W.M. Priestley

CALCULUS: A LIBERAL ART

Second Edition

Springer

Undergraduate Texts in Mathematics

Springer
New York
Berlin
Heidelberg
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

Undergraduate Texts in Mathematics

Anglin: Mathematics: A Concise History and Philosophy.
Readings in Mathematics.

Anglin/Lambek: The Heritage of Thales.
Readings in Mathematics.

Apostol: Introduction to Analytic Number Theory. Second edition.

Armstrong: Basic Topology.

Armstrong: Groups and Symmetry.

Axler: Linear Algebra Done Right. Second edition.

Beardon: Limits: A New Approach to Real Analysis.

Bak/Newman: Complex Analysis. Second edition.

Banchoff/Wermer: Linear Algebra Through Geometry. Second edition.

Berberian: A First Course in Real Analysis.

Brémaud: An Introduction to Probabilistic Modeling.

Bressoud: Factorization and Primality Testing.

Bressoud: Second Year Calculus.
Readings in Mathematics.

Brickman: Mathematical Introduction to Linear Programming and Game Theory.

Browder: Mathematical Analysis: An Introduction.

Buskes/van Rooij: Topological Spaces: From Distance to Neighborhood.

Cederberg: A Course in Modern Geometries.

Childs: A Concrete Introduction to Higher Algebra. Second edition.

Chung: Elementary Probability Theory with Stochastic Processes. Third edition.

Cox/Little/O'Shea: Ideals, Varieties, and Algorithms. Second edition.

Croom: Basic Concepts of Algebraic Topology.

Curtis: Linear Algebra: An Introductory Approach. Fourth edition.

Devlin: The Joy of Sets: Fundamentals of Contemporary Set Theory. Second edition.

Dixmier: General Topology.

Driver: Why Math?

Ebbinghaus/Flum/Thomas: Mathematical Logic. Second edition.

Edgar: Measure, Topology, and Fractal Geometry.

Elaydi: Introduction to Difference Equations.

Exner: An Accompaniment to Higher Mathematics.

Fine/Rosenberger: The Fundamental Theory of Algebra.

Fischer: Intermediate Real Analysis.

Flanigan/Kazdan: Calculus Two: Linear and Nonlinear Functions. Second edition.

Fleming: Functions of Several Variables. Second edition.

Foulds: Combinatorial Optimization for Undergraduates.

Foulds: Optimization Techniques: An Introduction.

Franklin: Methods of Mathematical Economics.

Gordon: Discrete Probability.

Hairer/Wanner: Analysis by Its History.
Readings in Mathematics.

Halmos: Finite-Dimensional Vector Spaces. Second edition.

Halmos: Naive Set Theory.

Hämmerlin/Hoffmann: Numerical Mathematics.
Readings in Mathematics.

Hijab: Introduction to Calculus and Classical Analysis.

Hilton/Holton/Pedersen: Mathematical Reflections: In a Room with Many Mirrors.

Iooss/Joseph: Elementary Stability and Bifurcation Theory. Second edition.

Isaac: The Pleasures of Probability.
Readings in Mathematics.

(continued after index)

William McGowen Priestley

Calculus: A Liberal Art

Second Edition

With 242 illustrations

Springer

W.M. Priestley
Department of Mathematics and Computer Science
The University of the South
Sewanee, TN 37383
USA

Mathematics Subject Classification (1991): 26-01, 26A06

Library of Congress Cataloging-in-Publication Data
Priestley, William McGowen, 1940–
Calculus : a liberal art / W.M. Priestley. – 2nd ed.
p. cm. – (Undergraduate texts in mathematics)
Includes index.
Contents: v. 1. A liberal art
ISBN 0-387-98379-1 (v. 1 : hardcover : alk. paper)
1. Calculus. I. Title. II. Series.
QA303.P92 1998
515–dc21 97-41813

Production managed by Bill Imbornoni; manufacturing supervised by Joe Quatela.
Typeset by Asco Trade Typesetting Ltd., Hong Kong.

9 8 7 6 5 4 3 2 1

ISBN 0-387-98379-1 Springer-Verlag New York Berlin Heidelberg SPIN 10656730

To Patten

What is the opposite of *two*?
A lonely me. A lonely you.

Even now there is a very wavering grasp of the true position of mathematics as an element in the history of thought. I will not go so far as to say that to construct a history of thought without profound study of the mathematical ideas of successive epochs is like omitting Hamlet from the play which is named after him. That would be claiming too much. But it is certainly analogous to cutting out the part of Ophelia. This simile is singularly exact. For Ophelia is quite essential to the play, she is very charming—and a little mad. Let us grant that the pursuit of mathematics is a divine madness of the human spirit, a refuge from the goading urgency of contingent happenings.

Alfred North Whitehead
from *Mathematics as an Element in the History of Thought*

Preface to the Second Edition

This second edition of *Calculus: An Historical Approach* has been slimmed down for use in a one-semester calculus course intended primarily for liberal arts students seeking to fulfil general education requirements in mathematics. Retitled *Calculus: A Liberal Art* to reflect this more specialized purpose, it includes everything from its predecessor that would normally be met in the first semester. New material has been added to give the instructor more freedom in determining the mathematical level at which to pitch the course and in choosing what emphasis to place upon historical and philosophical issues connected with the development of calculus and the nature of mathematics. Those who wish to place less emphasis upon such issues, for example, or who wish to discuss them only after first jumping into calculus as quickly as possible, can jump from Chapter 1 to Chapter 4, reviewing intermittently, as needed, topics from Chapters 2 and 3.

What emphasis should be placed upon writing in a course like this? The classical liberal arts included grammar, rhetoric, and logic—disciplines that are still related to our modern notion of a liberal arts education. Some of the exercises and problems in Chapter 2 are designed to reinforce this relation and to recall the larger kinship with mathematics that is often overlooked today. Even a failed attempt to teach writing skills in mathematics may have a beneficial result. Students who try to learn how to write mathematics may inadvertently learn how to *read* mathematics. Appendix 4, entitled "Clean Writing in Mathematics", may be useful in this connection.

All the material here could—in theory, at least—be presented early in secondary school following courses in algebra and geometry. The main reason for delaying its study has to do with the question of mathematical maturity.* No use is made here of trigonometric, logarithmic, or exponential functions except in occasional optional material indicating how such functions can be handled.

A perceptive remark made by George Pólya suggests how we can simultaneously learn mathematics and learn "about" mathematics—i.e., about the nature of mathematics and how it is developed:

> *If the learning of mathematics reflects to any degree the invention of mathematics, it must have a place for guessing, for plausible inference.*

The reader will find plenty of opportunity here for guessing. The early chapters go at a gentle pace and invite the reader to enter into the spirit of the investigation. Exercises asking the reader to "make a guess" should be taken in this spirit—as simply an invitation to speculate about what is the likely truth in a given situation without feeling any pressure to guess "correctly". Readers will soon realize that a matter about which they are asked to guess will likely be a topic of serious discussion later on.

The last couple of full sections in each chapter, after the first, often include several exercises designated as optional. Sometimes they offer brief glimpses of deeper ideas of real analysis. Likewise, the latter problems in most problem sets at the ends of chapters are generally more demanding. Readers can omit these if they wish and still find it easy to go on to study the next chapter. This challenging material is included only in the hope that it may encourage some more ambitious students to continue their study of calculus at the next level. The final appendix, "From Freely Falling Bodies to Taylor series", is included solely for this reason.

I wish to thank Hardy Grant for generously offering to read early drafts of much of the new material and for giving me the benefit of his sound judgment. I am grateful also to Bill Imbornoni for smoothly overseeing Springer-Verlag's production of this second edition with the same care that Joyce Schanbacher bestowed upon its predecessor some twenty years ago.

W.M.P.

January, 1998
Sewanee

* See p. xi.

Preface to the First Edition

This book is for students being introduced to calculus, and it covers the usual topics, but its spirit is different from what might be expected. Though the approach is basically historical in nature, emphasis is put upon ideas and their place—not upon events and their dates. Its purpose is to have students to learn calculus first, and to learn incidentally something about the nature of mathematics.

Somewhat to the surprise of its author, the book soon became animated by a spirit of opposition to the darkness that separates the sciences from the humanities. To fight the spell of that darkness anything at hand is used, even a few low tricks or bad jokes that seemed to offer a slight promise of success. To lighten the darkness, to illuminate some of the common ground shared by the two cultures, is a goal that justifies almost any means. It is possible that this approach may make calculus more fun as well.

Whereas the close ties of mathematics to the sciences are well known, the ties binding mathematics to the humanities are rarely noticed. The result is a distorted view of mathematics, placing it outside the mainstream of liberal arts studies. This book tries to suggest gently, from time to time, where a kinship between mathematics and the humanities may be found.

There is a misconception today that mathematics has mainly to do with scientific technology or with computers, and is thereby unrelated to humanistic thought. One sees textbooks with such titles as *Mathematics for Liberal Arts Majors*, a curious phrase that seems to suggest that the liberal arts no longer include mathematics.

No discipline has been a part of liberal arts longer than mathematics.

Three—logic, arithmetic, and geometry—of the original seven liberal arts are branches of mathematics. Plato's friend Archytas, who helped develop the whole idea of liberal education, was a distinguished mathematician. No true student of liberal arts can neglect mathematics.

How did it happen that mathematics, in the public eye, became dissociated from the humanities? In brief, the emergence and growth of scientific knowledge in the seventeenth century led to a polarization in academic circles. Science went one way, the humanities went another. Mathematics, at first in the middle, seems now to be more commonly identified with the sciences and with the technology they engendered.

Today in some academic institutions the state is not healthy. The ground between the sciences and the humanities is so dark that many well-meaning members on each side lack the education to see the most valuable contributions of the other. To the disadvantage of students, this is sometimes the case even among the faculties of so-called "liberal arts" colleges.

In the seventeenth century mathematics was a bridge between the two kinds of knowledge. Thus, for example, Isaac Newton's new physics could be read by Voltaire, who was at home both with Homer and with Archimedes. Voltaire even judged Archimedes to be superior, in imagination, to Homer.

The unity of knowledge which seemed attainable in the seventeenth century, and which has long been an ideal of liberal education, is still worth seeking. Today as in the time of Voltaire, and in the time of Plato, mathematics calls us to eye this goal.

For Anyone Afraid of Mathematics

Maturity, it has been said, involves knowing when and how to delay succumbing to an urge, in order by doing so to attain a deeper satisfaction. To be immature is to demand, like a baby, the immediate gratification of every impulse.

Perhaps happily, none of us is mature in every respect. Mature readers of poetry may be immature readers of mathematics. Statesmen mature in diplomacy may act immaturely in dealing with their own children. And mature mathematicians may on occasion act like babies when asked to listen to serious music, to study serious art, or to read serious poetry.

What is involved in many such cases is how we control our natural urge to get directly to the point. In mathematics, as in serious music or literature, the point sometimes simply cannot be attained immediately, but only by indirection or digression.

The major prerequisite for reading this book is a willingness to cultivate some measure of maturity in mathematics. If you get stuck, be willing to forge ahead, with suspended disbelief, to see where the road is leading. "Go forward, and faith will follow!" was d'Alembert's advice in the eighteenth century to those who would learn the calculus. Your puzzlement may vanish upon turning a page.

All that will be assumed at the outset is a nodding acquaintance with some elementary parts of arithmetic, algebra, and geometry, most of which was developed long before A.D. 1600. There will be some review in the early chapters, offering us as well a chance to outline the early history of mathematics.

I wish to thank Mary Priestley for helping me in this enterprise and for sharing with me its ups and downs. I am grateful also to Paul Halmos for his interest and encouragement.

W.M.P.

May, 1978
Sewanee

Contents

Acknowledgments

Full acknowledgment for a quotation from copyrighted sources is generally given on the same page as the quotation, but usually only the publisher holding U.S. rights is cited. More complete acknowledgments are required in the following cases, where the given page numbers refer to the location of the quotation in the present text.

George Allen & Unwin, Ltd., for permission to quote from Tobias Dantzig, *Number, the Language of Science* (p. 310); from A.N. Whitehead, *The Aims of Education* (pp. 111–12, 353–4); and from Bertrand Russell, *Our Knowledge of the External World* (p. 303).

Cambridge University Press, for permission to quote from A.N. Whitehead, *Science and the Modern World* (p. vi); and from G.H. Hardy, *A Mathematician's Apology* (p. 312).

The Macmillan Publishing Company, for permission to quote from Tobias Dantzig, *Number, the Language of Science*, copyright 1930, 1933, 1939, and 1954 by Macmillan Publishing Co., Inc., and renewed 1958, 1961, and 1967 by Anna G. Dantzig (p. 310); from A.N. Whitehead, *The Aims of Education*, copyright 1929 by Macmillan Publishing, Inc., and renewed 1957 by Evelyn Whitehead (pp. 111–12, 353–4); and from A.N. Whitehead, *Science and the Modern World*, copyright 1925 by Macmillan Publishing Co. Inc. and renewed 1953 by Evelyn Whitehead (p. vi).

Oxford University Press, for permission to quote from Leo Tolstoy, *War and Peace*, translated by Louise and Aylmer Maude, copyright 1970 (p. 306); and from Edmund Wilson, *The Triple Thinkers*, copyright 1984 (p. 74).

Rathbone Books Limited, for permission to quote from Bertrand Russell, *Wisdom of the West*, copyright 1959 (p. 303).

Simon and Schuster, for permission to quote from Bertrand Russell, *The Basic Writings of Bertrand Russell*, edited by Robert E. Egner and Lester E. Denonn, copyright 1961 (p. 190); and from Will Durant, *The Life of Greece*, copyright 1939 (p. 44, p. 77).

The quotation by Richard Courant (p. 287) is reproduced by kind permission of the author's son.

Anecdote of the Jar

Wallace Stevens

I placed a jar in Tennessee
And round it was, upon a hill.
It made the slovenly wilderness
Surround that hill.

The wilderness rose up to it,
And sprawled around, no longer wild.
The jar was round upon the ground
And tall and of a port in air.

It took dominion everywhere.
The jar was gray and bare.
It did not give of bird or bush,
Like nothing else in Tennessee.

1 CHAPTER

Tokens from the Gods

A calculus is a pebble, or small stone.* Playing with pebbles, or "calculating", is a primitive form of arithmetic. *The calculus*, or *calculus*, refers to some mathematics that was developed principally in the seventeenth century.

Today the calculus can be seen as a natural result of a certain point of view. This point of view is reached in three steps. One begins by inventing the notion of a *variable* and trying to see situations, where possible, in terms of variables. The second step is to focus attention upon the relationship between the variables arising in a particular situation. This leads to the idea of a *function.* The third step involves the notion of the *limit* of a function. This simple yet subtle notion, which makes it all work, was recognized in the seventeenth century as being a key idea.

We shall discuss limits a little later. Right now, let us look at a couple of concrete situations where we can get hold of the idea of a variable and of a function relating one variable to another.

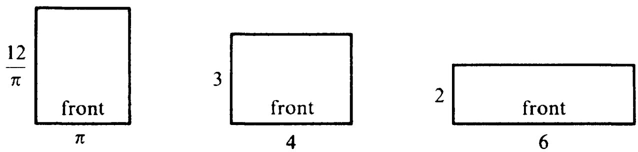
§1. A Calculus Problem

Let us become acquainted with a type of problem that calculus can handle. We shall not be able to solve this problem until certain tools are developed in a later chapter.

* Physicians still use the word *calculus* in this sense, to describe an unwelcome presence in the kidney or bladder. The success of a textbook on calculus is measured by the degree to which its contents are *not* described by the physician's usage of the word.

EXAMPLE 1
A small rectangular pen containing 12 square yards is to be fenced in. The front, to be made of stone, will cost \$5 per yard of fencing, while each of the other three wooden sides will cost only \$2 per yard. What is the least amount of money that will pay for the fencing?

In this example, the total *cost* of the fencing obviously will vary in terms of the design of the rectangle. Our job is to become familiar enough with *how the cost varies* in order to recognize the least possible cost. Toward this end we first pick at random a few possible designs and calculate their corresponding costs. There are lots of ways to enclose 12 square yards:



Exercises

1.1. Suppose the front is 1 yard in length. Find the cost. *Hint.* The cost is the sum of the costs of each of the four sides. First find the lengths of the sides, remembering that the area must be 12 square yards.

1.2. Suppose the front is 2 yards. Find the cost.

1.3. Suppose the front is 3 yards. Find the cost. *Answer*: \$37.00.

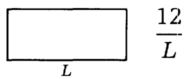
1.4. Suppose the front is π yards. Find the cost. *Answer*: $7\pi + (48/\pi)$ dollars.

§2. Variables and Functions

The information obtained in the exercises above may be conveniently summarized in a table. Here, L is an abbreviation for the length in yards of the front, and C stands for the cost of the fencing in dollars.

L	C
1	55
2	38
3	37
π	$7\pi + (48/\pi)$

We have seen, in the exercises above, that the value of C is entirely determined by the value of L. In other words, there is a *rule* by which one gets from L to C. This rule is simply given by

$C =$ cost in dollars of  (1)

$$
\begin{aligned}
&= \text{cost of front, plus cost of other sides} \\
&= 5L + 2L + 2\left(\frac{12}{L}\right) + 2\left(\frac{12}{L}\right) \\
&= 7L + \frac{48}{L}. \qquad (2)
\end{aligned}
$$

Because the cost C varies in terms of the length L, it is natural to speak of C as a *variable* whose value is determined by the value of the variable L. In other words (and more explanation will be forthcoming *below*), *C is a function of L*, which we express succinctly by writing

$$C = f(L) \qquad (3)$$

(read "C equals f of L"). The symbol f denotes the function, or rule, by which C is given in terms of L. Putting lines (2) and (3) together shows that the rule f can be expressed by the equation

$$f(L) = 7L + \frac{48}{L}.$$

The notation $f(L)$ does not, of course, denote multiplication, but rather denotes the effect of the rule f acting upon the variable L. For example, by this rule,

$$
\begin{aligned}
f(\pi) &= 7\pi + \frac{48}{\pi}, \\
f(3) &= 7 \cdot 3 + \frac{48}{3} = 21 + 16 = 37, \\
f(2) &= 7 \cdot 2 + \frac{48}{2} = 14 + 24 = 38, \\
f(1) &= 7 \cdot 1 + \frac{48}{1} = 7 + 48 = 55.
\end{aligned}
$$

Since the equation $C = 7L + (48/L)$ says virtually the same thing as the equation $f(L) = 7L + (48/L)$, one might ask the reason for introducing this new symbol f. The reason is that we shall need to have a name for the mechanism, or rule, by which one gets from the left column above to the right column. It is, after all, this mechanism f that we want to study in order to recognize the least possible value of the cost C.

Note that f is *not* a variable, but stands for a fixed rule relating the two variables C and L.

Exercises

2.1. Use the rule given by $f(L) = 7L + (48/L)$ to find each of the following.

(a) $f(4)$. (b) $f(5)$.
(c) $f(x)$. (d) $f(\sqrt{2})$.
(e) $f(3+\sqrt{2})$. (f) $f(3+h)$.
(g) $f(x+h)$. (h) $f(\pi^2)$.
(i) $f(x^2)$. (j) $f(4\pi)$.
(k) $f(4t)$. (l) $f(8/7)$.
(m) $f(6)$.

Answers: (c) $7x + (48/x)$. (g) $7(x+h) + (48/(x+h))$. (k) $28t + (12/t)$.

2.2. In the following table, fill in the question marks appropriately. (Your answers to the preceding exercise may be helpful here.)

$L \xrightarrow{f}$	C
4	?
2.5	?
x	?
x^2	?
$x+h$	?
?	50
?	50

§3. Three Ways of Looking at a Function

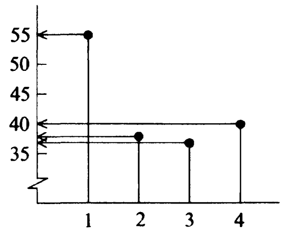
In this book, the word *function* will be used a little loosely and may have either of these three meanings:

(A) A function is a pair of columns of numbers. Not just any pair of columns, but a pair *whose first column has no number repeated.* We speak of the function as being *from* the first column *to* the second.

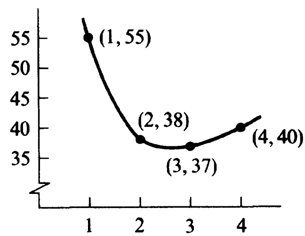
1	55
2	38
3	37
4	40

(etc.)

(B) A function is a rule of correspondence. Not just any rule, but a rule which associates to each number *exactly one* second number. We picture the correspondence as going *from* a horizontal number line *to* a vertical number line.



(C) A function is a curve in the plane. Not just any curve, but a curve that *no vertical line crosses more than once*. (Occasionally, instead of calling the curve a function, we call it the *graph* of a function.)



Do you agree that (A), (B), and (C) are, at heart, expressions of the same idea? Is it not remarkable that the same idea can be thought of—as in (A)—as a *static* notion or—as in (B)—as a *kinematic* notion or—as in (C)—as a *geometric* notion? This remarkable feature is one reason why the idea of a function is an important one. Already the reader may expect that the study of functions will have a bearing on the study of kinematics (that is, motion), and on the study of curves in the plane. If the reader has also the feeling that the idea of a function can change a moving, or fluid, situation into a more easily scrutinized static situation, then much of what the ensuing chapters hold has been foreseen.

A surprising amount of mathematics consists in simply saying the same thing in many different ways, until it is finally said in a way that makes it simple. The problem in Example 1 of finding the least possible cost could be rephrased as either of the following problems:

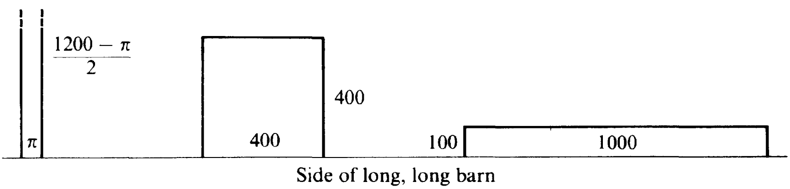
(1) Find the least number that can possibly occur in the second column in (A).
(2) Find the lowest point ever hit on the vertical axis by $f(L)$ in (B).
(3) Find the second coordinate of the lowest point on the curve in (C).

Calculus will teach us how to do the third of these problems. In Chapter 4 we shall begin the study of a technique that often enables one to find with ease the lowest point on a curve.

