nothing about exponentials. Let h be a number, now not necessarily rational, for which $0 < h < 1$ and let n be the greatest integer for which $nh \leqq 1$. The equality

$$(a^h - 1)a^{(k-1)h} = a^{kh} - a^{(k-1)h} \tag{9.23}$$

still holds when $k = 1, 2, 3, \cdots, n$, and summing over these values of k gives, as in (9.153),

$$(a^h - 1)\left[\sum_{k=1}^{n} a^{(k-1)h}\right] = a^{nh} - 1 \tag{9.231}$$

and hence

$$\frac{a^h - 1}{h} = \frac{a^{nh} - 1}{\sum_{k=1}^{n} a^{(k-1)h}h}. \tag{9.232}$$

The numerator of the right side converges to $a - 1$ as $h \to 0$. If to the denominator of the right side we add the negligible term $a^{nh}(1 - nh)$, the new sum will be a Riemann sum, with partition points

$$0 < h < 2h < 3h < \cdots < nh < 1$$

in the interval $0 \leqq x \leqq 1$, which converges to the positive number $\int_0^1 a^x\,dx$ as h, the norm of the partition, approaches 0. Because it is sometimes extremely helpful to be able to recognize Riemann sums when they appear in somewhat disguised forms, we give careful attention to the details. If we set $f(x) = a^x$, set $x_k = kh$ when $0 \leqq k \leqq n$, set $x_{n+1} = 1$, *set* $x_k^* = x_k$ when $1 \leqq k \leqq n + 1$, set $\Delta x_k = h$ when $1 \leqq k \leqq n$, and set $\Delta x_{n+1} = 1 - nh$, then the new sum becomes the Riemann sum

$$\sum_{k=1}^{n+1} f(x_k^*)\,\Delta x_k \tag{9.233}$$

and as h, the norm of the partition, approaches 0 the Riemann sum approaches $\int_0^1 f(x)\,dx$, that is $\int_0^1 a^x\,dx$. Therefore,

$$\lim_{h\to 0}\frac{a^h - 1}{h} = \frac{a - 1}{\int_0^1 a^x\,dx} = A, \tag{9.24}$$

where A is the constant defined by the last equality. It is possible to squeeze information from this formula, but we obtain a formula giving a simpler relation between a and A before actually showing that there is one and only one number a for which $A = 1$.

We have obtained the first of the formulas

$$\frac{d}{dx}a^x = Aa^x, \qquad \frac{d}{dx}\log_a x = \frac{1}{Ax}. \tag{9.25}$$

Since a^x has a positive derivative, Theorem 8.33 implies that its inverse $\log_a x$, defined over the infinite interval $x > 0$, is differentiable. Hence we can differentiate the first of the formulas

$$a^{\log_a x} = x, \qquad A a^{\log_a x} \frac{d}{dx} \log_a x = 1 \tag{9.251}$$

with the aid of the chain rule to obtain the second and hence obtain the second formula in (9.25). Conversely, if the second formula in (9.25) is known to hold, then differentiation of the members of the formula $\log_a a^x = x$ gives the first formula in (9.25). Since $\log_a 1 = 0$, replacing x by t in the second of the equivalent formulas (9.25) and integrating over the interval from 1 to x gives the formula

$$A \log_a x = \int_1^x \frac{1}{t}\, dt. \tag{9.252}$$

Since $\log_a a = 1$, putting $x = a$ gives the formula

$$A = \int_1^a \frac{1}{t}\, dt \tag{9.253}$$

which clarifies the relation between a and A.

Exponential functions and their derivatives and integrals occur so often that the constant A in the above formulas would, if allowed to survive, be an insufferable nuisance. Hence we can relish the proof that we can choose a to make $A = 1$. If we let $f(x)$ denote the right member of (9.252), then f is continuous and increasing over the whole interval $x > 0$, $f(1) = 0$, and

$$f(4) = \int_1^4 \frac{1}{t}\, dt > \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} > 1. \tag{9.254}$$

There is therefore exactly one positive number e for which $1 < e < 4$ and the constant A in (9.253) and preceding formulas is 1 when $a = e$. Thus all of the above formulas are correct when $a = e$ and $A = 1$. In particular, we have proofs of the basic formulas

$$\frac{d}{dx} e^x = e^x, \qquad \frac{d}{dx} \log x = \frac{1}{x} \tag{9.26}$$

and hence also of their companions

$$\int e^x\, dx = e^x + c, \qquad \int \frac{1}{x}\, dx = \log x + c. \tag{9.261}$$

Setting $a = e$ and $A = 1$ in (9.24) and (9.253) gives the formulas

$$\int_0^1 e^x\, dx = e - 1, \qquad \int_1^e \frac{1}{x}\, dx = 1, \tag{9.262}$$

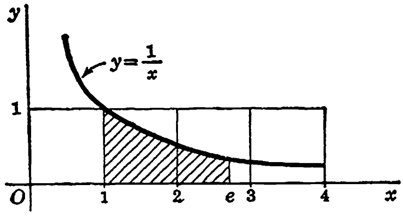


Figure 9.263

which we can now obtain very easily by evaluating the integrals. Each of these formulas actually determines e, and the second one says that e is so determined that the area A of the region shaded in Figure 9.263 is 1. Many other formulas involving e will appear later, and one of them will enable us to calculate e correct to 5 or 10 decimal places with surprisingly little effort. Meanwhile, we can remark that many persons remember the 16D (16-decimal, but only 15 after the decimal point) approximation to e by mentally grouping digits in the form

$$e = 2.7\ 1828\ 1828\ 45\ 90\ 45 \tag{9.264}$$

so we can visualize the repeated 1828 followed by 45 and twice 45 and 45.

The number e is the *natural base* of exponentials and logarithms. Other bases sometimes appear. The base 10 is used when we give a number n and say that "there are 10^n atoms in the universe," but not even Eddington suggested that this should be differentiated with respect to n. We *never* differentiate or integrate a^{k_1x} unless a is e. If, as never or rarely happens outside misguided examinations in calculus, we are called upon to differentiate or integrate a^{k_1x} where $a \neq e$, we write

$$a^{k_1x} = e^{k_2x} \tag{9.265}$$

and take logarithms with base e to obtain $k_1x \log a = k_2x$ and hence $k_2 = k_1 \log a$. We then work with e^{k_2x} instead of a^{k_1x}.

Our theory of exponentials and logarithms enables us to provide the promised proof of the power formula which, for convenience of reference, we put in a theorem.

Theorem 9.27 *If n is a constant, integer or not, then*

$$\frac{d}{dx}x^n = nx^{n-1}$$

when $x > 0$.

To prove this theorem, let $y = x^n$ and take logarithms to obtain $\log y = n \log x$. Since $\log y$ is differentiable, we can, because $y = e^{\log y}$, conclude that y itself is differentiable and can use the chain rule to obtain $(1/y)dy/dx = n/x$ so $dy/dx = nx^{n-1}$. This proves Theorem 9.27. We are now able to prove the last of the basic limit theorems, Theorem 3.288, which we now restate.

Theorem 9.271 *If p is a constant positive exponent, then the first of the formulas*

$$\lim_{x\to a} x^p = a^p, \qquad \lim_{x\to a+} x^p = a^p \tag{9.272}$$

holds when $a > 0$ and the second holds when $a \geqq 0$.

To prove this theorem, let p be a given positive constant and let f be the function for which $f(x) = x^p$ when $x \geqq 0$. That f is continuous at a when $a > 0$ is a consequence of the fact (see Theorem 9.27) that $f'(a)$ exists when $a > 0$. This proves that the formulas (9.272) are valid when $a > 0$. Since $0^p = 0$, it remains to be proved that

$$\lim_{x\to 0+} x^p = 0. \tag{9.273}$$

With the aid of the facts that

$$\lim_{x\to 0+} \log x = -\infty, \qquad \lim_{h\to -\infty} e^h = 0, \tag{9.274}$$

this follows from the equality

$$x^p = e^{p \log x}. \tag{9.275}$$

We conclude with some remarks that lead to a formula which has not appeared in our work but which nevertheless has much more than historical interest. If we know the basic properties of logarithms, we can let L be the function for which $L(x) = \log x$ when $x > 0$ and obtain the formula

$$\begin{aligned}\frac{L(x+h) - L(x)}{h} &= \frac{\log(x+h) - \log x}{h} = \frac{1}{h}\log\left(1 + \frac{h}{x}\right) \\ &= \frac{1}{x}\frac{x}{h}\log\left(1 + \frac{h}{x}\right) = \frac{1}{x}\log\left(1 + \frac{h}{x}\right)^{x/h}\end{aligned} \tag{9.28}$$

which is of interest from more than one point of view. If we can show, without reference to our previous results, that there is a number e for which

$$\lim_{h\to 0}\left(1 + \frac{h}{x}\right)^{x/h} = e, \tag{9.281}$$

then we can adopt this new number e for the base of our logarithms and use (9.281) and the fact that L is continuous and $L(e) = 1$ to obtain another proof of the formula $L'(x) = 1/x$. This formula then provides a back-handed proof of the formula $de^x/dx = e^x$. The difficulty in this view of (9.28) lies in the fact that "direct" proofs of existence of the limit in (9.281) are neither brief nor quickly comprehended. A proof is given at the end of the problems of this section. We now look again at (9.28) and realize that we have proved that $L'(x) = 1/x$ and hence that the members of (9.28) must converge to $1/x$ as $h \to 0$. Since L

is a continuous increasing function for which $L(e) = 1$, the formula (9.281) must be correct. Thus we have an "indirect" proof of (9.281). Replacing x and h by their reciprocals shows that (9.281) implies the more interesting formulas

$$(9.282) \qquad \lim_{|h|\to\infty}\left(1+\frac{x}{h}\right)^{h/x} = e, \qquad \lim_{|h|\to\infty}\left(1+\frac{x}{h}\right)^{h} = e^x.$$

The most famous of these formulas, obtained by setting $x = 1$ and considering h to be a positive integer, is

$$(9.283) \qquad \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n} = e.$$

Since $1 + 1/n > 1$, the result in (9.283) furnishes a very interesting compromise between the exponent n whose growth tends to make the quantity large and the term $1/n$ whose decrease tends to make the quantity near 1.

Problems 9.29

1 When we know some calculus, we can quickly dispose of problems that were troublesome. Show that the functions having values e^x and $\log x$ are increasing.

2 With the aid of information yielded by first and second derivatives, sketch a graph of $y = e^x$ and the tangent to the graph at the point (0,1).

3 With the aid of information yielded by first and second derivatives, sketch a graph of $y = \log x$ and the tangent to the graph at the point (1,0).

4 Let $y = xe^{-x}$. Show that

$$y'(x) = (1 - x)e^{-x}. \qquad \frac{d(e^{-x})}{dx} = -e^{-x}$$

Show that $y'(1) = 0$, $y'(x) > 0$ when $x < 1$, and $y'(x) < 0$ when $x > 1$. Show that

$$y''(x) = (x - 2)e^{-x}.$$

Determine when the slope of the graph is increasing and when it is decreasing. Use your information to sketch the graph.

5 Show that

(*a*) $\dfrac{d}{dx}e^{-x^2} = -2xe^{-x^2}$ (*b*) $\dfrac{d}{dx}e^{\sin x} = e^{\sin x}\cos x$

(*c*) $\dfrac{d}{dx}e^{e^x} = e^{x+e^x}$ (*d*) $\dfrac{d}{dx}e^{ae^{bx}} = abe^{bx+ae^{bx}}$

6 Show that

(*a*) $\dfrac{d}{dx}(\log x)^2 = \dfrac{2\log x}{x}$ (*b*) $\dfrac{d}{dx}\log\log x = \dfrac{1}{x\log x}$

(*c*) $\dfrac{d}{dx}\log\sin x = \cot x$ (*d*) $\dfrac{d}{dx}\log(1 + x^2) = \dfrac{2x}{1+x^2}$

7 Textbooks in differential equations provide substantial information about some of the situations in which a "population" or something else depends upon

t (time) in such a way that its rate of change with respect to t is proportional to it. There are in fact many situations in which it is assumed that y is a function of t such that, for some constant k,

$$\frac{dy}{dt} = ky \tag{1}$$

at each time t. We have thought about this matter before, and we shall think about it again. The pedestrian way to start to extract information from (1) is to divide by y. Without assuming that $y \neq 0$, show (with complete attention to each detail) that if (1) is true, then we can transpose ky and muliply by e^{-kt} to obtain the first of the formulas

$$\frac{d}{dt}e^{-kt}y = 0, \qquad e^{-kt}y = A, \qquad y = Ae^{kt} \tag{2}$$

and then the remaining ones in which A is a constant. Continue the work to show that if y satisfies (1) and the boundary condition

$$y = y_0 \text{ when } t = 0, \tag{3}$$

then

$$y = y_0e^{kt}. \tag{4}$$

Remark: This problem provides a major reason why exponential functions are important. In many cases, y decreases with passage of time and k is negative. In such cases there is a positive number T such that

$$\tfrac{1}{2}y_0e^{kt} = y_0e^{k(t+T)} \tag{5}$$

for each t. This number T, which is called the *half-life* of y, is the number of units of time required for half of y to fade away. From (5) we find that $\frac{1}{2} = e^{kT}$ or $e^{-kT} = 2$ or $-kT = \log 2$ or

$$T = -\frac{\log 2}{k} = -\frac{0.69315}{k}. \tag{6}$$

This formula can be used to determine the half-life T when we know k, and it can be used to determine k when we know the half-life.

8 Supposing that $0 \leqq x \leqq A$, observe that

$$0 \leqq e^x \leqq e^A. \tag{1}$$

Replace x by t and integrate over the interval from 0 to x to obtain

$$0 \leqq e^x - 1 \leqq e^Ax. \tag{2}$$

Replace x by t and integrate over the interval from 0 to x to obtain

$$0 \leqq e^x - 1 - x \leqq e^A\frac{x^2}{2!}. \tag{3}$$

Repeat the process to obtain

$$0 \leqq e^x - 1 - x - \frac{x^2}{2!} \leqq e^A\frac{x^3}{3!} \tag{4}$$

$$0 \leqq e^x - 1 - x - \frac{x^2}{2!} - \frac{x^3}{3!} \leqq e^A\frac{x^4}{4!}. \tag{5}$$

With or without more attention to details, jump to the conclusion that, for each $n = 1, 2, 3, \cdots$,

$$(6) \qquad 0 \leqq e^x - \sum_{k=0}^{n} \frac{x^k}{k!} \leqq e^A \frac{x^{n+1}}{(n+1)!} \leqq e^A \frac{A^{n+1}}{(n+1)!}.$$

While this may not be a suitable time to worry about the details, we can be sure that if $A = 400$, then the quantity

$$\frac{A^{n+1}}{(n+1)!} = \frac{A}{1}\frac{A}{2}\frac{A}{3}\frac{A}{4}\frac{A}{5} \cdots \frac{A}{n-1}\frac{A}{n}\frac{A}{n+1}$$

will be large when n is about A. The quantity will nevertheless be near 0 when n is sufficiently great. Therefore,

$$(7) \qquad e^x = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots.$$

While (7) is more spectacular than (6), it is not always as useful. Show that putting $x = A = 1$ in (6) gives the formula

$$(8) \qquad 0 \leqq e - \left(1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots + \frac{1}{n!}\right) \leqq \frac{3}{(n+1)!}.$$

Verify that

$$\begin{aligned} 1 &= 1. \\ 1 &= 1. \\ 1/2! &= 0.5 \\ 1/3! &= 0.16666\ 66666 \\ 1/4! &= 0.04166\ 66666 \\ 1/5! &= 0.00833\ 33333 \\ 1/6! &= 0.00138\ 88888 \end{aligned}$$

and that the next term is obtained by dividing by 7. Continue the work to obtain a decimal approximation to e that is correct to 6D (5 decimal places after the decimal point).

9 Supposing that $x < 0$, observe that

$$(1) \qquad 0 \leqq e^x \leqq 1.$$

Replace x by t and integrate over the interval from x to 0 to obtain

$$(2) \qquad 0 \leqq 1 - e^x \leqq -x.$$

Replace x by t and integrate over the interval from x to 0 to obtain

$$(3) \qquad 0 \leqq e^x - 1 - x \leqq \frac{x^2}{2!}.$$

Repeat the process to obtain

$$(4) \qquad 0 \leqq 1 + x + \frac{x^2}{2!} - e^x \leqq -\frac{x^3}{3!}$$

$$(5) \qquad 0 \leqq e^x - 1 - x - \frac{x^2}{2!} - \frac{x^3}{3!} \leqq \frac{x^4}{4!}.$$

Note that considerations very similar to those in Problem 8 establish validity of the formula

$$(6) \qquad e^x = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

when $x < 0$. For some purposes, (4) and (5) and their extensions are more useful than (6).

10 Have another look at the formula

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

and remember it. Then write the formula and use it to obtain an approximation to $e^{1/2}$ which agrees with the idea that $(e^{1/2})^2 = e$.

11 Prove one of the inequalities

$$e^x \geqq 1 + x, \qquad \log(1 + x) \leqq x, \qquad \log x \leqq x - 1$$

and show that each implies the other two. *Hint:* If no better idea appears, find the minimum value of $e^x - 1 - x$.

12 Prove that if $x > 0$ and k is a positive integer, then

$$0 < \frac{x^k}{e^x} < \frac{(k+1)!x^k}{x^{k+1}} = \frac{(k+1)!}{x}.$$

Use this result to show that

$$\lim_{x\to\infty} \frac{x^k}{e^x} = 0.$$

13 We can become accustomed to the formulas

$$(1) \qquad \frac{1 - t^n}{1 - t} = 1 + t + t^2 + \cdots + t^{n-1}$$

$$(2) \qquad \frac{1}{1 - t} = 1 + t + t^2 + \cdots + t^{n-1} + \frac{t^n}{1 - t}$$

if we see them often enough. Prove that if $-1 \leqq x < 1$, then integration from 0 to x gives the formula

$$(3) \qquad \log \frac{1}{1 - x} = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots + \frac{x^n}{n} + R_n$$

where

$$(4) \qquad R_n = \int_0^x \frac{t^n}{1 - t}\, dt.$$

In case $0 \leqq x < 1$, show that

$$(5) \qquad 0 \leqq R_n \leqq \int_0^x \frac{t^n}{1 - x}\, dt = \frac{x^{n+1}}{(n+1)(1 - x)} < \frac{1}{(n+1)(1 - x)}$$

and in case $-1 \leqq x < 0$, show that

$$(6) \qquad |R_n| = \left| \int_x^0 \frac{t^n}{1 - t}\, dt \right| \leqq \left| \int_x^0 t^n\, dt \right| = \frac{|x|^{n-1}}{n+1} \leqq \frac{1}{n+1}.$$

Hence prove that

$$\log \frac{1}{1-x} = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots \tag{7}$$

when $-1 \leqq x < 1$ and that

$$\log 2 = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \cdots . \tag{8}$$

Show that replacing x by $(-x)$ in (7) gives the formula

$$\log (1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \tag{9}$$

and that addition gives the formula

$$\log \frac{1+x}{1-x} = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots\right) \tag{10}$$

which holds when $|x| < 1$. Use (10) and a little imagination to obtain the formula

$$\log 2 = 2\left(\frac{1}{3} + \frac{1}{3\cdot 3^3} + \frac{1}{5\cdot 3^5} + \cdots\right) \tag{11}$$

from which we could, with the aid of a calculator or computer, calculate log 2 correct to many decimal places. Some persons need approximations as good as that in

$$\log 2 = 0.69314\ 71805\ 59945\ 30941. \tag{12}$$

If we want log 10, we could get log 8 by multiplying (12) by 3 and then add the result to $\log (\frac{10}{8})$ which can be calculated from the formula

$$\log \frac{10}{8} = 2\left(\frac{1}{9} + \frac{1}{3\cdot 9^3} + \frac{1}{5\cdot 9^5} + \frac{1}{7\cdot 9^7} + \cdots\right) \tag{13}$$

in which the series converges quite rapidly. The result is

$$\log 10 = 2.30258\ 50929\ 94045\ 68402. \tag{14}$$

To get log 3, we could let $x = \frac{1}{2}$ in (9), but this series converges rather slowly. It is much better to calculate log 9 and then log 3 from (14) and the formula

$$\log \frac{10}{9} = 2\left(\frac{1}{19} + \frac{1}{3\cdot 19^3} + \frac{1}{5\cdot 19^5} + \cdots\right). \tag{15}$$

This gives

$$\log 3 = 1.09861\ 22886\ 68109\ 69140. \tag{16}$$

14 Elementary combinations of the logarithms of 2, 3, and 10 give the logarithms of 2, 3, 4, 5, 6, 8, 9, 10, but 7 is missing. Show how the proximity of 49 to 50 enables us to calculate log 7 with the aid of a series that converges rapidly.

15 This book does not recommend learning a formula for dy/dx when $y = f(x)^{g(x)}$, where f and g are given functions. If (as frequently happens in good old-fashioned mathematics examinations) we are required to produce the derivative, we can begin by being irked by the fact that the base is not e. There are two superficially different ways to put the given equation in a form to which basic rules are applicable. Taking logarithms (with base e) gives

$$\log y = g(x) \log f(x) \tag{1}$$

and, with or without using (1), we can put the equation in the form

$$y = e^{g(x) \log f(x)}. \tag{2}$$

When (1) is used, we differentiate to obtain

$$\frac{1}{y}\frac{dy}{dx} = g(x)\frac{1}{f(x)}f'(x) + g'(x) \log f(x) \tag{3}$$

and, with the aid of the fact that $y = f(x)^{g(x)}$,

$$\frac{dy}{dx} = f(x)^{g(x)}\left[\frac{g(x)f'(x)}{f(x)} + g'(x) \log f(x)\right]. \tag{4}$$

When (2) is used, we differentiate to obtain

$$\frac{dy}{dx} = e^{f(x) \log g(x)}\left[f(x)\frac{1}{g(x)}g'(x) + f'(x) \log g(x)\right] \tag{5}$$

and then use (2) to obtain the result (4). With the above formulas out of sight, use one of the above ideas to obtain a formula for dy/dx when

(*a*) $y = x^x$ *Ans.*: $\dfrac{dy}{dx} = x^x(1 + \log x)$

(*b*) $y = (1 + x)^{1/x}$ *Ans.*: $\dfrac{dy}{dx} = (1 + x)^{1/x}\dfrac{x - (1 + x) \log (1 + x)}{x^2(1 + x)}$

(*c*) $y = (\sin x)^{\sin x}$ *Ans.*: $\dfrac{dy}{dx} = \cos x(1 + \log \sin x)(\sin x)^{\sin x}$

(*d*) $y = (\sin x)^{\tan x}$ *Ans.*: $\dfrac{dy}{dx} = (1 + \sec^2 x \log \sin x)(\sin x)^{\tan x}$

(*e*) $y = x^{\log x}$ *Ans.*: $\dfrac{dy}{dx} = 2x^{\log x - 1} \log x$

(*f*) $y = a^{b^x}$ *Ans.*: $\dfrac{dy}{dx} = a^{b^x}b^x \log b \log a$

(*g*) $y = x^{b^x}$ *Ans.*: $\dfrac{dy}{dx} = x^{b^x}b^x \log x\left[\log b + \dfrac{1}{x \log x}\right]$

16 One who is interested in solving a puzzle with an interesting answer may undertake to determine the nature of the graph of the equation $y = x^x$. One who has never thought about the orders of magnitudes of the numbers $(0.01)^{0.01}$ and $(\frac{1}{2})^{1/2}$ may even be surprised by the results.

17 If the result has not already been obtained, let $y = x^x$ and show that

$$y''(x) = x^x(1 + \log x)^2 + x^{x-1} > 0.$$

Tell what this says about the graph of $y = x^x$.

18 Derive the formula

$$\int \frac{1}{e^x + 1}\,dx = x - \log(e^x + 1) + c$$

with the aid of the identity

$$\frac{1}{e^x + 1} = \frac{e^x + 1 - e^x}{e^x + 1} = 1 - \frac{e^x}{e^x + 1}.$$

19 Letting $F(0) = 0$ and $F(x) = e^{-1/x^2}$ when $x \neq 0$, show that the formula

(1) $$F^{(k)}(x) = \frac{P_k(x)}{x^{3k}}\,e^{-1/x^2}$$

holds when $x \neq 0$, $k = 1$, and $P_1(x) = 2$. Supposing that (1) is valid when k is a given positive integer n and $P_n(x)$ is a particular polynomial in x, differentiate (1) to obtain a formula which shows that (1) holds when $k = n + 1$ and

(2) $$P_{n+1}(x) = (2 - 3nx^2)P_n(x) - x^3P_n'(x),$$

so that P_{n+1} is a polynomial in x. This shows (the principle involved being called mathematical induction) that (1) holds for each k, P_k being a polynomial in x.

20 Prove that if P and Q are polynomials in x for which $Q(x)$ is not always zero, then

(1) $$\lim_{x\to 0} \frac{P(x)}{Q(x)}\,e^{-1/x^2} = 0.$$

Solution: We can determine constants such that $b_0 \neq 0$, q is an integer, and

(2) $$\frac{P(x)}{Q(x)} = \frac{a_0 + a_1x + \qquad + a_mx^m}{b_0 + b_1x + \cdots + b_nx^n}\,x^q.$$

It therefore suffices to prove that

(3) $$\lim_{x\to 0} |x|^q e^{-1/x^2} = 0,$$

or, as we see by setting $t = 1/x^2$,

(4) $$\lim_{t\to\infty} t^{-q/2}e^{-t} = 0.$$

But when $t > 1$ and k is a positive integer for which $-q/2 < k$,

(5) $$|t^{-q/2}e^{-t}| < t^ke^{-t} = \frac{t^k}{e^t}.$$

Our conclusion is therefore a consequence of the formula

(6) $$\lim_{t\to\infty} \frac{t^k}{e^t} = 0$$

which is proved in Problem 12.

21 Let F be the function for which $f(0) = 0$ and

(1) $$F(x) = e^{-1/x^2}$$

when $x \neq 0$. With the aid of results of the two preceding problems, prove that

F and each of the derivatives $F^{(n)}$, $n = 1, 2, 3, \cdots$, is continuous over the whole infinite interval $-\infty < x < \infty$ and, moreover,

$$F^{(n)}(0) = 0 \qquad (n = 1, 2, 3, \cdots). \tag{2}$$

Remark: This is one of the famous functions of mathematical analysis. Note that we have no formulas into which we can put $x = 0$ to obtain the numbers $F^{(n)}(0)$. After we have learned that $F^{(k)}(0) = 0$ when $k = n$, we must use the definition

$$F^{(n+1)}(0) = \lim_{x\to 0} \frac{F^{(n)}(x) - F^{(n)}(0)}{x}$$

to learn that $F^{(k)}(0) = 0$ when $k = n + 1$.

22 Supposing that $a_1, a_2, a_3, \cdots$ are positive numbers, that n is an integer for which $n > 1$, that $x > 0$, and that

$$f_n(x) = \frac{\dfrac{a_1 + a_2 + \cdots + a_{n-1} + x}{n}}{\sqrt[n]{a_1a_2 \cdots a_{n-1}x}}, \tag{1}$$

prove that $f_n(x)$ is a minimum when and only when

$$x = \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}. \tag{2}$$

Hint: Use the fact that $f_n(x)$ is a minimum when $\log f_n(x)$ is a minimum. Show that

$$\frac{d}{dx}\log f_n(x) = \frac{n-1}{nx(a_1 + a_2 + \cdots + a_{n-1} + x)}\left[x - \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}\right]. \tag{3}$$

Remark: Our result shows that

$$f_n\left(\frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}\right) \leqq f_n(a_n) \tag{4}$$

and that equality holds only when a_n is the arithmetic mean (or ordinary average) of the numbers $a_1, a_2, \cdots, a_{n-1}$. Interest in this matter can start to develop when we observe that

$$f_n(a_n) = \frac{\dfrac{a_1 + a_2 + \cdots + a_n}{n}}{\sqrt[n]{a_1a_2 \cdots a_n}} \tag{5}$$

and hence that $f_n(a_n)$ is the ratio of the arithmetic mean of the numbers $a_1, a_2, \cdots, a_n$ to the geometric mean of the same numbers. By using (1) to obtain an expression for the left member of the formula

$$f_n\left(\frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}\right) = \left[\frac{\dfrac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}}{(a_1a_2 \cdots a_{n-1})^{1/(n-1)}}\right]^{(n-1)/n} = [f_{n-1}(a_{n-1})]^{(n-1)/n} \tag{6}$$

and applying some quite elementary algebraic operations to simplify the result, we can derive the formula. From (6) and (4) we obtain the formula

$$[f_{n-1}(a_{n-1})]^{n-1} \leqq [f_n(a_n)]^n. \tag{7}$$

Since $f_1(a_1) = 1$, this gives the remarkable parade of inequalities

$$(8)\quad 1 \leqq \left(\frac{\frac{a_1 + a_2}{2}}{\sqrt{a_1a_2}}\right)^2 \leqq \left(\frac{\frac{a_1 + a_2 + a_3}{3}}{\sqrt[3]{a_1a_2a_3}}\right)^3 \leqq \left(\frac{\frac{a_1 + a_2 + a_3 + a_4}{4}}{\sqrt[4]{a_1a_2a_3a_4}}\right)^4 \leqq \cdots .$$

The denominator in these quotients cannot exceed the numerators, and therefore

$$(9)\qquad \sqrt[n]{a_1a_2 \cdots a_n} \leqq \frac{a_1 + a_2 + \cdots + a_n}{n} \qquad (n = 1,2,3, \cdots).$$

The simple corollary (9) of the more spectacular result (8) is a statement of a very famous theorem which says that *if* $a_1, a_2, \cdots, a_n$ *are positive numbers, then their geometric mean is less than or equal to their arithmetic mean.* Our work enables us to show that *equality holds only when the numbers* $a_1, a_2, \cdots, a_n$ *are equal.* Many proofs of this theorem are known, and it is quite appropriate to become interested in the matter by looking at the special case

$$(10)\qquad 27a_1a_2a_3 \leqq (a_1 + a_2 + a_3)^3$$

and seeking ways to prove it. Sometimes scientists say that boys work with equalities and men work with inequalities.

23 For those who are interested in the matter, we present a direct proof that

$$(1)\qquad \lim_{|x| \to \infty} f(x) = e \text{ when } f(x) = \left(1 + \frac{1}{x}\right)^x.$$

When n is a positive integer, putting $a = 1$ and $b = 1/n$ in the binomial formula

$$(2)\quad (a + b)^n = a^nb^0 + na^{n-1}b^1 + \frac{n(n-1)}{2!}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3!}a^{n-3}b^3 + \cdots + \frac{n(n-1)(n-2)\cdots 2\cdot 1}{n!}a^0b^n$$

gives

$$(3)\quad f(n) = 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \frac{1}{4!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right) + \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{n-1}{n}\right).$$

Hence

$$(4)\quad f(n) \leqq 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} < 1 + \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n}\right) < 3.$$

It follows from Theorem 5.65 that the series in

$$(5)\qquad f(n) < 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots = e$$

converges to a number which we can call e and that (5) holds. If $m > n$, then replacing n by m in (3) shows that $f(m)$ consists of $(m + 1)$ terms of which the first $(n + 1)$ equal or exceed those in the right member of (3) and the remaining $m - n$ terms are positive. Hence $f(m) > f(n)$ when $m > n$. As we can see by putting $s_n = f(n)$ in Theorem 5.651, this and (5) imply existence of a number L for which

$$\lim_{n\to\infty} f(n) = L \leqq e. \tag{6}$$

To prove that $L = e$, we can use (3) and (5) to obtain

$$1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots$$
$$+ \frac{1}{N!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{N-1}{n}\right) \leqq f(n) \leqq e$$

when $n > N$. Letting $n \to \infty$ gives

$$1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{N!} \leqq L \leqq e, \tag{7}$$

and letting $N \to \infty$ shows that $L = e$. Supposing now that $x > 2$ and that x is not necessarily an integer, we can let n be the greatest integer in x so that $n \leqq x \leqq n + 1$. Then it is easily seen that

$$\left(1 + \frac{1}{n+1}\right)^n \leqq f(x) \leqq \left(1 + \frac{1}{n}\right)^{n+1}. \tag{8}$$

The first and last members of this inequality converge to e as $n \to \infty$ because they are respectively equal to

$$\left(1 + \frac{1}{n+1}\right)^{n+1}\left(1 + \frac{1}{n+1}\right)^{-1}, \quad \left(1 + \frac{1}{n}\right)^n\left(1 + \frac{1}{n}\right) \tag{9}$$

and the second factors converge to 1 as $n \to \infty$. Therefore, $f(x) \to e$ as $x \to \infty$. Proof of the part of (1) involving negative values of x is provided by the calculation

$$\lim_{x\to-\infty}\left(1 + \frac{1}{x}\right)^x = \lim_{x\to\infty}\left(1 - \frac{1}{x}\right)^{-x} = \lim_{x\to\infty}\left(\frac{x}{x-1}\right)^x \tag{10}$$
$$= \lim_{x\to\infty}\left(1 + \frac{1}{x-1}\right)^x = \lim_{x\to\infty}\left(1 + \frac{1}{x-1}\right)^{x-1}\left(1 + \frac{1}{x-1}\right) = e\cdot 1 = e.$$

9.3 *Hyperbolic functions* We begin with a peek at some formulas that are very useful in more advanced mathematics and will appear again in Section 12.4. The formula

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots, \tag{9.31}$$

which has been proved to be valid when z is a real number, is used to define e^z when z is a complex number of the form $z = x + iy$, where x and y are real and i is the so-called *imaginary unit* for which $i^2 = -1$. Since $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$, $\cdots$, it can be shown that

$$e^{iz} = \left(1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots\right) \tag{9.311}$$
$$+ i\left(z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots\right)$$

and hence that

$$e^{iz} = \cos z + i \sin z \tag{9.312}$$
$$e^{-iz} = \cos z - i \sin z. \tag{9.313}$$

Adding and subtracting give

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin z = \frac{e^{iz} - e^{-iz}}{2i}. \tag{9.314}$$

These are *Euler* (1707–1783) *formulas.* They are widely used, particularly in electrical engineering, to replace calculations involving sines, cosines, and their positive powers by simpler calculations involving exponentials. We do not need to understand these matters now, but we can at least acquire a vague feeling that trigonometric functions are related to exponential functions by formulas very similar to those which relate hyperbolic functions to exponential functions. In any case, we can know that there is a reason why formulas involving hyperbolic functions are so similar to formulas involving trigonometric functions. Modern scientists know that their ancestors complicated many problems by habitually using trigonometric and hyperbolic functions in situations in which results are obtained much more neatly and quickly by use of exponentials. Thus hyperbolic functions are introduced to students with the hope that they will (insofar as they can control their own activities) use the functions only for purposes for which they are useful.

We now return to our usual situation in which all numbers are real. The *hyperbolic sine*, *hyperbolic cosine*, and other *hyperbolic functions* are defined by the first of the following equations, and calculation of the derivatives gives practice in differentiation.

$$\sinh x = \frac{e^x - e^{-x}}{2}, \qquad \frac{d}{dx}\sinh x = \cosh x \tag{9.32}$$
$$\cosh x = \frac{e^x + e^{-x}}{2}, \qquad \frac{d}{dx}\cosh x = \sinh x \tag{9.321}$$
$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \qquad \frac{d}{dx}\tanh x = \operatorname{sech}^2 x \tag{9.322}$$
$$\coth x = \frac{e^x + e^{-x}}{e^x - e^{-x}}, \qquad \frac{d}{dx}\coth x = -\operatorname{csch}^2 x \tag{9.323}$$
$$\operatorname{sech} x = \frac{2}{e^x + e^{-x}}, \qquad \frac{d}{dx}\operatorname{sech} x = -\operatorname{sech} x \tanh x \tag{9.324}$$
$$\operatorname{csch} x = \frac{2}{e^x - e^{-x}}, \qquad \frac{d}{dx}\operatorname{csch} x = -\operatorname{csch} x \coth x \tag{9.325}$$

These and many other formulas are very similar to trigonometric formulas, but differences in signs must be noted. As is the case for trigonometric functions, the first and last are reciprocals, the next to the first and the next to the last are reciprocals, and the middle two are reciprocals. With the aid of the fact that $e^x e^{-x} = 1$, we can square $\cosh x$ and $\sinh x$ and obtain the first of the formulas

$$\cosh^2 x - \sinh^2 x = 1 \tag{9.33}$$
$$\tanh^2 x + \operatorname{sech}^2 x = 1 \tag{9.331}$$
$$\coth^2 x - \operatorname{csch}^2 x = 1. \tag{9.332}$$

To obtain the second (or third) formula, we can divide by $\cosh^2 x$ (or by $\sinh^2 x$) and transpose some terms. Graphs of the first three hyperbolic functions are shown in Figures 9.34 and 9.341, and the others are easily drawn.

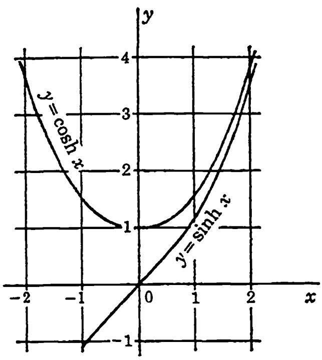


Figure 9.34

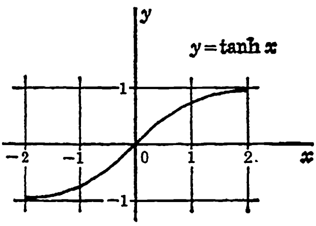


Figure 9.341

To work out formulas for the inverses of the first hyperbolic functions, we let

$$x = \cosh t = \frac{e^t + e^{-t}}{2}, \qquad y = \sinh t = \frac{e^t - e^{-t}}{2} \tag{9.35}$$

and observe that $\cosh t$ is increasing over the interval $t \geqq 0$ and $\sinh t$ is increasing over the whole infinite interval. The equations (9.35) can be put in the forms

$$e^{2t} - 2xe^t + 1 = 0, \qquad e^{2t} - 2ye^t - 1 = 0. \tag{9.351}$$

These equations are quadratic in e^t, and solving for e^t gives

$$e^t = x + \sqrt{x^2 - 1}, \qquad e^t = y + \sqrt{y^2 + 1}, \tag{9.352}$$

it being necessary to choose the positive sign in each case because e^t is positive and increasing as x and y increase. Taking logarithms and changing y to x gives the first items in the first two of the following formulas. Similar methods and differentiation give the remaining items.

(9.36) $\sinh^{-1} x = \log (x + \sqrt{x^2 + 1}),$ $\quad \dfrac{d}{dx} \sinh^{-1} x = \dfrac{1}{\sqrt{1 + x^2}}$

(9.361) $\cosh^{-1} x = \log (x + \sqrt{x^2 - 1}),$ $\quad \dfrac{d}{dx} \cosh^{-1} x = \dfrac{1}{\sqrt{x^2 - 1}}$ $\quad (x > 1)$

(9.362) $\tanh^{-1} x = \frac{1}{2} \log \dfrac{1 + x}{1 - x},$ $\quad \dfrac{d}{dx} \tanh^{-1} x = \dfrac{1}{1 - x^2}$ $\quad (|x| < 1)$

(9.363) $\coth^{-1} x = \frac{1}{2} \log \dfrac{x + 1}{x - 1},$ $\quad \dfrac{d}{dx} \coth^{-1} x = \dfrac{-1}{x^2 - 1}$ $\quad (|x| > 1)$

(9.364) $\operatorname{sech}^{-1} x = \log \dfrac{1 + \sqrt{1 - x^2}}{x},$ $\quad \dfrac{d}{dx} \operatorname{sech}^{-1} x = \dfrac{-1}{x\sqrt{1 - x^2}}$ $\quad (0 < x < 1)$

(9.365) $\operatorname{csch}^{-1} x = \log \dfrac{1 + \sqrt{1 + x^2}}{x},$ $\quad \dfrac{d}{dx} \operatorname{csch}^{-1} x = \dfrac{-1}{x\sqrt{1 + x^2}}$ $\quad (x > 0)$

Because of the growing tendency to use exponentials to eliminate calculations involving hyperbolic functions and even trigonometric functions, it seems unwise to devote more time to the formal aspects of the subject. It is, however, of interest to know how hyperbolic functions are related to hyperbolas. Setting

(9.37) $$x = \cosh t = \frac{e^t + e^{-t}}{2}, \qquad y = \sinh t = \frac{e^t - e^{-t}}{2},$$

we see that $x > 0$ and $x^2 - y^2 = 1$, so $P(x,y)$ lies on the branch of the rectangular hyperbola shown in Figure 9.371. A reasonable way to try

Figure 9.371

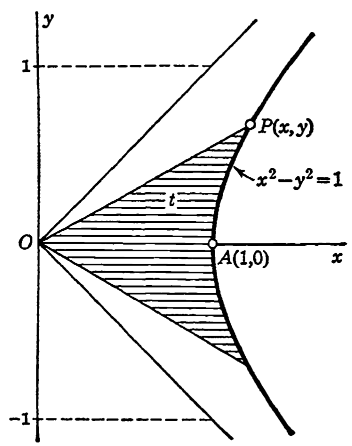
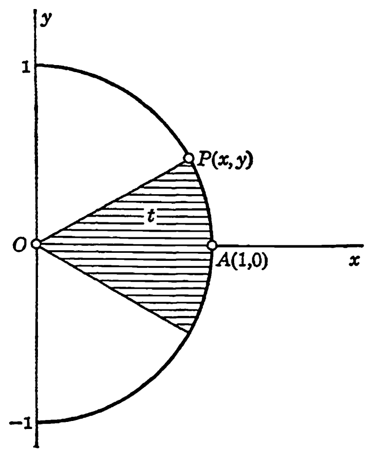


Figure 9.372



to discover the role of the parameter t is to let $\mathbf{r}$ be the vector running from the origin to a particle which occupies the position $(x(t),y(r))$ at time t. Then

$$\mathbf{r} = \cosh t\mathbf{i} + \sinh t\mathbf{j} \tag{9.381}$$

and differentiation gives the velocity and acceleration

$$\mathbf{v} = \sinh t\mathbf{j} + \cosh t\mathbf{j} \tag{9.382}$$
$$\mathbf{a} = \cosh t\mathbf{i} + \sinh t\mathbf{j}. \tag{9.383}$$

Thus $\mathbf{a} = \mathbf{r}$, so the particle is accelerated directly away from the origin, and the magnitude of the acceleration is proportional to (actually equal to) the first power of the distance. It is then known from classical physics that the particle must (as it does) move on a conic. Moreover, because the force is a central force which is always directed away from or toward the origin, the angular momentum of the particle must be constant and the vector from O to P must sweep over regions of equal area in equal time intervals. With (or perhaps even without) the aid of principles of physics that tell us to examine areas, we can calculate the area of the shaded region of Figure 9.371 and show that the area is $\cosh^{-1} x$ and hence is t. The problems show how the details can be handled. Thus the geometrical similarity between trigonometric functions (which used to be called circular functions) and hyperbolic functions (which still are called hyperbolic functions) is exposed. Trigonometric functions of t are "functions of the sector of area t of the unit circle shown in Figure 9.372." Hyperbolic functions of t are "functions of the sector of area t of the unit hyperbola shown in Figure 9.371." Those who have not peered into ancient mathematical tomes and are more accustomed to "sines of angles" than to "sines of arcs" and "sines of sectors" should quietly observe that the sector of the unit circle of Figure 9.372 has area t when the arc from A to P has length t and hence when the angle AOP "contains" t radians. Among other things, this little excursion into history explains the antics of those who write arcsin x in place of $\sin^{-1} x$ and talk about "the arc whose sine is x."

Problems 9.39

1 Verify the six formulas for derivatives of hyperbolic functions.

2 Verify the six formulas for derivatives of inverse hyperbolic functions.

3 Textbooks on differential equations show that if a flexible homogeneous cable or chain is suspended from its two ends and sags under the influence of a parallel force field (an idealized gravitational field) then the cable or chain must occupy a part of the graph of the equation

$$y = \frac{1}{2k}(e^{kx} + e^{-kx}) = \frac{1}{k}\cosh kx, \tag{1}$$

provided that the positive constant k and the coordinate system are suitably determined. Because of this fact and the fact that the Latin word for chain is *catenus*, the graph of (1) appearing in Figure 9.391 is called a *catenary*. Supposing

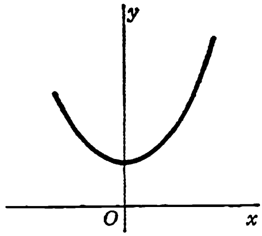


Figure 9.391

that $a > 0$, find the length L of the part of this catenary that hangs over the interval $0 \leqq x \leqq a$. *Ans.:* $L = \frac{1}{k} \sinh ka$.

4 Compare the graphs of $y = \operatorname{sech} x$ and $y = 1/(1 + x^2)$.

5 Starting with the formula

$$t = \log (x + \sqrt{x^2 - 1}),$$

show that $x = \cosh t$.

6 Prove the formula

$$\int \sqrt{x^2-1}\, dx = \tfrac{1}{2}x \sqrt{x^2 - 1} - \tfrac{1}{2} \cosh^{-1} x + c$$

with the aid of (8.488) and use it to show that

$$\cosh^{-1} x_0 = 2\left[\tfrac{1}{2}x_0 \sqrt{x_0^2 - 1} - \int_1^{x_0} \sqrt{x^2 - 1}\, dx\right].$$

Use this to show that $\cosh^{-1} x_0$ is the area of the shaded region of Figure 9.371 when $x = x_0$.

7 Evaluate the integral of the preceding problem with the aid of a hyperbolic function substitution.

8 Show that

$$e^x = \cosh x + \sinh x$$
$$e^{-x} = \cosh x - \sinh x.$$

Show that if there exist constants c_1 and c_2 for which

(1) $$f(t) = c_1 e^{(a+b)t} + c_2 e^{(a-b)t},$$

then there also exist constants C_1 and C_2 for which

(2) $$f(t) = e^{at}[C_1 \cosh bt + C_2 \sinh bt].$$

Use (1) to calculate formulas for $f'(t)$, $f''(t)$, $f'''(t)$, and $f^{(4)}(t)$. Use (2) to calculate $f'(t)$ and $f''(t)$ and observe that the hyperbolic functions are being nuisances.

9 Supposing that a and b are constants for which $b^2 \neq a^2$, obtain the formula

$$\int e^{at} \cosh bt\, dt = \frac{e^{at}(b \sinh bt - a \cosh bt)}{b^2 - a^2} + c$$

by integrating by parts twice. For the case where $b^2 = a^2$ see the next problem.

10 Considering separately the three cases in which $b^2 \neq a^2$, $b = a$, and $b = -a$, evaluate the integral

$$\tfrac{1}{2} \int [e^{(a+b)t} + e^{(a-b)t}]\, dt$$

and put the answers in terms of hyperbolic functions.

11 Determine whether

(*a*) $$\lim_{a \to -1} \int_1^x t^a\, dt = \int_1^x t^{-1}\, dt$$

(*b*) $$\lim_{b \to a} \int_0^x e^{at} \cosh bt\, dt = \int_0^x e^{at} \cosh at\, dt$$

9.4 *Partial fractions*

Let

$$(9.411) \qquad f(x) = x^2 - 2x + 1 + \frac{2}{x-1} + \frac{x}{x^2+1} + \frac{1}{x^2+1}.$$

Use of basic integration formulas then gives

$$(9.412) \qquad \int f(x)\, dx = \frac{x^3}{3} - x^2 + x + 2 \log |x - 1| + \tfrac{1}{2} \log (x^2 + 1) + \tan^{-1} x + c.$$

Adding the terms in the right member of (9.411) gives

$$(9.413) \qquad f(x) = \frac{x^5 - 3x^4 + 4x^3 - x^2 + 3x}{x^3 - x^2 + x - 1}$$

and shows us that the integral

$$(9.414) \qquad \int \frac{x^5 - 3x^4 + 4x^3 - x^2 + 3x}{x^3 - x^2 + x - 1}\, dx$$

is equal to the right member of (9.412). Interest in this business starts to develop when we wonder how we would evaluate the integral (9.414) if it were handed to us without the preceding formulas. The answer to this question is quite straightforward. We learn and use a procedure by which the preceding formulas can be worked out.

The first step is to look at (9.414). The integrand is a quotient of polynomials in x, and the degree of the numerator is not less than the degree of the denominator. In such cases we employ division (or long division) to obtain a polynomial and a new quotient in which the degree of the numerator is less than that of the denominator. Letting $f(x)$ be defined by (9.413), we divide to obtain

$$f(x) = x^2 - 2x + 1 + \frac{3x^2 + 1}{x^3 - x^2 + x - 1}. \tag{9.42}$$

Our difficulties will have been surmounted when we succeed in expressing the last quotient (or fraction) in (9.42) as a sum of simpler quotients (or fractions) that are called *partial fractions*. This brings us to the key problem of this section, namely, the problem of representing a given *proper rational function* (that is, a quotient of polynomials in which the degree of the numerator is less than that of the denominator) in terms of partial fractions. This key problem is important outside as well as inside the calculus. The problem is treated in some algebra books, but students normally make their first acquaintance with the problem in calculus books.

To begin discovery of the partial fractions whose sum is a given proper rational fraction, we must factor the given denominator and use these factors to determine the nature of the partial fractions. Electronic computers are often used to factor denominators, but factoring of the denominator of the quotient in (9.42) is quite easy if we happen to notice that

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

Thus the quotient in (9.42) is the left member of the equality

$$\frac{3x^2 + 1}{(x - 1)(x^2 + 1)} = \frac{A}{x - 1} + \frac{Bx + C}{x^2 + 1}. \tag{9.43}$$

While we are entitled to be quite mystified by the fact until the matter has been investigated, it is possible to determine three constants A, B, C such that (9.43) is true for each x for which the denominators are all different from zero. In fact, we can clear the denominators from (9.43) and determine the constants so that the formula

$$3x^2 + 1 = A(x^2 + 1) + (Bx + C)(x - 1) \tag{9.431}$$

is true for each x. One way to prove this and to find the constants is to put (9.431) in the form

$$3x^2 + 1 = (A + B)x^2 + (-B + C)x + (A - C)$$

and to observe that this is surely an identity in x if

$$A + B = 3, \qquad -B + C = 0, \qquad A - C = 1$$

and hence (as we show by solving these equations) if

$$A = 2, \qquad B = 1, \qquad C = 1. \tag{9.432}$$

There is, however, an easier way to find A, B, C directly from (9.431). Putting $x = 1$ in (9.431) shows immediately that $4 = A(2)$ and hence that $A = 2$. Putting $x = 0$ shows that $1 = A - C$ or $1 = 2 - C$, so $C = 1$. Finally, putting $x = 2$ shows that $13 = 2\cdot 5 + (2B + 1)$, so $B = 1$. Substituting from (9.432) into (9.43) gives

$$\frac{3x^2 + 1}{(x - 1)(x^2 + 2)} = \frac{2}{x - 1} + \frac{x + 1}{x^2 + 1}. \tag{9.433}$$

Since this is the last quotient in (9.42), we obtain (9.411). Thus we have succeeded in writing the given quotient in (9.413) as the sum of a polynomial and partial fractions.

Some problems are easier than the one we have solved, and some are more difficult. When we wish to integrate the left member of

$$\frac{2x + 6}{(x - 1)(x + 1)} = \frac{A}{x - 1} + \frac{B}{x + 1}, \tag{9.44}$$

we determine A and B so that (9.44) and

$$2x + 6 = A(x + 1) + B(x - 1) \tag{9.441}$$

hold. Putting $x = 1$ shows that $A = 4$, and putting $x = -1$ shows that $B = -2$. Thus

$$\frac{2x + 6}{(x - 1)(x + 1)} = \frac{4}{x - 1} - \frac{2}{x + 1},$$

and integration gives

$$\int \frac{2x + 6}{x^2 - 1}\, dx = 4 \log |x - 1| - 2 \log |x + 1| + c.$$

The result can, for better or for worse, be put in the form

$$\int \frac{2x + 6}{x^2 - 1}\, dx = \log \frac{(x - 1)^4}{(x + 1)^2} + c.$$

The problems at the end of this section provide additional clues to methods by which quotients are expressed in terms of partial fractions. Problem 11 proposes study of basic theory, and persons interested in more than the simplest mechanical aspects of our subject can do this studying at any time and perhaps even more than once.

Problems 9.49

1 Show that, when $p \neq q$ and $a > 0$,

$$\int \frac{1}{a(x - p)(x - q)}\, dx = \frac{1}{a(p - q)} \log \left| \frac{x - p}{x - q} \right| + c. \tag{1}$$

Remark: We should not be too busy to see how this formula leads to another that also appears in tables of integrals. Let

$$X = ax^2 + bx + c, \tag{2}$$

where a, b, c are constants for which $a > 0$ and $b^2 - 4ac > 0$. Observe that

$$X = a(x - p)(x - q), \tag{3}$$

where p and q are the values of x for which $X = 0$, so that

$$p = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \qquad q = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \qquad p - q = \frac{\sqrt{b^2 - 4ac}}{a}. \tag{4}$$

Substituting in (1) gives the formula

$$\int \frac{1}{X}\,dx = \frac{1}{\sqrt{b^2 - 4ac}} \log \left| \frac{2ax + b - \sqrt{b^2 - 4ac}}{2ax + b + \sqrt{b^2 - 4ac}} \right| + c. \tag{5}$$

We should know that additional tricks produce additional formulas that appear in tables of integrals. When $a > 0$, integrating the first and last members of the identity

$$\frac{x}{ax^2 + bx + c} = \frac{1}{2a}\frac{(2ax + b) - b}{ax^2 + bx + c} = \frac{1}{2a}\frac{2ax + b}{ax^2 + bx + c} - \frac{b}{2a}\frac{1}{ax^2 + bx + c} \tag{6}$$

gives the integral formula

$$\int \frac{x}{X}\,dx = \frac{1}{2a} \log |X| - \frac{b}{2a} \int \frac{1}{X}\,dx. \tag{7}$$

When X is defined by (2) and $b^2 - 4ac \neq 0$ and $n \neq 1$, differentiation and simplification give

$$\frac{d}{dx}\frac{2ax + b}{X^{n-1}} = -\frac{2a(2n - 3)}{X^{n-1}} - \frac{(n - 1)(b^2 - 4ac)}{X^n} \tag{8}$$

and hence

$$\int \frac{1}{X^n}\,dx = -\frac{2ax + b}{(n - 1)(b^2 - 4ac)X^{n-1}} - \frac{2a(2n - 3)}{(n - 1)(b^2 - 4ac)} \int \frac{1}{X^{n-1}}\,dx. \tag{9}$$

The formulas (7) and (9) are examples of *reduction formulas* that sometimes enable us to express given integrals in terms of other integrals that are more easily evaluated. Persons who like everything in mathematics can easily become interested in (9) and similar formulas that appear in books of tables. But the formulas are rarely used (most people never use them), and with only a twinge of regret we decline to invest our good time in consideration of examples more or less like $\int (x^2 - 5x - 1)^{-3}\,dx$.

2 Supposing that p, q, r are three different constants, determine A, B, C such that

$$\frac{1}{(x-p)(x-q)(x-r)} = \frac{A}{x-p} + \frac{B}{x-q} + \frac{C}{x-r}$$

and use the result to obtain the integral with respect to x of the left member. *Ans.:*

$$\frac{1}{(p-q)(p-r)}\log|x-p| + \frac{1}{(q-p)(q-r)}\log|x-q| + \frac{1}{(r-p)(r-q)}\log|x-r| + c.$$

3 Supposing that $p \neq q$, determine A, B, C such that

$$\frac{1}{(x-p)^2(x-q)} = \frac{A}{x-p} + \frac{B}{(x-p)^2} + \frac{C}{x-q}$$

and use the result to obtain the integral with respect to x of the left member. *Ans.:*

$$-\frac{1}{(q-p)^2}\log|x-p| + \frac{1}{q-p}\frac{1}{x-p} + \frac{1}{(q-p)^2}\log|x-q| + c.$$

4 Obtain the answer to Problem 3 by starting with the tricky calculation

$$\frac{1}{(x-p)^2(x-q)} = \frac{1}{q-p}\frac{q-p}{(x-p)^2(x-q)} = \frac{1}{q-p}\frac{(x-p)-(x-q)}{(x-p)^2(x-q)} = \frac{1}{q-p}\left[\frac{1}{(x-p)(x-q)} - \frac{1}{(x-p)^2}\right].$$

5 Assuming that p, q, and r are different constants, find the partial fraction expansions of the following quotients and check your answers.

(*a*) $\dfrac{x^3}{x^2-4}$ (*b*) $\dfrac{1}{x(x^2-4)}$

(*c*) $\dfrac{x}{(x-1)(x-2)}$ (*d*) $\dfrac{x}{(x-p)(x-q)}$

(*e*) $\dfrac{x}{(x-1)(x-2)(x-3)}$ (*f*) $\dfrac{x}{(x-p)(x-q)(x-r)}$

(*g*) $\dfrac{x^2}{(x-1)(x-2)(x-3)}$ (*h*) $\dfrac{x^2}{(x-p)(x-q)(x-r)}$

(*i*) $\dfrac{x}{(x-1)(x-2)^2}$ (*j*) $\dfrac{x}{(x-p)(x-q)^2}$

6 Show that

$$\int_1^\infty \frac{1}{t(1+t)^2}\,dt = \log 2 - \frac{1}{2}.$$

7 Show that

(1) $$\int_{-\infty}^\infty \frac{1}{(x^2+1)(x^2+4)}\,dx = \frac{\pi}{6}.$$

Remark: Determination of constants A, B, C, D for which

$$\frac{1}{(x^2+1)(x^2+4)} = \frac{Ax+B}{x^2+1} + \frac{Cx+D}{x^2+4}$$

or

$$1 = (Ax+B)(x^2+4) + (Cx+D)(x^2+1) \tag{2}$$

can be made in various ways. Particularly efficient work can be done by those fortunate persons who know about the algebra of complex numbers, including the "imaginary unit" i for which $i^2 = -1$. Putting $x = i$ in (2) gives $1 = (Ai + B)(3)$, so $A = 0$ and $B = \frac{1}{3}$. Putting $x = 2i$ in (2) gives $1 = (2Ci + D)(-3)$, so $C = 0$ and $D = -\frac{1}{3}$. Hence

$$\frac{1}{(x^2+1)(x^2+4)} = \frac{1}{3}\frac{1}{x^2+1} - \frac{1}{3}\frac{1}{x^2+4},$$

a result that is easily checked.

8 Obtain the partial fraction expansion of the left member of the formula

$$\frac{x^3}{(x+p)^5} = \frac{[(x+p)-p]^3}{(x+p)^5} = \frac{(x+p)^3 - 3p(x+p)^2 + \cdots}{(x+p)^5} = \cdots$$

by taking advantage of the broad hint in the formula.

9 We should not be too busy to see how Euler determined the constants in

$$\frac{x^2}{(1-x)^3(1+x^2)} = \frac{A}{(1-x)^3} + \frac{B}{(1-x)^2} + \frac{C}{1-x} + \frac{Dx+E}{1+x^2} \tag{1}$$

in his great textbook "Introductio in Analysin Infinitorum," Lausannae, 1748, volume 1, page 31. Clear of fractions to obtain

$$\begin{aligned} x^2 = A(1+x^2) &+ B(1-x)(1+x^2) \\ &+ C(1-x)^2(1+x^2) + (Dx+E)(1-x)^3. \end{aligned} \tag{2}$$

Put $x = 1$ to obtain $A = \frac{1}{2}$. Subtract $\frac{1}{2}(1 + x^2)$ from both members of (2) and divide by $(1 - x)$ to obtain

$$-\tfrac{1}{2}x - \tfrac{1}{2} = B(1+x^2) + C(1-x)(1+x^2) + (Dx+E)(1-x)^2. \tag{3}$$

Put $x = 1$ to obtain $B = -\frac{1}{2}$. Add $\frac{1}{2}(1 + x^2)$ and divide by $(1 - x)$ to obtain

$$-\tfrac{1}{2}x = C(1+x^2) + (Dx+E)(1-x). \tag{4}$$

Put $x = 1$ to obtain $C = -\frac{1}{4}$. Add $\frac{1}{4}(1 + x^2)$ and divide by $(1 - x)$ to obtain

$$\tfrac{1}{4} - \tfrac{1}{4}x = Dx + E, \tag{5}$$

and hence $D = -\frac{1}{4}$, $E = \frac{1}{4}$.

10 One who wishes to think about a partial fraction problem in arithmetic (or theory of numbers) may seek integers A and B for which

$$\frac{11}{15} = \frac{A}{3} + \frac{B}{5}.$$

11 Partial fraction expansions have their principal applications in electrical engineering and elsewhere where complex numbers (numbers of the form $a + ib$, where a and b are real and i is the imaginary unit for which $i^2 = -1$) are systematically used. When complex numbers are used, it is not necessary to bother with factors (like $x^2 + 1$ and $x^2 + x + 1$) that cannot be factored into real linear factors. Theory and applications can be based upon a version of the *fundamental theorem of algebra* which says that if

$$P(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n, \tag{1}$$

where n is a positive integer and $a_0 \neq 0$, then there exist k different numbers $x_1, x_2, \cdots, x_k$ (which are not necessarily real) and k exponents $p_1, p_2, \cdots, p_k$ (which are positive integers) such that

$$p_1 + p_2 + \cdots + p_k = n \tag{2}$$

and

$$P(x) = a_0(x - x_1)^{p_1}(x - x_2)^{p_2} \cdots (x - x_k)^{p_k}. \tag{3}$$

This theorem is a simple corollary of general theorems that appear in the theory of functions of a complex variable. Less sophisticated but more complicated proofs can be given. The numbers $x_1, x_2, \cdots, x_k$ are said to be *zeros* of the polynomial of *multiplicities* $p_1, p_2, \cdots, p_k$. It can be proved that if Q is a polynomial of degree less than that of P, then $Q(x)/P(x)$ is representable in the form

$$\begin{aligned} \frac{Q(x)}{P(x)} &= \frac{A_{1,1}}{(x - x_1)} + \frac{A_{1,2}}{(x - x_1)^2} + \frac{A_{1,3}}{(x - x_1)^3} + \cdots + \frac{A_{1,p_1}}{(x - x_1)^{p_1}} \\ &+ \frac{A_{2,1}}{(x - x_2)} + \frac{A_{2,2}}{(x - x_2)^2} + \frac{A_{2,3}}{(x - x_2)^3} + \cdots + \frac{A_{2,p_2}}{(x - x_2)^{p_2}} \\ &+ \cdots \\ &\qquad + \frac{A_{k,1}}{(x - x_k)} + \frac{A_{k,2}}{(x - x_k)^2} + \frac{A_{k,3}}{(x - x_k)^3} + \cdots + \frac{A_{k,p_k}}{(x - x_k)^{p_k}} \end{aligned} \tag{4}$$

where the p_1 constants $A_{1,1}, A_{1,2}, \cdots, A_{1,p_1}$ "go with the powers of the factor $(x - x_1)$," the p_2 constants $A_{2,1}, A_{2,2}, \cdots, A_{2,p_2}$ "go with the powers of the factor $(x - x_2)$," etcetera. Supply and demand do not generate heavy traffic in proofs of this partial fraction theorem, but we can pause to look at the three potent identities

$$\frac{x}{x - a}\frac{R(x)}{S(x)} = \frac{(x - a) + a}{x - a}\frac{R(x)}{S(x)} \tag{5}$$

$$\frac{1}{(x - a)(x - b)}\frac{R(x)}{S(x)} = \frac{1}{b - a}\frac{(x - a) - (x - b)}{(x - a)(x - b)}\frac{R(x)}{S(x)} \tag{6}$$

and

$$\frac{x}{(x - a)(x - b)}\frac{R(x)}{S(x)} = \frac{1}{b - a}\frac{b(x - a) - a(x - b)}{(x - a)(x - b)}\frac{R(x)}{S(x)} \tag{7}$$

in which it is supposed that a and b are real or complex numbers for which $b \neq a$. For example, use of (7) gives

$$\frac{x^5}{(x - x_1)(x - x_2)^2(x - x_3)^3} = \frac{x}{(x - x_1)(x - x_2)} \frac{x^4}{(x - x_2)(x - x_3)^3} \tag{8}$$
$$= \frac{1}{x_2 - x_1} \frac{x_2(x - x_1) - x_1(x - x_2)}{(x - x_1)(x - x_2)} \frac{x^4}{(x - x_2)(x - x_3)^3}$$
$$= \frac{1}{x_2 - x_1}\left[\frac{x_2}{x - x_2} - \frac{x_1}{x - x_1}\right] \frac{x^4}{(x - x_2)(x - x_3)^3}$$
$$= \frac{x_2}{x_2 - x_1} \frac{x^4}{(x - x_2)^2(x - x_3^3)} - \frac{x_1}{x_2 - x_1} \frac{x^4}{(x - x_1)(x - x_2)(x - x_3)^3}$$

Even when $Q(x)$ and $R(x)$ are values of functions that are not polynomials, (5), (6), and (7) enable us to express quotients as sums of simpler quotients. It is quite easy to see that the partial fraction theorem can be proved by repeated applications of (5), (6), and (7). Moreover, it is sometimes better to use (5) and (6) and (7) than to use other methods for obtaining partial fraction expansions. Textbooks in algebra show that if the coefficients $a_0, a_1, \cdots, a_n$ in (1) are real, then the zeros of the polynomial "come in conjugate pairs." This means that if $p + iq$ is a zero for which p and q are real and $q \neq 0$, then $p - iq$ is another zero. On account of this fact the right side of (4) can, when P is real, be represented as a sum of terms of the forms

$$\frac{B}{(x - C)^m}, \quad \frac{D}{(x^2 + Ex + F)^m}, \quad \frac{Gx}{(x^2 + Ex + F)^m}, \tag{9}$$

where m is a positive integer and the coefficients in the denominators are all real. The quadratic denominators are all real. The quadratic denominators arise because if H_1 and H_2 are constants, then there exist other constants H_3 and H_4 for which

$$\frac{H_1}{x - (p + iq)} - \frac{H_2}{x - (p - iq)} = \frac{H_3 + H_4 x}{x^2 - 2px + p^2 + q^2}. \tag{10}$$

In most practical applications of this material, the numbers $p_1, p_2, \ldots, p_k$ defined above are all 1 and (3) and (4) reduce to the much simpler formulas

$$P(x) = a_0(x - x_1)(x - x_2) \cdots (x - x_n) \tag{11}$$

$$\frac{Q(x)}{P(x)} = \frac{A_1}{x - x_1} + \frac{A_2}{x - x_2} + \cdots + \frac{A_n}{x - x_n}. \tag{12}$$

Except when $P(x)$ has only real zeros, we must use complex numbers to achieve simplicity.

9.5 ***Integration by parts*** In this book, and elsewhere in the scientific world, we frequently encounter situations where effective use can be made of the formula for integration by parts. Assuming that u and v

are functions having continuous derivatives over intervals appearing in our work, the formula which can be put in the forms

$$\frac{d}{dx}uv = u\frac{dv}{dx} + v\frac{du}{dx}, \qquad d(uv) = u\,dv + v\,du$$
$$u\frac{dv}{dx} = \frac{duv}{dx} - v\frac{du}{dx}, \qquad u\,dv = d(uv) - v\,du$$

implies the formula which can be put in the forms

$$\int u\frac{dv}{dx}\,dx = uv - \int v\frac{du}{dx}\,dx, \qquad \int u\,dv = uv - \int v\,du \tag{9.51}$$
$$\int u(x)v'(x)\,dx = u(x)v(x) - \int v(x)u'(x)\,dx \tag{9.52}$$

which involve the notations of Leibniz and Newton. We can prefer to use (9.52) when meanings of symbols are being explained and to use (9.51) when problems are being solved and the abbreviated notation expedites our work without confusing us.

We have already seen some of the reasons why the formula (9.51) or (9.52) for integration by parts is useful. As was pointed out following (8.483), efficient use of the formula is made by writing

$$u = u(x), \qquad dv = v'(x)\,dx \tag{9.53}$$
$$du = u'(x)\,dx, \qquad v = \int v'(x)\,dx = v(x), \tag{9.54}$$

where $u(x)$ and $v'(x)$ are chosen in such a way that the product $u(x)v'(x)$ is the integrand in the integral we wish to study. The integral of the product of the things in (9.53) is then the product of the things on the main diagonal minus the integral of the product of the things in (9.54).

The formula for integration by parts has so many applications that it is quite hopeless to undertake to tell when and how it is useful. In many (but not all) situations, the formula is useful when u and v' are chosen in such a way that $\int v'(x)\,dx$ is an elementary integral and the integral $\int v(x)u'(x)\,dx$ is simpler than the integral $\int u(x)v'(x)\,dx$. Our examples and problems will provide some ideas and information. Meanwhile, our guiding principle merits repetition. If we want to learn something about an integral and other methods fail to be helpful, we try integration by parts. Before turning to examples, we make a final observation. In order to apply the formula (9.52) for integration by parts, we need just one pair of functions u and v for which $u(x)v'(x)$ is a given integrand. It is therefore not necessary to insert an added constant c_1 of integration when we write a function v whose derivative is v'. One who wishes to do so may see what happens when we replace $v = -e^{-x}$ by $v = -e^{-x} + c_1$ in the following example. There are relatively few situations in which matters are simplified by inserting a c_1 which is different from 0.

Letting

$$I_1 = \int xe^{-x}\,dx, \tag{9.55}$$

we set

$$\begin{aligned} u &= x, & dv &= e^{-x}\,dx \\ du &= 1\,dx, & v &= \int e^{-x}\,dx = -e^{-x} \end{aligned}$$

and conclude that

$$I_1 = -xe^{-x} + \int e^{-x}\,dx = -xe^{-x} - e^{-x} + c.$$

The same idea is useful when n is a positive integer and

$$I_n = \int x^n e^{-x}\,dx. \tag{9.56}$$

Setting

$$\begin{aligned} u &= x^n, & dv &= e^{-x}\,dx \\ du &= nx^{n-1}\,dx, & v &= \int e^{-x}\,dx = -e^{-x} \end{aligned}$$

gives the reduction formula

$$I_n = -x^n e^{-x} + n\int x^{n-1}e^{-x}\,dx$$

which expresses I_n in terms of I_{n-1}. In particular,

$$\int x^2 e^{-x}\,dx = -x^2e^{-x} - 2xe^{-x} - 2e^{-x} + c.$$

If n is a positive integer and

$$J_n = \int x^n \sin x\,dx, \tag{9.57}$$

setting

$$\begin{aligned} u &= x^n, & dv &= \sin x\,dx \\ du &= nx^{n-1}\,dx, & v &= \int \sin x\,dx = -\cos x \end{aligned}$$

gives the formula

$$J_n = -x^n \cos x + n\int x^{n-1}\cos x\,dx.$$

If $n = 1$, the last integral is easily evaluated. In case $n > 1$, we can integrate by parts again. Setting

$$\begin{aligned} u &= x^{n-1}, & dv &= \cos x\,dx \\ du &= (n-1)x^{n-2}\,dx, & v &= \int \cos x\,dx = \sin x \end{aligned}$$

gives

$$J_n = -x^n\cos x + nx^{n-1}\sin x - n(n-1)\int x^{n-2}\sin x\,dx.$$

In particular,

$$\int x^2 \sin x\,dx = -x^2\cos x + 2x\sin x + 2\cos x + c.$$

We could feel that integration by parts would not enable us to simplify $\int \log x\, dx$, but we can set

$$u = \log x, \qquad dv = 1\, dx$$
$$du = \frac{1}{x}\, dx, \qquad v = x$$

to obtain

$$\int \log x\, dx = x \log x - \int 1\, dx = x \log x - x + c.$$

This result is easily checked by differentiation.

Problems 9.59

1 Derive the following formulas by one or more integrations by parts

(*a*) $\displaystyle\int xe^{ax}\, dx = \left(\frac{x}{a} - \frac{1}{a^2}\right) e^{ax} + c$

(*b*) $\displaystyle\int x \log x\, dx = \frac{x^2}{2} \log x - \frac{x^2}{4} + c$

(*c*) $\displaystyle\int x^n \log x\, dx = \frac{x^{n+1}}{n+1}\left(\log x - \frac{1}{n+1}\right) + c$

(*d*) $\displaystyle\int x^2 e^{ax}\, dx = \left(\frac{x^2}{a} - \frac{2x}{a^2} + \frac{2}{a^3}\right) e^{ax} + c$

(*e*) $\displaystyle\int x \sec^2 x\, dx = x \tan x + \log \cos x + c$

(*f*) $\displaystyle\int (\log x)^n\, dx = x(\log x)^n - n \int (\log x)^{n-1}\, dx$

(*g*) $\displaystyle\int \frac{\log x}{(x+1)^2}\, dx = \frac{x \log x}{x+1} - \log (x+1) + c$

(*h*) $\displaystyle\int \sin^{-1} x\, dx = x \sin^{-1} x + \sqrt{1 - x^2} + c$

(*i*) $\displaystyle\int \tan^{-1} x\, dx = x \tan^{-1} x - \tfrac{1}{2} \log (1 + x^2) + c$

2 Setting $u = t^p$, $dv = (1-t)^q\, dt$, derive the formula

$$\int_0^1 t^p(1-t)^q\, dt = \frac{p}{q+1} \int_0^1 t^{p-1}(1-t)^{q+1}\, dt.$$

Observe the fact that the result agrees with the beta integral formula

$$\int_0^1 t^p(1-t)^q\, dt = \frac{p!q!}{(p+q+1)!}$$

and the formula $r! = r[(r-1)!]$. *Remark:* In Problems 18 and 19 of Problems 13.49, we shall introduce the *Euler gamma integral formula*

$$z! = \int_0^\infty t^z e^{-t}\, dt \qquad (z > -1)$$

and show how it is used to derive the Euler beta integral formula.

3 Derive the formula

$$\int x^n \cos x\,dx = x^n \sin x + nx^{n-1} \cos x - n(n-1)\int x^{n-2} \cos x\,dx.$$

Use this formula to find $\int x^2 \cos x\,dx$, and check the result by differentiation. This formula and those of the next two problems are *reduction formulas*. In some cases, useful results are obtained by making repeated application of them.

4 Letting

$$I(p,q) = \int \sin^p x \cos^q x\,dx,$$

where p and q are constants for which $p \neq -1$ and $p + q \neq 0$, show that

$$I(p,q) = \frac{\sin^{p+1} x \cos^{q-1} x}{p+1} + \frac{q-1}{p+1} \int \sin^{p+2} x \cos^{q-2} x\,dx$$

and

$$I(p,q) = \frac{\sin^{p+1} x \cos^{q-1} x}{p+q} + \frac{q-1}{p+q} \int \sin^{p} x \cos^{q-2} x\,dx.$$

Hint: Start by writing $u = \cos^{q-1} x$, $dv = \sin^p x \cos x\,dx$.