For obvious reasons, Example 1 is called an *optimization problem*, where the optimum is achieved by *minimizing* a certain variable (the cost C). Let us now look at a second example, where the optimization problem that arises requires that a certain variable be *maximized*.

EXAMPLE 2
A farmer has a cow named Minerva. For her has been purchased 1200 feet of fencing to enclose *three* sides of a rectangular grazing area. The fourth side is bounded by a long barn and requires no fence. Find the largest possible grazing area that Minerva can have.

In this example the area varies with the design of the rectangle. Our task is to become familiar enough with *how* it varies in order to recognize the greatest possible area. We first pick at random a few possible designs and calculate the corresponding areas. There are lots of ways to use that 1200 feet of fencing.



Exercises

3.1. Suppose the side along the barn is 100 feet. Find the area enclosed. *Hint.* First figure out the lengths of the other sides.

3.2. Suppose the side along the barn is 400 feet. Find the area.

3.3. Suppose the side along the barn is 1000 feet. Find the area.

3.4. Suppose the side along the barn is π feet. Find the area. *Answer*: $600\pi - \frac{1}{2}\pi^2$ square feet.

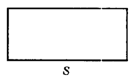
§4. Words versus Algebra

Letting s stand for the length, in feet, of the side along the barn, and letting A stand for the area enclosed, in square feet, we have the following table:

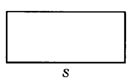
s	$\xrightarrow{g}$ A
100	55,000
400	160,000
1000	100,000
π	$600\pi - \frac{1}{2}\pi^2$

From the exercises above, it is clear that the value of s completely determines the value of A. This means that A is a function of s. We want to become familiar with this function in order to recognize the largest possible area A that it can produce for Minerva. We begin by giving it a name. Let us denote this function by g. (If we have a function pop up, we are free to baptize it with any name we choose. However, it is conventional in most books to reserve the letters f, g, F, and G to designate functions.)

We now have $A = g(s)$. That is, $g(s)$ is the area A corresponding to the rectangle whose length along the side of the barn is s feet. That is,

$g(s) =$ area, in square feet, of  (4)

Equation (4) defines the function g in *words*. It is perfectly proper to define a function by writing out its rule in words. However, if the rule is really an *algebraic* rule in disguise, it behooves us to recognize it. What is the height of the rectangle in (4) whose base is s feet? It is $\frac{1}{2}(1200 - s)$. *Reason*: Having used s feet opposite the barn, we have $1200 - s$ feet left, of which half must go on each of the other sides. Thus, from (4) we can go on:

$g(s) =$ area of  $\dfrac{1200 - s}{2}$

$$= \frac{s(1200 - s)}{2}$$

$$= 600s - \frac{s^2}{2}.$$

This shows that the function g, written out in words in equation (4), can be expressed as an equation in algebra:

$$g(s) = 600s - \frac{s^2}{2}. \tag{5}$$

For obvious reasons, such a function is called an *algebraic* function. Almost all the functions we shall encounter in the first six chapters of this book will be algebraic functions, and it is important to learn to convert an equation in words to an equation in algebra, whenever it is possible to do so. There arise many functions like g, whose rules are expressed in words, but whose rules are really algebraic rules in disguise.

Exercises

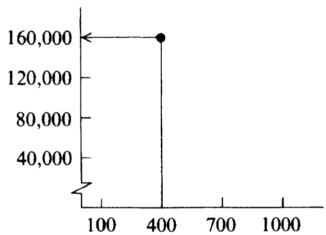
4.1. Use the algebraic rule $g(s) = 600s - \frac{1}{2}s^2$ to calculate

(a) $g(100)$. (b) $g(400)$.
(c) $g(700)$. (d) $g(1000)$.
(e) $g(\pi)$. (f) $g(x)$.
(g) $g(x+\pi)$. (h) $g(x+h)$.
(i) $g(2+3k)$. (j) $g(1/\pi)$.
(k) $g(1/x)$.

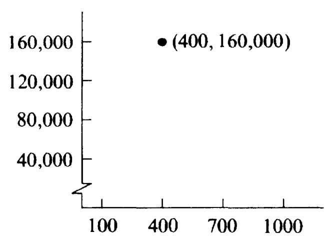
Answers: (c) 175,000. (g) $600(x+\pi) - \frac{1}{2}(x+\pi)^2$.

4.2 Read again the three ways (A), (B), and (C) of looking at a function.

(a) Draw a few arrows, as in (B), picturing the function g as a correspondence going from a horizontal number line to a vertical number line:



(b) Plot a few points, as in (C), lying on the curve g:



(c) In Chapter 4 we shall learn an easy way to find the highest point on the curve g. Can you *guess* what the highest point might be?

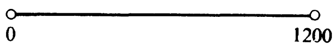
§5. Domain and Range

Look again at equations (4) and (5) above. There is a subtle difference between them, despite the fact that both equations describe exactly the same rule of correspondence. The difference is this: In equation (4) it would make no sense (why?) to let the variable s have a value greater than 1200. Nor would it make any sense in (4) to let s take on a negative value. On the other hand, the algebraic rule $600s - \frac{1}{2}s^2$, given in equation (5), is well defined for *any* value whatever of the variable s. For instance, when s is -2, this algebraic rule gives

$$600(-2) - \frac{(-2)^2}{2} = -1202,$$

even though it is impossible to have a rectangle whose area is negative.

A way to avoid such confusion is to agree to specify, *at the outset*, as soon as a function is introduced, the collection of numbers on which the function is defined. This collection is called the *domain* of the function. The domain of our function g, as specified in words in equation (4), is then the collection of all permissible values of the variable s, which may be pictured like this:



(The *open* circles at the endpoints 0 and 1200 indicate that these values are *excluded* from the domain. We cannot get an honest-to-goodness rectangle if we permit s to equal either 0 or 1200.) Instead of drawing a picture of the domain, one could equally well specify the domain by writing the inequality

$$0 < s < 1200,$$

which says that the values of the variable s are restricted to lie between 0 and 1200.*

Once the domain of a function has been specified, one can then speak of the *range* of values assumed by the function. For the function given by $A = g(s)$, the domain consists of all permissible values of s and the range consists of all corresponding values of A. Since there are three ways of looking at a function, there are three ways of thinking about a function's domain and range:

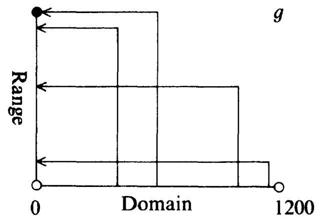
(A) If the function is thought of as a pair of columns, then its domain is the collection of all numbers allowed to go in the first column and its range is the collection of all numbers in the second.

* Had we wished (we did not) to include, say, the point 0 and exclude 1200, we would have written $0 \leq s < 1200$ or drawn the picture with a *closed* circle at 0 and an open circle at 1200.

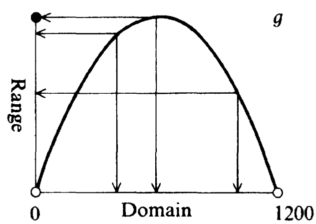
$$s \xrightarrow{g} A$$

Domain s	Range A
400	160,000
700	175,000
1000	100,000
(etc.)	

(B) If the function is thought of as a rule of correspondence, then its domain is the set of all numbers on which the rule acts, and its range is the set of all corresponding numbers.



(C) If the function is thought of as a curve, its domain is the projection of the curve on the horizontal axis and its range is the projection of the curve on the vertical axis.



The domain must be specified before it makes any sense to speak of the range of a function. If the domain is altered, then the range will likely change as well. To find the range of a given function is a problem we shall not discuss until Chapter 4. By (C) above, we see that finding the range involves finding the highest and lowest points on a curve, a topic we shall meet in Chapter 4.

It is usually easy to specify the domain of a function, however. In the function of Example 1, given by

$$f(L) = 7L + \frac{48}{L},$$

it is natural to take the domain to be specified by the inequality

$$0 < L$$

(or $L > 0$, if you prefer), which says that the values of the variable L are restricted to be positive. This restriction is forced by equation (1), where the rule for f is written out in words.

If one does not like to write inequalities, then one should learn to draw pictures. The domain of the function f of Example 1 can be pictured as follows:

(The arrow indicates that the domain is not bounded on the right, but continues to include all positive numbers.)

Suppose a function is specified simply by giving an algebraic rule, such as $\sqrt{x+1}$. (The radical sign $\sqrt{\ }$ denotes the positive square root of what follows.) What shall we understand to be its domain? We shall agree to the following convention.

Convention

Unless otherwise specified, the domain of an algebraic rule shall be understood to be the collection of all numbers for which the rule makes sense.

In applying this convention, one often has to remember two facts which ought to be familiar from arithmetic:

(1) It makes no sense to "divide by zero".
(2) It makes no sense to take the "square root" of a negative number.

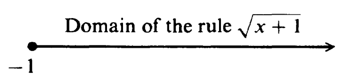
Thus, the domain of the algebraic rule given by $\sqrt{x+1}$, unless otherwise specified, shall be understood to be the collection of all numbers for which $x+1$ is not negative, that is, the collection of all numbers x for which

$$0 \leq x + 1,$$

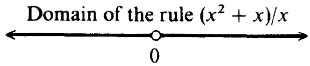
which is the same as saying

$$-1 \leq x,$$

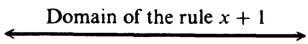
or drawing the picture



Since it makes no sense to divide by zero, the domain of the rule $(x^2 + x)/x$ is pictured as follows:



The rule given by $x + 1$, on the other hand, makes sense for *any number whatsoever*. By our convention, the domain of this rule (unless otherwise specified) shall be understood to be unrestricted:



We now make a point which the reader may think at first to be overly precise. The significance of this point will not be appreciated until later. The point is this: Although it is true that

$$\begin{aligned}\frac{x^2+x}{x} &= \frac{x(x+1)}{x} \\ &= x+1 \quad \text{if } x \neq 0,\end{aligned}$$

the functions given by

$$F(x) = \frac{x^2+x}{x}$$

and

$$G(x) = x+1$$

are *not* the same. *Reason*: The functions F and G do not have the same domain. To say two functions are the same means they have the same graph, and, in particular, they must have the same domain.

Exercises

5.1. In Example 1 we found that the numbers 37 and 55 were in the range of f. Do you believe that every number between 37 and 55 is also in the range? Why might you think so?

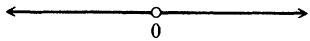
5.2. In Example 2 we found that it was possible to enclose an area of 100,000 square feet and also possible to enclose an area of 160,000 square feet. From these facts, given the nature of the problem raised in Example 2, can you conclude that it is possible to enclose 130,000 square feet?

5.3. Apply the convention above to specify the domain of each of the following algebraic rules. (You may specify the domain either by an inequality or by a picture.)

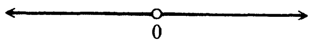
(a) $\sqrt{x}$. (b) $\sqrt{x-1}$.
(c) $1/x$. (d) $1/(x-1)$.
(e) $7x+(48/x)$. (f) $600x-\frac{1}{2}x^2$.
(g) $(h^2+2h)/h$. (h) $h+2$.
(i) $\sqrt{1+h^2}$. (j) $L^2/(L^2-1)$.
(k) $(s-1)(s-2)$. (l) $\sqrt{s-1}$.

Answers:

(e) $x \neq 0$:



(g) $h \neq 0$:



(h) h unrestricted:

(l) $1 \leq s$:

5.4. True or false? The function specified by the rule $(h^2+2h)/h$ is the same as the function specified by the rule $h+2$. *Hint.* Read the last paragraph preceding these exercises.

§6. Optimization

In Example 1, the problem of finding the least cost was seen to be the same as another problem, that of finding the least number in the range of possible costs. To answer the question raised in Example 1, we need to find the least number in the range of f, where f is the function whose rule of correspondence and whose domain are specified succinctly by writing

$$f(L) = 7L + \frac{48}{L}, \quad 0 < L.$$

We shall find this number, once we have developed the appropriate tools of calculus.

In Example 2, the problem of finding the biggest possible area was seen to be the same as another problem, that of finding the largest number in the range of possible areas. To answer the question raised in Example 2, we need to find the largest number in the range of g, where g is the function whose rule of correspondence and whose domain are specified succinctly by writing

$$g(s) = 600s - \frac{1}{2}s^2, \quad 0 < s < 1200.$$

We shall find this number later, using calculus.

In our discussion of Examples 1 and 2, we have seen the first step in how to handle *optimization* problems. An optimization problem can always be spotted by the presence of a superlative. Whenever a problem requires that we find the *least*, or *most*, or *cheapest*, or *best*, or *closest*, etc., we know that we have an optimization problem on our hands. From our discussions in Examples 1 and 2, we may expect that any optimization problem will give rise to a function, and that the solution to the problem will involve finding the highest (or lowest) point on the curve determined by the function. Thus, by seeing the optimization problem in terms of variables, and by getting an algebraic rule relating one variable to another, the optimization problem is transferred to another problem, that of studying the curve determined by the rule, or function, relating the variables. This is the first step in solving optimization problems. This step takes a little while to master. Once it is mastered, however, the second step of finding the highest (or lowest) point on a curve can often be done with the study of only a little calculus.

Must every curve necessarily have a highest point and a lowest point? Certainly not. The curve f of Example 1 has no highest point. *Reason*: The range of costs is not bounded above. There exists no most expensive way to build that fence. The curve g of Example 2 has no lowest point. *Reason*: The grazing area is to be a rectangle and thus cannot have an area of zero, yet the area A ranges arbitrarily close to zero. There is no least possible grazing area for Minerva.

Exercises

6.1. Suppose, in Example 1, the pen was to enclose 30 square yards instead of 12, the costs of stone and wood remaining the same. Find an algebraic rule giving the cost C in terms of the length L of the front, and specify the domain of this rule.

6.2. Suppose, in Example 2, the farmer had 2000 feet of fencing instead of 1200, the other conditions of the problem remaining unchanged. Find an algebraic rule giving the area A in terms of the length s of the side along the barn, and specify the domain of this rule. *Answer*: $A = 1000s - \frac{1}{2}s^2$, $0 < s < 2000$.

§7. Purpose

What follows, gentle reader, is an unorthodox introduction to the notion of a *limit*. (If this is frightening, then be assured that an orthodox discussion is given in Section 9.) Calculus is, in a sense,the study of limits, yet this simple notion is also easily misunderstood, unless the student can make the proper distinction between two things which are easy to confuse. These two things we might call "purpose" and "action". The analogy we shall make, in hopes that it will make the idea of a limit easier to grasp, is this:

The "limit" of a function, at a point in or near its domain, is like the purpose of a human being, at a point in time.

The reader may find that the word *limit* is almost exactly as easy (or as hard) to understand as the word *purpose*.

This analogy will be worth nothing at all unless the ordinary distinction between purpose and action is kept well in mind. These two notions, though often related, are quite different. Most of us can think of instances when our action did not reflect our purpose or of times when we wandered aimlessly to no purpose whatever. Sometimes, even with a purpose, one hesitates to act. Finally, there are the gratifying times when one has a purpose and acts accordingly.

A function, believe it or not, is just like a person in this respect, and one can learn a lot by inquiring into this aspect of the life of a function. At any point in the domain of a function we may compare its action (what it actually does at the point) with its purpose (what it seemed on the threshold of doing at the point). Often, just as in the lives of human beings, the action will agree with the purpose, giving a sense of "continuity". But there are several other possibilities that can occur. The action at some point may disagree with the purpose, or there may be no discernible purpose, or there may be purpose with no action, or there may be neither purpose nor action.

We study functions all the time in calculus, and we gradually learn that each function has a personality all its own. A function is something more than might be imagined from the description "a rule of correspondence", just as a human being is something more than "a featherless plantigrade biped mammal".

Let us try, while studying calculus, to feel ourselves into the world of functions, to see what they really are. Here is a fable. It is offered in fun. Take it seriously, but not too seriously.

Lim: A Fable

The gods did not reveal all things to men at the start; but, as time goes on, by searching, they discover more and more.

Xenophanes

The lord whose is the oracle at Delphi neither reveals nor hides, but gives tokens.

Heraclitus

On the First Day all functions were created, and solemnly told the harsh facts of a functional existence:

> *Each function has been assigned his domain, to which he will be restricted eternally to live in accordance with the rule he has been given. During eternity, he must contemplate the purpose of his being, knowing that on the Last Day the gods may require him to state his purpose at some troublesome point. At that point the function must either state his purpose, or reply that no purpose exists. For at each point some functions have been given a purpose and some have not.* Remember the words of Xenophanes and Heraclitus.

Among the multitude of functions trembling on the First Day was g of *Example* 2. Charged with a Herculean task, g of Example 2 must take each s between 0 and 1200 and throw it to the corresponding A. The throw must be in accordance with the god-given rule

$$A = \text{area of a rectangle of sides } s \text{ and } \frac{1}{2}(1200 - s).$$

Up and down his domain, g carefully moves, throwing his s's until he knows by heart where each little s is supposed to go. He is glad the gods did not ask him to throw -2, or any negative number, or to throw any number exceeding 1200, because he would have no clue where the gods might want these numbers thrown. At last, clever g realizes that he has no purpose at the point -2, or at any negative number, or at any number exceeding 1200. Should the gods ask him, on the fearsome Last Day, of his purpose at the point -2, g would reply in his best courtly fashion:

The purpose of g, *at the point* -2, *does not exist*! (6)

Confidence begins to well up in g.

Yet soon g realizes that the gods have played a trick on him. "Ye gods!" exclaims g, "Why did ye not give me a closed domain?" Poor g is tantalized whenever he moves near the ends of his domain. When he moves to his left, toward 0, he is allowed to throw numbers that lie arbitrarily close to 0; nevetheless, he is not allowed to throw 0, since 0 is not in his domain. A similar frustration is felt when he moves to the right toward 1200.

Night and day, for what seemed like half of eternity, g continuously worried about the points 0 and 1200. Finally the gods had pity upon g and sent down to him a messenger, named Lim.

"Hail, long-suffering g, most favored of Minerva, hail!" shouted Lim.

"Who that?" responded g, so startled that he began dropping his s's.

"I'm the One Who Knows", replied Lim, and smiled smugly. "Remember the words of Xenophanes and Heraclitus."

"Get off my domain!" shouted g, thinking his intruder to be an oracular fanatic.

"Now, now, calm down", said Lim. "I have been sent to help you find your purpose, if you should need help at any point. Do you know your purpose at 1200?"

"Since I am restricted to my domain for all of eternity," said g, "$g(1200)$ does not exist. I am not allowed to act at the point 1200."

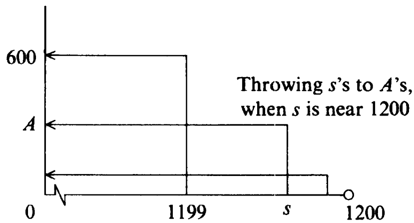
"It is true that you are not allowed action at 1200," responded Lim, "but it is still possible that you may have purpose at that point, not to be fulfilled before the Last Day. Have you no clue what the gods want you to do at 1200 on the Last Day?"

Long-suffering g thought and thought and thought. He thought about his s's near 1200, and about the A's that corresponded to them:

$s \xrightarrow{g} A$

s	A
1190	5950
1199	599.5
1199.9	59.995
1199.99	5.99995
1199.999	0.5999995

(etc.)



"As s gets closer to 1200, A gets closer to 0", exclaimed both Lim and g simultaneously. Then g, in deep tones, declaimed,

The purpose of g at 1200 *is to throw it to* 0. (7)

"Exactly," said Lim, "but why do you speak in such an old-fashioned way? The gods haven't talked like that for ages. Just use my name. Instead of your statement (7), just say,

Lim g *at* 1200 *is* 0,

and instead of (6), say

Lim g *at* -2 *does not exist.*

The gods will understand what you mean. They all know my name. I deliver their ambrosia on Thursdays."

EXERCISES

7.1. Is 1200 in the domain of g? *Answer*: No.

7.2. Does $g(1200)$ exist? *Answer*: No, $g(1200)$ is undefined, because 1200 is not in the domain of g. There is no action of the function g at the point 1200.

7.3. Does Lim g at 1200 exist? *Answer*: Yes. Lim g at 1200 is 0, because as $s \to 1200$, $g(s) \to 0$. (The arrow is an abbreviation for *approaches*, or *gets closer and closer to* or *tends to*.)

7.4. What is Lim g at -2? *Answer*: Lim g does not exist at -2, because $g(s)$ does not exist when s is close to -2, giving g no clue as to a purpose at -2.

7.5. What is Lim g at 1202?

7.6. What is $g(1202)$? *Answer*: Since the function g is not allowed to act at the point 1202, $g(1202)$ does not exist.

7.7. What is $g(0)$?

7.8. What is Lim g at 0?

7.9. What is Lim g at 1200.1? *Answer*: It does not exist.

7.10. What is Lim g at -0.1?

§8. Continuity: Purpose versus Action

"Aha!" said g, "I understand now everything about a function's purpose."

"That is doubtful," replied Lim, "for you are still likely to confuse *purpose* with *action*. What, for example, is your purpose at the point 600?"

"Lim g at 600 is 180,000," responded g without hesitation, "because $g(600)$ is 180,000."

"Aha!" said Lim, "A right answer, *but for a wrong reason*. Just as I expected. When the gods inquire about your purpose at 600, they have in mind something more subtle than you imagine. To reply that $g(600)$ is 180,000 is to state your *action* at the point 600. But action need not necessarily agree with purpose. (At the point 1200 you have no action, yet you do have purpose.)"

"To find your purpose at 600, the first thing you must do is to *forget entirely about your action* at 600. You may as well pretend that 600 has been removed from your domain. Then you proceed just as before. What does A approach as s approaches 600?"

Long-suffering g thought and thought and thought. What, indeed, would be his purpose at 600 if 600 were removed from his domain?

The point 600, being in the interior of the domain, can be approached by values of s either slightly smaller or slightly larger than 600:

$s \xrightarrow{g} A$

s	A
500	175,000
550	178,750
590	179,955
598	179,998
599	179,999.5
599.9	179,999.995

(etc.)

Letting s approach 600 from the left ($s \to 600^-$)

$s \xrightarrow{g} A$

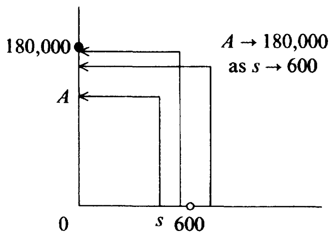
s	A
700	175,000
650	178,750
610	179,955
602	179,998
601	179,999.5
600.1	179,999.995

(etc.)