5 Letting

$$W(p,q) = \int \sin^p x \cos^q x\,dx,$$

where p and q are constants for which $q \neq -1$ and $p + q \neq 0$, show that

$$W(p,q) = -\frac{\sin^{p-1} x \cos^{q+1} x}{q+1} + \frac{p-1}{q+1} \int \sin^{p-2} x \cos^{q+2} x\,dx$$

and

$$W(p,q) = -\frac{\sin^{p-1} x \cos^{q+1} x}{p+q} + \frac{p-1}{p+q} \int \sin^{p-2} x \cos^{q} x\,dx.$$

Hint: Start by writing $u = \sin^{p-1} x$, $dv = \cos^q x \sin x\,dx$.

6 Supposing that p is an integer for which $p \geqq 2$, show how a result of the preceding problem can be used to obtain the formula

$$\text{(1)} \qquad \int_0^{\pi/2} \sin^p x\,dx = \frac{p-1}{p} \int_0^{\pi/2} \sin^{p-2} x\,dx.$$

Supposing that n is a positive integer, show how repeated applications of (1) give the famous *Wallis* (1616–1703) *formulas*

$$\text{(2)} \qquad \int_0^{\pi/2} \sin^{2n} x\,dx = \frac{2n-1}{2n}\cdot\frac{2n-3}{2n-2}\cdots\frac{5}{6}\cdot\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{\pi}{2}$$

$$\text{(3)} \qquad \int_0^{\pi/2} \sin^{2n+1} x\,dx = \frac{2n}{2n+1}\cdot\frac{2n-2}{2n-1}\cdots\frac{6}{7}\cdot\frac{4}{5}\cdot\frac{2}{3}.$$

Observe that

$$\text{(4)} \qquad 2n(2n-2)\cdots 6\cdot 4\cdot 2 = 2^n n!$$

and show that multiplying the numerators and denominators of the right members of (2) and (3) by the left member of (4) gives the formulas

$$(5)\qquad \int_0^{\pi/2} \sin^{2n} x\,dx = \frac{(2n)!}{2^{2n}n!n!}\cdot\frac{\pi}{2}$$

$$(6)\qquad \int_0^{\pi/2} \sin^{2n+1} x\,dx = \frac{2^{2n}n!n!}{(2n)!}\cdot\frac{1}{2n+1}.$$

One reason for interest in these things lies in the fact that the number

$$\frac{(2n)!}{2^{2n}n!n!}$$

is the probability of finding exactly n heads and exactly n tails when $2n$ coins are tossed. *Remark:* We embark on a little excursion to see that these formulas have startling consequences. When $0 < x < \pi/2$, we have $0 < \sin x < 1$, so

$$(7)\qquad 0 < \sin^{2n+1} x < \sin^{2n} x < \sin^{2n-1} x < 1$$

and hence

$$(8)\qquad 0 < \int_0^{\pi/2} \sin^{2n+1} x\,dx < \int_0^{\pi/2} \sin^{2n} x\,dx < \int_0^{\pi/2} \sin^{2n-1} x\,dx.$$

Putting $p = 2n + 1$ in (1) gives the formula

$$(9)\qquad \int_0^{\pi/2} \sin^{2n-1} x\,dx = \left(1 + \frac{1}{2n}\right)\int_0^{\pi/2} \sin^{2n+1} x\,dx.$$

It follows from (8) and (9) that there is a number θ_n for which $0 < \theta_n < 1$ and

$$(10)\qquad \int_0^{\pi/2} \sin^{2n} x\,dx = \left(1 + \frac{\theta_n}{2n}\right)\int_0^{\pi/2} \sin^{2n+1} x\,dx.$$

Multiplying the members of (5) and (6) gives

$$(11)\qquad \left[\int_0^{\pi/2} \sin^{2n} x\,dx\right]\left[\int_0^{\pi/2} \sin^{2n+1} x\,dx\right] = \frac{\pi}{4n\left(1 + \frac{1}{2n}\right)}.$$

Multiplying the members of (10) and (11) leads to the formula

$$(12)\qquad \int_0^{\pi/2} \sin^{2n} x\,dx = \sqrt{1 - \frac{1-\theta_n}{2n+1}}\,\frac{\sqrt{\pi}}{2\sqrt{n}}.$$

Substituting this in (5) gives the formula

$$(13)\qquad \frac{(2n)!}{2^{2n}n!n!} = \sqrt{1 - \frac{1-\theta_n}{2n+1}}\,\frac{1}{\sqrt{n\pi}}.$$

Since $0 < \theta_n < 1$, the first factor in the right member is near 1 when n is large and is quite near 1 even when n is as small as 4 or 3. Thus, even when n is quite small, the number $1/\sqrt{n\pi}$ is a very good approximation to the probability of finding exactly n heads and exactly n tails when $2n$ coins are tossed. This astonishing result has very significant consequences, and persons who comprehend it may say that mathematics was never so beautiful before, and mathematics really is the Queen of the Sciences.

7 Derive the formula

$$\int e^{ax} \sin bx\, dx = \frac{e^{ax}(a \sin bx - b \cos bx)}{a^2 + b^2} + c \tag{1}$$

in the following way. Set

$$u = e^{ax}, \qquad dv = \sin bx\, dx$$

to obtain

$$\int e^{ax} \sin bx\, dx = -\frac{e^{ax} \cos bx}{b} + \frac{a}{b} \int e^{ax} \cos bx\, dx. \tag{2}$$

Then set

$$u = \sin bx, \qquad dv = e^{ax}\, dx$$

to obtain

$$\int e^{ax} \sin bx\, dx = \frac{e^{ax} \sin bx}{a} - \frac{b}{a} \int e^{ax} \cos bx\, dx. \tag{3}$$

Then combine (2) and (3) to obtain (1). Now derive (1) in the following way. Let I denote the left members of (1), (2), and (3). Set

$$u = e^{ax}, \qquad dv = \cos bx\, dx$$

to put (2) in the form

$$I = -\frac{e^{ax} \cos bx}{b} + \frac{a}{b}\left[\frac{e^{ax} \sin bx}{b} - \frac{a}{b} I\right]. \tag{4}$$

Then solve (4) for I and obtain (1).

8 Derive the formula

$$\int e^{ax} \cos bx\, dx = \frac{e^{ax}(b \sin bx + a \cos bx)}{a^2 + b^2} + c.$$

•*9* Sketch graphs of $y = e^x$ and $y = \log x$ in the same figure. Let R_1 be the region consisting of points (x,y) for which $x \leqq 0$ and $0 \leqq y \leqq e^x$. This unbounded region R_1 is said to be bounded by (or to have boundaries) the x axis, the y axis, and the graph of $y = e^x$. Let R_2 be the region consisting of points (x,y) for which $x = 0$ and $y \leqq 0$ together with points (x,y) for which $0 < x \leqq 1$ and $\log x \leqq y \leqq 0$. This region R_2 is said to be bounded by the x axis, the y axis, and the graph of $y = \log x$. Observe that R_1 and R_2 are congruent regions. Use the fact that R_1 and R_2 possess areas $|R_1|$ and $|R_2|$ for which

$$|R_1| = \lim_{h \to -\infty} \int_h^0 e^{-x}\, dx, \qquad |R_2| = \lim_{h \to 0} \int_h^1 (-\log x)\, dx$$

to show that $|R_1| = |R_2| = 1$.

10 Evaluate the integral in

$$I = \int \frac{x^3}{\sqrt{1 + x^2}}\,dx$$

in two different ways, and make the results agree. First, use the identity

$$\frac{x^3}{\sqrt{1+x^2}} = \frac{x(1 + x^2 - 1)}{\sqrt{1+x^2}} = x\sqrt{1+x^2} - \frac{x}{\sqrt{1+x^2}}.$$

Second, integrate by parts with

$$u = x^2, \qquad dv = x(1 + x^2)^{-1/2}\,dx.$$

11 With the aid of the substitution (or change of variable) $\sqrt{a - x} = t$, show that

$$\int_0^a x^2 \sqrt{a - x}\,dx = 2\int_0^{\sqrt{a}} (a - t^2)^2 t^2\,dt = \tfrac{16}{105}a^{7/2}.$$

Show that the first integral can be evaluated by integration by parts.

12 Derive the formulas

$$\int_x^\infty e^{-t^2/2}\,dt = x^{-1}e^{-x^2/2} - \int_x^\infty t^{-2}e^{-t^2/2}\,dt$$

$$\int_x^\infty e^{-t^2/2}\,dt = x^{-1}e^{-x^2/2} - x^{-3}e^{-x^2/2} + 3\int_x^\infty t^{-4}e^{-t^2/2}\,dt.$$

Remark: Formulas of this nature give useful information, the integrals on the right being small in comparison to those on the left when x exceeds 2 or 3 or 4.

13 Many problems in pure and applied mathematics involve "best fit" or "best approximation" in some sense or other. We can start picking up ideas by observing that if λ is near 1 and

$$(1) \qquad f(x) = \cos x,\ g(x) = \left(1 - \frac{4x^2}{\pi^2}\right)$$

then the graphs of $f(x)$ and $\lambda g(x)$ over the interval $-\pi/2 \leqq x \leqq \pi/2$ look much alike. The graph of $\lambda g(x)$ is the *best fit in the sense of least squares* to the graph of $f(x)$ when λ is chosen such that

$$(2) \qquad \int_{-\pi/2}^{\pi/2} [f(x) - \lambda g(x)]^2\,dx$$

is a minimum. Show that (2) will be a minimum when

$$(3) \qquad \lambda \int_0^{\pi/2} [g(x)]^2\,dx = \int_0^{\pi/2} f(x)g(x)\,dx$$

and hence when $\lambda = 30/\pi^3 = 0.967546$. *Remark:* Sketching graphs of $f(x)$ and $g(x)$ on a rather large scale indicates that $g(x) > f(x)$ when $0 < x < \pi/2$. In fact, we can put

$$(4) \qquad F(x) = g(x) - f(x)$$

and show that $F(0) = F(\pi/2) = 0$, F is increasing over the part of the interval $0 < x < \pi/2$ for which $\sin x > (8/\pi^2)x$, and F is decreasing over the remaining part of the interval.

10 Polar, cylindrical, and spherical coordinates

10.1 Geometry of coordinate systems We begin with a glimpse of a (or the) major reason why polar coordinates should be studied. In many problems involving functions defined over all or portions of E_3, there is a line L which is particularly significant. This line L may, for example, be an axis of symmetry or a wire carrying an electric current or charge. When we want to use coordinates, we can let the line L be the z axis of a rectangular x, y, z coordinate system as in Figure 10.11. When we are interested in the cylinder consisting of points at a particular distance ρ_0 from the z axis, we can correctly describe this set as being the set of points having rectangular coordinates x, y, z for which $x^2 + y^2 = \rho_0^2$. It is, however, much simpler to take the distance ρ from the z axis to a point P to be one of the coordinates of P so that the equation of the cylinder is simply $\rho = \rho_0$. This one coordinate ρ is, however, not enough to determine the position of a point P. It turns out to be convenient to determine the position of P by use of ρ, the angle ϕ, and the rectangular coordinate z of Figure 10.11. The three numbers ρ, ϕ, z are called *cylindrical coordinates* of P.

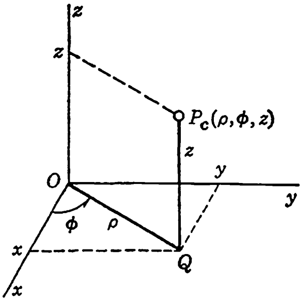


Figure 10.11

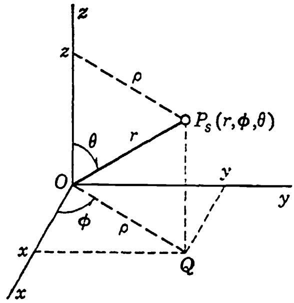


Figure 10.12

In many other problems involving functions defined over all or portions of E_3, there is a point P_0 (instead of a line L) which is particularly significant. This point P_0 may, for example, be the center of a spherical or nonspherical earth or may be a point at which an electric charge is supposed to be concentrated. When we want to use coordinates, we can let P_0 be the origin O of a rectangular x, y, z coordinate system as in Figure 10.12. When we are interested in the sphere consisting of points at a particular distance r_0 from the origin, we can correctly describe this set as being the set of points having rectangular coordinates x, y, z for which $x^2 + y^2 + z^2 = r_0^2$. It is, however, much simpler to take the distance r from the origin to a point P to be one of the coordinates of P so that the equation of the sphere is simply $r = r_0$. This one coordinate r is, however, not enough to determine the position of a point P. It turns out to be convenient to determine the position of P by use of r, the angle ϕ, and the angle θ of Figure 10.12. The three numbers r, ϕ, θ are called *spherical coordinates* of P.

We really should write $P_r(x,y,z)$, $P_c(\rho,\phi,z)$, and $P_s(r,\phi,\theta)$ to denote, respectively, the points having rectangular coordinates x, y, z, cylindrical coordinates ρ, ϕ, z, and spherical coordinates r, ϕ, θ. The coordinates $x, y, z, \rho, \phi, r, \theta$ are numbers, and it is precarious to confuse relations among coordinates and functions of these coordinates by allowing $P(0.7,0.6,0.5)$ and $f(0.7,0.6,0.5)$ to have ambiguous meanings. The subscripts are included in Figures 10.11 and 10.12 because they should be included. To be reasonable about this matter, we can allow $P(a,b,c)$ to be the point having cylindrical coordinates a, b, c while we are solving a problem or constructing a theory in which cylindrical coordinates and no other coordinates appear. We eliminate confusion, however, by agreeing that, except in special situations where there is an explicit agreement to the contrary, $P(a,b,c)$ is always the point having rectangular coordinates a, b, c.

In ordinary useful applications of the cylindrical coordinates of Figure 10.11, it is always supposed that $\rho \geqq 0$ and it is sometimes supposed that $\rho > 0$ and $-\pi < \phi \leqq \pi$. Sometimes the latter restriction on ϕ is removed so that ϕ can vary continuously as a particle having cylindrical coordinates (ρ,ϕ,z) makes excursions around the z axis. In those situations in which it is supposed that $\rho > 0$ or $\rho \geqq 0$, the graph in cylindrical coordinates of the equation $\phi = \phi_0$ is a half-plane (not a whole plane) having an edge on the z axis. In those situations in which ϕ is unrestricted, a point P does not determine its cylindrical coordinates because the two points

$$P_c(\rho,\ \phi + 2n\pi,\ z), \qquad P_c(\rho,\ \phi,\ z)$$

are identical when n is an integer. In ordinary useful applications of spherical coordinates, it is always supposed that $r \geqq 0$ and is sometimes supposed that $r > 0$, $0 \leqq \theta \leqq \pi$, and $-\pi < \phi \leqq \pi$. In ordinary geographical terms the coordinate θ, which is 0 when $P_s(r,\phi,\theta)$ is at the north pole and is $\pi/2$ when $P_s(r,\phi,\theta)$ is on the equator and is π when $P_s(r,\phi,\theta)$ is at the south pole, determines the *latitude* of $P_s(r,\phi,\theta)$. The coordinate ϕ determines the *longitude*.

Partly because the endeavor helps us to understand cylindrical and spherical coordinates, we turn to the study of polar coordinates of points in a plane. The basic idea behind the concept of polar coordinates is both simple and attractive. Suppose we are located at a point O, an origin or *pole*, in a plane and we wish to give explicit instructions telling how to make a pilgrimage to a point P in the same plane. We begin by constructing a half-line OA with an end at O as in Figure 10.13 and calling this half-line the *initial line* from which angles are to be measured. In case P is not the origin, instructions for reaching P are now easily given. Start at the origin looking in the direction of the initial line, turn in the positive (counterclockwise) direction through the angle ϕ until facing P, and then travel the appropriate distance ρ from O to P. We could (and sometimes do) end the matter here and say that ρ and ϕ are *the* polar coordinates of $P_p(\rho,\phi)$, the point having polar coordinates ρ and ϕ. Sometimes we restrict ϕ to the domain $-\pi < \phi \leqq \pi$ and end the matter in another way. While recognition of the fact is sometimes irksome, it is nevertheless true that if n is an integer which may be negative as well as positive or zero and if we turn through the angle $2n\pi + \phi$, then we will be facing toward P and can travel the distance ρ to reach P. When we take this possibility into account, we find that, for each $n = 0, \pm 1,$

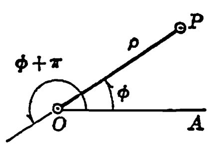


Figure 10.13

$\pm 2, \cdots$, the numbers ρ, $\phi + 2n\pi$ constitute a set of polar coordinates of P. We could (and sometimes do) end the matter here. It is not always easy to know when we are being wise, but we can recognize one more possibility. After turning through the angle $\phi + \pi$ or $\phi - \pi$ or $\phi + (2n + 1)\pi$, where n is an integer, we will be facing away from P and we can reach P by going backwards a distance ρ. When we take this last possibility into account, we find that, for each $n = 0, \pm 1, \pm 2, \cdots$, the numbers $(-\rho, \phi + \pi + 2n\pi)$ constitute a set of polar coordinates of P. When polar coordinates of this variety are permitted to appear in our work, we abandon the idea that ρ is a distance and take ρ to be a coordinate that can be negative. Thus when $\rho < 0$, the point $P_p(\rho,\phi)$ having polar coordinates ρ, ϕ is the same as the point $P_p(|\rho|, \phi + \pi)$ having the more normal polar coordinates $|\rho|$, $\phi + \pi$. We still have to consider the polar coordinates of P when P is the origin O. It turns out to be best to agree that, for each number ϕ, the numbers 0 and ϕ are polar coordinates of the origin.†

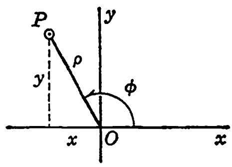


Figure 10.14

Let polar and rectangular coordinate systems be superimposed in such a way that, as in Figure 10.14, the initial line of the former coincides with the nonnegative x axis of the latter. When P is a point different from O, it is easy to obtain formulas relating the rectangular coordinates (x,y) of P and each set (ρ,ϕ) of polar coordinates of P for which $\rho > 0$. The definitions

$$\cos \phi = \frac{x}{\rho}, \qquad \sin \phi = \frac{y}{\rho}$$

of the trigonometric functions give the formulas

$$(10.141) \qquad x = \rho \cos \phi, \qquad y = \rho \sin \phi$$

which uniquely determine x and y in terms of ρ and ϕ. On the other hand, the formulas

$$(10.142) \quad \rho = \sqrt{x^2 + y^2}, \qquad \cos \phi = \frac{x}{\sqrt{x^2 + y^2}}, \qquad \sin \phi = \frac{y}{\sqrt{x^2 + y^2}}$$

uniquely determine ρ in terms of x and y and uniquely determine an angle ϕ_0 such that $-\pi < \phi_0 \leqq \pi$ and ϕ must have the form $\phi_0 + 2n\pi$, where n is an integer. In case $\rho > 0$ and $-\pi/2 < \phi < \pi/2$, the last

† In this chapter, the coordinates ρ, ϕ, r, θ have the classical significance they usually have in mathematical physics and elsewhere when Legendre polynomials and such things appear. Some textbooks use r and θ for polar coordinates and, with a shift in meaning of coordinates, use r, θ, ϕ for spherical coordinates as we do. Sometimes ρ is used for a spherical coordinate. Persons who stray from one book or one classroom to another do not always appreciate modifications of classical notation, but they are rarely if ever seriously injured.

two of the formulas can be replaced by the single formula $\phi = \tan^{-1} y/x$. These formulas enable us to make substitutions which transform formulas involving coordinates of one brand into formulas involving coordinates of the other brand.

The remainder of the text of this section is devoted to development of the art of sketching polar coordinate graphs of given equations for which it is more or less appropriate to operate without restricting ρ to nonnegative values and without restricting ϕ to an interval such as the interval $-\pi < \phi \leqq \pi$. In this situation, the relation between polar graphs and equations is complicated by the fact that each point has an infinite set of polar coordinates and we need a definition. The *polar coordinate graph* of an equation of the form $f(\rho,\phi) = 0$ is the set S consisting of points P each of which has at least one set (ρ,ϕ) of polar coordinates for which $f(\rho,\phi) = 0$. The polar graph of the equation $\rho = -1$ is then the unit circle C with center at the origin because $P_p(-1,\phi)$ is, for each ϕ, the same as the point $P_p(1, \phi + \pi)$. The polar graphs of the two equations $\rho = -1$ and $\rho = 1$ therefore coincide, even though there is clearly no single pair (ρ,ϕ) of numbers for which the two equations are simultaneously satisfied. Figure 10.143 can promote understanding of this matter. When we allow ρ to be negative, the polar graph of the equation $\phi = 0$ is more than the initial half-line of the polar coordinate system; it is the entire line upon which the initial half-line lies.

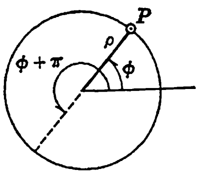


Figure 10.143

Supposing that a is a given positive constant, we undertake to determine the nature of the polar graph of the equation

$$\rho = a \sin 2\phi \tag{10.15}$$

without laboriously locating many points. Observing that $|\rho| \leqq a$ always and $|\rho| = a$ sometimes, we draw the circle of radius a with center at the origin to help us. Our knowledge of the sine tells us that $\sin 2\phi$ increases from 0 to 1 and ρ increases from 0 to a as 2ϕ increases from 0 to $\pi/2$ and hence as ϕ increases from 0 to $\pi/4$. This information enables us to sketch the first part of the first leaf, or loop, of Figure 10.151. Similarly, ρ decreases from a to 0 as 2ϕ increases from $\pi/2$ to π and hence as ϕ increases from $\pi/4$ to $\pi/2$. Now we complete the first leaf. Continuing to decrease, ρ decreases from 0 to $-a$ as 2ϕ increases from π to $3\pi/2$ and hence as ϕ increases from $\pi/2$ to $3\pi/4$. During this operation, the terminal side of the angle ϕ is in the second quadrant and negativeness of ρ throws the graph into the fourth quadrant to give the first half of the second leaf. Then ρ increases from $-a$ to 0 as 2ϕ increases from $3\pi/2$ to 2π and hence as ϕ increases from $3\pi/4$ to π. This gives the second half of the second leaf. Continuing its increase, ρ increases from

0 to a as 2ϕ increases from 2π to $5\pi/2$ and hence as ϕ increases from π to $5\pi/4$. This gives the first half of the third leaf. Three more increases in 2ϕ and ϕ complete the third and fourth leaves as ϕ increases to 2π. Increasing ϕ beyond 2π yields more curve but no more graph, since the graph is retraced. The full graph is shown in Figure 10.151. In the good old days, perhaps before clover was invented and when roses were primitive, someone called this graph the *rose with four leaves*.

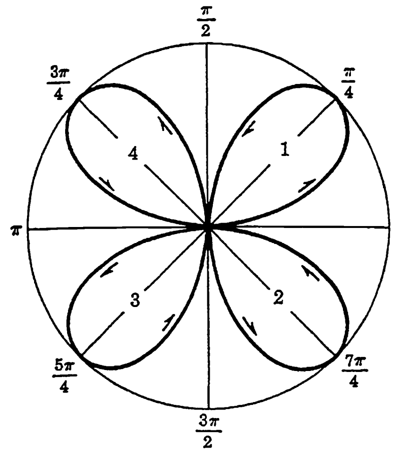


Figure 10.151

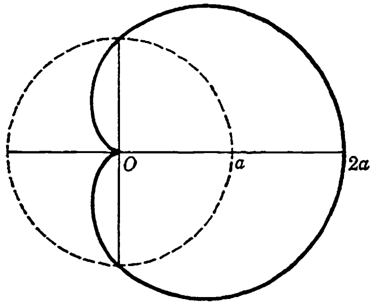


Figure 10.152

The polar graph of the equation

$$\rho = a(1 + \cos \phi) \tag{10.16}$$

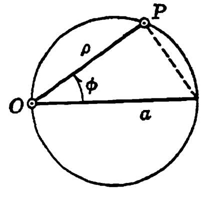
is obtained much more easily. As ϕ increases from 0 to π and then to 2π, ρ decreases from $2a$ to 0 and then increases to $2a$. This graph, which is called a *cardioid*, is shown in Figure 10.152.

Let a be a positive constant. When $0 < \phi < \pi/2$, it follows from elementary geometry and trigonometry that the point P having polar coordinates (ρ,ϕ) for which

$$\rho = a \cos \phi \tag{10.161}$$

lies on the circle of Figure 10.162. Consideration of other angles shows that the circle is the complete graph (which, of course, means *the* graph) of the equation.

Figure 10.162



A bit of novelty appears when we undertake to sketch a graph of the equation

$$\rho^2 = a^2 \cos 2\phi. \tag{10.17}$$

As 2ϕ increases from $-\pi/2$ to 0 and then to $\pi/2$, and hence as ϕ increases from $-\pi/4$ to 0 and then to

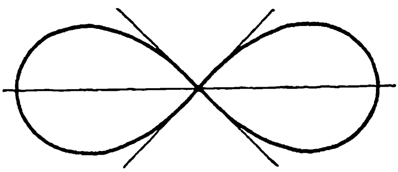


Figure 10.171

$\pi/4$, $a^2 \cos 2\phi$ increases from 0 to a^2 and then decreases to 0. This information enables us to sketch the two loops of Figure 10.171, ρ being positive on one loop and negative on the other. As 2ϕ increases from $\pi/2$ to $3\pi/2$, and hence as ϕ increases from $\pi/4$ to $3\pi/4$, $\cos 2\phi$ is negative and no values of ρ are obtained. Further investigation shows that the graph already drawn is complete. It is called a *lemniscate*.

When $a > 0$, the polar graph of the equation $\rho = a\phi$ is called a *spiral of Archimedes*. This graph is shown in Fig. 10.181, the part for which

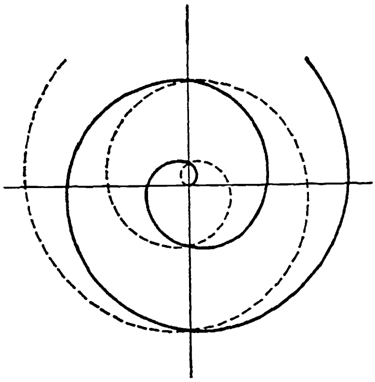


Figure 10.181

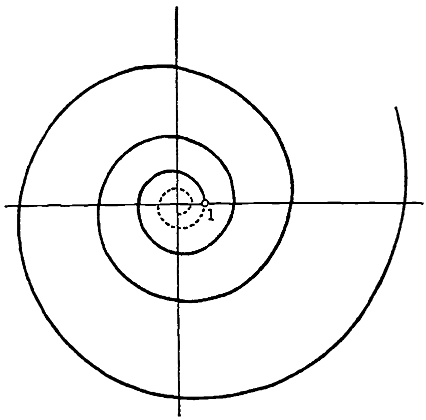


Figure 10.182

$\phi < 0$ being dotted. When spirals are being graphed, and at some other times, the approximations $\pi = 3.1416, \pi/2 = 1.5708, \pi/4 = 0.7854$, and similar others are used. It is very often necessary to know relations akin to the relations 2π radians $= 360°$, π radians $= 180°$, $\pi/2$ radians $= 90°$, and $\pi/4$ radians $= 45°$. It is sometimes useful to know that 1 radian is $180/\pi$ degrees or about 57 degrees, but degrees and minutes and seconds play minor roles in our work. The polar graph of the equation $\rho = e^{a\phi}$ is an *exponential spiral* which is commonly called a logarithmic spiral. The graph is shown in Figure 10.182. When $a > 0$, the dotted part for which $\phi < 0$ spirals inward around the origin.

Problems 10.19

1 With the aid of Figure 10.11, show that the formulas giving the rectangular coordinates x, y, z of a point in terms of the cylindrical coordinates ρ, ϕ, z of the same point are

$$x = \rho \cos \phi, \qquad y = \rho \sin \phi, \qquad z = z.$$

2 With the aid of Figure 10.12, show that the formulas giving the cylindrical coordinates ρ, ϕ, z of a point in terms of the spherical coordinates r, ϕ, θ of the same point are

$$\rho = r \sin \theta, \qquad \phi = \phi, \qquad z = r \cos \theta.$$

3 With the aid of Problems 1 and 2, show that the formulas giving the rectangular coordinates of a point in terms of the spherical coordinates of the same point are

$$x = r \cos \phi \sin \theta, \qquad y = r \sin \phi \sin \theta, \qquad z = r \cos \theta.$$

4 Transform the following equations from rectangular to polar coordinates

(*a*) $x^2 + y^2 = a^2$ *Ans.:* $\rho = a$
(*b*) $x = a$ *Ans.:* $\rho \cos \phi = a$
(*c*) $(\cos \alpha)x + (\sin \alpha)y = a$ *Ans.:* $\rho \cos (\phi - \alpha) = a$
(*d*) $xy = 1$ *Ans.:* $\rho^2 \sin 2\phi = 2$
(*e*) $x^2 - y^2 = a^2$ *Ans.:* $\rho^2 \cos 2\phi = a^2$

5 In Problem 31 of Section 6.4, we said that the rectangular graph of the equation

$$(x^2 + y^2)^2 = a^2(x^2 - y^2)$$

is a lemniscate. Show that the polar equation is $\rho^2 = a^2 \cos 2\phi$. The graph appears in Figure 10.171.

6 Lemniscates have a simple geometric property. Let b be a positive number and let F_1 and F_2 be the points (sometimes called foci) having the rectangular coordinates $(-b,0)$ and $(b,0)$. Let S be the set of points P for which

$$|\overrightarrow{F_1P}|\,|\overrightarrow{F_2P}| = b^2. \tag{1}$$

Show that the rectangular equation of S can be put in the form

$$(x^2 + y^2)^2 = 2b^2(x^2 - y^2) \tag{2}$$

and hence that S is a lemniscate. Show, in one of the various possible ways, that the polar equation of this lemniscate is

$$\rho^2 = 2b^2 \cos 2\phi. \tag{3}$$

7 Transform the following equations from polar to rectangular coordinates.

(*a*) $\rho = 3$ *Ans.:* $x^2 + y^2 = 9$
(*b*) $\rho^2 = a^2 \sin 2\phi$ *Ans.:* $(x^2 + y^2)^2 = 2a^2xy$
(*c*) $\rho = a \cos \phi$ *Ans.:* $x^2 + y^2 = ax$
(*d*) $\rho = 2a(1 - \cos \phi)$ *Ans.:* $(x^2 + y^2 + 2ax)^2 = 4a^2(x^2 + y^2)$

8 Show that when

$$\rho = 2a \cos \phi,$$

the point with polar coordinates (ρ,ϕ) runs once in the positive direction around a curve C as ϕ increases from $-\pi/2$ to $\pi/2$. Show, in one way or another or in more than one way, that C is the circle of radius a having its center at the point with rectangular coordinates $(a,0)$

9 Sketch polar graphs of the equations

(*a*) $\rho = 3 + 2 \cos \phi$ (*b*) $\rho = 3 + 4 \cos \phi$
(*c*) $\rho = a \sec \phi$ (*d*) $\rho = \tan \phi$
(*e*) $\rho = a \sin 3\phi$ (*f*) $\rho = a \sin \dfrac{\phi}{3}$
(*g*) $\rho = 1/\phi, (\phi > 0)$ (*h*) $\rho = 1/(1 + \phi^2), (\phi > 0)$

10 Sketch rectangular graphs of $y = \sqrt{\cos x}$, $y = \cos x$, and $y = \cos^2 x$ in one figure, and then sketch polar graphs of $\rho = \sqrt{\cos \phi}$, $\rho = \cos \phi$, and $\rho = \cos^2 \phi$ in another figure. *Partial solution:* Good polar graphs of $\rho = \cos \phi$ and $\rho = \sqrt{\cos \phi}$ and the right-hand half of the graph of $\rho = \cos^2 \phi$ are shown in Figure 10.191.

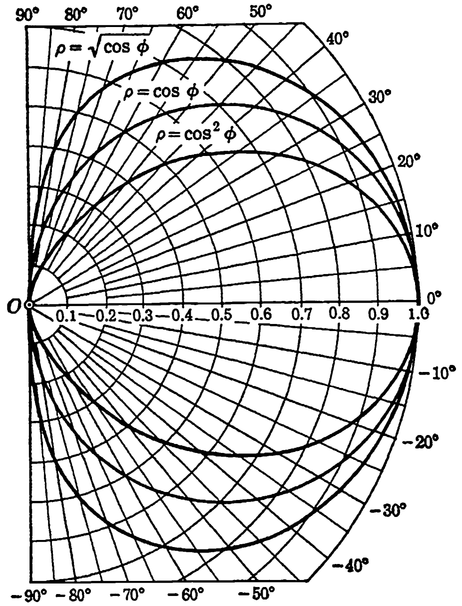


Figure 10.191

11 Sketch rectangular graphs of $x^2 + y^2 = 1$ and $x^2 + y^2 = 25$ in one figure and then sketch polar graphs of $\rho^2 + \phi^2 = 1$ and $\rho^2 + \phi^2 = 25$ in another figure.

12 In case $\rho_1 > 0$ and $\rho_2 > 0$, a straightforward application of the law of cosines gives the formula

$$d^2 = \rho_1^2 + \rho_2^2 - 2\rho_1\rho_2 \cos (\phi_2 - \phi_1)$$

for the square of the distance d between points having polar coordinates (ρ_1, ϕ_1) and (ρ_2, ϕ_2). Show that the formula is also valid when $\rho_1 \leqq 0$ or $\rho_2 \leqq 0$ or both.

13 This is a rather heroic problem that requires careful and accurate applications of rules for differentiation and attention to algebraic details. No vectors or figures or tricks of any kind are to be used; just differentiate and substitute. Using the formulas

$$x(t) = \rho(t) \cos \phi(t), \qquad y(t) = \rho(t) \sin \phi(t)$$

and the formula (7.389) for curvature in rectangular coordinates, show that if ρ and ϕ are functions having two continuous derivatives and if the point $P(t)$ having polar coordinates $\rho(t)$, $\phi(t)$ traverses the curve C as t increases, then the curvature K of C at $P(t)$ is

$$K = \frac{[\rho(t)]^2[\phi'(t)]^3 + \rho(t)\rho'(t)\phi''(t) - \rho(t)\rho''(t)\phi'(t) + 2[\rho'(t)]^2\phi'(t)}{\{[\rho(t)\phi'(t)]^2 + [\rho'(t)]^2\}^{3/2}}$$

provided the denominator is not zero.

14 Show that letting $\phi(t) = t$ and replacing t by ϕ in the last formula of the preceding problem gives the much simpler formula

$$K = \frac{[\rho(\phi)]^2 + 2[\rho'(\phi)]^2 - \rho(\phi)\rho''(\phi)}{\{[\rho(\phi)]^2 + [\rho'(\phi)]^2\}^{3/2}}$$

for the curvature of the graph of the polar equation $\rho = \rho(\phi)$ oriented in the direction of increasing ϕ.

15 The two equations

$$(1) \qquad \sqrt{x^2 + y^2} + |x| + |y| = a$$

$$(2) \qquad \sqrt{x^2 + y^2} + x + y = a,$$

in which a is a given positive constant, have respective graphs G_1 and G_2. Show that the polar coordinate equations of these graphs are

$$(3) \qquad \rho = \frac{a}{1 + |\sin \phi| + |\cos \phi|}$$

and

$$(4) \qquad \rho = \frac{a}{1 + \sin \phi + \cos \phi}$$

or

$$\rho = \frac{a}{1 + \sqrt{2} \sin\left(\phi + \frac{\pi}{4}\right)}$$

with $\rho > 0$. Obtain more or less complete information about G_1 and G_2.

16 The *cissoid of Diocles* is the set of points $P(x,y)$ obtained in the following way. Let $a > 0$. As in Figure 10.192, let C be the circle with center at $(a,0)$ having radius a. Let $0 < x < 2a$. Let Q_1 and Q_2 be points on the upper part of C having x coordinates x and $2a - x$. Then $P(x,y)$ is the intersection of the line OQ_2 and the vertical line through Q_1. Letting ϕ be the angle which OQ_2 makes with the x axis and with the line Q_2Q_1, we see that $y = a$ when $x = a$ and, when $x \neq a$,

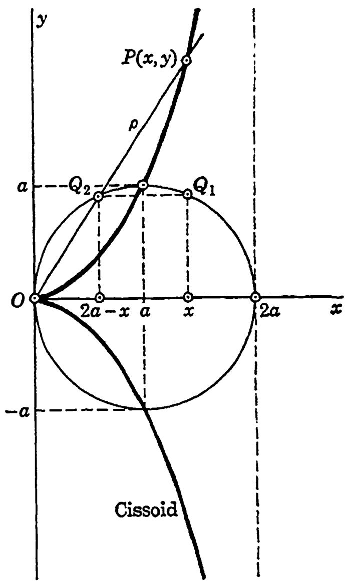


Figure 10.192

$$(1) \qquad \tan \phi = \frac{y}{x} = \frac{|\overrightarrow{Q_1P}|}{|\overrightarrow{Q_2Q_1}|} = \frac{y - \sqrt{a^2 - (x-a)^2}}{2(x-a)}.$$

This gives

$$(2) \qquad (2a - x)y = x\sqrt{x(2a - x)},$$

and squaring gives

$$(3) \qquad y^2 = \frac{x^3}{2a - x}.$$

The graph of this equation, including points for which $x = a$, $x = 0$, and $y < 0$, is the cissoid. We now come to the polar coordinate problem. With the aid of the fact that Q_2 has polar coordinates (ρ_1,ϕ) for which $\rho_1 = 2a \cos \phi$, try to use Figure 10.192 to derive the polar equation of the cissoid. If unsuccessful, use (2) and formulas which give x and y in terms of ρ and ϕ. *Ans.:*

$$(4) \qquad \rho = 2a \sin \phi \tan \phi.$$

Remark: Diocles employed the cissoid to "duplicate a cube," the problem being to start with some line segment of length x (the length of an edge of a particular cube) and construct a segment of length $\sqrt[3]{2}\, x$ (the length of the edge of a cube having double the volume of the original one). The line L through the points $(a,2a)$ and $(2a,0)$ has the equation $(2a - x) = y/2$ and intersects the cissoid at a point (x,y) for which $y^3 = 2x^3$ and hence $y = \sqrt[3]{2}\, x$. What Diocles really wanted to do was duplicate a cube by ruler-and-compass construction. This has been proved to be impossible. It is possible to construct the line L with a ruler and compass, and it is possible to construct points on the cissoid one by one with a ruler and compass. The reason why Diocles failed to accomplish his purpose should be explained. Life is too short to enable us to produce ruler-and-compass constructions of all of the points on the cissoid, and there is no way to prescribe rules for a ruler-and-compass construction of the particular point where L intersects the cissoid.

17 The *conchoids of Nicomedes* provide a method (but not a ruler-and-compass method, because no such method exists) for trisecting angles. Let p and q be given positive numbers. Let O (the pole of the conchoid) be the origin and let L (the directrix of the conchoid) be the line having the rectangular equation $x = p$ and the polar equation $\rho \cos \phi = p$ or $\rho = p \sec \phi$ as in Figure 10.193. The conchoid consists of two parts or branches. When $-\pi/2 < \phi < \pi/2$, the line OM of the figure meets the line L at the point M and meets the far branch (the branch most remote from the pole) at a point P whose distance from M is q and whose distance from the origin is $p \sec \phi + q$. The polar equation of the far branch is therefore

$$(1) \qquad \rho = p \sec \phi + q.$$

The same line OM meets the near branch (the branch nearer to the pole) at a point P' whose distance from M is q and whose polar equation is

$$(2) \qquad \rho = p \sec \phi - q.$$

The equation

$$(3) \qquad \rho = p \sec \phi \pm q$$

or the equation

$$(4) \qquad (\rho - p \sec \phi)^2 = q^2$$

is the polar equation of the conchoid The graph consisting of the two solid branches nearest the line L is a conchoid for which $q < p$. The dotted graph consisting of the two outer branches is a conchoid for which $q > p$. It can be observed that finding the polar equation of the conchoid was no problem; the equation came as the conchoid was being defined. Now comes the problem. By direct use of Figure 10.193 or, alternatively, by using (4) and transformation

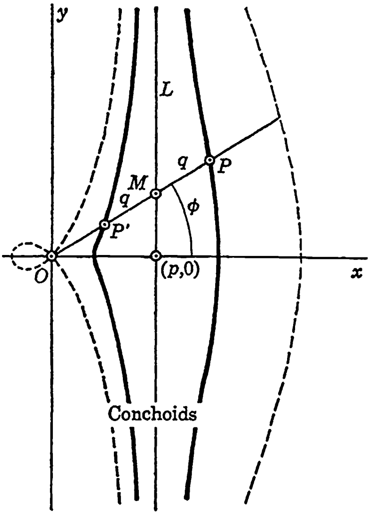


Figure 10.193

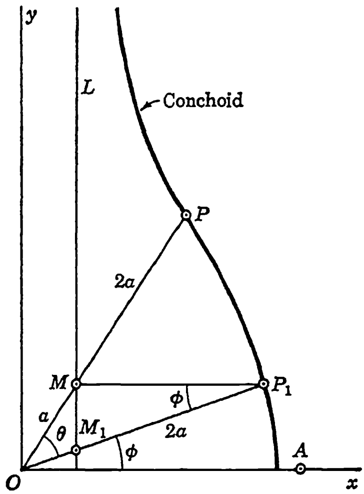


Figure 10.194

formulas, find the rectangular equation of the conchoid which applies to the primed coordinate system for which the x' and y' axes are the x axis and the line having the equation $x = p$. *Ans.:*

$$x'^2y'^2 = (q^2 - x'^2)(x' + p)^2. \tag{5}$$

Remark: Conchoids are interesting examples of graphs of *quartic equations*, that is, equations of the form $f(x,y) = g(x,y)$, where f and g are polynomials in x and y one of which has degree 4 and the other of which has degree not exceeding 4. To trisect the given angle AOP of Figure 10.194 with the aid of a conchoid, let M be the point at which the line OP intersects the line L, and let $a = |\overrightarrow{OM}|$. Let C be the far branch of the particular conchoid for which $q = 2a$. Let P_1 be the point at which the horizontal line through M intersects C and let M_1 be the point at which the line OP_1 intersects the line L so that $|\overrightarrow{M_1P_1}| = 2a$. To begin our attack upon angles, let ϕ be the angle AOP_1 and let θ be the angle P_1OP, so that the given angle AOP is $\phi + \theta$. Applying the law of sines to the triangle OMP_1 gives the first of the equations

$$\frac{a}{\sin\phi} = \frac{|\overrightarrow{MP_1}|}{\sin\theta}, \qquad \frac{a}{\sin\phi} = \frac{2a\cos\phi}{\sin\theta}$$

and the second follows because $|\overrightarrow{MP_1}| = 2a\cos\phi$. Thus

$$\sin\theta = 2\sin\phi\cos\phi = \sin 2\phi,$$

so $\theta = 2\phi$ and the given angle AOP is 3ϕ. Thus the line OP_1 trisects the given angle. It is possible to use ruler-and-compass constructions to locate many points on the conchoid C, but it is impossible to give explicit instructions for producing the particular point P_1 needed for trisection of the given angle.

18 Let O be a point on a circle of diameter a, and let b be a positive number. A set S of points P (a *limaçon of Pascal*) is determined in the following way. If L is a line through O and if Q is either the second point in which L intersects the circle or is O itself if L is tangent to the circle, then S contains the two points on L at distance b from Q. Sketch some figures and investigate these limaçons. *Remark:* With suitable coordinates, the equation can be put in the forms

$$\rho = a \cos \theta \pm b, \qquad \rho = a \cos \theta + b, \qquad \rho = b - a \cos \theta.$$

The fact that $\cos (\theta + \pi) = -\cos \theta$ is important.

19 Let a and b be positive constants. Let F_1 and F_2 be two points having polar coordinates (a,π) and $(a,0)$ and rectangular coordinates $(-a,0)$ and $(a,0)$. The set S of points for which $|\overrightarrow{F_1P}|\,|\overrightarrow{F_2P}| = b^2$ is called an *oval of Cassini*. Investigate these ovals. *Remark:* If $b \gg a$ (read "if b is much greater than a"), then S is closely approximated by a large circle. If $b = a$, then S is a "figure eight" which is, in fact, a lemniscate; see Problem 6. If $b \ll a$ (read "if b is much smaller than a"), then S consists of two small ovals that are closely approximated by small circles.

10.2 Polar curves, tangents, and lengths

As our discussion of coordinate systems may have indicated, polar coordinates can be particularly useful in situations where distances from an origin are particularly significant. It turns out that the polar equation of a conic is exceptionally neat and attractive when we put the conic in the "standard position." As in Figure 10.21, let the focus and directrix of a conic K having eccentricity e be placed upon a polar coordinate system in such a way that a focus is at the origin and the directrix is perpendicular to the initial line and intersects the extended initial line at the point having polar coordinates (p,π). The intrinsic equation of the conic K, which first appeared in our work in (6.23), is then

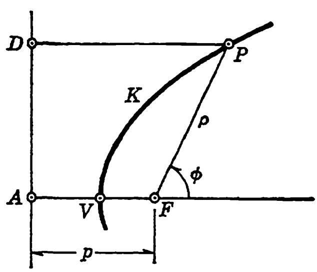


Figure 10.21

$$|\overrightarrow{FP}| = e|\overrightarrow{PD}|. \tag{10.22}$$

While it can be presumed that we know something about conics and can proceed without the result, it is nevertheless interesting to use the intermediate-value theorem to prove that if $\pi/2 < |\phi_0| \leqq \pi$, then there is exactly one point P_0 with polar coordinates (ρ_0,ϕ_0) for which $|\overrightarrow{FP_0}| = e|\overrightarrow{D_0P_0}|$. In any case, we consider only values of ρ for which

$\rho > 0$, so there can be no possible objection to use of Figure 10.21. The intrinsic equation (10.22) then gives

$$\rho = e(p + \rho \cos \phi).$$

When $\cos \phi < 1/e$, we can solve for ρ to obtain the important "standard form" equation

$$\rho = \frac{ep}{1 - e \cos \phi}. \tag{10.23}$$

To illustrate the fact that polar equations are used as sources of information, we proceed to study (10.23). When $e < 1$, the condition $\cos \phi < 1/e$ is satisfied for each ϕ and the polar graph of (10.23) is an ellipse. In this case the point P with polar coordinates (ρ,ϕ) runs repeatedly around the ellipse in the positive direction as ϕ increases over intervals of length 2π. In case $e = 1$, the condition $\cos \phi < 1$ means that ϕ is not an integer mutiple of 2π, and the polar graph of (10.23) is a parabola. In this case $P_p(\rho,\phi)$, the point having *polar* coordinates ρ and ϕ, runs in the positive direction along arcs of the parabola as ϕ increases over subintervals of the interval $0 < \phi < 2\pi$. In case $e > 1$, the restriction $\cos \phi < 1/e$ is more severe, and the polar graph of (10.23) is one branch of a hyperbola. In this case, we confine ϕ to the interval $\phi_0 < \phi < 2\pi - \phi_0$, where ϕ_0 is the first-quadrant angle for which $\cos \phi_0 = 1/e$. The point $P_p(\rho,\phi)$ runs over arcs of the hyperbola as ϕ increases over subintervals of the interval $\phi_0 < \phi < 2\pi - \phi_0$. Because particles moving along one branch of a hyperbola do not suddenly jump to another branch, and for other reasons, we are usually content to work with only one branch of a hyperbola. We are, therefore, usually not interested in the fact that if ϕ is an angle for which $|\phi| < \phi_0$ so that $e \cos \phi > 1$, then the formula (10.23) determines a negative number ρ and the point $P_p(\rho,\phi)$ lies on the other branch of the hyperbola. As Figure 10.21 indicates, each conic K intersects the axis of K at a vertex V between the focus and directrix. Putting $\phi = \pi$ in (10.23) shows that the distance from F to V is $ep/(1 + e)$. In case $e = 1$, this reduces, as it should, to $p/2$. In case $e \neq 1$, another vertex V' is obtained by setting $\phi = 0$. The polar coordinates of V' are $ep/(1 - e)$ and 0. In case $e < 1$, our formulas give

$$|\overrightarrow{VV'}| = |\overrightarrow{VF}| + |\overrightarrow{FV'}| = \frac{ep}{1 + e} + \frac{ep}{1 - e} = \frac{2ep}{1 - e^2}. \tag{10.24}$$

The center of the conic lies midway between the vertices, and it is a straightforward matter to continue this investigation to obtain additional information.

In connection with a conic or other curve C having a manageable polar equation, it is of interest to have information about the tangent to C

at a point P on it and, in particular, to have information about the angle ψ (psi) between this tangent and the vector running from the origin to P. The most informative way to attack these and related questions is by use of vectors; in fact it is not improbable that, in the long run, experience gained by working with vectors may be more valuable than information about ψ. We may start with a curve having the polar equation $\rho = f(\phi)$, where f is supposed to have a continuous derivative. We may suppose that a particle P moves along C in such a way that its polar coordinates at time t are $f(\phi(t))$ and $\phi(t)$. When this is so, we can set $\rho(t) = f(\phi(t))$ and say that P has polar coordinates $\rho(t)$ and $\phi(t)$ at time t. We now free ourselves from the supposition that ρ was a function of ϕ in the first place, and we consider the general situation in which a particle P has polar coordinates $\rho(t)$ and $\phi(t)$ at time t. Whenever we wish to do so, we can reduce our work to the special case simply by setting $\phi(t) = t$, but it is very much worthwhile to handle the more general situation. Moreover, we can still further increase the applicability of our work by studying a still more general situation.

One who needs the medicine can free himself from the notion that matters have become mysterious by supposing that P is a bumblebee or electron that is buzzing around in E_3 in such a way that its cylindrical coordinates at time t are $\rho(t)$, $\phi(t)$, $z(t)$. One who wishes to consider only polar coordinates can put $z(t) = 0$ at all times. The rectangular coordinates $x(t)$, $y(t)$, $z(t)$ are determined in terms of the cylindrical coordinates by the formulas

$$x(t) = \rho(t) \cos \phi(t), \qquad y(t) = \rho(t) \sin \phi(t), \qquad z(t) = z(t). \tag{10.25}$$

Thus, in terms of the standard unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, the vector $\mathbf{r}(t)$ running from the origin to P at time t is

$$\mathbf{r}(t) = \rho(t)[\cos \phi(t)\mathbf{i} + \sin \phi(t)\mathbf{j}] + z(t)\mathbf{k}. \tag{10.26}$$

In all of the following work, we suppose that t is confined to an interval over which ρ, ϕ, and z are functions having continuous derivatives. Let C be the curve (or arc) traversed by P as t increases over this interval. Differentiating (10.26) gives the formula

$$\begin{aligned}\mathbf{v}(t) = \rho'(t)&[\cos \phi(t)\mathbf{i} + \sin \phi(t)\mathbf{j}] \\ &+ \rho(t)\phi'(t)[-\sin \phi(t)\mathbf{i} + \cos \phi(t)\mathbf{j}] + z'(t)\mathbf{k}\end{aligned} \tag{10.261}$$

for the vector $\mathbf{v}(t)$ which is the velocity of P at time t and is also the forward tangent to C at P. This can be put in the form

$$\mathbf{v}(t) = \rho'(t)\mathbf{u}_1(t) + \rho(t)\phi'(t)\mathbf{u}_2(t) + z'(t)\mathbf{k} \tag{10.262}$$

where

$$\begin{cases} \mathbf{u}_1(t) = \cos \phi(t)\mathbf{i} + \sin \phi(t)\mathbf{j} \\ \mathbf{u}_2(t) = -\sin \phi(t)\mathbf{i} + \cos \phi(t)\mathbf{j}. \end{cases} \tag{10.263}$$

The vector $\mathbf{u}_1(t)$ is the unit vector in the direction of the projection of the vector $\mathbf{r}(t)$ upon the xy plane. The vector $\mathbf{u}_2(t)$ is easily seen to be another unit vector, and it is orthogonal to both $\mathbf{u}_1(t)$ and $\mathbf{k}$ because

$$\mathbf{u}_1(t)\cdot\mathbf{u}_2(t) = 0, \qquad \mathbf{k}\cdot\mathbf{u}_2(t) = 0.$$

Moreover, as we see by introducing vector products,

$$\mathbf{u}_1(t) \times \mathbf{u}_2(t) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cos\phi(t) & \sin\phi(t) & 0 \\ -\sin\phi(t) & \cos\phi(t) & 0 \end{vmatrix} = \mathbf{k},$$

so the three vectors $\mathbf{u}_1(t)$, $\mathbf{u}_2(t)$, $\mathbf{k}$, in this order, constitute a right-handed orthonormal system of vectors. Figure 10.27 shows these things

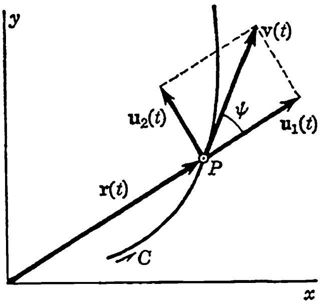


Figure 10.27

for the special case in which C, and hence the vector $\mathbf{r}(t)$, lies in the xy plane of the paper and the unit vector $\mathbf{k}$ therefore extends vertically upward from the plane of the paper. Use of the right-hand finger and thumb rule shows that Figure 10.27 would be wrong if the direction of the vector $\mathbf{u}_2(t)$ were reversed, and this shows that our excursion into E_3 helps us to see how things are oriented in the plane.

Thus, (10.261) and (10.262) are remarkably simple and informative formulas which display the scalar and vector components of $\mathbf{v}(t)$, the velocity vector or the forward tangent vector, in terms of the three orthonormal vectors $\mathbf{u}_1(t)$, $\mathbf{u}_2(t)$, and $\mathbf{k}$. The orthonormality of these vectors enables us to use (10.26) and (10.261) to obtain the formulas

$$|\mathbf{r}(t)| = \sqrt{[\rho(t)]^2 + [z(t)]^2} \tag{10.271}$$

$$|\mathbf{v}(t)| = \sqrt{[\rho'(t)]^2 + [\rho(t)\phi'(t)]^2 + [z'(t)]^2} \tag{10.272}$$

$$\mathbf{r}(t)\cdot\mathbf{v}(t) = \rho(t)\rho'(t) + z(t)z'(t). \tag{10.273}$$

The angle ψ between the vectors $\mathbf{r}(t)$ and $\mathbf{v}(t)$ can, when $|\mathbf{r}(t)| \neq 0$ and $|\mathbf{v}(t)| \neq 0$, be calculated from the basic formula

$$\mathbf{r}(t)\cdot\mathbf{v}(t) = |\mathbf{r}(t)|\,|\mathbf{v}(t)| \cos\psi. \tag{10.274}$$

For the case in which $z(t)$ is identically zero and $\rho(t) > 0$, this gives the formula

$$\cos \psi = \frac{\rho'(t)}{\sqrt{[\rho'(t)]^2 + [\rho(t)\phi'(t)]^2}}. \tag{10.275}$$

In case $\rho(t) > 0$, $\phi'(t) > 0$, and $\rho'(t) \neq 0$, this enables us to prove the first formula in

$$\tan \psi = \frac{\rho(t)\phi'(t)}{\rho'(t)}, \qquad \tan \psi = \rho \frac{d\phi}{d\rho} = \frac{\rho}{\dfrac{d\rho}{d\phi}}. \tag{10.276}$$

In appropriate circumstances, the second formula follows from the first.

In accordance with conclusions reached at the end of Section 7.2, we can put (10.272) in the form

$$s'(t) = \sqrt{[\rho'(t)]^2 + [\rho(t)\phi'(t)]^2 + [z'(t)]^2}, \tag{10.277}$$

where $s(t)$ is the coordinate at time t which is obtained by measuring distance along the curve C. When $z'(t) = 0$ for each t and $\phi(t) = t$ so $\phi'(t) = 1$, this is often put in the form

$$\frac{ds}{d\phi} = \sqrt{\rho^2 + \left(\frac{d\rho}{d\phi}\right)^2}. \tag{10.278}$$

Matters relating to lengths of curves are of sufficient interest to justify close scrutiny of the following theorem.

Theorem 10.28 *If ρ and ϕ are functions having continuous derivatives over $a \leqq t \leqq b$, then the integral in the formula*

$$L = \int_a^b \sqrt{[\rho'(t)]^2 + [\rho(t)\phi'(t)]^2}\, dt \tag{10.281}$$

is the length of the curve C consisting of the ordered set of points P having polar coordinates $(\rho(t), \phi(t))$ for which $a \leqq t \leqq b$.

The simplest proof of this theorem is obtained by setting

$$x(t) = \rho(t) \cos \phi(t), \qquad y(t) = \rho(t) \sin \phi(t), \qquad z(t) = 0 \tag{10.282}$$

in the formula (7.26) which was thoroughly discussed and proved in Section 7.2. The formulas

$$x'(t) = -\rho(t) \sin \phi(t)\phi'(t) + \cos \phi(t)\rho'(t) \tag{10.283}$$
$$y'(t) = \rho(t) \cos \phi(t)\phi'(t) + \sin \phi(t)\rho'(t) \tag{10.284}$$

enable us to convert (7.26) into (10.281). For the special case in which $\phi(t) = t$ and $\phi'(t) = 1$, it is standard practice to put (10.281) in the form

$$L = \int_\alpha^\beta \sqrt{\rho^2 + \left(\frac{d\rho}{d\phi}\right)^2}\, d\phi \tag{10.285}$$

in which the variable of integration is ϕ and the limits of integration are called α and β (instead of a and b) because α and β "look more like angles."

It is worthwhile to know a little trick by which the above formulas involving polar coordinates can be remembered. We can look at Figure 10.286 which shows, among other things, an arc of length Δs joining two

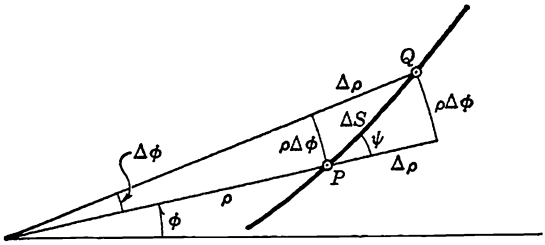


Figure 10.286

points P and Q which have polar coordinates ρ, ϕ and $\rho + \Delta\rho$, $\phi + \Delta\phi$. We can feel that the outer part of the figure resembles a rectangle enough to enable us to write the approximate formulas

$$\Delta s = \sqrt{\rho^2\,\Delta\phi^2 + \Delta\rho^2}, \qquad \cos\psi = \frac{\Delta\rho}{\sqrt{\rho^2\,\Delta\phi^2 + \Delta\rho^2}}, \qquad \tan\psi = \frac{\rho\,\Delta\phi}{\Delta\rho} \tag{10.287}$$

and expect that correct results should be obtained by dividing by Δt or by $\Delta\phi$ and taking limits. We can know that this optimism does not prove formulas, but it can help us to recall the formulas when we have forgotten them. When we wish to calculate the length L of the curve having the polar equation $\rho = f(\phi)$ with $\alpha \leqq \phi \leqq \beta$, we can, when f' is continuous, sketch a figure more or less like Figure 10.286 and use the optimistic calculation

$$\begin{aligned} L &= \lim \sum \Delta s = \lim \sum \sqrt{\rho^2\,\Delta\phi^2 + \Delta\rho^2} \\ &= \lim \sum \sqrt{\rho^2 + \left(\frac{\Delta\rho}{\Delta\phi}\right)^2}\,\Delta\phi = \int_\alpha^\beta \sqrt{\rho^2 + \left(\frac{d\rho}{d\phi}\right)^2}\, d\phi \end{aligned} \tag{10.288}$$

to lead us to the correct formula (10.285).

Problems 10.29

1 Obtain the standard polar equation of the conic K and use it to sketch the major and minor (or conjugate) axis of K when the eccentricity e and distance p from the focus to the directrix are

(*a*) $e = \frac{1}{2}$, $p = 6$ (*b*) $e = 2$, $p = 3$

2 Make a hasty sketch of the cardioid having the polar equation

$$\rho = 1 + \cos \phi.$$

Show that $\tan \psi = -(1 + \cos \phi)/\sin \phi = -\cot \frac{1}{2}\phi$. Calculate $\tan \psi$ when $\phi = \pi/2$ and when $\phi = \pi$ and make any repairs in your figure that this information may require.

3 Find the length of the cardioid of the preceding problem. *Ans.:* 8.

4 There is something unique about angles at which radial lines from the origin intersect the exponential spiral having the polar equation $\rho = e^{a\phi}$. What is it? *Ans.:* The angles are all equal.

5 As ϕ increases from 0, the point $P_P(\rho,\phi)$ on the polar graph of $\rho = e^{-\phi}$ spirals around the origin. Since π is about 3 and e^3 is about 20, the point P spirals toward the origin so rapidly that the length of the whole path may be not much greater than the distance from the starting point to the origin. What are the facts? *Ans.:* To try to preserve good ideas and perhaps create more, put the matter this way: If ϕ starts at time $t = 0$ and increases at a constant rate, the point P must keep moving forever, but its speed decreases so rapidly that the total distance traveled is always less than $\sqrt{2}$ and only approaches $\sqrt{2}$ as $t \to \infty$ and $\phi \to \infty$. It makes sense to say that the total length of the path is $\sqrt{2}$.

6 Let C be the polar graph of $\rho = f(\phi)$, where $f(0) = 1$, $f(2\pi) = 2$, and f is continuous and monotone increasing over the interval $0 \leqq \phi \leqq 2\pi$. Try to decide whether it is easy or difficult or impossible to prove that C must have finite length.

7 The curve C of the preceding problem lies between the polar graphs of the equations $\rho = 1$ and $\rho = 2$. Try to decide whether it is easy or difficult or impossible to prove that the length of C lies between the lengths of the inner and outer circles.

8 Let the displacement vector of a particle P at time t be

$$\mathbf{r}(t) = \rho(t)[\cos \phi(t)\mathbf{i} + \sin \phi(t)\mathbf{j}], \tag{1}$$

where it is supposed that ρ and ϕ have two derivatives. Forgetting formulas which we have derived but remembering rules for differentiating products, derive the velocity and acceleration formulas

$$\mathbf{v}(t) = \rho'(t)[\cos \phi(t)\mathbf{i} + \sin \phi(t)\mathbf{j}] + \rho(t)\phi'(t)[-\sin \phi(t)\mathbf{i} + \cos \phi(t)\mathbf{j}] \tag{2}$$

$$\begin{aligned}\mathbf{a}(t) = {} & [\rho''(t) - \rho(t)(\phi'(t))^2][\cos \phi(t)\mathbf{i} + \sin \phi(t)\mathbf{j}] \\ & + [\rho(t)\phi''(t) + 2\rho'(t)\phi'(t)][-\sin \phi(t)\mathbf{i} + \cos \phi(t)\mathbf{j}].\end{aligned} \tag{3}$$

Observe anew that the two vectors

$$[\cos \phi(t)\mathbf{i} + \sin \phi(t)\mathbf{j}], \qquad [-\sin \phi(t)\mathbf{i} + \cos \phi(t)\mathbf{j}] \tag{4}$$

are orthonormal vectors because they are unit vectors and their scalar (or dot) product is zero. Henceforth we consider only time intervals over which $\rho(t) > 0$. When motion of the particle P is being considered, the first vector in (4) is said to be *radial* because it has the direction of the "radius vector" from the origin to P, and the second vector is *transverse* because it is orthogonal to the radius vector. Thus the right member of (3) displays, in order, the radial and the transverse vector components of the acceleration of P. This acceleration is said to be radial (the kind produced by a "central force field" having its center at the origin) when its transverse component is zero, that is,

$$(5)\qquad \rho(t)\phi''(t) + 2\rho'(t)\phi'(t) = 0.$$

This is another one of those derivative equations that is called a differential equation and from which information can be extracted. Since $\rho(t) > 0$, the left member of (5) is zero if and only if the product of it and $\rho(t)$ is zero, that is,

$$(6)\qquad [\rho(t)]^2\phi''(t) + 2\rho(t)\rho'(t)\phi'(t) = 0.$$

The virtue of (6) lies in the fact that it can be put in the form

$$(7)\qquad \frac{d}{dt}\{[\rho(t)]^2\phi'(t)\} = 0,$$

and this fact should be carefully checked. The virtue of (7) lies in the fact that it holds over an interval of values of t if and only if there is a constant c such that

$$(8)\qquad \tfrac{1}{2}[\rho(t)]^2\phi'(t) = c$$

for each t in the interval. The physical significance of (8) will be revealed in Section 10.3; it is an important fact that (8) holds if and only if the radius vector from the origin to the particle P sweeps over regions of equal area in time intervals of equal lengths.

9 A circular race track has cylindrical equations $\rho = a$ and $z = 0$, and it has rectangular equations $x^2 + y^2 = a^2$ and $z = 0$. A wheel of radius b rolls, without slipping, around the track. The center of the wheel is always above the track, it travels with constant speed, and it makes a circuit of the track in T minutes. At time $t = 0$ the center of the wheel is going in the direction of the positive y axis, and a pink tack P in the tire on the wheel is at the point having rectangular coordinates $(a,0,0)$. Letting $\mathbf{r}_1(t)$ denote the vector running from the origin O to the center Q of the wheel at time t, show that

$$(1)\qquad \mathbf{r}_1(t) = a\left(\cos\frac{2\pi t}{T}\,\mathbf{i} + \sin\frac{2\pi t}{T}\,\mathbf{j}\right) + b\mathbf{k}.$$

Letting $x(t)$, $y(t)$, $z(t)$ denote the rectangular coordinates of the pink tack P at time t, show that

$$(2)\qquad z(t) = b\left(1 - \cos\frac{2\pi at}{bT}\right).$$

Using the fact that the line PQ (a spoke of the wheel) is perpendicular to the horizontal line from Q to the z axis, obtain the formula

$$(3)\qquad x(t)\cos\frac{2\pi t}{T} + y(t)\sin\frac{2\pi t}{T} = a$$

which may help us determine $x(t)$ and $y(t)$. Using the fact that the distance from P to Q is always b (the radius of the wheel), obtain the formula

$$\left[x(t) - a\cos\frac{2\pi t}{T}\right]^2 + \left[y(t) - a\sin\frac{2\pi t}{T}\right]^2 = b^2\sin^2\frac{2\pi at}{bT}. \tag{4}$$

Show that

$$[x(t)]^2 + [y(t)]^2 = a^2 + b^2\sin^2\frac{2\pi at}{bT}. \tag{5}$$

With or without assistance from these formulas, derive the formulas

$$\begin{aligned} x(t) &= a\cos\frac{2\pi t}{T} + b\sin\frac{2\pi t}{T}\sin\frac{2\pi at}{bT} \\ y(t) &= a\sin\frac{2\pi t}{T} - b\cos\frac{2\pi t}{T}\sin\frac{2\pi at}{bT} \\ z(t) &= b - b\cos\frac{2\pi at}{bT} \end{aligned}$$

and use them to find the velocity and acceleration of P when $t = 0$. *Remark:* Mechanisms involving rolling wheels (or gears) appear in machinery in various ways, and we have the preliminary idea that we can start studying these things.

10 Let a be a positive number. The point P lies on a line through the origin which intersects the line having the rectangular equation $x = a$ at a point Q, and $|\overrightarrow{PQ}|$ is equal to the distance from Q to the x axis. The set of such points P is a *strophoid*. Find the polar equation of the strophoid. *Ans.:*

$$\rho^2\cos\phi - 2a\rho + a^2\cos\phi = 0.$$

11 Supposing that $a > 0$, prove that the line having the rectangular equation $y = a$ is an asymptote of the *hyperbolic spiral* having the polar equation $\rho\phi = a$.

12 Supposing that $a > 0$, prove that the x axis is an asymptote of the *lituus* having the polar equation $\rho^2\phi = a^2$.

13 Show that transforming the first of the equations

$$\rho = 4a\cos\phi - a\sec\phi, \qquad x^4 + xy^2 + ay^2 - 3ax^2 = 0 \tag{1}$$

from polar to rectangular equations gives the second. *Remark:* The graph of these equations is a *trisectrix of Maclaurin*. It is possible to use a formula for $\tan 3\phi$ to show that if O is the origin, if Q is the point $(2a, 0)$, if P is a point on the trisectrix in the first quadrant, if θ is the angle in the interval $0 < \theta < \pi$ which the line OQ makes with the positive x axis, and if ϕ is the acute angle which OP makes with the positive x axis, then $\theta = 3\phi$. Thus the trisectrix provides a method (but not a ruler-and-compass method) for trisecting angles. Particularly when $0 \leqq \phi < \pi/2$, the number ρ in (1) has an elegant geometric interpretation. It is $|\overrightarrow{OB}| - |\overrightarrow{OA}|$ where A and B are the points where a line through O meets the line having the rectangular equation $x = a$ and the circle of radius $2a$ having center Q.

14 The innocent graph of the equation

$$y = \frac{1}{1 + x^2}$$

is called the *witch of Agnesi* because Maria Gaetena Agnesi (1718–1799) discovered a spooky ruler-and-compass method for constructing points on it. Let C be the circle of diameter 1 with center at $(0, \frac{1}{2})$. Let T be a point on the line tangent to the circle at the point $(0, 1)$, and let Q be the point, different from the origin O, at which the line OT intersects the circle. Show that the vertical line through T and the horizontal line through Q intersect at a point on the witch.

10.3 Areas and integrals involving polar coordinates Problems involving "areas in polar coordinates" can be formulated in different ways. To begin, we suppose that f is a nonnegative integrable function of ϕ over some interval $\alpha \leqq \phi \leqq \beta$ for which $\beta \leqq \alpha + 2\pi$. Let S be the set of points having polar coordinates ρ, ϕ for which $\alpha \leqq \phi \leqq \beta$ and $0 \leqq \rho \leqq f(\phi)$. In case f is continuous, S may be described as the set S bounded by the polar graphs of the equations $\phi = \alpha$, $\phi = \beta$, and

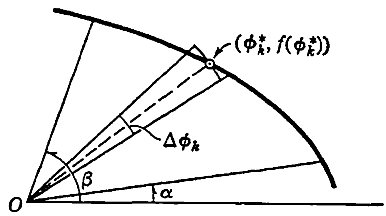


Figure 10.31

$\rho = f(\phi)$. The schematic Figure 10.31 should not be misleading. To find the area $|S|$ of S, we make a partition of the interval $\alpha \leqq \phi \leqq \beta$ into subintervals of which a representative one of length $\Delta\phi_k$ contains a particular ϕ_k^*. When there is a necessity for being precise about this matter, we set $\Delta\phi_k = \phi_k - \phi_{k-1}$ and choose ϕ_k^* such that $\phi_{k-1} \leqq \phi_k^* \leqq \phi_k$. The area of the subset of S containing points for which $\phi_{k-1} \leqq \phi \leqq \phi_k$ can now be approximated by the area $\frac{1}{2}[f(\phi_k^*)]^2\,\Delta\phi_k$ of the sector of radius $f(\phi_k^*)$ having central angle $\Delta\phi_k$. When we are hurried, we need not make the usual remarks about the way in which the approximation depends upon the choice of ϕ_k^*, but in any case we should know why we are using the area of the sector instead of its perimeter. An application of fundamental ideas about estimating, summing, and taking limits to set up integrals then gives

$$|S| = \lim \sum \tfrac{1}{2}[f(\phi_k^*)]^2\,\Delta\phi_k = \tfrac{1}{2}\int_\alpha^\beta [f(\phi)]^2\,d\phi, \tag{10.32}$$

where the integral is a Riemann integral. When we are not required to write detailed explanations of the reasons for doing what we do, we replace (10.32) by the simpler calculation

$$|S| = \lim \sum \tfrac{1}{2}[f(\phi)]^2\,\Delta\phi = \tfrac{1}{2}\int_\alpha^\beta [f(\phi)]^2\,d\phi \tag{10.33}$$

in which the subscripts and star superscripts do not appear. In applications, we often write ρ or $\rho(\phi)$ in place of $f(\phi)$.

To add variety to our acquaintance with problems, we define two numbers I_1 and I_2 by the formulas

$$(10.34)\qquad I_1 = \tfrac{1}{2}\int_{t_1}^{t_2} [\rho(t)]^2\phi'(t)\,dt, \qquad I_2 = \tfrac{1}{2}\int_{t_1}^{t_2} [\rho(t)]^2|\phi'(t)|\,dt$$

and ask how I_1 and I_2 may be interpreted in terms of polar coordinates and areas when $t_1 < t_2$. In order that the integrands and integrals exist, it is necessary that ρ and ϕ be functions for which $\rho(t)$, $\phi(t)$, and $\phi'(t)$ exist when $t_1 \leqq t \leqq t_2$, and we simplify matters by supposing that ρ, ϕ, and ϕ' are all continuous over $t_1 \leqq t \leqq t_2$. As t increases over the interval $t_1 \leqq t \leqq t_2$, the point $P(t)$ having polar coordinates $(\rho(t), \phi(t))$ traces a curve (or an oriented curve) C from $P(t_1)$ to $P(t_2)$ that could look like that shown in Figure 10.35 or like that shown in Figure 10.36.

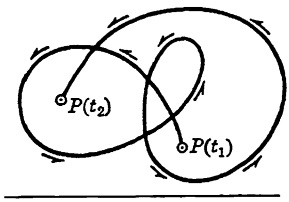


Figure 10.35

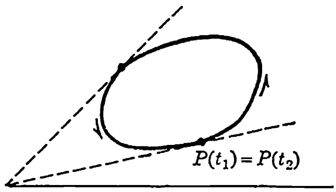


Figure 10.36

Of course, there are other possibilities, and we exclude more elaborate ones by supposing that the interval $t_1 \leqq t \leqq t_2$ can be separated into a finite set of subintervals such that, over each subinterval, ϕ is either monotone increasing or monotone decreasing. Let P be a partition of the interval $t_1 \leqq t \leqq t_2$ such that, in each subinterval, $\phi(t)$ is either monotone increasing or monotone decreasing. Let t_k^* be chosen in the kth subinterval in such a way that

$$(10.37)\qquad \phi(t_k) - \phi(t_{k-1}) = \phi'(t_k^*)(t_k - t_{k-1}) = \phi'(t_k^*)\,\Delta t_k,$$

and build the Riemann sum

$$(10.38)\qquad \Sigma\tfrac{1}{2}[f(t_k^*)]^2\phi'(t_k^*)\,\Delta t_k$$

which approximates I. A particular term in this sum is an approximation to the area of the region swept over by the vector from O to $P(t)$ as t increases from t_{k-1} to t_k provided $\phi(t_k) > \phi(t_{k-1})$, that is, provided the vector rotates in the positive direction. Similarly, the term is an approximation to the negative of the area if the vector rotates in the negative direction. Thus I_1 is the sum of areas of regions swept over in the positive direction and the negative of areas of regions swept over in the negative direction. We never have negative areas, but we can subtract areas because areas are numbers. In case the path of $P(t)$ is the closed curve C of Figure 10.36, I is the area of the region bounded by C. If the path of $P(t)$ is the curve obtained by reversing the direction of the arrows in Figure 10.36, then I_1 is the negative of the area of the region enclosed by the curve. Except in cases where the vector from O to $P(t)$ always rotates in the same direction, the number I_2 in (10.34) is usually less interesting than I_1. It is the sum of the areas of all of the regions swept over by the rotating vector.

Problems 10.39

1 Find the area of the region bounded by the cardioid having the polar equation

$$\rho = a(1 + \cos \phi).$$

Ans.: $3\pi a^2/2$.

2 Using the polar equation $\rho = 2a \cos \phi$ of a circle of radius a, and noting that a particle with polar coordinates (ρ,ϕ) traverses the circle once as ϕ increases from $-\pi/2$ to $\pi/2$, work out the familiar formula for the area of the circular disk bounded by the circle.

3 The graph of the polar equation $\rho^2 = a^2 \cos 2\phi$ is the lemniscate shown in Figure 10.171. Find the area of the bipartite set which it bounds *Ans.:* a^2.

4 Find the area of the region bounded by the graph of $\rho = \sqrt{\cos \phi}$. *Ans.:* 1.

5 Supposing that $a > 1$ and that n is a positive integer, find the area of the region bounded by the polar graph of the equation

$$\rho = a + \cos n\phi.$$

Ans.: $\pi a^2 + \pi/2$.

6 Use integration to find the area A_1 of the smaller region which the line with rectangular equation $y = x$ slices from the interior of the circle having the polar equation $\rho = 2a \cos \phi$. Then calculate A_1 from the fact that the interior of the circle is the union of the interior of an inscribed square and four slices each having area A_1. Make the results agree.

7 Sketch a polar graph of the equations

$$\rho = 4 + \cos t, \qquad \phi = \frac{\pi}{2} \sin t$$

and make a rough estimate of the area of the region enclosed by the graph. Then calculate the area.

8 Set up an integral for the area A of the region bounded by the ellipse having the standard polar equation

$$\rho = \frac{ep}{1 - e \cos \phi}$$

and show that the result can be put in the form

$$A = p^2 \int_0^{\pi} \frac{1}{(e^{-1} - \cos \phi)^2}\, d\phi.$$

9 It is much easier to learn to play a violin than to acquire competence to give basic definitions and theorems involving areas of patches of curved surfaces. Some problems are so simple, however, that elementary methods yield answers that are universally considered to be correct. As Figure 10.391 suggests, a

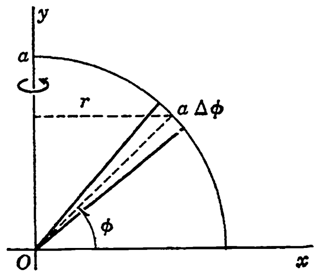


Figure 10.391

hemisphere (or hemispherical surface) of radius a is generated by rotating a quadrant of radius a about the y axis. To find the area A of this hemisphere, we make a partition of the interval $0 \leqq \phi \leqq \pi/2$. Rotating the segment of length $a\,\Delta\phi$ about the y axis gives a part of the hemisphere that can be roughly described as a ribbon of width $a\,\Delta\phi$ and length $2\pi r$, where r is $a \cos \phi$ The process for setting up and evaluating integrals then gives

$$A = \lim \sum 2\pi a^2 \cos \phi\, \Delta\phi = 2\pi a^2 \int_0^{\pi/2} \cos \phi\, d\phi = 2\pi a^2.$$

These preliminaries can be ended with the remark that the area of the hemisphere ought to be about double the area of the equatorial circular disk and that the world is so simple that the factor is exactly 2. Now comes the problem. Supposing that $0 \leqq c < c + h \leqq a$, find the area of the zone generated by rotating (about the y axis) the part of the arc between the lines having the equations $y = c$ and $y = c + h$.

10 Supposing that $r(t)$ and $\phi(t)$ have continuous derivatives and that

$$x(t) = r(t) \cos \phi(t), \qquad y(t) = r(t) \sin \phi(t),$$

calculate $x'(t)$ and $y'(t)$ and show that

$$x(t)y'(t) - y(t)x'(t) = [r(t)]^2\phi'(t).$$

11 Making use of the method involving (10.38), put appropriate hypotheses on functions $x(t)$ and $y(t)$ and discover geometric interpretations of the integrals

$$\int_{t_1}^{t_2} x(t)y'(t)\, dt, \qquad \int_{t_1}^{t_2} y(t)x'(t)\, dt.$$

12 Show that if a particle P moves on the conic having the polar coordinate equation

(1) $$\rho(t) = \frac{a}{1 - e \cos \phi(t)}$$

in such a way that ϕ is a function of t having two derivatives, then the displacement vector and velocity of P are

$$(2)\quad \mathbf{r}(t) = \frac{a}{1 - e\cos\phi(t)}[\cos\phi(t)\mathbf{i} + \sin\phi(t)\mathbf{j}]$$

$$(3)\quad \mathbf{v}(t) = \frac{-ea\phi'(t)\sin\phi(t)}{[1 - e\cos\phi(t)]^2}[\cos\phi(t)\mathbf{i} + \sin\phi(t)\mathbf{j}] + \frac{a\phi'(t)}{1 - e\cos\phi(t)}[-\sin\phi(t)\mathbf{i} + \cos\phi(t)\mathbf{j}].$$

For the case in which

$$(4)\quad \tfrac{1}{2}[\rho(t)]^2\phi'(t) = c,$$

so that the radius vector from O to P sweeps over equal areas in time intervals having equal lengths, show that

$$(5)\quad \phi'(t) = \frac{2c}{[\rho(t)]^2} = \frac{2c}{a^2}[1 - e\cos\phi(t)]^2$$

$$(6)\quad \mathbf{v}(t) = -\frac{2ce}{a}\sin\phi(t)[\cos\phi(t)\mathbf{i} + \sin\phi(t)\mathbf{j}] + \frac{2c}{a}[1 - e\cos\phi(t)][-\sin\phi(t)\mathbf{i} + \cos\phi(t)\mathbf{j}]$$

$$(7)\quad \mathbf{a}(t) = -\frac{4c^2}{a}\frac{1}{[\rho(t)]^2}[\cos\phi(t)\mathbf{i} + \sin\phi(t)\mathbf{j}].$$

Remark: This shows that if a particle P moves along a conic in such a way that the radius vector from the focus to P sweeps over equal areas in time intervals having equal lengths, then the particle is always accelerated toward the focus and the magnitude of the acceleration is inversely proportional to the square of the distance from the focus to the particle. Kepler discovered that, except for minor perturbations, the planets move in ellipses with the sun at a focus and that the "equal areas" property holds. As we did in this problem, Newton used these laws to derive his famous inverse square law of gravitation.

13 Let S be a bounded convex set in the xy plane and let the origin O be an inner point of S. (Basic definitions are given in a remark at the end of Problems 5.19.) Prove that to each ϕ there corresponds exactly one positive number $f(\phi)$ such that the point P having polar coordinates $(f(\phi), \phi)$ lies on the boundary B of S. *Solution:* Choose positive numbers δ and R such that the circular disk with center at O and radius 2δ is a subset of S and S is a subset of the circular disk with center at O and radius R. Let ϕ be a given angle. Let $f(\phi)$ be the least upper bound of numbers ρ_0 such that S contains each point having polar coordinates (ρ,ϕ) for which $0 < \rho < \rho_0$. Then $2\delta \leqq f(\phi) \leqq R$. Let P be the point having polar coordinates $(f(\phi), \phi)$ and, as in Figure 10.392, let Q_1 and Q_2 be the points of tangency of the lines through P tangent to the circle of radius δ having its center at the origin. With the possible exception of the point P itself, each point of the line segment OP is a point of S. Since Q_1 and Q_2 are also points of S and since S is convex, it follows that each point inside the triangle

Q_1PQ_2 is a point of S. No interior point Q of the shaded sector of Figure 10.392 can be a point of S. Otherwise, the point P, being an inner point of the inside of the triangle Q_1QQ_2, would be a point of S and so would also each point of some

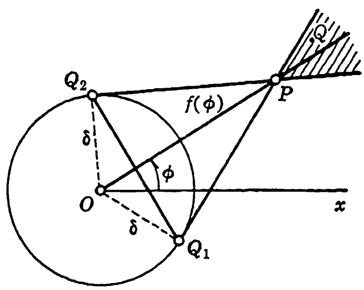


Figure 10.392

circular disk having its center at P. There would then be a number ρ_0 greater than $f(\theta)$ for which S contains each point having polar coordinates (ρ,ϕ) for which $0 \leqq \rho \leqq \rho_0$, and this is impossible. We now know that each point inside the circle with center at the origin and radius δ is a point of S, that each point inside the triangle Q_1PQ_2 is a point of S, and that each point Q interior to the shaded sector is not a point of S. It follows from this that $f(\phi)$ is the one and only positive number such that the point P having polar coordinates $(f(\phi),\ \phi)$ is a point on the boundary B of S *Remark:* Without going into details we observe that this proof provides supplementary information that enables us to relate $f(\phi + h)$ to $f(\phi)$ It can be shown that f is continuous and hence that the boundary B becomes a simple closed curve C when its points are so ordered that the point having polar coordinates $(f(\phi_1),\ \phi_1)$ precedes the point having polar coordinates $(f(\phi_2),\ \phi_2)$ when $0 \leqq \phi_1 < \phi_2 \leqq 2\pi$. The objection that curves were defined in terms of rectangular coordinates is overcome by the observation that if f is continuous, then so also are the functions g_1 and g_2 defined by

$$x = g_1(\phi) = f(\phi)\cos\phi$$
$$y = g_2(\phi) = f(\phi)\sin\phi.$$

It can be proved that the curve C has finite length L.

11 *Partial derivatives*

11.1 Elementary partial derivatives Leaving consideration of more complicated situations to later sections, we confine attention in this section to examples and problems in which the fundamental ideas can be stated quite simply and it is relatively easy to be completely sure of the meanings of all of the symbols that are used. We begin with an example. Suppose a copper rod occupies the interval $x_1 \leqq x \leqq x_2$ of an x axis and that we are interested in the temperature u (measured in degrees centigrade) at points of the rod at various times t. To be precise about the matter, we may suppose that the "space coordinate" x is measured in centimeters with $x = 0$ at some "space origin" and that the "time coordinate" t is measured in seconds with $t = 0$ at some "time origin" which could be the time at which some particular stop watch was started. In some problems, it is not presumed that x and t are positive. For present purposes, we suppose that to each pair of numbers x and t for which $x_1 \leqq x \leqq x_2$ and $t \geqq 0$ there corresponds exactly one number u which we may denote by $f(x,t)$ or by $u(x,t)$ and which is the temperature at the "space-time place" having "space-time coordinates" x and t. Thus the temperature u is a function of the two "variables" x and t. We are now in a realm where ideas of major importance can appear all over the place. Some students who are now dubious about the possibility of doing anything useful or interesting with these functons will, in two or three years, have substantial information about the possibility of starting with positive numbers a and L and a given function g and then determining a positive integer n and constants $A_1, A_2, \cdots, A_n$ such that the particular function u defined by

$$u(x,t) = \sum_{k=1}^{n} A_k e^{-\frac{k^2\pi^2 a^2}{L^2}t} \sin \frac{k\pi}{L} x \tag{11.11}$$

will be a good approximation to $g(x)$ when $t = 0$. In any case, (11.11) exhibits important examples (one for each choice of n and $A_1, A_2, \cdots,$ A_n) of functions u to which our work applies.

We continue study of our example in which temperature u (measured in degrees) is a function of x (measured in centimeters) and t (measured in seconds). If we wish to study the temperature of the rod at some particular time t_0, we can set $t = t_0$ and, without bothering to be fussy about the distinction between a function and values of the function, consider $u(x,t_0)$ to be a function of x alone. If the graph of $u(x,t_0)$ versus x happens to be that shown in Figure 11.12, we can look at the graph to see where the temperature is increasing, but it would not be too easy to determine the rate of change of u with respect to x. To do this and get a number of degrees per centimeter, we would want to differentiate $u(x,t_0)$ with respect to x. Thus we are led to a very important idea. When u is a function of x and some other variables (in our case, just one other variable), it may make sense to assign fixed values to all of the variables except x and differentiate the result with respect to x. In particular, the idea does make sense when we know what these other variables are and, in addition, we know that the resulting function of x is a differentiable function of x. When we know what we are doing (this is a conservative statement providing for the possibility that we may sometimes be puzzled by situations in thermodynamics) the resulting derivative is called the *partial derivative of u with respect to x.* While there are other and more informative symbols for partial derivatives, the simplest and most ingenious one is $\partial u/\partial x$. The "curly dees" in this symbol are unusual Greek deltas, and the symbol is usually read "partial of u with respect to x."

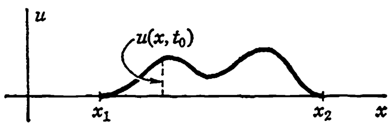


Figure 11.12

We are already in a position to understand statements and make calculations. Partial derivatives are important things that abound in books and on blackboards and in notebooks and on scratch pads. Wherever we find $\partial w/\partial q$, we automatically know that w is a function of q and some other variables, we know that fixed values have been assigned to all variables except q, and we know that $\partial w/\partial q$ stands for the result of differentiating the resulting function of q with respect to q. If w is measured in gees and q is measured in haws, then $\partial w/\partial q$ is measured in gees per haw. When we see the formula

$$(11.13) \quad u = x^2 + y^2 + e^{ax}\sin by \qquad \text{or} \qquad u(x,y) = x^2 + y^2 + e^{ax}\sin by,$$

we can compute partial derivatives with respect to x by supposing that all variables except x are assigned fixed values, so that they are to be regarded as constants when we differentiate with respect to x to obtain

$$(11.131) \quad \frac{\partial u}{\partial x} = 2x + ae^{ax}\sin by \qquad \text{or} \qquad u_x(x,y) = 2x + ae^{ax}\sin by.$$

The second one of these formulas involves the standard subscript notation for partial derivatives. Similarly,

$$(11.132) \quad \frac{\partial u}{\partial y} = 2y + be^{ax}\cos by \qquad \text{or} \qquad u_y(x,y) = 2y + be^{ax}\cos by.$$

It should now be apparent that calculating these partial derivatives is equivalent to evaluating the limits in

$$(11.14) \quad u_x(x,y) = \lim_{\Delta x \to 0} \frac{u(x + \Delta x, y) - u(x,y)}{\Delta x},$$
$$u_y(x,y) = \lim_{\Delta y \to 0} \frac{u(x, y + \Delta y) - u(x,y)}{\Delta y}.$$

Taking the partial derivatives with respect to x of the members of (11.131) and (11.132) gives

$$(11.141) \quad \frac{\partial^2 u}{\partial x^2} = 2 + a^2e^{ax}\sin by \qquad \text{or} \qquad u_{xx}(x,y) = 2 + a^2e^{ax}\sin by$$

and

$$(11.142) \quad \frac{\partial^2 u}{\partial x\,\partial y} = abe^{ax}\cos by \qquad \text{or} \qquad u_{yx}(x,y) = abe^{ax}\cos by.$$

Taking partial derivatives with respect to y of the members of (11.131) and (11.132) gives

$$(11.143) \quad \frac{\partial^2 u}{\partial y\,\partial x} = abe^{ax}\cos by \qquad \text{or} \qquad u_{xy}(x,y) = abe^{ax}\cos by$$

and

$$(11.144) \quad \frac{\partial^2 u}{\partial y^2} = 2 - b^2e^{ax}\sin by \qquad \text{or} \qquad u_{yy}(x,y) = 2 - b^2e^{ax}\sin by.$$

In the above formulas, we have used the formulas

$$(11.145) \qquad \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial^2 u}{\partial y\,\partial x}, \qquad (u_x)_y(x,y) = u_{xy}(x,y),$$

which serve a dual purpose: they provide abbreviations for the expressions on the left sides, and they provide meanings for the abbreviations on the right sides. It could be supposed that we should insert a comma between the subscripts x and y to write $u_{x,y}$ in place of u_{xy}, but it is customary to consider the commas to be superfluous.

The result of setting $x = y = 0$ in (11.131) is

$$\frac{\partial u}{\partial x} = 0 \qquad \text{or} \qquad u_x(0,0) = 0, \tag{11.146}$$

and this shows that the "curly dee" notation for partial derivatives is, after all, a miserably poor purveyor of information. When we see the first formula in (11.146), there is nothing to tell us that u depends upon exactly two variables x and y and that 0 is the result of setting $x = y = 0$ in $\partial u/\partial x$. When we keep the curly dees, it is often necessary to put (11.146) in the form

$$\left.\frac{\partial u}{\partial x}\right|_{(0,0)} = 0, \qquad u_x(0,0) = 0 \tag{11.147}$$

so the first formula, like the second, can really mean something. As this one example may suggest, the curly dees really should be banished from the universe because they have the habit of giving incomplete and sometimes misleading information. There can be no doubt, however, that they are so pretty and give information so quickly that they will continue to survive and be used.

Relatively simple fundamental calculations yield formulas in which the first two or the second two of the quantities

$$\frac{\partial}{\partial y}\frac{\partial u}{\partial x}, \qquad \frac{\partial}{\partial x}\frac{\partial u}{\partial y}, \qquad u_{xy}(x,y), \qquad u_{yx}(x,y) \tag{11.15}$$

both appear and will cancel out if we can be sure that they are equal. Taking a partial derivative with respect to x is, like putting on our shoes, a procedure that is called an *operation*. Taking a partial derivative with respect to y is, like putting on our socks, another operation. The question whether the mixed derivatives in (11.15) are equal is therefore the question whether two operations *commute*, that is, whether the result of performing the two operations in tandem (that is, one after the other) is independent of the order in which the operations are performed. Correct ideas about the problem involving shoes and socks can be obtained by experimentation. The remainder of the text of this section is devoted to the problem involving partial derivatives. We begin with some definitions in which n can be 1 or 2 or 3 or 416 and the variables are usually denoted by x, y or x, y, z instead of $x_1, x_2, \cdots, x_n$ when there are only two or three of them.

Definition 11.16 *A function f of n variables $x_1, x_2, \cdots, x_n$ is said to have the limit L as $x_1, x_2, \cdots, x_n$ approaches $a_1, a_2, \cdots, a_n$ and we write*

$$\lim_{x_1,x_2,\cdots,x_n\to a_1,a_2,\cdots,a_n} f(x_1,x_2,\cdots,x_n) = L$$

if to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that

$$|f(x_1,x_2,\cdots,x_n) - L| < \epsilon$$

whenever

$$0 < \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2 + \cdots + (x_n - a_n)^2} < \delta.$$

Definition 11.17 *A function f of n variables $x_1, x_2, \cdots, x_n$ is said to be continuous at $a_1, a_2, \cdots, a_n$ if*

$$\lim_{x_1,x_2,\cdots,x_n\to a_1,a_2,\cdots,a_n} f(x_1,x_2,\cdots,x_n) = f(a_1,a_2,\cdots,a_n).$$

We are now prepared to state a fundamental theorem which guarantees equality of f_{yx} and f_{xy} whenever these and some other derivatives exist and are continuous; the fact that the theorem gives additional information is interesting but less important.

Theorem 11.18 *If $u(x,y)$, $u_x(x,y)$, $u_y(x,y)$, and $u_{xy}(x,y)$ all exist and are continuous over some circular disk consisting of points (x_1,y_1) for which $(x_1 - x)^2 + (y_1 - y)^2 < \delta$, then $u_{yx}(x,y)$ exists and*

$$u_{yx}(x,y) = u_{xy}(x,y). \tag{11.181}$$

Proof of this theorem is quite tricky because it requires two applications of the mean-value theorem 5.52, and the first of these applications must be made in a particular special way to produce the required result. We shun the curly dees and use the subscript notation so we can know what we are doing. To approach the derivative $u_{yx}(x,y)$ about which we must learn, we observe that the derivatives in

$$u_y(x,y) = \lim_{\Delta y\to 0} \frac{u(x, y + \Delta y) - u(x,y)}{\Delta y} \tag{11.182}$$

$$u_y(x + \Delta x, y) = \lim_{\Delta y\to 0} \frac{u(x + \Delta x, y + \Delta y) - u(x + \Delta x, y)}{\Delta y} \tag{11.183}$$

exist when $|\Delta x|$ is sufficiently small. We must prove that the limit in

$$u_{yx}(x,y) = \lim_{\Delta x\to 0} \frac{u_y(x + \Delta x, y) - u_y(x,y)}{\Delta x} \tag{11.184}$$

exists and is $u_{xy}(x,y)$. Substituting for the terms in the numerator of the right member gives

(11.185)
$$u_{yx}(x,y) = \lim_{\Delta x\to 0}\lim_{\Delta y\to 0} \frac{u(x+\Delta x, y+\Delta y) - u(x+\Delta x, y) - u(x, y+\Delta y) + u(x,y)}{\Delta x\, \Delta y}.$$

Supposing for the moment that y and $y + \Delta y$ remain fixed like good numbers usually do, we define a function ϕ by the formula

$$(11.186) \qquad \phi(t) = u(t, y + \Delta y) - u(t,y).$$

The numerator N of the ponderous quotient in (11.185) is then found to be the second member of the formula

$$(11.187) \qquad \begin{aligned} N &= \phi(x + \Delta x) - \phi(x) = \phi'(\xi)\,\Delta x \\ &= [u_x(\xi, y + \Delta y) - u_x(\xi,y)]\,\Delta x, \end{aligned}$$

and an application of the mean-value theorem and (11.186) then gives the rest of the formula in which ξ is a number between x and $x + \Delta x$. Substituting the last member of (11.187) for the numerator in (11.185) gives

$$(11.188) \qquad u_{yx}(x,y) = \lim_{\Delta x \to 0} \lim_{\Delta y \to 0} \frac{u_x(\xi, y + \Delta y) - u_x(\xi,y)}{\Delta y}.$$

Since $u_x(\xi,t)$ is a differentiable function of t over the interval from y to $y + \Delta y$, another application of the mean-value theorem gives

$$(11.189) \qquad u_{yx}(x,y) = \lim_{\Delta x \to 0} \lim_{\Delta y \to 0} u_{xy}(\xi,\eta),$$

where η lies between y and $y + \Delta y$. Matters are complicated by the fact that both ξ and η can depend upon Δy, and the next step is a delicate one that demands careful attention in a rigorous course in advanced calculus. The fact that the limit in the left member of the formula

$$(11.1891) \qquad \lim_{\Delta y \to 0} u_{xy}(\xi,\eta) = u_{xy}(\xi^*,y)$$

exists and the fact that u_{xy} is continuous imply that there is a number ξ^* for which $|x - \xi^*| \leqq |\Delta x|$ and the formula holds. Since u_{xy} is continuous, this and (11.189) give

$$(11.1892) \qquad u_{yx}(x,y) = \lim_{\Delta x \to 0} u_{xy}(\xi^*,y) = u_{xy}(x,y).$$

Thus the limits exist and have the required value. It is not expected that the above proof of Theorem 11.18 will be "learned" in this course, but we need not be blissfully unaware of the facts that the theorem is important and that we could learn very much about partial derivatives and limits if we would (as is often done in advanced calculus) invest enough time to make a thorough study of the proof.

Two observations can be made. In the first place, it is easy to insert an extra variable z in Theorem 11.18 and its proof to obtain the formula

$$(11.1893) \qquad u_{xy}(x,y,z) = u_{yx}(x,y,z)$$

when the appropriate functions and derivatives are continuous. In the second place, continuity of appropriate functions and derivatives allows us to make any change we please in order of differentiation when u is differentiated more than once with respect to variables upon which it depends.

Problems 11.19

1 Let

$$f(x,y) = x^2 + y^2$$

and observe that $f_x(x,y) = 2x$. Then perform every single step required to show in the most tedious possible way that

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x,y)}{\Delta x} = 2x.$$

Then think about the whole business.

2 If

$$u(x,y,z) = x^2 + 2y^2 + 3z^2,$$

show that

$$\begin{array}{lll} u_x(x,y,z) = 2x & u_y(x,y,z) = 4y & u_z(x,y,z) = 6z \\ u_x(1,1,1) = 2, & u_y(1,1,1) = 4, & u_z(1,1,1) = 6. \end{array}$$

3 Supposing that $\rho > 0$ and

$$\rho^2 = x^2 + y^2, \qquad \phi = \tan^{-1} \frac{y}{x},$$

find the first partial derivatives of ρ and ϕ with respect to x and y and then use the formulas $x = \rho \cos \phi$, $y = \rho \sin \phi$ to put the results in the form

$$\frac{\partial \rho}{\partial x} = \cos \phi, \qquad \frac{\partial \rho}{\partial y} = \sin \phi, \qquad \frac{\partial \phi}{\partial x} = -\frac{\sin \phi}{\rho}, \qquad \frac{\partial \phi}{\partial y} = \frac{\cos \phi}{\rho}.$$

4 Obtain the simplest possible expression for

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$$

when

(a) $u = x^2 - y^2$ (b) $u = 3x^2y - y^3$

(c) $u = \log (x^2 + y^2)$ (d) $u = e^x \cos y$

(e) $u = x \sin y$ (f) $u = \tan^{-1} \dfrac{y}{x}$

(g) $u = \log \sqrt{(x - a)^2 + (y - b)^2}$

Ans.: With one exception, each answer is 0. *Remark:* The equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

and those that appear in the following problems are called *partial differential equations*. It is never too early to start learning that the above equation is the *Laplace equation*. A function u is said to be *harmonic* over a region if it satisfies the Laplace equation and is continuous and has continuous partial derivatives at each point of the region. Problem 12 gives an example of a function which satisfies the Laplace equation over the entire plane but is, nevertheless, not harmonic over regions containing the origin.

5 Prove that

$$x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = 0 \tag{1}$$

when $u = y/x$, when $u = \log(y/x)$, and when $u = \sin(y/x)$. Continue operations to prove that if f is a differentiable function of one variable, then

$$\frac{\partial}{\partial x}f\left(\frac{y}{x}\right) = f'\left(\frac{y}{x}\right)\frac{-y}{x^2}, \qquad \frac{\partial}{\partial y}f\left(\frac{y}{x}\right) = f'\left(\frac{y}{x}\right)\frac{1}{x}$$

and (1) still holds.

6 Prove that the *wave equation*

$$a^2\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2}$$

is satisfied when

(*a*) $u = (x + at)^3$ (*b*) $u = (x - at)^5$
(*c*) $u = e^{x+at}$ (*d*) $u = \sin(x - at)$
(*e*) $u = f(x + at)$ (*f*) $u = g(x - at)$,

it being supposed that f and g are twice-differentiable functions.

7 Show that the function defined by (11.11) satisfies the *heat equation*

$$a^2\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}.$$

8 Show that each of the following functions satisfies the equation written opposite it:

(*a*) $u = ax + by$ $\qquad x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = u$

(*b*) $u = \sin(x \sin y)$ $\qquad \cos y\frac{\partial u}{\partial x} - \sin y\frac{\partial u}{\partial y} = 0$

(*c*) $u = \dfrac{1}{\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}}$ $\qquad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$

(*d*) $u = (x - y)(y - z)(z - x)$ $\qquad \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z} = 0$

9 To simplify matters, we suppose that each function appearing in our work is continuous and has continuous partial derivatives of first and second orders. We begin acquaintance with the idea that if F or $F(x,y,z)$ is a scalar function, then the vector function ∇F defined by

$$\nabla F = \frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j} + \frac{\partial F}{\partial z}\mathbf{k} \tag{1}$$

is called the *gradient* of F. If $\mathbf{V}$ is a vector function defined by

$$\mathbf{V}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}, \tag{2}$$

then the scalar function $\nabla\cdot\mathbf{V}$ defined by

$$\nabla\cdot\mathbf{V} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \tag{3}$$

is called the *divergence* of $\mathbf{V}$. Finally, the vector function $\nabla \times \mathbf{V}$ defined by

(4) $$\nabla \times \mathbf{V} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$

is called the *curl* of $\mathbf{V}$. The expanded form of (4) is

(5) $$\nabla \times \mathbf{V} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}.$$

The inverted delta is called "del," and we shall hear more about the formula

(6) $$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}.$$

Meanwhile, use the definitions to show that

(7) $$\nabla \cdot (\nabla F) = \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2}$$

(8) $$\nabla \times (\nabla F) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \dfrac{\partial F}{\partial x} & \dfrac{\partial F}{\partial y} & \dfrac{\partial F}{\partial z} \end{vmatrix} = 0.$$

10 Start with the formulas

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}$$
$$F(x,y,z) = x^2 + y^2 + z^2$$
$$\mathbf{V}(x,y,z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

and calculate all of the following things that are meaningful:

(*a*) ∇F (*b*) $\nabla \cdot F$ (*c*) $\nabla \times F$
(*d*) $\nabla \mathbf{V}$ (*e*) $\nabla \cdot \mathbf{V}$ (*f*) $\nabla \times \mathbf{V}$
(*g*) $\nabla \cdot (\nabla F)$ (*h*) $\nabla \times (\nabla F)$ (*i*) $\nabla(\nabla F)$
(*j*) $\nabla(\nabla \mathbf{V})$ (*k*) $\nabla \times (\nabla \times \mathbf{V})$ (*l*) $\nabla \cdot (\nabla \cdot \mathbf{V})$

11 Letting u be the thoroughly reasonable function having the value $\frac{1}{2}\rho \sin 2\phi$ or $\rho \sin \phi \cos \phi$ at the point having polar coordinates ρ and ϕ, show that, in rectangular coordinates,

(1) $$u(0,0) = 0, \qquad u(x,y) = \frac{xy}{(x^2 + y^2)^{1/2}} \qquad (x^2 + y^2 \neq 0).$$

Show that

(2) $$\frac{\partial u}{\partial x} = \frac{y^3}{(x^2 + y^2)^{3/2}}, \qquad \frac{\partial u}{\partial y} = \frac{x^3}{(x^2 + y^2)^{3/2}} \qquad (x^2 + y^2 \neq 0).$$

Show that

(3) $$\lim_{y \to 0+} u_x(0,y) = 1, \qquad \lim_{y \to 0-} u_x(0,y) = -1.$$

Show that $u_x(x,0) = 0$ for each x and that $u_y(0,y) = 0$ for each y.

12 The function u for which $u(0,0) = 0$ and

(1) $$u(x,y) = \frac{2xy}{(x^2+y^2)^2} \qquad (x^2+y^2 \neq 0)$$

can be considered more than once. Show that

(2) $$u_{xx}(x,y) = 24xy\frac{x^2-y^2}{(x^2+y^2)^4}, \qquad u_{yy}(x,y) = 24xy\frac{y^2-x^2}{(x^2+y^2)^4}$$

and hence

(3) $$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = u_{xx}(x,y) + u_{yy}(x,y) = 0$$

when $x^2 + y^2 \neq 0$. Show that (3) holds when $x = y = 0$. Show that

(4) $$\lim_{x\to 0} u(x,x) = \infty.$$

13 Formulas more or less like

(1) $$F(x) = \int_a^b f(x,y)\,dy$$

often appear in pure and applied mathematics. We suppose that, for each x in some interval, the integral in the right member of (1) exists and is a number $F(x)$. More advanced courses set forth conditions under which $F'(x)$ exists and can be obtained by "differentiating with respect to x under the integral sign" so that

(2) $$F'(x) = \int_a^b \left[\frac{\partial}{\partial x} fx(x,y)\right] dy.$$

When this procedure is correct, we can combine (1) and (2) to obtain the formula

(3) $$\frac{d}{dx}\int_a^b f(x,y)\,dy = \int_a^b \left[\frac{\partial}{\partial x} f(x,y)\right] dy.$$

Verify that (3) is correct when

(*a*) $a = 0$, $b = 1$, $f(x,y) = x + y$
(*b*) $a = 0$, $b = 1$, $f(x,y) = x^2 + y^2$
(*c*) $a = 1$, $b = 2$, $f(x,y) = \dfrac{\sin xy}{x}$

11.2 *Increments, chain rule, and gradients*

Throughout this section, we suppose that u is a function of three variables x, y, z and we restrict attention to a region R in E_3 over which u is continuous and has continuous partial derivatives u_x, u_y, and u_z. In many examples the region R is the whole E_3, but this is not necessarily so. Whenever a useful purpose is served, we can regard $u(x,y,z)$ as being the temperature or pressure or potential or density or humidity at the point $P(x,y,z)$. While $u(x,y,z)$ cannot be a vector, it can be the scalar component in some particular direction of a vector. In any case, it should be recognized that our function can be of interest to men as well as to boys. Everything we do will depend upon the fundamental fact that there is an astonishingly effective way of estimating the number Δu, defined by

$$\Delta u = u(x + \Delta x, y + \Delta y, z + \Delta z) - u(x,y,z), \tag{11.21}$$

which represents the difference (or increment) between the values of u at two places. The basic trick is to divide and conquer the huge discrepancy between the natures of the two terms in the right member of (11.21) by subtracting and adding terms to obtain the telescopic sum in

$$\begin{aligned}\Delta u &= u(x + \Delta x, y + \Delta y, z + \Delta z) - u(x, y + \Delta y, z + \Delta z) \\ &\quad + u(x, y + \Delta y, z + \Delta z) - u(x, y, z + \Delta z) \\ &\quad + u(x, y, z + \Delta z) - u(x,y,z).\end{aligned} \tag{11.22}$$

Defining a function $\phi(t)$ by the formula

$$\phi(t) = u(t, y + \Delta y, z + \Delta z) \tag{11.221}$$

enables us to apply the mean-value theorem to the difference in the first line, and similar functions apply to the other two. Thus we obtain $\Delta u = u_x(x^*, y + \Delta y, z + \Delta z)\,\Delta x + u_y(x, y^*, z + \Delta z)\,\Delta y + u_z(x,y,z^*)\,\Delta z$, where x^* lies between x and $x + \Delta x$, y^* lies between y and $y + \Delta y$, and z^* lies between z and $z + \Delta z$. Our hypothesis that the derivatives are continuous allows us to put this in the form

(11.222)

$$\Delta u = [u_x(x,y,z) + \epsilon_1]\,\Delta x + [u_y(x,y,z) + \epsilon_2]\,\Delta y + [u_z(x,y,z) + \epsilon_3]\,\Delta z$$

where ϵ_1, ϵ_2, ϵ_3 are quantities which approach zero as Δx, Δy, Δz approach zero. This formula can be put in the form

$$\Delta u = \left[\frac{\partial u}{\partial x} + \epsilon_1\right]\Delta x + \left[\frac{\partial u}{\partial y} + \epsilon_2\right]\Delta y + \left[\frac{\partial u}{\partial z} + \epsilon_3\right]\Delta z. \tag{11.223}$$

It is quite possible to rub out the epsilons, replace deltas by dees, write the formula

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy + \frac{\partial u}{\partial z}\,dz, \tag{11.224}$$

and then undertake to explain the antics. At least for the present we adopt different tactics.

Let three functions x, y, z be defined and differentiable over some interval T of values of t, and let the point $P(t)$ having coordinates $x(t)$, $y(t)$, $z(t)$ trace a curve C in R as t increases over T. Let

$$w(t) = u(x(t), y(t), z(t)). \tag{11.225}$$

Looking forward to derivation of a formula (the chain formula) for $w'(t)$ we write

$$w(t + \Delta t) = u(x(t + \Delta t), y(t + \Delta t), z(t + \Delta t)).$$

Thus

$$w(t + \Delta t) = u(x(t) + \Delta x, y(t) + \Delta y, z(t) + \Delta z),$$

where

$$\Delta x = x(t + \Delta t) - x(t), \quad \Delta y = y(t + \Delta t) - y(t), \quad \Delta z = z(t + \Delta t) - z(t).$$

Applying (11.222) then gives

$$\begin{aligned} w(t + \Delta t) - w(t) = {} & [u_x(x(t), y(t), z(t)) + \epsilon_1][x(t + \Delta t) - x(t)] \\ & + [u_y(x(t), y(t), z(t)) + \epsilon_2][y(t + \Delta t) - y(t)] \\ & + [u_z(x(t), y(t), z(t)) + \epsilon_3][z(t + \Delta t) - z(t)], \end{aligned}$$

where ϵ_1, ϵ_2, ϵ_3 are quantities which approach zero as Δt approaches zero. Dividing by Δt and taking limits as Δt approaches zero gives the *chain formula* (11.232). The following theorem sets forth conditions under which the chain formula is correct.

Theorem 11.23 (**chain rule**) *If u is continuous and has continuous partial derivatives u_x, u_y, u_z over a region R in E_3, and if*

$$w(t) = u(x(t), y(t), z(t)), \tag{11.231}$$

where x, y, z are differentiable functions of t over some interval T, and if the point $P(t)$ having coordinates $x(t)$, $y(t)$, $z(t)$ traverses a curve C in R as t increases over T, then w is differentiable over T and

$$\begin{aligned} w'(t) = {} & u_x(x(t), y(t), z(t))x'(t) \\ & + u_y(x(t), y(t), z(t))y'(t) \\ & + u_z(x(t), y(t), z(t))z'(t) \end{aligned} \tag{11.232}$$

when t is in T.

In case $[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2 = 1$ so that $P(t)$ moves along C with unit speed, the number $w'(t)$ is called the *directional derivative* of u in the direction of the forward tangent to C at P. For this and other reasons, some of which will appear later, the chain formula (11.232) is extremely important. We temporarily suspend production of mathematics to ponder consequences of the rude fact that it takes a long time to write the formulas (11.231) and (11.232). We can wonder how much we can abbreviate these formulas without abbreviating the life out of them and without creating confusion that wastes more of our time than the abbreviations save. We can write (11.231) in the form

$$u = u(x,y,z) \tag{11.233}$$

and carry in our minds the idea that the left side stands for the value of a function of t and so also does the right side but, on the right side, u is linked to t by the intermediate variables x, y, z which are functions of t. We can then abbreviate (11.232) to

$$\frac{du}{dt} = \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial y}\frac{dy}{dt} + \frac{\partial u}{\partial z}\frac{dz}{dt}. \tag{11.234}$$

It turns out that, in practice, the version (11.234) of the chain formula is often very convenient. For example, if

$$u = x^2 + y^2 + z^2, \qquad x = \sin t, \qquad y = e^t, \qquad z = t^2,$$

then

$$\frac{du}{dt} = 2x \cos t + 2ye^t + 4zt$$

and it is, in fact, not always necessary or even desirable to express the right side entirely in terms of t. The abbreviations are not always so agreeable, however. In case the parameter t is x itself, we must employ considerable fortitude to comprehend the sentence containing (11.233) and the formula

$$\frac{du}{dx} = \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\frac{dy}{dx} + \frac{\partial u}{\partial z}\frac{dz}{dx}. \tag{11.235}$$

It is awkward and perhaps even undesirable to be required to think of u depending upon x in two different ways. It is easier to let

$$w(x) = u(x, y(x), z(x))$$

and to write

$$\frac{dw}{dx} = \frac{\partial u}{\partial x}\frac{dx}{dx} + \frac{\partial u}{\partial y}\frac{dy}{dx} + \frac{\partial u}{\partial z}\frac{dz}{dx}$$

and then observe that $dx/dx = 1$. While Theorem 11.23 is the fully meaningful theorem to which we can refer when meanings of symbols must be carefully set forth, we give a restatement of the theorem in terms of the simpler curly dee notation.

Theorem 11.24 (**chain rule, second version**) *If u is continuous and has continuous partial derivatives $\partial u/\partial x$, $\partial u/\partial y$, $\partial u/\partial z$ over a region R in E_3, and if*

$$u = u(x,y,z), \tag{11.241}$$

where x, y, z are differentiable functions of t over some interval T, and if the point P having coordinate (x,y,z) traverses a curve C in R as t increases over T, then u is differentiable over T and

$$\frac{du}{dt} = \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial y}\frac{dy}{dt} + \frac{\partial u}{\partial z}\frac{dz}{dt}. \tag{11.242}$$

It is not often that we have an opportunity to make an observation as valuable as the one that the right member of (11.242) is the scalar product of two vectors. Thus

$$(11.25)\qquad \frac{du}{dt} = \left(\frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j} + \frac{\partial u}{\partial z}\mathbf{k}\right)\cdot\left(\frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k}\right).$$

The remarkable feature of this product is the fact that, for each t, the first vector depends only upon the first partial derivatives of u at the point $P(x,y,z)$ and the second vector is (when it is not 0) simply a forward tangent to C at P.

Information is ready to gush from (11.25), and we make progress by learning about the first vector which is called the *gradient* of u at P and is denoted by ∇u so that

$$(11.26)\qquad \nabla u = \frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j} + \frac{\partial u}{\partial z}\mathbf{k}.$$

The symbol ∇, an inverted delta, is read "del" and ∇u is read "del u." We must always remember that ∇u is a vector. For many purposes, it is very helpful to consider ∇ itself to be an operator,

$$(11.261)\qquad \nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k},$$

which can be applied to a scalar function u having continuous partial derivatives to produce the gradient vector ∇u. In case $\nabla u = 0$, this is the whole story and there is nothing more to be learned. Henceforth we suppose that $\nabla u \neq 0$. Let the last vector in (11.25) be a unit vector $\mathbf{v}$ so that du/dt is the directional derivative of u at P in the direction of $\mathbf{v}$. Then

$$(11.262)\qquad \frac{du}{dt} = (\nabla u)\cdot\mathbf{v} = |\nabla u| \cos\theta,$$

where $|\nabla u|$ is the length of the gradient ∇u and θ is the angle between the vectors ∇u and $\mathbf{v}$. Since $-1 \leqq \cos\theta \leqq 1$ and $\cos 0 = 1$, it follows that the direction of ∇u is the direction in which the directional derivative of u at P attains its maximum value and that the length of ∇u is this maximum. This is the fundamental intrinsic property of the gradient of u at P. Since $\cos\pi = -1$, the direction opposite to that of ∇u is the direction in which the directional derivative of u at P attains its minimum value, and $-|\nabla u|$ is the minimum. Some applications are easy to understand. If u is temperature and Mr. Walker is at a place that is too cold to suit him, he will walk in the direction of the gradient of u, and the length of the gradient will tell him the rate (in degrees per meter, for example) at which his position becomes more comfortable.

Since $\cos\pi/2 = 0$, directional derivatives in directions orthogonal to the gradient ∇u will be 0, and this can make us think about *level surfaces* (isothermal surfaces or equipotential surfaces, for example) upon which u has a particular constant value. If $P_0(x_0,y_0,z_0)$ and some sphere with center at P_0 lie in our region R, then our hypotheses (including the hypothesis that $\nabla u \neq 0$ at P_0) imply that there is a surface S consisting of points $P(x,y,z)$ for which $u(x,y,z) = u(x_0,y_0,z_0)$. Let C be a curve which lies on S and passes through P_0 and has the vector **v** for its forward tangent at P_0. Supposing that a point moving along this curve has coordinates $x(t)$, $y(t)$, $z(t)$, we have

$$u(x(t),\, y(t),\, z(t)) = u(x_0,\, y_0,\, z_0) \tag{11.27}$$

and hence $du/dt = 0$. Therefore, $(\nabla u)\cdot\mathbf{v} = 0$. Thus ∇u is orthogonal to each line which passes through P_0 and which is tangent to a curve on S. As Figure 11.271 may suggest, this is just what we mean when we say that ∇u is a normal to the surface S at the point P_0. Therefore, we can find the normal to the surface S having the equation $u(x,y,z) = c$, at a point P_0 on S, by finding the gradient of u at P_0. To find the plane tangent to S at P_0 is easy; it is the plane through P_0 orthogonal to the gradient.

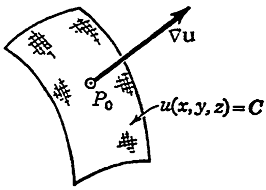


Figure 11.271

A review of matters relating to equations of lines and tangent planes may be in order. If the gradient of u at a point $P_0(x_0,y_0,z_0)$ is

$$A\mathbf{i} + B\mathbf{j} + C\mathbf{k},$$

then the equations of the line through P_0 upon which this gradient lies are

$$\frac{x - x_0}{A} = \frac{y - y_0}{B} = \frac{z - z_0}{C} \tag{11.28}$$

and the equation of the plane through P_0 normal to the gradient is

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0. \tag{11.281}$$

The equations (11.28) are correct because they say that the scalar components of the vector from $P_0(x_0,y_0,z_0)$ to $P(x,y,z)$ are proportional to the scalar components of the gradient. The equation (11.281) is correct because it says that the vector from $P_0(x_0,y_0,z_0)$ to $P(x,y,z)$ is orthogonal to the gradient.

Problems 11.29

1 This problem requires us to think about some ways in which the formulas (11.21) to (11.25) can be used to solve problems. Suppose f is a given function having continuous partial derivatives. Suppose we want $f(x,y,z)$ but we calculate $f(x_0,y_0,z_0)$ because it is easier to calculate $f(x_0,y_0,z_0)$ or because x, y, z are unknown and x_0, y_0, z_0 are the numbers we got when we measured them. How can we estimate the error resulting from using $f(x_0,y_0,z_0)$ instead of $f(x,y,z)$? *Ans.:* Put

$$x = x_0, \qquad y = y_0, \qquad z = z_0, \qquad \Delta x = x - x_0, \qquad \Delta y = y - y_0, \qquad \Delta z = z - z_0$$

in formulas from (11.21) to (11.223). *Remark:* Very often we do not want to hunt up books and copy formulas from them. See the next problem.

2 Remember the following *modus operandi* because it is useful when properly used. As an approximation to the number Δu defined by

$$\Delta u = u(x + dx,\, y + dy,\, z + dz) - u(x,y,z) \tag{1}$$

use the number du defined by

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy + \frac{\partial u}{\partial z}\,dz. \tag{2}$$

Remember that (2) can, in appropriate circumstances, be obtained by differentiating with respect to t by the chain rule (Theorem 11.24) and multiplying by dt. Note the similarity between this *modus operandi* and the one involving (3.96)

3 Supposing that $y = \rho \sin \phi$, derive the formula

$$|dy| \leqq |\rho \cos \phi|\,|d\phi| + |\sin \phi|\,|d\rho|$$

which gives information about the error in y resulting from use of erroneous values of ρ and ϕ.

4 Supposing that $y = \rho \cos \phi$, where ρ and ϕ are functions of t, find a formula for dy/dt in two different ways. First use partial derivatives and the chain rule. Then use ordinary (not partial) derivatives and the rule for differentiating products of functions of t. Make the results agree.

5 Formulate and solve another problem more or less like the preceding one.

6 Supposing that A, B, C are constants for which $A^2 + B^2 + C^2 = 1$, find the directional derivative of the function (or potential function)

$$V = \frac{1}{\sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2}} \tag{1}$$

at the point (x_0,y_0,z_0) in the direction of the vector

$$\mathbf{D} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k} \tag{2}$$

by two different (or superficially different) methods. In the first place, put

$$x = x_0 + At, \qquad y = y_0 + Bt, \qquad z = z_0 + Ct \tag{3}$$

in (1) and find dV/dt by differentiating with respect to t without use of partial derivatives. Then put $t = 0$. In the second place, show that

$$\nabla V = -\frac{(x - x_1)\mathbf{i} + (y - y_1)\mathbf{j} + (z - z_1)\mathbf{k}}{[(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2]^{3/2}}$$

and calculate the scalar product $(\nabla V)\cdot\mathbf{D}$. Then put $x = x_0$, $y = y_0$, $z = z_0$. Make the results agree. *Remark:* It is worthwhile to observe that, by introducing summation signs, we can easily extend all of these calculations to cover more complex situations in which n is a positive integer and

$$V = \sum_{k=1}^{n} \frac{m_k}{\sqrt{(x - x_k)^2 + (y - y_k)^2 + (z - z_k)^2}}.$$

There are times when it is not unreasonable to start with a special situation in which $n \geqq 3$ and make quite extensive calculations to learn about directional derivatives, gradients, and equipotential surfaces. An equipotential surface is a surface consisting of points (x,y,z) such that, for some constant c, $V(x,y,z) = c$.

7 Let the temperature u at the point (x,y) in an xy plane be defined by

$$u = e^{-x} \sin y.$$

Modify the procedure of Problem 6 to obtain, by two methods, the directional derivative of u at (x_0,y_0) in the direction of the unit vector

$$(\cos \alpha)\mathbf{i} + (\sin \alpha)\mathbf{j}.$$

Make the results agree.

8 Supposing that

$$u = Ax + By + Cz,$$

show that

$$\nabla u = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}.$$

Use this result to show that the equation of the plane tangent to the graph of the equation

$$Ax + By + Cz = D$$

at a point $P_0(x_0,y_0,z_0)$ on the graph is

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$$

or

$$Ax + By + Cz = D.$$

9 Supposing that

$$u = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}, \tag{1}$$

show that

$$\nabla u = \frac{2x}{a^2}\mathbf{i} + \frac{2y}{b^2}\mathbf{j} + \frac{2z}{c^2}\mathbf{k}. \tag{2}$$

Use this result to show that the equation of the plane tangent to the graph of the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{3}$$

at the point (x_0,y_0,z_0) on the graph is

$$\frac{2x_0}{a^2}(x - x_0) + \frac{2y_0}{b_0}(y - y_0) + \frac{2z_0}{c^2}(z - z_0) = 0 \tag{4}$$

or

$$\frac{x_0x}{a^2} + \frac{y_0y}{b^2} + \frac{z_0z}{c^2} = 1.$$

Remark: In case $a = b = c$, the graph of (3) is a sphere. Otherwise, the graph is an ellipsoid; see Figure 6.631 and the accompanying discussion.

10 When

$$u = x^2 + y^2 - z^2,$$

the graph of the equation $u = 0$ is a cone having its vertex at the origin. If $P_0(x_0,y_0,z_0)$ is a point on the cone which is not the origin, show that the equation of the plane tangent to the cone at P_0 is

$$x_0x + y_0y - z_0z = 0.$$

Note that the gradient at the origin is 0 and that we have no information about planes (if any) that are tangent to a cone at its vertex.

11 The graph of the equation

$$(1) \qquad z = \frac{y^2}{b^2} - \frac{x^2}{a^2}$$

is a *hyperbolic paraboloid* or *saddle surface* more or less like that shown in Figure 6.672. The sections in planes parallel to the xz and yz planes are parabolas while other sections are hyperbolas. Letting

$$(2) \qquad u = z + \frac{x^2}{a^2} - \frac{y^2}{b^2},$$

find the gradient of u at the point $P_0(x_0,y_0,z_0)$ on the graph of (1) and show that the equation of the plane tangent to the graph at P_0 is

$$(3) \qquad \frac{x_0x}{a^2} - \frac{y_0y}{b^2} + \frac{z + z_0}{2} = 0.$$

12 Letting u be the left member of the equation

$$(1) \quad a_{11}xx + a_{12}xy + a_{13}xz + b_1(x + x) + a_{21}yx + a_{22}yy + a_{23}yz + b_2(y + y) \\ + a_{31}zx + a_{32}zy + a_{33}zz + b_3(z + z) = c,$$

where $a_{21} = a_{12}$, $a_{31} = a_{13}$, and $a_{32} = a_{23}$, show that

$$(2) \quad \nabla u = 2[a_{11}x + a_{12}y + a_{13}z + b_1]\mathbf{i} + 2[a_{21}x + a_{22}y + a_{23}z + b_2]\mathbf{j} \\ + 2[a_{31}x + a_{32}y + a_{33}z + b_3]\mathbf{k}.$$

Supposing that the graph of (1) is a quadric surface S containing a point $P_0(x_0,y_0,z_0)$ at which $\nabla u \neq 0$, write the equation of the plane tangent to S at P_0. Put the equation in the form obtained from (1) by the following ritual. Wherever the product or sum of two variables appears, award a subscript to the first factor or term.

13 Prove that each plane tangent to the surface having the equation

$$x^{2/3} + y^{2/3} + z^{2/3} = a^{2/3}$$

intersects the coordinate axes in points that are projections upon these axes of a point on the sphere having the equation

$$x^2 + y^2 + z^2 = a^2.$$

14 Find the gradient of the function F for which

$$F(x,y,z) = xyz - 1,$$

find equations of the line normal to the surface having the equation

$$xyz = 1$$

at the point (x_0,y_0,z_0), and find the equation of the plane tangent to the latter surface at the latter point. *Ans.:*

$$\nabla F = yz\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$$
$$\frac{x - x_0}{y_0z_0} = \frac{y - y_0}{x_0z_0} = \frac{z - z_0}{x_0y_0}$$
$$y_0z_0(x - x_0) + x_0z_0(y - y_0) + x_0y_0(z - z_0) = 0.$$

15 Prove that if the plane π is tangent to the graph of the equation $xyz = a^3$, then π and the three coordinate planes are the boundaries of a tetrahedron having volume $9a^3/2$.

16 Prove that if u, u_x, u_y, and u_z are continuous over a spherical ball having its center at $P_0(x_0,y_0,z_0)$, then u cannot have even a local minimum or a local maximum at P_0 unless

$$u_x(x_0,y_0,z_0) = u_y(x_0,y_0,z_0) = u_z(x_0,y_0,z_0) = 0.$$

Hint: Tell what can be done when the gradient at P_0 is not 0. If necessary, look at (11.262).

17 Using the result of Problem 16, show that if

$$u = \frac{x + y + z}{1 + x^2 + y^2 + z^2}$$

attains a maximum at $P(x_0,y_0,z_0)$, then $x_0 = y_0 = z_0$. With this assistance, find the place where u is maximum and show that the maximum value of u is $\sqrt{3}/2$.

18 *Remark:* This is a remark for those who wish to see that gradients can be used to introduce a fundamental idea that is often used to solve more difficult problems. Suppose we are required to find numbers x, y, z for which $f(x,y,z)$ is a minimum or maximum when x, y, z are required to satisfy a supplementary condition of the form $u(x,y,z) = c$. To catch the idea, we suppose that

$$f(x,y,z) = x^2 + y^2 + z^2$$

and that the graph of $u = c$ is a surface more or less like that shown in Figure 11.291. It is easy to guess that if a maximum or a minimum occurs at $P(x,y,z)$, then the gradients of f and u at P must have the same or opposite directions and hence that there must be a constant λ such that

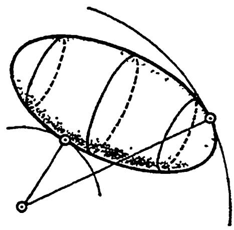


Figure 11.291

$$\nabla f + \lambda \nabla u = 0$$

or

$$\nabla(f + \lambda u) = 0. \tag{1}$$

We now make a profound observation. Finding values of x, y, z, and λ for which (1) holds is equivalent to finding values of x, y, z, and λ for which the first partial derivatives of the function w define by

$$w(x,y,z) = f(x,y,z) + \lambda[u(x,y,z) - c] \tag{2}$$

are 0. All this leads us to an idea that can sound very strange and which might be quite useless if it were not for the fact that it is the very useful idea that the

Lagrange multiplier λ should be used in the following way. When we want to find numbers x, y, z for which $f(x,y,z)$ takes extreme values when x, y, z satisfy the supplementary condition $u(x,y,z) = c$, we seek numbers x, y, z, λ for which the function w defined by (2) takes extreme values. We illustrate use of Lagrange multipliers by using them to obtain information about a thoroughly important problem. Letting

$$\begin{aligned} u = {} & a_{11}xx + a_{12}xy + a_{13}xz \\ & + a_{21}yx + a_{22}yy + a_{23}yz \\ & + a_{31}zx + a_{32}zy + a_{33}zz, \end{aligned}$$

where $a_{21} = a_{12}$, $a_{31} = a_{13}$, and $a_{32} = a_{23}$, we seek points on the quadric surface having the equation $u = c$ which lie at least and (if they exist) greatest distances from the origin. In this and similar problems, we systematically use the idea that distances have extreme values when their squares do. To make answers come out in the forms that are familiar when this problem is attacked by other methods, we modify (2) by writing $-\lambda^{-1}$ in place of λ and define w by the formula

$$w = x^2 + y^2 + z^2 - \lambda^{-1}[u(x,y,z) - c].$$

Equating the first-order partial derivatives to 0 gives the equations

$$\begin{aligned} \frac{\partial w}{\partial x} &= 2x - 2\lambda^{-1}(a_{11}x + a_{12}y + a_{13}z) = 0 \\ \frac{\partial w}{\partial y} &= 2y - 2\lambda^{-1}(a_{21}x + a_{22}y + a_{23}z) = 0 \\ \frac{\partial w}{\partial z} &= 2z - 2\lambda^{-1}(a_{31}x + a_{32}y + a_{33}z) = 0 \end{aligned}$$

$$\begin{aligned} (a_{11} - \lambda)x + a_{12}\,y + a_{13}\,z &= 0 \\ a_{21}\,x + (a_{22} - \lambda)y + a_{23}\,z &= 0 \\ a_{31}\,x + a_{32}\,y + (a_{33} - \lambda)z &= 0 \end{aligned}$$

which are necessary for extrema. When $c \neq 0$, these equations and the equation $u = c$ cannot be simultaneously satisfied by (x,y,z) unless

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = 0.$$

In many important cases, values of λ satisfying this equation can be found (usually only approximately) and the problem can be finished.

19 To become acquainted with Lagrange multipliers by solving easy problems, find extrema of the first function when the second equation is required to be satisfied:

(*a*) $x^2 + y^2 + z^2$	$ax + by + cz + d = 0$
(*b*) $x^2 + y^2$	$y = x^2 - 4$
(*c*) $x + y + z$	$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$

Solution of part (*b*): Let

$$w = x^2 + y^2 + \lambda(y - x^2 + 4).$$

The extrema (if any) occur when x, y, λ are such that

$$\frac{\partial w}{\partial x} = 2x - 2x\lambda = 0$$

$$\frac{\partial w}{\partial y} = 2y + \lambda = 0$$

and

$$y = x^2 - 4.$$

These relations imply that $\lambda = 1$, $y = -\frac{1}{2}$, $x = \pm\sqrt{3.5}$, and $x^2 + y^2 = 3.75$.

20 Formulas more or less like

$$w = \int_u^v f(x,y)\,dy \tag{1}$$

often appear in pure and applied mathematics. It is supposed that u and v are functions of x and that, for each x in some interval, the integral in the right member exists and is a number w. More advanced courses set forth conditions under which dw/dx can be obtained from the chain formula

$$\frac{dw}{dx} = \frac{\partial w}{\partial v}\frac{dv}{dx} + \frac{\partial w}{\partial u}\frac{du}{dx} + \frac{\partial w}{\partial x}. \tag{2}$$

When appropriate conditions are satisfied, applications of the fundamental theorem of the calculus give

$$\frac{\partial}{\partial v}\int_u^v f(x,y)\,dy = f(x,v) \tag{3}$$

$$\frac{\partial}{\partial u}\int_u^v f(x,y)\,dy = -\frac{\partial}{\partial u}\int_v^u f(x,y)\,dy = -f(x,u).$$

When (see the last of Problems 11.19) we can differentiate under the integral sign, we get

$$\frac{\partial w}{\partial x} = \int_u^v \frac{\partial}{\partial x} f(x,y)\,dy. \tag{4}$$

Substituting in (2) then gives the formula

$$\frac{d}{dx}\int_u^v f(x,y)\,dy = f(x,v)\frac{dv}{\partial x} - f(x,u)\frac{du}{dx} + \int_u^v \frac{\partial f(x,y)}{dx}\,dy. \tag{5}$$

Verify that (5) is correct when

(*a*) $u = x$, $v = 2x$, $f(x,y) = x + y$
(*b*) $u = x$, $v = 2x$, $f(x,y) = x^2 + y^2$
(*c*) $u = x$, $v = a + x$, $f(x,y) = 1/y$
(*d*) $u = x^2$, $v = x^3$, $f(x,y) = (x + y)e^{-y}$
(*e*) $u = x$, $v = x^2$, $f(x,y) = \log y$

21 It can be observed that our proof of the chain rule for functions of more than one variable is more straightforward than our proof of the chain rule (Theorem 3.65) for functions of one variable. Can this phenomenon be explained? *Ans.:* Yes. When we proved Theorem 3.65, we did not know about the mean-value theorem and, moreover, the mean-value theorem was inapplicable because we did not have the hypothesis that the derivatives exist over intervals and are continuous. When we proved Theorem 11.23, we knew about the mean-value theorem and had enough hypotheses to enable us to apply it.

22 This problem provides preliminary information about a way in which surfaces can be determined and studied. The circle in the xy plane having its center at the origin and radius a has the simple rectangular equation $x^2 + y^2 = a^2$. We have seen, however, that it is often convenient to use the parametric equations

$$x = a \cos u, \qquad y = a \sin u \tag{1}$$

and to recognize that, when $\mathbf{r}$ is the vector running from the origin to $P(x,y)$, we have

$$\mathbf{r} = (a \cos u)\mathbf{i} + (a \sin u)\mathbf{j} \tag{2}$$

and the tip of $\mathbf{r}$ runs once in the positive direction around the circle as u increases over the interval $0 \leqq u \leqq 2\pi$. This remark, in which u has been used where we ordinarily use θ or ϕ, is designed to lead us gently to the idea that if f_1, f_2, f_3 are suitable functions of two parameters u and v, then the vector $\mathbf{r}$ defined by

$$\mathbf{r} = f_1(u,v)\mathbf{i} + f_2(u,v)\mathbf{j} + f_3(u,v)\mathbf{k} \tag{3}$$

will be the vector running from the origin to the point $P(x,y,z)$ on a surface S for which $x = f_1(u,v)$, $y = f_2(u,v)$, $z = f_3(u,v)$. For example, when we use spherical coordinates r, ϕ, θ as in Section 10.1, the equation of the sphere S having its center at the origin and radius a has the spherical equation $r = a$. The formulas of Problem 3 of Section 10.1 then show that the point P on S having spherical coordinates r, θ, ϕ has rectangular coordinates x, y, z, where

$$x = a \cos \phi \sin \theta, \qquad y = a \sin \phi \sin \theta, \qquad z = a \cos \theta. \tag{4}$$

Except that the parameters are called ϕ and θ instead of u and v, we obtain a special case of (3) by setting

$$\mathbf{r} = a(\cos \phi \sin \theta\mathbf{i} + \sin \phi \sin \theta\mathbf{j} + \cos \theta\mathbf{k}), \tag{5}$$

and (5) is a two-parameter parametric equation of the sphere S a part of which is shown in Figure 11.292. When $0 < \theta_0 < \pi$ and $\theta = \theta_0$, (5) is the parametric equation of a circle $C_P(\theta_0)$, a geographic parallel, on S which the tip of $\mathbf{r}$ traces as ϕ increases from $-\pi$ to π. When $-\pi < \phi_0 \leqq \pi$ and $\phi = \phi_0$, (5) is the parametric equation of the semicircle $C_M(\phi_0)$, a geographic meridian, which the tip of $\mathbf{r}$ traces as θ increases from 0 to π. We make only a few calculations to illustrate the utility of these things. When $0 < \theta_0 < \pi$, the foward tangent $\mathbf{t}_1$ to $C_P(\theta_0)$ at the point P_0 for which $\phi = \phi_0$ and $\theta = \theta_0$ is obtained by putting $\theta = \theta_0$ in (5), differentiating with respect to ϕ, and putting $\phi = \phi_0$ in the result. Thus,

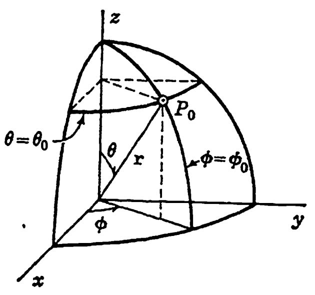


Figure 11.292

$$\mathbf{t}_1 = a(-\sin \phi_0 \sin \theta_0\mathbf{i} + \cos \phi_0 \sin \theta_0\mathbf{j}). \tag{6}$$

Similarly, when $-\pi < \phi_0 \leqq \pi$, the forward tangent $\mathbf{t}_2$ to $C_M(\phi_0)$ at the point

P_0 is obtained by putting $\phi = \phi_0$ in (5), differentiating with respect to θ, and putting $\theta = \theta_0$ in the result. Thus,

$$\text{(7)} \qquad \mathbf{t}_2 = a(\cos\phi_0 \cos\theta_0\mathbf{i} + \sin\phi_0 \cos\theta_0\mathbf{j} - \sin\theta_0\mathbf{k}).$$

While the north and south poles of a sphere are as good as any other points on the sphere, it often happens that methods involving spherical coordinates refuse to give information about them. When $0 < \theta_0 < \pi$, the two vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ are nonzero nonparallel tangents to S at the point P_0, and with the aid of the fact that S has a tangent plane and a normal at P_0 we conclude that the vector product of $\mathbf{t}_2$ and $\mathbf{t}_1$ must be a normal to S at P_0. Letting $\mathbf{N} = \mathbf{t}_2 \times \mathbf{t}_1$, we find that $\mathbf{N} = a^2 \sin\theta\mathbf{r}_0$, where $\mathbf{r}_0$ is the vector running from the origin to P_0. This shows that $\mathbf{r}_0$ is normal to S at P_0. In working out this elementary fact we have shown how, in at least one case, (3) can be used to obtain information about the surface which it represents.

23 As in (5) of the preceding problem, let $\mathbf{r}$ have its tail at the origin and let

$$\mathbf{r} = a(\cos\phi \sin\theta\mathbf{i} + \sin\phi \sin\theta\mathbf{j} + \cos\theta\mathbf{k}).$$

Let ϕ and θ be differential functions of t so the tip of $\mathbf{r}$ traverses a curve C on the sphere S as t increases. Find $\mathbf{r}'(t)$. *Ans.:*

$$\mathbf{r}'(t) = a\theta'(t)[\cos\phi\cos\theta\mathbf{i} + \sin\phi\cos\theta\mathbf{j} - \sin\theta\mathbf{k}] + a\phi'(t)\sin\theta[-\sin\phi\mathbf{i} + \cos\phi\mathbf{j}].$$

24 Let $\mathbf{r}$ have its tail at the origin and let

$$\text{(1)} \qquad \mathbf{r} = (b + a\cos\theta)\cos\phi\mathbf{i} + (b + a\cos\theta)\sin\phi\mathbf{j} + a\sin\theta\mathbf{k}.$$

Let ϕ and θ be differentiable functions of t so that, as we can see with the aid of Problem 22 at the end of Section 2.2, the tip of $\mathbf{r}$ traverses a curve C on a torus T as t increases. Show that

$$\text{(2)} \qquad \mathbf{r}'(t) = a\theta'(t)[-\sin\theta\cos\phi\mathbf{i} - \sin\theta\sin\phi\mathbf{j} + \cos\theta\mathbf{k}] + (b + a\cos\theta)\phi'(t)[-\sin\phi\mathbf{i} + \cos\phi\mathbf{j}].$$

Show that the two vectors in brackets are orthogonal. Work out the formula

$$\mathbf{n} = \cos\phi\cos\theta\mathbf{i} + \sin\phi\cos\theta\mathbf{j} + \sin\theta\mathbf{k}$$

for the unit normal to the torus.

25 This problem illustrates the fact that there are situations in which the elegant and useful chain formula

$$\text{(1)} \qquad \frac{d}{dt}u(x,y) = u_x(x,y)\frac{dx}{dt} + u_y(x,y)\frac{dy}{dt}$$

cannot be applied with impunity. Let

$$\text{(2)} \qquad u(0,0) = 0, \qquad u(x,y) = \frac{xy^2}{x^2 + y^2} \qquad (x^2 + y^2 \neq 0).$$

Let a and b be nonzero constants and let $x = at$ and $y = bt$. Show that, when $t = 0$, the left member of (1) is $ab^2/(a^2 + b^2)$ and hence is not 0 while the right member is 0. *Remark:* Because (1) is invalid, Theorem 11.24 implies that u_x and u_y cannot be continuous at the origin.

11.3 Formulas involving partial derivatives This section and its problems require us to learn more about partial derivatives and some formulas that have important applications. The first part of the section is a rather dismal discussion of unlovely terminology designed to promote understanding of curly dee abbreviations. To begin, let f be a function of two "secondary variables" x and y, and let g_1 and g_2 be functions of a single "primary variable" α. Then, in appropriate circumstances, we can set $x = g_1(\alpha)$, $y = g_2(\alpha)$, and define a function F of the single primary variable α by the formula

$$(11.31) \qquad F(\alpha) = f(x,y) = f(g_1(\alpha), g_2(\alpha)).$$

In this and all similar situations in this section, we suppose that the arguments of functions of one "variable" are confined to intervals over which the functions are differentiable and that the arguments of functions of more than one "variable" are confined to regions over which the functions have continuous partial derivatives of first order. Differentiating (11.31) with the aid of the chain rule then gives a result that can be written in ways that look very different. Using notation of one brand gives

$$F'(\alpha) = f_x(g_1(\alpha), g_2(\alpha))g_1'(\alpha) + f_y(g_1(\alpha), g_2(\alpha))g_2'(\alpha).$$

This can be put in the form

$$(11.321) \qquad F'(\alpha) = f_x(x,y)g_1'(\alpha) + f_y(x,y)g_2'(\alpha)$$

and we are responsible for remembering that the secondary variables are linked to the primary variable by the formulas $x = g_1(\alpha)$, $y = g_2(\alpha)$. Next, we can put this in the form

$$(11.322) \qquad F'(\alpha) = \frac{\partial f}{\partial x}\frac{dx}{d\alpha} + \frac{\partial f}{\partial y}\frac{dy}{d\alpha}.$$

Finally, as we usually do when the numbers in (11.31) represent temperature or something having recognizable significance, we denote the members of (11.31) by a single appropriately chosen letter, say u, and write

$$(11.33) \qquad \frac{du}{d\alpha} = \frac{\partial u}{\partial x}\frac{dx}{d\alpha} + \frac{\partial u}{\partial y}\frac{dy}{d\alpha}.$$

The abbreviated formula (11.33) is now expected to tell us that u is linked to a primary variable α by the secondary varables x and y that are identified by the fact that $\partial u/\partial x$ and $\partial u/\partial y$ appear in the formula. This formula makes sense and enables us to make calculations when, for example, we have the formulas $u = x^2 + y$, $x = \cos\alpha$, $y = \sin\alpha$.

One application of this material is worthy of mention. Suppose that a function $F(x,y)$ of two "secondary variables" and a constant c are given and that $y(x)$ is a function of the "primary variable" x such that

$$(11.34) \qquad F(x,y) = c$$

when $y = y(x)$. In this case, x is both a "primary variable" and a "secondary variable." Differentiating with respect to the "primary variable" x then gives the formula

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0 \tag{11.341}$$

from which dy/dx can be calculated when $\partial F/dy \neq 0$. This is a fact involving "implicit functions," the idea being that (11.34) is not a formula that gives an explicit formula for y in terms of x, but (11.34), and perhaps some further restrictions, may nevertheless imply that y must be the unique member or one of the members of a class of differentiable functions. The real significance of (11.341) lies in the fact that it often enables us to obtain a useful formula for dy/dx without undertaking the sometimes difficult or impossible task of "solving" (11.34) to obtain a useful formula for y. Problem 7 at the end of this section gives substantial information about this matter.

A satisfactory development of our subject must call attention to the fact that the symbol $\partial u/\partial x$ in (11.33) loses its unambiguous meaning when it is taken out of its context and we are not sure that the "independent variables" are x and y. To prove this, we construct Figure 11.35, in which P is supposed to be a point in the first quadrant. With each such point P, we may associate the circle through P with center at the origin and let A be the area of the disk bounded by this circle. If we consider x and y to be the independent variables that determine P, we obtain the first and then the second of the formulas

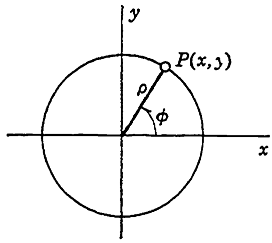


Figure 11.35

$$A = \pi(x^2 + y^2), \qquad \frac{\partial A}{\partial x} = 2\pi x. \tag{11.351}$$

If we consider x and ρ to be the independent variables, then

$$A = \pi\rho^2, \qquad \frac{\partial A}{\partial x} = 0. \tag{11.352}$$

If we consider x and ϕ to be the independent variables, then

$$A = \pi\frac{x^2}{\cos^2\phi}, \qquad \frac{\partial A}{\partial x} = \frac{2\pi x}{\cos^2\phi}. \tag{11.353}$$

Since the symbol $\partial A/\partial x$ has different values in different contexts, the symbol must be discarded or embellished when there is no clear specification of the identities of the independent variables. The first of the symbols

$$\left.\frac{\partial A}{\partial x}\right|_y, \qquad \left.\frac{\partial A}{\partial x}\right|_{x=0,\, y=0}$$

can tell us that A is a function of the independent variables x and y, the "fixed variables" being displayed at the bottom of the vertical line. The second symbol can tell us that the independent variables are x and y and that the partial derivative with respect to x is to be evaluated at the place for which $x = y = 0$. No sane person will bother with the embellishments in a very long calculation in which $\partial u/\partial x$ and $\partial u/\partial y$ must be written many times and their meanings are perfectly clear, but the embellishments are available when they are needed.