Letting s approach 600 from the right ($s \to 600^+$)

Whether s tends to 600 from the left, or the "minus side" ($s \to 600^-$), or whether s tends to 600 from the "plus side" ($s \to 600^+$), the corresponding values of A tend to 180,000. Since $A \to 180{,}000$ as $s \to 600$ (from either side),

$$\text{Lim}\, g \text{ at } 600 \text{ is } 180{,}000. \tag{8}$$



"Now let me get this straight", said g. "To find my purpose at 600, I first pretend that 600 has been removed from my domain, and then see what happens to A as s tends to 600. This is the way I figure out that statement (8) is true. Isn't there an easier way to do it?"

"Yes," said Lim, "if you are not afraid to use your common sense. Just look at the rule you were given, $g(s) = 600s - \frac{1}{2}s^2$, and note that it is described in terms of some simple algebraic operations. Look at what happens to each of them in turn, as $s \to 600$. Common sense should tell you that, as $s \to 600$, it must follow that $s^2 \to (600)^2$ and $600s \to 600(600)$. Therefore, as $s \to 600$,

$$A = g(s) = 600s - \frac{1}{2}s^2 \to 600(600) - \frac{1}{2}(600)^2.$$

Thus, $A \to 180{,}000$."

"I really feel great at 600," said g, "whereas at 1200 I become so frustrated."

"That is because, at 600, your action agrees with your purpose:

$$\begin{aligned} g(600) &= 180{,}000 \quad \text{(by applying the rule } g \text{ to } 600) \\ &= \text{Lim}\, g \text{ at } 600 \quad [\text{by } (8)]. \end{aligned}$$

Like any creature, you experience the wholesome feeling of *continuity* at any point where action and purpose exist and agree. Whenever there is not agreement between action and purpose, or whenever one or both are missing, the anxieties of discontinuity emerge. At 1200, friend g, you behave discontinuously. You have a purpose:

$$\text{Lim}\, g \text{ at } 1200 \text{ is } 0,$$

but you do not act accordingly:

$$g(1200) \text{ does not exist.}$$

Everyone is frustrated by discontinuity."

"Let me leave you with this idea, to ponder as you will. To say that a function is continuous at a certain point means that, at the point, the function has both purpose and action, and they agree."

Definition. A function G is said to be **continuous** at a point x provided that the following three conditions are satisfied:

(1) $G(x)$ *exists.*
(2) Lim G *at* x *exists.*
(3) $G(x) =$ Lim G *at* x.

"That is all the help I can give you," said Lim, "for I must now depart. I have to collect 37 chariotloads of ambrosia before Thursday."

"Wait!" shouted g. "Can't you help my friend f of Example 1? She lives on the domain $0 < L$. What a curve!"

"Remember the words of Xenophanes and Heraclitus!" said Lim, and departed without another word.

EXERCISES

8.1. What is $g(400)$?

8.2. What is Lim g at 400?

8.3. Is g continuous at the point 400?

8.4. Is g continuous at the point 0?

8.5. Is g continuous at the point -2?

8.6. Is g continuous at every point in its domain? *Answer*: Yes.

8.7. Consider the function f of Example 1. Is f continuous
(a) at 0?
(b) at -2?
(c) at 2?
Give reasons justifying your answers.

§9. Limits

In everyday language the word *limit* has virtually the same meaning as *bound*. In calculus, howevr, it has a rather different meaning. The *limit* of a function, at a certain point, is (roughly speaking) what the function, at that point, is on the threshold of doing.* If c is the point in question, then the limit of f at c is symbolized by

$$\operatorname*{Limit}_{x \to c} f(x), \tag{9}$$

or by

$$\text{Lim } f \text{ at } c,$$

and is found by investigating the action of f at points near c, *while completely ignoring the value of f at c.* Before any further explanation is given,

* The word *limit* is kin to the Latin word *limen*, which means "threshold".

it should be emphasized that $f(c)$, the value of f at c, may well be entirely unrelated to the limit of f at c. [If it happens that they are the same, that is, that $f(c) = \text{Limit}_{x\to c} f(x)$, then the function f is said to be **continuous** at the point c.]

How does one find the limit of a function at a point? The symbolism (9) is designed to suggest the method of doing this. We simply ask for the limiting value of $f(x)$, as we imagine the values of the variable x taken closer and closer to (but not equal to) the number c. The arrow in "$x \to c$" is supposed to suggest x approaching c ever more closely.

Some examples should serve to clarify things. The reader is asked simply to use common sense in thinking about what happens as the value of a variable gets closer and closer to a fixed number c.

EXAMPLE 3
Let $F(x) = (x^2 + x)/x$, with domain specified by the inequality $x \neq 0$. Find Lim F at 4.

Here we are asked to find

$$\underset{x\to 4}{\text{Limit}} \frac{x^2 + x}{x}, \tag{10}$$

and it is obvious how this is to be done, simply by reading the formula (10) in words: We are asked to find the limiting value of the expression $(x^2 + x)/x$ as x tends to 4. What happens to this expression as $x \to 4$? Common sense tells us that $x^2 \to 16$, so that the expression $(x^2 + x)/x$ approaches $(16 + 4)/4$, which is equal to 5. Therefore,

$$\underset{x\to 4}{\text{Limit}} \frac{x^2 + x}{x} = 5,$$

which answers the question raised in Example 3, and also shows, incidentally, that F is continuous at 4. (Why?) □

EXAMPLE 4
Let $G(x) = x + 1$, with unrestricted domain. Find Lim G at 0.

This is even easier than the preceding example. As $x \to 0$, common sense says that $(x + 1) \to 1$. Therefore,

$$\underset{x\to 0}{\text{Limit}} (x + 1) = 1. \qquad \square$$

EXAMPLE 5
Let $F(x) = (x^2 + x)/x$, with domain specified by $x \neq 0$. Find Lim F at 0.

Here we are asked to find

$$\underset{x\to 0}{\text{Limit}} \frac{x^2 + x}{x},$$

and it is *not* obvious, at first, how this is to be done. As $x \to 0$, both the numerator $x^2 + x$ and the denominator x also approach 0. What is to be done?

What is to be done is to realize that the question raised in Example 5 is exactly the same as the question raised in Example 4, whose answer, we have seen, is 1. Why are these two question exactly the same? Because Lim F at 0, remember, is to be found by letting x tend to 0, but never allowing x to equal 0. If $x \neq 0$, though, then

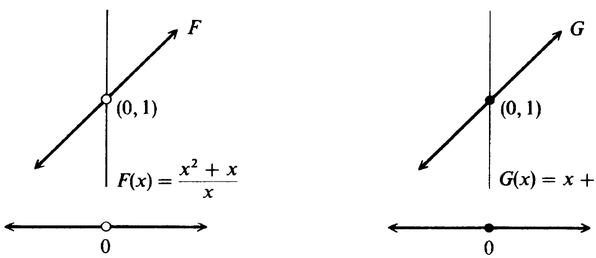
$$F(x) = \frac{x(x+1)}{x} = x + 1,$$

so that the limit of F at 0 is the same as the limit of $x + 1$:

$$\operatorname*{Limit}_{x\to 0} \frac{x^2 + x}{x} = \operatorname*{Limit}_{x\to 0} (x+1) = 1,$$

by Example 4. (What has just been illustrated in this example is not hard, but it is subtle. Reread this example, and also the last remark in Section 5, to make sure you understand it.) □

A picture is the best way to illustrate why Examples 4 and 5 must have the same answer:



The curves F and G are identical, except when x is 0. Since the limit of a function at the point 0 is independent of the action at 0, F and G have the same limit at 0.

Here is another example, using h instead of x as the variable.

EXAMPLE 6

Find $\text{Limit}_{h\to 0}((h^2 + 2h)/h)$.

This is like Example 5, where it is not immediately obvious whether the limit exists at 0. Both the numerator $h^2 + 2h$ and the denominator h tend to 0 as $h \to 0$. However, *if h is not equal to* 0, then we may divide by h, so

$$\frac{h^2+2h}{h} = \frac{h(h+2)}{h} = h+2.$$

This shows that the algebraic rule given by $(h^2 + 2h)/h$ is exactly the same as the algebraic rule $h + 2$, *provided h is not* 0. Since the limit at 0 is independent of the action at 0, these two rules have the same limit at 0:

$$\underset{h\to 0}{\text{Limit}} \frac{h^2+2h}{h} = \underset{h\to 0}{\text{Limit}}\,(h+2) = 2. \qquad \square$$

When investigating the limit of a function at a point, one may encounter any of the following situations:

(I) The limit exists and agrees with the action of the function at the point.
(II) The limit exists, but the function does not act accordingly.
(III) The limit does not exist.

If case (I) occurs, the function is said to be continuous at the point. This is illustrated in Examples 3 and 4. Case (II) is illustrated in Examples 5 and 6. Case (III) will be illustrated in Examples 7 and 8.

EXAMPLE 7
Find $\text{Limit}_{x\to 0}(7x + (48/x))$.

This limit does not exist. As $x \to 0$, the first term, $7x$, is "well-behaved", tending to 0, but the second term, $48/x$, does not tend to a limit, since it becomes large-positive as x tends to 0 from the right, and it becomes large-negative as x approaches 0 from the left:

x	$7x + 48/x$
1	55
0.1	480.7
0.01	4800.07

(etc.)

Letting x approach 0 from the right ($x \to 0^+$)

x	$7x + 48/x$
-1	-55
-0.1	-480.7
-0.01	-4800.07

(etc.)

Letting x approach 0 from the left ($x \to 0^-$)

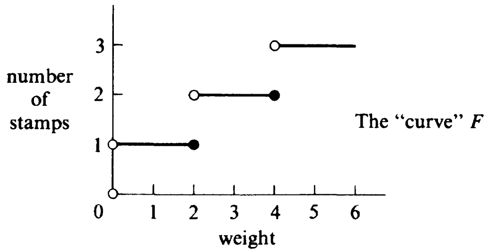
$\square$

EXAMPLE 8
The Post Office has discovered that the cost of sending a letter by mail varies in terms of the weight w of the letter. Accordingly, the number of stamps to be affixed to a letter is a function of w. One stamp is required if the weight w is 2 ounces or less; two stamps if $2 < w \leq 4$; three stamps if

$4 < w \leq 6$; etc. Let us call this function F, so that

$$F(w) = \text{the number of stamps on a letter of weight } w.$$

Find Lim F at 4.



Lim F at 4 does not exist. As w tends to 4 from the left, the number of stamps $F(w)$ tends to 2; whereas, when w tends to 4 from the right, $F(w)$ tends to 3:

w	$F(w)$
3.9	2
3.99	2
3.999	2

(etc.)

$[F(w) \to 2, \text{ as } w \to 4^-]$

w	$F(w)$
4.1	3
4.01	3
4.001	3

(etc.)

$[F(w) \to 3, \text{ as } w \to 4^+]$

The limit does not exist at 4, because we get different "answers" when we approach 4 from different sides. □

Exercises

9.1. Limits are sometimes described as "simple, yet subtle". The purpose of this exercise is to make sure that both these features are noticed. Here is an evaluation of a limit, correctly written as a chain of three equalities:

$$\operatorname*{Limit}_{x \to 0} \frac{x^2 - 7x}{x} = \operatorname*{Limit}_{x \to 0} \frac{x(x-7)}{x} = \operatorname*{Limit}_{x \to 0} (x - 7) = -7.$$

To justify the answer of -7, we must justify each of the three equalities.

(a) Which of the three equalities are "simple"? *Answer*: The first equality and the last require almost no thought.

(b) Explain the subtlety in the second equality. *Answer*: Remember that to get the limit at 0 of a function, we are allowed to consider the action of the function only at *nonzero* values. But when x is nonzero then x/x is always 1, so the function sending x to $x(x-7)/x$ produces the same action as the function sending x to $x-7$. Because both functions have the same action for nonzero values—and only the action at nonzero values determines the limit at zero—they have the same limit at zero, that is,

$$\operatorname*{Limit}_{x\to 0} \frac{x(x-7)}{x} = \operatorname*{Limit}_{x\to 0}(x-7).$$

This justifies the second equality. (It may be helpful to note further that the function sending x to $x-7$ is continuous at 0, whereas the function sending x to $x(x-7)/x$ is not continuous at 0 because it has no action at 0.)

9.2. Here is an evaluation of a limit, correctly written as a chain of equalities:

$$\operatorname*{Limit}_{x\to 0} \frac{x}{x^2-7x} = \operatorname*{Limit}_{x\to 0} \frac{x}{x(x-7)} = \operatorname*{Limit}_{x\to 0} \frac{1}{x-7} = -\frac{1}{7}.$$

To justify the answer of $-1/7$, we must justify each of the three equalities above.

(a) Which of the three equalities are "simple"?
(b) Explain the subtlety in the second equality.

9.3. Evaluate each of the indicated limits (or state that the limit does not exist). Write a chain of equalities to arrive at your answer, as illustrated in exercise 9.2. *Take care to include the "subtle" equality—if one is needed—in your chain.*

(a) $\operatorname*{Limit}_{x\to 0} \dfrac{x^2-4x}{x}$.

(b) $\operatorname*{Limit}_{x\to 0} (x-4)$.

(c) $\operatorname*{Limit}_{x\to 1} \dfrac{x^2-1}{x-1}$.

(d) $\operatorname*{Limit}_{h\to 0} \dfrac{5h^2}{h}$.

(e) $\operatorname*{Limit}_{h\to 0} \dfrac{5h}{h}$.

(f) $\operatorname*{Limit}_{h\to 0} \dfrac{5}{h}$.

(g) $\operatorname*{Limit}_{t\to 3} \dfrac{t-3}{t^2-9}$.

(h) $\operatorname*{Limit}_{t\to 3} \dfrac{5}{6+t}$.

9.4. Consider each of the following algebraic rules, and tell whether it is continuous at the indicated point c:

(a) $(x^2-4x)/x$; $c=0$. *Answer*: Not continuous.
(b) $x-4$; $c=0$.
(c) $(x^2-1)/(x-1)$; $c=1$.
(d) $5/(6+t)$; $c=3$. *Answer*: Continuous.

9.5. Consider the "Post Office function" defined in Example 8.

(a) The function F is defined by a rule stated in words. Do you think it is likely that this rule is an algebraic rule in disguise? *Hint*. Do you think an algebraic rule could have a graph like the "curve" F?
(b) Does Limit F at 2 exist?
(c) Does $F(2)$ exist?
(d) Is F continuous at 2?
(e) Is F continuous at 3? *Answer*: Yes.

(f) A politician asserts that "the scale of charges imposed by the Post Office upon its customers exhibits unnatural and unjustifiable discontinuities at 2-ounce intervals." Explain, in more detail, what the politician means.

9.6. (*A philosophical question to be pondered for a while before being answered*) Is discontinuity unnatural? That is, must the rules that come from laws of nature necessarily be continuous? Man-made rules like the Post Office function, are often discontinuous, at least at some points. (One philosopher's answer to this question is discussed in Section 3 of Chapter 6.)

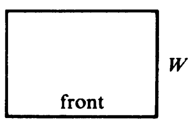
§10. Summary

Variables, *functions*, and *limits* were ideas that came of age in the seventeenth century. Fermat (pronounced fer-MAH) was probably the first to see the real importance of limits. *Continuity* is an old philosophical term that drew new interest from Leibniz (pronounced LĪP-nits), who was the first to use the word *function*.

These notions were not particularly well defined by their inventors, who were content to describe things in intuitive terms. The word *function* at first referred only to an *algebraic* rule, which is automatically continuous at each point in its domain.

Problem Set for Chapter 1

1. Consider Example 1 once more. We chose to look at it in terms of the variables C and L. The cost variable C cannot be avoided, since the problem involves finding the minimum of this variable. However, instead of choosing L, the length of the front, as our second variable, we might just as well have chosen W, the depth of the pen.

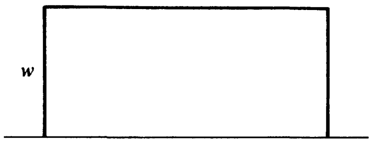


(a) Write an algebraic rule expressing C in terms of W.
(b) What is the domain of the rule in (a)?
(c) Plot a few points on the graph of the equation in (a) that expresses C in terms of W.

(d) Write an equation that relates W and L. *Hint.* What is their product?
(e) Go back to equation (2) and, in it, replace L by $12/W$ and simplify. Do you get the same equation as you got in part (a) above? Why did it work out that way?

2. In Example 1, change the word *least* to *greatest*. With this modification, respond to the question raised.

3. Consider Example 2 once more. We chose to look at it in terms of the variables A and s. There is no getting around the variable A, since it must be maximized in order to answer the question raised. However, instead of s, we might just as well have chosen the other dimension w of the rectangle to be our other variable.



(a) Write an algebraic rule expressing A in terms of w.
(b) What is the domain of the function whose rule is given in part (a)? (Be careful.)
(c) Plot a few points on the graph of the equation in (a)
(d) Write an equation relating w and s. *Hint.* What is the sum of $2w$ and s?
(e) In the equation $A = 600s - \frac{1}{2}s^2$, replace s by $1200 - 2w$, and simplify. You should get the same answer as in part (a). Why?

4. In Example 2, change the word *largest* to *smallest*. With this modification, respond to the question raised, bearing in mind that no honest-to-goodness rectangle has an area of 0.

5. Some curves determine functions and some do not. Does a circle ever determine a function?

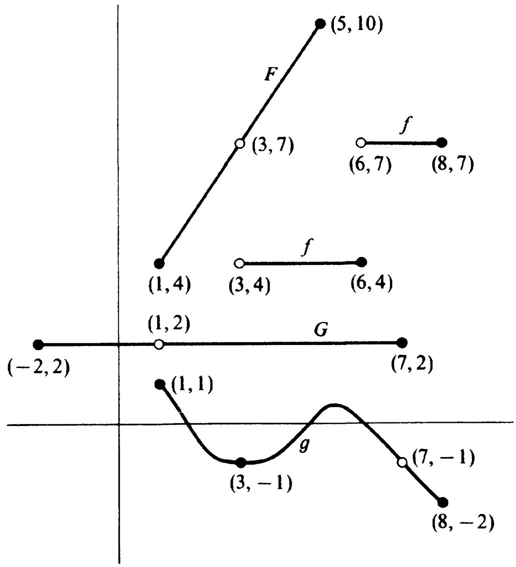
6. Do all straight lines determine functions? If not, give an example of one that doesn't.

7. Some algebraic equations determine functions and some do not. Consider the algebraic equation $x^2 + y^2 = 1$.
(a) Is $(0, 1)$ on the graph of this equation?
(b) Is $(0, -1)$ on the graph of this equation?
(c) Does the algebraic equation $x^2 + y^2 = 1$ determine a function?

8. Does the algebraic equation $y = \sqrt{1 - x^2}$ determine a function? If so, what is its domain?

9. One way to specify a function is to draw the curve it determines. For each of the curves below, specify the domain and the range. Specify either by draw-

ing pictures or by writing an inequality, whichever is easier. (Note that the curve f consists of two pieces.)

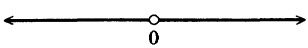


10. Referring to the functions F, f, G, and g pictured in the preceding problem, find
    (a) Limit F at 3.
    (b) Limit f at 6.
    (c) Limit G at 1.
    (d) Limit g at 3.
    (e) Limit g at 7.
    (f) $F(3)$.
    (g) $f(6)$.
    (h) $G(1)$.
    (i) $g(3)$.
    (j) $g(7)$.

11. Still referring to the functions F, f, G, and g of problem 9, answer the following questions.
    (a) Is F continuous at 3?
    (b) Is f continuous at 6?
    (c) Is G continuous at 1?
    (d) Is g continuous at 3?
    (e) Is g continuous at 7?

12. The domain of the "Post Office function" of Example 8 is specified by the inequality $0 < w$. What is its range?

13. The functions of Examples 5 and 6 (in Section 9) have the same domain. It has a hole in it, at the point 0:

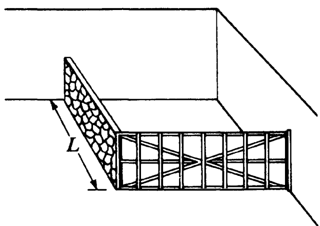


(a) What is the range of the function of Example 5?
(b) What is the range of the function of Example 6?

14. Suppose the numbers 1 and 3 are known to be in the range of a certain function. Must the range then necessarily contain all numbers between 1 and 3? *Hint.* Look at your answer to problem 12.

15. Suppose the numbers 1 and 3 are known to be in the range of a certain function, and suppose the function is continuous at every point in its domain. Must the range necessarily contain all numbers between 1 and 3? *Hint.* Look at your answer to problem 13(b). What if, in addition, the domain has no "holes" in it?

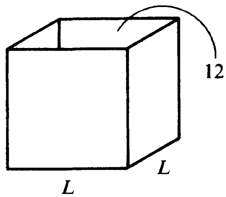
16. In the corner of a large courtyard a rectangular enclosure is to be built. To pay for the material, \$240 has been allocated. This is to be used to pay for both the stone fence, which costs \$6 per meter, and the wood fence, which costs \$2 per meter. The area A of the enclosure will vary with the way the enclosure is built.



(a) Let L be the length of the stone fence. How much money will be left to spend for wood?
(b) Let L be the length of the stone fence. How long will the wood fence be? *Hint.* The answer to part (a) tells you how much money is left for wood.
(c) Let L be the length of the stone fence. Find an algebraic rule giving the area A in terms of L, and specify the domain of this rule.

17. There is often more than one way to choose your variables. In problem 16,
(a) Let x be the amount of money spent on stone. How much is left to spend on wood, and how long, therefore, is the wooden fence?
(b) Let x be the amount of money spent on stone. Find an algebraic rule giving A in terms of x, and specify the domain of this rule.

18. A metal container, in the form of a rectangular solid, is to be constructed. The base is to be square, there is to be no top, and the volume of the container (the product of its three dimensions) is to be 12 cubic meters. Suppose the material for the sides costs \$2 per square meter, and the material for the base costs \$3 per square meter.



(a) Let C be the cost of the material for the container, and let L be the length of a side of the square base (in meters). Find the cost C if L is 2. *Hint.* First note that the height of the container must be 3 (why?) if L is 2.
(b) Find the cost C if L is π. *Hint.* First note that the height of the container must be $12/\pi^2$ if L is π.
(c) Find an algebraic rule giving C in terms of L, and specify its domain.