Without introducing hordes of symbols for functions, we look briefly at applications of the chain rule in situations in which u is a function of x, y, z and, in addition, x, y, z are functions of α, β, γ. The assumption following (11.31) applies here, and accordingly u is also a function of α, β, γ. While such agreements are usually made without explicit mention, we solemnly proclaim that whenever the partial derivative with respect to a Roman coordinate x or y or z (or a Greek coordinate α or β or γ) appears, the thing being differentiated must be a function of those coordinates and the other coordinates of the same nationality are the "fixed variables." If we keep β and γ so rigidly fixed that it is unnecessary to take this fact into account in our notation, then u is a function of x, y, z and x, y, z are functions of α and the version of the chain rule in Theorem 11.24 enables us to write

$$\frac{du}{d\alpha} = \frac{\partial u}{\partial x}\frac{dx}{d\alpha} + \frac{\partial u}{\partial y}\frac{dy}{d\alpha} + \frac{\partial u}{\partial z}\frac{dz}{d\alpha}.$$

When we use partial derivative notation to convey the information that β and γ are fixed, we obtain the first of the formulas

$$\begin{aligned} \frac{\partial u}{\partial \alpha} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial \alpha} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial \alpha} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial \alpha} \\ \frac{\partial u}{\partial \beta} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial \beta} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial \beta} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial \beta} \\ \frac{\partial u}{\partial \gamma} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial \gamma} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial \gamma} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial \gamma}, \end{aligned} \tag{11.36}$$

and the next two are obtained by similar processes. If, as usually happens in applications, we confine attention to regions over which the equations giving x, y, z in terms of α, β, γ can be solved to give α, β, γ in terms of x, y, z then the same procedure gives the equations

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial u}{\partial \alpha}\frac{\partial \alpha}{\partial x} + \frac{\partial u}{\partial \beta}\frac{\partial \beta}{\partial x} + \frac{\partial u}{\partial \gamma}\frac{\partial \gamma}{\partial x} \\ \frac{\partial u}{\partial y} &= \frac{\partial u}{\partial \alpha}\frac{\partial \alpha}{\partial y} + \frac{\partial u}{\partial \beta}\frac{\partial \beta}{\partial y} + \frac{\partial u}{\partial \gamma}\frac{\partial \gamma}{\partial y} \\ \frac{\partial u}{\partial z} &= \frac{\partial u}{\partial \alpha}\frac{\partial \alpha}{\partial z} + \frac{\partial u}{\partial \beta}\frac{\partial \beta}{\partial z} + \frac{\partial u}{\partial \gamma}\frac{\partial \gamma}{\partial z}. \end{aligned} \tag{11.37}$$

Teachers and textbooks providing instruction in more advanced mathematics like to presume that their disciples know enough about functions and partial derivatives to be able to write the equations (11.36) and (11.37). We are not, at the present time, required to comprehend the various reasons why these equations are important, but we can make an observation. In case the first of the conditions

$$(11.38)\qquad \begin{vmatrix} \dfrac{\partial x}{\partial \alpha} & \dfrac{\partial y}{\partial \alpha} & \dfrac{\partial z}{\partial \alpha} \\ \dfrac{\partial x}{\partial \beta} & \dfrac{\partial y}{\partial \beta} & \dfrac{\partial z}{\partial \beta} \\ \dfrac{\partial x}{\partial \gamma} & \dfrac{\partial y}{\partial \gamma} & \dfrac{\partial z}{\partial \gamma} \end{vmatrix} \neq 0, \qquad \begin{vmatrix} \dfrac{\partial \alpha}{\partial x} & \dfrac{\partial \beta}{\partial x} & \dfrac{\partial \gamma}{\partial x} \\ \dfrac{\partial \alpha}{\partial y} & \dfrac{\partial \beta}{\partial y} & \dfrac{\partial \gamma}{\partial y} \\ \dfrac{\partial \alpha}{\partial z} & \dfrac{\partial \beta}{\partial z} & \dfrac{\partial \gamma}{\partial z} \end{vmatrix} \neq 0,$$

is satisfied, the system (11.36) of equations can be solved to obtain formulas expressing $\partial u/\partial x$, $\partial u/\partial y$, $\partial u/\partial z$ in terms of $\partial u/\partial \alpha$, $\partial u/\partial \beta$, $\partial u/\partial \gamma$. Similarly, if the second condition in (11.38) is satisfied, then the system (11.37) of equations can be solved to obtain formulas expressing $\partial u/\partial \alpha$, $\partial u/\partial \beta$, $\partial u/\partial \gamma$ in terms of $\partial u/\partial x$, $\partial u/\partial y$, $\partial u/\partial z$. The determinants in (11.38) are called *functional determinants* or *Jacobi determinants* or *Jacobians*, and those who are destined to encounter them later may become thankful for this preliminary glimpse of them. Those accustomed to use of matrices prefer to see the systems (11.36) and (11.37) of equations written in the forms

$$\begin{pmatrix} \dfrac{\partial u}{\partial \alpha} \\ \dfrac{\partial u}{e\beta} \\ \dfrac{\partial u}{\partial \gamma} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial x}{\partial \alpha} & \dfrac{\partial y}{\partial \alpha} & \dfrac{\partial z}{\partial \alpha} \\ \dfrac{\partial x}{\partial \beta} & \dfrac{\partial y}{\partial \beta} & \dfrac{\partial z}{\partial \beta} \\ \dfrac{\partial x}{\partial \gamma} & \dfrac{\partial y}{\partial \gamma} & \dfrac{\partial z}{\partial \gamma} \end{pmatrix} \begin{pmatrix} \dfrac{\partial u}{\partial x} \\ \dfrac{\partial u}{\partial y} \\ \dfrac{\partial u}{\partial z} \end{pmatrix} \qquad \begin{pmatrix} \dfrac{\partial u}{\partial x} \\ \dfrac{\partial u}{\partial y} \\ \dfrac{\partial u}{\partial z} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \alpha}{\partial x} & \dfrac{\partial \beta}{\partial x} & \dfrac{\partial \gamma}{\partial x} \\ \dfrac{\partial \alpha}{\partial y} & \dfrac{\partial \beta}{\partial y} & \dfrac{\partial \gamma}{\partial y} \\ \dfrac{\partial \alpha}{\partial z} & \dfrac{\partial \beta}{\partial z} & \dfrac{\partial \gamma}{\partial z} \end{pmatrix} \begin{pmatrix} \dfrac{\partial u}{\partial \alpha} \\ \dfrac{\partial u}{\partial \beta} \\ \dfrac{\partial u}{\partial \gamma} \end{pmatrix}.$$

This permits use of the algebra of matrices to which an introduction was given at the end of Section 2.5.

In the following problems, it is assumed that each of the given functions has all of the derivatives and partial derivatives we want to use and that these derivatives are continuous.

Problems 11.39

1 Assume that x, y, z are functions of α, β, γ such that

$$f_1(x,y,z,\alpha,\beta,\gamma) = 0 \qquad f_2(x,y,z,\alpha,\beta,\gamma) = 0 \qquad f_3(x,y,z,\alpha,\beta,\gamma) = 0.$$

Write the equations obtained by taking partial derivatives with respect to α and then find the condition under which these equations uniquely determine

$\partial x/\partial\alpha$, $\partial y/\partial\alpha$, $\partial z/\partial\alpha$ in terms of partial derivatives of f_1, f_2, f_3. *Partial ans.:* The first equation can be put in the form

$$\frac{\partial f_1}{\partial x}\frac{\partial x}{\partial\alpha}+\frac{\partial f_1}{\partial y}\frac{\partial y}{\partial\alpha}+\frac{\partial f_1}{\partial z}\frac{\partial z}{\partial\alpha}=-\frac{\partial f_1}{\partial\alpha}$$

and the required condition is

$$\begin{vmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} & \dfrac{\partial f_1}{\partial z} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} & \dfrac{\partial f_2}{\partial z} \\ \dfrac{\partial f_3}{\partial x} & \dfrac{\partial f_3}{\partial y} & \dfrac{\partial f_3}{\partial z} \end{vmatrix} \neq 0.$$

2 Supposing that x and y are differentiable functions of α and β for which

$$\begin{aligned} 2x^2+3y-2\alpha^2-3\beta &= 0\\ x^2+2y^3-\alpha-2\beta^2 &= 0, \end{aligned}$$

calculate $\partial x/\partial\alpha$ and $\partial y/\partial\alpha$. *Ans.:*

$$\frac{\partial x}{\partial\alpha}=\frac{8\alpha y^2-1}{8xy^2-2x},\qquad \frac{\partial y}{\partial\alpha}=\frac{2-4\alpha}{12y^2-3}.$$

3 Supposing that $\rho > 0$ and that ρ and ϕ are functions of x and y for which

$$\rho\cos\phi = x,\qquad \rho\sin\phi = y,$$

differentiate with respect to x and then with respect to y to obtain

$$\begin{aligned} &\frac{\partial\rho}{\partial x}\cos\phi-\rho\sin\phi\frac{\partial\phi}{\partial x}=1, &&\frac{\partial\rho}{\partial y}\cos\phi-\rho\sin\phi\frac{\partial\phi}{\partial y}=0\\ &\frac{\partial\rho}{\partial x}\sin\phi+\rho\cos\phi\frac{\partial\phi}{\partial x}=0, &&\frac{\partial\rho}{\partial y}\sin\phi+\rho\cos\phi\frac{\partial\phi}{\partial y}=1 \end{aligned}$$

and solve for the derivatives to obtain

$$\frac{\partial\rho}{\partial x}=\cos\phi,\qquad \frac{\partial\rho}{\partial y}=\sin\phi,\qquad \frac{\partial\phi}{\partial x}=-\frac{\sin\phi}{\rho},\qquad \frac{\partial\phi}{\partial y}=\frac{\cos\phi}{\rho}.$$

4 Copy the formulas at the conclusion of Problem 3 and use them in appropriate places in the process of deriving the following formulas that are used to make transformations from rectangular to polar and cylindrical coordinates. Supposing that u is a function of x and y and that x and y are functions of ρ and ϕ for which

$$x=\rho\cos\phi,\qquad y=\rho\sin\phi, \tag{1}$$

show that

$$\frac{\partial u}{\partial x}=\frac{\partial u}{\partial\rho}\frac{\partial\rho}{\partial x}+\frac{\partial u}{\partial\phi}\frac{\partial\phi}{\partial x} \tag{2}$$

$$\frac{\partial u}{\partial y}=\frac{\partial u}{\partial\rho}\frac{\partial\rho}{\partial y}+\frac{\partial u}{\partial\phi}\frac{\partial\phi}{\partial y} \tag{3}$$

and hence

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial \rho}\cos\phi - \frac{\partial u}{\partial \phi}\frac{\sin\phi}{\rho} \tag{4}$$

$$\frac{\partial u}{\partial y} = \frac{\partial u}{\partial \rho}\sin\phi + \frac{\partial u}{\partial \phi}\frac{\cos\phi}{\rho}. \tag{5}$$

Then show that the formulas

$$\frac{\partial^2 u}{\partial x^2} = \left[\frac{\partial}{\partial \rho}\left(\frac{\partial u}{\partial x}\right)\right]\frac{\partial \rho}{\partial x} + \left[\frac{\partial}{\partial \phi}\left(\frac{\partial u}{\partial x}\right)\right]\frac{\partial \phi}{\partial x}$$

$$\frac{\partial^2 u}{\partial y^2} = \left[\frac{\partial}{\partial \rho}\left(\frac{\partial u}{\partial y}\right)\right]\frac{\partial \rho}{\partial y} + \left[\frac{\partial}{\partial \phi}\left(\frac{\partial u}{\partial y}\right)\right]\frac{\partial \phi}{\partial y}$$

can be put in the forms

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial \rho^2}\cos^2\phi - 2\frac{\partial^2 u}{\partial \rho\,\partial \phi}\frac{\sin\phi\cos\phi}{\rho} + \frac{\partial u}{\partial \rho}\frac{\sin^2\phi}{\rho} + \frac{\partial^2 u}{\partial \phi^2}\frac{\sin^2\phi}{\rho^2} + 2\frac{\partial u}{\partial \phi}\frac{\sin\phi\cos\phi}{\rho^2} \tag{6}$$

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 u}{\partial \rho^2}\sin^2\phi + 2\frac{\partial^2 u}{\partial \rho\,\partial \phi}\frac{\sin\phi\cos\phi}{\rho} + \frac{\partial u}{\partial \rho}\frac{\cos^2\phi}{\rho} + \frac{\partial^2 u}{\partial \phi^2}\frac{\cos^2\phi}{\rho^2} - 2\frac{\partial u}{\partial \phi}\frac{\sin\phi\cos\phi}{\rho^2} \tag{7}$$

Show finally that

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \phi^2} \tag{8}$$

and hence

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \phi^2} + \frac{\partial^2 u}{\partial z^2}. \tag{9}$$

For use in the next remark, we note also the formula

$$\frac{1}{y}\frac{\partial u}{\partial y} = \frac{1}{\rho}\frac{\partial u}{\partial \rho} + \frac{\cos\phi}{\rho^2\sin\phi}\frac{\partial u}{\partial \phi} \tag{10}$$

which comes from (5) and (1).

5 This remark can be very helpful to those who will study brands of physics and engineering in which the Laplace equation, the heat equation, and the wave equation appear. While the operation might be tedious and need not be performed, we could copy all of Problem 4 with x, y, ρ, and ϕ respectively replaced by z, ρ, r, and θ. This shows us that if u is a function of z and ρ and if

$$z = r\cos\theta, \qquad \rho = r\sin\theta, \tag{1}$$

then

$$\frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 u}{\partial \rho^2} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} \tag{2}$$

$$\frac{1}{\rho}\frac{\partial u}{\partial \rho} = \frac{1}{r}\frac{\partial u}{\partial r} + \frac{\cos\theta}{r^2\sin\theta}\frac{\partial u}{\partial \theta}. \tag{3}$$

These formulas and the simple formula

$$(4) \qquad \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \phi^2} = \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 u}{\partial \phi^2}$$

enable us to transform important expressions from cylindrical coordinates ρ, ϕ, z to spherical coordinates r, ϕ, θ. To put fundamental consequences of our results in the compact form which is very often used, we define $\nabla^2 u$ (read del squared u), the Laplacian of u, by the first of the formulas

$$(5) \qquad \nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

$$(6) \qquad \nabla^2 u = \frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \phi^2} + \frac{\partial^2 u}{\partial z^2}$$

$$(7) \qquad \nabla^2 u = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\cos\theta}{r^2\sin\theta}\frac{\partial u}{\partial \theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial \phi^2}.$$

Then (5) gives $\nabla^2 u$ in terms of rectangular coordinates x, y, z. As Problem 4 showed, (6) gives $\nabla^2 u$ in terms of the cylindrical coordinates ρ, ϕ, z of Figure 10.11. As we see by adding (2), (3), and (4), the formula (7) gives $\nabla^2 u$ in terms of the spherical coordinates r, ϕ, θ of Figure 10.12.

6 The results of Problems 3, 4, and 5 are important. Instead of proposing that similar but less important problems be solved, the author suggests that these problems be solved again and again.

7 This long problem involves a theorem which is called an implicit function theorem. The equation

$$(1) \qquad x^3 + xy + y^3 - 31 = 0,$$

which happens to be satisfied when $x = 3$ and $y = 1$, provides an introduction to the subject. If we know that y is a differentiable function of x for which (1) holds, then we can differentiate with respect to x to obtain

$$(2) \qquad 3x^2 + x\frac{dy}{dx} + y + 3y^2\frac{dy}{dx} = 0$$

or

$$(3) \qquad (3x^2 + y) + (x + 3y^2)\frac{dy}{dx} = 0$$

and hence

$$(4) \qquad \frac{dy}{dx} = -\frac{3x^2 + y}{x + 3y^2}$$

provided $x + 3y^2 \neq 0$. We can be pleased by our abilities to calculate derivatives of differentiable functions, but we can also be irked and frustrated when we realize that we do not know whether there is a differentiable function f for which

$$(5) \qquad x^3 + xf(x) + [f(x)]^3 - 31 = 0.$$

With the aid of partial derivatives, we can obtain very satisfying information about hordes of problems of which the one considered above is a special case.

(6) ***Implicit function theorem*** *Let G be a given function of two variables x and y such that $G(x_0,y_0) = 0$ and, moreover, the function G and its two partial derivatives G_x and G_y are continuous and $G_y(x,y) \neq 0$ over some rectangular region R_1 having its center at the point (x_0,y_0). Then, for some positive number h, there is one and only one function f, defined and differentiable over the interval $x_0 - h < x < x_0 + h$, for which $f(x_0) = y_0$ and*

(7) $$G(x, f(x)) = 0 \qquad (x_0 - h < x < x_0 + h).$$

Moreover, when $x_0 - h < x < x_0 + h$ and $f(x) = y$, $f'(x)$ and $\dfrac{dy}{dx}$ can be obtained by differentiating (7) *or*

(8) $$G(x,y) = 0$$

to obtain

(9) $$G_x(x, f(x)) + G_y(x, f(x))f'(x) = 0$$

or

(10) $$\frac{\partial G}{\partial x} + \frac{\partial G}{\partial y}\frac{dy}{dx} = 0$$

and solving to obtain

(11) $$f'(x) = -\frac{G_x(x, f(x))}{G_y(x, f(x))}$$

or

(12) $$\frac{dy}{dx} = -\frac{\dfrac{\partial G}{\partial x}}{\dfrac{\partial G}{\partial y}}.$$

Before proving the theorem, we observe that it provides very solid information about the graph of the equation $G(x,y) = 0$; the graph must contain a curve C which contains the point $P_0(x_0,y_0)$, and this curve C has a tangent at the point (x_0,y_0) which has slope $f'(x_0)$. As sometimes happens in other cases, our proof of the theorem reveals some facts that are not stated in the conclusion of the theorem. To prove the theorem, we suppose that $G_y(x_0,y_0) > 0$; in case $G_y(x_0,y_0) < 0$, the proof is similar. Then the hypothesis that $G_y(x,y) \neq 0$ over R_1 and the intermediate-value theorem imply that $G_y(x,y) > 0$ over R_1. Choose numbers y_1 and y_2 such that $y_1 < y_0 < y_2$ and R_1 contains the line segment consisting of points (x_0,y) for which $y_1 \leqq y \leqq y_2$. Since $G_y(x_0,y) > 0$ when $y_1 \leqq y \leqq y_2$, it follows that $G(x_0,y)$ is an increasing function of y over the interval $y_1 \leqq y \leqq y_2$. But $G(x_0,y_0) = 0$, and therefore

(13) $$G(x_0,y_1) < 0 < G(x_0,y_2).$$

Since G is continuous, we can choose a positive number h such that the rectangular region R of Figure 11.391 is a subset of the given rectangular region R_1 and, moreover,

$$G(x,y_1) < 0 \qquad (x_0 - h < x < x_0 + h) \tag{14}$$
$$G(x,y_2) > 0 \qquad (x_0 - h < x < x_0 + h). \tag{15}$$

The negative and positive signs in Figure 11.391 serve to remind us that $G(x,y) < 0$ at points (x,y) on the lower edge of R and that $G(x,y) > 0$ at points (x,y) on the upper edge of R. We are now prepared to obtain results. Let $x_0 - h < x < x_0 + h$. Since $G_y(x,y) > 0$ when $y_1 \leqq y \leqq y_2$, we conclude that $G(x,y)$ is increasing over the interval $y_1 \leqq y \leqq y_2$. Since $G(x,y_1) < 0$ and $G(x,y_2) > 0$, we conclude, with the aid of the intermediate-value theorem, that there is one and only one number $f(x)$ for which $y_1 < f(x) < y_2$ and $G(x, f(x)) = 0$.

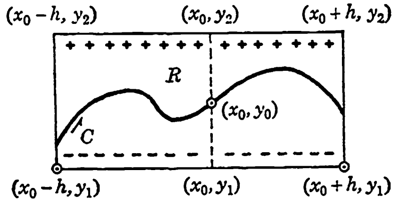


Figure 11.391

Our proof is now about half done; we have found our f but we need information about $f'(x)$. To start getting this information, suppose that $x_0 - h < x < x_0 + h$ and $x_0 - h < x + \Delta x < x_0 + h$. Then

$$G(x, f(x)) = 0, \qquad G(x + \Delta x, f(x + \Delta x)) = 0 \tag{16}$$

and consequently

$$[G(x + \Delta x, f(x + \Delta x)) - G(x, f(x + \Delta x))] + [G(x, f(x + \Delta x)) - G(x, f(x))] = 0. \tag{17}$$

Applying the mean-value theorem (Theorem 5.52) to these differences shows that there exists a number ξ between x and $x + \Delta x$ and a number η between $f(x)$ and $f(x + \Delta x)$ such that

$$G_x(\xi, f(x + \Delta x))\,\Delta x + G_y(x,\eta)[f(x + \Delta x) - f(x)] = 0 \tag{18}$$

and hence

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = -\frac{G_x(\xi, f(x + \Delta x))}{G_y(x,\eta)}. \tag{19}$$

Since G_x and G_y are, by hypothesis, continuous at the point $(x, f(x))$, the desired result (11) will follow from this if we show that f is continuous at x, that is,

$$\lim_{\Delta x \to 0} f(x + \Delta x) = f(x). \tag{20}$$

To prove (20), let ϵ be a positive number for which $f(x) + \epsilon$ and $f(x) - \epsilon$ lie between y_1 and y_2. Since $G(x, f(x)) = 0$ and $G_y(x,y)$ is an increasing function of y, we find that

$$G(x, f(x) - \epsilon) < 0 < G(x, f(x) + \epsilon).$$

Since G is continuous, we can choose a positive number δ such that

$$G(x + \Delta x, f(x) - \epsilon) < 0 < G(x + \Delta x, f(x) + \epsilon)$$

whenever $|\Delta x| < \delta$. Since $G(x + \Delta x, y)$ is an increasing function of y and $G(x + \Delta x, f(x + \Delta x)) = 0$, it follows that

$$f(x) - \epsilon < f(x + \Delta x) < f(x) + \epsilon$$

whenever $|\Delta x| < \delta$. This establishes (20) and our theorem is proved. Since f is continuous over the interval $x_0 - h < x < x_0 + h$, the graph of $y = f(x)$ over this interval is a curve C. Moreover, our proof shows that this curve C is the only part of the graph of the equation $F(x,y) = 0$ which lies inside the rectangular region R shown in Figure 11.391.

8 This is another long problem. State and prove a theorem, similar to that of the preceding problem, in which it is assumed that $G_x(x,y) \neq 0$ and the conclusion involves a function ϕ for which $G(\phi(y), y) = 0$.

9 Prove that if G is a function of x and y such that G, G_x, and G_y are everywhere continuous, and if (x_0,y_0) is a point on the graph Γ of the equation $G(x,y) = 0$ for which $G_x(x_0,y_0)$ and $G_y(x_0,y_0)$ are not both 0, then the point (x_0,y_0) is a "simple point" on the graph. *Remark:* This is a theorem in geometry. The conclusion means that if R is a rectangular (or circular) region which has its center at (x_0,y_0) and which has a sufficiently small diameter, then the points of Γ that lie in R can be ordered in such a way that they constitute a simple curve or a Jordan arc; see the last two of Problems 7.19. The theorem implies that multiple points and isolated points of Γ can occur only at places where G_x and G_y are both zero.

10 For each of the equations

(*a*) $xy = 0$ (*b*) $x^2 + y^2 = 0$
(*c*) $y^2 - x^3 = 0$ (*d*) $y^2 - x^2(x^2 - 1) = 0$
(*e*) $y^2 - x^2(x^2 + 1) = 0$

the preceding problem allows the possibility that the origin may be an isolated point or a multiple point. What are the facts?

11 Let

$$u = [(x - 1)^2 + (y - 1)^2][x^2 + y^2 + 3]$$

and observe that the graph in the xy plane of the equation $u = 0$ contains only one point P_1. Observe that this result and Problem 9 imply that the two first-order partial derivatives of u at P_1 must be zero. Calculate these derivatives and show that it is so.

12 Let f be a vector-to-scalar function for which $f(\mathbf{r})$ is a number (or scalar) whenever $\mathbf{r}$ is a vector in the domain of f. We do not need a coordinate system to define a number $D(f, \mathbf{r}, \mathbf{u})$ by the formula

$$D(f, \mathbf{r}, \mathbf{u}) = \lim_{h\to 0} \frac{f(\mathbf{r} + h\mathbf{u}) - f(\mathbf{r})}{h} \tag{1}$$

whenever $\mathbf{r}$ and $\mathbf{u}$ are vectors such that the limit exists. In case $\mathbf{u}$ is a unit vector, (1) provides an intrinsic definition of the *derivative of f at* $\mathbf{r}$ *in the direction of* $\mathbf{u}$. To start acquaintance with this matter, introduce a coordinate system and notation such that

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \qquad f(\mathbf{r}) = f(x, y, z) \tag{2}$$

and hence $f(\mathbf{r} + h\mathbf{u}) = f(x + hu_1, y + hu_2, z + hu_3)$ when $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$. Then, assuming that f has continuous partial derivatives, prove that

$$D(f, \mathbf{r}, \mathbf{u}) = \nabla f \cdot \mathbf{u}. \tag{3}$$

Observe that the scalar components of the vectors $\mathbf{r}$, $\mathbf{u}$, and ∇f depend upon the coordinating system which was chosen, but that $\mathbf{r}$, $\mathbf{u}$, $D(f, \mathbf{r}, \mathbf{u})$ and ∇f do not. *Remark:* The intrinsic definition (1) is particularly convenient when "general" or "abstract" theories are being developed. In fact some parts of modern mathematics involve "vectors" that are "abstract elements" of "abstract spaces" for which appropriate axioms are valid. Some elegant theories are developed without use of coordinate systems. At the other extreme, some elementary developments of vectors in E_3 are tied so rigidly to a single sublime coordinate system that vectors are identified with ordered sets of numbers (*the* scalar components of the vectors). One virtue of (1) lies in the fact that it can be used when f is a vector-to-vector function of which the domain and range are both sets of vectors. Nobody ever learns all about all of these things on a windy Wednesday, but people who keep studying mathematics do keep picking up ideas.

12 *Series*

12.1 Definitions and basic theorems Even though we have already had experiences with series, we start *ab initio* to develop the subject. An array of numbers and plus signs of the form

$$(12.11) \qquad u_1 + u_2 + u_3 + \cdots$$

is called a *simple infinite series* or simply a *series*. The numbers s_1, s_2, s_3, $\cdots$ defined by $s_1 = u_1$, $s_2 = u_1 + u_2$, $s_3 = u_1 + u_2 + u_3$, $\cdots$ and, in general,

$$(12.12) \qquad s_n = \sum_{k=1}^{n} u_k \qquad (n = 1, 2, 3, \cdots),$$

constitute the *sequence of partial sums* of the series. If it happens that

$$(12.121) \qquad \lim_{n\to\infty} s_n = s,$$

then the series is said to be *convergent* and is said to *converge* to s. If the limit does not exist, the series is said to be *divergent*. A series which converges to s is not the number s, just as a hand that contains 5 fingers is not the number 5. Nevertheless, we find it very convenient to abbreviate the statement that the series converges to s by writing one or the other of

$$(12.13) \qquad s = u_1 + u_2 + u_3 + \cdots, \qquad s = \sum_{k=1}^{\infty} u_k.$$

The significance of this matter is usually not fully comprehended by unfortunate people who have not digested the contents of Problems 6 and 7 of Problems 5.69. For present purposes, it is important to recognize that the equality signs in (12.13) do not have meanings like those of the equality signs in ordinary arithmetic and algebra, and that we cannot get s by "adding up" all of the terms of the series. Passage of time may possibly bring extinction to the habit of calling s the *sum* of the series. The trouble is that the habit makes the theory of series seem too easy for quick-witted superficial people, and, at the same time, seem too mysterious and difficult for everyone else. Keeping the importance of (12.121) constantly in mind enables us to make rapid progress with the elementary theory of series.

There are numerous reasons why the geometric series

$$(12.131) \qquad a + ar + ar^2 + ar^3 + \cdots$$

is important in advanced as well as elementary mathematics. We should always be well aware of the fact that if s_n is the sum of n terms of this series, then, when $r \neq 1$,

$$(12.132) \qquad s_n = a(1 + r + r^2 + \cdots + r^{n-1}) = a\frac{1 - r^n}{1 - r}.$$

If $|r| < 1$, then $\lim_{n\to\infty} r^n = 0$, so $\lim_{n\to\infty} s_n = a/(1 - r)$ and hence

$$(12.133) \qquad \frac{a}{1 - r} = a + ar + ar^2 + ar^3 + \cdots \qquad (|r| < 1).$$

If $|r| \geqq 1$, the series diverges.

The series

$$(12.134) \qquad 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \cdots,$$

in which the terms are alternately $+1$ and -1, has partial sums

$$(12.135) \qquad 1, 0, 1, 0, 1, 0, 1, 0, \cdots$$

which are alternately 1 and 0, and we start cultivating a good habit by plotting the points $s_1 = 1$, $s_2 = 0$, $s_3 = 1$, $s_4 = 0$, $\cdots$ as in Figure 12.136. There is clearly no s such that $\lim s_n = s$, and therefore the

Figure 12.136

series is divergent. A little thought about this matter can lead us to the idea that a series Σu_k cannot be convergent unless $u_n \to 0$ as $n \to \infty$. To prove that the idea is correct, suppose Σu_k converges to s. Then $s_n \to s$ and $s_{n-1} \to s$ as $n \to \infty$ and, since $u_n = s_n - s_{n-1}$, it follows that $u_n \to 0$ as $n \to \infty$. Several ideas can be obtained by investigating the series

$$\frac{1}{2} + \frac{1}{2} - \frac{1}{3} - \frac{1}{3} - \frac{1}{3} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} - \frac{1}{5} - \frac{1}{5} - \frac{1}{5} - \frac{1}{5} - \frac{1}{5} + \cdots \tag{12.137}$$

and its sequence of partial sums. The nth term of the series approaches 0 as $n \to \infty$. As we plot the points $s_1, s_2, s_3, s_4, \cdots$, we find ourselves hopping to and fro between 0 and 1. The sequence of partial sums is bounded, but the series is not convergent. The possibility of learning about convergence of series will be enhanced if we obtain a full appreciation of the way in which the following fundamental theorem is proved.

Theorem 12.14 *If*

$$s = u_1 + u_2 + u_3 + \cdots \tag{12.141}$$

and

$$t = v_1 + v_2 + v_3 + \cdots \tag{12.142}$$

and if a and b are constants, then

$$as + bt = (au_1 + bv_1) + (au_2 + bv_2) + (au_3 + bv_3) + \cdots. \tag{12.143}$$

To prove this theorem let, for each $n = 1, 2, 3, \cdots$,

$$s_n = u_1 + u_2 + \cdots + u_n, \qquad t_n = v_1 + v_2 + \cdots + v_n.$$

Rules of arithmetic (or possibly algebra) allow us to multiply by a and b, respectively, and add the results to obtain

$$as_n + bt_n = (au_1 + bv_1) + (au_2 + bv_2) + \cdots + (au_n + bv_n).$$

Thus $as_n + bt_n$ is the sum of n terms of the series in (12.143), and to prove (12.143), it is necessary to prove that

$$\lim_{n\to\infty} (as_n + bt_n) = as + bt. \tag{12.144}$$

It is now very easy to see how to finish the proof. The hypothesis (12.141) means that the first of the formulas

$$\lim_{n\to\infty} s_n = s, \qquad \lim_{n\to\infty} t_n = t \tag{12.145}$$

holds, and the hypothesis (12.142) means that the second holds. Finally (12.145) implies (12.144) and the proof is finished. The above theorem and the next two lie at the foundation of the theory of series.

Theorem 12.15 *If Σu_k is a series of nonnegative terms, so that $u_k \geqq 0$ for each k, and if the sequence $s_1, s_2, \cdots$ of partial sums is bounded, so that, for some constant M,*

$$s_n = u_1 + u_2 + \cdots + u_n \leqq M \quad (n = 1, 2, 3, \cdots), \tag{12.151}$$

then Σu_k converges.

This theorem was proved in Section 5.6, but the theorem is so important that we think about it some more. The hypothesis that $u_1 \geqq 0$, $u_2 \geqq 0$, $u_3 \geqq 0, \cdots$ implies that $s_1 = u_1 \geqq 0$, $s_2 = s_1 + u_2 \geqq s_1$, $s_3 = s_2 + u_3 \geqq s_2$, and, in fact, that

$$0 \leqq s_1 \leqq s_2 \leqq s_3 \leqq \cdots \leqq s_n \leqq M \tag{12.152}$$

for each $n = 1, 2, 3, \cdots$. Since the set E consisting of the numbers $s_1, s_2, s_3, \cdots$ is nonempty and has an upper bound, it must, according to Theorem 5.46, have a least upper bound which we can call s. Then $s_n \leqq s$ for each n. Let $\epsilon > 0$. There must be an index N for which $s_N > s - \epsilon$, since otherwise $s - \epsilon$ would be an upper bound of the set of numbers $s_1, s_2, s_3, \cdots$ and s would not be the *least* upper bound. It follows from (12.152) that

$$s - \epsilon \leqq s_n \leqq s \qquad (n > N), \tag{12.153}$$

and hence that $\lim s_n = s$. Therefore Σu_n converges to s and Theorem 12.15 is proved. In case the terms of a series Σu_k are all positive and $s_n \leqq M$ for each n, the figure obtained by plotting the partial sums $s_1, s_2, s_3, \cdots$, the upper bound M, and the number s to which Σu_k converges must look essentially like Figure 12.154. On the other hand, if

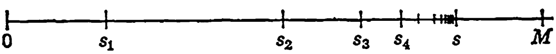


Figure 12.154

Σu_k is a series of nonnegative terms for which the sequence of partial sums is not bounded, then $s_n \to \infty$ as $n \to \infty$ and the series diverges. For series of nonnegative terms, and for such series only, it is convenient to use the first of the formulas

$$\Sigma u_k < \infty, \qquad \Sigma u_k = \infty$$

to abbreviate the statement that the series is convergent and to use the second of the formulas to abbreviate the statement that the series is divergent. In particular, a series Σu_k is said to *converge absolutely*, or to be *absolutely convergent*, if $\Sigma |u_k| < \infty$.

Before stating the next theorem, we look at the two series

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots, \qquad 1 + x + x^2 + x^3 + \cdots$$

in which it is supposed that x is a positive number and $a_0, a_1, a_2, \cdots$ are numbers, not necessarily positive, for which $|a_k| \leqq 1$ for each $k = 0, 1, 2, \cdots$. Employing the admirable terminology of the following definition, we can see that the first of the two series is dominated by the second.

Definition 12.16 *The series $u_1 + u_2 + u_3 + \cdots$ is said to be dominated by the series $M_1 + M_2 + M_3 + \cdots$ if $|u_k| \leqq M_k$ for each $k = 1, 2, \cdots$.*

Observe that the terms of a dominating series must be nonnegative; the inequality $|u_k| \leqq M_k$ can never be satisfied unless $M_k \geqq 0$. The following theorem is known as the *comparison test* for convergent series; it tells us that we can be sure that a given series is convergent if we can find a convergent series that dominates it.

***Theorem 12.17* (*comparison test*)** *If the series $u_1 + u_2 + u_3 + \cdots$ is dominated by a convergent series $M_1 + M_2 + M_3 + \cdots$, then the dominated series must be convergent and, moreover, must be absolutely convergent.*

This theorem assures us, in slightly different words, that a given series must be convergent if we can find a series of bigger fellows that is convergent. For the case in which the terms are nonnegative, it assures us that a given series must be divergent if we can find a series of smaller fellows that is divergent. Our proof of the theorem depends upon use of two series Σp_k and Σq_k, with terms defined by

$$p_k = \frac{|u_k| + u_k}{2}, \qquad q_k = \frac{|u_k| - u_k}{2} \tag{12.171}$$

that are useful for other purposes. Observe that $p_k = u_k$ when $u_k \geqq 0$, that $p_k = 0$ when $u_k \leqq 0$, and that

$$0 \leqq p_k \leqq |u_k| \leqq M_k$$

in each case. Observe also that $q_k = 0$ when $u_k \geqq 0$, that $q_k = -u_k$ when $u_k \leqq 0$, and that

$$0 \leqq q_k \leqq |u_k| \leqq M_k$$

in each case. Observe finally that adding and subtracting the formulas in (12.171) gives

$$u_k = p_k - q_k, \qquad |u_k| = p_k + q_k. \tag{12.172}$$

Letting M be the number to which the series ΣM_k of nonnegative terms converges, we see that

$$p_1 + p_2 + \cdots + p_n \leqq M_1 + M_2 + \cdots + M_n \leqq M$$
$$q_1 + q_2 + \cdots + q_n \leqq M_1 + M_2 + \cdots + M_n \leqq M.$$

It follows from Theorem 12.15 that the series Σp_k and Σq_k are both convergent. It then follows from Theorem 12.14 that the series $\Sigma(p_k - q_k)$ and $\Sigma(p_k + q_k)$ are both convergent and hence that the series Σu_k and $\Sigma|u_k|$ are both convergent. This completes the proof of Theorem 12.17.

Let Σu_k be a series that converges absolutely so that the series $\Sigma|u_k|$ is convergent. Since the series Σu_k is dominated by $\Sigma|u_k|$, an application of Theorem 12.17 (the comparison test) gives the following nontrivial theorem.

Theorem 12.18 *If a series converges absolutely, then it converges.*

From the point of view of ordinary elementary mathematics, absolutely convergent series are the ones most easily manipulated. Series that converge but do not converge absolutely are quite respectable but can be troublesome. In this course, we learn relatively little about divergent series and (except for a brief excursion in Problems 5.69) we never assign values to them. From our present point of view, the assertion "$\Sigma u_k = \infty$" does not mean that the series has a value; it means that the terms are nonnegative and that the series has partial sums $s_1, s_2, s_3, \cdots$ for which $\lim\limits_{n\to\infty} s_n = \infty$.

Problems 12.19

1 Tell the meaning of the statement

$$0 = 0 + 0 + 0 + 0 + \cdots$$

and prove the statement.

2 Using the approximations

$$\log 2 = 0.693, \qquad \pi/4 = 0.785, \qquad e = 2.71, \qquad e^{-1} = 0.37,$$

draw the interval $0 \leqq x \leqq 1$ on a rather large scale and mark the points whose coordinates are the partial sums $s_1, s_2, \cdots$ of the series in the formula

(*a*) $0 = 0 + 0 + 0 + 0 + 0 + \cdots$
(*b*) $1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \cdots$
(*c*) $\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots$
(*d*) $\pi/4 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots$
(*e*) $e - 2 = \dfrac{1}{2!} + \dfrac{1}{3!} + \dfrac{1}{4!} + \dfrac{1}{5!} + \dfrac{1}{6!} + \dfrac{1}{7!} + \cdots$
(*f*) $\dfrac{1}{e} = 1 - \dfrac{1}{1!} + \dfrac{1}{2!} - \dfrac{1}{3!} + \dfrac{1}{4!} - \dfrac{1}{5!} + \cdots$

3 Sketch figures indicating the natures of the partial sums of the series

(*a*) $\frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{6} + \cdots$
(*b*) $1 - 2 + 3 - 4 + 5 - 6 + 7 - 8 + \cdots$
(*c*) $1 + 0 + 2 + 0 + 3 + 0 + 4 + 0 + \cdots$
(*d*) $1 - 1 + 2 - 2 + 3 - 3 + 4 - 4 + \cdots$
(*e*) $(1 - 1) + (2 - 2) + (3 - 3) + (4 - 4) + \cdots$

4 Considering separately the cases in which $x = 0$, $0 < x < 1$, and $-1 < x < 0$, use one or the other or both of the formulas

$$\frac{1 - x^{n+1}}{1 - x} = 1 + x + x^2 + \cdots + x^n, \qquad \frac{1}{1 - x} = 1 + x + x^2 + x^3 + \cdots$$

to determine the nature of the sequence of partial sums of the geometric series.

5 Prove that if $|a_k| \leqq 1$ for each k, then the series

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$$

converges when $|x| < 1$. *Hint:* Use a dominating series.

6 The preceding problem is important and must be thoroughly understood. Tell what is meant by the statement that the series

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots \tag{1}$$

is dominated by the series

$$1 + |x| + |x|^2 + |x|^3 + \cdots. \tag{2}$$

Tell why the series (2) converges when $|x| < 1$. Give a full statement of the comparison test for convergence of series. *Remark:* We should hear very often that (1) is called a *power series*, that the numbers $a_0, a_1, a_2, \ldots$ are called constants, and that the number x is called a variable. We should not, however, allow the terminology to interfere with our understanding of Problem 5.

7 Prove that if A and ρ are positive constants for which

$$|a_k| \leqq A\rho^k \qquad (k = 0, 1, 2, 3, \cdots)$$

then the series

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$$

converges when $|x| < 1/\rho$. *Solution:* In the stated circumstances $|a_kx^k| \leqq A|\rho x|^k$ and the series is dominated by the convergent series in

$$\frac{A}{1 - |\rho x|} = A + A|\rho x| + A|\rho x|^2 + A|\rho x|^3 + \cdots.$$

8 Prove that if the series

$$c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots$$

converges when $x = x_0$, then it also converges when $|x| < |x_0|$. *Solution:* This problem is much like the preceding one. The hypothesis implies that $\lim_{k\to\infty} c_kx_0^k = 0$ and hence that there is a constant M for which

$$|c_kx_0^k| \leqq M \qquad (k = 0, 1, 2, \cdots).$$

If $|x| < |x_0|$, then

$$|c_kx^k| \leqq M|x/x_0|^k \qquad (k = 0, 1, 2, \cdots)$$

and the series is dominated by a convergent geometric series.

9 Show that

$$1 = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} + \frac{1}{5\cdot 6} + \cdots$$

by showing that if s_n is the sum of n (that is, the first n) terms of the series, then

$$s_n = (1 - \tfrac{1}{2}) + (\tfrac{1}{2} - \tfrac{1}{3}) + (\tfrac{1}{3} - \tfrac{1}{4}) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1 - \frac{1}{n+1}.$$

Note that the middle sum is a telescopic sum.

10 With the aid of a comparison of the two series

$$\frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} + \frac{1}{8^2} + \cdots \qquad \frac{1}{3\cdot4} + \frac{1}{4\cdot5} + \frac{1}{5\cdot6} + \frac{1}{6\cdot7} + \frac{1}{7\cdot8} + \cdots$$

prove the first of the formulas

$$\sum_{k=1}^{\infty} \frac{1}{k^2} < 1 + \tfrac{1}{4} + \tfrac{1}{9} + \tfrac{1}{3} = \tfrac{61}{36}, \qquad \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}.$$

The second formula is a simple consequence of basic theory of series known as Fourier series. The result we have obtained is significant because π^2 is about 10 and $\dfrac{\pi^2}{6}$ is about $\frac{10}{6}$ or $\frac{60}{36}$.

11 The first of the series in

$$q = \frac{1}{1\cdot3} + \frac{1}{3\cdot5} + \frac{1}{5\cdot7} + \frac{1}{7\cdot9} + \cdots \quad \text{and} \quad 1 = \frac{1}{1\cdot2} + \frac{1}{2\cdot3} + \frac{1}{3\cdot4} + \frac{1}{4\cdot5} + \cdots$$

converges because it is dominated by the second and the second is convergent. Show that $0 < q < 1$ and, if possible, find q.

12 Prove that if $a_k \geqq 0$ and $\Sigma a_k < \infty$, then $\Sigma a_k^2 < \infty$.

13 This is a preliminary skirmish with the *harmonic series*

$$(1) \qquad 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8} + \tfrac{1}{9} + \cdots.$$

This series is divergent because its terms and partial sums are greater than or equal to those of the divergent series

$$(2) \qquad 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{16} + \cdots$$

of positive terms. Let $H(1)$, $H(2)$, $H(3)$, $\cdots$ denote the partial sums of the harmonic series and, with the aid of (2), show that $H(2^0) = 1$, $H(2^1) = \frac{3}{2}$, $H(2^2) > \frac{4}{2}$, $H(2^3) > \frac{5}{2}$, and, in general,

$$H(2^n) > (n+2)/2.$$

14 Note that the series in

$$s = \frac{1}{1\cdot2\cdot3} + \frac{1}{2\cdot3\cdot4} + \frac{1}{3\cdot4\cdot5} + \frac{1}{4\cdot5\cdot6} + \frac{1}{5\cdot6\cdot7} + \cdots$$

is dominated by the series in

$$1 = \frac{1}{1\cdot2} + \frac{1}{2\cdot3} + \frac{1}{3\cdot4} + \frac{1}{4\cdot5} + \frac{1}{5\cdot6} + \cdots$$

and hence must be convergent. Try to find simple reasons why $\frac{1}{6} < s < \frac{1}{3}$. If more time is available, show that $s = \frac{1}{4}$. *Remark:* Proof that $s = \frac{1}{4}$ can be based upon the identity

$$\frac{1}{(p-1)p(p+1)} = \frac{1}{p(p^2-1)} = \frac{1}{2}\frac{(p+1)-(p-1)}{p(p^2-1)}$$
$$= \frac{1}{2}\left[\frac{1}{(p-1)p} - \frac{1}{p(p+1)}\right].$$

15 With the possibility of using consequences of the facts that

$$1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \cdots = \infty,$$
$$\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} + \cdots = 1,$$
$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x} \qquad (|x| < 1).$$

when these things are helpful, tell whether and why the following series are convergent or divergent:

(*a*) $1 + \frac{2}{2^2} + \frac{4}{4^2} + \frac{8}{8^2} + \frac{16}{16^2} + \cdots$

(*b*) $1 + \frac{2}{2^2} + \frac{3}{4^2} + \frac{4}{8^2} + \frac{5}{16^2} + \cdots$

(*c*) $\frac{0^2}{1^2\cdot 2^2} + \frac{1^2}{2^2\cdot 3^2} + \frac{2^2}{3^2\cdot 4^2} + \frac{3^2}{4^2\cdot 5^2} + \frac{4^2}{5^2\cdot 6^2} + \cdots$

(*d*) $1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{5}} + \cdots$

(*e*) $\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \cdots$

(*f*) $\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{10} + \cdots$

(*g*) $\frac{0}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{2}{3\cdot 4} + \frac{3}{4\cdot 5} + \frac{4}{5\cdot 6} + \cdots$

(*h*) $1 + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \frac{1}{5^3} + \cdots$

(*i*) $\frac{1}{1+1^2} + \frac{1}{1+2^2} + \frac{1}{1+3^2} + \frac{1}{1+4^2} + \frac{1}{1+5^2} + \cdots$

(*j*) $\frac{\log 2}{2} + \frac{\log 3}{3} + \frac{\log 4}{4} + \frac{\log 5}{5} + \frac{\log 6}{6} + \cdots$

(*k*) $\frac{\sin x}{1^2} + \frac{\sin 2x}{2^2} + \frac{\sin 3x}{3^2} + \frac{\sin 4x}{4^2} + \frac{\sin 5x}{5^2} + \cdots$

Hint: The comparison test is important. Seek a convergent series that dominates your series (so you will know your series is convergent) or seek a divergent series which your series dominates (so you will know that your series cannot be convergent and must be divergent).

16 Prove that

$$2\sum_{k=1}^{\infty} |u_k v_k| \leq \sum_{k=1}^{\infty} |u_k|^2 + \sum_{k=1}^{\infty} |v_k|^2$$

whenever the series on the right are convergent. *Solution:* The inequality

$$0 \leqq (|u_k| - |v_k|)^2 = |u_k|^2 - 2|u_k|\,|v_k| + |v_k|^2$$

implies that

$$2|u_k|\,|v_k| \leqq |u_k|^2 + |v_k|^2$$

and hence

$$2\sum_{k=1}^{n} |u_k|\,|v_k| \leqq \sum_{k=1}^{n} |u_k|^2 + \sum_{k=1}^{n} |v_k|^2 \leqq U + V,$$

where U and V are the numbers to which $\Sigma|u_k|^2$ and $\Sigma|v_k|^2$ converge. The result follows.

17 Imagine that a coin is tossed repeatedly and that we let $x_k = 1$ if the kth toss produces a head and let $x_k = 0$ if the kth toss produces a tail. Tell why the series in

$$x = \frac{x_1}{2^1} + \frac{x_2}{2^2} + \frac{x_3}{2^3} + \cdots$$

must be convergent to a number x for which $0 \leqq x \leqq 1$. Show that

$$x - \frac{1}{2} = \frac{x_1 - \frac{1}{2}}{2^1} + \frac{x_2 - \frac{1}{2}}{2^2} + \frac{x_3 - \frac{1}{2}}{2^3} + \cdots$$

and

$$x = \frac{1}{2} + \frac{2x_1 - 1}{2^2} + \frac{2x_2 - 1}{2^3} + \frac{2x_3 - 1}{2^4} + \cdots.$$

Remark: This problem can steer our thoughts toward the *Rademacher functions* $r_1(t), r_2(t), \cdots$ for which

$$t = \frac{r_1(t)}{2} + \frac{r_2(t)}{2^2} + \frac{r_3(t)}{2^3} + \cdots$$

when $0 \leqq t \leqq 1$. Figures 12.191, 12.192, and 12.193 exhibit graphs of the first

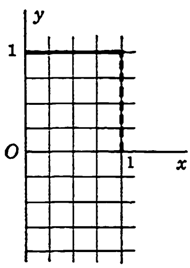


Figure 12.191

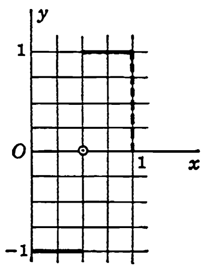


Figure 12.192

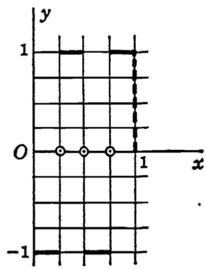


Figure 12.193

three Rademacher functions. These things are important in the theory of probability and elsewhere.

18 Look briefly at the following outline of a proof that e is irrational and then, with the textbook out of sight, write a proof in which more details are given. If we suppose that $e = m/n$, where m and n are integers, we can suppose that $n > 0$ and put $x = 1$ in the formula

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

to obtain

$$\frac{m}{n} = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \cdots$$

and then multiply by $n!$ to obtain the formula

$$M = N + \frac{1}{n+1} + \frac{1}{(n+1)(n+2)} + \frac{1}{(n+1)(n+2)(n+3)} + \cdots,$$

where M and N are integers. Thus the quantity Q defined by

$$Q = \frac{1}{n+1} + \frac{1}{(n+1)(n+2)} + \frac{1}{(n+1)(n+2)(n+3)} + \cdots$$

is the difference of two integers and must therefore be an integer. But

$$\frac{1}{n+1} < Q < \frac{1}{n+1} + \frac{1}{(n+1)^2} + \frac{1}{(n+1)^3} + \cdots = \frac{1}{n},$$

so Q cannot be an integer.

19 Give a reasonable definition setting forth conditions under which a given series

$$\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \cdots$$

of vectors in E_3 is said to be convergent. Show that the series will be convergent if the series

$$|\mathbf{v}_1| + |\mathbf{v}_2| + |\mathbf{v}_3| + \cdots$$

is convergent.

20 This problem involves rearrangements of series of nonnegative terms. Let $u_k \geqq 0$ for each k and let

$$s = u_1 + u_2 + u_3 + u_4 + \cdots.$$

Let $m_1, m_2, m_3, \cdots$ be a sequence of positive integers, not necessarily in their natural order, in which each positive integer appears exactly once. Prove that

$$s = u_{m_1} + u_{m_2} + u_{m_3} + u_{m_4} + \cdots.$$

Hint: Let

$$t_n = u_{m_1} + u_{m_2} + \cdots + u_{m_n}.$$

Show that $t_n \leqq s$ and that if $\epsilon > 0$, then $t_n > s - \epsilon$ whenever n is sufficiently great.

21 This problem has a preamble. To pour acid upon the idea that each rearrangement of the series

$$(1) \qquad 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \tfrac{1}{7} - \tfrac{1}{8} + \cdots$$

must converge to log 2, we enter the construction business. Let $Q_1, Q_2, Q_3, \cdots$ be a sequence of numbers. Let $u(1), u(2), \cdots, u(n_1)$ be, in order, the first positive term of (1) together with just as many of the following positive terms as are necessary to obtain a cumulated sum $s(n_1)$ which exceeds Q_1. Let $u(n_1 + 1), u(n_1 + 2), \cdots, u(n_2)$ be the first negative term of (1) together with just as many of the following negative terms as are necesary to obtain a cumulated sum $s(n_2)$ less than Q_2. Let $u(n_2 + 1), u(n_2 + 2), \cdots, u(n_3)$ be the first unused positive term of (1) together with just as many of the following positive terms as are necessary to obtain a cumulated sum $s(n_3)$ which exceeds Q_3. Let $u(n_3 + 1), u(n_3 + 2), \cdots, u(n_4)$ be the first unused negative term of (1) together with just as many of the following negative terms as are necessary to obtain a cumulated sum $s(n_4)$ less than Q_4, and then continue the process. Now comes the problem. Give precise information about the series $u(1) + u(2) + u(3) + \cdots$ and its sequence $s(1), s(2), s(3), \cdots$ of partial sums when, for each nonnegative integer m,

(*a*) $Q_m = 416$, (*b*) $Q_m = Q$, (*c*) $Q_m = 10^m$
(*d*) $Q_m = -10^m$ (*e*) $Q_m = (-1)^m$ (*f*) $Q_m = (-10)^m$

22 Show that the first of the formulas

$$\begin{aligned}
\log 2 &= 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \tfrac{1}{7} - \tfrac{1}{8} + \tfrac{1}{9} - \tfrac{1}{10} + \tfrac{1}{11} - \tfrac{1}{12} + \cdots \\
\tfrac{1}{2}\log 2 &= 0 + \tfrac{1}{2} + 0 - \tfrac{1}{4} + 0 + \tfrac{1}{6} + 0 - \tfrac{1}{8} + 0 + \tfrac{1}{10} + 0 - \tfrac{1}{12} + \cdots \\
\tfrac{3}{2}\log 2 &= 1 + 0 + \tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{5} + 0 + \tfrac{1}{7} - \tfrac{1}{4} + \tfrac{1}{9} + 0 + \tfrac{1}{11} - \tfrac{1}{6} + \cdots \\
\tfrac{3}{2}\log 2 &= 1 + \tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{5} + \tfrac{1}{7} - \tfrac{1}{4} + \tfrac{1}{9} + \tfrac{1}{11} - \tfrac{1}{6} + \tfrac{1}{13} + \tfrac{1}{15} - \tfrac{1}{8} + \cdots
\end{aligned}$$

implies the remaining ones. What, if any, new ideas appear in this problem?

23 Let $0 < \lambda < 1$. Suppose that a steel ball dropped from height h hits a steel plate $\sqrt{h/16}$ seconds later and immediately (without wasting time compressing and then expanding to reverse its direction) starts to rebound to height λh to begin a similar bounce. Suppose that the ball continues to bounce in this way. Find the total distance D traveled by the bouncing ball. Find also the total time T, discovering that the small bounces occur so rapidly that the ball does not bounce forever. *Ans.:*

$$D = h\frac{1 + \lambda}{1 - \lambda}, \qquad T = \frac{\sqrt{h}}{4}\,\frac{1 + \sqrt{\lambda}}{1 - \sqrt{\lambda}}.$$

24 A sequence $x_1, x_2, x_3, \cdots$ is called a *Cauchy sequence* if

$$\lim_{m,n\to\infty} (x_m - x_n) = 0, \tag{1}$$

that is, if to each positive number ϵ there corresponds a number P such that

$$|x_m - x_n| < \epsilon \qquad (m,n > P). \tag{2}$$

It is easy to prove that ***each convergent sequence is a Cauchy sequence.*** Suppose

$$\lim_{n\to\infty} x_n = L. \tag{3}$$

Let $\epsilon > 0$. Choose an index P such that

$$|x_n - L| < \epsilon/2 \qquad (n > P) \tag{4}$$

Then, when both m and n exceed P,

$$\begin{aligned}(5)\quad |x_m - x_n| = |(x_m - L) - (x_n - L)| &\leqq |x_m - L| + |x_n - L| \\ &< \epsilon/2 + \epsilon/2 = \epsilon,\end{aligned}$$

and the conclusion follows. It is also true that *each Cauchy sequence is convergent* Proofs of this fact are much more difficult because they must, in one way or another, use completeness of the real-number system to produce the number to which the sequence converges. One lively proof starts with the choice of an increasing sequence $P_1, P_2, \cdots$ of integers such that, for each $k = 1, 2, 3, \cdots$,

$$|x_m - x_n| < \frac{1}{2^k} \qquad (m,n \geqq P_k). \tag{6}$$

Then

$$|x_{P_{k+1}} - x_{P_k}| < \frac{1}{2^k} \qquad (k = 1, 2, 3, \cdots). \tag{7}$$

It follows from the comparison test for convergence of series that the series

$$x_{P_1} + (x_{P_2} - x_{P_1}) + (x_{P_3} - x_{P_2}) + (x_{P_4} - x_{P_3}) + \cdots \tag{8}$$

is convergent, say to L. Since the sum of n terms of this series is x_{P_n}, it follows that

$$\lim_{n\to\infty} x_{P_n} = L. \tag{9}$$

This shows that a subsequence of the given sequence converges to L. To prove that the whole sequence converges to L, let $\epsilon > 0$. Choose a positive number N such that

$$|x_n - x_m| < \epsilon/2 \qquad (m,n \geqq N) \tag{10}$$

and

$$|x_{P_n} - L| < \epsilon/2 \qquad (n \geqq N). \tag{11}$$

Then, when $n > N$, we can use the above inequalities and the fact that $P_n \geqq n$ to obtain

$$\begin{aligned}(12)\quad |x_n - L| &= |(x_n - x_{P_n}) + (x_{P_n} - L)| \\ &\leqq |x_n - x_{P_n}| + |x_{P_n} - L| < \epsilon/2 + \epsilon/2 = \epsilon,\end{aligned}$$

and our result is established.

12.2 Ratio test and integral test This section contains more theorems about convergence of series. These theorems, like hammers and saws and other tools in carpenter shops, have their usefulnesses and we can cultivate abilities to make effective use of appropriate ones at appropriate times. We begin by fumbling with the question whether the series

$$1^2x + 2^2x^2 + 3^2x^3 + 4^2x^4 + 5^2x^5 + \cdots \tag{12.21}$$

converges when $x = 0.99$. We set $u_n = n^2x^n$ and obtain the first and then the second of the formulas

$$u_n = n^2(1 - \tfrac{1}{100})^n, \qquad u_{100} = 10{,}000(1 - \tfrac{1}{100})^{100}.$$

Since $(1 - 1/n)^n \to 1/e$ as $n \to \infty$, it is easy to reach the correct conclusion that u_{100} is of the order of magnitude of $10{,}000/e$ and that there are several values of n for which $u_n > 1000$. This can make us suspect that the series is not convergent, but it is still possible that the series may converge to some relatively large number of the order of 10^6 or 10^{12}. Appreciation of usefulness of the ratio test can now be gained by noticing that the simple calculation

$$\begin{aligned} \lim_{n\to\infty} \left| \frac{u_{n+1}}{u_n} \right| &= \lim_{n\to\infty} \left| \frac{(n+1)^2x^{n+1}}{n^2x^n} \right| \\ &= \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^2 |x| = |x| \end{aligned} \tag{12.22}$$

shows that the series (12.21) converges when $|x| < 1$ and hence when $x = 0.99$.

Theorem 12.23 (*ratio test*) *Let $u_1 + u_2 + u_3 + \cdots$ be a series of nonzero terms and suppose that*

$$\lim_{n\to\infty} \left| \frac{u_{n+1}}{u_n} \right| = \rho. \tag{12.231}$$

In case $\rho < 1$, the series Σu_n is absolutely convergent and $\lim_{n\to\infty} u_n = 0$. *In case $\rho > 1$, the series diverges and* $\lim_{n\to\infty} |u_n| = \infty$.

To prove the first part of the theorem, choose a number r for which $\rho < r < 1$. There is then an index N such that

$$\left| \frac{u_{n+1}}{u_n} \right| < r \qquad (n \geqq N). \tag{12.232}$$

Giving n successively the values N, $N + 1$, $N + 2$, $\cdots$ yields the formulas

$$\begin{aligned} |u_{N+1}| &< |u_N|r \\ |u_{N+2}| &< |u_{N+1}|r < |u_N|r^2 \\ |u_{N+3}| &< |u_{N+2}|r < |u_N|r^3 \\ |u_{N+4}| &< |u_{N+3}|r < |u_N|r^4 \end{aligned} \quad \cdot$$

etcetera. Thus the series

$$u_N + u_{N+1} + u_{N+2} + u_{N+3} + \cdots$$

is dominated by the convergent series

$$|u_N| + |u_N|r + |u_N|r^2 + |u_N|r^3 + \cdots$$

and the conclusion follows. To prove the second part of the theorem, choose a number R for which $1 < R < \rho$. There is then an index N such that $|u_{n+1}/u_n| > R$ when $n \geqq N$ and hence

(12.233) $$|u_{N+p}| > |u_N|R^p \qquad (p = 0, 1, 2, \cdots).$$

Since $u_N \neq 0$ and $R > 1$, this shows that $|u_n| \to \infty$ as $n \to \infty$ and completes the proof of Theorem 12.23.

The remainder of the text of this section involves a connection between series and integrals which is both interesting and important. Everything that we do can be easily understood and permanently remembered

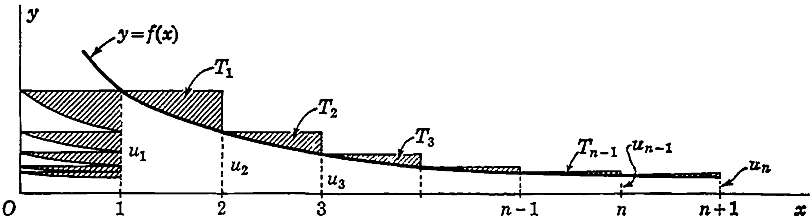


Figure 12.24

with the aid of Figure 12.24. We suppose that the terms of a series Σu_k and the values of a function f are related by the formula

$$u_k = f(k) \qquad (k = 1,2,3, \cdots),$$

that f is positive and continuous and decreasing over the interval $x \geqq 1$, and that $f(x) \to 0$ as $x \to \infty$. The left member of the formula

(12.241) $$\sum_{k=1}^{n} u_k = \int_1^n f(x)\,dx + |T_1| + |T_2| + |T_3| + \cdots + |T_{n-1}| + u_n$$

is then the sum of the areas of the rectangles of heights $u_1, u_2, \cdots, u_n$ that stand upon the unit intervals with left (or left-hand) end points at $1, 2, 3, \cdots, n$. The first term of the right member is the area of the region bounded by the graphs of the equations $x = 1$, $x = n$, $y = 0$, and $y = f(x)$. For each k, $|T_k|$ is the area of the triangular patch T_k bounded by the graphs of $x = k$, $x = k + 1$, $y = f(x)$, and $y = u_k$. Elementary bookkeeping shows that the members of (12.241) are equal. The easiest way to appraise the sum of the areas of the triangular patches is to put duplicates of these patches in the rectangle having opposite vertices at the origin and the point $(1,u_1)$. Setting

$$A_n = |T_1| + |T_2| + \cdots + |T_{n-1}|,$$

we see that $0 < A_2 < A_3 < \cdots < A_n < u_1$. There is therefore a number C such that

$$0 < \lim_{n\to\infty} [|T_1| + |T_2| + \cdots + |T_{n-1}|] = C \leqq u_1.$$

Putting $C_n = A_n + u_n$ gives the following theorem.

Theorem 12.25 (integral test) *If f is positive and continuous and decreasing over the interval $x \geqq 1$, if $f(x) \to 0$ as $x \to \infty$, and if $u_k = f(k)$ for each $k = 1, 2, 3, \cdots$, then the sequence $C_1, C_2, C_3, \cdots$ of constants defined by*

$$\sum_{k=1}^{n} u_k = \int_1^n f(x)\,dx + C_n$$

is convergent and $0 \leqq C_n \leqq u_1$ and there is a constant C for which

$$0 \leqq \lim_{n\to\infty} C_n = C \leqq u_1.$$

This theorem clearly implies the following theorem, which is known as the *integral test* for convergence of series.

Theorem 12.251 (integral test) *If f is positive and continuous and decreasing over the interval $x \geqq 1$, if $f(x) \to 0$ as $x \to \infty$, and if $u_k = f(k)$ for each $k = 1, 2, 3, \cdots$, then $\Sigma u_k < \infty$ if and only if $\int_1^\infty f(x)\,dx < \infty$.*

It can be shown that $C_1 \geqq C_2 \geqq C_3 \geqq \cdots$, and this result is sometimes useful. The most important application of Theorem 12.25 involves the case in which $f(x) = 1/x$, Σu_k is the harmonic series, and the constants C_n and C are called γ_n and γ (gamma). This application gives

$$\sum_{k=1}^{n} \frac{1}{k} = \log n + \gamma_n, \tag{12.26}$$

where the constant γ for which

$$\gamma = \lim_{n\to\infty} \gamma_n = 0.57721\ 56649\ 01532\ 86061 \tag{12.261}$$

is called the *Euler constant*. This constant γ is, after π and e, the most important mathematical constant not appearing in elementary arithmetic.

Putting $f(x) = 1/x^s$, where $s > 1$, gives

$$\sum_{k=1}^{n} \frac{1}{k^s} = \frac{1}{s-1}\left[1 - \frac{1}{n^{s-1}}\right] + C_n(s), \tag{12.271}$$

where $0 < C_n(s) < 1$. Letting $n \to \infty$ and using the definition

(12.272) $$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}$$

of the *Riemann zeta function* $\zeta(s)$ gives the nontrivial formula

(12.273) $$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s} = \frac{1}{s-1} + C(s),$$

where $0 < C(s) < 1$. The above results and the results obtained in the problems at the end of this section imply that

(12.281) $$\sum_{k=1}^{\infty} \frac{1}{k^s} < \infty \qquad (s > 1)$$

(12.282) $$\sum_{k=1}^{\infty} \frac{1}{k^s} = \infty \qquad (s \leqq 1)$$

(12.283) $$\sum_{k=2}^{\infty} \frac{1}{k(\log k)^s} < \infty \qquad (s > 1)$$

(12.284) $$\sum_{k=2}^{\infty} \frac{1}{k(\log k)^s} = \infty \qquad (s \leqq 1).$$

These series are often used with the comparison test to determine whether other given series are convergent.

Problems 12.29

1 Use the ratio test to show that the series in

(*a*) $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$

(*b*) $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$

(*c*) $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$

converge for each x, and that the geometric series in

(*d*) $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$

converges when $|x| < 1$ and diverges when $|x| > 1$.

2 Use the ratio test to show that the series

$$1!x + 2!x^2 + 3!x^3 + 4!x^4 + \cdots$$

diverges for each x for which $x \neq 0$.

3 Prove that if s is a constant, then the series

$$1^s x + 2^s x^2 + 3^s x^3 + 4^s x^4 + \cdots$$

converges when $|x| < 1$ and diverges when $|x| > 1$.

4 When $x = 1$, the ratio test does not tell whether the series of the preceding problem is convergent. Using this hint, give an example of a divergent series Σu_n for which

$$\lim_{n\to\infty} \frac{u_{n+1}}{u_n} = 1, \tag{1}$$

and then give an example of a convergent series for which (1) holds.

5 The nth term u_n of the series

$$\frac{2!}{(1!)^2} + \frac{4!}{(2!)^2} + \frac{6!}{(3!)^2} + \cdots$$

is $(2n)!/(n!)^2$. Prove that

$$\lim_{n\to\infty} \frac{u_{n+1}}{u_n} = \lim_{n\to\infty} \frac{(2n+2)(2n+1)}{(n+1)(n+1)} = 4$$

and hence that the series is divergent.

6 What information does the ratio test give about convergence of the series

$$\frac{(1!)^2}{2!}x + \frac{(2!)^2}{4!}x^2 + \frac{(3!)^2}{6!}x^3 + \frac{(4!)^2}{8!}x^4 + \cdots ?$$

Ans.: The series converges when $|x| < 4$ and diverges when $|x| > 4$. The ratio test gives no information when x is 4 or -4.

7 Supposing that $m \neq a$, show that the series in

$$\frac{1}{x-m} = -\frac{1}{m-a} - \frac{x-a}{(m-a)^2} - \frac{(x-a)^2}{(m-a)^3} - \frac{(x-a)^3}{(m-a)^4} - \cdots$$

is a geometric series and that it converges to $1/(x-m)$ when $|x-a| < |m-a|$.

8 Supposing again that $m \neq a$, show how the calculation

$$\frac{1}{x-m} = \frac{-1}{m-x} = \frac{-1}{(m-a)-(x-a)} = \frac{-1}{m-a}\,\frac{1}{1-\dfrac{x-a}{m-a}}$$

can be used to obtain the formula of the preceding problem when $|x-a| < |m-a|$. *Hint:* We must always know that the geometric series

$$1 + r + r^2 + r^3 + \cdots$$

converges to $1/(1-r)$ when $|r| < 1$, and we must sometimes be wise enough to start with $1/(1-r)$ and write the geometric series that converges to it when $|r| < 1$.

9 Write a complete proof of the fact that the formula

$$\frac{x}{x^2-9} = \frac{1}{x}\frac{1}{1-\dfrac{9}{x^2}} = \frac{1}{x} + \frac{9}{x^3} + \frac{9^2}{x^5} + \frac{9^3}{x^7} + \cdots$$

is valid when $|x| > 3$. Obtain a similar expansion of $x/(x^2+1)$.

10 For each $n = 1, 2, 3, \cdots$ let $d(n)$ be the number of positive integer divisors of n, including 1 and n, so that $d(1) = 1$, $d(2) = 2$, $d(3) = 2$, $d(4) = 3$, $d(5) = 2$, $d(6) = 4$, etcetera. Tell why the ratio test does not provide a useful source of information about convergence of the series

$$d(1)x + d(2)x^2 + d(3)x^3 + d(4)x^4 + \cdots .$$

Tell why $d(n) \geqq 1$ and the series diverges when $x \geqq 1$. Tell why $1 \leqq d(n) \leqq n$ and the series converges when $0 \leqq x < 1$. *Hint:* The series $x + 2x^2 + 3x^3 + 4x^4 + \cdots$ is convergent when $|x| < 1$.

11 Give two or more examples of convergent series $u_1 + u_2 + u_3 + \cdots$ of positive terms for which $\lim_{n\to\infty} u_{n+1}/u_n$ does not exist. *Ans.:* The series

$$\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{16} + \tfrac{1}{16} + \cdots$$
$$\frac{1}{2} + \frac{1}{10} + \frac{1}{2^2} + \frac{1}{10^2} + \frac{1}{2^3} + \frac{1}{10^3} + \frac{1}{2^4} + \frac{1}{10^4} + \cdots$$

are simple examples.

12 Show that the series in

$$\frac{1}{e^x-1} = e^{-x} + e^{-2x} + e^{-3x} + e^{-4x} + \cdots \tag{1}$$

is a geometric series that converges to the left member when $x > 0$. Use this result to show that, when $x > 0$,

$$\frac{x}{e^x-1} = \lim_{n\to\infty}\sum_{k=1}^{n} xe^{-kx}. \tag{2}$$

Show that if the manipulations

$$\int_0^\infty \frac{x}{e^x-1}\,dx = \int_0^\infty \lim_{n\to\infty}\sum_{k=1}^{n} xe^{-kx}\,dx = \lim_{n\to\infty}\int_0^\infty \sum_{k=1}^{n} xe^{-kx}\,dx \tag{3}$$

are valid, then

$$\int_0^\infty \frac{x}{e^x-1}\,dx = \sum_{k=1}^{\infty}\frac{1}{k^2}. \tag{4}$$

Remark: We shall soon start hearing that the last series converges to $\pi^2/6$.

13 Supposing that n is a positive integer, sketch a graph of

$$y = \frac{1}{n^2+x^2} \tag{1}$$

and show that

$$(2)\qquad \sum_{k=0}^{n-1} \frac{1}{n^2 + k^2} > \int_0^n \frac{1}{n^2 + x^2}\,dx = \frac{1}{n}\tan^{-1}\frac{x}{n}\Big]_0^n = \frac{\pi}{4n}.$$

Use this result to show that

$$(3)\qquad \sum_{n=1}^{\infty}\sum_{k=0}^{n-1} \frac{1}{n^2 + k^2} = \infty.$$

Remark: This result is of interest in cosmology. Suppose a particular universe contains an earth at the origin of a plane x, y coordinate system and contains a star like our sun at each point (x,y) for which x and y are integers not both zero. The rate (in appropriate units) at which the earth receives radiated energy from the star at the point (n,k) is then

$$\frac{1}{n^2 + k^2}$$

provided the inverse square law is applicable These hypotheses and (3) imply that the earth would receive energy at an infinite rate and hence would burn up instantly. This result implies that either the stars cannot be so uniformly distributed or that (perhaps because other stars and interstellar material interfere with transmission of energy) the inverse square law is inapplicable.

14 Supposing that $0 < s < 1$, show that

$$\sum_{k=1}^{n} \frac{1}{k^s} = \frac{n^{1-s} - 1}{1 - s} + C_n(s),$$

where $0 \leqq C_n(s) \leqq 1$. Show that

$$\sum_{k=1}^{n} \frac{1}{\sqrt{k}} = 2\sqrt{n} - 2 + C_n,$$

where $0 \leqq C_n \leqq 1$. Check the last result when $n = 1$ and when $n = 4$.

15 Prove that, when $n \geqq 2$,

$$\sum_{k=2}^{n} \frac{1}{k \log k} = \sum_{k=1}^{n-1} \frac{1}{(k+1)\log(k+1)}$$
$$= \log\log n - \log\log 2 + C_n,$$

where $0 \leqq C_n \leqq 1/(2 \log 2)$.

16 Prove that, when $s > 1$ and $n \geqq 2$,

$$\sum_{k=2}^{n} \frac{1}{k(\log k)^s} = \frac{1}{s-1}\left[\frac{1}{(\log 2)^{s-1}} - \frac{1}{(\log n)^{s-1}}\right] + C_n(s),$$

where $0 \leqq C_n(s) \leqq 1/2(\log 2)^s$, and that

$$\sum_{k=2}^{\infty} \frac{1}{k(\log k)^s} = \frac{1}{s-1}\,\frac{1}{(\log 2)^{s-1}} + C(s),$$

where $0 \leqq C(s) \leqq 1/2(\log 2)^s$.

17 With the aid of an appropriate figure show that, when $s \geqq 3$,

$$\frac{1}{3^s} + \frac{1}{4^s} + \cdots + \frac{1}{n^s} < \int_2^\infty \frac{1}{x^s}\,dx = \frac{1}{s-1}\frac{1}{2^{s-1}} \leqq \frac{1}{2^s}$$

and hence that

$$\frac{1}{2^s} < \zeta(s) - 1 < \frac{1}{2^{s-1}},$$

and

$$\lim_{s\to\infty} \zeta(s) = 1.$$

18 Sketch the graph of $f(x) = x^{-1/2}$ over the interval $0 < x < 1$ and observe that, even though f is unbounded and therefore not Riemann integrable, it can be suspected that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left(\frac{k}{n}\right)^{-1/2} \frac{1}{n} = \int_0^1 x^{-1/2}\,dx = 2, \tag{1}$$

where the integral is a Cauchy extension of a Riemann integral. In any case, use a result of Problem 14 to show that formula (1) is correct. *Remark:* One who undertakes to prove (1) without use of Problem 14 does so at his own peril.

19 Suppose that f is nonnegative and continuous and increasing over the interval $x \geqq 1$ and that $f(x) \to \infty$ as $x \to \infty$. Let

$$u_k = f(k) \qquad (k = 1, 2, \cdots).$$

With the aid of a figure which is in some respects like Figure 12.24, show that to each n there corresponds a number A_n such that $0 < A_n < 1$ and

$$u_1 + u_2 + \cdots + u_n = \int_1^n f(x)\,dx + u_n - A_n(u_n - u_1).$$

Applying this to the case in which $f(x) = \log x$ and

$$\sum_{k=1}^{n} u_k = \sum_{k=1}^{n} \log k = \log n!,$$

obtain the formula

$$\log n! = n \log n - n + 1 + (1 - A_n) \log n,$$

and hence

$$n! = n^n e^{-n} e n^{1-A_n} = (n/e)^n e n^{1-A_n}.$$

Remark: This elementary calculation gives an introduction to the important idea that $n!$ is of the order of magnitude of $(n/e)^n$. It is easy to see with the aid of a figure that A_n is a little greater than $\frac{1}{2}$ and hence that n^{1-A_n} is less than $\sqrt{n}$. In Problem 4 of Section 12.6, we shall discover (among other things) that if n is a positive integer, then

$$n! = \sqrt{2n\pi}\, n^n e^{-n} e^{\theta_n/12n},$$

where θ_n is a number for which

$$1 - \frac{1}{30n^2} < \theta_n < 1 - \frac{1}{30n^2} + \frac{1}{105n^4}$$

and hence θ_n is quite close to 1 even when $n = 1$. The above formula for $n!$ is a *Stirling formula*, and it is very useful.

20 We are familiar with the fact that

$$(1) \qquad \int_1^\infty \frac{1}{x^p}\,dx = \lim_{h\to\infty}\int_1^h x^{-p}\,dx = \lim_{h\to\infty}\frac{x^{1-p}}{1-p}\Big]_1^h = \frac{1}{p-1}$$

when $p > 1$ and

$$(2) \qquad \int_1^\infty \frac{1}{x}\,dx = \lim_{h\to\infty}\int_1^h \frac{1}{x}\,dx = \lim_{h\to\infty}\log x\Big]_1^h = \infty.$$

Some questions and answers involving existence (or convergence) of Riemann-Cauchy integrals are quite analogous to questions and answers involving infinite series.

(3) **Theorem** *Let f and g be Riemann integrable over each finite interval $a \leqq x \leqq h$ for which $h > a$ and let*

$$(4) \qquad 0 \leqq f(x) \leqq g(x) \qquad (x \geqq a).$$

Then

$$(5) \qquad 0 \leqq \int_a^\infty f(x)\,dx \leqq \int_a^\infty g(x)\,dx$$

when (5) *is interpreted to mean that* $\int_a^\infty f(x)\,dx \leqq M$ *whenever* $\int_a^\infty g(x)$ *exists and is M, and that* $\int_a^\infty g(x)\,dx = \infty$ *if* $\int_a^\infty f(x)\,dx = \infty$.

Proof of this theorem depends upon the fact that

$$(6) \qquad 0 \leqq \int_a^h f(x)\,dx \leqq \int_a^h g(x)\,dx \qquad (h \geqq a).$$

The functions in (6) are monotone increasing functions of h. In case $\int_a^\infty g(x)\,dx = M$, the function $\int_a^h g(x)\,dx$ has the least upper bound M. The function $\int_a^h f(x)\,dx$ then has an upper bound M and hence must have a least upper bound M_1 for which $M_1 \leqq M$ and

$$(7) \qquad \int_a^\infty f(x)\,dx = \lim_{h\to\infty}\int_a^h f(x)\,dx = \underset{h\geqq a}{\text{l.u.b.}}\int_a^h f(x)\,dx = M_1 \leqq \int_a^\infty g(x)\,dx.$$

In case $\int_a^\infty f(x)\,dx = \infty$, the function $\int_a^h f(x)\,dx$ does not have an upper bound, so $\int_a^h g(x)\,dx$ cannot have an upper bound and hence $\int_a^\infty g(x)\,dx = \infty$. This proves the theorem. Supposing that $p > 1$ and that q is real, prove that

$$\int_2^\infty \frac{(\log x)^q}{x^p}\,dx < \infty$$

by choosing a number r for which $1 < r < p$ and showing that there is a constant A for which

$$0 < \frac{(\log x)^q}{x^p} = \frac{(\log x)^q}{x^{p-r}}\frac{1}{x^r} \leqq A\frac{1}{x^r}$$

when $x \geqq 2$. Prove that

$$\int_3^\infty \frac{\log x}{x}\,dx = \infty.$$

21 Using results of the preceding problem when and if they are helpful, prove that the first of the integrals

$$\int_0^1 \frac{x^p}{(1+x)^q}\,dx, \qquad \int_1^\infty \frac{x^p}{(1+x)^q}\,dx, \qquad \int_0^\infty \frac{x^p}{(1+x)^q}\,dx$$

exists when $p > -1$ and fails to exist when $p \leqq -1$. Prove that the second integral exists when $q - p > 1$ and fails to exist when $q - p \leqq 1$. For what pairs of values of p and q does the third integral exist?

22 The first of the two integrals

$$(1) \qquad \int_a^\infty f(x)\,dx, \qquad \int_a^\infty |f(x)|\,dx$$

is sometimes said to *converge absolutely* if the second one exists. Prove the following theorem.

(2) ***Theorem*** *If f is Riemann integrable over each finite interval $a \leqq x \leqq h$ for which $h > a$ and if $\int_a^\infty |f(x)|\,dx < \infty$, then $\int_a^\infty f(x)\,dx$ exists.*

Solution: This theorem and its proof are very similar to Theorem 12.17 and its proof. Let

$$p(x) = \tfrac{1}{2}[|f(x)| + f(x)], \qquad q(x) = \tfrac{1}{2}[|f(x)| - f(x)],$$

so that $0 \leqq p(x) \leqq |f(x)|$ and $0 \leqq q(x) \leqq |f(x)|$. It then follows from the theorem of Problem 20 that the limits in

$$\lim_{h\to\infty} \int_0^h p(x)\,dx = L_1, \qquad \lim_{h\to\infty} \int_0^h q(x)\,dx = L_2$$

exist. Hence

$$\lim_{h\to\infty} \int_0^h f(x)\,dx = \lim_{h\to\infty} \left[\int_0^h p(x)\,dx - \int_0^h q(x)\,dx\right] = L_1 - L_2.$$

23 Does existence of $\int_0^\infty |f(x)|\,dx$ imply existence of $\int_0^\infty f(x)\,dx$? *Ans.:* No. For example, $f(x)$ might be e^{-x} when x is rational and $-e^{-x}$ when x is irrational. In this case $|f(x)| = e^{-x}$ and $\int_0^\infty |f(x)|\,dx = 1$ but f is everywhere discontinuous and there is no interval over which f is Riemann integrable.

24 Prove that if f and g are both Riemann integrable over each finite interval $a \leqq x \leqq h$ for which $h > a$, and if

$$|f(x)| \leqq |g(x)| \qquad (x \geqq a)$$

then existence of $\int_a^\infty |g(x)|\,dx$ implies existence of $\int_a^\infty f(x)\,dx$. *Solution:* The theorem of Problem 20 implies that $\int_a^\infty |f(x)|\,dx$ exists and the theorem of Problem 22 then gives the required result.

25 Prove that if f is integrable over each finite interval and the first of the integrals

$$\int_0^\infty |f(x)|\,dx, \qquad \int_0^\infty f(x)\sin x\,dx, \qquad \int_0^\infty f(x)\cos x\,dx$$

exists, then the other two also exist. *Solution:* Since $\sin x$ and $\cos x$ are continuous, $f(x)\sin x$ and $f(x)\cos x$ are integrable over each finite interval. Since also

$$|f(x)\sin x| \leqq |f(x)|, \qquad |f(x)\cos x| \leqq |f(x)|,$$

the results follow from the results of the preceding problem.

12.3 Alternating series and Fourier series

The following theorem embodies the alternating series test for convergence of series.

Theorem 12.31 *If the terms of a series Σu_k are alternately positive and negative and if their absolute values decrease and have the limit* 0 *so that*

$$|u_1| > |u_2| > |u_3| > \cdots, \qquad \lim_{n\to\infty} |u_n| = 0,$$

then the series converges to a number s for which

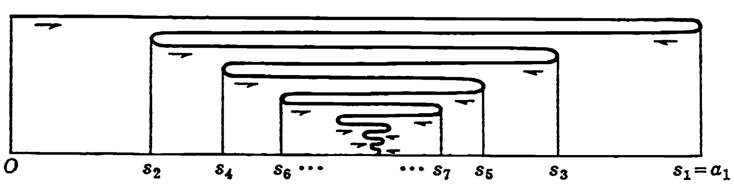
$$\left| s - \sum_{k=1}^{n} u_k \right| < |u_{n+1}| \qquad (n = 1,2,3,\cdots). \tag{12.311}$$

The inequality (12.311) tells us that if we use a particular partial sum as an approximation to s, then the error will be less than the absolute value of the first term of the series not included in the partial sum. This information is very useful. To prove the theorem, we suppose that the given series has the form

$$a_1 - a_2 + a_3 - a_4 + a_5 - a_6 + \cdots, \tag{12.312}$$

where $a_1 > a_2 > a_3 > \cdots$ and $a_k \to 0$ as $k \to \infty$. To locate the partial sums $s_1, s_2, \cdots$ shown in Figure 12.313, we start at the origin, go to the right the distance a_1 to reach s_1, then go left the smaller distance a_2 to reach s_2, then go right the still smaller distance a_3 to reach s_3, and

Figure 12.313



so on. Because the quantities in parentheses are positive, the formulas

(12.314) $$s_{2k} = (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2k-1} - a_{2k}) \qquad (k = 1,2, \cdots)$$

show that $0 < s_2 < s_4 < x_6 < \cdots$ and the formulas

(12.315) $$s_{2k-1} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - (a_{2k-2} - a_{2k-1}) \qquad (k = 1,2, \cdots)$$

show that $a_1 = s_1 > s_3 > s_5 > s_7 > \cdots$. When $n > k$, these facts and the formula

(12.316) $$s_{2n+1} = s_{2n} + a_{2n+1} > s_{2n}$$

imply that

(12.317) $$0 < s_{2k} < s_{2n} < s_{2n+1} < s_{2k-1} \leqq a_1$$

as Figure 12.313 indicates. The bounded increasing sequence s_2, s_4, s_6, $\cdots$ and the bounded decreasing sequence s_1, s_3, s_5, $\cdots$ must have limits and these limits must be equal, since $s_{2n+1} = s_{2n} + a_{2n+1}$ and $a_{2n+1} \to 0$ as $n \to \infty$. Letting s be the value of these limits, we have

(12.318) $$\lim_{n\to\infty} s_n = s.$$

In case n is even, we have $s_n < s < s_{n+1}$ and hence

$$0 < s - s_n < s_{n+1} - s_n = |a_{n+1}|.$$

This formula and a similar one holding when n is odd give the conclusion of Theorem 12.31.

The remainder of the text of this section gives a preview of fundamental ideas about series that are called *Fourier* (1768–1830) *series*. While snatches of the story can be understood by everyone, most of the results are given without proof and it is necessary to study pure and applied mathematics for a few years to obtain a full appreciation of the whole story. Let L be a given positive number, and let functions ϕ_1, ϕ_2, ϕ_3, $\cdots$ be defined by

(12.32) $$\phi_k(x) = \sqrt{\frac{2}{L}} \sin \frac{k\pi}{L} x \qquad (k = 1,2,3, \cdots).$$

Let E be the closed interval $0 \leqq x \leqq L$. Because the little trick enables us to obtain formulas that have many other applications, we write $\int_E$ instead of $\int_0^L$. It can then be shown that

(12.33) $$\int_E |\phi_k(x)|^2\,dx = 1, \qquad \int_E \phi_j(x)\phi_k(x)\,dx = 0 \quad (j \neq k).$$

On this account, we say that the functions ϕ_1, ϕ_2, ϕ_3, $\cdots$ constitute an *orthonormal set* over E. Now let f be a function which is defined over

E and is such that the two integrals

$$\int_E f(x)\,dx, \qquad \int_E |f(x)|^2\,dx \tag{12.34}$$

both exist as Riemann integrals or as Cauchy extensions of Riemann integrals. Supposing that n is a positive integer, we seek constants $c_1, c_2, \cdots, c_n$ for which the integral in the left member of the formula

$$\int_E \left|f(x) - \sum_{k=1}^{n} c_k\phi_k(x)\right|^2 dx = \int_E |f(x)|^2\,dx - \sum_{k=1}^{n} |a_k|^2 + \sum_{k=1}^{n} |c_k - a_k|^2 \tag{12.35}$$

will attain the least possible value. The first big step in the theory is made by working out the formula (12.35) in which the constants $a_1, a_2, \cdots$ are the *Fourier coefficients* of f defined by the formulas†

$$a_k = \int_E f(x)\phi_k(x)\,dx \qquad (k = 1,2, \cdots). \tag{12.36}$$

The series

$$a_1\phi_1(x) + a_2\phi_2(x) + a_3\phi_3(x), \cdots, \tag{12.361}$$

in which the coefficients are defined by (12.36), is called the *Fourier series* of f. It is very easy to see that the c's for which the right side (and hence also the left side) of (12.35) is a minimum are those for which $c_k = a_k$. Putting $c_k = a_k$ in (12.35) gives the key formula

$$\int_E \left|f(x) - \sum_{k=1}^{n} a_k\phi_k(x)\right|^2 dx = \int_E |f(x)|^2\,dx - \sum_{k=1}^{n} |a_k|^2. \tag{12.362}$$

Since the left side of (12.35) cannot be negative, we obtain the first and then the second of the inequalities

$$\sum_{k=1}^{n} |a_k|^2 \leqq \int_E |f(x)|^2\,dx, \qquad \sum_{k=1}^{\infty} |a_k|^2 \leqq \int_E |f(x)|^2\,dx. \tag{12.363}$$

The second inequality is called the *Bessel* (1784–1846) *inequality*. For most purposes, the important orthonormal sets $\phi_1, \phi_2, \cdots$ are those for which the members of (12.362) converge to 0 as $n \to \infty$ so that

$$\lim_{n\to\infty} \int_E \left|f(x) - \sum_{k=1}^{n} a_k\phi_k(x)\right|^2 dx = 0, \qquad \sum_{k=1}^{\infty} |a_k|^2 = \int_E |f(x)|^2\,dx. \tag{12.364}$$

† These formulas were known by Euler. Fourier contributed very little to the theory of Fourier coefficients and Fourier series. The things bear his name, not because he invented them, but because he advertised their formal usefulness in problems of mathematical physics.

Such sets are said to be *complete*. While proofs of such things are so long and devious that nobody should expect to be able to originate them in a few days, it can be proved that the set in (12.32) is complete. The set

$$(12.365)\qquad \frac{1}{\sqrt{L}}, \sqrt{\frac{2}{L}}\cos\frac{\pi}{L}x, \sqrt{\frac{2}{L}}\cos\frac{2\pi}{L}x, \sqrt{\frac{2}{L}}\cos\frac{3\pi}{L}x, \cdots$$

is also a complete orthonormal set over E when E is the interval $0 \leqq x \leqq L$. The set

$$(12.366)\qquad \frac{1}{\sqrt{2L}}, \frac{1}{\sqrt{L}}\cos\frac{\pi}{L}x, \frac{1}{\sqrt{L}}\sin\frac{\pi}{L}x, \frac{1}{\sqrt{L}}\cos\frac{2\pi}{L}x, \frac{1}{\sqrt{L}}\sin\frac{2\pi}{L}x, \cdots$$

is a complete orthonormal set over E when E is an interval of length $2L$. The world's mathematical storehouse contains many other useful complete orthonormal sets, and the above formulas have many important applications. Henceforth we suppose that $\phi_1, \phi_2, \phi_3, \cdots$ is the trigonometric orthonormal set appearing in (12.32) or (12.365) or (12.366) and that f has period $2L$ so that $f(x + 2L) = f(x)$ for each x. Even in this case, fundamental problems involving validity of the formula

$$(12.37)\qquad f(x) = a_1\phi_1(x) + a_2\phi_2(x) + a_3\phi_3(x) + \cdots$$

remain unsolved. It is, however, known that (12.37) is valid over $-\infty < x < \infty$ provided (i) f has period $2L$, (ii) f is bounded and piecewise monotone over $-L \leqq x \leqq L$, (iii)

$$(12.371)\qquad \lim_{h\to 0}\frac{f(x+h) + f(x-h)}{2} = f(x),$$

(iv) f is odd so that $f(-x) = -f(x)$ in case the orthonormal set is (12.32), and (v) f is even so that $f(-x) = f(x)$ in case the orthonormal set is (12.365).

The most illuminating batch of applications of the above ideas involves the Bernoulli functions $B_0(x)$, $B_1(x)$, $B_2(x)$, $\cdots$ that appeared in Section 4.3, Problem 10, and in Section 5.3, Problem 19. These are the functions of period 1 for which $B_0(x) = 1$,

$$(12.381)\qquad B_n'(x) = B_{n-1}(x) \qquad (n = 1,2,3, \cdots)$$

$$(12.382)\qquad \int_0^1 B_n(x)\,dx = 0 \qquad (n = 1,2,3, \cdots)$$

except that (12.381) fails to hold when n is 1 or 2 and x is an integer. In particular, $B_1(x)$ is the saw-tooth function for which $B_1(x) = 0$ when x is an integer and

$$(12.383)\qquad B_1(x) = x - [x] - \tfrac{1}{2}$$

when x is not an integer and $[x]$ denotes the greatest integer less than or equal to x. Problem 8 at the end of this section will show where the first of the formulas

(12.384)
$$\begin{aligned}
B_1(x) &= \frac{-2}{2\pi}\left[\frac{\sin 2\pi x}{1} + \frac{\sin 4\pi x}{2} + \frac{\sin 6\pi x}{3} + \frac{\sin 8\pi x}{4} + \cdots\right]\\
B_2(x) &= \frac{2}{(2\pi)^2}\left[\frac{\cos 2\pi x}{1^2} + \frac{\cos 4\pi x}{2^2} + \frac{\cos 6\pi x}{3^2} + \frac{\cos 8\pi x}{4^2} + \cdots\right]\\
B_3(x) &= \frac{2}{(2\pi)^3}\left[\frac{\sin 2\pi x}{1^3} + \frac{\sin 4\pi x}{2^3} + \frac{\sin 6\pi x}{3^3} + \frac{\sin 8\pi x}{4^3} + \cdots\right]\\
B_4(x) &= \frac{-2}{(2\pi)^4}\left[\frac{\cos 2\pi x}{1^4} + \frac{\cos 4\pi x}{2^4} + \frac{\cos 6\pi x}{3^4} + \frac{\cos 8\pi x}{4^4} + \cdots\right]\\
B_5(x) &= \frac{-2}{(2\pi)^5}\left[\frac{\sin 2\pi x}{1^5} + \frac{\sin 4\pi x}{2^5} + \frac{\sin 6\pi x}{3^5} + \frac{\sin 8\pi x}{4^5} + \cdots\right]
\end{aligned}$$

comes from. The remaining formulas come by successive integration; it can be proved that the series for $B_2(x)$, $B_3(x)$, $B_4(x)$, $\cdots$ can be differentiated and integrated termwise [except that the series for $B_2(x)$ is not termwise differentiable when x is an integer] and hence that the fundamental formulas (12.381) and (12.382) hold. Since Section 4.3, Problem 10, shows that

(12.385)
$$B_2(0) = \frac{B_2}{2!} = \frac{1}{12}, \qquad B_4(0) = \frac{B_4}{4!} = \frac{-1}{720},$$

putting $x = 0$ in the formulas for $B_2(x)$ and $B_4(x)$ gives

(12.386)
$$\zeta(2) = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}, \qquad \zeta(4) = \sum_{k=1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{90}.$$

Balth Van der Pol used to claim that persons who know these formulas are mathematicians and persons who do not are not.

Problems 12.39

1 Use Theorem 12.31 to show that each of the following series is convergent.

(*a*) $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots$

(*b*) $\frac{1}{\log 3} - \frac{1}{\log 4} + \frac{1}{\log 5} - \frac{1}{\log 6} + \cdots$

(*c*) $\frac{1}{\log\log 20} - \frac{1}{\log\log 21} + \frac{1}{\log\log 22} - \frac{1}{\log\log 23} + \cdots$

(*d*) $\frac{1}{1+x^2} - \frac{1}{2+x^2} + \frac{1}{3+x^2} - \frac{1}{4+x^2} + \frac{1}{5+x^2} - \cdots$

(*e*) $\frac{\log 3}{3} - \frac{\log 4}{4} + \frac{\log 5}{5} - \frac{\log 6}{6} + \frac{\log 7}{7} - \cdots$

2 Find, correct to four decimal places, the numbers to which the following series are convergent.

(*a*) $\frac{1}{10} - \frac{2}{10^2} + \frac{3}{10^3} - \frac{4}{10^4} + \frac{5}{10^5} - \cdots$

(*b*) $\frac{1^2}{10} - \frac{2^2}{10^2} + \frac{3^2}{10^3} - \frac{4^2}{10^4} + \frac{5^2}{10^5} - \cdots$

3 Show that the series

$$\frac{x}{3} + \frac{x^2}{5} + \frac{x^3}{7} + \frac{x^4}{9} + \cdots,$$

for which the nth term is $x^n/(2n + 1)$, converges when $-1 \leqq x < 1$ and diverges when $x < -1$ and when $x \geqq 1$. *Hint:* Some but not all of the information is revealed by the ratio test.

4 With the aid of basic information about alternating series show that the series in

$$S = \frac{1}{1^2} - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \frac{1}{5^2} - \cdots \tag{1}$$

converges to a number S for which $0 < S < 1$. Then show that, correct to 5D (5 decimal places),

$$S > 0.75000 \qquad S < 0.86111 \qquad S > 0.79861 \qquad S < 0.83861$$

Remark: One who wishes to invest a moment to pick up some ideas may start with the esoteric but important formula

$$\frac{\pi^2}{6} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \cdots \tag{2}$$

and obtain the formula

$$\frac{\pi^2}{24} = 0 + \frac{1}{2^2} + 0 + \frac{1}{4^2} + 0 + \frac{1}{6^2} + \cdots. \tag{3}$$

Subtracting twice (3) from (2) then gives (1) with $S = \pi^2/12$. Subtracting (3) from (2) gives the formula

$$\frac{\pi^2}{8} = \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \frac{1}{9^2} + \cdots \tag{4}$$

which sparkles almost as brightly as (2).

5 Supposing that $0 < x < 1$, use the formula

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

to show that

$$x - \log(1 + x) = x^2\left(\frac{1}{2} - \frac{x}{3} + \frac{x^2}{4} - \frac{x^3}{5} + \cdots\right)$$

and

$$0 < x - \log(1 + x) < \tfrac{1}{2}x^2.$$

6 Supposing that $x = \pi/4$, give one or more reasons why the series

$$\frac{\sin x}{1} - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \frac{\sin 4x}{4} + \cdots$$

is not an alternating series to which Theorem 12.31 applies.

7 The function ϕ_k being defined by (12.32), prove the formulas (12.33).

8 It is not expected that the theory of Fourier series and its formulas have been learned, but it is expected that we can start solving problems when suitable formulas and instructions are given. Write the formulas (12.32) for the case in which $L = 1$ and show that the formula (12.36) for the Fourier coefficients becomes

$$a_k = \sqrt{2} \int_0^1 f(x) \sin k\pi x \, dx.$$

Letting $f(x)$ be the Bernoulli function $B_1(x)$ so that $f(x) = x - \frac{1}{2}$ when $0 < x < 1$, show that

$$a_k = \sqrt{2} \int_0^1 (x - \tfrac{1}{2}) \sin k\pi x \, dx.$$

Show that integration by parts gives

$$\begin{aligned} a_k &= \sqrt{2} \left[\left(x - \frac{1}{2} \right) \frac{-1}{k\pi} \cos k\pi x \right]_0^1 + \sqrt{\frac{2}{k\pi}} \int_0^1 \cos k\pi x \, dx \\ &= \frac{-\sqrt{2}}{k\pi} \frac{1 + \cos k\pi}{2} = \frac{-\sqrt{2}}{k\pi} \frac{1 + (-1)^k}{2}, \end{aligned}$$

so that $a_k = 0$ when K is odd and $a_k = -\sqrt{2}/k\pi$ when k is even. Observe that the conditions in the sentence following (12.37) are satisfied and hence that (12.37) must be valid. Then show that substituting in (12.37) gives the first of the formulas (12.384).

9 Another particularly important example involves the *square sine* function (or *square wave* function) defined by

$$\mathbf{Sin}\, x = \operatorname{sgn} \sin x.$$

The graph of this function is shown in Figure 12.391. To find the trigonometric

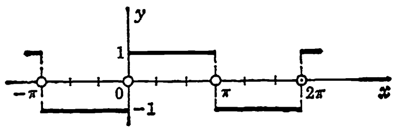


Figure 12.391

Fourier series of the odd function $\mathbf{Sin}\,(\pi x/L)$ which has period $2L$, use the orthonormal set (12.32) and, after observing that $\mathbf{Sin}\,(\pi x/L) = 1$ when $0 < x < L$, calculate the Fourier coefficients of $\mathbf{Sin}\,(\pi x/L)$ from the formula

$$a_k = \int_0^L \sqrt{\frac{2}{L}} \sin \frac{k\pi x}{L} \, dx.$$

Then tell why (12.37) must be valid and substitute in it to obtain

$$\operatorname{Sin}\frac{\pi x}{L} = \frac{4}{\pi}\left[\sin\frac{\pi x}{L} + \frac{1}{3}\sin\frac{3\pi x}{L} + \frac{1}{5}\sin\frac{5\pi x}{L} + \frac{1}{7}\sin\frac{7\pi x}{L} + \cdots\right].$$

Observe that this implies that

$$\frac{\pi}{4} = \sin\frac{\pi x}{L} + \frac{1}{3}\sin\frac{3\pi x}{L} + \frac{1}{5}\sin\frac{5\pi x}{L} + \frac{1}{7}\sin\frac{7\pi x}{L} + \cdots$$

when $0 < x < L$ and that putting $x = L/2$ gives the formula

$$\frac{\pi}{4} = 1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} + \tfrac{1}{9} - \cdots$$

which we have seen before.

10 Engineers who are interested in fully rectified (or full-wave rectified) alternating currents would want x replaced by ωt in the formula

$$|\sin x| = \frac{4}{\pi}\left[\frac{1}{2} - \frac{\cos 2x}{1\cdot 3} - \frac{\cos 4x}{3\cdot 5} - \frac{\cos 6x}{5\cdot 7} - \frac{\cos 8x}{7\cdot 9} - \cdots\right],$$

but this shift is easily made. Work out the formula and observe that it is correct when $x = 0$ because

$$\frac{1}{2} = \frac{1}{1\cdot 3} + \frac{1}{3\cdot 5} + \frac{1}{5\cdot 7} + \frac{1}{7\cdot 9} + \cdots.$$

11 Sketch a graph of the even function f of period 2π for which $f(x) = x$ when $0 \leqq x \leqq \pi$. Show that

$$f(x) = \frac{\pi}{2} - \frac{4}{\pi}\left(\cos x + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \frac{\cos 7x}{7^2} + \cdots\right).$$

12 Use the formulas following (12.384) to show that, when n is a positive integer,

$$B_{2n}(0) = (-1)^{n+1}\frac{2}{(2\pi)^{2n}}\zeta(2n)$$

and hence that $B_{2n}(0)$ is a small number like $1/6^{2n}$ when n is large. Use the formula $B_k = k!\,B_k(0)$ and the Stirling formula of Section 12.2, Problem 19, to show that

$$B_{2n} = (-1)^{n+1}\,4\sqrt{n\pi}\left(\frac{n}{\pi e}\right)^{2n}\zeta(2n)e^{\theta_{2n}/24n}.$$

Remark: Even crude estimates show that $|B_{2n}|$ is very large when n is large. Since $\pi e < 9$, we have $n/\pi e > 10$ and $|B_{2n}| > 10^{2n}$ when $n \geqq 90$. Similarly, $|B_{2n}| > 1000^{2n}$ when $n \geqq 900$.

13 Our work with Fourier series has involved *Fourier analysis*. We started with given functions and found their Fourier series. While proofs of results lie beyond the scope of this course, we take brief cognizance of a problem in *Fourier* *synthesis.* Let $\phi_1, \phi_2, \phi_3, \cdots$ be a set of functions orthonormal over a set E. Let $a_1, a_2, a_3, \cdots$ be given coefficients for which $\Sigma|a_k|^2 < \infty$. Then, when Lebesgue integrals are used, there is a function f for which the formulas

$$(1) \qquad \int_E |f(x)|^2\,dx < \infty, \qquad \lim_{n\to\infty} \int_E \left| f(x) - \sum_{k=1}^{n} a_k\phi_k(x) \right|^2 dx = 0$$

are valid. Moreover the numbers $a_1, a_2, a_3, \cdots$ are the Fourier coefficients of f, that is,

$$(2) \qquad a_k = \int_E f(x)\overline{\phi_k(x)}\,dx \qquad (k = 1,2,3, \cdots).$$

In this formula we have recognized the fact that, in many important applications, $\phi_k(x)$ is a complex number and "$\phi_k(x)$ bar" is the complex conjugate of $\phi_k(x)$. It has not been asserted (and is in fact sometimes untrue) that the series in

$$(3) \qquad f(x) \sim a_1\phi_1(x) + a_2\phi_2(x) + a_3\phi_3(x) + \cdots$$

converges to $f(x)$. However, as the second formula in (1) shows, the sum of the first n terms of the series must be a good global approximation to $f(x)$ whenever n is large. Persons who study Lebesgue integration and Fourier series for a year or two can learn all about these things.

14 Supposing that $A_0, A_1, A_2, \cdots$ and $B_1, B_2, \cdots$ are bounded sequences of constants, let

$$(1) \qquad f(x) = A_0 + \sum_{k=1}^{\infty} (A_k \cos kx + B_k \sin kx)$$

for those values of x (if any) for which the series is convergent. The series in the right member of (1) is called a *trigonometric series.* Some profound studies of the series in (1) depend upon use of the function F defined by

$$(2) \qquad F(x) = \tfrac{1}{2}A_0x^2 - \sum_{k=1}^{\infty} \frac{A_k \cos kx + B_k \sin kx}{k^2},$$

the series being convergent for each x because it is dominated by a convergent series of constants. Show that, when $t \neq 0$,

$$(3) \qquad \frac{F(x-2t) - 2F(x) + F(x+2t)}{4t^2} = A_0 + \sum_{k=1}^{\infty} [A_k \cos kx + B_k \sin kx] \left(\frac{\sin kt}{kt}\right)^2.$$

Remark: Riemann discovered this formula and used it to solve some difficult problems. In more advanced mathematics, it is proved that the formula

$$(4) \qquad \lim_{t\to 0} \left[u_0 + \sum_{k=1}^{\infty} u_k \left(\frac{\sin kt}{kt}\right)^2 \right] = u_0 + \sum_{k=1}^{\infty} u_k$$

is valid whenever the series on the right is convergent. From this it follows that

$$\lim_{t\to 0} \frac{F(x-t) - 2F(x) + F(x+t)}{t^2} = f(x) \tag{5}$$

for each x for which the series in (1) is convergent. We conclude with a brief outline of the sophisticated steps by which this sophisticated result is used to prove the following difficult theorem. *If the series in* (1) *converges to* 0 *for each* x, *then* $A_k = B_k = 0$ *for each* k. The present hypotheses imply that F is continuous and that the left member of (5), the generalized second derivative of F at x, is 0 for each x. These facts can be used to prove that F must be a linear function, and further arguments involving (2) can be used to prove that $F(x) = 0$ for each x. Since $F(0) = F(2\pi) = 0$, use of (2) shows that $A_0 = 0$. Further arguments involving (2) show that $A_k = B_k = 0$ for each k. The theorem is called a uniqueness theorem because it implies that if f is a given function, then there can be at most one collection of constants $A_0, A_1, A_2, \cdots$ and $B_1, B_2, \cdots$ for which the series in (1) converges to $f(x)$ for each x. This means that if the formulas (1) and

$$f(x) = C_0 + \sum_{k=1}^{\infty} (C_k \cos kx + D_k \sin kx) \tag{6}$$

are both valid for each x, then $C_k = A_k$ and $D_k = B_k$ for each k. Our brief glimpse of the Riemann theory of trigonometric series can make us aware of the fact that the uniqueness theorem for trigonometric series has been proved, and the proof involves mathematical ideas that we have not yet assimilated. This matter is important, because trigonometric series appear even in quite elementary applied mathematics and we need some authoritative information to help us appraise the revelations appearing in textbooks that give superficial treatments of the subject.

12.4 Power series

The series

$$c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots, \tag{12.41}$$

in which a and $c_0, c_1, c_2, \cdots$ are constants, is called a *power series* in $(x - a)$. The fundamental reason for importance of these things lies in the fact that powers of $(x - a)$ are relatively easy to calculate, to differentiate, and to integrate. Some power series, like the series

$$0! + 1!(x-a) + 2!(x-a)^2 + 3!(x-a)^3 + \cdots \tag{12.411}$$

converge only when $x = a$. Others, like those in the important formulas

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \tag{12.412}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \tag{12.413}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots, \tag{12.414}$$

converge for each x. The geometric series in the formula

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \qquad (|x| < 1) \tag{12.415}$$

is an example of a power series in $(x - 0)$ which converges for at least one x different from 0 and diverges for at least one x. With each power series in $(x - a)$ which converges for at least one x different from a and diverges for at least one x, there is associated a positive number R, called the *radius of convergence* of the power series, such that the series converges absolutely for each x for which $|x - a| < R$ and diverges for each x for which $|x - a| > R$. As Figure 12.42 indicates, the interval $|x - a| < R$ is called the interval of convergence of the series.

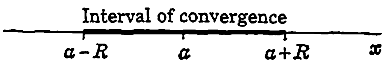


Figure 12.42

We now give, without proof, some very useful information about power series. Those who are interested in proofs should study the theory of functions of a complex variable. If the power series in (12.43) converges when $|x - a| < r$ and if for each such x we let $f(x)$ be the number to which the series converges, then

$$f(x) = c_0 + c_1(x - a) + c_2(x - a)^2 + c_3(x - a)^3 + \cdots . \tag{12.43}$$

The function f thus defined is continuous over the interval $|x - a| < r$ and, moreover, has derivatives of all positive orders which "can be obtained by termwise differentiation," that is, when $|x - a| < r$,

$$f'(x) = c_1 + 2c_2(x - a) + 3c_3(x - a)^2 + 4c_4(x - a)^3 + \cdots \tag{12.441}$$

$$f''(x) = 1\cdot 2c_2 + 2\cdot 3c_3(x - a) + 3\cdot 4c_4(x - a)^2 + 4\cdot 5c_5(x - a)^3 + \cdots \tag{12.442}$$

$$f^{(3)}(x) = 1\cdot 2\cdot 3c_3 + 2\cdot 3\cdot 4c_4(x - a) + 3\cdot 4\cdot 5c_5(x - a)^2 + 4\cdot 5\cdot 6c_6(x - a)^3 + \cdots \tag{12.443}$$

$$f^{(4)}(x) = 1\cdot 2\cdot 3\cdot 4c_4 + 2\cdot 3\cdot 4\cdot 5c_5(x - a) + 3\cdot 4\cdot 5\cdot 6c_6(x - a)^2 + \cdots \tag{12.444}$$

and so on, so that for each $n = 1, 2, 3, \cdots$

*(12.445)

$$f^{(n)}(x) = n!c_n + \frac{(n+1)!}{1!}c_{n+1}(x - a) + \frac{(n+2)!}{2!}c_{n+2}(x - a)^2 + \cdots . \tag{12.445}$$

Termwise integration as well as termwise differentiation is permissible, that is,

$$(12.45)\quad \int_a^x f(t)\,dt = \int_a^x c_0\,dt + \int_a^x c_1(t-a)\,dt + \int_a^x c_2(t-a)^2\,dt + \int_a^x c_3(t-a)^3\,dt + \cdots$$
$$= c_0(x-a) + c_1\frac{(x-a)^2}{2} + c_2\frac{(x-a)^3}{3} + c_3\frac{(x-a)^4}{4} + \cdots$$

when $|x-a| < r$. Moreover, we can multiply (12.43) by $g(x)$ and, provided g is Riemann integrable over the interval from a to x, integrate termwise to obtain

$$(12.46)\quad \int_a^x f(t)g(t)\,dt = \int_a^x c_0 g(t)\,dt + \int_a^x c_1(t-a)g(t)\,dt + \int_a^x c_2(t-a)^2 g(t)\,dt + \cdots .$$

Putting $x = a$ in (12.43) and the formulas that follow it gives the remarkable formulas

$$(12.47)\quad f(a) = c_0,\quad f'(a) = c_1,\quad f''(a) = 2!c_2,\quad f^{(3)}(a) = 3!c_3,\quad f^{(4)}(a) = 4!c_4,\quad \cdots .$$

Solving these equations for $c_0, c_1, c_2, \cdots$ and putting the results in (12.43) gives the more remarkable formula

$$(12.48)\quad f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots .$$

These formulas show one of the ways in which the coefficients $c_0, c_1, \cdots$ in a convergent power series can be determined in terms of the function to which the series converges. The series in (12.48) is the *Taylor series*, or *Taylor expansion*, of f in powers of $(x-a)$, and our work shows that *each convergent power series is the Taylor series of the function to which it converges.* In case $a = 0$, the Taylor series is sometimes called a *Maclaurin series*, but the practice has little justification and is being slowly abandoned.

The following *uniqueness theorem* is used very often.

Theorem 12.481 *If $r > 0$ and if the two power series $\Sigma b_k(x-a)^k$ and $\Sigma c_k(x-a)^k$ both converge to the same $f(x)$ when $|x-a| < r$, so that*

$$(12.482)\quad f(x) = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots \qquad (|x-a| < r)$$
$$(12.483)\quad f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots \qquad (|x-a| < r),$$

then $b_k = c_k$ for each k.

For example, the series in (12.412), (12.413), and (12.414) are the only power series in x that converge to e^x, $\cos x$, and $\sin x$. To prove this theorem, we can start with (12.482) and show that $b_k = f^{(k)}(a)/k!$ just as we started with (12.43) and showed that $c_k = f^{(k)}(a)/k!$.

Problems 12.49

1 Learn the formulas

$$(a)\ e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

$$(b)\ \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

$$(c)\ \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

$$(d)\ \frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

Write the four formulas obtained by differentiating formulas (*a*) to (*d*).

2 Explain the steps by which the series

$$\frac{1}{1-t} = 1 + t + t^2 + t^3 + \cdots \qquad (|t| < 1)$$

and modifications of it can be used to obtain the formulas

$$(a)\ \frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots \qquad (|x| < 1)$$

$$(b)\ \log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \qquad (|x| < 1)$$

$$(c)\ \frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots \qquad (|x| < 1)$$

$$(d)\ \tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots \qquad (|x| < 1)$$

$$(e)\ \frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots \qquad (|x| < 1)$$

$$(f)\ \lim_{x\to 0} \frac{\log(1+x)}{x} = 1, \qquad \lim_{x\to 0} \frac{x - \log(1-x)}{x^2} = \frac{1}{2}$$

$$(g)\ \int_0^x \frac{\log(1+t)}{t}\,dt = x - \frac{x^2}{2^2} + \frac{x^3}{3^2} - \frac{x^4}{4^2} + \cdots \qquad (|x| < 1)$$

3 We can object to the general principle that problems should be solved in inefficient ways, but nevertheless we can sometimes profitably sacrifice a few square feet of paper to promote understanding of a subject. Assuming that the series in

$$\sin 2x = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + \cdots$$

converges to $\sin 2x$, find c_0, c_1, c_2, $\cdots$ in the following way. Put $x = 0$ to find c_0. Differentiate once and put $x = 0$ to find c_1. Differentiate once more and put $x = 0$ to find c_2. Continue the process until c_5 has been obtained. Finally, see whether the result agrees (as it should) with the result of replacing x by $2x$ in the basic formula

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots .$$

4 Repeat the operation of Problem 3 to find the expansion of e^{bx} in powers of x, it being assumed that $b > 0$ and that there is a power series in x that converges to e^{bx}. Tell how your answer can be checked.

5 Assuming that there exist constants $c_0, c_1, c_2, \cdots$ for which

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots ,$$

do enough differentiating and substituting to find the first few c's when

(*a*) $f(x) = (1-x)^{-1}$, $a = 0$
(*b*) $f(x) = x^{-1}$, $a = 1$
(*c*) $f(x) = \log (1+x)$, $a = 0$
(*d*) $f(x) = \log x$, $a = 1$

6 Without bothering to write derivatives of the right member of the formula

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots ,$$

suppose that the series converges to $f(x)$ and find the first few of the c's with the aid of the formula

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$

when

(*a*) $f(x) = \sin x$, $a = 0$
(*b*) $f(x) = (1-x)^{-1}$, $a = 0$
(*c*) $f(x) = (1+x)^3$, $a = 0$
(*d*) $f(x) = x^3 - 2x^2 + x - 1$, $a = 1$

7 It is possible to apply the methods of the preceding problems to calculate a few coefficients in cases where the complexities of formulas for $f^{(n)}(x)$ increase very rapidly as n increases. In some such cases, it is worthwhile to know the numerical values of the first one or two or three nonzero coefficients. Verify the first two nonzero coefficients in each of the formulas

(*a*) $\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \frac{62}{2835}x^9 + \cdots \qquad \left(|x| < \frac{\pi}{2}\right)$

(*b*) $\sec x = 1 + \frac{1}{2}x^2 + \frac{5}{24}x^4 + \frac{61}{720}x^6 + \frac{277}{8064}x^8 + \cdots \qquad \left(|x| < \frac{\pi}{2}\right)$

(*c*) $(1 + x + x^2)^{1/2} = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \cdots$

8 It can be proved that if the series in the first two formulas

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots \tag{1}$$
$$g(x) = b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + \cdots \tag{2}$$
$$f(x)g(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + \cdots \tag{3}$$

converge to $f(x)$ and $g(x)$ when $|x| < r$, then the series in (3) will converge to $f(x)g(x)$ when $|x| < r$ provided the constants $c_0, c_1, c_2, \cdots$ are determined by the formulas

$$\begin{aligned} c_0 &= a_0b_0 \\ c_1 &= a_0b_1 + a_1b_0 \\ c_2 &= a_0b_2 + a_1b_1 + a_2b_0 \\ c_3 &= a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0 \end{aligned}$$

etcetera. Observe that this is precisely the way we would write the product of the right members of (1) and (2) if they were polynomials. To obtain a bit of experience with these formulas, write the formulas to which (1), (2), and (3) reduce when $a_k = b_k = 1$ for each k. Check your work by obtaining the third formula from the first in another way.

9 Prove that if the series $u_0 + u_1 + u_2 + \cdots$ has partial sums $s_0, s_1, s_2, \cdots$ and if the series in

$$f(x) = u_0 + u_1x + u_2x^2 + u_3x^3 + \cdots$$

converges to $f(x)$ when $|x| < 1$, then

$$f(x) = (1 - x)(s_0 + s_1x + s_2x^2 + s_3x^3 + \cdots)$$

when $|x| < 1$. *Hint:* Use the information given at the start of the preceding problem, putting $b_k = 1$ for each k.

10 Write two more terms of each of the series

$$\begin{aligned} e^{-x} &= 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots \\ \frac{1}{1+x} &= 1 - x + x^2 - x^3 + \cdots \\ \frac{e^{-x}}{1+x} &= 1 - 2x + \frac{5x^2}{2} - \cdots . \end{aligned}$$

11 It can be proved that if the series in the first of the formulas

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots \tag{1}$$

$$\frac{1}{f(x)} = b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + \cdots \tag{2}$$

converges to $f(x)$ when $|x| < r_1$ and if $a_0 \neq 0$, then there exist numbers $r_2, b_0, b_1, b_2, \cdots$ such that the series in (2) converges to $1/f(x)$ when $|x| < r_2$. Since the product of the left members of (1) and (2) has the power series expansion

$$1 + 0x + 0x^2 + 0x^3 + 0x^4 + \cdots ,$$

the coefficients $b_0, b_1, b_2, \cdots$ can be calculated from the formulas

$$\begin{aligned} 1 &= a_0b_0 \\ 0 &= a_0b_1 + a_1b_0 \\ 0 &= a_0b_2 + a_1b_1 + a_2b_0 \\ 0 &= a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0. \end{aligned}$$

Use this idea and the known power series expansion of $\cos x$ to obtain some of the coefficients in the expansion

$$\sec x = 1 + \tfrac{1}{2}x^2 + \tfrac{5}{24}x^4 + \tfrac{61}{720}x^6 + \tfrac{277}{8064}x^8 + \cdots.$$

12 We can be agreeably surprised by the simplicity of the operations which determine the first three or four of the b's in the formulas

$$\frac{\sin x}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots$$

$$\frac{x}{\sin x} = b_0 + b_2x^2 + b_4x^4 + b_6x^6 + \cdots$$

and yield the formula

$$\csc x = \frac{1}{x} + \frac{x}{6} + \frac{7x^3}{360} + \frac{31x^5}{15{,}120} + \frac{127}{604{,}800}x^7 + \cdots$$

which is valid when $0 < |x| < \pi$.

13 Start with the power series expansion of e^x and use it to obtain the formula

$$\frac{e^x - 1}{x} = 1 + \frac{x}{2!} + \frac{x^2}{3!} + \frac{x^3}{4!} + \frac{x^4}{5!} + \cdots.$$

Find the formula obtained by equating the derivatives of the members of this formula and putting $x = 1$ in the result. *Ans.:*

$$1 = \frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \frac{4}{5!} + \cdots.$$

14 Determine the first six of the coefficients in the formula

$$\frac{1}{1 + x + x^2} = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots.$$

Hint: Start by writing

$$\begin{aligned} 1 = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots \\ + a_0x + a_1x^2 + a_2x^3 + a_3x^4 + \cdots \\ + a_0x^2 + a_1x^3 + a_2x^4 + \cdots \end{aligned}$$

and obtaining the formulas in the first column

$$\begin{aligned} 1 &= a_0 & \qquad a_0 &= 1 \\ 0 &= a_0 + a_1 & a_1 &= -1 \\ 0 &= a_0 + a_1 + a_2 & a_2 &= 0 \\ 0 &= a_1 + a_2 + a_3 & a_3 &= 1 \\ &\cdots & &\cdots \end{aligned}$$

which determine the answers in the second column.

15 Obtain the given coefficients and two more coefficients in the formula

$$\frac{1 + 2x}{1 - x - x^2} = 1 + 3x + 4x^2 + 7x^3 + \cdots.$$

Hint: Start by writing

$$\frac{1+2x}{1-x-x^2} = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$$

and

$$\begin{aligned} 1 + 2x = a_0 + a_1x &+ a_2x^2 + a_3x^3 + a_4x^4 + \cdots \\ - a_0x &- a_1x^2 - a_2x^3 - a_3x^4 - \cdots \\ &- a_0x^2 - a_1x^3 - a_2x^4 - \cdots . \end{aligned}$$

Then obtain the formulas in the first column

$$\begin{array}{lcr} 1 = a_0 & \qquad\qquad & a_0 = 1 \\ 2 = a_1 - a_0 & & a_1 = 3 \\ 0 = a_2 - a_1 - a_0 & & a_2 = 4 \\ 0 = a_3 - a_2 - a_1 & & a_3 = 7 \\ \cdots & & \cdots \end{array}$$

which determine the answers in the second column.

16 The sequences

$$(1) \qquad 1, 1, 2, 3, 5, 8, 13, 21, \cdots$$

$$(2) \qquad 1, 3, 4, 7, 11, 18, 29, 47, \cdots$$

are examples of *Fibonacci sequences*, that is, sequences for which each element after the first two is the sum of its two nearest predecessors. A little work with power series reveals some surprising information about these famous sequences. Let $F_0, F_1, F_2, \cdots$ be a Fibonacci sequence. Letting g be defined by the first of the formulas

$$(3) \qquad g(x) = F_0 + F_1x + F_2x^2 + F_3x^3 + F_4x^4 + \cdots$$

$$(4) \qquad xg(x) = F_0x + F_1x^2 + F_2x^3 + F_3x^4 + \cdots$$

$$(5) \qquad x^2g(x) = F_0x^2 + F_1x^3 + F_2x^4 + \cdots,$$

tell how the next two are obtained. Subtract the last two formulas from the first and use the result to show that

$$(6) \qquad g(x) = \frac{F_0 + (F_1 - F_0)x}{1 - x - x^2}.$$

New and illuminating formulas are obtained by expressing $g(x)$ as a sum of partial fractions and expanding these fractions into power series. To simplify writing, let

$$(7) \qquad A = \frac{\sqrt{5}-1}{2} = 0.618034, \qquad B = \frac{\sqrt{5}+1}{2} = 1.618034.$$

Observe that $AB = 1$. Show (the details are a bit onerous) that

$$(8) \qquad \frac{1}{1-x-x^2} = \frac{1}{\sqrt{5}}\left[\frac{B}{1-Bx} + \frac{A}{1+Ax}\right]$$

and hence that, when $|Bx| < 1$,

$$(9) \qquad \frac{1}{1-x-x^2} = \sum_{n=0}^{\infty} \frac{1}{\sqrt{5}}[B^{n+1} + (-1)^nA^{n+1}]x^n.$$

For the case in which $F_0 = F_1 = 1$ as in (1), the formulas (3) and (6) and (9) show that

$$F_n = \frac{1}{\sqrt{5}}\,[B^{n+1} + (-1)^n A^{n+1}]. \tag{10}$$

Show that, when $n > 1$,

$$\frac{F_n}{F_{n-1}} = \frac{B^{n+1} + (-1)^n A^{n+1}}{B^n - (-1)^n A^n} = \frac{B + (-1)^n A^{n+1}/B^n}{1 - (-1)^n A^n/B^n} \tag{11}$$

and hence that

$$\lim_{n\to\infty} \frac{F_n}{F_{n-1}} = B = \frac{\sqrt{5}+1}{2}. \tag{12}$$

Remark: The function g in (3) is called the *generating function* of the sequence $F_0, F_1, F_2, \cdots$. The things which we have done are of interest in many branches of mathematics and can be extended in many ways.

17 Suggest a few ways in which the expansion

$$\sin^2 x = \frac{2x^2}{2!} - \frac{2^3x^4}{4!} + \frac{2^5x^6}{6!} - \frac{2^7x^8}{8!} + \cdots$$

can be obtained. *Remark:* Perhaps the simplest way involves the identities

$$\sin^2 x = \tfrac{1}{2} - \tfrac{1}{2}\cos 2x = \tfrac{1}{2} - \tfrac{1}{2}\left(1 - \frac{(2x)^2}{2!} + \frac{(2x)^4}{4!} - \frac{(2x)^6}{6!} + \cdots\right).$$

18 Supposing that $x > 1$, obtain the formula

$$\tan^{-1} x = \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \frac{1}{7x^7} - \cdots$$

in two different ways. First, use the formula $\tan^{-1} x = \pi/2 - \tan^{-1}(1/x)$ and then use a modification of part (*d*) of Problem 2. Secondly, start with the identity

$$\frac{1}{1+t^2} = \frac{1}{t^2}\,\frac{1}{1+\frac{1}{t^2}} = \frac{1}{t^2}\left[1 - \frac{1}{t^2} + \left(\frac{1}{t^2}\right)^2 - \left(\frac{1}{t^2}\right)^3 + \cdots\right]$$

and integrate over the interval $t \geq x$.

19 We often use the fact that the elementary expression for the left member of the formula

$$\int_1^x u^s\,du = \frac{\log x}{1!} + \frac{(s+1)(\log x)^2}{2!} + \frac{(s+1)^2(\log x)^3}{3!} + \frac{(s+1)^3(\log x)^4}{4!} + \cdots, \tag{1}$$

in which $x > 0$, has one form when $s \neq -1$ and has another form when $s = -1$. Prove that (1) is correct for each s. *Hint:* For the case in which $s \neq -1$, evaluate the integral and expand the result into a series with the aid of the fact that $x^{s+1} = e^{(s+1)\log x}$. Treat the case in which $s = -1$ separately.

20 Without pretending to give a reasonable introduction to complex numbers, we take a hasty look at some remarkable formulas that involve these numbers. Let i, the so-called *imaginary unit*, be a number for which $i^2 = -1$. It is possible to set forth rules for operating with *complex numbers* of the form $x + iy$, where x and y are real. For example, if $z = x + iy$, then

$$z^2 = (x + iy)^2 = x^2 + 2ixy + i^2y^2 = (x^2 - y^2) + 2ixy.$$

A complex number $x + iy$ for which x and y are real and $y = 0$ is identified with the real number x, so the set of real numbers is a subset of the set of complex numbers. The series in the right members of the formulas

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \frac{z^5}{5!} + \cdots \tag{1}$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots \tag{2}$$

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots \tag{3}$$

are power series in z. With the agreement that a sequence $x_1 + iy_1$, $x_2 + iy_2$, $\cdots$ of complex numbers converges to $L_1 + iL_2$ if $\lim_{n\to\infty} x_n = L_1$ and $\lim_{n\to\infty} y_n = L_2$, it is possible to construct a theory of series of complex numbers. It can then be shown that, for each z, the series in (1), (2), and (3) converge to complex numbers which are real only in special cases and which can be denoted by e^z, $\cos z$, and $\sin z$ even when they are not real. Thus e^z, $\cos z$, and $\sin z$ are defined for each complex z and the formulas (1), (2), (3) are valid for each z. If w is a complex number, which may be real but is not necessarily so, we can put $z = iw$ in (1) to obtain

$$\begin{aligned} e^{iw} &= 1 + 0 + \frac{(iw)^2}{2!} + 0 + \frac{(iw)^4}{4!} + 0 + \cdots \\ &\quad + 0 + iw + 0 + \frac{(iw)^3}{3!} + 0 + \frac{(iw)^5}{5!} + \cdots . \end{aligned} \tag{4}$$

Since $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$, $i^6 = -1$, $\cdots$, this formula can be put in the form

$$e^{iw} = \left(1 - \frac{w^2}{2!} + \frac{w^4}{4!} - \frac{w^6}{6!} + \cdots\right) + i\left(w - \frac{w^3}{3!} + \frac{w^5}{5!} - \frac{w^7}{7!} + \cdots\right). \tag{5}$$

Since (2) and (3) show that the series in parentheses converge to $\cos w$ and $\sin w$, we obtain the first of the four *Euler formulas*

$$e^{iw} = \cos w + i \sin w \tag{6}$$

$$e^{-iw} = \cos w - i \sin w \tag{7}$$

$$\cos w = \frac{e^{iw} + e^{-iw}}{2}, \qquad \sin w = \frac{e^{iw} - e^{-iw}}{2i}. \tag{8}$$

Since (2) and (3) show that $\cos(-w) = \cos w$ and $\sin(-w) = -\sin w$, we can replace w by $-w$ in (6) to obtain (7). Adding and subtracting (6) and (7) enables us to solve for $\cos w$ and $\sin w$ to obtain (8). Thus we have proved the four Euler formulas. These formulas are sometimes said to be the most remarkable formulas in mathematics. They have very many important applications, and it is worthwhile to be able to start with a clean sheet of paper and write all of the formulas needed to derive them.

21 If $|a_n|^{1/n} < R_1$ for each sufficiently great n, prove that $\Sigma a_n x^n$ converges when $|x| < 1/R_1$. *Solution:* When $|a_n|^{1/n} < R_1$, we find that

$$|a_n|^{1/n}|x| \leqq |xR_1|, \qquad |a_n x^n| \leqq |xR_1|^n.$$

The hypothesis that $|x| < 1/R_1$ implies that $|xR_1| < 1$, so $\Sigma|xR_1|^n < \infty$ and convergence of $\Sigma a_n x^n$ follows from the comparison test.

22 If $R_2 > 0$ and $|a_n|^{1/n} \geqq R_2$ for an infinite set of values of n, prove that $\Sigma a_n x^n$ diverges when $|x| \geqq 1/R_2$. *Solution:* When $|a_n|^{1/n} \geqq R_2$, we find that

$$|a_n|^{1/n}|x| \geqq R_2|x|, \qquad |a_n x^n| \geqq (R_2|x|)^n.$$

The hypothesis that $|x| \geqq 1/R_2$ implies that $R_2|x| \geqq 1$. Hence $|a_n x^n| \geqq 1$ for an infinite set of values of n. Thus it cannot be true that

$$\lim_{n\to\infty} a_n x^n = 0,$$

and therefore $\Sigma a_n x^n$ must be divergent.

23 Prove the following famous theorem, which is known as the Abel power series theorem. *If the series in*

$$s = a_0 + a_1 + a_2 + a_3 + \cdots \tag{1}$$

converges to s, *then the series in*

$$f(r) = a_0 + a_1 r + a_2 r^2 + a_3 r^3 + \cdots \tag{2}$$

converges when $0 < r < 1$ *and*

$$\lim_{r\to 1-} f(r) = s. \tag{3}$$

Solution: Let $s_n = a_0 + a_1 + \cdots + a_n$, so that

$$\lim_{n\to\infty} s_n = s. \tag{4}$$

There must be a constant M such that $|a_k| \leqq M$ and $|s_k| \leqq M$ for each k. When $|r| < 1$, the series in (2) and the series in

$$f(r) = s_0 + (s_1 - s_0)r + (s_2 - s_1)r^2 + (s_3 - s_2)r^3 + \cdots \tag{5}$$

are therefore both convergent and they both converge to the same number $f(r)$ because $s_0 = a_0$, $s_1 - s_0 = a_1$, $s_2 - s_1 = a_1$, $\cdots$. Because the separate series are both convergent, it follows from (5) that

$$f(r) = (s_0 + s_1 r + s_2 r^2 + \cdots) - (0 + s_0 r + s_1 r^2 + \cdots) \tag{6}$$

and hence that

$$f(r) = (1 - r)(s_0 + s_1 r + s_2 r^2 + \cdots). \tag{7}$$

But

$$s = (1 - r)(s + sr + sr^2 + \cdots) \tag{8}$$

and hence

$$f(r) - s = (1 - r)\sum_{k=0}^{\infty} (s_k - s)r^k. \tag{9}$$

Let $\epsilon > 0$. Choose an integer N such that $|s_k - s| < \epsilon/2$ when $n > N$ and let $C = \sum_{k=0}^{N} |s_k - s|$. Then, when $0 < r < 1$,

$$\begin{aligned}(10)\qquad |f(r) - s| &\leqq (1 - r)C + (1 - r) \sum_{k=N+1}^{\infty} |s_k - s|r^k \\ &< (1 - r)C + (1 - r) \sum_{k=N+1}^{\infty} \frac{\epsilon}{2} r^k \\ &< (1 - r)C + \frac{\epsilon}{2}.\end{aligned}$$

If we choose δ such that $(1 - r)C < \epsilon/2$ when $1 - \delta < r < 1$, we will have

$$(11)\qquad |f(r) - s| < \epsilon \qquad (1 - \delta < r < 1).$$

This gives (3) and the theorem is proved.

24 Tell why the Abel theorem of the preceding problem implies that

$$\lim_{x\to 1-} \left(x - \frac{x^2}{2^2} + \frac{x^3}{3^2} - \frac{x^4}{4^2} + \cdots\right) = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \cdots$$

and hence that part (g) of Problem 2 implies that

$$\int_0^1 \frac{\log(1 + t)}{t}\, dt = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \cdots.$$

25 Some of the most honorable parts of mathematics involve connections between the Riemann zeta function and prime numbers. Deriving a basic formula is a good exercise for us. Let $p_1, p_2, p_3, \cdots$ denote in order the prime numbers $2, 3, 5, 7, 11, 13, \cdots$. Euclid proved that the set of primes is infinite. We use the *fundamental theorem of arithmetic* which says that if n is an integer for which $n > 1$ and if p_k is the greatest prime factor of n, then n is uniquely representable in the form

$$n = p_1^{\lambda_1} p_2^{\lambda_2} \cdots p_k^{\lambda_k},$$

where the exponents $\lambda_1, \lambda_2, \cdots, \lambda_k$ are nonnegative integers. For example, $504 = 2^3 3^2 5^0 7^1$. Show that, when k is a positive integer and $s > 1$,

$$(1)\qquad \frac{1}{1 - \dfrac{1}{2^s}} = 1 + \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{8^s} + \cdots$$

$$(2)\qquad \frac{1}{1 - \dfrac{1}{3^s}} = 1 + \frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{27^s} + \cdots$$

$$(3)\qquad \frac{1}{1 - \dfrac{1}{5^s}} = 1 + \frac{1}{5^s} + \frac{1}{25^s} + \frac{1}{125^s} + \cdots$$

$$(4)\qquad \frac{1}{1 - \dfrac{1}{p_k^s}} = 1 + \frac{1}{p_k^s} + \frac{1}{p_k^{2s}} + \frac{1}{p_k^{3s}} + \cdots.$$

Making wholesale use of the rule of Problem 8 for multiplication of series, show that

(5) $$\frac{1}{1-\dfrac{1}{p_1^s}}\,\frac{1}{1-\dfrac{1}{p_2^s}}\cdots\frac{1}{1-\dfrac{1}{p_k^s}} = \sum_{n=1}^{\infty}{}^{*}\,\frac{1}{n^s},$$

where the star on the sigma means that some of the terms for which $n > p_k$ are omitted from the series. Prove the formula

(6) $$\sum_{n=1}^{p_k}\frac{1}{n^s} \leqq \sum_{n=1}^{\infty}{}^{*}\,\frac{1}{n^s} \leqq \sum_{n=1}^{\infty}\frac{1}{n^s} = \zeta(s)$$

and use it to prove that

(7) $$\lim_{k\to\infty}\frac{1}{1-\dfrac{1}{p_1^s}}\,\frac{1}{1-\dfrac{1}{p_2^s}}\,\frac{1}{1-\dfrac{1}{p_3^s}}\cdots\frac{1}{1-\dfrac{1}{p_k^s}} = \zeta(s).$$

Use (5) to obtain the inequality

(8) $$\frac{p_1^s}{p_1^s-1}\,\frac{p_2^s}{p_2^s-1}\cdots\frac{p_k^s}{p_k^s-1} \geqq \sum_{n=1}^{p_k}\frac{1}{n^s}$$

and show that taking the limit as $s \to 1+$ gives the inequality

(9) $$\frac{p_1}{p_1-1}\,\frac{p_2}{p_2-1}\cdots\frac{p_k}{p_k-1} \geqq \gamma + \log p_k$$

where γ is the Euler constant. Show that neglecting the γ and taking logarithms gives

(10) $$\sum_{k=1}^{n}\log\left(1+\frac{1}{p_k-1}\right) > \log\log p_n.$$

Show that $x > \log(1+x)$ when $x > 0$ and hence that

(11) $$\sum_{k=1}^{n}\frac{1}{p_k-1} > \log\log p_n$$

and

(12) $$\sum_{k=1}^{n}\frac{1}{p_k} > \frac{1}{2}\log\log p_n.$$

Show finally that

(13) $$\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}+\cdots = \infty.$$

26 The theory of functions of a complex variable provides an elegant proof of the fact that if

(1) $$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^2 + \cdots$$

when $|x| < R$, then

(2) $$f'(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots$$

when $|x| < R$. Persons willing to go fishing even when no fish are caught can try to prove the result with rudimentary equipment. To simplify writing, let $a = 0$. Then, when $|x_0| < R$ and $|x| < R$,

$$f(x) = \sum_{k=0}^{\infty} c_k x^k = \lim_{n\to\infty} \sum_{k=0}^{n} c_k x^k,$$

$$f(x_0) = \sum_{k=0}^{\infty} c_k x_0^k = \lim_{n\to\infty} \sum_{k=0}^{n} c_k x_0^k,$$

so

$$f(x) - f(x_0) = \lim_{n\to\infty} \sum_{k=1}^{n} c_k(x^k - x_0^k)$$

and

$$\frac{f(x) - f(x_0)}{x - x_0} = \lim_{n\to\infty} \left[c_1 + \sum_{k=2}^{n} c_k(x^{k-1} + x^{k-2}x_0 + \cdots + x_0^{k-1})\right].$$

If we can prove that the limits exist, we can take limits as x approaches x_0 to obtain

$$f'(x_0) = \lim_{x\to x_0} \lim_{n\to\infty} \left[c_1 + \sum_{k=2}^{n} c_k(x^{k-1} + x^{k-2}x_0 + \cdots + x_0^{k-1})\right].$$

If we can prove that the same result is obtained by interchanging the order in which limits are taken, we obtain

$$f'(x_0) = \lim_{n\to\infty} \lim_{x\to x_0} \left[c_1 + \sum_{k=2}^{n} c_k(x^{k-1} + x^{k-2}x_0 + \cdots + x_0^{k-1})\right]$$

and hence

$$f'(x_0) = \lim_{n\to\infty} \left[c_1 + \sum_{k=2}^{n} kc_k x_0^{k-1}\right] = \sum_{k=1}^{\infty} kc_k x_0^{k-1}.$$

Our fishing expedition can be ended with the remark that we ran into questions involving iterated limits and change of order of limits that swamped us.

12.5 Taylor formulas with remainders In Section 12.4, we started with given convergent power series and found that these series are the Taylor series of the functions to which they converge. In this section we start with a given function f and study the general aspects and further applications of a method we have previously employed in special cases to obtain power series expansions of e^x, $\cos x$, and $\sin x$. We suppose that a and x are confined to an interval over which f has all of the continuous derivatives we want to use. Then

$$f(x) = f(a) + \int_a^x f'(t)\,dt. \tag{12.51}$$

Integrating by parts with

$$u = f'(t) \qquad dv = dt$$
$$du = f''(t)\,dt, \qquad v = -(x - t)$$

gives

$$f(x) = f(a) + [-f'(t)(x - t)]_a^x + \int_a^x f''(t)\frac{(x - t)}{1!}\,dt$$

or

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \int_a^x f''(t)\frac{(x - t)}{1!}\,dt. \tag{12.52}$$

Another integration by parts with

$$u = f''(t) \qquad dv = \frac{(x - t)}{1!}\,dt$$
$$du = f^{(3)}(t)\,dt, \qquad v = -\frac{(x - t)^2}{2!}$$

gives

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \int_a^x f^{(3)}(t)\frac{(x - t)^2}{2!}\,dt.$$

One more integration by parts gives the result of putting $n = 3$ in the formula

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + R_n(x), \tag{12.53}$$

where

$$R_n(x) = \frac{1}{n!}\int_a^x f^{(n+1)}(t)(x - t)^n\,dt, \tag{12.54}$$

and further integrations by parts give the results when $n = 4, 5, 6, \cdots$. The formula (12.53) is a *Taylor formula with remainder* $R_n(x)$. The right member of (12.54) is the *integral form* of the remainder. In some cases, (12.54) and other remainder formulas can be used to determine values of x for which

$$\lim_{n \to \infty} R_n(x) = 0. \tag{12.55}$$

For such values of x the *Taylor formula*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots \tag{12.56}$$

is valid. The right member of (12.56) is called *the Taylor expansion of* f *in powers of* $(x - a)$. Problem 5 gives an example of a function f whose Taylor expansion in powers of x exists and converges for each x to a number which differs from $f(x)$ when $x \neq 0$. For this and other reasons, it is sometimes necessary to use (12.54) and other remainder formulas to obtain numerical estimates of $|R_n(x)|$.

The fact that $(x - t)^n$ is either always positive or always negative when t lies between a and x enables us to use (12.54) to obtain other formulas for $R_n(x)$ that are sometimes, but by no means always, more easily used than (12.54) itself. The simplest and most widely used of these formulas is obtained by the observation that the value of $R_n(x)$ lies between the numbers obtained by replacing the factor $f^{(n+1)}(t)$ by its minimum and maximum values over the interval from a to x. Hence the intermediate-value theorem implies existence of a number x^* between a and x such that

$$R_n(x) = \frac{1}{n!}\int_a^x f^{(n+1)}(x^*)(x - t)^n\,dt$$

and hence

$$R_n(x) = \frac{f^{(n+1)}(x^*)}{(n+1)!}(x - a)^{n+1}. \tag{12.57}$$

This is the *Lagrange form of the remainder.*

The *binomial formula*

$$(1 + x)^q = 1 + \frac{q}{1}x + \frac{q(q-1)}{2!}x^2 + \frac{q(q-1)(q-2)}{3!}x^3 + \frac{q(q-1)(q-2)(q-3)}{4!}x^4 + \frac{q(q-1)(q-2)(q-3)(q-4)}{5!}x^5 + \cdots, \tag{12.58}$$

in which the exponent q is not necessarily an integer and $|x| < 1$, is used so often that many persons find it worthwhile to remember the rule by which the coefficients can be written. To prove the formula, we let $f(x) = (1 + x)^q$ and calculate the derivatives

$$f'(x) = q(1 + x)^{q-1}, \qquad f''(x) = q(q-1)(1 + x)^{q-2},$$
$$f'''(x) = q(q-1)(q-2)(1 + x)^{q-3},$$
$$f^{(n)}(x) = q(q-1)(q-2)\cdots(q-n+1)(1 + x)^{q-n}.$$

Since $f(0) = 1$, $f'(0) = q$, $f''(0) = q(q-1)$, $\cdots$, the series in (12.58) is indeed the Taylor expansion of f in powers of x. To prove that the series converges when $|x| < 1$, we can apply the ratio test, but to prove that the series converges to $(1 + x)^q$, we must show that $\lim R_n(x) = 0$. While we could (without being completely unfashionable) find that the Lagrange form of the remainder will work when $0 < x < 1$ and that another special form will work when $-1 < x < 0$, we shun these things and use the integral form (12.54) to obtain

$$(12.581)\quad R_n(x) = \frac{q(q-1)(q-2)\cdots(q-n)}{n!}\int_0^x (1+t)^{q-1}\left(\frac{x-t}{1+t}\right)^n dt.$$

The function ϕ for which $\phi(t) = (x-t)/(1+t)$ is monotone over the interval from 0 to x and $\phi(x) = 0$, so $|\phi(t)|$ must attain its maximum over the interval from 0 to x when $t = 0$. This maximum is therefore $|x|$. Hence

$$(12.582)\quad |R_n(x)| \leqq \frac{|q(q-1)(q-2)\cdots(q-n)|}{n!}|x|^n\left|\int_0^x (1+t)^{q-1}\,dt\right|.$$

In case q is a nonnegative integer or $x = 0$, it is easy to see that

$$(12.583)\qquad \lim_{n\to\infty} R_n(x) = 0,$$

because $R_n(x) = 0$ for each sufficiently great n. When q is not a nonnegative integer and $0 < |x| < 1$, an application of the ratio test gives (12.583). This establishes the binomial formula (12.58) for the case in which $|x| < 1$.

Problems 12.59

1 With the aid of Taylor formulas with remainders, obtain the expansions of f in powers of $x - a$ when

(*a*) $f(x) = e^x$, $a = 0$ (*b*) $f(x) = e^x$, $a = 1$

(*c*) $f(x) = \sin x$, $a = 0$ (*d*) $f(x) = \sin x$, $a = \dfrac{\pi}{4}$

(*e*) $f(x) = \cos x$, $a = 0$ (*f*) $f(x) = \cos x$, $a = \dfrac{\pi}{2}$

2 Supposing that $|x| < 1$, write two more terms in each series appearing in the calculations

$$\sin^{-1} x = \int_0^x \frac{1}{\sqrt{1-t^2}}\,dt$$

$$(1-x)^{-1/2} = 1 + \frac{-\frac{1}{2}}{1}(-x) + \frac{(-\frac{1}{2})(-\frac{3}{2})}{1\cdot 2}(-x)^2 + \cdots$$

$$= 1 + \frac{1}{2}x + \frac{1\cdot 3}{2\cdot 4}x^2 + \cdots$$

$$\sin^{-1} x = \int_0^x \left[1 + \frac{1}{2}t^2 + \frac{1\cdot 3}{2\cdot 4}t^4 + \cdots\right] dt$$

$$\sin^{-1} x = x + \frac{1}{2\cdot 3}x^3 + \frac{1\cdot 3}{2\cdot 4\cdot 5}x^5 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 7}x^7 + \cdots.$$

3 Write two more terms of each series appearing in the calculations

$$K = \int_0^{\pi/2} \frac{1}{\sqrt{1 - k^2 \sin^2 \theta}}\, d\theta$$

$$(1 - x)^{-1/2} = 1 + \frac{-\frac{1}{2}}{1}(-x) + \frac{(-\frac{1}{2})(-\frac{3}{2})}{1\cdot 2}(-x)^2 + \cdots$$

$$= 1 + \frac{1}{2}x + \frac{1\cdot 3}{2\cdot 4}x^2 + \cdots$$

$$K = \int_0^{\pi/2} \left[1 + \frac{1}{2}k^2 \sin^2 \theta + \frac{1\cdot 3}{2\cdot 4}k^4 \sin^2 \theta + \cdots\right] d\theta$$

in which $0 < k < 1$. Use of the formula

$$\int_0^{\pi/2} \sin^{2p} \theta\, d\theta = \frac{1\cdot 3\cdot 5 \cdots (2p - 1)}{2\cdot 4\cdot 6 \cdots (2p)} \frac{\pi}{2} \qquad (p = 1,2,3, \cdots)$$

then gives

$$K = \frac{\pi}{2}\left(1 + \frac{1^2}{2^2}k^2 + \frac{1^2\cdot 3^2}{2^2\cdot 4^2}k^4 + \frac{1^2\cdot 3^2\cdot 5^2}{2^2\cdot 4^2\cdot 6^2}k^6 + \cdots\right).$$

4 Supposing that $0 < k < 1$, let

$$E = \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta$$

and modify the work of the preceding problem to obtain

$$E = \frac{\pi}{2}\left(1 - \frac{1}{2^2}k^2 - \frac{1^2\cdot 3}{2^2\cdot 4^2}k^4 - \frac{1^2\cdot 3^2\cdot 5}{2^2\cdot 4^2\cdot 6^2}k^6 - \cdots\right).$$

It is not easy to know everything, and we may be unable to say whether our formula for E is valid when $k = 1$. Show that if it is, then

$$\frac{1}{2^2} + \frac{1^2\cdot 3}{2^2\cdot 4^2} + \frac{1^2\cdot 3^2\cdot 5}{2^2\cdot 4^2\cdot 6^2} + \frac{1^2\cdot 3^2\cdot 5^2\cdot 7}{2^2\cdot 4^2\cdot 6^2\cdot 8^2} + \cdots = 1 - \frac{2}{\pi}.$$

5 While this may not be an appropriate time to enter into details of proofs, the function f for which $f(0) = 0$ and

$$f(x) = e^{-1/x^2} \qquad (x \neq 0)$$

has continuous derivatives of all orders over the whole infinite interval $-\infty < x < \infty$. Moreover, $f^{(k)}(0) = 0$ for each $k = 1, 2, 3, \cdots$. In this case the Taylor formula (12.53) with $a = 0$ becomes

$$f(x) = 0 + 0 + \cdots + 0 + R_n(x).$$

The Taylor expansion of f in powers of x is therefore the series

$$0 + 0 + 0 + \cdots$$

which converges for each x but converges to $f(x)$ only when $x = 0$.