19. *(This problem is like the preceding one, except that we have a specified amount of material, instead of a specified volume.)* Suppose that we have 120 square feet of material, out of which is to be constructed a square base and four sides of a rectangular container. (The container is to have no top.) Let L be the length of a side of the base.
(a) If H is the height of the container, then it is true that $120 = L^2 + 4LH$. (Why?) Solve this equation for H, to get H in terms of L.
(b) The volume V of the container is the product of its three dimensions, so $V = L \cdot L \cdot H$. Use part (a) to get V in terms of L alone.
(c) In part (b) we have V as a function of L. What is the domain of this function?

20. *(Magic)* We shall define two functions, f and g, specifying the first as an algebraic rule and specifying the second as a rule of correspondence. We define f by $f(x) = (x^2 + x)/2$, so that, for example, $f(5) = (25 + 5)/2 = 15$. We define g by specifying that only positive integers go in the left column of g, while in the right column goes the corresponding sum of all the positive integers up to and including the integer upon which g acts. Thus, for example, $g(5)$ is defined to be sum of all the positive integers up to and including 5, i.e., $g(5) = 1 + 2 + 3 + 4 + 5 = 15$.
(a) Make a table of two columns representing f and make another table of two columns representing g. Choose a few small positive integers, say 5, 6, 7, and 10, to put in the left-hand column of each function and calculate the corresponding numbers in the right-hand column for f and for g.
(b) You should notice something striking about your two tables and (if you are willing to make the "philosopher's leap") you will probably be able to guess correctly what is the sum of the first thousand positive integers. *What is it?* $1 + 2 + 3 + \cdots + 1000 = ??!$

21. *(More magic)* Virtually anything you can do to a number to make a new number (square it, double it, halve it, etc.) you can do to a function to make a new function.
(a) Make a table of two columns representing the "square" of the function f of the preceding problem. (Since $f(5) = 15$, it follows that $f^2(5) =$

$(15)^2 = 225$; more generally, since $f(x) = (x^2 + x)/2$, it follows that $f^2(x) = (x^2 + x)^2/4$.) Choose a few small positive integers, say 5, 6, 7, and 10 to put into the left-hand column of f^2 and calculate the corresponding numbers in the right-hand column.

(b) Let us define a function G by specifying that only positive integers go in the left column of G, while in the right column goes the corresponding sum of the *cubes* of the positive integers up to and including the integer upon which G acts. Thus, for example, $G(5)$ is defined to be the sum of the cubes of all the positive integers up to and including 5, i.e., $G(5) = 1^3 + 2^3 + 3^3 + 4^3 + 5^3 = 1 + 8 + 27 + 64 + 125 = 225$. Choose a few small positive integers, say 5, 6, 7, and 10 to put into the left-hand column of G and calculate the corresponding numbers in the right-hand column.

(c) You should notice something striking about your two tables and (if you are willing to make the "philosopher's leap") you will probably be able to guess correctly what is the sum of the cubes of the first thousand positive integers. *What is it?* $1^3 + 2^3 + 3^3 + \cdots 1000^3 = ??!$

(*This problem and the preceding problem are previews of some ideas to be discussed more fully in Chapter 2.*)

22. Find each of the following limits (or state that the limit does not exist).

(a) $\underset{x\to 5}{\text{Limit}}\, \dfrac{x^2-25}{x-5}$. (b) $\underset{x\to 1}{\text{Limit}}\, \dfrac{x^2-2x+1}{x-1}$. (c) $\underset{x\to \pi}{\text{Limit}}\, \dfrac{x-\pi}{x^2-\pi^2}$.

(d) $\underset{x\to 5}{\text{Limit}}\, \dfrac{x^2-25}{x^2-5x}$. (e) $\underset{x\to 7}{\text{Limit}}\, \dfrac{3x-21}{x^2-7x}$. (f) $\underset{x\to 0}{\text{Limit}}\, \dfrac{3x}{x^2-7x}$.

23. (a) Suppose it is known that $f(x) = (x^2 - 7x)/3x$ for all *nonzero* values of x. Suppose it is further known that f is continuous at 0 (so that, in particular, $f(0)$ exists). What is $f(0)$?

(b) Suppose a function has both action and limit at a certain point. Must the function necessarily be continuous at that point? If not, give an example of a function with both action and limit at a point, where the action is not the same as the limit.

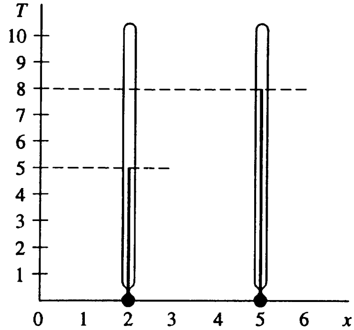
(c) (*Your answer to this vague question may be connected with your answers to questions 24 and 25 that follow.*) Can you give an example of a discontinuous function defined on an unrestricted domain that would not appear "unnatural", "contrived", or "strange" to an impartial observer?

24. ("*Thought-experiment*") A very thin metal rod extends from $x = 0$ to $x = 6$ on the horizontal axis. Suppose at a particular instant its temperature T has been measured at every point except its midpoint $x = 3$ and is given in degrees Celsius by the formula

$$T = \frac{x^2 - 9}{x - 3}. \quad \textit{[See figure below.]}$$

This formula gives no answer when x is 3, but the rod must, of course, have *some* temperature at its midpoint. What must its temperature be when $x = 3$? *Hint.* If you use this formula to calculate its temperature when $x = 2.9$ or when $x = 3.01$, you'll have a couple of clues as to its *approximate* temperature when x is 3. What do you think its exact temperature must be? More importantly, *why must its temperature be what you think?*

x	T
2	5.00
2.9	5.90
3	?
3.01	6.01
5	8.00



25. (*A more challenging thought-experiment*) Suppose, at a particular instant, we somehow know the temperature in degrees Celsius at each point of the space occupied by some physical entity such as a very thin metal rod, the surface of a calm lake, or a house on fire. We then know the rule of correspondence that, at a fixed instant, assigns to each point in this space its temperature. Must this "temperature function" necessarily be continuous? Respond to this question by either
   (a) describing a situation *that could actually happen* where the temperature at some point is not the limit of the temperature function at that point; or
   (b) writing a paragraph explaining why it is *physically impossible* for the temperature at any point P of a physical entity to be anything other than the limit of the temperatures attained at points arbitrarily close to P.

Is your answer to part (c) of problem 23 related to your answer to this problem?

2 CHAPTER Rational Thoughts

The purpose of this chapter and the next is to remind ourselves of the elements of precalculus mathematics by studying a little history. Some readers may have the urge instead to jump flat-footed into an attack upon the problem that arose in Chapter 1, the problem of how the highest and lowest points of a curve may easily be found. Those readers may scan these short chapters quickly and jump into Chapter 4 if they wish, but they are warned that flat-footed jumps are awkward without a firm foundation from which to leap. Studying history builds foundations.

As we shall see, the Greeks' interest in mathematics for its own sake, beginning as early as the sixth century B.C., played a crucial role in the development of science, philosophy, and liberal education. In earlier periods the importance of mathematics seems to have been attached almost wholly to its practical value in measurement, as in astronomy or commerce. The geometry of magnitudes was valuable to the first stargazers, as was the arithmetic of pebbles to the first tradesmen. Why should practical tools of measurement such as magnitude and number be studied in abstraction for their own sake? The birth of pure mathematics some 2500 years ago still raises intriguing questions about the human spirit that do not lend themselves well to ordinary discourse.

Unfortunately, the Greeks, who gave us half-serious myths to account for various aspects of human nature, left us here to mythologize on our own. Let us then say, half-seriously, that our story begins in the twilight before the dawn of our day, when Pythagoras of Samos walked upon the rocky shore by Homer's fabled wine-dark sea and found himself dreaming of a mystical union between heaven and earth. The vision he beheld would inform his whole life. Arithmetic, with grand contempt for the slippery pebbles' uncertain support, boldly vaulted from the earth. Geometry was drawn out of the stars.

Mathematics sprang from this marriage, to be nurtured in the bright morning of a new era, when Greeks began to walk like giants, to wrest secrets from the gods. Why did Pythagoras lengthen his stride? What spirit moved the train—Eudoxus to Archimedes—that followed Pythagoras down the shore? We are still transported by this caravan of Greeks.

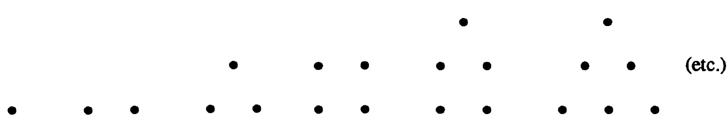
§1. The Philosophy of Pythagoras

Real mathematics—mathematics studied for its own sake as a worthy human interest—comes into particular prominence in Greek civilization with Thales (ca. 624–547 B.C.) and Pythagoras (ca. 572–497 B.C.), both of whom are reported to have traveled widely and learned much in Egypt, Babylonia, and perhaps in the Orient. Why the spark of mathematics should then glow so brightly in Greece, and why the flame should die some three or four centuries later with the coming of the Romans, is still not widely understood.

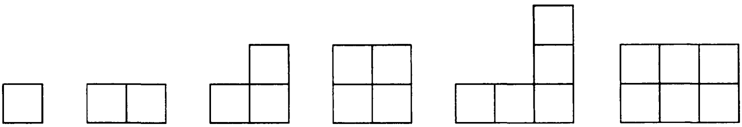
Geometry became increasingly the dominant theme in Greek mathematics, but Pythagoras was at first more attracted to arithmetic. He and his followers, the *Pythagoreans*, formed a society of men and women—Pythagoras laid down and practiced the principle of equal opportunity for both sexes—that virtually worshipped numbers. One short sentence is all it takes to sum up the philosophy of Pythagoras:

All is number.

According to tradition Pythagoras was the one who first put together two Greek words to make the word *philosophy*, which literally means "love of wisdom". He sought wisdom by studying numbers. Number, to the Pythagoreans, meant a "multitude of units", the first few numbers being represented by the sequence beginning



The **unit** need not be a point or pebble as pictured here, but might be an arbitrarily chosen geometric magnitude, such as a line segment of a fixed length. Today we might choose a foot or a meter as our unit. If our measurements are of two-dimensional magnitudes, it would be natural to use a square, perhaps a square foot. In three dimensions a cube, perhaps a cubic meter, might serve as the unit of reference. The essential thing is that, once a unit is established for discussion, the numbers of which we speak consist of multitudes of identical copies of the established unit.



(etc.)

The unit itself was not considered to be a number by the Greeks since "number" referred to a multitude rather than to a single object. This view of the unit did not completely fade away until the seventeenth century, when it became obsolete with the use of decimal fractions, where every point on the "number line" represents a number. The point representing the unit length then lost forever any special status it might have earlier enjoyed. Now, of course, the Hindu-Arabic numerals 1, 2, 3, etc. are universally used to symbolize the unit and the numbers it generates, which we now call *positive integers*. Needless to say, the Greeks did not consider negative integers or zero as possibilities for numbers.

What about fractions, like 3/2, that we use today? It may be helpful to recall first that the word *fraction* has to do with fracturing or breaking. Today we consider 3/2 to be the number we reach on the number line by putting three unit lengths together and breaking this sum in two. The length that results has the same relation to the unit length as 3 has to 2. Its decimal representation of 1.50 has the same size relative to the unit 1.00 as 3 has to 2, which is, of course, why we say today that 3/2 *is* 1.50 and consider 3/2 as a perfectly good number. As a result of adopting such decimal representations for numbers, we have conflated the idea of number with the idea of a point on the number line. For us today, of course, fractions are perfectly good numbers.

The Greeks resisted such a conflation by maintaining a clear distinction between numbers (multitudes of units to be studied in arithmetic) and magnitudes (such as line segments to be studied in geometry). But they were continually on the lookout for comparisons between things in the form of a **ratio**—"the size of one thing relative to another"—which enabled them to make many remarkable connections. It was the notion of a ratio such as 3:2 (which is essentially equivalent to our fraction 3/2 or our decimal 1.50) that enabled the Greeks to deal with fractions much as we do.

For the Greeks, the unit generates all numbers, whose ratios—according to the Pythagorean faith—have the power to measure all things. The testing of this faith helped to promote a great deal of research in mathematics. Many contributions were made during the first hundred years or so after the death of Pythagoras by Hippocrates and Archytas, two of the most notable Pythagoreans. Profound advances were made in the middle of the fourth century B.C. by Eudoxus, the most brilliant of the mathematicians at Plato's Academy in Athens. Following Eudoxus a great body of research was produced by the long train of mathematicians who studied for a while—or worked for a lifetime—at the museum and library in Alexandria, built near the end of the fourth century by order of Alexander the Great. Among these were Euclid, Archimedes, Eratosthenes, and Apollonius, who continually advanced mathematics throughout the third century B.C. As we shall see, Archimedes came close to developing the calculus.

Exercises

1.1. The Pythagoreans contributed significantly to the rise of mathematics, science, philosophy, and liberal education, but in addition to these "rational" pursuits they developed religious beliefs tending toward mysticism. In their numerology odd numbers were "male" and even numbers were "female". Two symbolized Woman. Three symbolized Man. Five was then the number of Marriage, the union of man and woman. Four symbolized Justice. And so on.

(a) Why did the unit symbolize God? *Hint.* Read the beginning of the last paragraph preceding these exercises.

(b) Why did Four symbolize Justice? *Hint.* Justice has to do with "squarely" balancing the claims of one against the claims of another. What does this have to do with Four? (See exercise 1.3.)

(c) Ten was a sacred number to the Pythagoreans but, perhaps surprisingly, not because we have ten fingers. Ten symbolized the cosmos—the Greek name Pythagoras gave to the universe. In what way is this symbolism appropriate? *Hint.* Though there are a couple of ways one might respond, it is relevant to note that the figurate representation of Ten as a *triangular* number (see exercise 1.5) shows Ten = One + Two + Three + Four.

(d) Pythagoras preached that human beings have a *psyche*—a "soul"—that is reborn in another body upon the death of the body in which it resides. *Metempsychosis* is the Greek word for such a passing. Can you think of anything in Pythagoras's study of mathematics that might have led him to such a fantastic idea? *Hint.* See the hint for exercise 1.2.

1.2. A "figurate number" refers to the representation of a positive integer by a figure made up of dots or pebbles. An **even** integer can be represented as a rectangular array with two dots on one side of the rectangle; otherwise the integer is called **odd**. Using these definitions, write an argument proving that

(a) the sum of two even numbers is even.
(b) the sum of two odd numbers is even.
Hint. Reread the second sentence in Chapter 1. Then play with some pebbles, imagining how Pythagoras might have played with pebbles 2500 years ago on a beach near Crotona, Italy. Perhaps (through metempsychosis?) the same idea will come to you as came to him.

1.3. The **square** of a figurate number is, of course, the number made by "squaring", e.g., the square of · · is ∷. A positive integer is a square if it is possible to represent it as a square array of dots. Prove that
(a) the square of an even number is even.
(b) the square of an odd number is odd.
Hint. If you play with dots in a few examples, you'll see how to make a general proof.

1.4. The sum of consecutive odd numbers is easy to evaluate, once you see the pattern.
(a) What is the sum of the first two odd numbers?
Answer: · plus ⋰ is ∷, the square of two.
(b) What is the sum of the first three odd numbers?
Answer: · plus ⋰ plus ⁙ is ⁞⁞⁞, the square of three.
(c) *Do you see the pattern?* The sum of the first four odd numbers will turn out to be the square of ... (?)
(d) The sum $1 + 3 + 5 + 7 + 9$, which is the sum of the first five odd numbers, will turn out to be the square of ... (?)
(e) What is the sum of the first hundred odd numbers?
(f) What is the sum of the first n odd numbers?
(g) How many odd numbers are in the sum $1 + 3 + 5 + \cdots + 99$? What is the sum?

1.5. The sum of consecutive integers is almost as easy to see. The Pythagoreans knew that an easy way to count the number of pebbles arranged in the shape of a triangle is to view the triangle as half of a rectangle:
(a) What is the sum of the first two positive integers?
Answer: · plus : is ⋰, which is half of ⁞⁞, so the sum of the first two positive integers is $\frac{1}{2}(2)(3)$.
(b) What is the sum of the first three positive integers? *Answer*: It is $\frac{1}{2}(3)(4)$, or half the dots in a 3 by 4 rectangle.
(c) *Do you see the pattern?* The sum $1 + 2 + 3 + 4 + 5 + 6 + 7$ is half the dots in a 7 by 8 rectangle, so this sum is equal to ... ?
(d) What is the sum $1 + 2 + 3 + \cdots + 100$? *Hint.* This is half the dots in a rectangle of size 100 by 101, so this sum is equal to ... ?
(e) What is the sum $1 + 2 + 3 + \cdots + 1000$?
(f) What, then, is the formula for the sum of the first n positive integers?
(g) Does your work in this exercise destroy the magic in problem 20 at the end of Chapter 1?

1.6. Exercises 1.4 and 1.5 show that the square numbers are the sums of consecutive odd integers and that the triangular numbers are the sums of consecutive integers (both sums beginning with unity). You probably know that the prime numbers are 2, 3, 5, 7, 11, 13, 17, etc. A **prime** is an integer exceeding unity that has no proper divisors except itself and unity. Describe the prime numbers in terms of their figurate representations. *Hint.* What connection do the *rectangular* numbers have with the numbers that are *not* prime?

1.7. (a) The square numbers are 1^2, 2^2, 3^2, etc. so it is obvious that the n-th square number is n^2. The triangular numbers are 1, 3, 6, 10, etc. Prove that the n-th triangular number is $n(n+1)/2$. *Hint.* This is easy if you see the connection between this question and the last part of exercise 1.5.

(b) Show, using figurate numbers, that any square (except for 1^2) is the sum of two consecutive triangular numbers. (For example, the third square is $3^2 = 9 = 3 + 6$, which is the sum of the second and third triangular numbers.) *Hint.* Given a square array of dots, split it into an upper triangle and a lower triangle, with the bottom of the upper triangle consisting of the dots on the square's main diagonal.

(c) Write 10,000 ($= 100^2$) as the sum of two consecutive triangular numbers. *Answer*: $10,000 = 4950 + 5050 =$ sum of the 99th and 100th triangular numbers.

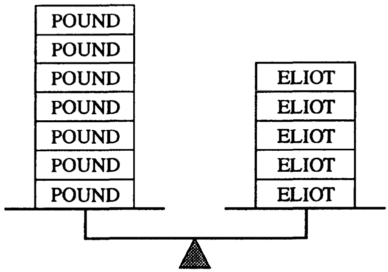
(d) Write 1,000,000 ($= 1000^2$) as the sum of two consecutive triangular numbers.

§2. Six Famous Ratios—Are they Ratios of Integers?

What the Pythagoreans mean by asserting that all is number is not entirely clear. Perhaps they mean that ultimately everything is determined in some fashion by the positive integers. At the least they mean that numbers are connected with many things that, at first, seem totally unrelated to numbers. For example, the musical tones produced by plucked strings seem at first to have nothing to do with numbers. Yet it was Pythagoras himself, so legend has it, who discovered the number ratios associated with the relative pitches that make up the easily recognized musical intervals of thirds, fifths, octaves, etc. When plucked, two lengths of the same string (pulled taut with the same tension) sound exactly an octave apart if their lengths are in the ratio 2:1. Elementary facts about music are such common knowledge today that we surely underrate their significance. In the sixth century B.C. their discovery must have been astonishing. Imagine! Numbers have something to do with music.

It is then easy to take the philosopher's leap: If numbers have something to with music, perhaps number is the fundamental principle behind all things. Perhaps everything is number. Here Pythagoras does not follow his teacher Thales, who had proposed that water, because of all its various manifestations, might be the basis for all things.

The fact that *ratios* of string lengths—rather than simply string lengths alone—enter into the theory of music reminds us that measurement deals not with size itself but with the comparison of sizes. The same is true in measuring time or in measuring weights. When we say, for example, that a book of T.S. Eliot's poetry weighs 1.4 pounds we mean that the ratio 14:10, or 7:5, describes the weight of the book in proportion to a pound weight, our standard unit at the moment. This in turn means, of course, that a stack of 7 of these standard units would exactly balance with a stack of 5 copies of the book.



A perfect balance if one book weighs 1.4 pounds

We easily overlook how much is involved in the process of measurement. A single fraction like 7/5 measures nothing until the unit of measurement is revealed, in which case we have a **proportion** (the proportion being, of course, that the ratio of the object measured to the unit is 7:5). Measurement cannot be done by a number without knowledge of what unit is being taken for consideration, and measurement also requires an understanding of ratio and proportion. At the risk of belaboring the point, let us repeat that to say a book weighs 1.4 pounds is to say that the following proportion holds:

$$\text{book's weight} : \text{pound weight} :: 7:5.$$

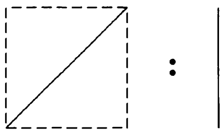
The proportion we have just written in compact symbols would have been expressed by the Greeks in a long sentence: "The ratio of the book's weight to the pound weight is 'squarely balanced' by the ratio of seven units to five units." The Greeks rarely used abbreviative symbolism.

Let us repeat the same thing, but in greater generality. If A and B are two weights (or *volumes*, or *times*, or *areas*, or *lengths*), then when we say that the ratio $A : B$ can be expressed as $m : n$, where m and n are positive integers, we mean that we think n copies of A is "equivalent" to m copies of B. In other words, if we have the proportion

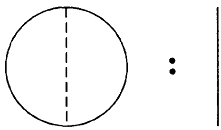
$$A : B :: m : n,$$

then we can find a unit such that A is equivalent to m copies of the unit and B is equivalent to n copies. We simply declare that the unit be taken as what we get by dividing A into m equal parts—or, equivalently, by dividing B into n equal parts. In this "co-measurable" case, when A and B are each measured by taking appropriate multitudes of the same unit, then A and B are said to be **commensurable**.

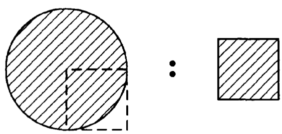
We can hardly say that we know much about squares, circles, cylinders, spheres, pyramids, and cubes unless we can measure their sizes relative to each other, or the ratios between the sizes of parts of any one of them. Here are six ratios of magnitudes. The Pythagoreans believed that each of these ratios of magnitudes could be expressed as ratios of integers. But how can one possibly find such ratios of integers?



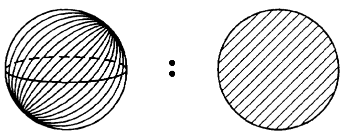
What is the ratio $r_1 = d : s$ of the length of a diagonal of a square to the length of its side?



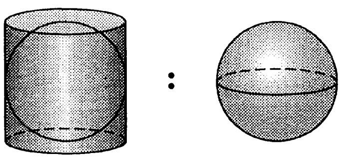
What is the ratio $r_2 = C : D$ of the circumference of a circle to its diameter?



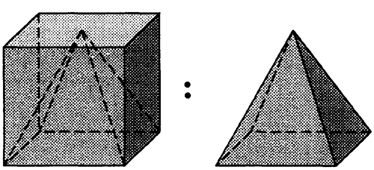
What is the ratio $r_3 = A : r^2$ of the area of a circle to the area of a square built upon its radius?



What is the ratio $r_4 = S : A$ of the surface area of a sphere to the area of a circle through its "equator"?



What is the ratio $r_5 = V_C : V_S$ of the volume of a cylinder to the volume of an inscribed sphere?



What is the ratio $r_6 = V_Q : V_P$ of the volume of a cube to the volume of an inscribed pyramid?

Pondering the six questions above will give the reader an idea of what research in Greek mathematics was like. *The reader is not expected to know the answer to any of these six questions at this point.* In fact, for a couple of these ratios the Greeks themselves never knew for certain whether they were equivalent to a ratio of integers or not.

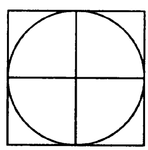
Notice that the first two ratios each deal with a pair of "one-dimensional" magnitudes, the next two deal with "two-dimensional" magnitudes, and the last two with "three-dimensional" magnitudes. It was Pythagoras's belief that any two *comparable* geometric magnitudes (meaning, roughly, that the magnitudes were both finite in size and of the same dimension) are commensurable, i.e., their ratio is equal to the ratio of some pair of integers. It is, of course, not immediately clear whether this is true, or how one might go about deciding whether this is true.

Yet the later Pythagoreans and others made much progress on these and other questions. Perhaps this had something to do with the fact that, around 400 B.C., the Greek word *mathemata*—which referred to "knowledge" or "learning in general"—was contracted to *mathema*, and the discipline now known as mathematics received its proper name.

Exercises

(*Most of these exercises deal with the ratios* $r_1, r_2, \ldots, r_6$ *defined above.*)

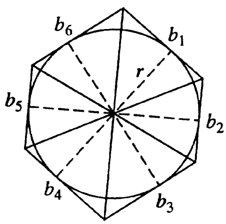
2.1. Prove that the ratio r_3 is less than 4:1. *Hint.* Let the unit for this discussion be the square built on the radius of the circle. All you have to do is to explain why the area of the circle is less than 4 of these unit squares. Nothing could be simpler.



2.2. Prove that the ratio r_2 is less than 4:1. *Hint.* Let the unit for this discussion be the length of one side of a square that is circumscribed about the circle. All you have to do is to explain why the path around the circumference is less than 4 of these units. You can use the same picture as in exercise 2.1, but the unit now is a line segment rather than a square.

2.3. Consider a square circumscribed about a circle. Prove that the ratio $A : r^2$ of the square's area to the square of the circle's radius is equal to the ratio $P : D$ of the square's perimeter to the circle's diameter. *Hint.* In the course of working exercises 2.1 and 2.2 you have already found both ratios. Are they equal?

2.4. (*Generalizing exercise 2.3*) Consider a polygon of a large number of (not necessarily equal) sides that is circumscribed about a circle. Prove that the ratio $A : r^2$ of the polygon's area to the square of the circle's radius is equal to the ratio $P : D$ of the polygon's perimeter to the circle's diameter, i.e., prove that the proportion $A : r^2 :: P : D$ is true. *Hint.* First consider a polygon of, say, six sides of lengths $b_1, b_2, \ldots, b_6$, as in the figure here. Let A_6 denote the area of the polygon and let P_6 denote its perimeter (so $P_6 = b_1 + b_2 + \cdots + b_6$). By splitting up the polygon into triangles as indicated, show first that $A_6 = \frac{1}{2}rb_1 + \frac{1}{2}rb_2 + \cdots + \frac{1}{2}rb_6 = \frac{1}{2}rP_6$. Then divide both sides of the equation $A_6 = \frac{1}{2}rP_6$ by r^2. Does this show that $A_6 : r^2 :: P_6 : D$, where $D = 2r =$ diameter of circle? Do you see now how to make a proof if the polygon has more than six sides?



2.5. (*You are simply asked to make guesses here, in the hope that it will make you curious to learn how the Greeks were able to answer such questions without guessing.*) At this stage you are not expected to be acquainted with the exact numerical value of any of the six geometric ratios r_1, r_2, r_3, r_4, r_5, r_6 defined above. Nevertheless, your intuition about relative sizes should enable you to make fairly confident guesses.

(a) Archimedes proved that one of these six ratios is expressible as 4 : 1. Can you guess which one?

(b) Archimedes also proved that one of them is the ratio 3 : 2. Guess which one.

(c) Eudoxus proved one of them is 3 : 1. Guess which one.

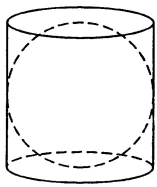
(d) Do you think any of the ratios is equal to 2 : 1? Which ones, if any?

(e) Ratios can be ordered in an obvious way. (3 : 2 is "smaller" than 2 : 1 since $1.50 < 2.00$.) Make a guess as to which of the six ratios is the smallest.

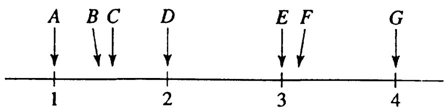
(f) Guess which is the largest.

(g) Archimedes proved that a certain pair of the six ratios are actually identical. (He proved that neither is larger than the other.) Can you guess which pair? *Hint.* The point of exercises 2.3 and 2.4 was to give you clues for part (g).

2.6. (*One more guess*) Let S_1 denote the surface area of a cylinder—just the "side" of the cylinder, not including the area of its circular top and bottom. Let S_2 denote the surface area of a sphere inscribed within the cylinder. *Make a guess* as to the size of the ratio r_7 defined as $S_1 : S_2$. In particular, guess which of S_1 and S_2 is larger. Or are they equal? (It was nearly 300 years after Pythagoras before Archimedes settled this question.)



2.7. The modern way to look at a ratio is as a real number identified with a decimal expansion, which in turn is identified with a point on the "number line". Thus the ratio 3:2 is identified with 3/2 or 1.50. The ratio 4:1 is identified with the point 4/1 or 4.00. Try to match the ratios r_1 through r_7 with the appropriate points. *Caution*: Not all seven points A through G pictured here will be needed, for the ratios are not all different. *Just make a guess here. We cannot be sure of the answers until we develop the calculus to find them.*



2.8. (*The greatest Greek mathematician*) From the information in exercises 2.5 and 2.6 it should not be hard to guess the name of the greatest Greek mathematician. He is, of course, the one who did the deepest research on the famous ratios introduced in this section, even though others had already had some 250 years since Pythagoras to think about them before he came along. What is his name?

§3. The Spirit of the Liberal Arts

Thales is said to have attached great importance to logical arguments that begin from clear premises, and Pythagoras wholeheartedly endorsed this principle. An argument is now seen as having a life of its own, independent of its origin, so that truth arrived at by argument—unlike the 'truth' of a mysterious oracular pronouncement—may thereby appear to have an existence independent of human beings. Such new attitudes are expressions in mathematics and philosophy of the rise of a civilization whose works of sculpture, architecture, and literature are also so celebrated that one is tempted to ascribe almost superhuman power to the Greeks. But much, if not all, of that power can be seen to flow instead from the fullness with which they enjoyed their very human nature.

Indeed, much of the initial growth of mathematics and philosophy can be seen as an overflow into mental activity of certain human characteristics long present in the physical arena. The spirit of competitiveness and play is at least as old as mankind, and the Greeks—who had established the Olympic games in 776 B.C.—were second to none in their celebration of athletic prowess. Over the course of the next couple of centuries, various bands of Greeks (the Pythagoreans were only one of several) became intrigued by new kinds of competitive challenges inherent in word-play, dialogues, problems, and paradoxes. The emergence and rapid growth of Greek interest in serious thought owes much to this popularization of the sportive aspect of the life of the mind.

The concurrent development of an unmistakably modern sense of humor is further evidence of a growing delight in the play of the mind. Consider this short tale about Dionysius I (ca. 430–363 B.C.), dictator of Syracuse, who fancied himself to be a poet.

> When the poet Philoxenus, asked by the dictator for his opinion of the royal verses, pronounced them worthless, Dionysius sentenced him to the quarries. The next day the King repented, had Philoxenus released, and gave a banquet in his honor. But when Dionysius read more of his poetry, and asked Philoxenus to judge it, Philoxenus bade the attendants take him back to the quarries.

There are many such wonderfully droll stories that have come down to us because the Greeks loved to repeat them. Wit and vitality of mind are mutually reinforcing, and the Greeks profited greatly from this symbiotic relationship.

The life of the mind must have expanded even further with the realization that it furnishes a new arena for the display of courage—for many Greeks, the greatest of virtues. No longer restricted by purely physical considerations, courage can now be seen in the perseverance through mental strength to the end of an argument without fear of where the argument may lead. A voyage through strange seas is just as demanding when it is taken in the mental world. It is also just as prone to shipwreck. We may lose everything when an argument leads to absurdity and thus produces the conclusion that one of its premises must be false. On the other hand, when we exercise proper foresight (as in the logical technique of **reductio ad absurdum**, to be introduced in the exercises below), that conclusion may be exactly what we seek. The play of the mind is the greatest of all open-ended games. Each unexpected twist and turn we take in wrestling with a significant problem brings to us the possibility of a new vision of the world.

Historians already credit the spirit of competition and play with a profound effect upon art and literature, for by the sixth century B.C. the Greek games had expanded far beyond athletics. Eventually the Olympic torch sparked public competitions in pottery, poetry, sculpture, painting, choral singing, and drama. Mathematics and philosophy were, of course, even more directly encouraged by this expansion of the Olympic spirit.

Was this growth of spirit an inevitable step in the evolution of human beings? Or was it something that could have happened only at this particular time and place? Who can best explain what animated the Greeks, and especially the Pythagoreans, despite (or was it because of?) their mysticism, numerology, and other idiosyncrasies. History, anthropology, psychology, evolutionary biology, and theology would doubtless give characteristically different responses.

Most people would probably agree, however, that it was a quintessentially human spirit that lifted the Greeks above provincialism and toward the kind of universality reflected by their establishment of democracy in government and of the liberal arts in education. The word *liberal* as used here is supposed to have the central connotation of freedom and to denote those arts whose study will produce a liberated mind worthy of a free citizen of a democracy, as opposed to one enslaved by ignorance, prejudice, superstition, or fear. Tradition holds that Archytas of Tarentum, a Pythagorean born around 428 B.C., conceived the idea of what was much later called the **quadrivium** (the "four-fold way")—consisting of arithmetic, geometry, music, and astronomy—as the basis for an education with the power to make free and independent citizens out of *hoi polloi*.

The Pythagoreans revolutionized the study of arithmetic and geometry, and were so enthralled by the science of music that they undertook the technical study of musical scales, harmony, etc. They saw arithmetic playing a role in music analogous to the role of geometry in astronomy. Arithmetic studies numbers "at rest"; whereas music—in studying the harmony of tones produced by vibrating strings of proper ratios—is the study of numbers "in motion". Similarly, geometry studies magnitudes at rest; whereas astronomy—in seeking order in the movements of the heavenly bodies—is the study of magnitudes in motion.

Despite his mystical leanings Pythagoras is sometimes regarded as the founder of Western science because he continually promoted mathematics as a means of finding harmony and order in the natural world. He even "objectified" the notion of order by taking the Greek word for it, *cosmos*, and giving this name to the universe. Such audacity may have helped inspire Plato's attempt, much more radical and two centuries later, to objectify things like Truth and Beauty.

The Greeks emphasized the central role played by ratios in liberal education. Ratios not only permeate the "scientific" quadrivium but also appear, loosely disguised, in the type of discourse that would later be called humanistic. For example, the Pythagorean analogy between pairs of members of the quadrivium may be expressed "rationally" by writing

ARITHMETIC : MUSIC : : GEOMETRY : ASTRONOMY,

which shows how an analogy (a rhetorical device) is strikingly like a proportion (a mathematical device). "To measure is to know" makes as much sense in rhetoric as it does in mathematics.

These is no doubt the Greeks saw the connection between a proportion and an analogy because they used the same word (*analogos*) to refer to both. Other connections between mathematics and rhetoric were probably much more apparent to the Greeks than to us today. The Greeks expressed almost all their mathematics rhetorically, i.e., they wrote nearly everything out in words, rarely using abbreviative symbolism. For example, in proportions they never employed the now-familiar symbol made by a square of four dots to signify equivalence. This convention was not introduced until the seventeenth century, even though the idea of a delicate balancing—as on the scales of Justice—of one side "squarely" against another hearkens back to the Pythagoreans, whose numerology identified Four with Justice. The seventeenth-century proliferation of abbreviative notation in mathematics and the consequent large-scale algebraic and logical manipulation of symbols has proven to be enormously efficient, but some of the ties between mathematics and rhetoric have thereby been obscured.

Skillful use of analogy is central to rhetoric, yet every analogy is implicitly dependent upon the notion of a ratio. There are similar parallels between logical arguments and mathematical proofs, and between grammatical clarity and mathematical precision. The quadrivium was naturally conjoined with the **trivium** of grammar, rhetoric, and dialectic (or logic) to form the Greek *enkuklios paideia*, a phrase from which we get "encyclopedia", but which is better translated as "general education" or "the usual everyday education received by all". Since *kuklios* means "circle", it is tempting to read into this phrase the connotations of unity and well-roundedness, two attributes that, at any rate, characterize the program. The Greeks used generalized notions of *ratio*, *proportion*, *rhythm*, and *harmony* to expose common threads that tie these seven disciplines into a whole.

Integrity—"wholeness"—in education is a Greek ideal that seems less highly prized today. Even the Greeks were strained, however, to study the quadrivium as thoroughly as the trivium. Some elements of the trivium were clearly prerequisites to the study of the arts (later to be called "sciences") of the quadrivium.

Greek democracy was relatively short-lived, but the *enkuklios paideia* survived the long period of Roman domination. The Romans adopted Greek education along with some other aspects of Greek culture, although they virtually ignored all mathematics that was not immediately practical and therefore paid only a little attention to the quadrivium.

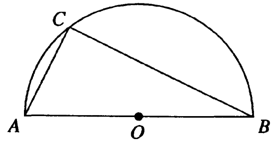
Despite its "pagan" origin this Greco-Roman system of education—now known by its Latin name, *artes liberales*—eventually received the grudging endorsement of the Roman church. With the church it survived the fall of Rome and came to flourish in late antiquity as the **seven liberal arts**. The word "art" as used here, of course, means "discipline" and has little to do with the modern sense of reference to art as a kind of personal or subjective creativity. In medieval European universities the Bachelor of Arts degree for four years' study of the trivium, followed by a Master of Arts degree for three additional years of the quadrivium became the usual prerequisite for advanced study in such things as law, medicine, science, philosophy, or theology. The liberal arts tradition today can thus trace its existence back continuously for nearly 2500 years.

Exercises

3.1. (*What sort of arguments did Thales make?*) According to tradition Thales was known for emphasizing the importance of justifying theorems by showing how they follow logically from simple premises. Here are two examples.

(a) *In a triangle the angles opposite equal sides are equal.* Prove this theorem the way you did in geometry class. Thales was supposedly the first to feel the need of proving this simple theorem about isosceles triangles, and the first to give a proof.

(b) The most striking theorem associated with Thales' name is one he may, in fact, have never proved: *Any angle inscribed in a semicircle is a right angle.* Assuming that the line AB in the figure below is the diameter of a circle and that the point C lies on the circle, prove that the angle ACB is a right angle. *Hint.* Draw the line segment OC, where O is the center of the circle and notice that two isosceles triangles are formed, so you can immediately use the result of part (a) on each one. The rest of the argument is up to you.



3.2. (*A taste of grammar*) In the classical liberal arts, grammar included etymology. Explain how the hypotenuse got its name. *Answer*: Suppose you lived in Thales' time and you wanted to give an appropriate name to the side of a triangle opposite a right angle. If you are in the habit of drawing your right triangles lying in a semicircle as in the figure of exercise 3.1, this is the side "stretched beneath", or "subtended". The Greek word for "beneath" is *hypo*—as in *hypodermic*, meaning "beneath the skin". The Greek word meaning "to stretch" is *tenein*.

3.3. (*A taste of logic*) In the answers to exercise 3.1 we have examples of direct arguments. They begin from simple statements and deduce more subtle assertions in a straightforward way. An indirect argument, or an argument by *reductio ad absurdum*—also called a **proof by contradiction**—proves an assertion by showing that an absurdity (or contradiction) follows from supposing the assertion to be false. It is often a useful technique for proving that something does not exist or that something cannot happen. Give a short *reductio ad absurdum* argument to prove each of the following propositions:

(a) There is no largest integer. *Answer*: Suppose there were a largest integer. Call it N. Let M denote the integer we get by adding another unit to N. Thus $M = N + 1$. But then M is an integer larger than the largest integer N, an absurd conclusion. Since an absurd conclusion follows from the supposition that there is a largest integer, this supposition must be false. Therefore, there is no largest integer, Q.E.D.

(b) There is no largest even integer. *Hint*. Suppose there were; call it N. let $M = N + 2$, and argue much as in part (a) to reach an absurd conclusion. You may need to use the result of exercise 1.2(a) to nail down the contradiction.

(c) There is no largest multiple of 10. *Hint*. Part (b) showed there was no largest multiple of 2. This can be proved in the same way. Just change "2" to "10" and argue as in part (b). [Or take Ten as the unit and argue as in part (a).]

3.4. (*More logic*) Give *reductio ad absurdum* proofs of each of the following.

(a) If the square of an integer is even, then the integer itself must be even. *Answer*: Suppose n is an integer whose square is even and suppose n is odd. Then n^2 is even and n is odd. But we know [by exercise 1.3(b)] that n^2 must be odd if n is odd. This is a contradiction, for n^2 cannot be both odd and even. Therefore, if n^2 is even, then n must be even, Q.E.D.

(b) If the square of an integer is odd, then the integer itself must be odd. *Hint*. Exercise 1.3(a) can be used here the way 1.3(b) was used in the answer to part (a).

3.5. ("*Scarecrow Logic*") In the movie *The Wizard of Oz*, the scarecrow—having just been given a brain—spouts nonsense, something like the following: "The square root of the hypotenuse of an isosceles triangle is equal to the sum of the square roots of the two equal sides." Show that such a triangle does not exist. *Hint*. Suppose there were such a triangle and let its sides have lengths a, a, and b, with b as the length of the "hypotenuse". Then according to the scarecrow, $\sqrt{b} = \sqrt{a} + \sqrt{a}$, so $\sqrt{b} = 2\sqrt{a}$, and by squaring both sides we have $b = 4a$. Thus the sides of the triangle are a, a, and $4a$. This is absurd! (*Why?*)

3.6. (*A taste of rhetoric*) Just for fun, try to complete the following analogies. The first is adapted from a playful remark of Will Durant's; the second alludes to something Mark Twain said about writing; the third is something Archimedes proved about a circle.

(a) "Just as astronomy is a trick that geometry plays upon the eye, so music is a trick that ____ plays upon the ____."

(b) A good word : the best word : : a lightning bug: ____.

(c) area of a circle : square of its radius : : circumference of a circle: ____.

Answer: (b) the lightning.

3.7. What justification is there for intelligence tests to rely upon questions about analogies?

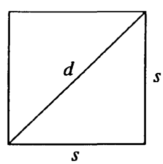
3.8. (*Suggestions for serious readers*) Howard DeLong perceptively noted the importance of the sportive aspect of the life of the mind in the first chapter of his much-praised *Profile of Mathematical Logic*, Addison-Wesley, 1970. The account of Dionysius and Philoxenus given in this section comes from the lively pen of Will Durant in his *Life of Greece*, Simon and Schuster, 1939. A beautifully written book describing the origin of the liberal arts is H.I. Marrou's *History of Education in Antiquity*, translated by George Lamb, Sheen and Ward, 1956. Check out one of these from a library and read ten pages.

§4. The Pythagorean Theorem and Its Modern Consequences

The most famous theorem associated with the Pythagoreans is one with which the reader is doubtless already familiar and whose content was at least partially known in Babylonia, China, and India long before Pythagoras. The Babylonians, for example, had earlier made tables listing many of the possible integral values a, b, c for the sides of a right triangle satisfying the relation $c^2 = a^2 + b^2$, including the most familiar "Pythagorean triple" of 3, 4, and 5.

By virtue of the Pythagorean theorem we can give another name to the first of the six ratios introduced in Section 2. The ratio r_1 of the length d of a diagonal of a square to the length s of its side can be numerically expressed as $\sqrt{2}$. That is, if we were to conflate the ancient and modern ways of dealing with numbers, fractions, and ratios, we would write

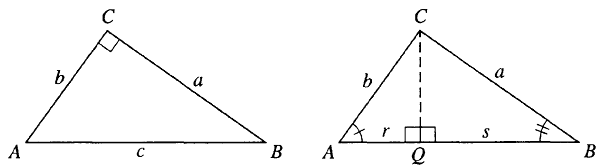
$$\begin{aligned} r_1 &= d : s \\ &= d/s \\ &= \sqrt{2}. \end{aligned}$$



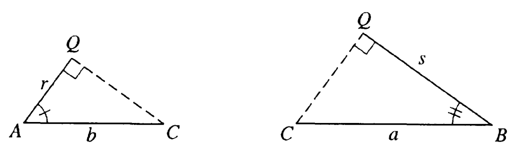
To see this, simply note that d is the hypotenuse of an isosceles right triangle whose two legs have length s so the Pythagorean theorem says that $d^2 = s^2 + s^2 = 2s^2$. From the fact that $d^2 = 2s^2$, we get $d^2/s^2 = 2$, which means $d/s = \sqrt{2}$.