6 Little things like the formula

$$(1)\quad f(1) = f(0) + \frac{f'(0)}{1!}(1-0) + \frac{f''(0)}{2!}(1-0)^2 + \cdots + \frac{f^{(n)}(0)}{n!}(1-0)^n + \frac{f^{(n+1)}(t^*)}{(n+1)!}(1-0)^{n+1}$$

and the capacitors that appear in electrical networks have surprising applications. As we shall see, simple applications of (1) give Taylor formulas for functions of "several variables," "several" meaning more than one. Extensions to functions of more variables being easily made, we suppose that G is a function of two variables x and y. We suppose that (x_0,y_0) and (x,y) are interior points of some convex region, a circular disk, for example, over which G is continuous and has all of the continuous partial derivatives we want to use. Supposing finally that $0 \leqq t \leqq 1$, let

$$(2)\qquad f(t) = G(x_0 + t(x-x_0),\ y_0 + t(y-y_0)).$$

Then $f(1) = G(x,y)$ and $f(0) = G(x_0,y_0)$ and we can start production of the Taylor formulas. Differentiating (2) with the aid of the chain rule of Theorem 11.23 gives

$$(3)\quad f'(t) = G_x(x_0+t(x-x_0),\ y_0+t(y-y_0))(x-x_0) + G_y(x_0+t(x-x_0),\ y_0+t(y-y_0))(y-y_0).$$

We can use (1) with $n = 0$ and obtain the primitive but nevertheless useful Taylor formula

$$(4)\qquad G(x,y) = G(x_0,y_0) + G_x(x^*,y^*)(x-x_0) + G_y(x^*,y^*)(y-y_0),$$

where

$$(5)\qquad x^* = x_0 + t^*(x-x_0),\quad y^* = y_0 + t^*(y-y_0).$$

To prepare for more elaborate Taylor formulas, we put $t = 0$ in (3) to obtain

$$(6)\qquad f'(0) = G_x(x_0,y_0)(x-x_0) + G_y(x_0,y_0)(y-y_0)$$

and differentiate (3) with the aid of the chain rule to obtain

$$\begin{aligned}(7)\qquad f''(t) &= G_{xx}(x_0+t(x-x_0),\ y_0+t(y-y_0))(x-x_0)^2\\ &+ G_{xy}(x_0+t(x-x_0),\ y_0+t(y-y_0))(x-x_0)(y-y_0)\\ &+ G_{xy}(x_0+t(x-x_0),\ y_0+t(y-y_0))(y-y_0)(x-x_0)\\ &+ G_{yy}(x_0+t(x-x_0),\ y_0+t(y-y_0))(y-y_0)^2.\end{aligned}$$

We now have the material required to use (1) with $n = 1$. It is easy to continue the procedure, but the expressions we write will become more and more ponderous unless we introduce simplifying notation. We begin by abbreviating (3) to the form

$$(8)\qquad f'(t) = \left[(x-x_0)\frac{\partial}{\partial x} + (y-y_0)\frac{\partial}{\partial y}\right]G$$

and observing that, on account of the equality of the mixed partial derivatives,

(7) can be put in the form

(9) $$f^{(k)}(t) = \left[(x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right]^k G$$

when $k = 2$. It turns out that Taylor formulas with remainders are obtained by substituting (9) into (1), the partial derivatives being evaluated at (x_0,y_0) when $k \leqq n$ and at (x^*,y^*) when $k = n + 1$.

7 Let G be a function having continuous partial derivatives of first and second orders over a neighborhood of (x_0,y_0) in which (x,y) is supposed to lie. The Taylor formula of Problem 6 which terminates with second derivatives then takes the form

(1) $$G(x,y) = G(x_0,y_0) + G_x(x_0,y_0)(x - x_0) + G_y(x_0,y_0)(y - y_0) + \tfrac{1}{2}[G_{xx}(x^*,y^*)(x - x_0)^2 + 2G_{xy}(x^*,y^*)(x - x_0)(y - y_0) + G_{yy}(x^*,y^*)(y - y_0)^2],$$

where x^* lies between x_0 and x and y^* lies between y_0 and y except that $x^* = x_0$ when $x = x_0$ and $y^* = y_0$ when $y = y_0$. This formula is useful. For example, it provides an easy way of estimating the difference between $G(x,y)$ and $G(x_0,y_0)$ that is especially useful when $|x - x_0|$ and $|y - y_0|$ are so small that the last term is negligible in comparison to the two preceding terms. In particular, (1) gives us a chance to estimate the magnitude of the error involved when the number dz defined by

(2) $$dz = G_x(x_0,y_0)(x - x_0) + G_y(x_0,y_0)(y - y_0)$$

is taken as an approximation to the number Δz defined by

(3) $$\Delta z = G(x,y) - G(x_0,y_0).$$

It is quite possible to spend a few days solving problems involving these ideas, and the investment of time might even be a reasonably good one. We invest a few minutes to study extrema (local and global minima and maxima) of G. If $G(x,y)$ has an extremum at (x_0,y_0), then $G(x,y_0)$ must have an extremum at x_0 and hence $G_x(x_0,y_0) = 0$. Similarly, if $G(x,y)$ has an extremum at (x_0,y_0), then $G(x_0,y)$ must have an extremum at y_0 and hence $G_y(x_0,y_0) = 0$. Thus G cannot have an extremum at (x_0,y_0) unless

(4) $$G_x(x_0,y_0) = G_y(x_0,y_0) = 0.$$

To investigate the question whether G has an extremum at a point (x_0,y_0) for which (4) holds, we put (1) in the form

(5) $$G(x,y) - G(x_0,y_0) = \tfrac{1}{2}[(A + \epsilon_1)h^2 + 2(B + \epsilon_2)hk + (C + \epsilon_3)k^2],$$

where

(6) $$A = G_{xx}(x_0,y_0), \qquad B = G_{xy}(x_0,y_0), \qquad C = G_{yy}(x_0,y_0),$$

$h = x - x_0$, $k = y - y_0$, and the numbers ϵ_1, ϵ_2, ϵ_3 depend upon h and k and are small when $|h|$ and $|k|$ are small. In case $A \neq 0$ we can, when $|h|$ and $|k|$ are small enough to make $A + \epsilon_1 \neq 0$, put (5) in the form

(7) $$G(x,y) - G(x_0,y_0) = \frac{1}{2(A + \epsilon_1)}\{[(A + \epsilon_1)h + (B + \epsilon_2)k]^2 + [(A + \epsilon_1)(C + \epsilon_3) - (B + \epsilon_2)^2]k^2\}.$$

In case $AC - B^2 > 0$, the quantity in braces will be nonnegative whenever $|h|$ and $|k|$ are small enough to make

$$(A + \epsilon_1)(C + \epsilon_3) - (B + \epsilon_2)^2 > 0. \tag{8}$$

It follows that *if* (4) *holds and*

$$G_{xx}(x_0,y_0) > 0, \qquad G_{xx}(x_0,y_0)G_{yy}(x_0,y_0) - [G_{x,y}(x_0,y_0)]^2 > 0, \tag{9}$$

then G must have a minimum at (x_0,y_0) because in this case $A + \epsilon_1$ must have the same sign as A and the right member of (7) must be nonnegative when $|h|$ and $|k|$ are sufficiently small. Similarly *if* (4) holds and

$$G_{xx}(x_0,y_0) < 0, \qquad G_{xx}(x_0,y_0)G_{yy}(x_0,y_0) - [G_{xy}(x_0,y_0)]^2 > 0, \tag{10}$$

then G must have a maximum at (x_0,y_0) because in this case the right member of (7) must be nonpositive when $|h|$ and $|k|$ are sufficiently small. *In case* (4) *holds and* $AC - B^2 < 0$, *that is,*

$$G_{xx}(x_0,y_0)G_{yy}(x_0,y_0) - [G_{xy}(x_0,y_0)]^2 < 0, \tag{11}$$

the function G cannot have an extremum at (x_0,y_0). We omit proof of this fact, and we also omit discussion of the way in which Taylor formulas having more terms can (when the required derivatives exist) be used to discuss cases in which $AC - B^2 = 0$.

8 Supposing that $a \neq 0$ and $b^2 - 4ac \neq 0$, find the extrema of the function G for which

$$G(x,y) = ax^2 + bxy + cy^2. \tag{1}$$

Remark: When solving problems of this nature, it is usually safe to set

$$z = ax^2 + bxy + cy^2$$

and use the curly dee notation for partial derivatives. *Outline of solution:* The system of equations $\dfrac{\partial z}{\partial x} = 0$, $\dfrac{\partial z}{\partial y} = 0$ is satisfied only when $x = y = 0$. The formulas

$$\frac{\partial^2 z}{\partial x^2} = 2a, \qquad \frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial y\, \partial x}\right)^2 = 4ac - b^2 \tag{2}$$

and the italicized statements of Problem 7 show that z has a minimum at (0,0) if $a > 0$ and $4ac - b^2 > 0$, that z has a maximum at (0,0) if $a < 0$ and $4ac - b_0 > 0$, and that z has neither a maximum nor a minimum at (0,0) if $4ac - b^2 < 0$. *Remark:* The examples

$$z = x^2 + y^2, \qquad z = -x^2 - y^2, \qquad z = x^2 - y^2 \tag{3}$$

illustrate the three phenomena. The examples

$$z = 0, \qquad z = (2x - y)^2 \tag{4}$$

illustrate the cases excluded from this problem by the condition $b^2 - 4ac \neq 0$.

9 Supposing that $a \neq 0$, determine the points (x,y), if any, at which the function G defined by

$$G(x,y) = x^3 - 3axy + y^3$$

takes extreme values. *Ans.:* The critical points where the first-order partial derivatives both vanish are (0,0) and (a,a). At (0,0), G has neither a maximum nor a minimum. At (a,a), G has a minimum if $a > 0$ and a maximum if $a < 0$ Moreover, $G(a,a) = -a^3$.

10 Assuming that

$$G(x,y) = a_0 + b_1x + b_2y + c_1x^2 + c_2xy + c_3y^2 + d_1x^3 + d_2x^2y + d_3xy^3 + d_4x^3 + e_1x^4 + \cdot$$

over some neighborhood of the origin, and that the series can be differentiated termwise with respect to x and y as often as desired, determine enough coefficients to verify three or four terms in the expansion

$$\begin{aligned} G(x,y) = G(0,0) &+ [G_x(0,0)x + G_y(0,0)y] \\ &+ \frac{1}{2!}[G_{xx}(0,0)x^2 + 2G_{xy}(0,0)xy + G_{yy}(0,0)y^2] \\ &+ \frac{1}{3!}[G_{xxx}(0,0)x^3 + 3G_{xxy}(0,0)x^2y + 3G_{xyy}(0,0)xy^2 + G_{yyy}(0,0)y^3] + \cdots . \end{aligned}$$

11 Suggest two or more ways to obtain the power series in x and y which converges to e^{x+y} and use each method to obtain a few of the terms.

12.6 *Euler-Maclaurin summation formulas*† As we shall see, we need only one very simple idea to obtain some remarkably useful and important formulas involving the Bernoulli functions and numbers. While the index reveals locations of more information about Bernoulli functions and numbers, we start with the facts that

$$B_0(x) = 1 \tag{12.611}$$

$$B_n'(x) = B_{n-1}(x) \qquad (n = 1,2,3, \cdots) \tag{12.612}$$

$$\int_0^1 B_n(x)\,dx = 0 \qquad (n = 1,2,3, \cdots) \tag{12.613}$$

$$B_n(x+1) = B_n(x) \qquad (n = 0,1,2, \cdots) \tag{12.614}$$

over the interval $-\infty < x < \infty$, except that (12.612) fails to hold when n is 1 or 2 and x is an integer. The function $B_1(x)$ is the saw-tooth function having the graph shown in Figure 12.62. The Bernoulli numbers

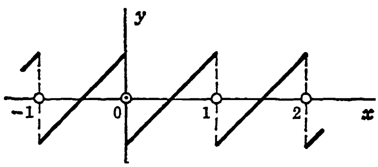


Figure 12.62

† This section can be omitted. It is not claimed that the section is easy. It is not claimed that the material can be thoroughly digested in a day or a week or a year. It is, however, claimed that students of calculus should see a substantial and useful application of calculus. Even though we build only more modest structures in examinations in this course, we should see at least one cathedral.

B_n are defined by the formulas

$$B_n = n!B_n(0) \qquad (n \geqq 2) \tag{12.63}$$

and

$$B_0 = 1,\ B_1 = -\tfrac{1}{2},\ B_2 = \tfrac{1}{6},\ B_3 = 0,\ B_4 = -\tfrac{1}{30},\ B_5 = 0,$$
$$B_6 = \tfrac{1}{42},\ B_7 = 0,\ B_8 = -\tfrac{1}{30},\ B_9 = 0,\ B_{10} = \tfrac{5}{66},\ \cdots. \tag{12.631}$$

Without yet knowing what is going to happen, let p and q be integers and let f be continuous and have all of the derivatives we want to use over the interval $p \leqq x \leqq q$. Letting k be an integer for which $p \leqq k < q$, we start with the simple idea that

$$\int_k^{k+1} f(x)\,dx = \int_k^{k+1} f(x)B_0(x)\,dx \tag{12.64}$$

and that we can modify the right side by integrating by parts with the aid of (12.612). However, since $B_1(x)$ is discontinuous at k and $k+1$, we must be careful. Accordingly, we put (12.64) in the form

$$\int_k^{k+1} f(x)\,dx = \lim_{\epsilon\to 0+} \int_{k+\epsilon}^{k+1-\epsilon} f(x)B_0(x)\,dx. \tag{12.65}$$

Setting

$$\begin{aligned} u &= f(x) & dv &= B_0(x)\,dx \\ du &= f'(x)\,dx, & v &= B_1(x) \end{aligned}$$

gives

$$\int_k^{k+1} f(x)\,dx = \lim_{\epsilon\to 0+} \left\{ f(x)B_1(x)\Big]_{k+\epsilon}^{k+1-\epsilon} - \int_{k+\epsilon}^{k+1-\epsilon} f'(x)B_1(x)\,dx \right\} \tag{12.651}$$

and hence

$$\int_k^{k+1} f(x)\,dx = \frac{f(k)+f(k+1)}{2} - \int_k^{k+1} f'(x)B_1(x)\,dx. \tag{12.652}$$

Adding the members of (12.652) for integer values of k for which $p \leqq k \leqq q-1$ gives the more useful identity

$$\int_p^q f(x)\,dx = \sum_{k=p}^{q} f(k) - \frac{f(p)+f(q)}{2} - \int_p^q f'(x)B_1(x)\,dx. \tag{12.653}$$

Transposing gives the basic Euler-Maclaurin formula

$$\sum_{k=p}^{q} f(k) = \int_p^q f(x)\,dx + \frac{f(p)+f(q)}{2} + \int_p^q f'(x)B_1(x)\,dx \tag{12.66}$$

which is used to estimate sums. More Euler-Maclaurin formulas

$$(12.661)\quad \sum_{k=p}^{q} f(k) = \int_p^q f(x)\,dx + \frac{f(p)+f(q)}{2} + \frac{f'(q)-f'(p)}{12} - \int_p^q f''(x)B_2(x)\,dx,$$

$$(12.662)\quad \sum_{k=p}^{q} f(k) = \int_p^q f(x)\,dx + \frac{f(p)+f(q)}{2} + \frac{f'(q)-f'(p)}{12} + \int_p^q f'''(x)B_3(x)\,dx,$$

etcetera, are easily obtained by further integrations by parts. The formula obtained after m integrations by parts is

$$(12.663)\quad \sum_{k=p}^{q} f(k) = \int_p^q f(x)\,dx + \frac{f(p)+f(q)}{2} + \sum_{j=2}^{m} [f^{(j-1)}(q) - f^{(j-1)}(p)]\frac{B_j}{j!} + (-1)^{m+1}\int_p^q f^{(m)}(x)B_m(x)\,dx.$$

This formula reduces to (12.66), (12.661), and (12.662) when m is 1, 2, and 3. We must observe that $B_j = 0$ when j is odd and $j \geqq 3$; otherwise, some of the signs in (12.663) would be wrong.

In some important applications, $f^{(m)}(x) \to 0$ and as $x \to \infty$ and the integral in

$$(12.664)\quad C_p = \frac{f(p)}{2} - \sum_{j=2}^{m} f^{(j-1)}(p)\frac{B_j}{j!} + (-1)^{m+1}\int_p^\infty f^{(m)}(x)B_m(x)\,dx$$

exists when m is sufficiently great, say $m \geqq m_0 \geqq 1$. In such cases, we can define the constant C_p by the formula (12.664), the right member being independent of m because integration by parts shows that it is unchanged when m is replaced by $m+1$. Subtracting (12.664) from (12.663) gives the formula

$$(12.665)\quad \sum_{k=p}^{q} f(k) = C_p + \int_p^q f(x)\,dx + \frac{f(q)}{2} + \sum_{j=2}^{m} f^{(j-1)}(q)\frac{B_j}{j!} + (-1)^m\int_q^\infty f^{(m)}(x)B_m(x)\,dx.$$

Solving this formula for C_p gives the formula

$$(12.666)\quad C_p = \sum_{k=p}^{q} f(k) - \int_p^q f(x)\,dx - \frac{f(q)}{2} - \sum_{j=2}^{m} f^{(j-1)}(q)\frac{B_j}{j!} - (-1)^m\int_q^\infty f^{(m)}(x)B_m(x)\,dx,$$

which is sometimes used to calculate approximations to C_p.

In many practical applications, the values of the integrals

$$(12.67)\qquad \int_p^q f^{(m)}(x)B_m(x)\,dx, \qquad \int_q^\infty f^{(m)}(x)B_m(x)\,dx$$

are not known, but this makes no difference because their algebraic signs can be determined and we discover that numbers we want to calculate lie between known numbers that are extremely close together. For example, the natures of $B_3(x)$ and $B_5(x)$ are such that if $f^{(3)}(x)$ and $f^{(5)}(x)$ are positive and decreasing, then the integrals in (12.67) are positive when $m = 3$ and are negative when $m = 5$.

The first problems at the end of this section give some of the simple applications of the Euler-Maclaurin summation formulas. In Problem 4 we shall derive some very important Stirling formulas. In Problem 5 we shall give an elementary proof of the fact that if $\omega_1 < \omega_2$ and

$$(12.68)\qquad P_n(\omega_1,\omega_2) = \sum_{\frac{n}{2}+\frac{\omega_1}{2}\sqrt{n} \leq k \leq \frac{n}{2}+\frac{\omega_2}{2}\sqrt{n}} \frac{1}{2^n}\frac{n!}{k!(n-k)!}$$

then

$$(12.681)\qquad \lim_{n\to\infty} P_n(\omega_1,\omega_2) = \sqrt{\frac{1}{2\pi}}\int_{\omega_1}^{\omega_2} e^{-x^2/2}\,dx.$$

This and related formulas are very important in probability and statistics. Persons who have or will have interest in these matters are well advised to complete this course and proceed to study an authoritative textbook by Feller† which many people read just for the fun of it.

Problems 12.69

1 Show that putting $p = 0$, $q = n$, and $f(x) = x^2$ in (12.662) gives the formula

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

2 Supposing that s is a positive integer, show that putting $p = 0$, $q = n$, $m = s$, and $f(x) = x^s$ in (12.663) gives the formula

$$\sum_{k=1}^{n} k^s = \frac{n^{s+1}}{s+1} + \frac{n^s}{2} + \sum_{j=2}^{s} s(s-1)(s-2)\cdots(s-j+2)n^{s-j+1}\frac{B_j}{j!}.$$

Remark: The result can be put in the neater form

$$\sum_{k=1}^{n} k^s = n^s + \frac{1}{s+1}\sum_{j=0}^{s}\binom{s+1}{j} n^{s+1-j}B_j$$

† William Feller, "An Introduction to Probability Theory and its Applications," John Wiley & Sons, Inc., New York, 1957.

involving binomial coefficients. For example, putting $s = 3$ gives

$$\sum_{k=1}^{n} k^3 = n^3 + \tfrac{1}{4}[n^4(1) + 4n^3(-\tfrac{1}{2}) + 6n^2(\tfrac{1}{6}) + 4n(0)],$$

so

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

Putting $s = 4$ gives

$$\sum_{k=1}^{n} k^4 = n^4 + \tfrac{1}{5}[n^5 + 5n^4(-\tfrac{1}{2}) + 10n^3(\tfrac{1}{6}) + 10n^2(0) + 5n(-\tfrac{1}{30})],$$

so

$$\sum_{k=1}^{n} k^4 = \frac{n(6n^4 + 15n^3 + 10n^2 - 1)}{30} = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}.$$

Some people spend huge amounts of time working out these formulas by other methods.

3 Show that setting $p = 1$, $q = n$, $f(x) = x^{-1}$, $m = 3$, and $C_1 = \gamma$ in (12.665) gives the formula

$$(1) \qquad \sum_{k=1}^{n} \frac{1}{k} = \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + 6\int_n^\infty \frac{B_3(x)}{x^4}\,dx,$$

where γ is the Euler constant. Show that setting $p = 1$, $q = 10$, $f(x) = x^{-1}$, and $C_1 = \gamma$ in (12.666) gives the formula

$$(2) \qquad \gamma = \sum_{k=1}^{10} \frac{1}{k} - \log 10 - \tfrac{1}{20} + \sum_{j=2}^{m} \frac{1}{10^j}\frac{B_j}{j} - \int_{10}^{\infty} \frac{m!}{x^{m+1}} B_m(x)\,dx.$$

Remark: Knowing that

$$\log 10 = 2.30258\ 50929\ 94046 \cdots,$$

it is not difficult to push a pencil through the calculations by which (2) is used with $m = 10$ to obtain the first 10 decimal places in

$$\gamma = 0.57721\ 56649\ 01532\ 86061 \quad \cdots.$$

When γ is known, (1) enables us to make very close estimates of the sum of 10 or more terms of the harmonic series.

4 Supposing that $z > -1$, show that putting $p = 1$, $q = n$, and $f(x) = \log(z + x)$ in (12.663) gives, when $m \geqq 2$,

$$(1) \quad \sum_{k=1}^{n} \log(z+k) = (z + n + \tfrac{1}{2})\log(z+n) - (z + \tfrac{1}{2})\log(z+1) - n + 1$$
$$+ \sum_{j=2}^{m} \frac{B_j}{(j-1)j}\left[\frac{1}{(z+n)^{j-1}} - \frac{1}{(z+1)^{j-1}}\right] + \int_1^n \frac{(m-1)!B_m(x)}{(z+x)^m}\,dx.$$

Remark: We proceed to show how this formula can be used to derive very important formulas involving factorials. Putting $z = 0$ in (1) gives

$$(2)\quad \log n! = (n + \tfrac{1}{2}) \log n - n + 1 + \sum_{j=2}^{m} \frac{B_j}{(j-1)j}\left[\frac{1}{n^{j-1}} - 1\right] + \int_1^n \frac{(m-1)!B_m(x)}{x^m}\,dx$$

One of the truly great mathematical discoveries is the fact that (2) can be improved with the aid of the formula

$$(3)\quad \lim_{n\to\infty} \frac{(2n)!\sqrt{n\pi}}{2^{2n}(n!)^2} = 1;$$

see Problems 9.59, Problem 6, equation (13). From (3) we obtain

$$(4)\quad \lim_{n\to\infty} \{\log (2n)! + \log \sqrt{n\pi} - 2n \log 2 - 2 \log n!\} = 0.$$

In this formula we substitute the expression for $\log n!$ given in (2) and the expression for $\log (2n)!$ obtained by replacing n by $2n$ in (2). The result should not overwhelm us, because we can overwhelm it. Many of the terms cancel out, the remaining ones have limits, and (4) reduces to

$$(5)\quad \log \sqrt{2\pi} - 1 + \sum_{j=2}^{m} \frac{B_j}{(j-1)j} - \int_1^\infty \frac{(m-1)!B_m(x)}{x^m}\,dx = 0.$$

Since

$$(6)\quad \int_1^n \frac{(m-1)!B_m(x)}{x^m}\,dx - \int_1^\infty \frac{(m-1)!B_m(x)}{x^m}\,dx = -\int_n^\infty \frac{(m-1)!B_m(x)}{x^m}\,dx = -\int_0^\infty \frac{(m-1)!B_m(x)}{(x+n)^m}\,dx,$$

adding the left side of (5) to the right side of (2) gives the remarkable *Stirling formula*

$$(7)\quad \log n! = \log \sqrt{2\pi} + (n + \tfrac{1}{2}) \log n - n + \sum_{j=2}^{m} \frac{B_j}{(j-1)j}\frac{1}{n^{j-1}} - \int_0^\infty \frac{(m-1)!B_m(x)}{(n+x)^m}\,dx.$$

To produce a Stirling formula applying to factorials of noninteger numbers, we suppose that $z > -1$ and use the definition

$$(8)\quad z! = \lim_{n\to\infty} \frac{n!n^z}{(z+1)(z+2)\cdots(z+n)},$$

which appeared in Section 3.3, Problem 11. This is equivalent to

$$(9)\quad \log z! = \lim_{n\to\infty} [\log n! + z \log n - \sum_{k=1}^{n} \log (z+k)].$$

In this formula we substitute the expression for $\log n!$ given in (7) and the expression for the last sum given in (1). With the aid of the fact that

$$\lim_{n\to\infty} (z + n + \tfrac{1}{2})[\log (z + n) - \log n] = z, \tag{10}$$

we find that

$$\log z! = \log \sqrt{2\pi} + (z + \tfrac{1}{2}) \log (z + 1) - (z + 1) + \sum_{j=2}^{m} \frac{B_j}{(j-1)j} \frac{1}{(z+1)^{j-1}} - \int_1^\infty \frac{(m-1)!B_m(x)}{(z+x)^m}\, dx. \tag{11}$$

In case $z > 0$, we can add $\log (z + 1)$ to both sides of (11) and then replace z by $z - 1$ to obtain the alternative formula

$$\log z! = \log \sqrt{2\pi} + (z + \tfrac{1}{2}) \log z - z + \sum_{j=2}^{m} \frac{B_j}{(j-1)j} \frac{1}{z^{j-1}} - \int_0^\infty \frac{(m-1)!B_m(x)}{(z+x)^m}\, dx, \tag{12}$$

which reduces to (7) when $z = n$. The derivation of Stirling formulas is not yet complete. To finish the task, we must study the theory of analytic functions of a complex variable. It will then be possible to observe that the members of (12) are analytic over the set consisting of complex numbers which are neither 0 nor negative. The principle of analytic extension then implies that the members of (12) are equal for each z which is neither 0 nor negative. We conclude with some remarks about (12). Let

$$E(z) = \sum_{j=2}^{m} \frac{B_j}{(j-1)j} \frac{1}{z^{j-1}} - \int_0^\infty \frac{(m-1)!B_m(x)}{(z+x)^m}\, dx. \tag{13}$$

We can then put (12) in the forms

$$\log z! = \log \sqrt{2z\pi} + \log z^z - z + E(z) \tag{14}$$

and

$$z! = \sqrt{2z\pi}\, z^z e^{-z} e^{E(z)}. \tag{15}$$

The formulas (12) and (14) are *Stirling formulas* for $\log z!$, and (15) is the *Stirling formula* for $z!$. In many applications, $E(z)$ is so near 0 that it can be neglected. Information about $E(z)$ is obtained from (13). When $n = 1$, the first sum in the right member of (13) contains no terms. Hence putting $m = 1$ in (13) gives

$$E(z) = -\int_0^\infty \frac{B_1(x)}{z+x}\, dx. \tag{16}$$

Putting $m = 3$ and 5 and 7 in (13) gives

$$E(z) = \frac{1}{12z} - \int_0^\infty \frac{B_3(x)}{(z+x)^3}\, dx \tag{17}$$

$$E(z) = \frac{1}{12z} - \frac{1}{360z^3} - \int_0^\infty \frac{6B_5(x)}{(z+x)^5}\, dx \tag{18}$$

$$E(z) = \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1260z^5} - \int_0^\infty \frac{120B_7(x)}{(z+x)^7}\, dx, \tag{19}$$

and more formulas can be produced by giving greater values to m. While these formulas have important applications to cases in which z is a complex number, we confine our attention to the case in which z is real and $z > 0$. When p is a positive integer, the function $(z + x)^{-p}$ is then positive and decreasing over the infinite interval $x \geqq 0$. Hence properties of the Bernoulli functions (those revealed in a problem at the end of Section 5.3) enable us to show that the integrals in (16), (17), (18), (19) are respectively negative, positive, negative, positive. Hence

$$E(z) > 0, \qquad E(z) < \frac{1}{12z}, \qquad E(z) > \frac{1}{12z} - \frac{1}{360z^3}, \tag{20}$$

and

$$E(z) < \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1260z^5}. \tag{21}$$

Even when z is as small as 2 or 3, this gives remarkably precise information about $E(z)$. In very many applications of these things, z is a positive integer n and (15) is put in the form

$$n! = \sqrt{2n\pi}\, n^n e^{-n} e^{\theta_n/12n}, \tag{22}$$

where $0 < \theta_n < 1$ and θ_n is near 1 whenever n is large. In fact, putting $E(z) = E(n) = \theta_n/12n$ in (20) and (21) shows that

$$1 - \frac{1}{30n^2} < \theta_n < 1 - \frac{1}{30n^2} + \frac{1}{105n^4}. \tag{23}$$

Thus θ_n is quite close to 1 even when $n = 1$.

5 When n is a positive integer, we can put $x = y = 1$ in the binomial formula

$$(x + y)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \tag{1}$$

to obtain the formula

$$\sum_{k=0}^{n} \frac{1}{2^n} \binom{n}{k} = 1. \tag{2}$$

To obtain more information about the terms in this sum, particularly when k is roughly $n/2$, we start with the formula

$$\log \frac{1}{2^n} \binom{n}{k} = -n \log 2 + \log n! - \log k! - \log (n-k)!. \tag{3}$$

The last three terms can be calculated from the Stirling formula

$$\log n! = \log \sqrt{2\pi} + (n + \tfrac{1}{2}) \log n - n + E_n \tag{4}$$

and the results of replacing n by k and by $(n - k)$ in it. The error term E_n, which can be approximated very closely with the aid of formula (13) of the preceding problem, is about $1/12n$ even when n is quite small. Except when n

and k are quite small, it is advisable to substitute in (3) and simplify the result before making numerical calculations. Thus we can put (3) in the form

$$(5)\quad \log \frac{1}{2^n}\binom{n}{k} = \log\sqrt{\frac{2}{n\pi}} - \left(k+\frac{1}{2}\right)\log\frac{2k}{n} - \left(n-k+\frac{1}{2}\right)\log 2\left(1-\frac{k}{n}\right) + E_{n,k},$$

where $E_{n,k} = E_n - E_k - E_{n-k}$ Much progress in probability and statistics is based upon the idea that when k is near $n/2$, we can represent it in the form

$$(6)\qquad k = \frac{n}{2} + \lambda\sqrt{n},$$

where λ is a number that depends upon k and n. When (6) holds, we find that

$$(7)\qquad \frac{2k}{n} = 1 + \frac{2\lambda}{\sqrt{n}},\qquad 2\left(1-\frac{k}{n}\right) = 1 - \frac{2\lambda}{\sqrt{n}}$$

and we can put (5) in the form

$$(8)\quad \log\frac{1}{2^n}\binom{n}{k} = \log\sqrt{\frac{2}{n\pi}} - \left(\frac{n}{2}+\lambda\sqrt{n}+\frac{1}{2}\right)\log\left(1+\frac{2\lambda}{\sqrt{n}}\right) - \left(\frac{n}{2}-\lambda\sqrt{n}+\frac{1}{2}\right)\log\left(1-\frac{2\lambda}{\sqrt{n}}\right) + E_{n,k}$$

and hence

$$(9)\quad \log\frac{1}{2^n}\binom{n}{k} = \log\sqrt{\frac{2}{n\pi}} - \frac{n+1}{2}\log\left(1-\frac{4\lambda^2}{n}\right) - \lambda\sqrt{n}\left[\log\left(1+\frac{2\lambda}{\sqrt{n}}\right) - \log\left(1-\frac{2\lambda}{\sqrt{n}}\right)\right] + E_{n,k}.$$

Now let α be a positive number and suppose that k is close enough to $n/2$ to make

$$(10)\qquad \frac{n}{2} - \alpha\sqrt{n} \leqq k = \frac{n}{2} + \lambda\sqrt{n} \leqq \frac{n}{2} + \alpha\sqrt{n}.$$

Then $|\lambda| \leqq \alpha$ and we will have

$$(11)\qquad \frac{4\lambda^2}{n} < \left|\frac{2\lambda}{\sqrt{n}}\right| \leqq \frac{2\alpha}{\sqrt{n}} < \frac{1}{2}$$

provided n is sufficiently great, say $n \geqq n_0$. It can be shown that there is a constant M_1 such that, when $|x| < \frac{1}{2}$,

$$(12)\qquad \log(1-x) = -x + F_1(x)|x|^2$$

$$(13)\qquad \log(1+x) - \log(1-x) = 2x + F_2(x)|x|^3,$$

where $|F_1(x)| \leqq M_1$ and $|F_2(x)| \leqq M_1$. These facts and (9) yield the conclusion that there are constants M and D_{nk} such that $|D_{nk}| \leqq M$ and

$$\log \frac{1}{2^n}\binom{n}{k} = \log \sqrt{\frac{2}{n\pi}} - 2\lambda^2 + \frac{D_{nk}}{n}. \tag{14}$$

This shows that if (10) and (11) hold, then

$$e^{-M/n}\sqrt{\frac{2}{n\pi}}\,e^{-2\lambda^2} \leqq \frac{1}{2^n}\binom{n}{k} \leqq e^{M/n}\sqrt{\frac{2}{n\pi}}\,e^{-2\lambda^2}. \tag{15}$$

Suppose now that x_1 and x_2 are two numbers, not necessarily positive, for which $x_1 < x_2$ and suppose that α has been chosen such that $|x_1| < \alpha$ and $|x_2| < \alpha$ Let

$$P_n(x_1,x_2) = \sum_{\frac{n}{2}+x_1\sqrt{n} \leqq k \leqq \frac{n}{2}+x_2\sqrt{n}} \frac{1}{2^n}\binom{n}{k}. \tag{16}$$

Since $\lim_{n\to\infty} e^{-M/n} = 1$ and $\lim_{n\to\infty} e^{M/n} = 1$, it follows from (15) that

$$\lim_{n\to\infty} P_n(x_1,x_2) = \lim_{n\to\infty} \sum\nolimits^* \sqrt{\frac{2}{n\pi}}\,e^{-2\lambda^2} \tag{17}$$

provided the limit on the right exists. In (17) and elsewhere, a star on a sigma indicates that the range of summation is the same as that in (16). For present purposes, let the number λ in (6), (15), and (17) be denoted by λ_{nk}, so that

$$k = \frac{n}{2} + \lambda_{nk}\sqrt{n}.$$

Then $\lambda_{n,k} - \lambda_{n,k-1} = 1/\sqrt{n}$, and consequently

$$\sum\nolimits^* \sqrt{\frac{2}{n\pi}}\,e^{-2\lambda_{nk}^2} = \sum\nolimits^* \sqrt{\frac{2}{\pi}}\,e^{-2\lambda_{nk}^2}(\lambda_{n,k} - \lambda_{n,k-1}). \tag{18}$$

The right member of (18) is, except for negligible discrepancies at the ends of the interval $x_1 \leqq x \leqq x_2$, a Riemann sum which converges to the right member of the formula

$$\lim_{n\to\infty} \sum\nolimits^* \sqrt{\frac{2}{n\pi}}\,e^{-2\lambda_{nk}^2} = \sqrt{\frac{2}{\pi}}\int_{x_1}^{x_2} e^{-2x^2}\,dx. \tag{19}$$

Therefore, (19) holds. From (19), (17), and (16) we obtain the formula

$$\lim_{n\to\infty} \sum_{\frac{n}{2}+x_1\sqrt{n} \leqq k \leqq \frac{n}{2}+x_2\sqrt{n}} \frac{1}{2^n}\binom{n}{k} = \sqrt{\frac{2}{\pi}}\int_{x_1}^{x_2} e^{-2x^2}\,dx. \tag{20}$$

In order to compare this result with other statistical results, we replace x_1 and x_2 by $\omega_1/2$ and $\omega_2/2$ in (20). Changing the variable of integration by setting

$x = t/2$ then gives the formula

$$(21)\qquad \lim_{n\to\infty} \sum_{\frac{n}{2}+\frac{\omega_1}{2}\sqrt{n} \leqq k \leqq \frac{n}{2}+\frac{\omega_2}{2}\sqrt{n}} \frac{1}{2^n}\binom{n}{k} = \frac{1}{\sqrt{2\pi}} \int_{\omega_1}^{\omega_2} e^{-t^2/2}\, dt.$$

The right side of (21) can be evaluated with the aid of tables giving values of

$$\frac{1}{\sqrt{2\pi}} \int_0^{\omega} e^{-t^2/2}\, dt$$

for various values of ω.

6 Our educations are not quite complete until we have seen the formula by which the power series expansion of $\tan x$ is expressed in terms of Bernoulli numbers Letting $B_0(x)$, $B_1(x)$, $\cdots$ be the Bernoulli functions, we start by deriving the formula

$$(1)\qquad \frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x)t^n,$$

which holds when $0 < x < 1$ and $0 < |t| < 2\pi$. To simplify matters, we can suppose at first that $0 < |t| < 1$. When $y(x)$ denotes the right side of (1), we can differentiate termwise to obtain

$$(2)\qquad \begin{aligned} y'(x) &= \sum_{n=1}^{\infty} B_n'(x)t^n = \sum_{n=0}^{\infty} B_{n+1}'(x)t^{n+1} \\ &= \sum_{n=0}^{\infty} B_n(x)t^{n+1} = t\sum_{n=0}^{\infty} B_n(x)t^n = ty(x). \end{aligned}$$

Therefore,

$$(3)\qquad \frac{d}{dx}[e^{-xt}y(x)] = e^{-xt}[y'(x) - ty(x)] = 0,$$

and it follows that for each t there is a constant $c(t)$ for which

$$(4)\qquad e^{-xt}y(t) = c(t) \qquad \text{or} \qquad y(t) = c(t)e^{xt}.$$

Therefore,

$$(5)\qquad c(t)e^{xt} = \sum_{n=0}^{\infty} B_n(x)t^n.$$

Integrating (5) over the interval $0 \leqq x \leqq 1$ and using (12.611) and (12.613) give the first and then the second of the formulas

$$(6)\qquad c(t)\frac{e^t - 1}{t} = 1, \qquad c(t) = \frac{t}{e^t - 1}.$$

Substituting in (5) then gives (1). Since $B_1(x) = x - \frac{1}{2}$ when $0 < x < 1$, we can subtract $(x - \frac{1}{2})t$ from both sides of (1) to obtain the formula

$$(7)\qquad \frac{t}{2}\,\frac{e^t + 2e^{xt} - 1 - 2x(e^t - 1)}{e^t - 1} = \sum_{n=0}^{\infty}{}^{*} B_n(x)t^n$$

in which the star on the Σ means that the term for which $n = 1$ is omitted from the series. More advanced mathematics contains theorems which allow us to take termwise limits, as x approaches zero through positive values, of the members of (7). This, the fact that

$$\lim_{x\to 0} B_n(x) = B_n(0) = \frac{B_n}{n!} \tag{8}$$

when $n \neq 1$, and the fact that $B_n = 0$ when n is odd and $n \neq 1$, give the formula

$$\frac{t}{2}\frac{e^t+1}{e^t-1} = B_0 + \frac{B_2}{2!}t^2 + \frac{B_4}{4!}t^4 + \frac{B_6}{6!}t^6 + \cdots = \sum_{k=0}^{\infty} \frac{B_{2k}}{(2k)!}t^{2k} \tag{9}$$

which is valid when $0 < |t| < 2\pi$. Putting $t = 2z$ in (9) gives the formula

$$z\frac{e^z+e^{-z}}{e^z-e^{-z}} = \sum_{k=0}^{\infty} \frac{2^{2k}B_{2k}}{(2k)!}z^{2k} \tag{10}$$

which is valid when $0 < |z| < \pi$. The theory of functions of a complex variable provides reasons why (10) is valid when z is a complex number for which $0 < |z| < \pi$, and we can put $z = i\theta$ to obtain

$$i\theta\frac{e^{i\theta}+e^{-i\theta}}{e^{i\theta}-e^{-i\theta}} = \sum_{k=0}^{\infty} \frac{2^{2k}B_{2k}}{(2k)!}(i\theta)^{2k} \tag{11}$$

when $0 < |\theta| < \pi$. Since $i^{2k} = (i^2)^k = (-1)^k$, use of the Euler formulas given in Problem 20 of Problems 12.49 enables us to put (11) in the form

$$\theta \cot \theta = \sum_{k=0}^{\infty} (-1)^k \frac{2^{2k}B_{2k}}{(2k)!}\theta^{2k} \tag{12}$$

when $0 < |\theta| < \pi$. Having established (12), we can show that

$$\theta \tan \theta = \theta \cot \theta - 2\theta \cot 2\theta \tag{13}$$

when $0 < |\theta| < \pi/2$ and use (12) to obtain our final formula

$$\tan \theta = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{2^{2k}(2^{2k}-1)B_{2k}}{(2k)!}\theta^{2k-1} \tag{14}$$

which is valid when $|\theta| < \pi/2$. The above formulas and modifications of them appear in quite elementary tables, but we must always be prepared to observe that some brief treatments of Bernoulli numbers use B_k to denote the number $B_{2k}(0)/(2k)!$ which we have called B_{2k}.

13 Iterated and multiple integrals

13.1 Iterated integrals When we differentiate a function f having values $f(x)$ and then iterate (or repeat) the process, we obtain functions having values denoted by $f'(x)$, $f''(x)$, $f'''(x)$, $\cdots$. When we iterate the process of integrating a function f over an interval from a to x, we need more elaborate notation to express the results in terms of the given function. To begin operations, we suppose that

$$(13.11) \quad f_1(x) = \int_a^x f(t)\,dt, \qquad f_2(x) = \int_a^x f_1(t)\,dt, \qquad f_3(x) = \int_a^x f_2(t)\,dt, \qquad \cdots .$$

To express f_2 in terms of f, we can avoid snarls of various kinds by replacing t by t_1 and then x by t to obtain

$$f_1(t) = \int_a^t f(t_1)\,dt_1.$$

Substituting this in the formula for $f_2(x)$ then gives

$$(13.12) \qquad f_2(x) = \int_a^x \left(\int_a^t f(t_1)\, dt_1 \right) dt.$$

Replacing t_1 by t_2 and then t by t_1 and x by t gives

$$f_2(t) = \int_a^t \left(\int_a^{t_1} f(t_2)\, dt_2 \right) dt_1$$

and substituting this into the formula for $f_3(x)$ gives

$$(13.13) \qquad f_3(x) = \int_a^x \left\{ \int_a^t \left(\int_a^{t_1} f(t_2)\, dt_2 \right) dt_1 \right\} dt.$$

The integrals in (13.12) and (13.13) are examples of iterated integrals and, for each n, we could write a formula for $f_n(x)$ which involves n of these integrals.

In case $f(x) = 1$ for each x, we do not need the iterated integrals to obtain formulas for $f_1(x), f_2(x), \cdots$; we can use the formulas (13.11) one after another to obtain

$$(13.14) \quad f_1(x) = x - a, \qquad f_2(x) = \frac{(x-a)^2}{2!}, \qquad f_3(x) = \frac{(x-a)^3}{3!},$$
$$f_4(x) = \frac{(x-a)^4}{4!}, \qquad \cdots .$$

We must, however, learn how iterated integrals are manipulated when they appear in our work and cannot be avoided. The way in which (13.12) was obtained tells us that to get $f_2(x)$ we should integrate first with respect to t_1 to evaluate the integral in parentheses to obtain a function of t which is integrated with respect to t to obtain $f_2(x)$. A tempest in a teapot appears when we, like everyone else, adopt the view that parentheses are nuisances and write (13.12) in the form

$$(13.15) \qquad f_2(x) = \int_a^x \int_a^t f(t_1)\, dt_1\, dt$$

and insist that we must find $f_2(x)$ by integrating first with respect to t_1, the limits of integration being a and t, and integrating last with respect to t from a to x. The difficulty lies in the fact that there is always the possibility of constructing a theory of iterated integrals in such a way that, for example,

$$(13.16) \qquad \int_A^B \int_C^D F(x,y)\, dx\, dy$$

means the result of integrating first with respect to y from C to D (*not* from A to B) and then integrating last with respect to x from A to B. It is equally sensible to insist on one hand that we should "work outward from the middle," so that $\int_C^D$ goes with dx and $\int_A^B$ goes with dy, and to insist on the other hand that we should "work from right to left," so that $\int_C^D$ goes with dy and $\int_A^B$ goes with dx. Of course, we could lengthen a long story by insisting that we should always keep the parentheses and avoid the tempests and the stories, but this is impractical. While we reserve the option of using parentheses whenever we wish to do so, we ordinarily remove parentheses and ambiguities from iterated integrals by writing the integrals in such a way that each integral sign except the one on the right is immediately followed by the symbol showing the variable of integration which has the limits of integration appearing on the integral sign. Thus, for example,

$$\begin{aligned}(13.17)\quad \int_1^2 dx \int_3^4 xy^2\,dy &= \int_1^2 dx\left[x\frac{y^3}{3}\right]_{y=3}^{y=4}\\ &= \frac{37}{3}\int_1^2 x\,dx = \frac{37}{3}\frac{x^2}{2}\bigg]_{x=1}^{x=2} = \frac{111}{6}\end{aligned}$$

$$\begin{aligned}(13.18)\quad \int_1^2 dy \int_3^4 xy^2\,dx &= \int_1^2 dy\left[\frac{x^2y^2}{2}\right]_{x=3}^{x=4}\\ &= \frac{7}{2}\int_1^2 y^2\,dy = \frac{7}{2}\frac{y^3}{3}\bigg]_1^2 = \frac{49}{6}\end{aligned}$$

and

$$\begin{aligned}(13.181)\quad \int_0^x dt \int_0^t (t-u)\,du &= \int_0^x dt\left[-\frac{(t-u)^2}{2}\right]_{u=0}^{u=t}\\ &= \frac{1}{2}\int_0^x t^2\,dt = \frac{1}{2}\frac{t^3}{3}\bigg]_{t=0}^{t=x} = \frac{x^3}{6}.\end{aligned}$$

Note that, in each case, the integral appearing on the right is evaluated first. Note also that when we are in the process of integrating with respect to a particular variable, all other variables are temporarily considered to be constants. Opportunities to become familiar with these things are provided by the following problems.

Problems 13.19

1 Show that

(*a*) $\int_0^1 dx \int_0^x (x^2+y^2)\,dy = \frac{1}{3}$ (*b*) $\int_0^1 dt \int_0^t (t^n+u^n)\,du = \frac{1}{n+1}$ $(n > -1)$

(*c*) $\int_0^1 dx \int_0^x dy \int_0^{xy} dz = \frac{1}{8}$ (*d*) $\int_0^\infty dx \int_x^{x+1} e^{-y}\,dy = 1 - e^{-1}$

2 By evaluating all of the integrals involved, show that

(*a*) $\int_0^1 dx \int_0^2 x\,dy = \int_0^2 dy \int_0^1 x\,dx$ (*b*) $\int_0^1 dx \int_1^2 x\,dy = \int_1^2 dy \int_0^1 x\,dx$

3 By evaluating all of the integrals involved, show that

$$\int_0^t dx \int_0^x f(y)\, dy = \int_0^t (t - y)f(y)\, dy$$

when

(*a*) $p > -1$ and $f(y) = y^p$
(*b*) $k > 0$ and $f(y) = e^{-ky}$
(*c*) $\omega \neq 0$ and $f(y) = \sin \omega y$

4 The formula

$$\int_0^t u\, dv = uv\Big]_0^t - \int_0^t v\, du,$$

which abbreviates the formula

$$\int_0^t u(x)v'(x)\, dx = u(x)v(x)\Big]_{x=0}^{x=t} - \int_0^t v(x)u'(x)\, dx$$

for integration by parts, has unexpected applications. Assuming that f is continuous and

$$I = \int_0^t dx \int_0^x f(y)\, dy,$$

find the result of integrating by parts with

$$u(x) = \int_0^x f(y)\, dy \qquad v'(x) = 1$$
$$u'(x) = f(x), \qquad v(x) = -(t - x).$$

5 Calculate the two integrals I and J defined by

$$I = \int_0^a dx \int_0^x f(x,y)\, dy, \qquad J = \int_0^a dy \int_y^a f(x,y)\, dx$$

and show that they are equal when

(*a*) $p > -1$, $q > -1$ and $f(x,y) = x^p y^q$
(*b*) $f(x,y) = e^{x+y}$

6 Show that, when $n > -1$,

$$\int_0^1 dx \int_0^1 (x + y)^n\, dy = \frac{2^{n+2} - 2}{(n + 1)(n + 2)}.$$

7 Show that

$$\int_0^1 dx \int_0^1 \frac{1}{x + y}\, dy = 2 \log 2.$$

8 Show that

$$\int_0^1 dx \int_0^1 \frac{1}{(x + y)^2}\, dy = \infty.$$

9 Supposing that $0 < p < 2$ and $p \neq 1$, show that

$$\int_0^1 dx \int_0^1 \frac{1}{(x + y)^p}\, dy = \frac{2^{2-p} - 2}{(2 - p)(1 - p)}.$$

10 Supposing that $n > -2$ and $n \neq 1$, show that

$$\int_0^1 dx \int_0^x (x + y)^n \, dy = \frac{2^{n+1} - 1}{(n + 1)(n + 2)}.$$

Investigate the case in which $n = -1$.

11 Show that

$$(a)\ \int_0^1 dx \int_x^{2x} \frac{y}{x} \, dy = \frac{3}{4} \qquad (b)\ \int_0^1 dx \int_x^1 \frac{y}{x} \, dy = \infty$$

$$(c)\ \int_0^1 dx \int_x^{2x} \frac{x}{y} \, dy = \frac{\log 2}{2} \qquad (d)\ \int_0^1 dx \int_x^1 \frac{x}{y} \, dy = \frac{1}{4}$$

12 Show that making one integration gives the formula

$$2 \int_0^1 dx \int_0^x \frac{1}{1 - xy} \, dy = 2 \int_0^1 \frac{1}{x} \log \frac{1}{1 - x^2} \, dx$$

which, so far as we know, has dubious validity because the last integrand is meaningless when $x = 0$ and when $x = 1$. Show that

$$\int_0^x \frac{1}{1 - xy} \, dy$$

has the value 0 when $x = 0$ and does not exist (or has the value $+\infty$) when $x = 1$. Then proceed to the next problem.

13 The integrals in

$$(1) \qquad 2 \int_0^1 dx \int_0^x \frac{1}{1 - xy} \, dy$$

cannot exist as iterated Riemann integrals because the integral in

$$(2) \qquad f(x) = \int_0^x \frac{1}{1 - xy} \, dy$$

does not exist as a Riemann integral when $x = 1$. However, when $0 \leqq x < 1$,

$$(3) \qquad \begin{aligned} f(x) &= \int_0^x (1 + xy + x^2y^2 + x^3y^3 + \cdots) \, dy \\ &= \left[y + \frac{xy^2}{2} + \frac{x^2y^3}{3} + \frac{x^3y^4}{4} + \cdots \right]_{y=0}^{y=x} \\ &= x + \frac{x^3}{2} + \frac{x^5}{3} + \frac{x^7}{4} + \cdots . \end{aligned}$$

To each positive integer n there corresponds a positive number δ such that

$$(4) \qquad f(x) \geqq x + \frac{x^3}{2} + \frac{x^5}{3} + \cdots + \frac{x^{2n-1}}{n} > \left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) - 1$$

when $1 - \delta < x < 1$ (why?) and hence (why?)

$$(5) \qquad \lim_{x \to 1-} f(x) = \infty .$$

The integral in (1) will exist as a Riemann-Cauchy integral and will have the value V if

$$(6)\qquad \begin{aligned} V &= \lim_{r\to 1-} 2\int_0^r \left(x + \frac{x^3}{2} + \frac{x^5}{3} + \frac{x^7}{4} + \cdots\right) dx \\ &= \lim_{r\to 1-} \left(\frac{r^2}{1^2} + \frac{r^4}{2^2} + \frac{r^6}{3^2} + \frac{r^8}{4^2} + \cdots\right). \end{aligned}$$

With the aid of the basic fact that $\Sigma(1/n^2) = \pi^2/6$, prove that $V = \pi^2/6$. *Hint:* Supposing that $0 < r < 1$, let the last series in (6) converge to $g(r)$ and begin by showing that $g(r) < \pi^2/6$. Then compare $g(r)$ with $\sum_{k=1}^{n} 1/k^2 - \epsilon/2$

14 Prove that the formula

$$\int_a^b dx \int_c^d f(x,y)\, dy = \int_a^b dx \int_c^d g(x,y)\, dy + \int_a^b dx \int_c^d h(x,y)\, dy$$

is valid provided (i) $f(x,y) = g(x,y) + h(x,y)$ when $a \leqq x \leqq b$ and $c \leqq y \leqq d$ and (ii) the integrals in the right member exist. *Hint:* Let

$$F(x) = \int_c^d f(x,y)\, dy, \qquad G(x) = \int_c^d g(x,y)\, dy, \qquad H(x) = \int_c^d h(x,y)\, dy.$$

and use known facts about simple integrals.

15 Let

$$(1)\qquad \mathbf{F}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k},$$

where the functions are continuous over some spherical ball B with center at (x_0,y_0,z_0), be the force on a particle when the particle is at the point (x,y,z). Show that the work $W_1(x,y,z)$ done by the force in moving the particle along line segments from (x_0,y_0,z_0) to (x_0,y,z_0) and then to (x_0,y,z) and finally to the point (x,y,z) in the ball is

$$(2)\qquad W_1(x,y,z) = \int_{y_0}^{y} Q(x_0,\beta,z_0)\, d\beta + \int_{z_0}^{z} R(x_0,y,\gamma)\, d\gamma + \int_{x_0}^{x} P(\alpha,y,z)\, d\alpha$$

and that

$$(3)\qquad \frac{\partial W_1}{\partial x} = P(x,y,z).$$

Show that the work $W_2(x,y,z)$ done by the force in moving the particle along line segments from (x_0,y_0,z_0) to (x,y_0,z_0) and then to (x,y_0,z) and then to (x,y,z) is

$$(4)\qquad W_2(x,y,z) = \int_{x_0}^{x} P(\alpha,y_0,z_0)\, d\alpha + \int_{z_0}^{z} R(x,y_0,\gamma)\, d\gamma + \int_{y_0}^{y} Q(x,\beta,z)\, d\beta$$

and that

$$(5)\qquad \frac{\partial W_2}{\partial y} = Q(x,y,z).$$

Show that the work $W_3(x,y,z)$ done by the force in moving the particle along line segments from (x_0,y_0,z_0) to (x,y_0,z_0) and then to (x,y,z_0) and then to (x,y,z) is

$$(6)\quad W_3(x,y,z) = \int_{x_0}^{x} P(\alpha,y_0,z_0)\,d\alpha + \int_{y_0}^{y} Q(x,\beta,z_0)\,d\beta + \int_{z_0}^{z} R(x,y,\gamma)\,d\gamma$$

and that

$$(7)\qquad \frac{\partial W_3}{\partial z} = R(x,y,z).$$

Remark: A force field is called *conservative* if the work done in moving a particle around a closed curve is zero or (what amounts to the same thing) if the work done in moving the particle from one point to another is the same for all paths running from the first point to the second. If $\mathbf{F}$ is conservative, the functions in (2), (4), and (6) are equal and we may set

$$(8)\qquad W(x,y,z) = W_1(x,y,z) = W_2(x,y,z) = W_3(x,y,z).$$

Then (3), (5), and (7) give

$$(9)\qquad P = \frac{\partial W}{\partial x},\qquad Q = \frac{\partial W}{\partial y},\qquad R = \frac{\partial W}{\partial z}$$

so

$$(10)\qquad \mathbf{F} = \frac{\partial W}{\partial x}\mathbf{i} + \frac{\partial W}{\partial y}\mathbf{j} + \frac{\partial W}{\partial z}\mathbf{k} = \nabla W$$

and $\mathbf{F}$ is the gradient of W. If in addition, the particle being moved is a unit mass in a gravitational field or a unit charge in an electric field, then W is called a *potential function*, the potential at the point (x,y,z) being $W(x,y,z)$ or $W(x,y,z) + C$, where C is a constant. It is sometimes important to know that if P, Q, R are the scalar components of a conservative vector function and if they have continuous partial derivatives of first order, then (9) implies that

$$(11)\qquad \frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y},\qquad \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z},\qquad \frac{\partial R}{\partial y} = \frac{\partial Q}{\partial z}.$$

More information about this matter appears in Problem 10 of Section 13.3.

16 Assuming that all of the integrals exist, tell why

$$\int_a^b dx \int_{g_1(x)}^{g_2(x)} dy \int_{f_1(x,y)}^{f_2(x,y)} F_1(x)F_2(x,y)F_3(x,y,z)\,dz = \int_a^b F_1(x)\,dx \int_{g_1(x)}^{g_2(x)} F_2(x,y)\,dy \int_{f_1(x,y)}^{f_2(x,y)} F_3(x,y,z)\,dz.$$

17 Calculate

$$\frac{\partial^2}{\partial x\,\partial y}\int_a^x ds \int_b^y u(s,t)\,dt$$

for the case in which

(*a*) $p \geqq 0$, $q \geqq 0$, $u(s,t) = s^p t^q$
(*b*) $u(s,t) = s + t$
(*c*) $u(s,t) = e^s \sin t$

Ans.: In each case the answer is $u(x,y)$, and it may be worthwhile to try to understand why this should be so.

18 Try to understand the formulas

$$\int_b^y u_y(x,t)\,dt = u(x,t)\Big]_{t=b}^{t=y} = u(x,y) - u(x,b)$$

and

$$\begin{aligned}\int_a^x ds \int_b^y u_{x,y}(s,t)\,dt &= \int_a^x ds\left[u_x(s,t)\right]_{t=b}^{t=y}\\ &= \int_a^x \left[u_x(s,y) - u_x(s,b)\right]ds = \left[u(s,y) - u(s,b)\right]_{s=a}^{s=x}\\ &= u(x,y) - u(a,y) - u(x,b) + u(a,b).\end{aligned}$$

13.2 *Iterated integrals and volumes*

In the study of iterated and multiple integrals and their applications, we continually need uninherited skills and information that can be efficiently acquired by making a calm and thorough examination of matters relating to Figure 13.21.

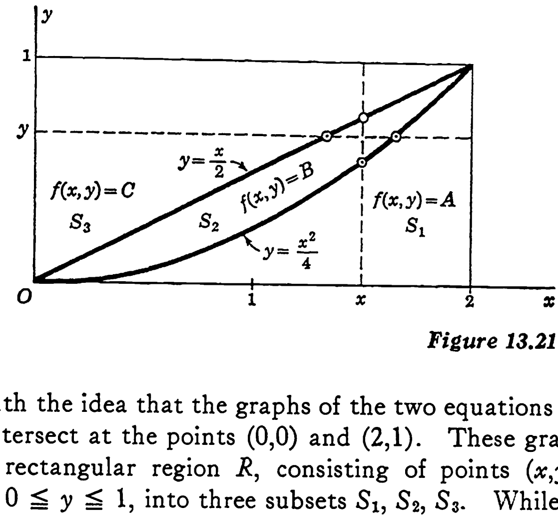


Figure 13.21

We start with the idea that the graphs of the two equations $y = x/2$ and $y = x^2/4$ intersect at the points (0,0) and (2,1). These graphs separate the closed rectangular region R, consisting of points (x,y) for which $0 \leqq x \leqq 2$, $0 \leqq y \leqq 1$, into three subsets S_1, S_2, S_3. While it makes no difference how disputes over ownership of boundaries are resolved, we want them resolved in some way and we suppose that S_2 contains each of its boundary points. Thus S_2 is the set of points (x,y) for which $0 \leqq x \leqq 2$ and $x^2/4 \leqq y \leqq x/2$. Then S_1 is the set of points (x,y) for which $0 < x \leqq 2$ and $0 \leqq y < x^2/4$, and S_3 is the set of points (x,y) for which $0 \leqq x < 2$ and $x/2 < y \leqq 1$.

While more recondite modifications of the construction are easily made, we keep our example simple by supposing that A, B, C are three constants, that $f(x,y) = A$ when (x,y) is a point of S_1, that $f(x,y) = B$ when (x,y) is a point of S_2, and that $f(x,y) = C$ when (x,y) is a point of S_3. Subject to existence of the integrals involved, we investigate the number I defined by

$$I = \int_0^2 dx \int_0^1 f(x,y)\, dy. \tag{13.22}$$

Our first step is to recognize that, for each fixed x in the interval $0 \leqq x \leqq 2$, the integrand in the integral

$$\int_0^1 f(x,y)\, dy \tag{13.221}$$

has values that depend upon y. To make effective use of our information, we mark a point x on the x axis between 0 and 2 and then draw a line through this point parallel to the y axis. A part of this line is in S_1 where $f(x,y) = A$, another part is in S_2 where $f(x,y = B$, a third part is in S_3 where $f(x,y) = C$, and, moreover, the end points of these parts depend upon the fixed x. We must understand our situation so thoroughly that we see that

$$\begin{aligned} f(x,y) &= A \text{ when } 0 \leqq y < x^2/4, \\ f(x,y) &= B \text{ when } x^2/4 \leqq y \leqq x/2, \\ f(x,y) &= C \text{ when } x/2 < y \leqq 1, \end{aligned}$$

and that

$$\int_0^1 f(x,y)\, dy = \int_0^{x^2/4} A\, dy + \int_{x^2/4}^{x/2} B\, dy + \int_{x/2}^1 C\, dy \tag{13.222}$$

so

$$\int_0^1 f(x,y)\, dy = A\frac{x^2}{4} + B\left(\frac{x}{2} - \frac{x^2}{4}\right) + C\left(1 - \frac{x}{2}\right). \tag{13.223}$$

Substituting this in (13.22) gives

$$I = \int_0^2 \left[A\frac{x^2}{4} + B\left(\frac{x}{2} - \frac{x^2}{4}\right) + C\left(1 - \frac{x}{2}\right)\right] dx \tag{13.224}$$

and hence

$$I = \tfrac{2}{3}A + \tfrac{1}{3}B + C. \tag{13.225}$$

Supposing that A, B, C are nonnegative constants, we proceed to show that the number I is the volume of the solid block H which stands upon the rectangular base R of Figure 13.21 and has, at each point (x,y) of R, height $f(x,y)$. It is quite possible to imagine that the rectangular set R of Figure 13.21 lies in a horizontal plane beneath our eyes and that the block H has its base on R and extends upward toward our eyes. If the operation is helpful, we should imagine that we are in an airplane and are looking down upon a hotel built upon the set or site R;

the parts covering S_1, S_2, S_3 have heights A, B, C. We should know that we can undertake to start with an x, y, z coordinate system oriented in the usual way and to sketch a figure like Figure 13.23 showing the block (or hotel) H.

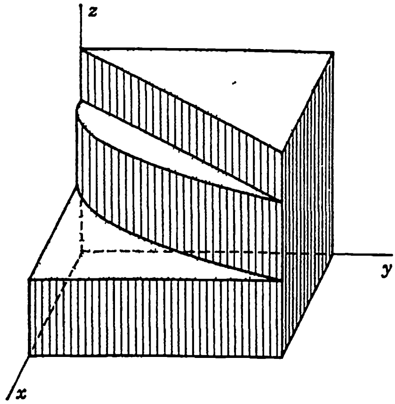


Figure 13.23

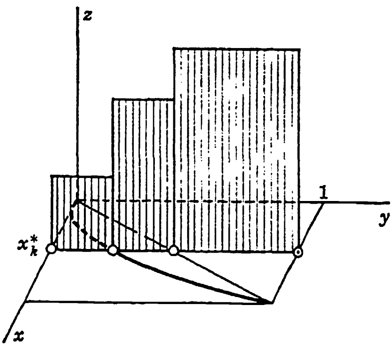


Figure 13.24

To find the volume of the block H, we do not need a figure in which an architect could take pride. It is sufficient to use the slab method which was candidly presented and employed in Section 4.5. We make a partition P of the interval $0 \leqq x \leqq 2$ into subintervals of lengths Δx_1, Δx_2, $\cdots$, Δx_n and let x_k^* be a point in the kth subinterval so that $x_{k-1} \leqq x_k^* \leqq x_k$. For each k, the number

(13.241) $$\int_0^1 f(x_k^*,y)\,dy$$

is the area of the intersection of the plane $x = x_k^*$ and the body H. Depending upon the choice of x_k^*, the number

(13.242) $$\Delta x_k \int_0^1 f(x_k^*,y)\,dy$$

is exactly equal to or is an approximation to the volume of the slab of H between the planes $x = x_{k-1}$ and $x = x_k$. The sum in the formula

(13.243) $$V = \lim \sum_{k=1}^{n} \Delta x_k \int_0^1 f(x_k^*,y)\,dy$$

is then exactly equal to the volume V of H or is an approximation to V, and in any case the limit as $|P| \to 0$ is V. Thus

(13.244) $$V = I = \int_0^2 dx \int_0^1 f(x,y)\,dy$$

and we have shown that, in the special cases being considered, our iterated integral is the volume of the block H.

Our work involving Figure 13.21 is only half finished because we must investigate the integral J defined by

$$J = \int_0^1 dy \int_0^2 f(x,y)\,dx \tag{13.25}$$

and must assimilate some new ideas. For each y in the interval $0 \leqq y \leqq 1$, the integrand in the integral

$$\int_0^2 f(x,y)\,dx \tag{13.251}$$

has values that depend upon x. As in Figure 13.21 we mark a point y on the y axis between 0 and 1 and draw a line through this point parallel to the x axis. A part of this line is in S_3 where $f(x,y) = C$, another part is in S_2 where $f(x,y) = B$, a third part is in S_1 where $f(x,y) = A$, and, moreover, the end points of these parts depend upon the fixed y. We must examine Figure 13.21 carefully enough to see that

$$\begin{aligned} f(x,y) &= C \text{ when } 0 \leqq x < 2y,\\ f(x,y) &= B \text{ when } 2y \leqq x \leqq 2\sqrt{y},\\ f(x,y) &= A \text{ when } 2\sqrt{y} < x \leqq 2, \end{aligned}$$

and that

$$\int_0^2 f(x,y)\,dx = \int_0^{2y} C\,dx + \int_{2y}^{2\sqrt{y}} B\,dx + \int_{2\sqrt{y}}^2 A\,dx, \tag{13.252}$$

so

$$\int_0^2 f(x,y)\,dx = 2A(1-\sqrt{y}) + 2B(\sqrt{y}-y) + 2Cy. \tag{13.253}$$

Substituting this in (13.25) gives

$$J = 2\int_0^1 [A(1-\sqrt{y}) + B(\sqrt{y}-y) + Cy]\,dy$$

and hence

$$J = \tfrac{2}{3}A + \tfrac{1}{3}B + C. \tag{13.26}$$

Comparison of (13.225) and (13.26) shows that I and J are equal. When A, B, C are nonnegative, I is the volume of the block H and it follows that J must also be this volume. We need the experience gained by using the slab method to prove that J, like I, is the volume of H so we will have another and more informative proof that $J = I$. Let P be a partition of the interval $0 \leqq y \leqq 1$ into subintervals of lengths Δy_1, Δy_2, $\cdots$, Δy_n and let y_k^* be a point in the kth subinterval such that $y_{k-1} \leqq y_k^* \leqq y_k$. While we should acquire the ability to do our chores without benefit of elegant figures, we can use Figure 13.27 to help us see what we are doing.

For each k, the number

$$(13.271) \qquad \int_0^2 f(x,y_k^*)\,dx$$

is the area of the intersection of the plane $y = y_k^*$ and the body H. Depending upon the choice of y_k^*, the number

$$(13.272) \qquad \Delta y_k \int_0^2 f(x,y_k^*)\,dx$$

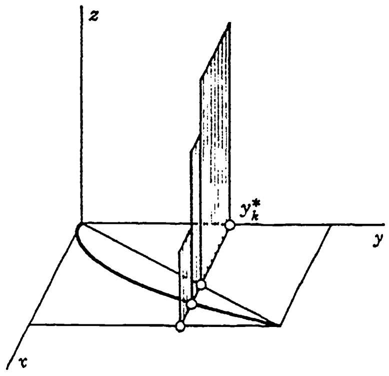


Figure 13.27

is exactly equal to or is an approximation to the volume of the slab of H between the planes $y = y_{k-1}$ and $y = y_k$. The sum in the formula

$$(13.273) \qquad V = \lim \sum_{k=1}^{n} \Delta y_k \int_0^2 f(x,y_k^*)\,dx$$

is then exactly equal to the volume V of H or is an approximation to V, and in any case the limit as $|P| \to 0$ is V. Thus

$$(13.274) \qquad V = J = \int_0^1 dy \int_0^2 f(x,y)\,dx.$$

For the special case being considered, equality of the last two members of the formula

$$(13.275) \qquad V = \int_0^2 dx \int_0^1 f(x,y)\,dy = \int_0^1 dy \int_0^2 f(x,y)\,dx$$

is a consequence of the fact that each member is equal to the volume V of a solid body H.

Some special applications of these ideas and formulas are particularly worthy of notice. In case $A = C = 0$ and $B > 0$, the number V is the volume of a body which rests upon the base S_2 and (13.275) reduces to the formula

$$(13.276) \qquad V = \int_0^2 dx \int_{x^2/4}^{x/2} f(x,y)\,dy = \int_0^1 dy \int_{2y}^{2\sqrt{y}} f(x,y)\,dx.$$

In case $A = B = C = 1$, the solid H has unit height over the whole rectangular set R and the volume of H must be the same as the area $|R|$ of R. We could therefore be sure that there would be a mistake in our work if it were not true that the numbers I and J in (13.225) and (13.26) reduce to 2 when $A = B = C = 1$. In case $A = C = 0$ and $B = 1$, the numbers V, I, and J reduce to the area $|S_2|$ of the set S_2.

Among other things, the above example leads us to the following idea. Suppose we want to find the volume V of a solid body K which rests upon a set S which is a subset of the rectangular set R containing points (x,y) for which $a \leqq x \leqq b$, $c \leqq y \leqq d$, the height of K at each point (x,y) in S being $f(x,y)$. We can undertake to solve the problem in the following way. Let f be extended in such a way that $f(x,y) = 0$ when (x,y) is a point not in the set S. Then, unless the set S and the function f are much more tortuous than those appearing in elementary nonpathological problems, two applications of the slab method for finding the volume of V yield the formula

$$V = \int_a^b dx \int_c^d f(x,y)\,dy = \int_c^d dy \int_a^b f(x,y)\,dx. \tag{13.28}$$

It is very important to be aware that, in all ordinary and many extraordinary circumstances, the last two members of (13.28) are equal even when f is a discontinuous function for which $f(x,y) = 0$ when the point (x,y) is not in a particular set S in which we are interested. More information about this matter will appear later.

Problems 13.29

1 Let

$$I = \int_{-40}^{40} dx \int_{-40}^{40} f(x,y)\,dy, \qquad J = \int_{-40}^{40} dy \int_{-40}^{40} f(x,y)\,dx,$$

where f is continuous over the region R bounded by the graphs of the lines having the equations $y = -1$, $y = x$, and $x = 1$ and $f(x,y) = 0$ when the point (x,y) is not in R. Show that

$$I = \int_{-1}^{1} dx \int_{-1}^{x} f(x,y)\,dy, \qquad J = \int_{-1}^{1} dy \int_{y}^{1} f(x,y)\,dx.$$

Evaluate I and J and show that they are equal in case

(*a*) $f(x,y) = 1$ (*b*) $f(x,y) = x$ (*c*) $f(x,y) = y$
(*d*) $f(x,y) = x + y$ (*e*) $f(x,y) = xy$ (*f*) $f(x,y) = x^2 + y^2$

when (x,y) is in R and $f(x,y) = 0$ when (x,y) is not in R.

2 For each of the formulas

(*a*) $\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 dx \int_0^{x^2} f(x,y)\,dy$

(*b*) $\int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 dy \int_0^{\sqrt{y}} f(x,y)\,dx$

(*c*) $\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^4 dx \int_x^{2x} f(x,y)\,dy$

(*d*) $\int_0^a dx \int_0^a f(x,y)\,dy = \int_0^a dx \int_0^{\sqrt{a^2-x^2}} f(x,y)\,dy$

find a region R such that the formula is valid whenever $f(x,y)$ is continuous over R and $f(x,y) = 0$ when the point (x,y) is not in R.

3 Find a region R such that the formula

$$\int_0^1 dx \int_x^{2x} f(x,y)\, dy = \int_0^1 dy \int_{y/2}^{y} f(x,y)\, dx + \int_1^2 dy \int_{y/2}^{1} f(x,y)\, dx$$

is valid when f is continuous over R. Evaluate all of the integrals and make the results agree when

(*a*) $f(x,y) = 1$ (*b*) $f(x,y) = x$ (*c*) $f(x,y) = y$
(*d*) $f(x,y) = x + y$ (*e*) $f(x,y) = xy$ (*f*) $f(x,y) = x^2 + y^2$

4 A particular solid body K can be described as the set in E_3 which rests upon the base S in the xy plane bounded by the plane graphs of the equations $y = x/2$ and $y = x^2/4$ and has, at each point (x,y) in S, height $x^2 + y^2$. The same body K can be described as the set in E_3 which is bounded by the graphs (they are all surfaces) in E_3 of the equations

$$y = x/2, \qquad y = x^2/4, \qquad z = 0, \qquad z = x^2 + y^2.$$

This book tries to be too honest to pretend that it is easy to sketch a good figure showing the body K. The book does insist, however, that we should have picked up ideas enough to enable us to use iterated integrals in two different ways to find the volume $|K|$ of K. Do it. *Remark:* The answers should agree with each other. Moreover, since the area of the base is $\frac{1}{3}$ and the height varies from 0 to 5, the answers should be between 0 and $\frac{5}{3}$.