Note that the ratio of diagonal to side of a square is $\sqrt{2}$ *regardless of what size the square is*. This is, of course, because any two squares are **similar**, i.e., one is the "same shape" as the other, just pictured in a different scale. We shall take it as intuitively clear—though the Greeks would prefer to discuss this point much more carefully—that the ratio of two parts of a figure is unchanged if the figure is simply rescaled to make something similar. (Think about what happens when a photograph of your face is enlarged: The ratio of the size of your smile to the size of your face must be the same in both pictures.)

If there is any justification for naming this theorem after the Pythagoreans, it is that they were probably the first to give an argument to prove it. Today we know many different proofs. The sophisticated argument presented below is rather like the one Euclid gave in his *Elements* some two hundred years after the death of Pythagoras, except we have modified it a bit to emphasize how ratios can be brought to bear. The key to this proof is the equivalence of ratios of certain pairs of sides when a perpendicular is dropped from the vertex of a right angle upon the hypotenuse.

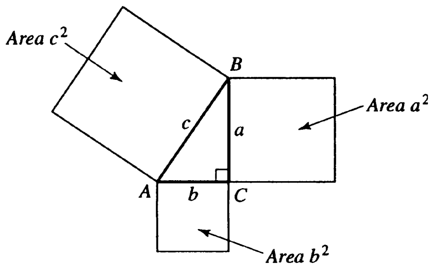


It is obvious that the two smaller right triangles AQC and BQC each have two angles of the same size as those of the original triangle ABC. It follows (*why?*) that all three angles must be the same. Thus all three triangles are similar—each is the same shape as the other two, simply drawn to a different scale.



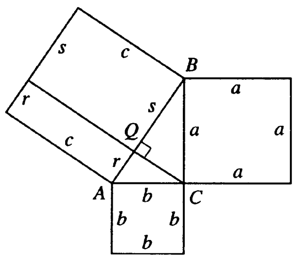
The similarity of the two triangles just above with the right triangle originally given means that ratios of comparable pairs of sides in each are equivalent. Thus we see that $r : b :: b : c$ and $s : a :: a : c$. But it is easier for modern readers to follow Euclid's argument if we re-express the equivalence of these ratios in modern terms as equalities of fractions, as we do in equations (2) and (3) in the proof below.

The Pythagorean Theorem
In a right triangle, the square built on the hypotenuse has the same area as the combined area of the squares built on the other two sides.



Proof

From the vertex of the right angle, drop a perpendicular to the hypotenuse, hitting the hypotenuse at Q.



The hypotenuse is then split in two, as indicated, so that

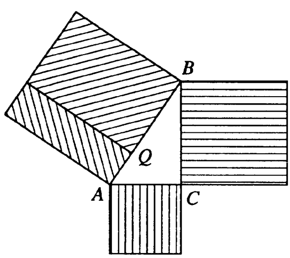
$$c = s + r. \tag{1}$$

Since $\triangle AQC$ is similar to $\triangle ACB$, we have

$$\frac{r}{b} = \frac{b}{c}. \tag{2}$$

Since $\triangle BQC$ is similar to $\triangle BCA$, it follows that

$$\frac{s}{a} = \frac{a}{c}. \tag{3}$$



From equation (2) it follows that $b^2 = rc$ (showing that the two figures with vertical markings are equal in area). From equation (3) it follows that $a^2 = sc$ (showing that the two figures with horizontal markings are equal in area). From the equations $a^2 = sc$ and $b^2 = rc$, together with equation (1), we have

$$a^2 + b^2 = sc + rc = (s + r)c = c \cdot c = c^2.$$

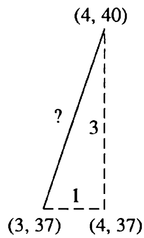
Therefore, $a^2 + b^2 = c^2$. □

The proof just given shows exactly how the square built on the hypotenuse can be split into two areas that are equal, respectively, to the squares built on the sides. Pythagoras probably gave a more elementary proof, perhaps like the one that is outlined in problem 3 at the end of this chapter.

The Pythagorean theorem is applicable to a surprising variety of situations, as indicated in the exercises that follow.

Exercises

4.1. In the plane, indicate the positions of the points P and Q whose coordinates are given below, and use the Pythagorean theorem to find the distance between P and Q.
(a) $P = (3, 37)$, $Q = (4, 40)$. *Answer*: Dist P to Q is $\sqrt{10}$ (*see figure below*).
(b) $P = (1, 55)$, $Q = (6, 50)$.
(c) $P = (4, 40)$, $Q = (2, 38)$.
(d) $P = (\pi, \pi^2)$, $Q = (\pi^3, \pi^4)$. *Answer*: Dist P to Q is $\sqrt{(\pi^3 - \pi)^2 + (\pi^4 - \pi^2)^2}$.



4.2. Use the Pythagorean theorem to find a formula for the distance from (x_1, y_1) to (x_2, y_2). *Answer*: Dist (x_1, y_1) to (x_2, y_2) is $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. (This is called the **distance formula** and should be memorized.)

4.3. (a) Use the distance formula to find the distance from $(0, 0)$ to $(3, 4)$.
(b) There are three "equals" signs in the calculation below. Tell which of the three are correct and which, if any, is incorrect.

$$\text{Dist}(0, 0) \text{ to } (3, 4) = \sqrt{3^2 + 4^2} = 3 + 4 = 7.$$

4.4. Use the distance formula to find the distance from $(0,0)$ to (x,y). Be sure you know why you answer is *not* equal to $x+y$. *Answer*: Dist $(0,0)$ to $(x,y)=\sqrt{(x-0)^2+(y-0)^2}=\sqrt{x^2+y^2}$.

4.5. Consider this sentence:

The distance from $(0,0)$ to (x,y) is 5.

Use your answer to exercise 4.4 to write this sentence as an equation using only algebraic symbols. *Hint.* In algebra the word "is" may be translated "equals".

4.6. Consider this sentence:

The point (x,y) lies on the circle of radius 5 with center at $(0,0)$.

(a) Explain why this sentence has exactly the same meaning as the sentence in exercise 4.5.
(b) Rewrite this sentence in algebra. *Hint.* In view of part (a) your answer to exercise 4.5 is also an answer to this question.
(c) Is an algebraic equation a sentence? (Does it have a subject, verb, and predicate?)

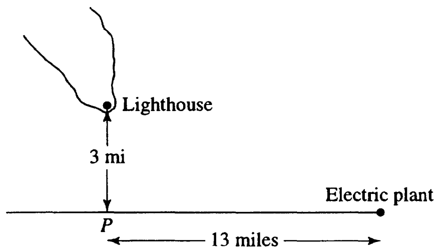
4.7. Consider this sentence: $x^2+y^2=25$.
(a) Explain why this sentence means that $\sqrt{x^2+y^2}=5$.
(b) Explain why the sentence "$\sqrt{x^2+y^2}=5$" means that (x,y) lies on the circle of radius 5 with center at $(0,0)$.
(c) Put your answers to parts (a) and (b) together to explain why "$x^2+y^2=25$" means that (x,y) lies on the circle of radius 5 with center at $(0,0)$.
(d) Explain carefully in what sense we might say that the equation $x^2+y^2=25$ "is" the circle of radius 5 with center at $(0,0)$. *Hint.* It is not meant literally, of course, because an equation is something from algebra and a circle is something from geometry. See the remarks on *analytic geometry* at the beginning of the next section.

4.8. Consider each of the following equations and rewrite it as a sentence in words.
(a) $x^2+y^2=49$. *Answer*: (x,y) lies on a circle of radius 7 centered at $(0,0)$.
(b) $x^2+y^2=36$.
(c) $x^2+y^2=25$.
(d) $x^2+y^2=1$.

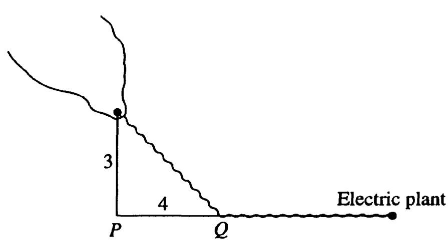
4.9. Rewrite each of the following sentences as an equation.
(a) The distance from (x,y) to $(\pi,3)$ is 5. *Answer*: $(x-\pi)^2+(y-3)^2=25$.
(b) (x,y) lies on the circle of radius 5 centered at $(\pi,3)$. *Answer*: Same as 4.9(a).
(c) The distance from (x,y) to $(2,3)$ is 5.
(d) (x,y) lies on the circle of radius 5 centered at $(2,3)$.
(e) (x,y) lies on the circle of radius $\sqrt{2}$ centered at $(1,0)$.
(f) (x,y) lies on the circle of radius 3 with center at $(-2,5)$.
(g) (x,y) lies on the circle of radius r with center at $(a.\,b)$.

4.10. Rewrite each of the following equations as sentences about circles.
(a) $(x+2)^2+(y-5)^2=9$. *Answer*: (x,y) lies on the circle of radius 3 centered at $(-2,5)$.
(b) $(x-2)^2+(y+5)^2=9$.
(c) $x^2+(y-2)^2=3$.
(d) $7x^2+7y^2=14$. *Hint*. First rewrite this as $x^2+y^2=2$. Then it is easier.
(e) $10(x+2)^2+10(y-4)^2=250$.

4.11. A cable is to be run from a large island to an electric power plant. Suppose a lighthouse is located at the point of the island nearest the power plant, which is 13 miles down a straight shoreline, while the lighthouse is 3 miles from the nearest point P on land. Suppose that undersea cable cost 7 thousand dollars per mile and underground cable costs 2 thousand per mile.



(a) What is the cost (in thousands) of running cable from the lighthouse undersea to P, then underground to the power plant?
(b) What is the cost (in thousands) of running cable undersea from the lighthouse to a point Q located 4 miles downshore, then underground to the plant?



(c) Find an algebraic rule giving the cost C (in thousands) of running the cable undersea from the lighthouse to a point located x miles downshore, then underground to the plant. *Answer*: $C=7\sqrt{9+x^2}+2(13-x)$.
(d) What is the *domain* of the function expressed in part (c)?

§5. Geometry versus Arithmetic

Pythagoras could have had little notion that the Pythagorean theorem would be relevant in such a variety of contexts as we have just seen in the exercises above. In particular, the idea that an *algebraic* equation (like $x^2 + y^2 = 25$) could be identified with a *geometric* curve (the circle of radius 5 with center at the origin) is an idea whose value the Greeks never fully realized. The importance of this interplay between algebra and geometry was first seen by two seventeenth-century Frenchmen, Pierre de Fermat and René Descartes. It was they who developed **analytic geometry**, the name given to the study of this interplay, whose goal is the attainment of a synthesis of algebra and geometry.

Ever since 1637, when Descartes wrote *La Géométrie*, it has been common knowledge that curves can have equations and that equations can determine curves. Why did the Greeks fail to utilize this means of approaching problems in geometry? The answer is simple. The Greeks knew that their curves had equations. They developed, however, little abbreviative symbolism and therefore had to write the "equations" out in words. For the Greeks, sometimes only a wondrous wealth of limiting clauses could adequately describe the mathematician's latest and prettiest discovery:

> Let a cone be cut by a plane through the axis, and let it also be cut by another plane cutting the base of the cone in a straight line perpendicular to the base of the axial triangle, and further let the diameter of the section be parallel to one side of the axial triangle; then if any straight line be drawn from the section of the cone parallel to the common section of the cutting plane and the base of the cone as far as the diameter of the section, its square will be equal to the rectangle bounded by the intercept made by it on the diameter in the direction of the vertex of the section and a certain other straight line....
>
> Apollonius, *Conics*, ca. 200 B.C.

See?

By contrast, our modern system, which uses Descartes's coordinates and his abbreviative notation, is almost magically efficient. In modern terms, the long statement of Apollonius simply says,

> Given a parabola, a Cartesian coordinate system can be introduced in which the parabola has an equation of the form $y = cx^2$, where c is a certain constant....

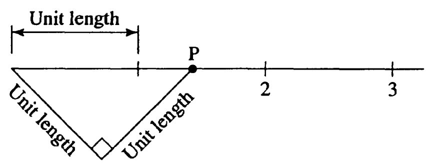
It is hard to overstate the value of appropriate symbolism. The Greeks never had it, and they developed only a little algebra. Their powers were concentrated upon geometry.

Why did the Greeks prefer to couch their mathematics in geometry? Why not let number play the key role in mathematics, particularly since Pythagoras would base everything upon numbers? The reason has to do with the discovery by the Pythagoreans of *irrational* quantities, a discovery that might be interpreted as disproving their own philosophy!

> It is told that those who first brought out the irrational from concealment into the open perished in shipwreck, to a man. For the unutterable and the formless must needs be concealed. And those who uncovered and touched this image of life were instantly destroyed and shall remain forever exposed to the play of the eternal waves.
>
> Proclus

Pythagoreans want to explain everything in terms of numbers. Trouble starts when one tries to explain the simplest elements of geometry by numbers. How does one account for points on a line in terms of numbers? This appears easy at first, but the appearance is deceptive. On a line segment a *unit length* is first chosen, and then to each ratio of integers is associated a point, in a way that is now familiar to every schoolchild. The ratio $\frac{3}{4}$, for example, names the point obtained by dividing the unit length into 4 equal parts and then taking 3 of them. At first it appears that every point on the line can be named in this way, by using ratios of integers, or **rational numbers**.

The Pythagoreans, however, discovered to their distress that there was a certain point P that could be accounted for by *no rational number whatever*!



Consider the point P situated on the line as indicated above. The number associated with P would measure our ratio r_1 introduced in Section 2, since it measures the length of the hypotenuse of a right triangle whose legs each have a length of one unit. By the Pythagorean theorem, *the square of the number associated with P must be equal to 2.* The shock was felt when somehow, out of the Pythagorean school, around 430 B.C., came the following remarkable theorem. The proof we give is essentially the same as Euclid's, which depends on the simple fact that no number can be both even and odd—the key to the original Pythagorean argument, according to Aristotle.

Theorem
There is no rational number whose square is 2.

Proof

The proof is by *reductio ad absurdum*. Suppose the theorem stated above is false, i.e., suppose there *is* a rational number whose square is 2. Then, by cancelling out any common factors in the numerator and denominator, we should have a rational number *a/b in lowest terms* whose square is 2. We should then have integers a and b satisfying

$$a \text{ and } b \text{ have no common divisor;} \tag{4}$$

$$\frac{a^2}{b^2} = 2. \tag{5}$$

From equation (5) it follows that

$$a \text{ is even.}$$

Reason: If a were odd, then a^2 would be odd as proved in exercise 1.3(b); but equation (5) tells us that $a^2 = 2b^2$, showing that a^2 is even (being twice another integer).

Putting this information together with that of (4) shows that

$$b \text{ is odd,} \tag{6}$$

since otherwise a and b would both be even and thus have a common divisor of 2.

Since a is even, a must be equal to twice some other integer. Calling this other integer k, we then have $a = 2k$, or $a^2 = 4k^2$, so that equation (5) becomes

$$\frac{4k^2}{b^2} = 2, \tag{7}$$

where k and b are integers. But from this equation it follows that

$$b \text{ is even.} \tag{8}$$

Reason: If b were odd, then b^2 would be odd—again by exercise 1.3(b)—yet equation (7) tells us, when it is solved for b^2, that $b^2 = 2k^2$, showing that b^2 is even.

The absurd conclusion is evident when one compares statements (6) and (8), since obviously no number is both odd and even. This shows that an absurd conclusion is a consequence of supposing the theorem false. Therefore, the theorem must be true. □

Our ratio r_1, introduced in Section 2, can thus not be accounted for by any ratio of integers! In modern terms we would say that the point P on the "number line" cannot be represented by a rational number. Today, must students have no qualms about associating the point P with the number defined by a *never-ending* decimal expansion beginning

$$1.414 \ldots.$$

Historians have observed that one of the characteristics of the Greeks, in art as well as in science, is a tendency to to shy away from the notion of the infinite. In all the liberal arts they sought to obtain perfected works like those achieved in Greek sculpture and architecture, rather than merely to suggest the beginning of an on-going process.

When the Greeks realized that they could base mathematics upon ratios of integers only by accepting never-ending processes, they sought refuge in the "safe" framework of geometry. The point P offers, of course, no difficulty at all to geometry. It is just as simple an object as any other geometric point. Although the difficulty with irrational numbers resurfaces in geometry as a problem with *incommensurable* line segments, the brilliant Eudoxus (ca. 408–355 B.C.) showed—as we shall see in the next section—how infinity might be tamed. It seemed that geometry could handle something that arithmetic could not, and the Greeks came to regard geometry more highly than arithmetic. For over two thousand years mathematics was couched largely in the framework of geometry.

The tendency of the ancients to couch mathematical facts in geometrical, rather than numerical, terms is one of several factors that held back the development of the calculus. What might have happened if the Greeks, or those who followed them, had emphasized measurement by numbers and had studied the way that numerical quantities relate to one another in a given setting? The notion of a *function* studying how inputs lead to outputs might have quickly arisen, giving birth to one of the fundamental viewpoints of experimental science and accelerating the development of the calculus. Could science have developed quickly enough to have landed men on the moon a thousand years ago? The irrationality of $\sqrt{2}$ may have had a profound effect upon the history of mankind.

Exercises

5.1. Is 1.414 a rational number? *Answer*: Yes, since it is equal to 1414/1000, a ratio of integers.

5.2. Is 2 a rational number? *Hint*. $2 = 2/1$.

5.3. Is it true that $\sqrt{2} = 1.414$? *Hint*. The number $\sqrt{2}$ signifies a number whose square is 2. Is the square of 1.414 exactly equal to 2?

5.4. Is $\sqrt{4}$ irrational?

5.5. Using the fact proved in this section (the fact that $\sqrt{2}$ is irrational), give a *reductio ad absurdum* proof that $\frac{3}{4}\sqrt{2}$ is also irrational. *Answer*: If it were rational it would be equal to some ratio m/n of integers, which leads to a contradiction, as follows: $\frac{3}{4}\sqrt{2} = m/n$ implies $\sqrt{2} = 4m/3n$ = a ratio of integers, contradicting the fact that $\sqrt{2}$ is irrational.

5.6. Using the fact that $\sqrt{2}$ is irrational, give a *reductio ad absurdum* proof that $\frac{5}{13}\sqrt{2}$ is irrational. Imitate the language used in the answer given to exercise 5.5, making appropriate modifications.

5.7. Prove that $5\sqrt{2}$ is irrational. Use language similar to your answer to 5.6.

5.8. Prove that $\frac{a}{b}\sqrt{2}$ is irrational if a and b are nonzero integers. Explain why this result is false when either a or b is zero.

5.9. Prove that $2+\sqrt{2}$ is irrational. *Answer*: If $2+\sqrt{2}$ were rational, we would have $2+\sqrt{2} = m/n$, which, when solved for $\sqrt{2}$, yields $\sqrt{2} = (m-2n)/n$ = a rational number, which contradicts that fact that $\sqrt{2}$ is irrational.

5.10. Prove that $2-\sqrt{2}$ is irrational. *Hint*. Modify appropriately the language of the given answer to exercise 5.9.

5.11. Is the sum of two irrationals numbers always irrational? *Hint*. Add the irrational number discovered in exercise 5.9 to the irrational number in exercise 5.10.

5.12. Is the sum of two rational numbers always rational? How about the product? Tell *why* or *why not*. Don't just answer *yes* or *no*.

5.13. Is the product of two irrationals always irrational? *Why* or *why not*?

5.14. Find an irrational number between 0 and 1/1000000000000. *Hint*. Set a equal to 1 in the expression studied in exercise 5.8 and try taking b as the present national debt, in dollars, of the United States. Does this work?

5.15. We have seen in exercise 3.3 how to prove there is no largest integer. Use *reductio ad absurdum* to prove there is no smallest positive irrational number.

5.16. Are there *infinitely many* irrational numbers between 0 and 1/1000000000000? Explain *why* or *why not*.

5.17. Are there infinitely many irrationals between any two rational numbers? Explain *why* or *why not*.

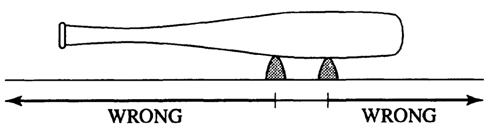
5.18. (*For more ambitious students*) The long statement of Apollonius given in this section is, believe it or not, one of the prettiest theorems of geometry, once it is understood. Read pp. 203–204 of *The World of Mathematics*, edited by James R. Newman, Simon and Schuster, 1956, where Apollonius' proof may be found.

§6. Turning Wrong Answers Inside Out: Eudoxus' Anticipation of Modern Analysis

Searching for wrong answers may seem foolish until one realizes that *reductio ad absurdum* arguments rely upon little else. The key to such indirect proofs is to demonstrate that one of two possible assertions leads to absurdity, thereby eliminating it from consideration because it is a wrong answer.

Similarly, it is sometimes useful to search for wrong answers when given a problem requiring a numerical answer. This is because a *thorough* search for wrong answers is equivalent to a search for the right answer. It is impossible to know all the wrong answers to a numerical problem without knowing the right answer—assuming, of course, that there is a "right answer". Some problems can best be attacked indirectly by finding a procedure to determine when an answer is too large or too small.

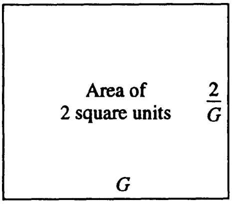
Let us first take an easy example from simple physics. Suppose you are trying to balance a baseball bat on a single finger. This is hard to do directly by simply guessing where to put your finger, but it is easy to balance the bat by resting it upon two fingers separated from each other. With virtually no effort you have thereby found a large number of wrong answers:



A baseball bat balanced upon two fingers
Wrong answers for the center of gravity of the bat

To get closer to the point of balance beneath the bat's center of gravity, move your fingers toward each other while ensuring that the bat continues to rest upon both. Though neither of your fingers will ever quite reach the "right answer" this way, the desired point will be determined as the limit of the positions of your fingers as they approach each other.

Mathematics can use the same idea to solve certain types of problems. To illustrate this, *let us search for a number whose square is 2.* The Pythagoreans, having proved that there is no rational number whose square is 2, would have thought such a number to be inconceivable. If we make a "guess" G that is a rational number then the square of G cannot be equal to 2. Geometrically, this means that if the length of a rectangle's base is measured by a rational number G and if the rectangle has an area of 2 square units, then the rectangle cannot be a square:



If this rectangle has area 2 and base G then its height must be 2/G. If G is rational the rectangle of area 2 cannot be a square.