5 As in Figure 13.291, let S be the closed set of points in the rectangle $a \leqq x \leqq b$, $c \leqq y \leqq d$ which is bounded below and above by the graphs of $y = f_1(x)$ and $y = f_2(x)$ and which is bounded on the left and right by graphs of $x = g_1(y)$ and $x = g_2(y)$ Let F be a function which is continuous over S and is such that $F(x,y) = 0$ when (x,y) is a point not in S Tell why the first integral in the formula

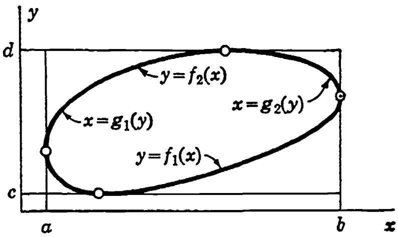


Figure 13.291

(1) $$\int_c^d F(x,y)\, dy = \int_{f_1(x)}^{f_2(x)} F(x,y)\, dy$$

exists and is equal to the second one when $a \leqq x \leqq b$. Tell why the first integral in the formula

(2) $$\int_a^b F(x,y)\, dx = \int_{g_1(y)}^{g_2(y)} F(x,y)\, dx$$

exists and is equal to the second one when $c \leqq y \leqq d$ Show that the formula

(3) $$\int_a^b dx \int_c^d F(x,y)\, dy = \int_c^d dy \int_a^b F(x,y)\, dx$$

reduces to

$$(4) \qquad \int_a^b dx \int_{f_1(x)}^{f_2(x)} F(x,y)\,dy = \int_c^d dy \int_{g_1(y)}^{g_2(y)} F(x,y)\,dx.$$

Finally, give geometric interpretations of the numbers in (4) when, for each point (x,y) in S,

(*a*) $F(x,y) = 1$
(*b*) $F(x,y) = x^2 + y^2$

6 By use of iterated integrals, find the volume V of the solid bounded by the three coordinate planes and the graph of the equation

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1.$$

Remark and ans.: The solid is a pyramid, and use of the fundamental fact that the volume of a pyramid is one-third of the product of its height and the area of its base gives $V = \frac{1}{6}abc$

7 Even when details are efficiently managed, it is not a short task to find the volume V of the solid in the first octant bounded by the three coordinate planes and the graph of the equation

$$\left(\frac{x}{a}\right)^{1/2} + \left(\frac{y}{b}\right)^{1/2} + \left(\frac{z}{c}\right)^{1/2} = 1.$$

It is, however, worthwhile to try to manage details efficiently and to earn the satisfaction involved in showing that $V = abc/90$. It can be added that the world is wide enough to accommodate and even need persons who run amok or amuck and become strong enough to solve the problems obtained by replacing the exponent $\frac{1}{2}$ by $\frac{1}{3}$ and $\frac{1}{4}$ and $\frac{1}{5}$.

8 Prove that if the formula

$$(1) \qquad \int_a^b dx \int_c^d [f(x,y) + B]\,dy = \int_c^d dy \int_a^b [f(x,y) + B]\,dx$$

is valid for some constant B, then

$$(2) \qquad \int_a^b dx \int_c^d f(x,y)\,dy = \int_c^d dy \int_a^b f(x,y)\,dx.$$

Remark: In case f is a bounded function for which $f(x,y)$ is sometimes positive and sometimes negative, we can choose B such that $f(x,y) + B$ is always positive. In appropriate circumstances, we can prove (2) by proving the first formula which involves only positive integrands.

9 Let a and x be confined to an interval over which a given function f is continuous. Let $f_0(x) = f(x)$ and let

$$(1) \qquad f_1(x) = \int_a^x f(t)\,dt, \qquad f_2(x) = \int_a^x f_1(t)\,dt, \qquad f_3(x) = \int_a^x f_2(t)\,dt$$

and so on so that, for each $n = 1, 2, 3, \cdots$,

$$(2) \qquad f_{n+1}(x) = \int_a^x f_n(t)\,dt.$$

Since $0! = 1$, the formula

$$(3) \qquad f_n(x) = \int_a^x \frac{(x-t)^{n-1}}{(n-1)!} f(t)\, dt$$

is certainly correct when $n = 1$. Assuming that (3) is correct for a particular positive integer n, show that

$$(4) \qquad f_n(u) = \int_a^u \frac{(u-t)^{n-1}}{(n-1)!} f(t)\, dt$$

and use (2) to obtain

$$(5) \qquad f_{n+1}(x) = \int_a^x du \int_a^u \frac{(u-t)^{n-1}}{(n-1)!} f(t)\, dt.$$

Use (5) and Figure 13.292 to obtain

$$(6) \qquad f_{n+1}(x) = \int_a^x dt \int_t^x \frac{(u-t)^{n-1}}{(n-1)!} f(t)\, du$$

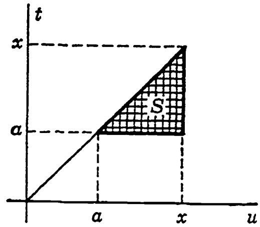


Figure 13.292

and then use (6) to obtain the result of replacing n by $n+1$ in (3). Since (3) is correct when $n = 1$, it must be correct when $n = 2$ and hence when $n = 3$ and hence when $n = 4$, and so on.

10 This section should not leave the impression that our ideas about Riemann integrals can always be applied to Riemann-Cauchy integrals. It can happen that

$$I = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} f(x,y)\, dy, \qquad J = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} f(x,y)\, dx$$

both exist but have unequal values. For example, let

$$\begin{aligned} f(x,y) &= 1 \text{ when } x \geqq 0,\ x-1 < y < x, \\ f(x,y) &= -1 \text{ when } x \geqq 0,\ x < y < x+1, \\ f(x,y) &= 0 \text{ otherwise.} \end{aligned}$$

Show that, in this case, $I = 0$ and $J = 1$.

13.3 Double integrals

Section 4.2 showed how we partition intervals into subintervals to form Riemann sums and how we use these sums to define Riemann integrals over one-dimensional intervals. Because the idea is important in both pure and applied mathematics, we must learn about the process by which a set S in a plane is partitioned into subsets in order to enable us to form Riemann sums and define Riemann integrals over S. To begin, let S be a set of points which may, for example, be the set of points inside and on a circle or an outer boundary curve such as that shown in Figure 13.31. The set S may be the set of points P which are neither outside the outer boundary nor inside the inner boundary of Figure 13.32. What we really require is that the set S have positive area and that the points of S lie inside or on some rectangle R so that the set S is bounded. In some applications the set S is regarded as a lamina (thin plate) or as a plane section of a three-dimensional solid in which we are interested. While we normally use coordinates (rectangular or polar, for example) to determine a point P of S, we do not at present allow any one brand of coordinates to dominate our work. We suppose that we have a bounded function f defined over S and use the symbol $f(P)$ to denote the value of f at P. For example, if S is a lamina, $f(P)$ could be the density (mass per unit area) at P or the product of the density at P and the specific heat at P and the temperature at P. If S is a lamina and we want to calculate its moment of inertia about a line L, $f(P)$ could be the product of the density at P and the square of the distance from P to L. It is often helpful to think of $|f(P)|$ as being the height at P of a solid which stands upon the base S.

The first step in our approach to a Riemann sum is to make a partition Q (the letter P has been preempted) of the set S into n subsets $S_1, S_2, \cdots, S_n$. As Figures 13.31 and 13.32 indicate, the result of partitioning

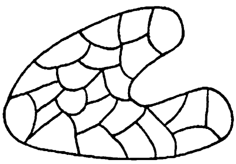


Figure 13.31

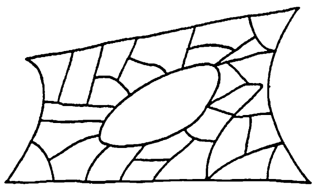


Figure 13.32

a set S in E_2 into smaller subsets is not as simple as the result of partitioning an interval in E_1 into subintervals. The only things we require of the sets $S_1, S_2, \cdots, S_n$ is that they be nonoverlapping, that their union be S, and that each of them have positive area. It turns out that the notational transition from Riemann sums to Riemann integrals will be facilitated by denoting the areas of the sets $S_1, S_2, \cdots, S_n$ by the symbols $\Delta S_1, \Delta S_2, \cdots, \Delta S_n$. The meanings of our symbols should be emphasized. For each $k = 1, 2, \cdots, n$, the symbol ΔS_k does not stand for a part of the set S; it stands for the area of a part of the set S. For each $k = 1, 2, \cdots, n$, let P_k be a point in the set S_k. The number RS (Riemann sum) defined by

$$\text{RS} = \sum_{k=1}^{n} f(P_k)\,\Delta S_k \tag{13.33}$$

is then a *Riemann sum* formed for the function f and for the partition Q of S.

In order to tell what we mean by the norm $|Q|$ of the partition Q, it is necessary to introduce a simple geometrical concept. The *diameter* of a set is the least upper bound of distances between pairs of points of a set. The *norm* $|Q|$ of the partition Q of S into subsets $S_1, S_2, \cdots, S_n$ is the greatest of the diameters of the sets $S_1, S_2, \cdots, S_n$. We are now ready to define the Riemann integral of f over S, the definition being analogous to that involving (4.23). If there is a number I such that to each $\epsilon > 0$

there corresponds a $\delta > 0$ such that

$$\left| I - \sum_{k=1}^{n} f(P_k)\, \Delta S_k \right| < \epsilon \tag{13.34}$$

whenever the sum is a Riemann sum formed for the function f and for a partition Q of S for which $|Q| < \delta$, then f is said to be *Riemann integrable* over S and I is said to be the *Riemann integral* of f over S. This integral is usually denoted by the symbol

$$\iint_S f(P)\, dS, \tag{13.35}$$

which displays the function f and the symbol S that represents the set which was partitioned to obtain the approximating Riemann sums. The integral is called a *double integral* because the set S is two-dimensional, that is, a set in E_2 having positive area. The two integral signs serve to remind us that S is two-dimensional, but sometimes one of them is omitted from the symbol. As was the case for simple (that is, one-dimensional) integrals, it is a convenience (and sometimes also a source of misunderstanding, confusion, and controversy) to drag in the notation of limits and write

$$\iint_S f(P)\, dS = \lim_{|Q|\to 0} \sum_{k=1}^{n} f(P_k)\, \Delta S_k \tag{13.36}$$

or

$$\iint_S f(P)\, dS = \lim \sum f(P)\, \Delta S. \tag{13.37}$$

When we are interested in problems in which a function f defined over a bounded set S is involved and rectangular coordinates are to be used, we can produce substantial simplifications of our work by letting R be a rectangular set which contains the set S and by extending the domain of f by putting $f(x,y) = 0$ when (x,y) is a point of R which is not in S. The following theorem then enables us to evaluate double integrals by evaluating iterated integrals.

Theorem 13.38 *If S is a subset of the rectangular region R consisting of points (x,y) for which $a \leqq x \leqq b$ and $c \leqq y \leqq d$, if $f(x,y) = 0$ when (x,y) is a point in R but not in S, and if the four integrals*

$$(13.381)\quad I_1 = \iint_S f(x,y)\, dS, \qquad I_2 = \iint_R f(x,y)\, dR,$$

$$(13.382)\quad I_3 = \int_a^b dx \int_c^d f(x,y)\, dy, \qquad I_4 = \int_c^d dy \int_a^b f(x,y)\, dx$$

all exist, then

$$I_1 = I_2 = I_3 = I_4, \tag{13.383}$$

that is, the four integrals are all equal.

Proof of this theorem lies far beyond the scope of this course. Persons who continue study of mathematical analysis until theories of Lebesgue measure and Lebesgue integrals (including a theorem known as the Fubini theorem) have been learned will find that validity of the theorem will be an easy consequence of fundamental relations between Riemann and Lebesgue integrals. For the present, we can be content with a hazy understanding of the fundamental fact that the double integrals I_1 and I_2 will exist if f is bounded and the set D of discontinuities of f has area (two-dimensional Lebesgue measure) 0 and, moreover, the iterated integrals I_3 and I_4 will also exist if it is also true that each horizontal line and each vertical line intersects D in a set having length (one-dimensional Lebesgue measure) 0. So far as elementary applications to elementary problems are concerned, we can be sure that if the set S and the function f are bounded, then the double integrals in (13.381) and the iterated integrals in (13.382) must exist and must have the same value.

Symbols used for iterated integrals were discussed in Section 13.1. In addition to the symbols used in this section for double integrals, those appearing in the formula

$$(13.384)\qquad \iint_S f(x,y)\,dS = \iint_S f(x,y)\,dx\,dy = \int_S f(x,y)\,dx\,dy$$

are sometimes used when rectangular coordinates are involved.

Problems 13.39

1 Let S be the set in E_2 bounded by the graphs of the equations

$$y = x^2, \qquad y = x + 2.$$

Supposing that f is continuous over S and that

$$J = \iint_S f(x,y)\,dS,$$

sketch a figure which displays the set S and an appropriate rectangular region R and then write complete and intelligible descriptions of the steps involved in using Theorem 13.38 to obtain the formulas

$$J = \int_{-1}^{2} dx \int_{x^2}^{x+2} f(x,y)\,dy$$

and

$$J = \int_0^1 dy \int_{-\sqrt{y}}^{\sqrt{y}} f(x,y)\,dx + \int_1^4 dy \int_{y-2}^{\sqrt{y}} f(x,y)\,dx.$$

2 Supposing that $0 < a < b$ and that f is continuous, determine a set S in the xy plane such that

$$\int_0^a dy \int_y^{y+b} f(x,y)\,dx = \iint_S f(x,y)\,dS.$$

Then show that

$$\iint_S f(x,y)\,dS = \int_0^a dx \int_0^x f(x,y)\,dy + \int_a^b dx \int_0^a f(x,y)\,dy + \int_b^{a+b} dx \int_{x-b}^{a} f(x,y)\,dy.$$

3 Supposing that S is a bounded set having positive area, interpret and prove the statement

$$|S| = \iint_S 1\,dS.$$

Hint: Look at the definition of double integrals.

4 Supposing that f is continuous over S, use Theorem 13.38 and the method of Problem 1 to obtain iterated integrals equal to

$$\iint_S f(x,y)\,dS$$

when S is the set in E_2 bounded by the graphs of the equations

(*a*) $y = x^2$, $y = mx + b$ $\quad (m,b > 0)$
(*b*) $y = x$, $y = 2x$, $x = 1$
(*c*) $y = 0$, $y = x$, $y = 3x - 2$
(*d*) $y = x^3$, $y = x$
(*e*) $y = e^x$, $y = 0$, $x = 0$, $x = 1$
(*f*) $y = x$, $y = \sin x$, $x = \pi/2$

5 Observe that $\sin(x + y)$ is continuous and nonnegative over the square S in the xy plane having opposite vertices at the points $(0,0)$ and $(\pi/2, \pi/2)$. Evaluate

$$\iint_S \sin(x + y)\,dx\,dy.$$

Ans.: 2.

6 Evaluate

$$\iint_Q \frac{1}{1 + x + y}\,dx\,dy$$

when Q is the square having opposite vertices at the points $(0,0)$ and $(1,1)$. *Ans.:* $3 \log 3 - 4 \log 2$.

7 Supposing that $0 < a < b$ and $0 < p < q$, evaluate

$$\iint_S e^{y/x}\,dS$$

when S is the region bounded by the lines having the equations $x = a$, $x = b$, $y = px$, and $y = qx$. *Ans.:* $\frac{1}{2}(e^q - e^p)(b^2 - a^2)$.

8 Supposing that S is the square having opposite vertices at the points $(0,0)$ and $(1,1)$, show that

$$\text{(1)} \qquad \iint_S \frac{1}{1 - xy}\,dx\,dy = \int_0^1 dx \int_0^1 \frac{1}{1 - xy}\,dy = \int_0^1 \frac{-\log(1 - x)}{x}\,dx$$

provided the integrals exist. With the aid of series and termwise integration (which can be justified), continue the work to obtain

$$\iint_S \frac{1}{1-xy}\,dx\,dy = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}. \tag{2}$$

Remark: Persons who study more advanced mathematical analysis may encounter the following elegant theorem. If S is a set in E_n, if S_1 is a subset of S having measure 0, if

$$f(P) = u_1(P) + u_2(P) + u_3(P) + \cdots \tag{3}$$

when P is in S but not in S_1, and if $u_k(P) \geqq 0$ for each k when P is in S, then the formula

$$\int_S f(P)\,dS = \int_S u_1(P)\,dS + \int_S u_2(P)\,dS + \int_S u_3(P)\,dS + \cdots \tag{4}$$

is valid provided the integrals exist as Riemann integrals or Riemann-Cauchy integrals or Lebesgue integrals.

9 Obtain the result (2) of the preceding problem by termwise integration (which can be justified) of the series in

$$\frac{1}{1-xy} = 1 + xy + x^2y^2 + x^3y^3 + \cdots.$$

10 This long problem is for persons who wish to become good mathematicians or physicists or engineers and who realize that the best ones start learning about important things while they are young and then continue to learn more. Make a large copy of Figure 13.291 showing the set S which is the set of points (x,y) for which $a \leqq x \leqq b$ and $f_1(x) \leqq y \leqq f_2(x)$ and is also the set of points (x,y) for which $c \leqq y \leqq d$ and $g_1(y) \leqq x \leqq g_2(y)$. Let C be the curve consisting of the boundary of S traversed once in the positive direction. Let F and G be continuous over S and let

$$I = \iint_S F(x,y)\,dS, \qquad J = \iint_S G(x,y)\,dS. \tag{1}$$

Show that

$$I = \int_c^d dy \int_{g_1(y)}^{g_2(y)} F(x,y)\,dx, \qquad J = \int_a^b dx \int_{f_1(x)}^{f_2(x)} G(x,y)\,dy. \tag{2}$$

We enter a gate to an important scientific garden when we suppose that P and Q are functions which are defined and have continuous first-order partial derivatives over S and put

$$F(x,y) = \frac{\partial Q}{\partial x} = Q_x(x,y), \qquad G(x,y) = \frac{\partial P}{\partial y} = P_y(x,y). \tag{3}$$

Show that in this case

$$\begin{aligned} I &= \int_c^d dy \int_{g_1(y)}^{g_2(y)} Q_x(x,y)\,dx = \int_c^d dy\Big[Q(x,y)\Big]_{x=g_1(y)}^{x=g_2(y)} \\ &= \int_c^d Q(g_2(y), y)\,dy - \int_c^d Q(g_1(y), y)\,dy. \end{aligned} \tag{4}$$

Be sure to realize that this is correct because, for each fixed y, $Q(x,y)$ is a function of x whose derivative with respect to x is the integrand $Q_x(x,y)$. We must now look at our figure. As y increases from c to d, the point with coordinates $(g_2(y), y)$ traverses, in the positive direction, the part C_1 of C which lies to the right of S. It is a consequence of the definition of curve integrals given in Problem 15 of Section 7.2 that

$$\int_c^d Q(g_2(y), y)\, dy = \int_{C_1} Q(x,y)\, dy, \tag{5}$$

the parameter now being y instead of t. The last term in (4) is more troublesome. As y increases from c to d, the point with coordinates $(g_1(y), y)$ traverses, in the *negative* direction, the part C_2 of C which lies to the left of S. This reversal of direction introduces a change in sign so that

$$-\int_c^d Q(g_1(y), y)\, dy = \int_{C_2} Q(x,y)\, dy. \tag{6}$$

Show that combining these results gives the formula

$$\iint_S \frac{\partial Q}{\partial x}\, dS = \int_C Q\, dy. \tag{7}$$

Tell why

$$\begin{aligned}(8)\quad J = \int_a^b dx \int_{f_1(x)}^{f_2(x)} P_y(x,y)\, dy &= \int_a^b dx \Big[P(x,y)\Big]_{y=f_1(x)}^{y=f_2(x)} \\ &= \int_a^b P(x, f_2(x))\, dx - \int_a^b P(x, f_1(x))\, dx\end{aligned}$$

and

$$-\iint_S \frac{\partial P}{\partial y}\, dS = \int_C P\, dy.$$

Note that combining (7) and (8) gives the formula

$$\iint_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dS = \int_C (P\, dx + Q\, dy). \tag{10}$$

This formula appeared in the works of George Green (1793-1841), a pioneer in applied mathematics who originated the term *potential function*, and the formula is called a (or, sometimes, *the*), *Green formula.*

The Green formula and its extensions have very important applications, some of which involve vectors. To be broad-minded about the matter, let

$$\mathbf{V}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}, \tag{11}$$

where $\mathbf{V}$ is a vector function having scalar components P, Q, R that are continuous and have continuous partial derivatives over a part of E_3 in which our sets and curves are supposed to lie. The *curl* of the vector function V is a vector function which is written $\nabla \times \mathbf{V}$, which is read "the curl of $\mathbf{V}$" or "del cross $\mathbf{V}$" and is defined by the formula

$$\nabla \times \mathbf{V} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} \tag{12}$$

in which the right member is a determinant. The expanded form of (12) is

$$\text{(13)} \qquad \nabla \times \mathbf{V} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}.$$

For the special case to which (10) applies, S is a patch of surface in the xy plane and $\mathbf{k}$ is the unit normal to S which lies in the direction of the thumb on a right hand when the fingers point in the direction in which C is oriented. If we denote this unit normal by $\mathbf{n}$, then in the special case (13) reduces to

$$\text{(14)} \qquad (\nabla \times \mathbf{V})\cdot\mathbf{n} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}.$$

Therefore, the left member of (10) is, in the special case, the left member of the formula

$$\text{(15)} \qquad \iint_S (\nabla \times \mathbf{V})\cdot\mathbf{n}\, dS = \int_C \mathbf{V}\cdot d\mathbf{r}.$$

Our next step is to show that the right of (10) is, in the special case, the right member of (15). This is quite easy. Let functions $x(t)$, $y(t)$, $z(t)$ be such that the point $P(t)$ having coordinates $x(t)$, $y(t)$, $z(t)$ traverses C once in the positive direction as t increases from t_1 to t_2 and let

$$\text{(16)} \qquad \mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}.$$

For the case in which these functions have piecewise continuous derivatives, differentiation gives

$$\text{(17)} \qquad r'(t) = x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k}.$$

Then

$$\text{(18)} \qquad \mathbf{V}\cdot\mathbf{r}'(t) = Px'(t) + Qy'(t) + Rz'(t),$$

where $\mathbf{V}$ stands for $\mathbf{V}(x(t), y(t), z(t))$, P stands for $P(x(t), y(t), z(t))$, and so on. Hence

$$\text{(19)} \qquad \int_{t_1}^{t_2} \mathbf{V}\cdot\mathbf{r}'(t)\, dt = \int_{t_1}^{t_2} [Px'(t) + Qy'(t) + Rz'(t)]\, dt$$

and these integrals are, by definitions of curve integrals, respectively equal to those in the formula

$$\text{(20)} \qquad \int_C \mathbf{V}\cdot d\mathbf{r} = \int_C (P\, dx + Q\, dy + R\, dz).$$

In the special cases where C lies in the xy plane, we have $z(t) = 0$ for each t, so the right member of (10) is equal to the right members of (19) and (20) and hence is equal to the left member of (20), which is the right member of (15).

The formula (15), that is,

$$\text{(21)} \qquad \iint_S (\nabla \times \mathbf{V})\cdot\mathbf{n}\, dS = \int_C \mathbf{V}\cdot d\mathbf{r},$$

which reduces to the Green formula (10) when S and C lie in the xy plane, is known as the *Stokes formula*. So far, (21) has been proved only when S is a set in the xy plane which satisfies the heavy restriction given at the beginning of this problem. One example is sufficient to expose an interesting idea that enables us to establish (21) for less simple sets S in the xy plane. To establish (21) for the set S of Figure 13.391, we split S into two subsets S_1 and S_2 as in

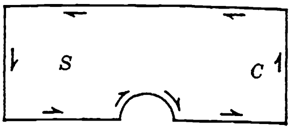


Figure 13.391

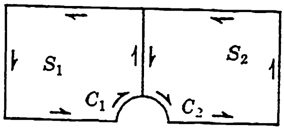


Figure 13.392

Figure 13.392. Writing the formulas obtained by applying (21) to the simpler sets S_1 and S_2 and adding the results shows that (21) is valid for S because the curve integrals over the common boundary of S_1 and S_2 come with opposite signs and cancel out of the sum. After having proved that (21) is valid for sets of particular types that lie in the xy plane, the next step is to recognize that, when C is a suitable plane curve, the right side of (21) has an intrinsic meaning which is independent of coordinate systems. As can be suspected, this fact can be used to show that the left member of (21) and the curl itself also have intrinsic meanings. To be appropriately narrow-minded about this matter, let S be a plane triangular set or plane circular disk in E_3 which is bounded by an oriented triangle or oriented circle C. Then (21) is valid because the simpler Green formula shows that it is valid when the coordinate system is chosen such that S lies in the xy plane. The method that was applied to the plane sets of Figures 13.391 and 13.392 can now be employed to prove that (21) holds when S is a triangulated oriented surface in E_3 consisting of a finite set of plane triangular faces bounded by oriented triangles provided the topological structure and orientations are such that if a side of a triangle is a part of the boundaries of more than one triangular face, then the side is a part of boundaries of exactly two such faces and, as in Figure 13.393, the side has opposite orientations in the two triangles that contain it. While a full treatment of the matter lies beyond the scope of this book, teachers of courses in electricity and magnetism and aerodynamics (among others) require knowledge of consequences of the idea that the Stokes formula (21) is valid when S is a patch of surface in E_3 that can be satisfactorily approximated by a triangulated oriented surface of the type described above.

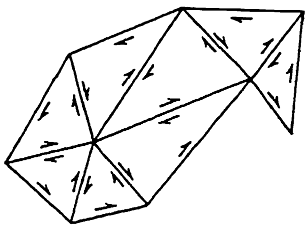


Figure 13.393

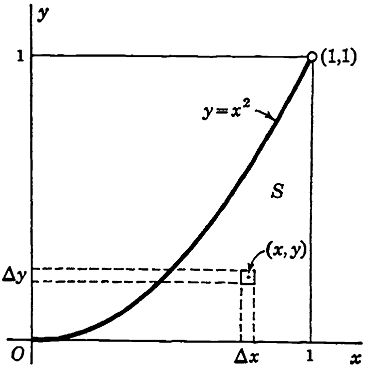
One of the important applications of the Green and Stokes formulas involves conservative force fields. Let the vector $\mathbf{V}(x,y,z)$ in (11) be the force on a particle when the particle is at the point (x,y,z). The *force field* determined by the vector function $\mathbf{V}$ is said to be *conservative* if the work done in moving the particle around a closed curve C is 0 when C belongs to a class of curves which is not always carefully defined but which certainly includes circles. If $\mathbf{V}$ is conservative over a region in E_3, then the right member of (21) must therefore be 0 when C is a circle in the region. Then the left member of (21) must be 0 when S is a plane circular disk in the region and therefore (as can be proved) the hypothesis that $\nabla \times \mathbf{V}$ is continuous over the region implies that $\nabla \times \mathbf{V} = 0$ over the region. On the other hand, if $\nabla \times \mathbf{V} = 0$, then the left member of (21) is clearly 0, so the right member must be 0 and $\mathbf{V}$ must be conservative This proves the very useful nontrivial fact that if $\mathbf{V}$ and its scalar components P, Q, R are continuous and have continuous partial derivatives over a region in E_3, then $\mathbf{V}$ is conservative over the region if and only if the formulas

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}, \qquad \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}, \qquad \frac{\partial R}{\partial y} = \frac{\partial Q}{\partial z}$$

hold over the region.

11 There is much to be learned about the process of Problem 10 by which the Stokes formula (21) is proved first for simpler surfaces composed of oriented plane triangles suitably joined at their edges and then for curved surfaces that can be satisfactorily approximated by the simpler surfaces Relatively few people undertake to master all of the details, but everybody can see that some quite delicate topological considerations are involved. Classical examples involve ordinary bands and Mobius bands. When the ends of a strip of paper a foot long and an inch wide are joined in the simplest way, the result is an ordinary curved band that has two edges (a top and a bottom) and two sides (an inside and an outside). It is easy to put a dozen diagonal creases in the paper to obtain a band composed of a dozen plane triangular patches. Let S_1 be the surface composed of the points on the outside of the latter band. It is easy to orient the triangles as in the discussion of Figure 13.393 and to obtain the Stokes formula for S_1. To make a Mobius band, we start with another strip of paper a foot long and an inch wide, but this time we put a twist (a half-turn) in one of the ends before the two ends are joined. This strip can be creased to obtain a band composed of plane triangles joined at their edges. It turns out that the Mobius band has just one edge and just one side, there being no "side" that is "an outside" that is different from "the inside." Inner secrets are revealed to those who try to color only "the outside" of this band. Persons interested in this matter may construct Möbius bands and study their properties. It is quite easy to obtain the correct idea that topological considerations form an essential part of rigorous (free from blunders) statements and proofs of theorems setting forth conditions under which the Stokes formula is valid.

Figure 13.41



13.4 Rectangular coordinate applications of double and iterated integrals This section illustrates ideas that are often used when problems are being solved with the aid of double and iterated integrals involving rectangular coordinates. The principal illustration involves a lamina (or flat plate) which, as in Figure 13.41, lies in the xy plane and is bounded by the graphs of the equations $y = 0$, $y = x^2$, and $x = 1$. We suppose that, at each point (x,y) of the lamina, the lamina has areal density (or mass per unit area) $\delta(x,y)$. This means that to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that

$$\left| \delta(x,y) - \frac{\Delta m}{\Delta S} \right| < \epsilon \tag{13.42}$$

whenever ΔS is the area of a part of the lamina containing the point (x,y) and having diameter less than δ and Δm is the mass of the part. In the simplest applications, there is a constant k, which may be 1, such that $\delta(x,y) = k$ whenever (x,y) is a point of the set S occupied by the lamina; in this case the lamina is said to be *homogeneous*. While the ideas can be applied in some other cases, we suppose that δ is continuous. Supposing that x_0 is a given number and that p is a given nonnegative integer that is 0 or 1 or 2 in most applications, we undertake to learn the techniques involved in setting up three different expressions for $M^{(p)}_{x=x_0}$, the pth moment of the lamina about the line $x = x_0$.

To set up a double integral for $M^{(p)}_{x=x_0}$, we chop the rectangle of Figure 13.41 into subrectangles by lines parallel to the coordinate axes. A particular subrectangle, such as the one shown in Figure 13.41, has area $\Delta x\, \Delta y$. Supposing that the subrectangle lies entirely within the lamina, we select a point (x,y) in the subrectangle and use the number $\delta(x,y)\, \Delta x\, \Delta y$ as an approximation to the mass of the part of the lamina within the subrectangle. If this total mass were concentrated at the point (x,y), its pth moment about the line $x = x_0$ would be

$$(x - x_0)^p \delta(x,y)\, \Delta x\, \Delta y. \tag{13.43}$$

We therefore use this number as an approximation to the pth moment about the line $x = x_0$ of the part of the lamina in the one subrectangle. The sum

$$\Sigma (x - x_0)^p \delta(x,y)\, \Delta x\, \Delta y, \tag{13.431}$$

which contains a term for each subrectangle in the lamina, should then be a good approximation to the total pth moment of the entire lamina whenever the diameters of the subrectangles are all small. This leads us to the formula

$$M^{(p)}_{x=x_0} = \lim \Sigma (x - x_0)^p \delta(x,y)\, \Delta x\, \Delta y, \tag{13.432}$$

the right side of which is taken to be the definition of the number $M^{(p)}_{x=x_0}$ which we are seeking. In accordance with the theory of double integrals involving (13.34) and Theorem 13.38, the right side of (13.432) is a double integral which we can denote by one or the other of the symbols in the formula

$$M^{(p)}_{x=x_0} = \iint_S (x - x_0)^p \delta(x,y)\, dS = \iint_S (x - x_0)^p \delta(x,y)\, dx\, dy. \tag{13.433}$$

With the aid of Theorem 13.38 we can quickly express the above double integral as an iterated integral in two different ways. It is, however, worthwhile to learn to use a procedure which leads directly to iterated integrals without making use of double integrals. As in the preceding paragraph, we observe that $\Delta x\ \Delta y$ is the area of a subrectangle and use the number $\delta(x,y)\ \Delta x\ \Delta y$ as an approximation to the mass of the part of the lamina within the subrectangle. Again we note that if this total mass were concentrated at the point (x,y), its pth moment about the line $x = x_0$ would be

$$(x - x_0)^p\delta(x,y)\ \Delta x\ \Delta y. \tag{13.44}$$

We then form the sum

$$\Delta x \sum_{x\text{ fixed}} (x - x_0)^p\delta(x,y)\ \Delta y \tag{13.441}$$

where the part "x fixed" of the symbol serves to inform us that the sum contains only terms arising from those subrectangles which comprise a vertical strip such as that shown in Figure 13.445. When the numbers Δx and Δy are all small, the coefficient of Δx in (13.441) is a Riemann sum which is a good approximation to the coefficient of Δx in the expression

$$\Delta x \int_0^{x^2} (x - x_0)^p\delta(x,y)\ dy. \tag{13.442}$$

Using this as an approximation to the pth moment about the line $x = x_0$ of the part of the lamina in one strip, we are led to expect that the sum in

$$M_{x=x_0}^{(p)} = \lim \sum \Delta x \int_0^{x^2} (x - x_0)^p\delta(x,y)\ dy \tag{13.443}$$

will be good approximation to the required total moment when the numbers Δx are all small and hence that (13.443) should be a valid formula. This gives

$$M_{x=x_0}^{(p)} = \int_0^1 dx \int_0^{x^2} (x - x_0)^p\delta(x,y)\ dy \tag{13.444}$$

because the sum in (13.443) is a Riemann sum which approximates the integral in (13.444).

Figure 13.445

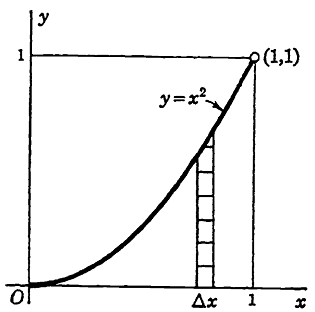
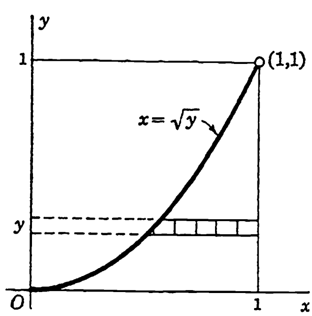


Figure 13.446



A simple modification of the preceding paragraph gives an iterated integral in which the first integration is with respect to x. Instead of (13.441), we form the sum

$$\Delta y \sum_{y \text{ fixed}} (x - x_0)^p \delta(x,y)\, \Delta x, \tag{13.45}$$

where the part "y fixed" of the symbol informs us that the sum contains only terms arising from subrectangles which comprise a horizontal strip such as that shown in Figure 13.446. When the numbers Δx and Δy are all small, (13.45) and

$$\Delta y \int_{\sqrt{y}}^{1} (x - x_0)^p \delta(x,y)\, dx \tag{13.451}$$

are good approximations to the pth moment about the line $x = x_0$ of the part of the lamina in one strip, and we are led to the formula

$$M_{x=x_0}^{(p)} = \lim \sum \Delta y \int_{\sqrt{y}}^{1} (x - x_0)^p \delta(x,y)\, dx \tag{13.452}$$

and hence to the formula

$$M_{x=x_0}^{(p)} = \int_0^1 dy \int_{\sqrt{y}}^{1} (x - x_0)^p \delta(x,y)\, dx \tag{13.453}$$

for the pth moment about the line $x = x_0$ of the whole lamina.

Several quite simple and obvious remarks can now be made. In order to obtain derivations of formulas for $M_{y=y_0}^{(p)}$, the pth moment of the lamina about the line $y = y_0$, it suffices to replace the factor $(x - x_0)^p$ by the factor $(y - y_0)^p$ in the above derivations. In case $p = 0$, the factors $(x - x_0)^p$ and $(y - y_0)^p$ are both equal to 1 and the numbers $M_{x=x_0}^{(p)}$ and $M_{y=y_0}^{(p)}$ are both equal to the mass M of the lamina. Thus

$$M = \iint_S \delta(x,y)\, dx\, dy, \tag{13.454}$$

and we can replace this double integral by iterated integrals.

In case $p = 1$, the moments are *first moments or moments of first order.* In case x_0 and y_0 are chosen such that $M_{x=x_0}^{(1)} = 0$ and $M_{y=y_0}^{(1)} = 0$, the point (x_0,y_0) is called the *centroid* of the lamina. It is customary to let $\bar{x}$ and $\bar{y}$ (x bar and y bar) denote the coordinates of the centroid. The equations which determine $\bar{x}$ and $\bar{y}$ then become $M_{x=\bar{x}}^{(1)} = 0$, $M_{y=\bar{y}}^{(1)} = 0$ or, as we see from (13.433) and the similar formula for $M_{y=y_0}^{(1)}$,

$$\iint_S (x - \bar{x})\delta(x,y)\, dx\, dy = 0, \qquad \iint_S (y - \bar{y})\delta(x,y)\, dy = 0. \tag{13.46}$$

Since $\bar{x}$ and $\bar{y}$ are constants that can be moved across integral signs, we can put these equations in the form

$$\bar{x} \iint_S \delta(x,y)\,dx\,dy = \iint_S x\delta(x,y)\,dx\,dy, \tag{13.461}$$
$$\bar{y} \iint_S \delta(x,y)\,dx\,dy = \iint_S y\delta(x,y)\,dx\,dy$$

or

$$\bar{x} = \frac{\iint_S x\delta(x,y)\,dx\,dy}{\iint_S \delta(x,y)\,dx\,dy}, \qquad \bar{y} = \frac{\iint_S y\delta(x,y)\,dx\,dy}{\iint_S \delta(x,y)\,dx\,dy}, \tag{13.462}$$

where the denominators are equal to the mass M of the lamina. It is sometimes helpful to know that if, as in Figure 13.463, the line $x = x_0$ is a line of symmetry of a homogeneous lamina, then $M^{(1)}_{x=x_0} = 0$ and hence $\bar{x} = x_0$. In order to find the first moment $M^{(1)}_{x=0}$ of the lamina about the y axis, it suffices to calculate the mass M of the lamina and

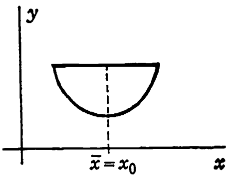


Figure 13.463

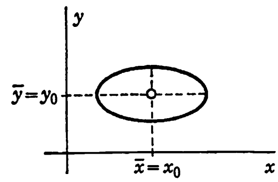


Figure 13.464

use the formula $M\bar{x} = M^{(1)}_{x=0}$. If, as in Figure 13.464, the lines $x = x_0$ and $y = y_0$ are both lines of symmetry of a homogeneous lamina, then $\bar{x} = x_0$ and $\bar{y} = y_0$, so the centroid of the lamina is the point (x_0,y_0).

In case $p = 2$, the number $M^{(p)}_{x=x_0}$ becomes

$$M^{(2)}_{x=x_0} = \iint_S (x - x_0)^2\delta(x,y)\,dx\,dy, \tag{13.47}$$

the *second moment* or *moment of inertia* of the lamina about the line $x = x_0$. When these things are being calculated and used in mechanics and elsewhere, information concerning moments of inertia about parallel lines (or axes) is very helpful. To obtain information of this nature, we let $\bar{x}$ be the x coordinate of the centroid and use the simple identity

$$\begin{aligned}(x - x_0)^2 &= [(x - \bar{x}) + (\bar{x} - x_0)]^2\\ &= (x - \bar{x})^2 + (\bar{x} - x_0)^2 + 2(\bar{x} - x_0)(x - \bar{x})\end{aligned}$$

to put (13.47) in the form

$$\begin{aligned}M^{(2)}_{x=x_0} = \iint_S (x - \bar{x})^2\delta(x,y)\,dx\,dy + (\bar{x} - x_0)^2 \iint_S \delta(x,y)\,dx\,dy\\ + 2(\bar{x} - x_0) \iint_S (x - \bar{x})\delta(x,y)\,dx\,dy.\end{aligned} \tag{13.471}$$

The first term in the right member is $M^{(2)}_{x=\bar{x}}$, the moment of inertia of the lamina about the line through the centroid parallel to the line $x = x_0$. The second term is $(\bar{x} - x_0)^2M$, where M is the mass of the lamina. The third term is 0 because the integral is the first moment of a lamina about a line through its centroid. Thus (13.471) reduces to the important formula

$$M^{(2)}_{x=x_0} = M^{(2)}_{x=\bar{x}} + (\bar{x} - x_0)^2M. \tag{13.472}$$

This gives the following *parallel axis theorem.*

Theorem 13.48 *The moment of inertia of a lamina about a line is equal to the sum of two terms, one being the moment of inertia of the lamina about the parallel line through the centroid and the other being the product of the mass M of the lamina and the square of the distance between the two lines*

Up to the present time, we have considered only moments of plane laminas about lines in the planes of the laminas. The second moment or moment of inertia of a lamina about a line L perpendicular to the plane of the lamina is called the *polar moment of inertia* of the lamina about the line L. As before, let the lamina cover a set S in the xy plane and let L be the line in E_3 having the equations $x = x_0$, $y = y_0$. Letting ΔS or $\Delta x\,\Delta y$ be the area of a part of the set S which contains the point (x,y) and letting $\delta(x,y)$ denote the density of the lamina at the point (x,y), we use the number

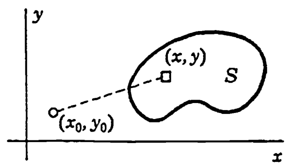


Figure 13.481

$$[(x - x_0)^2 + (y - y_0)^2]\delta(x,y)\,\Delta x\,\Delta y \tag{13.482}$$

as an approximation to the polar moment of inertia about L of the part of the lamina. The polar moment of inertia about L of the whole lamina may be denoted by the symbol $M^{(2)}_{x=x_0\;y=y_0}$. It is defined by the formula

$$M^{(2)}_{x=x_0,y=y_0} = \lim \Sigma[(x - x_0)^2 + (y - y_0)^2]\delta(x,y)\,\Delta x\,\Delta y \tag{13.483}$$

or

$$M^{(2)}_{x=x_0\;y=y_0} = \iint_S [(x - x_0)^2 + (y - y_0)^2]\delta(x,y)\,dx\,dy. \tag{13.484}$$

Comparing this with the formula (13.47) for $M^{(2)}_{x=x_0}$ and the corresponding formula for $M^{(2)}_{y=y_0}$ gives the formula

$$M^{(2)}_{x=x_0,y=y_0} = M^{(2)}_{x=x_0} + M^{(2)}_{y=y_0}, \tag{13.485}$$

which says that the polar moment of inertia of the lamina about the line $x = x_0$, $y = y_0$ perpendicular to the lamina is equal to the sum of the moments of inertia of the lamina about the two lines $x = x_0$ and $y = y_0$ in the plane of the lamina. Finally, we note that the parallel axis theorem, Theorem 13.48, holds for polar moments of inertia as well as for moments of inertia about lines in the plane of a lamina.

Problems 13.49

1 A rectangular lamina has opposite vertices at the origin and the point (a,b), and has areal density (mass per unit area) $\delta(x,y)$ at the point (x,y). Set up an iterated integral for the pth moment of the lamina about the x axis. Then evaluate the integral for the cases

(*a*) $\delta(x,y) = 1$ (*b*) $\delta(x,y) = kx$ (*c*) $\delta(x,y) = ky$

Ans.: The required integral is

$$\int_0^a dx \int_0^b y^p\delta(x,y)\,dy \qquad \text{or} \qquad \int_0^b dy \int_0^a y^p\delta(x,y)\,dx.$$

The required moments are respectively $\dfrac{ab^{p+1}}{p+1}, \dfrac{ka^2b^{p+1}}{2(p+1)}, \dfrac{kab^{p+2}}{p+2}$.

2 Supposing that $0 < p < q$, set up and evaluate an iterated integral for the area A of the region in the first quadrant bounded by the graphs of the equations $y = x^p$ and $y = x^q$. *Ans.:*

$$A = \int_0^1 dx \int_{x^q}^{x^p} dy = \frac{q-p}{(p+1)(q+1)}.$$

3 When $0 < p < q$, the region in the first quadrant bounded by the graphs of $y = x^p$ and $y = x^q$ has area $(q-p)/(p+1)(q+1)$. Find the coordinates of the centroid of this region. *Ans.:*

$$\bar{x} = \frac{(p+1)(q+1)}{(p+2)(q+2)}, \qquad \bar{y} = \frac{(p+1)(q+1)}{(2p+1)(2q+1)}.$$

4 Find the centroid of the region bounded by the x and y axes and the graph of $y = e^{-x}$. *Ans.:* $\bar{x} = 1$, $\bar{y} = \frac{1}{4}$.

5 Find the centroid of the long golf tee obtained by rotating the region of Problem 4 about the x axis. *Ans.:* $\bar{x} = \frac{1}{2}$, $\bar{y} = 0$, $\bar{z} = 0$.

6 Find the centroid of the region in the strip $0 \leqq x \leqq \pi$ bounded by the x axis and the graph of $y = \sin x$. *Ans.:* $\bar{x} = \pi/2$, $\bar{y} = \pi/8$.

7 Find the centroid of the region which lies in the interval $0 \leqq x \leqq 2a$ and is bounded by the graphs of the equations $y = 0$ and $y = b \sin \dfrac{\pi x}{2a}$. *Ans.:* $\bar{x} = a$, $\bar{y} = \dfrac{\pi}{8}b$.

8 A vertical face of a dam is bounded by the segment $0 \leqq x \leqq 2a$ of the x axis and the graph of the equation $y = -b \sin \dfrac{\pi x}{2a}$. The water level is at the top of the dam, and the weight per cubic unit of the water is w. Find the magnitude of the force on the dam. *Ans.:* $\frac{1}{2}wab^2$.

9 Find the centroid of the region in the first quadrant bounded by the coordinate axes and the hypocycloid having the equation $x^{2/3} + y^{2/3} = a^{2/3}$. *Ans.:* $\bar{x} = \bar{y} = \dfrac{256a}{315\pi}$.

10 Solve Problem 9 again, using the parametric equations

$$x = a \cos^3 t, \qquad y = a \sin^3 t$$

to obtain integrals involving t.

11 A lamina having density xy at the point $P(x,y)$ lies in the first quadrant and is bounded by the coordinate axes and the ellipse having the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Find its mass M and the coordinates of its centroid. *Ans.:* $M = \frac{1}{8}a^2b^2$, $\bar{x} = \frac{8}{15}a$, $\bar{y} = \frac{8}{15}b$.

12 For the region R bounded by the positive x axis and the graphs of $y = xe^{-x}$ and $x = a$, find each at the following:

(*a*) The area of R *Ans.:* $1 - (a + 1)e^{-a}$

(*b*) The volume of the solid obtained by rotating R about the x axis *Ans.:* $\frac{1}{4}\pi[1 - (1 + 2a + 2a^2)e^{-2a}]$

(*c*) The volume of the solid obtained by rotating R about the y axis *Ans.:* $2\pi[2 - (2 + 2a + a^2)e^{-a}]$

(*d*) The first moment of R about the x axis *Ans.:* $\frac{1}{8}[1 - (1 + 2a + 2a^2)e^{-2a}]$

(*e*) The first moment of R about the y axis *Ans.:* $2 - (2 + 2a + a^2)e^{-a}$

(*f*) The moment of inertia of R about the x axis *Ans.:* $\frac{1}{27}[2 - (2 + 6a + 9a^2 + 9a^3)e^{-3a}]$

(*g*) The moment of inertia of R about the y axis *Ans.:* $6 - (6 + 6a + 3a^2 + a^3)e^{-a}$

(*h*) The polar moment of inertia of R about the line through the origin perpendicular to the plane of R *Ans.:* Sum of answers to (*f*) and (*g*)

13 A triangular lamina has vertices at points $P_1(x_1,y_1)$, $P_2(x_2,y_2)$, $P_3(x_3,y_3)$ and has areal density (mass per unit area) $\delta(x,y)$ at the point (x,y). Assuming that the points are placed as in Figure 13.491 so that $x_1 < x_3 < x_2$ and $y_2 < y_1 <$

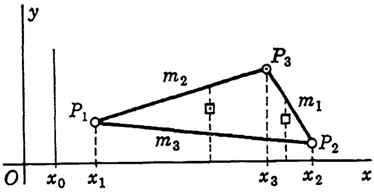


Figure 13.491

y_3, and letting m_k denote the slope of the side opposite P_k so that

$$m_1 = \frac{y_3 - y_2}{x_3 - x_2}, \qquad m_2 = \frac{y_3 - y_1}{x_3 - x_1}, \qquad m_3 = \frac{y_2 - y_1}{x_2 - x_1},$$

set up an iterated integral for the pth moment $M^{(p)}_{x=x_0}$ of the lamina about the line $x = x_0$. *Ans.:*

$$M^{(p)}_{x=x_0} = \int_{x_1}^{x_3} dx \int_{y_1+m_3(x-x_1)}^{y_1+m_2(x-x_1)} (x - x_0)^p \delta(x,y)\, dy + \int_{x_3}^{x_2} dx \int_{y_2+m_3(x-x_2)}^{y_2+m_1(x-x_2)} (x - x_0)^p \delta(x,y)\, dy.$$

14 Develop computational skill by simplifying the answer to Problem 13 for the case in which $\delta(x,y) = 1$ and $p = 0$ so the answer is numerically equal to the area of the lamina. *Ans.:*

$$\tfrac{1}{2}[(x_1y_2 - x_2y_1) - (x_1y_3 - x_3y_1) + (x_2y_3 - x_3y_2)] \qquad \text{or} \qquad \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

15 Figure 13.492 shows two parallel rods of lengths a and b. The rod on the left has linear density (mass per unit length) $\delta_1(t)$ at distance t from its lower

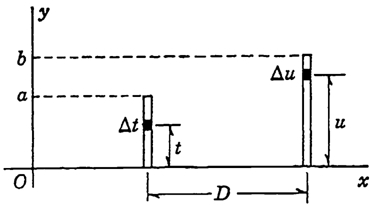


Figure 13.492

end, and the rod on the right has linear density $\delta_2(u)$ at distance u from its lower end. We undertake to learn about the total or resultant gravitational force $\mathbf{F}$ on the left-hand rod that is produced by the right-hand rod. Assume that particles of mass m_1 and m_2 at points P_1 and P_2 attract each other with a force of magnitude $Gm_1m_2/|\overrightarrow{P_1P_2}|^2$, where G is a gravitational constant, and the actual force pulling the particle at P_1 toward P_2 is obtained by multiplying this magnitude by $\overrightarrow{P_1P_2}/|\overrightarrow{P_1P_2}|$, the unit vector which has its tail at P_1 and points toward P_2. Derive the formula

$$G\frac{\delta_1(t)\,\Delta t\,\delta_2(u)\,\Delta u}{[D^2 + (u - t)^2]^{3/2}}[D\mathbf{i} + (u - t)\mathbf{j}]$$

for the force which an element (or subset) of the rod on the right exerts upon an element of the rod on the left. Then derive the formula

$$G\delta_1(t)\,\Delta t \int_0^b \frac{\delta_2(u)[D\mathbf{i} + (u - t)\mathbf{j}]}{[D^2 + (u - t)^2]^{3/2}}\, du$$

for the force which the whole rod on the right exerts upon the element of the rod on the left. Then derive the formula

$$\mathbf{F} = G\int_0^a \delta_1(t)\, dt \int_0^b \frac{\delta_2(u)[D\mathbf{i} + (u - t)\mathbf{j}]}{[D^2 + (u - t)^2]^{3/2}}\, du.$$

16 Two identical slender rods having constant linear density δ occupy the intervals $-a \leqq x \leqq -\epsilon$ and $\epsilon < x \leqq a$ of an x axis. It is supposed that $0 < \epsilon$

$< a$, and we can be interested in situations in which ϵ is small. Set up an integral for the gravitational force which the rod on the right exerts upon the rod on the left.

17 Figure 13.493 shows rods of lengths a and b that have constant linear density δ, that lie on the x and y axes, and that are hinged at their ends at the

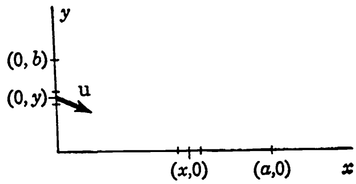


Figure 13.493

origin. A little segment of the horizontal rod in a neighborhood of $(x,0)$ produces a little gravitational force $\mathbf{u}$ on a little segment of the vertical rod in a neighborhood of $(0,y)$. This little force $\mathbf{u}$ has a little scalar horizontal component u_x which produces a little torque (or first moment) yu_x which tends to rotate the vertical rod toward the horizontal one. There are hordes of little torques. Set up an integral for the total torque.

18 Most people having serious interest in mathematics want to see and perhaps study the nontrivial steps by which the important *Euler gamma integral formula*

$$z! = \int_0^\infty t^z e^{-t}\,dt \qquad (z > -1) \tag{1}$$

is derived from the definition of $z!$ given in Problem 11 of Problems 3.39. We start with the fact that, when z is not a negative integer, $z!$ is defined by the formulas

$$z! = \lim_{n\to\infty} F_n(z),\ F_n(z) = \frac{n!n^z}{(z+1)(z+2)\cdots(z+n)}. \tag{2}$$

Expressing $[(z+1)(z+2)\cdots(z+n)]^{-1}$ as a sum of partial fractions leads to the formula

$$F_n(z) = n^{z+1}\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\frac{1}{z+k+1}.$$

To put this in a form that can be simplified by use of the binomial formula, we use the fact that

$$\frac{1}{z+k+1} = \int_0^1 u^{z+k}\,du = \int_0^1 u^z u^k\,du \tag{3}$$

when $z > -1$ and $k = 0, 1, 2, \ldots$. Assuming henceforth that $z > -1$, we find that

$$F_n(z) = n^{z+1}\int_0^1 u^z \sum_{k=0}^{n-1}\binom{n-1}{k}1^{n-k}(-u)^k\,du \tag{4}$$

and hence that

$$F_n(z) = n^{z+1}\int_0^1 u^z(1-u)^{n-1}\,du. \tag{5}$$

Changing the variable of integration by setting $u = t/(n-1)$ then gives

$$(6) \qquad F_n(z) = \left(1 - \frac{1}{n}\right)^{z+1} \int_0^{n-1} t^z \left(1 - \frac{t}{n-1}\right)^{n-1} dt$$

when $n > 1$. Therefore

$$(7) \qquad z! = \lim_{n\to\infty} \int_0^\infty G_n(t)\, dt$$

where

$$(8) \qquad G_n(t) = t^z\left(1 - \frac{t}{n-1}\right)^{n-1} \qquad (0 < t \leqq n-1)$$

and $G_n(t) = 0$ when $t \geqq n$. It can be shown that

$$(9) \qquad \lim_{n\to\infty} G_n(t) = t^z e^{-t}, \qquad |G_n(t)| \leqq |t^z e^{-t}|.$$

While full exploration of the matter lies beyond the scope of elementary calculus, (1) is a consequence of (7), (9), and the Lebesgue criterion of dominated convergence for taking limits under integral signs. When m and s are numbers for which $s > m$, we can put $t = (s - m)x$ in (1) to obtain the formula

$$(10) \qquad \frac{z!}{(s-m)^{z+1}} = \int_0^\infty e^{-sx}x^z e^{mx}\, dx.$$

Particularly when it is recognized that (10) is valid even when m is complex, this single formula (10) is the equivalent of a huge table of Laplace transforms and is therefore very important.

19 We examine the formulas by which ideas of this chapter are used to start with the Euler gamma integral formula

$$(1) \qquad z! = \int_0^\infty t^z e^{-t}\, dt$$

of the preceding problem and derive the beta integral formula

$$(2) \qquad \int_0^1 t^p(1-t)^q\, dt = \frac{p!q!}{(p+q+1)!}$$

It is supposed that z, p, and q are complex numbers having real parts exceeding -1. Use of (1) gives

$$(3) \qquad p!q! = \int_0^\infty x^p e^{-x}\, dx \int_0^\infty y^q e^{-y}\, dy$$

where the right side is the product of two integrals. Writing this as an iterated integral and putting $y = u - x$ give

$$(4) \qquad p!q! = \int_0^\infty dx \int_0^\infty x^p y^q e^{-(x+y)}\, dy = \int_0^\infty dx \int_x^\infty x^p(u-x)^q e^{-u}\, du$$

The Fubini theorem justifies change of order of integration to obtain the first equality in

$$(5) \quad p!q! = \int_0^\infty e^{-u}\, du \int_0^u x^p(u-x)^q\, dx = \int_0^\infty u^{p+q+1}e^{-u}\, du \int_0^1 t^p(1-t)^q\, dt$$

and putting $x = ut$ gives the second equality Therefore

$$p!q! = (p + q + 1)! \int_0^1 t^p(1 - t)^q \, dt \tag{6}$$

and (1) follows.

13.5 *Integrals in polar coordinates* In some cases a plane set S and a function f defined over S are such that the double integral defined as in (13.37) by

$$\iint_S f(P) \, dS = \lim \sum f(P) \, \Delta S \tag{13.51}$$

can be advantageously expressed in terms of polar coordinates ρ and ϕ. For example, suppose that, as in Figure 13.52, S is the set of points

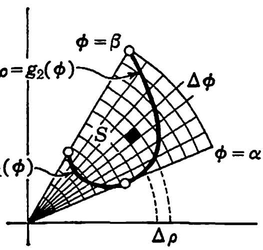


Figure 13.52

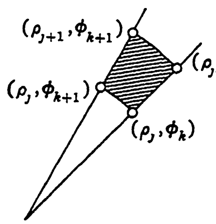


Figure 13.53

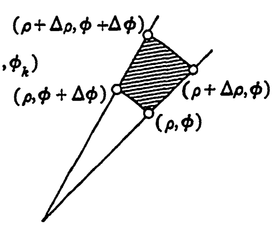


Figure 13.531

having polar coordinates ρ, ϕ for which $g_1(\phi) \leqq \rho \leqq g_2(\phi)$ and $\alpha \leqq \phi \leqq \beta$, where g_1 and g_2 are continuous functions for which $0 \leqq g_1(\phi) \leqq g_2(\phi)$ when $\alpha \leqq \phi \leqq \beta$. Our first step is to partition S into subsets S_1, S_2, $\cdots$, S_n by radial lines having the polar equations $\phi = \phi_0$, $\phi = \phi_1$, $\cdots$, $\phi = \phi_m$, where $\alpha = \phi_0 < \phi_1 < \cdots < \phi_m = \beta$, and by circles or circular arcs having the equations $\rho = \rho_0$, $\rho = \rho_1$, $\cdots$, $\rho = \rho_{m'}$, as in Figure 13.52. A typical one of the subsets is a part of a sector having corners at the points whose polar coordinates are shown in Figure 13.53. We set $\Delta\rho_j = \rho_{j+1} - \rho_j$ and $\Delta\phi_k = \phi_{k+1} - \phi_k$ and then simplify the things we write by discarding all subscripts to obtain the polar coordinates shown in Figure 13.531. The symbol ΔS then represents the area of the shaded set in Figure 13.531. This shaded set is not a rectangle, because the straight sides are not parallel, and the inner and outer sides are arcs of circles which are not parallel line segments. It is, however, thoroughly reasonable to have the opinion that, when $\rho > 0$ and $\Delta\rho$ and $\Delta\phi$ are small, the shaded set is "nearly" rectangular and that ΔS should be closely approximated by the product of $\Delta\rho$ (the length of one of the straight sides) and $\rho \, \Delta\phi$ (the length of the inner curved side). This suggests that we should be able to use $\rho \, \Delta\phi \, \Delta\rho$ as an approximation to ΔS. Much more can be said about this approximation business, and there are different ways of proving that, when $\rho > 0$,

$$\lim_{\Delta\rho\to 0,\ \Delta\theta\to 0} \frac{\Delta S}{\rho\Delta\,\rho\,\Delta\phi} = 1. \tag{13.532}$$

It is, in fact, easy to work out an exact formula for ΔS, because ΔS is the difference of the areas of two sectors having central angle $\Delta\phi$. The larger sector has radius $\rho + \Delta\rho$ and the smaller sector has radius ρ, so

$$\Delta S = \tfrac{1}{2}(\rho + \Delta\rho)^2\,\Delta\phi - \tfrac{1}{2}\rho^2\,\Delta\phi = \tfrac{1}{2}(2\rho\,\Delta\rho + (\Delta\rho)^2)\,\Delta\phi$$

and hence

$$\Delta s = \left(\rho + \frac{\Delta\rho}{2}\right)\Delta\phi\,\Delta\rho. \tag{13.533}$$

This can be used to prove (13.532), and it can also give us another idea. If we let $\rho^* = \rho + \Delta\rho/2$, then we obtain the exact formula

$$\Delta S = \rho^*\,\Delta\phi\,\Delta\rho. \tag{13.534}$$

One who wishes to do so may insist that the formula

$$\Delta S = \rho\,\Delta\phi\,\Delta\rho \tag{13.535}$$

is an exact formula obtained by setting $\Delta\rho_j = \rho_{j+1} - \rho_j$, $\Delta\phi_k = \phi_{k+1} - \phi_k$, $\rho_j^* = \rho_j + \frac{1}{2}(\rho_{j+1} - \rho_j)$, and discarding all subscripts and stars. When we put $f(P)$ in the form $f(\rho,\phi)$, where ρ and ϕ are polar coordinates, we are therefore able to put (13.51) in the form

$$\iint_S f(\rho,\phi)\rho\,d\phi\,d\rho = \lim \sum f(\rho,\phi)\rho\,\Delta\phi\,\Delta\rho. \tag{13.54}$$

Assuming that f is bounded over S and is sufficiently continuous to make all of the integrals exist, we can use Theorem 13.38 to express this double integral as an iterated integral. For example, when S is the set featured in Figure 13.52, we observe that the point having polar coordinates (ρ,ϕ) lies in S when $\alpha \leqq \phi \leqq \beta$ and, for each such ϕ, $g_1(\phi) \leqq \rho \leqq g_2(\phi)$, so

$$\iint_S f(\rho,\phi)\rho\,d\phi\,d\rho = \int_\alpha^\beta d\phi \int_{g_1(\phi)}^{g_2(\phi)} f(\rho,\phi)\rho\,d\rho. \tag{13.55}$$

Except in cases where S is a circular sector (which may be a whole circular disk) or the difference of two circular sectors (which may be a whole circular ring), it is usually not convenient to use formulas of the form

$$\iint_S f(\rho,\phi)\rho\,d\phi\,d\rho = \int_a^b d\rho \int_{h_1(\rho)}^{h_2(\rho)} f(\rho,\phi)\rho\,d\phi \tag{13.551}$$

in which the first integration is with respect to ϕ. The following three examples serve to show how double and iterated integrals in polar coordinates can be set up. It is not recommended that the formulas be remembered; we reconstruct them whenever we want to use them.

Example 1 Let $0 \leqq \alpha < \beta \leqq \pi$, let $f(\phi) \geqq 0$ when $\alpha \leqq \phi \leqq \beta$, let f be Riemann integrable over the interval $\alpha \leqq \phi \leqq \beta$, and let S be the plane set of points having polar coordinates (ρ,ϕ) for which $\alpha \leqq \phi \leqq \beta$ and $0 \leqq \rho \leqq f(\phi)$. In case f is continuous, S is the set bounded by the polar graphs of the equations $\phi = \alpha$, $\phi = \beta$, and $\rho = f(\phi)$. The problem is to find the volume $|B|$ of the solid body B that is "generated" by rotating the set S about the initial line (or x axis) of Figure 13.56. The first

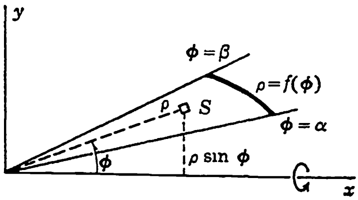


Figure 13.56

step is to partition S into subsets by radial lines through the origin and circular arcs having centers at the origin. Letting ρ and ϕ denote the polar coordinates of a point in a typical subset, we use the number $\rho\,\Delta\rho\,\Delta\phi$ as an approximation to the area of the subset. When the subset is rotated about the x axis, it generates a solid which may be thought of as a ring or hoop or gasket having radius $\rho \sin\phi$ and length $2\pi\rho\sin\phi$. The number

$$(2\pi\rho\sin\phi)(\rho\,\Delta\rho\,\Delta\phi),$$

or

$$2\pi\rho^2\sin\phi\,\Delta\rho\,\Delta\phi,$$

being the product of the length of the ring and the area of a cross section of the ring, is then used as an approximation to the volume of the ring. The sum in

$$(13.561)\qquad |B| = \lim \Sigma 2\pi\rho^2 \sin\phi\,\Delta\rho\,\Delta\phi$$

is then used as an approximation to the volume $|B|$ of the body B, and the limit is (without proof) taken to be the exact volume $|B|$. The right member of (13.561) is a double integral. Expressing this as an iterated integral gives

$$(13.562)\qquad |B| = \int_\alpha^\beta d\phi \int_0^{f(\phi)} 2\pi\rho^2 \sin\phi\,d\rho$$

or

$$(13.563)\qquad |B| = 2\pi \int_\alpha^\beta \sin\phi\,d\phi \int_0^{f(\phi)} \rho^2\,d\rho$$

or

$$(13.564)\qquad |B| = \frac{2\pi}{3} \int_\alpha^\beta [f(\phi)]^3 \sin\phi\,d\phi.$$

It is particularly easy to evaluate this integral for the special case in which there is a constant R for which $f(\phi) = R$ when $\alpha \leqq \phi \leqq \beta$. In this case, S is a circular sector and

(13.565) $$|B| = \frac{2\pi R^3}{3}[\cos\alpha - \cos\beta].$$

In case $\alpha = 0$ and $\beta = \pi$, the solid body B is a complete spherical ball of radius R and the right side reduces to the correct volume $\frac{4}{3}\pi R^3$.

Example 2 Let $\alpha < \beta \leqq \alpha + 2\pi$, let $f(\phi) \geqq 0$ when $\alpha \leqq \phi \leqq \beta$, let f be Riemann integrable over the interval $\alpha \leqq \phi \leqq \beta$, let S be the plane set of points having polar coordinates (ρ,ϕ) for which $\alpha \leqq \phi \leqq \beta$ and $0 \leqq \rho \leqq f(\phi)$, and let a lamina (or flat plate) cover S and have areal density (mass per unit area) $\delta(\rho,\phi)$ at the point (ρ,ϕ) of S. The problem is to find the pth moment of the lamina about the y axis shown in Figure

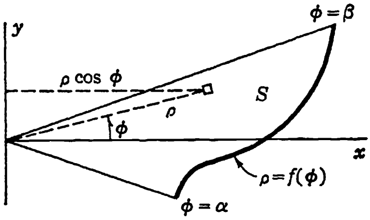


Figure 13.57

13.57. As before, we use $\rho\,\Delta\rho\,\Delta\phi$ as an approximation to the area of a subset of S. The next step is to use

$$\delta(\rho,\phi)\rho\,\Delta\rho\,\Delta\phi$$

as an approximation to the mass of a part of the lamina. Multiplying this by x^p or $(\rho\cos\phi)^p$, the pth power of the x coordinate of a point in the part, gives an approximation to the pth moment about the line $x = 0$ of the part of the lamina. This leads to the formula

$$M^{(p)}_{x=0} = \lim \Sigma(\rho\cos\phi)^p\delta(\rho,\phi)\rho\,\Delta\rho\,\Delta\phi$$

for the required pth moment. The right member is a double integral, and expressing it as an iterated integral gives

$$M^{(p)}_{x=x_0} = \int_\alpha^\beta \cos^p\phi\,d\phi \int_0^{f(\phi)} \delta(\rho,\phi)\rho^{p+1}\,d\rho.$$

Example 3 We now require an integral for the polar moment of inertia $M^{(2)}_{x=x_0,y=y_0}$ of the lamina of Example 2 about the line L through the origin perpendicular to the plane of the lamina. We use $\rho\,\Delta\rho\,\Delta\phi$ as an approximation to the area of a subset of the lamina and then use $\delta(\rho,\phi)\rho\,\Delta\rho\,\Delta\phi$ as an approximation to the mass of the subset. Multiplying this by ρ^2, the square of the distance from the line L to the subset, gives the approximation

(13.58) $$\delta(\rho,\phi)\rho^3\,\Delta\rho\,\Delta\phi$$

to the polar moment of the subset. This leads to the formulas

$$M^{(2)}_{x=x_0\ y=y_0} = \lim \Sigma\delta(\rho,\phi)\rho^3\,\Delta\rho\,\Delta\phi$$

and

$$M^{(2)}_{x=x_0\ y=y_0} = \int_\alpha^\beta d\phi \int_0^{f(\phi)} \delta(\rho,\phi)\rho^3\,d\rho.$$

Problems 13.59

1 Set up an iterated integral in polar coordinates for the volume V of the solid generated by rotating, about the y axis, the triangular set bounded by the lines having the polar equations $\phi = 0$ and $\phi = \beta$ (where $0 < \beta < \pi/2$) and the line having the rectangular equation $x = h$ (where $h > 0$). Then evaluate the integral and discover that $V = \frac{2\pi}{3} h^3 \tan\beta$ *Remark:* Correctness of the answer can be verified by use of elementary geometry, because the solid is obtainable by removing a part of a solid right circular cone from a segment of a solid right circular cylinder.

2 Find the distance from the vertex to the centroid of a lamina having the foım of a circular sector of radius R and central angle 2α when each of the following is true.

(*a*) The lamina is uniform. *Ans.:* $\frac{2}{3}\frac{\sin\alpha}{\alpha}R$

(*b*) The density is proportional to kth power of the distance from the vertex. *Ans.:* $\frac{k+2}{k+3}\frac{\sin\alpha}{\alpha}R$

3 Using the equation $\rho = 2a\cos\phi$, set up and evaluate an iterated integral in polar coordinates for the moment of inertia I_0 of a circular disk about a line perpendicular to the disk and containing a point on the boundary of the disk. *Ans.:*

$$I_0 = 2\int_0^{\pi/2} d\phi \int_0^{2a\cos\phi} \rho^3\,d\rho = \tfrac{3}{2}\pi a^4.$$

4 Using the equation $\rho = a$, set up an iterated integral in polar coordinates for the moment of inertia I_0 of a circular disk about the axis of the disk (the line through the center of the disk perpendicular to the plane of the disk) when the density of the disk at each point is the pth power of the distance from the point to the diameter on the initial line (or x axis). *Ans.:*

$$I_0 = 4\int_0^{\pi/2} \sin^p\phi\,d\phi \int_0^a \rho^{p+3}\,d\rho.$$

Remark: With the aid of the formula

$$\int_0^{\pi/2} \sin^p x\cos^q x\,dx = \frac{1}{2}\frac{\left(\frac{p-1}{2}\right)!\left(\frac{q-1}{2}\right)!}{\left(\frac{p+q}{2}\right)!}$$

appearing among the problems of Section 8.4, and the fact that $(-\frac{1}{2})! = \sqrt{\pi}$, we can put the result in the form

$$I_0 = \frac{4a^{p+4}}{p+4}\int_0^{\pi/2} \sin^p \phi\, d\phi = \frac{2\sqrt{\pi}\, a^{p+4}}{p+4} \frac{\left(\frac{p-1}{2}\right)!}{\left(\frac{p}{2}\right)!}$$

which is valid even when p is not an integer.

5 A circular disk has radius a and constant density δ. Find its moment of inertia about each of the following:

(*a*) A diameter (line, not number) *Ans.:* $\frac{1}{4}\pi a^4\delta$

(*b*) A tangent in its plane *Ans :* $\frac{5}{4}\pi a^4\delta$

6 Find the volume of the solid S obtained by rotating, about the x axis, the region in the first quadrant bounded by the x axis and the graph of the polar equation $\rho = a\sqrt{\cos\phi}$. *Outline of solution:* The volume $|S|$ of S is approximated by the sum of "elementary" rings a sample one of which has cross-sectional area $\rho_k^* \,\Delta\rho_k\,\Delta\phi_k$ and length $2\pi\rho_k^* \sin\phi_k$. This leads to the formula

$$|S| = 2\pi \int_0^{\pi/2} \sin\phi\, d\phi \int_0^{a\sqrt{\cos\phi}} \rho^2\, d\rho.$$

Ans.: $4\pi a^3/15$.

7 Supposing that the solid S of Problem 6 has constant density δ, find the gravitational force $\mathbf{F}$ which it exerts upon a particle m^* of mass m concentrated at the origin. *Outline of solution:* The sample elementary ring of the preceding problem has mass M_k, where

$$M_k = 2\pi\delta \sin\phi_k \rho_k^2\, \Delta\phi_k\, \Delta\rho_k.$$

The force $\Delta\mathbf{F}_k$ which this ring exerts upon m^* is the same as the $\mathbf{i}$ component of the force exerted upon m^* by a single particle of the same mass M_k concentrated at the point having polar coordinates (ρ_k^*,ϕ_k). Therefore,

$$\Delta\mathbf{F}_k = \mathbf{i}\,\frac{GmM_k}{\rho_k^2}\cos\phi_k.$$

This leads to the formula

$$\mathbf{F} = 2\pi Gm\delta\mathbf{i}\int_0^{\pi/2} \sin\phi\cos\phi \int_0^{a\sqrt{\cos\phi}} d\rho.$$

The answer is $\mathbf{F} = \frac{4}{5}\pi aGm\delta\mathbf{i}$. *Remark:* We embark on a little excursion to see that the solid S of this problem and the preceding one is a most remarkable solid. If a particle P of mass M is located at the point in our plane having polar coordinates (ρ,ϕ), or at a point which is obtained by rotating it part of the way around the x axis, then the scalar component F_x in the direction of the x axis which it exerts upon m^* is

$$F_x = \frac{GmM}{\rho^2}\cos\phi.$$

If P lies inside our solid S, then $\rho < a\sqrt{\cos\phi}$, so $\rho^2 < a^2\cos\phi$, so $(\cos\phi)/\rho^2 > 1/a^2$, so $F_x > GmM/a^2$. If P lies outside our solid S, then $F_x \leqq 0$ if $\pi/2 \leqq |\phi| <$

π and $\rho > a\sqrt{\cos\phi}$ if $|\phi| < \pi/2$. It follows that if P lies outside our solid, then $F_x < GmM/a^2$. These results show that if we transfer material from the inside to the outside of our solid, we decrease the force which pulls m^* in the direction of the positive x axis. This fact and the answer to Problem 6 give the following interesting conclusion. Of all solids having volume $4\pi a^3/15$ and uniform density δ, our solid S is the one and only solid which exerts, upon a particle at the origin, the greatest force in the direction of the positive x axis. The force $\mathbf{F}_B$ exerted upon m^* by a spherical ball B which has the same volume and density as S, and which has its center on the positive x axis and has the origin on its surface, is the first of the vectors

$$F_B = \sqrt[3]{25}\,\tfrac{4}{15}\pi aGm\delta\mathbf{i}, \qquad \mathbf{F}_S = 3\tfrac{4}{15}\pi aGm\delta\mathbf{i},$$

and the force exerted by our solid S is the second one. Thus we have a new proof of the inequality $\sqrt[3]{25} < 3$. Tables say that

$$\sqrt[3]{25} = 2.924018.$$

Spherical balls are not the best, but the best is only about 2.5 per cent better.

8 Partly because the result is thoroughly important in probability and statistics and elsewhere, and partly because understandings of multiple and iterated integrals should be developed, this problem requires learning a standard method by which the formulas

$$(1) \qquad \int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}, \qquad \frac{1}{\sqrt{\pi}}\int_{-\infty}^\infty e^{-x^2}\,dx = 1$$

and

$$(2) \qquad \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^\infty e^{-x^2/2\sigma^2}\,dx = 1, \qquad \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^\infty x^2e^{-x^2/2\sigma^2}\,dx = \sigma^2$$

are derived. Supposing that $h > 0$, we define $F(h)$ by the first of the two equivalent formulas

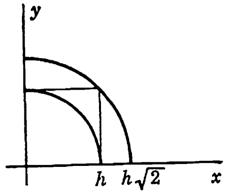
$$(3) \qquad F(h) = \int_0^h e^{-x^2}\,dx, \qquad F(h) = \int_0^h e^{-y^2}\,dy.$$

Then

$$(4) \qquad [F(h)]^2 = \left[\int_0^h e^{-x^2}\,dx\right]\left[\int_0^h e^{-y^2}\,dy\right] = \int_0^h dx\int_0^h e^{-(x^2+y^2)}\,dy$$
$$= \iint_{Q(h)} e^{-(x^2+y^2)}\,dx\,dy$$

where we have, in order, the product of two integrals, an iterated integral and, finally, a double integral over the square region $Q(h)$ of Figure 13.591. This

Figure 13.591



turns out to be useful because the double integral can be compared with other double integrals that are easily evaluated by use of polar coordinates. Let $D(h)$ be the quadrant of the circular disk consisting of points having polar coordinates (ρ,ϕ) for which $0 \leqq \rho \leqq h$ and $0 \leqq \phi \leqq \pi/2$ and let

$$G(h) = \iint_{D(h)} e^{-(x^2+y^2)}\,dx\,dy. \tag{5}$$

Then, because the integrands are everywhere positive,

$$G(h) \leqq [F(h)]^2 \leqq G(h\sqrt{2}). \tag{6}$$

Writing (5) in terms of polar coordinates and evaluating the result by use of an iterated integral gives

$$\begin{aligned} G(h) &= \iint_{D(h)} e^{-\rho^2}\rho\,d\phi\,d\rho = \int_0^{\pi/2} d\phi \int_0^h e^{-\rho^2}\rho\,d\rho \\ &= \int_0^{\pi/2} d\phi\,[-\tfrac{1}{2}e^{-\rho^2}]_{\rho=0}^h = \frac{\pi}{4}[1 - e^{-h^2}]. \end{aligned} \tag{7}$$

Since $F(h) > 0$, this and (6) give

$$\frac{\sqrt{\pi}}{2}\sqrt{1 - e^{-h^2}} \leqq \int_0^h e^{-x^2}\,dx \leqq \frac{\sqrt{\pi}}{2}\sqrt{1 - e^{-2h^2}}. \tag{8}$$

Taking the limit as $h \to \infty$ gives the first formula in (1), and the second formula in (1) follows from the first. The first formula in (2) is obtained from the second formula in (1) by a change of variable; the trick is to set $x = t/\sqrt{2}\,\sigma$ and then replace t by x in the new integral. The second formula in (2) is obtained by integrating by parts and using the first formula in (2). *Remark:* The formulas obtained by replacing x by $x - M$ in (2) are important. The function Φ defined by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-M)^2}{2\sigma^2}} \tag{9}$$

is the *Gauss probability density function* having *mean* (or average) M and *standard deviation* σ. In appropriate circumstances, the number

$$\int_a^b \Phi(x)\,dx \tag{10}$$

is, when $a < b$, used for the probability that a number x (which could be the number of red corpuscles per cubic centimeter in your blood) lies between a and b. The formula

$$\int_M^{M+\lambda\sigma} \Phi(t)\,dt = \frac{1}{\sqrt{2\pi}\,\sigma}\int_M^{M+\lambda\sigma} e^{-\frac{(x-M)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}}\int_0^\lambda e^{-t^2/2}\,dt, \tag{11}$$

which is proved by use of the substitution $(x - M)/\sigma = t$, facilitates calculations of probabilities because the last member is tabulated as a function of λ. Many students of anthropology, medicine, education, agronomy, and other branches of agriculture are required to study analytic geometry and calculus so they can start learning about these things.

13.6 Triple integrals; rectangular coordinates We have partitioned sets in E_1 and E_2 into subsets and used these partitions in the process of setting up Riemann sums and Riemann integrals of functions. One reason for the importance of what we have done lies in the fact that the methods are easily extended to provide information about triple integrals of the form

$$\iiint_S f(P)\, dS \tag{13.61}$$

in which S is a set in three-dimensional space E_3 in which it is sometimes convenient to suppose that we exist.

Let S be a bounded set in E_3 which may be a spherical ball (the set consisting of the points inside and the points on a sphere) or any other bounded set in E_3 which has a positive volume $|S|$. As was the case when we defined double integrals, we do not allow any one brand of coordinates to dominate our work. We suppose that we have a bounded function f defined over S and use the symbol $f(P)$ to denote the value of f at P. For example, $f(P)$ could be the density (mass per unit volume) at P or the product of the density at P and the specific heat at P and the temperature at P. The first step in our approach to a Riemann sum is to make a partition Q (again the letter P has been preempted) of the set S into n subsets $S_1, S_2, \cdots, S_n$. The only thing we require of the sets $S_1, S_2, \cdots, S_n$ is that they be nonoverlapping, that their union be S, and that each one of them have positive volume. The notational transition from Riemann sums to Riemann integrals is facilitated by denoting the volumes of the sets $S_1, S_2, \cdots, S_n$ by the symbols $\Delta S_1, \Delta S_2, \cdots, \Delta S_n$. Thus, for each $k = 1, 2, \cdots, n$, the symbol ΔS_k does not stand for a part of the set S; it stands for the volume of a part of S. For each $k = 1, 2, \cdots, n$, let P_k be a point in the set S_k. The number RS (Riemann sum) defined by

$$\text{RS} = \sum_{k=1}^{n} f(P_k)\, \Delta S_k \tag{13.62}$$

is then a Riemann sum formed for the function f and for the partition Q of S. The norm $|Q|$ of the partition Q is, as in previous cases, the greatest of the diameters of the subsets. If there is a number I such that to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that

$$\left| I - \sum_{k=1}^{n} f(P_k)\, \Delta S_k \right| < \epsilon \tag{13.621}$$

whenever the sum is a Riemann sum formed for the function f and for a partition Q of S for which $|Q| < \delta$, then f is said to be Riemann integrable over S and I is said to be the *Riemann integral* of f over S. The integral is usually denoted by the symbol

$$\iiint_S f(P)\,dS. \tag{13.622}$$

The integral is called a *triple integral*, and the three integral signs serve to remind us that S is a three-dimensional set, that is, a set in E_3 having positive volume. As in previous cases, it is a convenience (and sometimes also a source of confusion) to introduce the notation of limits and write

$$\iiint_S f(P)\,dS = \lim_{|Q|\to 0} \sum_{k=1}^{n} f(P_k)\,\Delta S_k \tag{13.623}$$

or

$$\iiint_S f(P)\,dS = \lim \sum f(P)\,\Delta S. \tag{13.624}$$

The following theorem, which is analogous to Theorem 13.38, is very useful.

Theorem 13.63 *If S is a subset of a region R consisting of points (x,y,z) for which $a_1 \leqq x \leqq a_2$, $b_1 \leqq y \leqq b_2$, $c_1 \leqq z \leqq c_2$, if $f(x,y,z) = 0$ when (x,y,z) is a point in R but not in S, and if the eight integrals*

$$I_1 = \iiint_S f(x,y,z)\,dS, \qquad I_2 = \iiint_R f(x,y,z)\,dS$$
$$I_3 = \int_{a_1}^{a_2} dx \int_{b_1}^{b_2} dy \int_{c_1}^{c_2} f(x,y,z)\,dz, \qquad I_4 = \int_{a_1}^{a_2} dx \int_{c_1}^{c_2} dz \int_{b_1}^{b_2} f(x,y,z)\,dy$$
$$I_5 = \int_{b_1}^{b_2} dy \int_{a_1}^{a_2} dx \int_{c_1}^{c_2} f(x,y,z)\,dz, \qquad I_6 = \int_{b_1}^{b_2} dy \int_{c_1}^{c_2} dz \int_{a_1}^{a_2} f(x,y,z)\,dx$$
$$I_7 = \int_{c_1}^{c_2} dz \int_{a_1}^{a_2} dx \int_{b_1}^{b_2} f(x,y,z)\,dy, \qquad I_8 = \int_{c_1}^{c_2} dz \int_{b_1}^{b_2} dy \int_{a_1}^{a_2} f(x,y,z)\,dx$$

all exist, then

$$I_1 = I_2 = I_3 = I_4 = I_5 = I_6 = I_7 = I_8,$$

that is, the eight integrals are all equal.

Remarks analogous to those following Theorem 13.38 apply here. Proof of the theorem lies far beyond the scope of this course. We can be content with a hazy understanding of the fact that the triple integrals I_1 and I_2 will exist if f is bounded and the set D of discontinuities of f has volume (three-dimensional Lebesgue measure) 0. So far as elementary applications to elementary problems are concerned, we can be sure that if the set S and the function f are bounded, then all of the integrals appearing in the theorem must exist and must have the same value.

To develop a technique for setting up triple and iterated integrals,

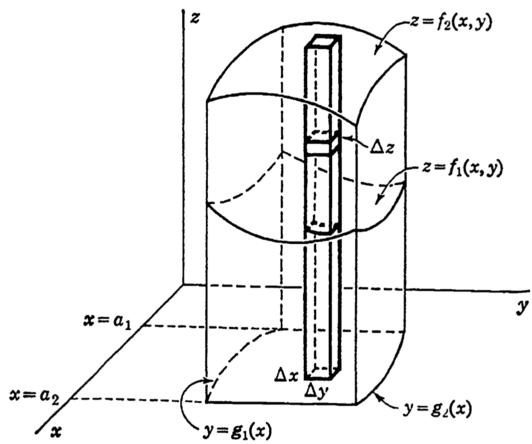


Figure 13.64

we consider an example. As in Figure 13.64, let S be the set or solid bounded by the surfaces having the equations $z = f_1(x,y)$, $z = f_2(x,y)$, $y = g_1(x)$, $y = g_2(x)$, $x = a_1$, and $x = a_2$. We suppose that, at each point (x,y,z) of the solid, the solid has density (mass per unit volume) $\delta(x,y,z)$. This means that to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that

(13.641) $$\left| \delta(x,y,z) - \frac{\Delta m}{\Delta S} \right| < \epsilon$$

whenever ΔS is the volume of a part of the solid containing the point (x,y,z) and having diameter less than δ, and Δm is the mass of the part. In case there is a constant k such that $\delta(x,y,z) = k$ whenever (x,y,z) is a point of S, the solid is said to be *homogeneous*. While the ideas can be applied in some other cases, we suppose that all of the functions which we have introduced are continuous. Supposing that x_0 is a given number and that p is a nonnegative integer that is 0 or 1 or 2 in most applications, we set up integrals for $M^{(p)}_{x=x_0}$, the pth moment of the solid about the plane $x = x_0$. Using rectangular coordinates, we slice S into subsets by planes parallel to the coordinate planes. A typical subset, such as one shown in Figure 13.64, has volume $\Delta x\, \Delta y\, \Delta z$. Letting (x,y,z) be a point in the subset, we use the number

(13.642) $$\delta(x,y,z)\, \Delta x\, \Delta y\, \Delta z$$

(the product of mass per unit volume and volume) as an approximation to the mass of the subset. If the total mass of the subset were concentrated at the point (x,y,z), its pth moment about the plane $x = x_0$ would be

(13.643) $$(x - x_0)^p \delta(x,y,z)\, \Delta x\, \Delta y\, \Delta z.$$

We therefore use this number as an approximation to the pth moment of the subset. The sum

$$\Sigma(x - x_0)^p\delta(x,y,z)\,\Delta x\,\Delta y\,\Delta z, \tag{13.644}$$

which contains a term for each subset, should then be a good approximation to the total pth moment of the whole solid whenever the diameters of the subsets are all small. This leads us to the formula

$$M^{(p)}_{x=x_0} = \lim \Sigma(x - x_0)^p\delta(x,y,z)\,\Delta x\,\Delta y\,\Delta z, \tag{13.645}$$

the right side of which is taken to be the definition of the number $M^{(p)}_{x=x_0}$ which we are seeking. In accordance with the definition of triple integrals, the right side of (13.645) is a triple integral which we can denote by one or the other of the symbols in the formula

$$\begin{aligned} M^{(p)}_{x=x_0} &= \iiint_S (x - x_0)^p\delta(x,y,z)\,dS \\ &= \iiint_S (x - x_0)^p\delta(x,y,z)\,dx\,dy\,dz. \end{aligned} \tag{13.646}$$

With the aid of Theorem 13.63, we can undertake to express the triple integral in terms of iterated integrals in various ways. It is, however, worthwhile to learn to use a procedure which leads directly to iterated integrals. As above, we build up the expression

$$(x - x_0)^p\delta(x,y,z)\,\Delta x\,\Delta y\,\Delta z \tag{13.65}$$

to serve as an approximation to the required moment of a single subset. We then form the sum

$$\Delta x\,\Delta y \sum_{x,y \text{ fixed}} (x - x_0)^p\delta(x,y,z)\,\Delta z, \tag{13.651}$$

where the part "x, y, fixed" of the symbol serves to inform us that the sum contains only terms arising from those subsets which comprise a single vertical column such as that shown in Figure 13.64. Thus (13.651) and the number

$$\Delta x\,\Delta y \int_{f_1(x,y)}^{f_2(x,y)} (x - x_0)^p\delta(x,y,z)\,dz \tag{13.652}$$

are approximations to the required moment of one column. Next we form the sum

$$\Delta x \sum_{x \text{ fixed}} \Delta y \int_{f_1(x,y)}^{f_2(x,y)} (x - x_0)^p\delta(x,y,z)\,dz, \tag{13.653}$$

where the part "x fixed" of the symbol serves to inform us that the sum contains only terms arising from columns that comprise a slab running from the cylinder on which $y = y_1(x)$ to the cylinder on which $y = y_2(x)$.

Thus (13.653) and

$$\Delta x \int_{g_1(x)}^{g_2(x)} dy \int_{f_1(x,y)}^{f_2(x,y)} (x - x_0)^p \delta(x,y,z)\, dz \tag{13.654}$$

are approximations to the required moment of one slab. Finally,

$$\sum \Delta x \int_{g_1(x)}^{g_2(x)} dy \int_{f_1(x,y)}^{f_2(x,y)} (x - x_0)^p \delta(x,y,z)\, dz \tag{13.655}$$

is an approximation to the required moment of the whole solid and replacing this by its limit gives the formula

$$M_{x=x_0}^{(p)} = \int_{a_1}^{a_2} dx \int_{g_1(x)}^{g_2(x)} dy \int_{f_1(x,y)}^{f_2(x,y)} (x - x_0)^p \delta(x,y,z)\, dz \tag{13.656}$$

for the required pth moment of the solid about the plane $x = x_0$.

Remarks very similar to those following (13.453) can now be made. Formulas for the pth moments of the solid about the planes $y = y_0$ and $z = z_0$ are obtained by replacing the factor $(x - x_0)^p$ by the factors $(y - y_0)^p$ and $(z - z_0)^p$ in the above derivations. In case $p = 0$, the factors $(x - x_0)^p$, $(y - y_0)^p$, $(z - z_0)^p$ are all 1 and the numbers $M_{x=x_0}^{(p)}$, $M_{y=y_0}^{(p)}$, $M_{z=z_0}^{(p)}$ are all equal to the mass M of the solid. Thus

$$M = \iiint_S \delta(x,y,z)\, dx\, dy\, dz, \tag{13.66}$$

and we can replace this triple integral by iterated integrals. Formulas very similar to (13.46) show that the formulas $M_{x=\bar{x}}^{(1)} = 0$, $M_{y=\bar{y}}^{(1)} = 0$, $M_{z=\bar{z}}^{(1)} = 0$ which determine the centroid $(\bar{x},\bar{y},\bar{z})$ of S can be put in the form

$$\bar{x} = \frac{\iiint_S x\delta(x,y,z)\, dx\, dy\, dz}{\iiint_S \delta(x,y,z)\, dx\, dy\, dz}, \qquad \bar{y} = \frac{\iiint_S y\delta(x,y,z)\, dx\, dy\, dz}{\iiint_S \delta(x,y,z)\, dx\, dy\, dz}, \tag{13.67}$$

$$\bar{z} = \frac{\iiint_S z\delta(x,y,z)\, dx\, dy\, dz}{\iiint_S \delta(x,y,z)\, dx\, dy\, dz},$$

where the denominators are equal to the mass of the solid.

For some purposes, the *polar moment of inertia* of a solid about a line L is of importance. When the line L is the line having the equations $x = x_0$, $y = y_0$, we may denote the polar moment of inertia about L by the symbol $M_{x=x_0,y=y_0}^{(2)}$ or $I_{x=x_0,y=y_0}$ and work out the formula

$$\begin{aligned} M_{x=x_0,y=y_0}^{(2)} &= I_{x=x_0,y=y_0} \\ &= \iiint_S [(x - x_0)^2 + (y - y_0)^2]\delta(x,y,z)\, dx\, dy\, dz \end{aligned} \tag{13.68}$$

which is analogous to (13.484). The formula

$$M_{x=x_0,y=y_0}^{(2)} = M_{x=x_0}^{(2)} + M_{y=y_0}^{(2)} \tag{13.681}$$

and the parallel axis theorem (Theorem 13.48) hold for solids as well as for lamina.

Problems 13.69

1 Set up and evaluate a threefold iterated integral for the volume V of the solid tetrahedron bounded by the coordinate planes and the plane having the equation

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1,$$

it being assumed that a, b, c are positive numbers. *Ans.:*

$$V = \int_0^a dx \int_0^{b\left(1-\frac{x}{a}\right)} dy \int_0^{c\left(1-\frac{x}{a}-\frac{y}{b}\right)} dz = \tfrac{1}{6}abc.$$

2 Supposing that f is continuous over the tetrahedron T of the preceding problem, set up a threefold iterated integral equal to the triple integral

$$\int_T f(x,y,z)\, dx\, dy\, dz,$$

where, as is often done when rectangular coordinates are involved, $dx\, dy\, dz$ is written instead of dT. *Ans.:* Same as answer to preceding problem except that the integrand is $f(x,y,z)$ instead of 1.

3 Set up a threefold iterated integral for the volume V of the solid in E_3 bounded by the parabolic cylinders having the equations $y = x^2$ and $x = y^2$ and the planes having the equations $z = 0$ and $x + y + z = 2$. *Ans.:*

$$V = \int_0^1 dx \int_{x^2}^{\sqrt{x}} dy \int_0^{2-x-y} dz.$$

4 A homogeneous cube has density δ and has edges of length a. Find its moment of inertia about an edge. *Ans.:* $\frac{2}{3}\delta a^5$ or $\frac{2}{3}Ma^2$, where M is the mass of the cube.

5 A long solid circular cylinder S of radius b has its axis on the y axis of an x, y, z coordinate system. A circular hole having radius a and having its axis on the z axis is drilled. Supposing that $0 < a \leqq b$, set up an integral for the volume V of the part of S that is drilled away. *Ans.:* Because of symmetry, V is 8 times the volume of the part in the first octant and

$$V = 8\int_0^a dx \int_0^{\sqrt{a^2-x^2}} dy \int_0^{\sqrt{b^2-x^2}} dz.$$

6 Let q be a nonnegative constant and let S be a spherical ball of radius R whose density is proportional to the qth power of the distance from the center. Taking the origin at the center of the ball, set up a triple integral for the polar moment of inertia of the ball about the z axis. Simplify matters by using the fact that the total moment is 8 times the moment of the part of the ball in the first octant. *Ans.:*

$$8\int_0^R dy \int_0^{\sqrt{R^2-y^2}} \int_0^{\sqrt{R^2-x^2-y^2}} (x^2+y^2)(x^2+y^2+z^2)^{q/2}\, dz.$$

Remark: Section 13.8 will enable us to avoid this and some other unpleasant integrals.

7 The tetrahedron (or pyramid) bounded by the planes $x = 0$, $y = 0$, $z = 0$, and

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1$$

has density (mass per unit volume) $\delta(x,y,z)$ at the point (x,y,z). Supposing that a, b, c are positive constants and that p is a nonnegative constant, set up two different iterated integrals for the pth moment $M_{x=0}^{(p)}$ of the solid about the plane $x = 0$. *Ans.:* One of the possibilities is

$$M_{x=x_0}^{(p)} = \int_0^a x^p \int_0^{b\left(1-\frac{x}{a}\right)} dy \int_0^{c\left(1-\frac{x}{a}-\frac{y}{b}\right)} \delta(x,y,z)\, dz.$$

8 For the case in which $\delta(x,y,z) = 1$, so that the solid is homogeneous, show that the last formula of Problem 7 can be put in the form

$$M_{x=0}^{(p)} = \frac{bc}{2}\int_0^a x^p \left(1 - \frac{x}{a}\right)^2 dx.$$

At least when $p = 0$ and $p = 1$, show that

$$M_{x=0}^{(p)} = \frac{p!}{(p+3)!} a^{p+1}bc = \frac{1}{(p+1)(p+2)(p+3)} a^{p+1}bc.$$

Finally, show that $\bar{x} = a/4$.

9 Let $0 < b \leqq a$. A spherical ball of radius a has its center at the origin. Set up a threefold iterated integral for the volume V of the part of the ball drilled away when a bit of radius b drills a cylindrical hole centered on the line having the equations $x = 0$, $y = a - b$. Symmetry may be used, and the integral need not be evaluated. *Ans.:*

$$V = 4\int_{a-2b}^{a} dy \int_0^{\sqrt{b^2-[x-(a-b)]^2}} dx \int_0^{\sqrt{a^2-x^2-y^2}} dz.$$

10 Show that when $b = a$, the answer to Problem 9 reduces, as it should, to an integral for the volume of the whole spherical ball.

11 Assuming that the spherical ball of Problem 9 has density $\delta(x,y,z)$ at $P(x,y,z)$, modify the integral of the preceding problem to obtain an integral for the amount by which the drilling of the hole decreases the polar moment of inertia about the z axis. *Ans.:* Multiply the invisible 1 preceding dz by the factor $\delta(x,y,z)(x^2 + y^2)$.

12 Modify the answer to Problem 9 to obtain an integral for the volume of the material drilled from the ball when the hole is centered on the line which meets the y axis where the surface of the ball does.

13 Supposing that $a > 0$, evaluate

$$\iiint_S \frac{1}{1 + x + y + z}\, dS,$$

where S is the cube having four of its vertices at the points (0,0,0), $(a,0,0)$, $(0,a,0)$, $(0,0,a)$. *Ans.:*

$\frac{1}{2}(1 + 3a)^2 \log(1 + 3a) - \frac{3}{2}(1 + 2a)^2 \log(1 + 2a) + \frac{3}{2}(1 + a)^2 \log(1 + a)$.

14 Supposing that a, b, c are positive numbers, evaluate

$$\iiint_T \sin\left\{\pi\left(\frac{x}{a} + \frac{y}{b} + \frac{z}{c}\right)\right\} dT,$$

where T is the tetrahedron having vertices at the points (0,0,0), $(a,0,0)$, $(0,b,0)$, $(0,0,c)$. *Ans.:* $(\pi^2 - 4)abc/2\pi^3$.

13.7 *Triple integrals; cylindrical coordinates* In some cases a solid S (or set S in E_3 having positive volume) and a function f defined over S are such that the triple integral defined as in (13.624) by the formula

$$\iiint_S f(P)\, dS = \lim \sum f(P)\, \Delta S \tag{13.71}$$

can advantageously be expressed in terms of cylindrical coordinates ρ, ϕ, z. When we use cylindrical coordinates, the set S is partitioned into subsets $S_1, S_2, \cdots, S_n$ by planes through the z axis having cylindrical equations $\phi = \phi_0$, $\phi = \phi_1, \cdots, \phi = \phi_{m}$, by circular cylinders having cylindrical equations $\rho = \rho_0$, $\rho = \rho_1, \cdots, \rho = \rho_{m'}$, and by planes parallel to the xy plane having the cylindrical equations $z = z_0$, $z = z_1, \cdots,$ $z = z_{m''}$. Figure 13.72 shows a typical subset containing a point having

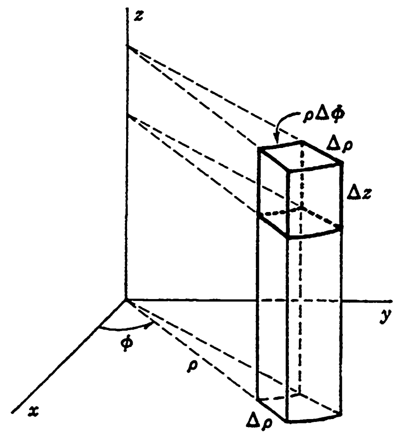


Figure 13.72

cylindrical coordinates ρ, ϕ, z. This subset has height Δz and, as we learned when studying polar coordinates, its base has area exactly or approximately equal to $\rho\, \Delta\phi\, \Delta\rho$. Thus we use the formula

$$\Delta S = \rho\, \Delta\phi\, \Delta\rho\, \Delta z \tag{13.73}$$

to put (13.71) in the form

$$(13.74) \qquad \iiint_S f(\rho,\phi,z)\rho \, d\phi \, d\rho \, dz = \lim \sum f(\rho,\phi,z)\rho \, \Delta\phi \, \Delta\rho \, \Delta z.$$

Assuming that f is bounded over S and that f is sufficiently continuous to make all of the integrals exist, we can use Theorem 13.63 to express this triple integral as an iterated integral. When limits of integration for iterated integrals are being determined, information obtained by looking at Figure 13.72 can be helpful. Adding subsets for which z varies (ρ and ϕ being fixed) yields a vertical column. Adding columns for which ρ varies (ϕ being fixed) yields a whole or a part of a wedge which in some cases looks like a conventional wedge of a cake or orange or lemon. Adding the wedges obtained for appropriate values of ϕ then gives the entire solid S. Results of performing summations and integrations in different orders are easily described. For example, adding subsets for which ϕ varies (a and ρ being fixed) yields all or part of a circular hoop or ring, and there are two ways in which these hoops can be added to yield more extensive parts of S.

Supposing that S is a right circular cylindrical solid bounded by the graphs of the equations $\rho = R$, $z = 0$, and $z = H$ and that the density (mass per unit volume) at the point having cylindrical coordinates (ρ,ϕ,z) is $\delta(\rho,\phi,z)$, we set up an integral for the polar moment of inertia $M^{(2)}_{y=0,z=0}$ of S about the x axis. For the volume ΔS of a subset of the

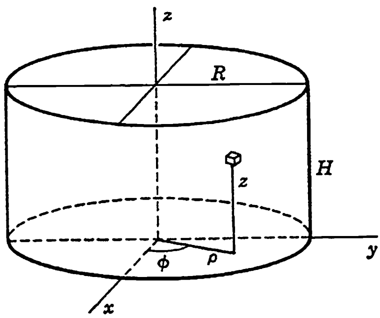


Figure 13.75

solid, which is shown in Figure 13.75, we use the formula

$$(13.76) \qquad \Delta S = \rho \, \Delta\phi \, \Delta\rho \, \Delta z.$$

To get the mass ΔM of the subset, we multiply by the density (mass per unit volume) $\delta(\rho,\phi,z)$ to obtain

$$(13.761) \qquad \Delta M = \delta(\rho,\phi,z)\rho \, \Delta\phi \, \Delta\rho \, \Delta z.$$

Then we must realize what we are trying to do and multiply this by

$$(13.762) \qquad z^2 + (\rho \sin \phi)^2,$$

the square of the distance from the subset to the x axis, to obtain the expression

$$[z^2 + \rho^2 \sin^2 \phi]\delta(\rho,\phi,z)\rho\, \Delta\phi\, \Delta\rho\, \Delta z \tag{13.763}$$

for the polar moment of the subset about the x axis. Adding these and taking the limit of the sum gives

$$M^{(2)}_{y=0,z=0} = \iiint_S [z^2 + \rho^2 \sin^2 \phi]\delta(\rho,\phi,z)\rho\, d\phi\, d\rho\, dz. \tag{13.764}$$

This is, as it should be, the sum of the moments of inertia of S about the two planes $z = 0$ and $y = 0$. The limits of integration being determined with the aid of Figure 13.75, we can obtain the iterated integral formula

$$M^{(2)}_{z=0,y=0} = \int_0^{2\pi} d\phi \int_0^R d\rho \int_0^H [z^2 + \rho^2 \sin^2 \phi]\delta(\rho,\phi,z)\rho\, dz. \tag{13.77}$$

In order to be able to evaluate this integral in decimal form, we must know R, H, and $\delta(\rho,\phi,z)$. For the special case in which q is a nonnegative constant and $\delta(\rho,\phi,z) = k\rho^q$, we can evaluate the integrals in terms of R, H, q, and k to obtain

$$M^{(2)}_{z=0,y=0} = k\left[2\pi \frac{H^3}{3}\frac{R^{q+2}}{q+2} + \pi H \frac{R^{q+4}}{q+4}\right]. \tag{13.78}$$

The result for the case in which the cylinder is homogeneous is obtained by setting $q = 0$.

Problems 13.79

1 Supposing that $0 < b \leqq a$, set up and evaluate a threefold iterated integral in cylindrical coordinates for the volume V of the solid lying inside the sphere and cylinder having the cylindrical equations $\rho^2 + z^2 = a^2$ and $\rho = b$. *Ans.:*

$$V = 8\int_0^{\pi/2} d\phi \int_0^b \rho\, d\rho \int_0^{\sqrt{a^2-\rho^2}} dz = \tfrac{4}{3}\pi[a^3 - (a^2 - b^2)^{3/2}].$$

2 Supposing that $0 < b < a$, set up and evaluate a threefold iterated integral for the mass M of the solid lying inside the sphere but outside the cylinder having the cylindrical equations $\rho^2 + z^2 = a^2$ and $\rho = b$, it being assumed that the density of the body at $P(\rho,\phi,z)$ is $|z|$. *Ans.:*

$$M = 8\int_0^{\pi/2} d\phi \int_b^a \rho\, d\rho \int_0^{\sqrt{a^2-\rho^2}} z\, dz = \frac{\pi}{2}(a^2 - b^2)^2.$$

3 Set up and evaluate a threefold iterated integral in cylindrical coordinates for the volume V of the solid bounded by the sphere and cylinder having the cylindrical equations $\rho^2 + z^2 = a^2$ and $\rho = a \cos \phi$. *Ans.:*

$$\begin{aligned} V &= 4\int_0^{\pi/2} d\phi \int_0^{a\cos\phi} \rho\, d\rho \int_0^{\sqrt{a^2-\rho^2}} dz \\ &= \tfrac{4}{3}a^3 \int_0^{\pi/2} (1 - \sin^3\phi)\, d\phi = \tfrac{4}{3}a^3\left[\frac{\pi}{2} - \frac{2}{3}\right]. \end{aligned}$$

4 Assuming that the solid cone shown in the upper part of Figure 13.791 has density (mass per unit volume) $\delta(\rho,\phi,z)$ at the point having cylindrical coordinates (ρ,ϕ,z), set up an iterated integral for the moment of inertia of the cone

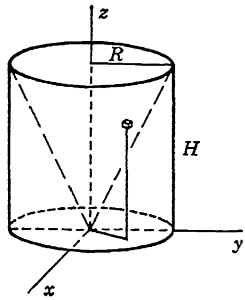


Figure 13.791

about the x axis. Then calculate the required moment for the case in which the density is proportional to the qth power of the distance from the axis of the cone, that is, $\delta(\rho,\phi,z) = k\rho^q$. *Ans.:* The integral is the same as that in (13.77) except that the lower limit of integration with respect to z is $(H/R)\rho$. The required moment is

$$k\left[\frac{2\pi}{(q+2)(q+5)}H^3R^{q+2} + \frac{\pi}{(q+4)(q+5)}HR^{q+4}\right].$$

5 Show how the preceding problem gives the conclusion that the moment of inertia of a homogeneous conical solid, having density δ and height H and base radius R, about a line through the vertex perpendicular to the axis is $\frac{1}{20}\pi R^2H(4H^2 + R^2)\delta$.

6 A solid cylinder having constant density δ is bounded by the cylinder and planes having the equations $\rho = a$, $z = 0$, and $z = h$. Set up and evaluate a threefold iterated integral in cylindrical coordinates for the moment of inertia I of the solid about the x axis. *Ans.:*

$$I = \delta\int_0^{2\pi} d\phi \int_0^a \rho\, d\rho \int_0^h (z^2 + \rho^2\sin^2\phi)\, dz = \pi a^2h\delta\left[\frac{h^2}{3} + \frac{a^2}{4}\right].$$

7 A conical solid has height h, base radius a, and density kz, where k is a constant and z is the distance from the base. Find the mass M of the solid and the distance $\bar{z}$ from the base to the centroid. *Ans.:* $M = \frac{1}{12}\pi a^2h^2k$, $z = \frac{2}{5}h$.

8 A conical solid has height h, base radius a, and density $k\rho$, where k is a constant and ρ is the distance from the axis. Find the mass M of the solid and the distance $\bar{z}$ from the base to the centroid. *Ans.:* $M = \frac{1}{6}\pi a^3hk$, $\bar{z} = \frac{1}{5}h$.

9 A cup-shaped solid S is obtained by rotation about the z axis of a region R in the yz plane bounded by the graphs of the equations

$$z = x^2, \qquad z = x^2 + 1, \qquad z = 10.$$

The density (mass per unit volume) of S at the point having cylindrical coordinates ρ, ϕ, z is $\delta(\rho,\phi,z)$. Set up a threefold iterated integral in polar coordinates for the total mass M of the solid. *Ans.:*

$$M = \int_0^{2\pi} d\phi \int_0^3 \rho\, d\rho \int_{\rho^2}^{\rho^2+1} \delta(\rho,\phi,z)\, dz + \int_0^{2\pi} d\phi \int_3^{\sqrt{10}} \rho\, d\rho \int_{\rho^2}^{10} \delta(x,y,z)\, dz.$$

10 As we near the end of our textbook, we can and should review and summarize some of our ideas about integrals. This problem does not require us to produce a specific numerical answer to a specific problem; it requires us to think in general terms about methods by which such answers are produced. With the understanding that the ideas have applications to more complicated situations as well as to simpler ones, we consider the gravitational force $\mathbf{F}$ exerted upon a particle of mass m at a point Q in E_3 by a body B. This body B may be one-dimensional, that is, it may be concentrated upon a one-dimensional set S which may be a line segment or an arc of a curve having positive length. In this case we suppose that the body has linear (or one-dimensional) density $\delta(P)$ at the point P in S. The body B may be two-dimensional, that is, it may be concentrated upon a two-dimensional set S which may be a circular disk or some other region (on a plane or curved surface) which has positive area. In this case, we suppose that the body has areal (or two-dimensional) density $\delta(P)$ at the point P in S. Finally, the body B may be an ordinary three-dimensional solid body, that is, it may occupy a set S in E_3 having positive volume. In this case we suppose the body has ordinary (mass per unit volume) density $\delta(P)$ at the point P in S. We simplify and unify our discussion of these things by considering length to be *one-dimensional measure*, area to be *two-dimensional measure*, and volume to be *three-dimensional measure*. Thus we handle all of our examples together by saying that we have a body B occupying an n-dimensional set S in E_3 having positive n-dimensional measure $|S|$ and that the body has n-dimensional density $\delta(P)$ at the point P in S. The integer n may be 1 or 2 or 3. To start the process of calculating $\mathbf{F}$, we make a partition of the set S into q (note that n has been preempted) measurable subsets $S_1, S_2, \cdots, S_q$. It would be thoroughly reasonable to denote the measures of these sets by $|S_1|, |S_2|, \cdots, |S_q|$, but we find it convenient to denote the measures by $\Delta S_1, \Delta S_2, \cdots, \Delta S_q$. Thus, for each k, ΔS_k is not a part of S; it is the measure† of a part of S. The next step is to select a point P_k (or P_k^*) in S_k. Note that P_k is not in ΔS_k but that P_k is a point in the set of which ΔS_k is the measure. We use the number $\delta(P_k)\, \Delta S_k$ to approximate the mass of the part of the body occupying the set S_k. For each k, we apply the Newton inverse square law to obtain the force $\Delta \mathbf{F}_k$ which a particle of mass $\delta(P_k)\, \Delta S_k$ at P_k exerts upon the particle of mass m at Q. Even small figures can be helpful, and we can look at Figure 13.792. We find that

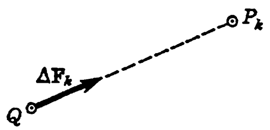


Figure 13.792

$$\Delta \mathbf{F}_k = \frac{Gm\delta(P_k)\, \Delta S_k}{|\overrightarrow{QP_k}|^2} \frac{\overrightarrow{QP_k}}{|\overrightarrow{QP_k}|}, \tag{1}$$

† Perhaps we should recognize the fact that the very useful standard notation is a relic of the good old days when it was not the fashion to recognize a difference between a set and the measure of the set.

where G is the gravitational constant and the last factor is a unit vector in the direction of $\overrightarrow{QP_k}$. Making a slight modification of the right member of (1) and adding give the right member of the formula

$$\text{RS} = Gm \sum_{k=1}^{q} \frac{\delta(P_k)\overrightarrow{QP_k}}{|\overrightarrow{QP_k}|^3} \Delta S_k, \tag{2}$$

which is a Riemann sum (RS) formed for the vector function having the value

$$\frac{\delta(P)\overrightarrow{QP}}{|\overrightarrow{QP}|^3} \tag{3}$$

at the point P of our set S. When we are in a hurry, we use the idea that the Riemann sum should be near the force $\mathbf{F}$ whenever the norm of the partition is small as a basis for introducing the definition

$$\mathbf{F} = \lim Gm \sum_{k=1}^{q} \frac{\delta(P_k)\overrightarrow{QP_k}}{|\overrightarrow{QP_k}|^3} \Delta S_k. \tag{4}$$

Since the limit of the Riemann sums is a Riemann integral, we write

$$\mathbf{F} = Gm \int_S \frac{\delta(P)\overrightarrow{QP}}{|\overrightarrow{QP}|^3}\, dS, \tag{5}$$

the symbol on the right being an orthodox symbol for the integral. When we wish our notation to be as informative as possible, we can use n integral signs when S is n-dimensional. The fact that different notations are used at different times need not disturb us, because in any particular application we can be expected to know the dimensionality of the set we partition. We can, when we are unhurried, be more precise about the meanings of (4) and (5). The integral is, when it exists, the one and only vector $\mathbf{F}$ such that to each positive number ϵ there corresponds a positive number δ such that

$$\left| \mathbf{F} - Gm \sum_{k=1}^{q} \frac{\delta(P_k)\overrightarrow{QP_k}}{|\overrightarrow{QP_k}|^3} \Delta S_k \right| < \epsilon \tag{6}$$

whenever the Riemann sum is formed for a partition whose norm is less than δ. For some purposes, it is important to observe that (5) is an intrinsic formula which does not depend upon any one particular coordinate system which may be used to specify the positions of the points involved in the problem. For other purposes, particularly when problems in elementary books are being solved, it is necessary to introduce a coordinate system. The *raison d'être* of different coordinate systems lies in the fact that different ones are most useful in different situations.

13.8 Triple integrals; spherical coordinates In some cases a solid S (or a set S in E_3 having positive volume) and a function f defined over S are such that the triple integral defined by the formula

$$\iiint_S f(P)\, dS = \lim \sum f(P)\, \Delta S \tag{13.81}$$

can advantageously be expressed in terms of spherical coordinates r, ϕ, θ. When we use spherical coordinates, the set S is partitioned into subsets $S_1, S_2, \cdots, S_n$ by spheres having spherical equations $r = r_0$, $r = r_1$, $\cdots$, $r = r_m$, by half-planes having spherical equations $\phi = \phi_0$, $\phi = \phi_1$, $\cdots$, $\phi = \phi_{m'}$, and by half-cones having the spherical equations $\theta = \theta_0$, $\theta = \theta_1$, $\cdots$, $\theta = \theta_{m''}$. Figure 13.82 shows a typical subset containing

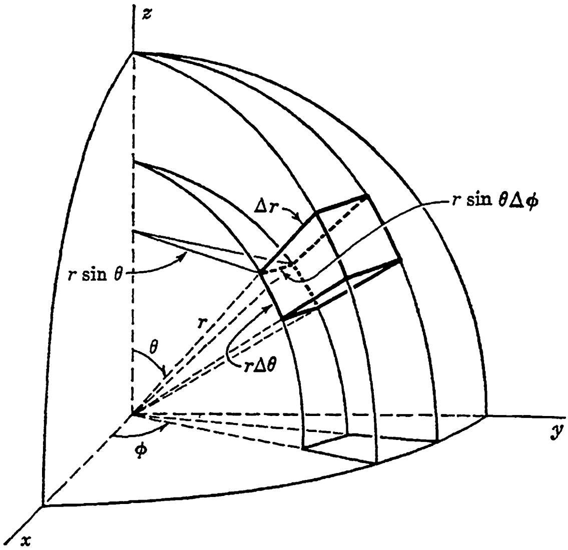


Figure 13.82

a point having spherical coordinates (r,ϕ,θ). When $r > 0$ and the numbers Δr, $\Delta\phi$, $\Delta\theta$ are all small, this subset closely approximates a rectangular parallelepiped one dimension of which is Δr, the difference of the radii of two spheres. The inner (or outer) face perpendicular to the ray from the origin to the point (r,ϕ,θ) closely approximates a rectangle one side of which has length $r\,\Delta\theta$ (the length of the arc of a sector having radius r and central angle $\Delta\theta$) and the perpendicular sides of which have length $r \sin \theta\, \Delta\phi$ (the length of the arc of a sector having radius $r \sin \theta$ and central angle $\Delta\phi$). Thus the area of the face is approximately $r^2 \sin \theta\, \Delta\phi\, \Delta\theta$. Thus we use the formula

(13.83) $$\Delta S = r^2 \sin \theta\, \Delta\phi\, \Delta\theta\, \Delta r$$

(which, depending upon the choice of r, ϕ, θ, is exactly or approximately correct) to put (13.81) in the form

(13.84) $$\iiint_S f(r,\phi,\theta)r^2 \sin \theta\, d\phi\, d\theta\, dr = \lim \sum f(r,\theta,\phi)r^2 \sin \theta\, \Delta\phi\, \Delta\theta\, \Delta r.$$

Assuming that f is bounded over S and that f is sufficiently continuous to make all of the integrals exist, we can use Theorem 13.63 to express this triple integral as an iterated integral. When limits of integration for iterated integrals are to be determined, information obtained by looking at Figure 13.82 can be helpful. Adding subsets for which r varies (ϕ and θ being fixed) yields a spike or tapered column. Adding spikes for which θ varies (ϕ being fixed) yields a whole or a part of a wedge which in some cases looks like a conventional wedge of an orange or lemon or cake. Adding the wedges obtained for the appropriate values of ϕ then gives the entire solid S. Articulate persons can describe results of performing summations and integrations in other orders. For example, adding subsets for which ϕ varies (r and θ being fixed) yields all or a part of a circular hoop or ring, and there are two ways in which these hoops can be added to yield more extensive parts of S.

Supposing that S is a spherical ball having center at the origin and radius R and that the density (mass per unit volume) at the point having spherical coordinates (r,ϕ,θ) is $\delta(r,\phi,\theta)$, we set up an integral for the polar moment of inertia $M^{(2)}_{x=0,y=0}$ of the ball about the z axis. For the volume ΔS of a subset of the ball, we use the formula

$$\Delta S = r^2 \sin \theta \,\Delta\phi\, \Delta\theta\, \Delta r \tag{13.85}$$

which, like the telephone number of a dentist, is sometimes needed but is usually not permanently remembered. To get the mass ΔM of the subset, we multiply by the density (mass per unit volume) $\delta(r,\phi,\theta)$ to obtain

$$\Delta M = \delta(r,\phi,\theta)r^2 \sin \theta \,\Delta\phi\, \Delta\theta\, \Delta r. \tag{13.86}$$

Then we must be wise and strong enough to multiply this by $(r \sin \theta)^2$, the square of the distance from the subset to the z axis, to obtain the expression

$$\delta(r,\phi,\theta)r^4 \sin^3 \theta \,\Delta\phi\, \Delta\theta\, \Delta r \tag{13.861}$$

for the polar moment of the subset about the z axis. Adding these and taking the limit of the sum gives

$$M^{(2)}_{x=0,y=0} = \iiint_S \delta(r,\phi,\theta)r^4 \sin^3 \theta \, d\phi\, d\theta\, dr. \tag{13.87}$$

Since S is an entire sphere with center at the origin and radius R,

$$M^{(2)}_{x=0,y=0} = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \int_0^{R} \delta(r,\phi,\theta)r^4 \sin^3 \theta \, dr. \tag{13.88}$$

Problems 13.89

1 Let S be a spherical ball of radius a. Supposing that the center of the sphere is at the origin and that the density is $\delta(r,\phi,\theta)$ at the point having spherical coordinates (r,ϕ,θ), set up an iterated integral for the mass M. Arrange the order of integration in such a way that the last integration is with respect to r and simplify the result as much as possible for the special case in which the density is a function of r alone, say $\delta(r,\phi,\theta) = f(r)$. *Ans.:*

$$M = \int_0^a dr \int_0^\pi d\theta \int_0^{2\pi} \delta(r,\phi,\theta)r^2 \sin\theta\, d\phi, \qquad M = 4\pi \int_0^a r^2 f(r)\, dr.$$

2 Show how the last result of Problem 1 can be obtained by direct use of spherical shells and without use of iterated integrals.

3 Solve the modification of Problem 1 in which S is a spherical shell bounded by concentric spheres having radii r_1 and r_2.

4 Let q be a nonnegative constant and let S be a spherical ball of radius a whose density is proportional the qth power of the distance from the center. Using the formula

$$r^2 \sin\theta\, \Delta\phi\, \Delta\theta\, \Delta r$$

for volume in spherical coordinates and taking the origin at the center of the ball, set up and evaluate a triple integral for the polar moment of inertia of S about the z axis. *Ans.:* The triple integral is obtained by setting $\delta(r,\phi,\theta) = kr^q$ in (13.88). The required moment is $\dfrac{8\pi k a^{q+5}}{3(q+5)}$.

5 Show how the preceding problem gives the conclusion that the moment of inertia, about a diameter, of a spherical ball having radius a and uniform density δ is $\frac{8}{15}\pi a^5\delta$.

6 A solid spherical ball of radius a has, at each point P, density equal to the product of the distances from P to the origin and to the axis from which θ is measured. Set up and evaluate a threefold iterated integral in spherical coordinates for the mass M of the ball. *Ans.:*

$$M = \int_0^{2\pi} d\phi \int_0^\pi \sin^2\theta\, d\theta \int_0^a r^4\, dr = \tfrac{1}{5}\pi^2 a^5.$$

7 This problem involves lengths of curves. Suppose that, as time t increases from a to b, a particle P moves along a curve C in such a way that its rectangular coordinates x, y, z, its cylindrical coordinates ρ, ϕ, z, and its spherical coordinates r, ϕ, θ are all functions of t having continuous derivatives. Start with the formula

$$(1) \qquad L = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt$$

giving the length L of the curve C as an integral involving rectangular coordinates. Use the formulas

$$(2) \qquad x = \rho\cos\phi, \qquad y = \rho\sin\phi, \qquad z = z$$

to obtain the formula

$$(3) \qquad L = \int_a^b \sqrt{\rho^2\left(\frac{d\phi}{dt}\right)^2 + \left(\frac{d\rho}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt$$

giving L as an integral involving cylindrical coordinates. Then use the formulas

$$(4) \qquad \rho = r\sin\theta, \qquad \phi = \phi, \qquad z = r\cos\theta$$

to obtain the formula

$$(5)\qquad L = \int_a^b \sqrt{r^2 \sin^2\theta\left(\frac{d\phi}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2 + \left(\frac{dr}{dt}\right)^2}\,dt$$

giving L as an integral involving spherical coordinates.

8 Using ideas and formulas from the preceding problem, start with the formula

$$(1)\qquad \mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

for the vector running from the origin to P at time t. Show that

$$(2)\qquad \mathbf{r} = r(\cos\phi\sin\theta\mathrm{i} + \sin\phi\sin\theta\mathbf{j} + \cos\theta\mathbf{k}).$$

Show that the velocity at time t is

$$(3)\qquad \mathbf{v} = \frac{dr}{dt}\mathbf{u}_1 + r\frac{d\theta}{dt}\mathbf{u}_2 + r\sin\theta\frac{d\phi}{dt}\mathbf{u}_3,$$

where

$$\begin{aligned}
&(4)\qquad \mathbf{u}_1 = \cos\phi\sin\theta\mathbf{i} + \sin\phi\sin\theta\mathbf{j} + \cos\theta\mathbf{k}\\
&(5)\qquad \mathbf{u}_2 = \cos\phi\cos\theta\mathbf{i} + \sin\phi\cos\theta\mathbf{j} - \sin\theta\mathbf{k}\\
&(6)\qquad \mathbf{u}_3 = -\sin\phi\mathbf{i} + \cos\phi\mathbf{j}\qquad .
\end{aligned}$$

Prove that the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, in that order, constitute a right-handed orthonormal system. Show, finally, that

$$(7)\qquad |\mathbf{v}| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{r^2\sin^2\theta\left(\frac{d\phi}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2 + \left(\frac{dr}{dt}\right)^2}.$$

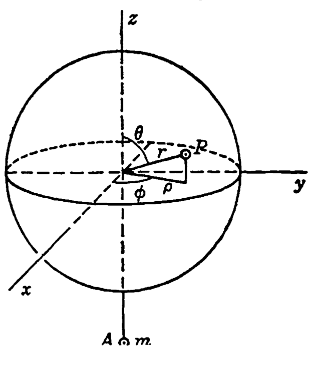
9 With the aid of hints that may be gleaned from the two preceding problems, tell the meanings of the things in the formula

$$\frac{ds}{dt} = \sqrt{r^2\sin^2\theta\left(\frac{d\phi}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2 + \left(\frac{dr}{dt}\right)^2}$$

and give conditions under which the formula is valid.

10 This problem provides an opportunity to fill in details and gain an understanding of a method by which a great problem in cosmology is solved with the aid of vectors and integrals that involve rectangular and spherical coordinates. Figure 13.891 shows a spherical ball S of radius a having its center at the origin.

Figure 13.891



It is supposed that the sphere has density $\delta(r,\phi,\theta)$ at the point P having spherical coordinates (r,ϕ,θ). A particle of mass m is supposed to be concentrated at a point A which lies outside the ball and on the negative z axis, the rectangular coordinates of A being $(0,0,-D)$, where $D > a$. We are required to determine and learn something about the gravitational force $\mathbf{F}$ upon the particle of mass m that is produced by the ball. We start with a basic idea of Newton that lies at the foundation of classical science. If particles of masses m and ΔM are located at points A and P, then each particle pulls the other toward it with a force of magnitude $Gm\,\Delta M/|\overrightarrow{AP}|^2$, where G is a universal gravitational constant whose numerical value depends only upon the units of force and distance that are used. The actual force upon the particle of mass m is obtained by multiplying this magnitude by $\overrightarrow{AP}/|\overrightarrow{AP}|$, the unit vector which has its tail at A and is pointed toward P. Letting $\Delta\mathbf{F}$ denote this force, we have

(1) $$\Delta\mathbf{F} = Gm\frac{\overrightarrow{AP}}{|\overrightarrow{AP}|^3}\,\Delta M.$$

We need a useful formula for the vector $\overrightarrow{AP}$. The rectangular coordinates x, y, z, the cylindrical coordinates ρ, ϕ, z, and the spherical coordinates r, ϕ, θ of the point P are related by the formulas

$$\begin{aligned} x &= \rho\cos\phi = r\sin\theta\cos\phi \\ y &= \rho\sin\phi = r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}$$

which can be found in this and other books and can be derived from Figure 13.891. Since A has rectangular coordinates $0, 0, -D$, we find that

(2) $$\overrightarrow{AP} = r\sin\theta\cos\phi\mathbf{i} + r\sin\theta\sin\phi\mathbf{j} + (D + r\cos\theta)\mathbf{k}$$

and hence that

(3) $$|\overrightarrow{AP}| = (D^2 + 2Dr\cos\theta + r^2)^{1/2}.$$

Using spherical coordinates, we employ the right member of the formula

(4) $$\Delta M = \delta(r,\phi,\theta)r^2\sin\theta\,\Delta r\,\Delta\phi\,\Delta\theta$$

to approximate the mass ΔM of a subset (or element) of the ball containing the point P. Substituting in (1) gives the formula

(5) $$\Delta\mathbf{F} = Gm\frac{\delta(r,\phi,\theta)r^2\sin\theta}{(D^2 - 2Dr\cos\theta + r^2)^{3/2}}\overrightarrow{AP}\,\Delta r\,\Delta\phi\,\Delta\theta$$

in which the right side is an approximation to the force upon the particle of mass m produced by one subset of the ball. In (5) and some of the following formulas, $\overrightarrow{AP}$ is written instead of the right member of (2) to save time and paper. Supposing that the density function δ is a reasonably decent function, we employ partitions of the ball and principles of the integral calculus to obtain

$$\mathbf{F} = \lim \Sigma \Delta\mathbf{F} \tag{6}$$

and

$$\mathbf{F} = Gm \iiint_S \frac{\delta(r,\phi,\theta)r^2 \sin\theta}{(D^2 + 2Dr\cos\theta + r^2)^{3/2}} \overrightarrow{AP}\, dr\, d\phi\, d\theta, \tag{7}$$

where the integral is a triple integral, S is the ball or the portion of E_3 occupied by the ball, and $\mathbf{F}$ is the total force on the particle of mass m. The formula

$$\mathbf{F} = Gm \int_0^a r^2\, dr \int_0^\pi \frac{\sin\theta}{(D^2 + 2Dr\cos\theta + r^2)^{3/2}}\, d\theta \int_0^{2\pi} \delta(r,\phi,\theta)\overrightarrow{AP}\, d\phi \tag{8}$$

shows one of the six ways in which $\mathbf{F}$ can be represented as an iterate integral.

The first phase of our work is done, and we proceed to see how (8) can be simplified when the density $\delta(r,\phi,\theta)$ is independent of ϕ so that $\delta(r,\phi,\theta) = \delta_1(r,\theta)$, where δ_1 is a function of r and θ only. In this case the last integral in (8) is

$$\int_0^{2\pi} \delta_1(r,\theta)[r\sin\theta\cos\phi\mathbf{i} + r\sin\theta\sin\phi\mathbf{j} + (D + r\cos\theta)\mathbf{k}]\, d\phi. \tag{9}$$

With the aid of the fact that $\int_0^{2\pi} \sin\phi\, d\phi = 0$ and $\int_0^{2\pi} \cos\phi\, d\phi = 0$, we see that this reduces to

$$2\pi\delta_1(r,\theta)(D + r\cos\theta)\mathbf{k} \tag{10}$$

and hence that (8) reduces to

$$\mathbf{F} = 2\pi Gm\mathbf{k} \int_0^a r^2\, dr \int_0^{2\pi} \frac{\delta_1(r,\theta)(D + r\cos\theta)\sin\theta}{(D^2 + 2Dr\cos\theta + r^2)^{3/2}}\, d\theta. \tag{11}$$

This shows that, when the density is independent of ϕ, the components of $\mathbf{F}$ in the directions of the x and y axes are 0. Of course, wise scientists always claim that this must be true "on account of symmetry."

Our final step is to make additional simplification of (11) for the case in which the density is a function of r only, so that $\delta_1(r,\theta) = \delta_2(r)$ and the ball is said to be radially homogeneous. One reason for interest in this case lies in the fact that suns and planets and moons are closely approximated by radially homogeneous balls unless rapid rotations about their axes produce nontrivial equatorial bulges. When $\delta_1(r,\theta) = \delta_2(r)$, we can put (11) in the form

$$\mathbf{F} = 2\pi Gm\mathbf{k} \int_0^a r\delta_2(r)f(r)\, dr \tag{12}$$

where

$$f(r) = \int_0^\pi \frac{(D + r\cos\theta)r\sin\theta}{(D^2 + 2Dr\cos\theta + r^2)^{3/2}}\, d\theta. \tag{13}$$

The integral in (13) may seem to be quite impenetrable until its fundamental weakness is discovered. If we set $u = r\cos\theta$, then (for each fixed r) $du = -r$

sin $\theta\, d\theta$ and, except for algebraic sign, the integrand becomes that in

$$f(r) = \int_{-r}^{r} \frac{D + u}{(D^2 + 2Du + r^2)^{3/2}}\, du. \tag{14}$$

When $\theta = 0$, we have $u = r$, and when $\theta = \pi$, we have $u = -r$, and we see that (14) is correct when we see that $-\int_{r}^{-r} = \int_{-r}^{r}$. We can discover that trading (14) for (13) was good business if we know or discover or are told that (14) can be demolished by the substitution

$$v = D^2 + 2Du + r^2, \qquad u = \frac{v - D^2 - r^2}{2D}$$

$$D + u = \frac{v + D^2 - r^2}{2D}, \qquad du = \frac{1}{2D}\, dv.$$

Since $v = (D - r)^2$ when $u = -r$ and $v = (D + r)^2$ when $u = r$, this substitution gives

$$\begin{aligned} f(r) &= \frac{1}{4D^2} \int_{(D-r)^2}^{(D+r)^2} \frac{v + D^2 - r^2}{v^{3/2}}\, dv \\ &= \frac{1}{4D^2} \int_{(D-r)^2}^{(D+r)^2} [v^{-1/2} + (D^2 - r^2)v^{-3/2}]\, dv. \end{aligned} \tag{15}$$

Since $0 < r < D$, this gives $f(r) = 2r/D^2$. Substitution in (12) then gives the first equality in

$$\mathbf{F} = \frac{Gm\mathbf{k}}{D^2} \int_0^a 4\pi r^2 \delta_2(r)\, dr = \frac{GmM}{D^2}\, \mathbf{k}. \tag{16}$$

As was shown in the first problem in this list, the integral in (16) is the total mass M of the sphere, and hence the second equality holds. The result embodied in (16) is the following famous theorem. *If S is a radically homogeneous spherical ball, then the gravitational force which S exerts upon a particle outside S is equal to the force resulting from the assumption that the total mass of S is concentrated at the center of S.* To help us understand the significance of this result, we should know some history. It is said that Newton mistrusted his whole theory of gravitational attraction (and therefore delayed publication of his theory for 20 years) until he was able to prove the theorem.

11 If S is a radially homogeneous spherical shell, then the gravitational force $\mathbf{F}$ which S exerts upon a particle inside S is 0. All scientists should know this fact and some should, when a suitable occasion comes, earn the satisfaction of discovering the modifications that must be made in the work of Problem 10 to prove the fact.

12 With the aid of results of Problems 10 and 11, suppose that the earth is a homogeneous spherical ball and discuss the gravitational force upon a particle at the bottom of a very deep well.

Appendix 1 -

Proofs of basic theorems on limits

This appendix contains proofs of the basic theorems on limits which were given without proof in Section 3.2. Persons having competence in mathematical analysis must know these theorems and be able to give their proofs as thoroughly and as expertly as competent violin and piano players know and can play their scales. Individuals having nontrivial mathematical ambitions must therefore study the material of this appendix more than once. Most of the proofs depend upon the fundamental fact that if x and y are numbers, then

$$(1) \qquad |x + y| \leqq |x| + |y|, \qquad |x - y| \leqq |x| + |y|.$$

The first basic theorem shows us that there can be at most one number L for which $\lim_{x\to a} f(x) = L$.

Theorem A

$$\text{(2)} \qquad \textit{If } \lim_{x\to a} f(x) = L_1,\ \lim_{x\to a} f(x) = L_2 \textit{ then } L_2 = L_1.$$

Let ϵ be a given positive number. Then $\epsilon/2$ is also a positive number which we could call ϵ_1. The first hypothesis of the theorem implies that there is a positive constant δ_1, such that

$$\text{(3)} \qquad |f(x) - L_1| < \frac{\epsilon}{2}$$

whenever $x \neq a$ and $|x - a| < \delta_1$. The second hypothesis implies that there is a positive number δ_2 such that

$$\text{(4)} \qquad |f(x) - L_2| < \frac{\epsilon}{2}$$

whenever $x \neq a$ and $|x - a| < \delta_2$. Let δ be the lesser of δ_1 and δ_2. Then, when $x \neq a$ and $|x - a| < \delta$, the two inequalities (3) and (4) both hold and hence

$$\text{(5)} \qquad \begin{aligned} |L_2 - L_1| &= |[f(x) - L_1] - [f(x) - L_2]| \\ &\leqq |f(x) - L_1| + |f(x) - L_2| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

If we suppose that $|L_2 - L_1| \neq 0$, then we could let ϵ be the positive number $\frac{1}{2}|L_2 - L_1|$ and reach the false conclusions that $|L_2 - L|_1 < \frac{1}{2}|L_2 - L_1|$ and $1 < \frac{1}{2}$ and $2 < 1$. Therefore, $|L_2 - L_1| = 0$ and hence $L_2 = L_1$. This proves Theorem A. The last part of the proof involves a principle that is very often used. If h is a number and if $|h| < \epsilon$ whenever $\epsilon > 0$, then $h = 0$.

Theorem B

If b is a constant, then

$$\text{(6)} \qquad \lim_{x\to a} b = b.$$

This theorem tells us that if $f(x) = b$, where b is a constant, then

$$\text{(7)} \qquad \lim_{x\to a} f(x) = b.$$

To prove the result, let $\epsilon > 0$. Since $f(x) - b = 0$ for each x, we can let $\delta = \epsilon$ and conclude that $|f(x) - b| < \epsilon$ when $x \neq a$ and $|x - a| < \delta$.

Theorem C

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

This theorem tells us that if $f(x) = x$, then

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

To prove the theorem, we observe that if $\epsilon > 0$ and we set $\delta = \epsilon$, then $|x - a| < \epsilon$ whenever $0 < |x - a| < \delta$.

Theorem D

If c is a constant, then

(10) $$\lim_{x \to a} cf(x) = c \lim_{x \to a} f(x)$$

provided the limit on the right exists.

To prove this theorem, let $\lim_{x \to a} f(x) = L$. In case $c = 0$, the result is a consequence of the fact that both sides of (10) are 0. In case $c \neq 0$, let $\epsilon > 0$ and choose a positive number δ such that

(11) $$|f(x) - L| < \frac{\epsilon}{|c|}$$

when $|x| \neq a$ and $|x - a| < \delta$. Then

(12) $$|cf(x) - c \lim f(x)| < \epsilon$$

when $x \neq a$ and $|x - a| < \delta$. This proves (10).

Theorem E

The formulas

(13) $$\lim_{x \to a} [f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$$

(14) $$\lim_{x \to a} [f(x)g(x)] = [\lim_{x \to a} f(x)][\lim_{x \to a} g(x)]$$

(15) $$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$$

are valid provided the limits on the right exist and, in the case of the last formula, $\lim_{x \to a} g(x) \neq 0$.

To prove these results, let

(16) $$\lim_{x \to a} f(x) = L, \qquad \lim_{x \to a} g(x) = M.$$

Let ϵ and ϵ_1 and ϵ_2 be positive numbers. Choose positive numbers δ_1 and δ_2 such that

$$|f(x) - L| < \epsilon_1 \qquad (0 < |x - a| < \delta_1), \tag{17}$$
$$|g(x) - M| < \epsilon_2 \qquad (0 < |x - a| < \delta_2). \tag{18}$$

Let δ be the lesser of δ_1 and δ_2. Then, when $0 < |x - a| < \delta$, we have

$$\begin{aligned} |[f(x) + g(x)] - [L + M]| &= |[f(x) - L] + [g(x) - M]| \\ &\leqq |f(x) - L| + |g(x) - M| < \epsilon_1 + \epsilon_2. \end{aligned} \tag{19}$$

If we set $\epsilon_1 = \epsilon_2 = \epsilon/2$, then (16) and (19) give

$$|[f(x) + g(x)] - [\lim_{x\to a} f(x) + \lim_{x\to a} g(x)]| < \epsilon \tag{20}$$

when $0 < |x - a| < \delta$. This proves (13).

To prove (14), we bridge the gap between $f(x)g(x)$ and LM by subtracting and adding the term $f(x)M$ and using (17) and (18) to obtain

$$\begin{aligned} |f(x)g(x) - LM| &= |[f(x)g(x) - f(x)M] + [f(x)M - LM]| \\ &\leqq |f(x)|\,|g(x) - M| + |f(x) - L|\,|M| \\ &\leqq (|L| + \epsilon_1)\epsilon_2 + \epsilon_1 M \end{aligned} \tag{21}$$

when $0 < |x - a| < \delta$. If we choose ϵ_1 such that $\epsilon_1|M| < \epsilon/2$ and afterward choose ϵ_2 such that $(|L| + \epsilon_1)\epsilon_2 < \epsilon/2$, then (21) gives

$$|f(x)g(x) - LM| < \epsilon \qquad (0 < |x - a| < \delta). \tag{22}$$

This proves (14). To prove (15), we begin by proving that

$$\lim_{x\to a} \frac{1}{g(x)} = \frac{1}{M} \tag{23}$$

when $\lim_{x\to a} g(x) = M$ and $M \neq 0$. The more general result (15) will then follow from (14) and the fact that

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} [f(x)] \frac{1}{[g(x)]} = [\lim_{x\to a} f(x)] \left[\lim_{x\to a} \frac{1}{g(x)} \right]. \tag{24}$$

To prove (23), we suppose that $M \neq 0$, that ϵ_2 has been chosen such that $0 < \epsilon_2 < |M|/2$, and that δ has been chosen such that

$$|g(x) - M| < \epsilon_2 \tag{25}$$

whenever $0 < |x - a| < \delta$. Then

$$|M| = |g(x) - M - g(x)| \leqq |g(x) - M| + |g(x)| < \epsilon_2 + |g(x)| \tag{26}$$

and hence

$$|g(x)| > |M| - \epsilon_2 > M - \frac{M}{2} = \frac{M}{2} \tag{27}$$

whenever $0 < |x - a| < \delta$. Therefore,

$$\left|\frac{1}{g(x)} - \frac{1}{M}\right| = \left|\frac{M - g(x)}{g(x)M}\right| \leqq \frac{\epsilon_2}{(M/2)M} = \frac{2\epsilon_2}{M_2} \tag{28}$$

when $0 < |x - a| < \delta$. If we choose ϵ_2 such that $2\epsilon_2/M^2 < \epsilon$, we will have

$$\left|\frac{1}{g(x)} - \frac{1}{\lim\limits_{x\to a} g(x)}\right| < \epsilon \tag{29}$$

whenever $0 < |x - a| < \delta$. This proves (15) and completes the proof of Theorem E.

Theorem F

If

$$\lim_{x\to a} f(x) = L \tag{30}$$

then

$$\lim_{x\to a} |f(x) - L| = 0 \tag{31}$$

and conversely.

The assertion (30) means that to each positive number ϵ there corresponds a positive number δ such that

$$|f(x) - L| < \epsilon \tag{32}$$

whenever $0 < |x - a| < \delta$. The assertion (31) means that to each positive number ϵ there corresponds a positive number δ such that

$$|\,|f(x) - L| - 0| < \epsilon \tag{33}$$

whenever $0 < |x - a| < \delta$. Since the left members of (32) and (33) are equal, each assertion implies the other.

Theorem G (sandwich theorem, or flyswatter theorem)

If, for some positive number p,

$$g(x) \leqq f(x) \leqq h(x) \tag{34}$$

when $a - p < x < a$ and when $a < x < a + p$, and if

$$\lim_{x\to a} g(x) = L, \qquad \lim_{x\to a} h(x) = L, \tag{35}$$

then

$$\lim_{x\to a} f(x) = L. \tag{36}$$

The primitive idea behind this theorem may be phrased as follows. If two slices of bread (or two books) are near Minneapolis and if a slice of ham (or a fly) is between them, then the thing that is caught in the middle must also be near Minneapolis. To prove this theorem, let $\epsilon > 0$. Choose δ such that $0 < \delta < p$ and the two inequalities

$$(37) \qquad L - \epsilon < g(x) < L + \epsilon, \qquad L - \epsilon < h(x) < L + \epsilon$$

hold when $0 < |x - a| < \delta$. This and (34) give

$$(38) \qquad L - \epsilon < g(x) \leqq f(x) \leqq h(x) < L + \epsilon$$

and hence

$$(39) \qquad |f(x) - L| < \epsilon$$

when $0 < |x - a| < \delta$. This proves Theorem G.

Theorem 3.288, the last one of the basic theorems of Section 3.2, asserts that if p is a constant positive exponent and $a \geqq 0$, then

$$\lim_{x \to a} x^p = a^p.$$

Proof of this theorem is much more difficult and is given in Section 9.2 after the theory of exponentials and logarithms has been developed; see Theorem 9.271.

We conclude this appendix with an indication of the extent to which mathematical fashions have changed. In a calculus textbook published in 1879 and cited in a footnote near the end of Chapter 3, W. E. Byerly says he "embodies the results of my own experience in teaching the calculus at Cornell and Harvard Universities." His preface claims that one of the "peculiarities" of his book is "rigorous use of the Doctrine of Limits as a foundation of the subject." His basic definition of limit appears on page 3. "*If a variable* which changes its value according to some law *can be made to approach some fixed, constant value as nearly as we please, but can never become equal to it*, the constant is called the *limit* of the variable under the circumstances in question." The "fundamental proposition" in the theory of limits is given as a theorem on page 5:

THEOREM. If two variables are so related that as they change they keep always equal to each other, and each approaches a limit, their limits are absolutely equal.

For two variables so related that they are always equal form but a single varying value, as at any instant of their change they are by hypothesis absolutely the same. A single varying value cannot be made to approach at the same time two different constant values as nearly as we please; for, if it could, it could eventually be made to assume a value between the two constants; and, after that, in approaching one it would recede from the other.

This appendix is based upon the premise that such "definitions" and "proofs" outlived their usefulness as their staunch defenders insisted that it is easier to learn them than to learn definitions and proofs involving epsilons and deltas.

Appendix 2

Volumes

This appendix involves volumes of sets in E_3. Its purpose is to show that the theory of volumes is not simple. We shall reveal the fact that there is something inherently contradictory in the combination of the following four assumptions:

(A_1) Each bounded set S in E_3 has a volume.
(A_2) If S_1 and S_2 are congruent bounded sets in E_3, then S_1 and S_2 have equal volumes.
(A_3) If a bounded set S in E_3 is composed of two or three or four or five separate and distinct subsets, then the volume of S is the sum of the volumes of its subsets.
(A_4) If S is a solid spherical shell having inner and outer radii for which $0 < r_1 < r_2$, then the volume of S is positive.

Figure A.1 shows three spherical shells the inner and outer radii of which are 1 and 2. A point P lies in one of these shells if its distance r

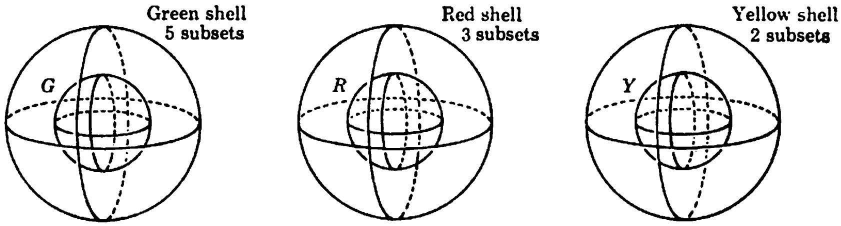


Figure A.1

from the center is such that $1 \leqq r \leqq 2$. The shells are identical (or congruent) except that the first one is green, the second one is red, and the third one is yellow. It has been proved to be possible to separate the green shell G into five separate and distinct parts or subsets G_1, G_2, G_3, G_4, G_5, to separate the red shell R into three separate and distinct parts or subsets R_1, R_2, R_3, and to separate the yellow shell Y into two separate and distinct parts or subsets Y_1 and Y_2 in such a way that

$$R_1 \sim G_1, \qquad R_2 \sim G_2, \qquad R_3 \sim G_3, \qquad Y_1 \sim G_4, \qquad Y_2 \sim G_5$$

where the symbol "$\sim$" means "is congruent to." If we make the first three assumptions listed above and use the symbol $|S|$ to denote the volume of a set S, we obtain

$$\begin{aligned}|G| &= |G_1| + |G_2| + |G_3| + |G_4| + |G_5| \\ &= |R_1| + |R_2| + |R_3| + |Y_1| + |Y_2| \\ &= |R| + |Y| = |G| + |G| = 2|G|.\end{aligned}$$

This implies that $|G| = 0$ and contradicts the fourth assumption (A_4).

Without undertaking to press very far into theories of volumes (these theories being a part of the more comprehensive theory of additive set functions in E_3), we point out that it is possible to assign numbers (called volumes) to some of the sets in E_3 in such a way that the following statements are true.

(B_1) Some sets in E_3, including solid spherical shells having inner and outer radii for which $0 < r_1 < r_2$, have positive volumes.

(B_2) If S is a set in E_3 which has a volume, then each set in E_3 which is congruent to S has a volume which is equal to the volume of S.

(B_3) If a set S in E_3 is the union of a finite collection of separate and distinct subsets each of which has a volume, then S has a volume and the volume of S is the sum of the volumes of the subsets.

The example involving the colored shells proves the following fundamental fact. Whenever volumes are assigned to sets in E_3 in such a way that (B_1), (B_2), and (B_3) are valid, a contradiction arises from the assumption that each bounded set in E_3 has a volume. Thus there exist bounded sets in E_3 that do not have volumes. We are doing "rigorous mathematics" when we give a definition of volume and prove that a given spherical shell has a volume. We are still doing "rigorous mathematics" when we make and use clear statements of provable facts but postpone or omit the proofs. We are deep in the depths of intellectual degradation when, without having a definition of volume, we hold aloft a brick or spherical ball and convey (either explicitly or implicitly) the impression that the thing "obviously" has a volume. We hit the bottom when we say that the thing *is* a volume. Unless we tolerate the idea that bad mathematics can be acceptable elementary calculus, we must avoid these degradations. Perhaps we can attain a reasonable view of this whole matter by recognizing the fact that modern set theory shakes the foundations of nineteenth-century mathematics as vigorously as modern atomic theory shakes the foundations of nineteenth-century physics and chemistry and engineering. Some of us will learn more about these matters than others, but we can all know that there is much to be learned.

The Greek Alphabet

Letters		*Names*	*Letters*		*Names*	*Letters*		*Names*
A	α	alpha	I	ι	iota	P	ρ	rho
B	β	beta	K	κ	kappa	Σ	σ	sigma
Γ	γ	gamma	Λ	λ	lambda	T	τ	tau
Δ	δ	delta	M	μ	mu	Υ	υ	upsilon
E	ϵ	epsilon	N	ν	nu	Φ	ϕ	phi
Z	ζ	zeta	Ξ	ξ	xi	X	χ	chi
H	η	eta	O	o	omicron	Ψ	ψ	psi
Θ	θ	theta	Π	π	pi	Ω	ω	omega

Index