In the figure above we see that if our guess G is rational then $2/G$ must be a number different from G. Of course, we can see the same thing algebraically: If $2/G$ were equal to G, then G^2 would equal 2, an impossibility if G is rational.

How will this help us manufacture $\sqrt{2}$, a number whose square is 2? *Answer*: It is obvious from studying the rectangle above that the numbers G and $2/G$ are on opposite sides of the right answer. That is, if the square of one of these is less than 2 then the square of the other will exceed 2. Thus the average of G and $2/G$, being halfway in between, will likely be a better guess at $\sqrt{2}$. Since the average of two numbers is half their sum, it follows that if we define B by the equation

$$B = \frac{1}{2}\left(G + \frac{2}{G}\right), \tag{9}$$

then B ought to be a better guess at $\sqrt{2}$ than was our original guess G.

There is evidence to indicate the Babylonians knew a geometric version of this recipe for cooking up better guesses at square roots at least as early as 1000 B.C. Given a guess G the Babylonians could turn it into a better guess B essentially by using the "divide and average" formula above. If we take $G = 2$ then by applying formula (9) we get $B = 3/2 = 1.5$. Now we can change our guess to $3/2$, divide and average, and repeat this process as often as we wish:

$$\text{If } G = 3/2, \text{ then } B = \frac{1}{2}\left(\frac{3}{2} + \frac{2}{3/2}\right) = 17/12 \qquad (= 1.4166\ldots).$$

$$\text{If } G = 17/12, \text{ then } B = \frac{1}{2}\left(\frac{17}{12} + \frac{2}{17/12}\right) = 577/408 \qquad (= 1.4142156\ldots).$$

We could continue this process *ad infinitum* to manufacture rational numbers closer and closer to $\sqrt{2}$.

Expressing the fractions in decimal expansions, we would say that $\sqrt{2}$ is approximated ever more closely as we march along in the sequence beginning

$$2, \quad 1.5, \quad 1.4166, \quad 1.4142156, \quad \text{etc.}, \tag{10}$$

where each succeeding term is manufactured by applying formula (9) to the term before. Since every member of this sequence will be a rational number (*why?*), we can never actually get to $\sqrt{2}$ no matter how far out we go in the sequence. The magnitude represented by $\sqrt{2}$ is not equal to any of the numbers in the sequence (10), no matter how many terms of the sequence we calculate, but is, so to speak, what these numbers are "trying to be". The "real number" we wish to manufacture to be attached to the symbol $\sqrt{2}$ is then, in a sense that is perhaps still rather vague, the limit of a sequence of "wrong answers" which approximate $\sqrt{2}$ ever so closely.

The decimal system is extraordinarily useful, partly because it (like calculus itself, incidentally) has the capacity to be used effectively even by students who do not completely understand it. To understand the decimal system it is necessary to understand that the decimal representation of an irrational magnitude essentially expresses the magnitude as the limit of rational numbers that approximate it ever more closely. Better and better rational approximations are manufactured as the expansion of the desired magnitude is calculated to more and more decimal places.

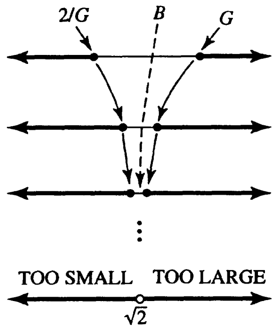
Eudoxus of Cnidus (ca. 408–355 B.C.) was jokingly called Endoxus ("the renowned") by his colleagues at Plato's Academy, for he was known in his time not only as a mathematician, but as an astronomer, physician, orator, and philosopher. Today he is still renowned in mathematics for a simple yet subtle observation. With no decimal system and no explicit notion of a limit in the mathematics of his time, he could nevertheless view the work above essentially as a procedure for finding all ratios larger—and all ratios smaller—than the size of the magnitude in question. *The magnitude whose square has an area of 2 is always between G and $2/G$.* As we let G successively be $3/2, 17/12, 577/408, \ldots$, we see that the successive G's form a sequence converging to $\sqrt{2}$ from above, while the successive values of $2/G$ (given by $4/3, 24/17, 816/577, \ldots$) converge from below. Each successive value of G here is the preceding value of B as calculated by formula (9).

Today we can express these observations in three ways. In the figure below on the left is a statement of the facts essentially as Eudoxus might have expressed them, i.e., in terms of ratios of whole numbers; in the center is a geometric picture; while on the right we catalogue the same data using modern decimal expansions unavailable to Eudoxus.

$\sqrt{2}$ is between $\frac{4}{3}$ and $\frac{3}{2}$

$\sqrt{2}$ is between $\frac{24}{17}$ and $\frac{17}{12}$

$\sqrt{2}$ is between $\frac{816}{577}$ and $\frac{577}{408}$



2/G	B	G
1.33333	1.41667	1.50000
1.41176	1.41422	1.41667
1.41421		1.41422
⋮	⋮	⋮

Searching for all the wrong ratios:
Rational approximations to $\sqrt{2}$ by the Babylonian Method

What sense can we make of this? Whereas the exact size of the magnitude in question here can be specified by no *single* rational number, its size can be determined exactly with reference to the collection of *all* rational numbers in an indirect way: by specifying all the rationals that are too small to measure it, together with all those that are too large. This revolutionary idea of the fourth century B.C.—specifying indirectly a numerical measurement of a geometric ratio in the case when no single ratio of integers will measure it—is that century's response to the previous century's discovery that not all ratios of geometric magnitudes could be measured by ratios of integers. The resulting new theory of proportions, traditionally attributed to Eudoxus, is described by Euclid in Book V of the *Elements*, but couched in geometric language difficult for modern readers to follow. What is expressed today by writing the seemingly simple inequality

$$\sqrt{2} < 3/2 \tag{11}$$

could not be said nearly so briefly before the use of the symbols "$\sqrt{\ }$", "<", and "/" (not to mention the Hindu-Arabic numerals) became commonplace. Eudoxus would have had to express the meaning of the compact inequality (11) in a lengthy rhetorical phrase referring to geometric magnitudes, perhaps something like this:

> The two diagonals of a square, when laid end to end, will be exceeded in length by three sides of the square laid end to end. (12)

It was not until the nineteenth century that the German mathematician Richard Dedekind (1831–1916), by translating the lengthy Greek phrases into modern compact inequalities, interpreted Eudoxus' insight in the following way: Every possible proportion—every *real number*, as we should say today—is entirely specified by the way it cuts the rational numbers into two segments.

This indirect method of getting what we want by throwing away (or "exhausting") the answers that are too large and those that are too small is an ancient secret that deserves to be more widely known. It turns out not to be so generally applicable as the modern method of limits, but when it can be used it may be expected to yield the same answer as would be obtained by taking a limit. By using this powerful **method of exhaustion**, which in essence consists of skillfully throwing out the wrong answers, the Greeks were able to uncover significant results of calculus some two thousand years before the subject was given its name. In Chapter 7 we shall re-examine this method in modern terms.

This has been the barest sketch of a profound theory that is the basis for the modern (nineteenth-century) development of the **real number system**, the most important structure in mathematics. The foundation for this structure was laid by Eudoxus of Cnidus, who is still renowned today for his strength in mathematics. He turned wrong answers inside out and made them tell the truth.

Exercises

6.1. The Greeks didn't speak in terms of functions, but today we would think of equation (9) as defining a function since it defines a rule that, given G, manufactures a unique number B. Calling this function f, we then have $B = f(G) = (G + 2/G)/2$, with the domain of f naturally taken to consist of all positive numbers. Let us catalogue the results of this section in a pair of columns representing the action of this function f. It is useful to make two tables, one using fractions and the other using decimals:

$G \xrightarrow{f} B$

G	B
2	3/2
3/2	17/12
17/12	577/408
577/408	———

$G \xrightarrow{f} B$

G	B
2	1.5
1.5	1.4166666...
1.4166666...	1.41421568...
1.41421568...	———

Carry the procedure one step further by filling in the two blanks correctly.

6.2. The function given by $B = f(G)$ of exercise 6.1 shows us how to modify a rectangle of area 2 to make it more and more like a square of area 2. The closer the "output" B is to the "input" G, the closer the rectangle of sides G and $2/G$ is to a square of area 2 (and therefore the closer both B and G are to measuring the size of $\sqrt{2}$). Thus we are interested in the difference between B and G.

(a) Find $B - G$, using equation (9). *Answer*: $B - G = (2 - G^2)/2G$.

(b) Use your answer to part (a) to show that equation (9) may be re-expressed as $B = G - (G^2 - 2)/2G$.

(c) A **fixed point** of a function is a point that is sent into itself by the function, i.e., a function has a fixed point where the output is the same as the input. Show that if G is a fixed point of the function f of exercise 6.1, then $G = \pm\sqrt{2}$. *Hint.* In your answer to part (a), set B equal to G and solve for G.

(d) A modern pocket calculator has a button with a odd-looking symbol on it. If the number 2 is displayed by the calculator and this odd-looking button is pressed, the calculator is programmed to do the following odd thing. Without displaying its action it calculates $f(2)$, then $f(f(2))$, then $f(f(f(2)))$, and so on, where f is the function of exercise 6.1. It keeps on with such calculations until it "thinks" it gets the same number on two successive calculations (because their difference is such a small number that the calculator considers it to be zero). The calculator then displays the number that is repeated. What is the symbol on the button?

6.3. Approximate $\sqrt{10}$ by the Babylonian method of "dividing and averaging". Here our rectangle has sides G and $10/G$, so B will be the average of G and $10/G$.

(a) Try $G = 3$ as your initial guess and give the successive approximations to $\sqrt{10}$ in both fractional and decimal form by filling in the six blanks below:

G	$\frac{1}{2}\left(G + \frac{10}{G}\right)$
3	19/6
19/6	———
———	———

G	$\frac{1}{2}\left(G + \frac{10}{G}\right)$
3	3.1666...
3.1666...	———
———	———

(b) The drawing in this section illustrates successive approximations to $\sqrt{2}$. Make a similar drawing to illustrate your approximations to $\sqrt{10}$ manufactured in part (a).

6.4. Of course we know that $\sqrt{9} = 3$, but it is interesting to see how the Babylonian method will force this fact upon us if we have forgotten it. Make three tables using decimal expansions like the table in exercise 6.3 but with "9" replacing "10", taking your initial guess to be (a) $G = 2$; (b) $G = 4$; (c) $G = 3$.

6.5. Approximate $\sqrt{3}$ by the Babylonian method. Choose your own initial guess G (the closer G is to $\sqrt{3}$ the better the approximation B will be), and calculate tables like those in exercises 6.3 and 6.4. *Suggestion.* Take G initially to be 1.7, or 17/10, and find at least three successive values of B. Use decimal expressions instead of fractions when the integers in your fractions get too large to handle. Do you see the advantage of the decimal system here?

6.6. Later on, when we re-visit Example 1 on page 2 of this text, we shall find ourselves interested in the square root of 48/7.

(a) Apply the Babylonian method to approximate $\sqrt{48/7}$ by filling in the ten blanks below, where we take $G = 2$ as our initial guess.

G	$\frac{1}{2}\left(G+\frac{48}{7G}\right)$
2	19/6
19/7	——
——	——
(revert to decimals)	(revert to decimals)

G	$\frac{1}{2}\left(G+\frac{48}{7G}\right)$
2	2.714...
2.714...	——
——	——
——	——

(b) Try $G = 3$ instead of $G = 2$ as your initial guess at $\sqrt{48/7}$ and see how the decimal approximations in the right-hand table above are changed.

6.7. (*A research question*) Suppose you want to approximate the *cube* root of 2. Extending the Babylonian ideas to 3 dimensions you might consider a *rectangular solid* instead of a rectangle and note that if G measures one dimension, then taking the other dimensions to be G and $2/G^2$ will ensure that the volume of the solid is 2 cubic units. Take B to be the average of the *three* quantities G, G, and $2/G^2$.

(a) What is the formula for B in terms of G?

(b) Take $G = 1$ as your initial guess and calculate the decimal expansions of several successive values of B until you get an answer that repeats to several decimal places. Have you found $\sqrt[3]{2}$? (Calculate the third power of your answer and see how close to 2 it is.)

(c) Read in the appendix on Archimedes about **the duplication of the cube** and explain the relevance of your work here to the solution of this ancient Greek problem.

6.8. (*Higher-dimensional research*) The *twelfth* root of 2 plays a crucial role in tuning a piano because the ratio of the frequency of each tone to the one below should be $\sqrt[12]{2}$ in order to ensure that after any 12 successive steps (through "white notes" and "black notes") we reach a pitch exactly twice as high as our initial pitch (i.e., we cover an *octave*).

(a) Reason by analogy with the method of exercise 6.7 to extend the Babylonian ideas to 12 dimensions. (This is really a modern problem. The Greeks never got above three dimensions.) *Hint.* Take B to be the average of 12 quantities, of which the first eleven are all equal to G. What should the twelfth quantity then be in order that the product of all 12 quantities be 2?

(b) Take $G = 1$ as your initial guess and calculate the decimal expansions of several successive values of B until you get an answer that repeats to several decimal places. Have you found $\sqrt[12]{2}$? (Calculate the twelfth power of your answer and see how close to 2 it is.)

6.9. (*Make a guess*) Eudoxus investigated the ratio of the volume of a cylinder to the volume of a cone of the same base and height as the cylinder. He proved this ratio is equivalent to one of the six ratios introduced in Section 2. Can you guess which one? (*We shall later use calculus to find this ratio, in exercise 9.2 of Chapter 7.*)

§7. Ratios of "Unlike" Magnitudes?

There is an interesting aspect of the Greek theory of ratios that would probably go unnoticed if not explicitly mentioned. It is the fact that only "like" magnitudes could be said to have a ratio. In Section 2 we have pictured examples of ratios of "one-dimensional" magnitudes, of "two dimensional" magnitudes, and of "three-dimensional" magnitudes. Of course, one could also speak of ratios of "zero-dimensional" magnitudes, such as the ratio of seven points to five points, but this hardly seems worth mentioning, as it is already—in essence—the ratio of one number to another.

In Greek mathematics one could not speak of a ratio between two geometric magnitudes of different dimensions, such as between a cube and a square, a square and a line segment, or a line segment and a point. To have a ratio between two things, according to an axiom introduced by Eudoxus, some multiple of each thing must exceed the other. Why can we not speak of the ratio of a line segment to a point? *Answer*: No matter how you "multiply" points (i.e., no matter what number of points you put together), the length of the resulting figure will not exceed that of a line segment. Today we might put this by saying that a point is "infinitesimal" when compared with a line segment, or—to put it in complementary terms—that a line segment is "infinitely greater" than a point.

In one of Plato's dialogues the speaker asserts that the philosopher, the statesman, and the sophist are separated by "an interval that no geometrical ratio can express", implying that the philosopher is infinitely greater than the statesman and the statesman is infinitely greater than the sophist. To say that "no geometrical ratio can express this" is simply to say that the axiom of Eudoxus effectively bans the notion of infinity (and the complementary notion of an infinitesimal) from being considered as a proper ratio.

Today we still ban infinity from being considered as a number. Where would it be on the number line? It would have to be "the point at the end", but a line has no end. In this respect we still follow the Greeks, but there are ways in which our use of number is much more general than the Greek use of ratios. We have no qualms today about speaking of the ratio of the distance travelled during a journey to the time taken. In fact, we call this number the *average speed* of the trip, and measure it in such units as miles per hour. The Greeks would have hesitated to speak of the ratio of such unlike things as distance and time, or to conceive of something like "miles per hour" as a unit of measurement. It would be like mixing apples and oranges.

The Greeks could speak of motion at a *constant* speed by saying that the ratio of the distance travelled during one time segment to the distance travelled during any other time segment is the same as the ratio of the time segments involved. They could also deal with uniformly accelerated motion without straining their language of ratios, but the physics of more general motion was beyond them. Aristotle wrote a book on physics, but as the son of a physician, he tended to think of physics first of all in terms of the growth processes studied by physicians. *Physics* comes from a Greek word whose root meaning is the study of nature.

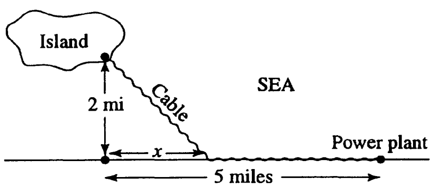
Aristotle remarked that, in general, "motion of motion" (acceleration) is beyond our ability to measure. Of course, the modern physics of motion rests upon Newton's law stating that the force on a body is proportional to its acceleration. Aristotle—who was, among many other things, a great biologist—obviously had little chance of becoming a great physicist. The restricted notion of number, which Aristotle endorsed, prevented progress in this area.

The modern physics of motion, in fact, is hardly possible to comprehend without using the techniques of calculus to make sense out of "instantaneous velocity"—a contradiction in terms to Aristotle since there is no motion in an instant—and to speak also of "instantaneous acceleration". In Chapter 6 we shall see how easily the seventeenth-century calculus developed by Newton and Leibniz enables us to handle such "contradictions in terms" by defining them naturally in terms of limits.

Problem Set for Chapter 2

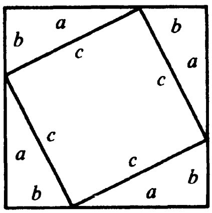
1. In a book of the late sixteenth century written by the Flemish engineer Simon Stevin, promoting the usefulness of the decimal representation of fractions, Stevin wrote, in large capitals at the top of a page, UNITY IS A NUMBER. Why did Stevin feel it necessary to emphasize something that is so obvious to us today? (*Stevin*, incidentally, is accented on the second syllable.)

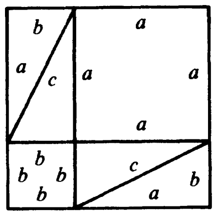
2. Undersea cable cost \$11,000 per mile, whereas underground cable costs \$7,000 per mile. An island and a power plant are located as indicated, and cable is to be run between them.



Let C denote the total cost (in thousands of dollars) of the undersea and underground cable, and let x be as indicated. Find an algebraic rule expressing C in terms of x, and specify the domain of this rule. (In Chapter 5 we shall find the value of x yielding the cheapest way of running the cable.)

3. (*Eureka!*) The following principle is self-evident: *If the same amount is taken away from two figures having equal area, then the two modified figures have equal area.* It is thought that Pythagoras might have employed this principle, as follows:





The two large squares have equal area. Take away the four right triangles from each of the large squares and use the principle above. Have you just proved the Pythagorean theorem? Explain why.

4. Find the center and the radius of the circle corresponding to each of the following equations.
   (a) $x^2 + (y-2)^2 = 7$.
   (b) $3(x+3)^2 + 3(y-\sqrt{2})^2 = 12$.
   (c) $x^2 + y^2 = 10$.
   (d) $5x^2 + 5y^2 = 10$.

5. Write an algebraic equation corresponding to each of the following circles: The circle with
   (a) radius 5, center at $(-3, 4)$.
   (b) radius $\sqrt{5}$, center at $(3, 0)$.
   (c) radius r, center at (a, b).

6. The method of *reductio ad absurdum*—also called "proof by contradiction" or "indirect proof"—is a technique that every student of the liberal arts should master.
   (a) In which of the seven liberal arts described in Section 3 would a student be introduced to the method of *reductio ad absurdum?*
   (b) Describe this method clearly. *Suggestion.* Finish this sentence: "*Reductio ad absurdum* is a technique of [your answer to part (a)] that proves an assertion to be true by demonstrating that an absurdity follows from the supposition that ..."
   (c) Taking it as a known fact that $\sqrt{2}$ is irrational, give a *reductio ad absurdum* proof that $\sqrt{2}/20{,}000$ is irrational.

7. For each of these, tell *why* or *why not.* (Don't just say *true* or *false.*)
   (a) The product of a rational number with an irrational number is irrational.
   (b) The product of a nonzero rational number with an irrational number is irrational.
   (c) The sum of an irrational number and a rational number is irrational.
   (d) The sum of two irrational numbers is irrational.

8. Give an example of a rational number lying between

(a) 0 and 1/10,000.
(b) 100 and 100.0001. *Hint.* Add 100 to your answer to part (a).

9. Give an example of an *irrational* number between 100 and 100.0001. *Hint.* Is $\sqrt{2}/20{,}000$ between 0 and 1/10,000? If so, use problem 6, 7(c) and the hint to 8(b).

10. Give a *reductio ad absurdum* argument to prove that
(a) $\sqrt{8}$ is irrational.
(b) $\sqrt{18}$ is irrational.
Hint. (a) First note that $\sqrt{8} = \sqrt{4 \cdot 2} = \sqrt{4}\sqrt{2} = 2\sqrt{2}$. Then proceed, using the known fact that $\sqrt{2}$ is irrational.

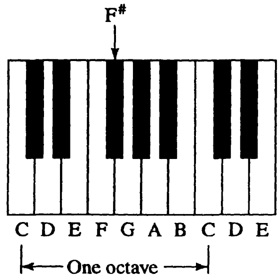
11. Prove by *reductio ad absurdum* that the square root of an irrational number must be irrational.

12. Consider the sequence of numbers $\sqrt{2}, \sqrt[4]{2}, \sqrt[8]{2}, \sqrt[16]{2}, \ldots$, where each number is the square root of the number before and the pattern continues without end, generating infinitely many real numbers.
(a) Are all the numbers in this sequence *irrational*? *Hint.* Use the result of the preceding problem.
(b) What is the *limit* of this sequence, i.e., what is the number being approached, as we go further and further out in the sequence? (If this is not obvious, it will be when you use a calculator to get the decimal expansions of the numbers in the sequence. To get them quickly, enter the number 2 in your calculator and repeatedly press the square root button.)
(c) Explain why there must be infinitely many irrational numbers lying between 1 and 1.1
(d) Explain why there must be infinitely many irrational numbers between 1 and 1.001.
(e) Explain why there must be infinitely many irrational numbers between 1 and $1 + \varepsilon$, no matter how small ε is, so long as ε is positive. (Think of ε—the Greek letter *epsilon*—as signifying an "error tolerance".)

13. (*Mathematics and music*) The Greeks regarded the ability to play a musical instrument (and to dance) as an indispensable social skill. Indeed it has been written that, as a people, they thought of themselves as musicians more than anything else. When they spoke of music as a liberal art, however, they meant the science of music—not the performance of it.
(a) (*Can you hear the square root of two?*) Explain how the musical interval from C to F# "is" the square root of two. *Hint.* You get from C to F# by traversing six keys on a piano. Each time you move from one to the next, as explained in exercise 6.8, you multiply the frequency of the pitch produced by a factor of $\sqrt[12]{2}$. What, then, is the ratio of the frequency of F# to the frequency of C? (The first two notes of the song "Maria" from Leonard Bernstein's *West Side Story* utilize this interval, which suggests a relationship of tenuous harmony.)
(a) Prove that $\sqrt[12]{2}$ is irrational. *Hint.* Use *reductio ad absurdum*. Suppose it were rational; then raise it to the sixth power and use part (a) to get an immediate contradiction.
(c) Explain how you can "hear" the *cube* root of two on the piano by finding

an interval that represents it. Note that the cube root of two happens to be very close to a ratio of 5:4. Does it strike you as being more "harmonious" than the interval from C to F♯?

(d) The Pythagoreans, who thought of musical tones as "numbers in motion" had a particular affinity for musical intervals represented by ratios of very small positive integers. Can you "hear" the ratio 3:2 as a musical interval on the piano? This interval should be the most harmonious of all, except for the octave, which is a ratio of 2:1.



14. Our perception of sound is connected with the mathematical theory of *logarithms*, which we will meet later. Roughly speaking, however, logarithms translate the operation of multiplication into the operation of addition. Give examples showing how *adding* musical intervals translates into *multiplying* the ratios associated with these intervals. *Hint.* The interval from C to F♯ is associated with the ratio $\sqrt{2}$, as is the interval from F♯ to high C. Adding these intervals gives us the interval from C to high C. What is the ratio associated with the interval from C to high C? Is it the product of the ratios associated with the two intervals that we added? Give other examples of this phenomenon.

15. In part (c) of problem 13 we noted that 5/4 is a very close approximation to $\sqrt[3]{2}$, and in exercise 6.7 we saw that the formula

$$B = \frac{1}{3}\left(G + G + \frac{2}{G^2}\right)$$

gives successively better approximations to $\sqrt[3]{2}$.

(a) Re-do exercise 6.7(b), taking $G = 5/4$. Then calculate, using fractions (not decimals) two succeeding values of B to produce the rational number 375047/297675.

(b) To how many places does the decimal representation of the rational number 375047/297675 agree with the decimal representation of $\sqrt[3]{2}$?

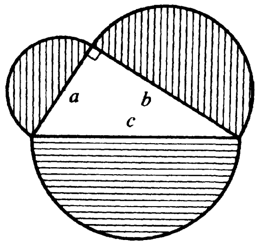
(c) (*Make a guess*) Do you think $\sqrt[3]{2}$ is *rational* or *irrational*?

16. (*Moving to higher dimensions unrecognized by the Greeks*) We have seen that in order to approximate the cube root of 2, we imagine, as in exercise 6.7, a cube of sides G, G, and $2/G^2$ and take the average of these three numbers to get out next approximation B, arriving at the formula displayed in the preceding problem.

(a) What formula should we use to approximate the cube root of 10?

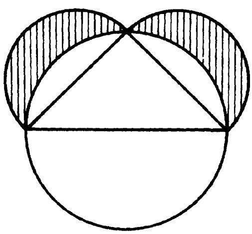
(b) What formula should we use to approximate the cube root of 100?
(c) What formula should we use to approximate the fourth root of 500?
(d) What formula should we use to approximate the seventh root of 2?

17. The Pythagorean theorem deals with squares constructed on the three sides of a right triangle. What about *semicircles* instead? Prove that the area of the semicircle with hypotenuse as diameter is equal to the sum of the areas of the semicircles on the other two sides. *Hint.* This is easy. Use the formula $\frac{1}{2}\pi r^2$ to find the area of each of the semicircles whose radii have lengths $c/2$, $a/2$, and $b/2$. Then use the Pythagorean theorem.

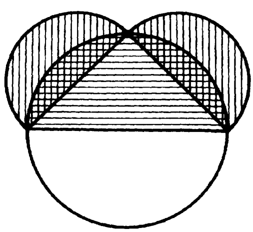


18. (*This is a famous result of Hippocrates of Chios, a member of the Pythagorean school who lived in the fifth century* B.C.) In the figure below, the hypotenuse of the right triangle is also the diameter of the circle in which the triangle is inscribed. Prove that the combined area of the two "lunes" (moon-shaped areas) with vertical markings is equal to the area of the right triangle.

Lunes of Hippocrates



Hint. This is astonishingly simple (once you "see" it). From the preceding problem we know that the area marked vertically in the lower figure is equal to the area with horizontal markings. Take away the cross-hatched area from both figures and use the principle given at the beginning of problem 3. *Eureka*!



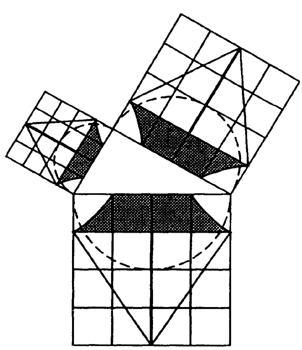
19. (*Reductio ad absurdum*) Show that it is impossible to place five points in, or on the boundary of, a square of size 2 units by 2 units in such a way that the distance between all pairs of these points exceeds $\sqrt{2}$ units. *Hint.* Suppose there were five points in a 2×2 square with the distance between all pairs exceeding $\sqrt{2}$. Divide the square into four unit squares. Since there are five points and only four unit squares, some pair of the points must be in the same unit square—and yet the distance between them exceeds $\sqrt{2}$. Use the Pythagorean theorem to get a contradiction.

20. (*Reductio ad absurdum*) Show that there are two trees in Brazil with the same number of leaves. *Hint.* Use the fact that Brazil has a billion trees, but no tree has a billion leaves.

21. Show that somewhere in the world there are two mice with the same number of hairs. *Hint.* Think of an analogy between mice, hairs, trees, and leaves; then proceed as in problem 20.

22. (*Stereometry*) Plato wanted to include stereometry (solid geometry) in the quadrivium. Show that it is impossible to place nine points in, or on the boundary of, a cube of size 2 units by 2 units by 2 units in such a way that the distance between all pairs of these points exceeds $\sqrt{3}$ units. *Hint.* Generalize to three dimensions the argument made in problem 19.

23. (*Red sails in the sunset*) Take a square photographic print of some appealing object—a sailboat at sunset, say—and have three different-sized enlargements made whose bases fit exactly on the three sides of some right triangle. Think about the sum of the areas taken up on your prints by the two smaller copies of the sailboat in comparison with the sailboat-shaped area on the largest print. The sum of these two smaller sailboat-shaped areas must be either *less than, equal to*, or *greater than* the area of the largest.
    (a) Which of these three possibilities would you guess must actually occur?
    (b) Give an argument to prove your guess is correct. *Hint.* Consider first a sailboat as a cubist might view it, made up of little squares or triangles.
    (c) Do you now see the "real reason" why problem 17 came out as it did? *Hint.* Just change your focus from the sailboat to the sunset.



24. (*A short essay question*) The Pythagoreans essentially died out in the fourth century B.C., though neo-Pythagoreans—"copycats"—continually pop up, even today. Was the nineteenth-century development of the periodic table inspired by neo-Pythagoreanism? If you have studied the periodic table in a chemistry class, write a short essay either attacking or defending the assertion that the existence of this table supports the Pythagorean belief that number is the basis of all things.

    If you know nothing about the periodic table, then write instead a five-paragraph essay describing the most significant contribution of the Pythagoreans to the development of each of these five areas: mathematics, science, philosophy, education, religion.

25. (*A longer essay question*) Copy down the following for your opening paragraph, then continue to complete an essay of four to seven pages on the topic "Mathematics and the Liberal Arts".

    > The rise of pure mathematics has never enjoyed anything like the attention historians have lavished upon the rise of science. While the Scientific Revolution of the seventeenth century is well known, the much earlier "mathematical revolution" is not. Yet the Greeks' interest in mathematics for its own sake promoted not only the eventual rise of science, but also the imminent rise of a system of education that initially shaped Western thought.

    Near the end of your essay, compare the role mathematics played in helping to shape Western thought with the role it played in shaping the thought of other cultures with which you are familiar.

26. (*An essay on integrity*) The kinship of the purposes of science and art—and the consequent integrity ("wholeness") of the liberal arts—is an ancient ideal. The literary critic Edmund Wilson explains how a Greek geometer and a Greek dramatist are working to make similar patterns:

    > In my view, all our intellectual activity, in whatever field it takes place, is an attempt to give a meaning to our experience. . .; for by understanding things we make it easier to survive and get around among them. The mathematician Euclid, working in a convention of abstractions, shows us relations between the distances of our unwieldy and cluttered-up environment upon which we are able to count. A drama of Sophocles also indicates relations between the various human impulses which appear so confused and dangerous, and it brings out a certain justice of Fate—that is to say, of the way in which the interaction of these impulses is seen in the long run to work out—upon which we can also depend. The kinship, from this point of view, of the purposes of science and art appears very clearly in the case of the Greeks, because not only do both Euclid and Sophocles satisfy us by making patterns, but they make much the same kind of patterns. Euclid's *Elements* takes simple theorems and by a series of logical operations builds them up to a climax in the square on the hypotenuse. A typical drama of Sophocles develops in the same way.
    >
    > from "The Historical Interpretation of Literature",
    > in *The Triple Thinkers*, Oxford, 1984, p. 269

    Is the integrity of the liberal arts today so highly prized? Write an essay comparing the integrity of the liberal arts today with the Greek ideal.

3 CHAPTER

To Measure Is to Know

As we have seen in the previous chapter, the word *ratio* is connected with the idea of rational thought or calculated study. The Greek word for the same notion is *logos*, which has similarly acquired overlays of meanings stemming from the idea of measurement. We find it as a suffix in many academic words derived from Greek: *anthropology* ("study of man"), *biology* ("study of life"), and so on. The word *logic*, of course, also comes from *logos*.

This notion presumably arises from our desire to see connections, to find some sense of unity or regularity in apparent diversity. Edmund Wilson's words quoted in the last problem of the preceding chapter make this point. We find ratios everywhere in the liberal arts, helping us to recognize (or to impose) order in potentially chaotic settings. The notion of a proportion—an equality of ratios—manifests itself in mathematics as *measurement*, in rhetoric as *analogy*, and in music as *harmony*. Proportion is, of course, closely tied to *beauty* in all the classical arts.

We have touched upon these things in the previous chapter. In this chapter let us go back to the fundamental mathematical meaning of ratio and see how the Greeks measured the earth, the circle, and the cone. Surprisingly, as we shall see, the earth poses less of a problem than either of the other two.

Let us first, however, take a brief look at the way mathematics influenced two great philosophers.

§1. Plato, Aristotle, and Mathematics

The safest general description of Western philosophy, according to a famous remark by Alfred North Whitehead, is that it consists of a series of footnotes to the writings of Plato (430–349 B.C.). Yet, as we know from the warning on his gate ("Let no one ignorant of geometry enter here"), Plato's philosophy is profoundly affected by his conception of mathematics. Since Plato learned much about mathematics from Archytas, it is no wonder that the Pythagorean spirit can be found in the writings of Plato.

There is a story that Archytas—the Pythagorean who gave mathematics a charter membership and a prominent role in the liberal arts—interceded with Dionysus I in 387 B.C. to prevent the outrageous poetaster-tyrant from selling Plato into slavery. Dionysius had supposedly lost all restraint when Plato told him to become a philosopher or to cease being a king. The role of Archytas as hero of this story is disputed, but there is no doubt that Plato strongly supported the quadrivium in education as well as the Pythagorean belief that the same basic education should be provided for both males and females. The purpose of education is to learn to tell the truth, and mathematics—if taught for its own sake—promotes this end by helping students to sharpen their intuition, to learn to reason better, to recognize valid reasoning, and to write and say more precisely what they intend. Such an education is essential to freedom, for without knowing how to tell the truth one is easily boxed in by sophistries. (A sophistry relevant to calculus may be found in the last problem at the end of this chapter.) Nothing is more abhorrent to Plato than the Sophists who use their art of persuasion to empower themselves through deliberately deceptive arguments with no concern for truth or other ultimate ends such as goodness and beauty.

Plato, on the other hand, was driven to know the Good, the True, and the Beautiful, and became more enamoured of mathematics when he perceived a resemblance (which at first must have been quite hazy, indeed) between such things and mathematical forms such as the circle and the triangle. The fact that mathematicians could use reason to test their intuition about mathematical forms apparently inspired Plato to believe there must be an analogous way that philosophers might learn to know the higher forms with greater certainty. "Geometry will draw the soul toward Truth," said Plato, "and create the spirit of philosophy." All knowledge might aspire to the state attained by mathematics. Here, beyond the realm of immediate practicality, lies the true spirit of pure thought.

For Plato then, the ultimate aim of education is the training of the mind to pass from the apparent and the ephemeral to the true and permanent. The *eternal* becomes of great interest to Plato, since it alone has the power to withstand the erosion of time, and here again he found virtue in mathematics. The theorems of the Pythagoreans will surely outlive not only the Pythagoreans, but the Greek language as well. Significant theorems will retain their value not just for the next few thousand years, but literally forever.

Plato went even further. As an illustration, consider the question whether the Pythagorean theorem was true *before Pythagoras came upon it*. Plato would reply strongly in the affirmative and would assert that the theorem had always been true. It had been "built into the universe", and Pythagoras was just one who saw it clearly enough to present a proof. In fact, as we have remarked, there is evidence that the Babylonians, Indians, and Chinese knew this result long before the Greeks, though they did not develop the Greeks' enthusiasm for demonstrating the truth of a proposition by finding an argument connecting it with simple premises.

In Plato's view the connections between the ideas involved have always been there. They are waiting to be discovered, just as (unbeknownst to Plato) the moons of Jupiter were awaiting Galileo and his telescope. Plato thought that all enduring knowledge must be like this. Knowledge consists of ideas, or eternal forms, and their great web of connections that form a realm of beauty beyond the comprehension of our senses alone:

> The laws whereby the stars are made
> are fairer than the stars.

The inspiration for scientific inquiry has perhaps never been better expressed than in this strikingly brief statement about truth and beauty, but Plato was interested in more than what we today consider as science. To put any significant piece of knowledge into down-to-earth terms, so that all can understand, is a noble undertaking. Socrates, Plato's teacher, undertook to explain the idea of Justice—an idea that is still imperfectly understood. Plato tackled the virtually impossible task of examining the Good, the True, and the Beautiful, and to see the interrelationship between these ideas. Not long after Plato, Euclid began to write down the interrelationship between all the ideas of geometry that were known up to his time. Such efforts as these have inspired to this day many more seemingly impossible undertakings.

The aspect of Plato's philosophy just described is sometimes pictured as follows. The ideas, or eternal forms, already exist, floating in the "Platonic heaven", just beyond our grasp. Perhaps, as the Pythagoreans believed, we ourselves existed in a former life when we might have known these ideas before, but we are born with only a hazy memory of where we come from. To know the forms fully, to "remember" them, is our most natural calling, which is why we must study philosophy. Only a lover of wisdom can climb high enough to swing around heaven and slide back down to earth with a new perspective. A Platonist today might hold the view that liberal arts consists of ideas brought down to earth by such swingers.

It should be emphasized that Plato was not a mathematician, but he saw to it that his Academy in Athens fostered the work of Eudoxus, whom we have already met, and of Theaetetus, who became famous in mathematics for being first to prove that the square root of every non-square integer is irrational. Plato must have been strongly affected by the fact that these mathematicians had created (or, is *discovered* a better word?) something of significance that would last forever.

The excitement that Plato lends to the Hellenic study of mathematics in Athens contrasts greatly with the tone taken by his teacher Socrates, who had little interest in mathematics, and by his great student Aristotle (384–322 B.C.), whose more scholarly and less speculative nature would influence for centuries the Hellenistic thought of Alexandria. Aristotle saw nothing in mathematics to inspire such a flight of imagination as is taken by Plato. The capacity to systematize knowledge, to bring order through reason, is of the highest importance. The value of mathematics, to Aristotle, lies in its exemplification of this capacity to a degree unmatched in any other discipline. He is more interested in logic than in mathematics.

Aristotle's views on logic have had great influence. They are reflected in the style of Euclid, whose *Elements* appeared in Alexandria not long after Aristotle's death. Euclid seems to show that the towering edifice of geometry is simply the consequence of logic unerringly applied to "self-evident" propositions, or *axioms*. The value of Euclid's work lies not in the announcement of previously unknown theorems (many if not most of the theorems in the *Elements* were known before Euclid was born), but rather in the masterful logical organization of a great body of knowledge by the *axiomatic method*. Aristotle endorsed this method, which seems to have been introduced two centuries earlier by Thales, who taught Pythagoras. The axiomatic method consists in stating clearly one's initial assumptions (axioms) and deducing all else by means of logic. The method results in a writing style that is demanding, austere, and—to some—supremely beautiful:

> Euclid alone has looked on Beauty bare.
> Let all who prate of Beauty hold their peace,
> And lay themselves prone upon the earth and cease
> To ponder on themselves, the while they stare
> At nothing, intricately drawn nowhere
> In shapes of shifting lineage; let geese
> Gabble and hiss, but heroes seek release
> From dusty bondage into luminous air.
> O blinding hour, O holy, terrible day,
> When first the shaft into his vision shone
> of light anatomized! Euclid alone
> Has looked on Beauty bare. Fortunate they

Who, though once only and then but far away,
Have heard her massive sandal set on stone.

—Edna St. Vincent Millay*

Western civilization has absorbed over a thousand editions of Euclid's *Elements*. It is no surprise that traces of the axiomatic method can be detected in many nonmathematical writings:

We hold these truths to be self-evident.

a new nation ... dedicated to the proposition that....

Thomas Jefferson and Abraham Lincoln were among Euclid's admirers, as was Benjamin Franklin, who suggested "self-evident" to Jefferson just before the final draft of the Declaration of American Independence. Lincoln considered his reading of Euclid an indispensable part of his education. The following passage is from a biographical sketch written for the 1860 presidential campaign.

> He studied and nearly mastered the six books of Euclid since he was a member of Congress.
>
> He began a course of rigid mental discipline with the intent to improve his faculties, especially his powers of logic and language. Hence his fondness for Euclid, which he carried with him on the circuit till he could demonstrate with ease all the propositions in the six books; often studying far into the night, with a candle near his pillow, while his fellow-lawyers, half a dozen in a room, filled the air with interminable snoring.

What is it about Euclid that attracts? Is it not the cold, unexcited certainty with which tower upon tower of seemingly irrefutable arguments are built? No work could be more dispassionate than Euclid's *Elements*. Yet this severe and solemn quality has probably repelled as often as it has attracted.

> Euclid's work ought to have been any educationist's nightmare. The work presumes to begin from a beginning; that is, it presupposes a certain level of readiness, but it makes no other prerequisites. Yet it never offers any "motivations", it has no illuminating "asides", it does not attempt to make anything "intuitive", and it avoids applications to a fault. It is so "humorless" in its mathematical purism that, although it is a book about "Elements", it nevertheless does not unbend long enough in its singlemindedness to make the remark, however incidentally, that if a rectangle has a base of 3 inches and a height of 4 inches then it has an area of 12 square inches. Euclid's work never mentions the name of a person; it never makes a statement about, or even an (intended) allusion to, genetic developments of mathematics.... In short, it is almost impossible to refute an assertion that the *Elements* is the work of an unsufferable pedant and martinet.
>
> —S. Bochner, *The Role of Mathematics in the Rise of Science*, Princeton, 1966, p. 35

*Sonnet XLV, from *Collected Poems*, Harper and Row. Copyright 1923, 1951 by Edna St. Vincent Millay and Norma Millay Ellis. Reprinted by permission.

Whatever excitement Euclid felt for mathematics he restrained in writing the *Elements*. Aristotle would have wanted it that way. Scholarship must stand up under the cold, steady eye.

Exercises

1.1. Do you think mathematics is *discovered* or *created*? (Or would you rather dodge the issue and simply speak of mathematics as being "developed"?)

1.2. Plato thought of mathematics as being *discovered*. What are the arguments *for* and *against* such a view?

1.3. Plato popularized—and thereby immortalized—the teaching method of his own teacher Socrates, who rarely made statements but instead masterfully led students toward the discovery of truth by continually asking them thoughtful questions. How is the Socratic method related to Plato's belief that learning is not simply processing information or mimicking proper behavioral responses, but must be essentially concerned with "remembering" or "re-awakening"?

1.4. In talking about "being" versus "non-being", Plato suggests that one knows what a thing is by knowing what it is not. How is this related to the elimination of wrong answers discussed in Section 6 of Chapter 2, which culminates in the subtle modern numerical characterization of an irrational magnitude like $\sqrt{2}$ in terms of "everything it is not"?

1.5. Plato made much of the fact that mathematical ideas like points and circles do not exist in the same physical sense as tables or chairs. We actually see no mathematical forms, but only physical objects that approximate them. A star is not a point, but we may "get the idea" of points by looking at stars. Similarly, a wedding band is not a circle, but as we conceive of ever thinner and rounder bands we "get the idea" of a circle.
   (a) Is the principle of elimination at work in our conception of points and circles? That is, do we learn what they are by saying what they are not?
   (b) Is the notion of a limit somehow at work here? Do we conceive of a circle as the "limit" in some sense of physical objects that approximate more and more the "form" of a circle?
   (c) Plato viewed "eternal forms" like Truth, Beauty, and Goodness as existing in somewhat the same sense as mathematical forms. Do you see any way that the principle of elimination or the notion of a limit might have influenced Plato to hold such a radical view?

1.6. Aristotle, Plato's student, dismissed his teacher's beloved eternal forms as "sound without sense". Aristotle thought of *circularity*, for example, as simply a property that might be inherent in a wedding band (which we "abstract" from it)—not as a "form" with independent existence in the Platonic heaven. Whose side do you take in this controversy between Plato and Aristotle, and why?

1.7. Aristotle made a sharp distinction between number—on the one hand—and geometric magnitudes on the other, and Euclid adopted Aristotle's outlook in his *Elements*. Did the sharpness of this distinction *help* or *hinder* the development of mathematics? (Our modern notion of a real number essentially conflates these two older notions, and we couch much of modern mathematics in terms of functions from real numbers to real numers. How important to the development of mathematics was this change in point of view?)

1.8. Latin translations of Euclid have made famous the abbreviation *Q.E.D.*, which stands for *quod erat demonstrandum*. Use a dictionary to find out what this means in English. Where does this phrase naturally appear in proofs?

1.9. In addition to studying ratios of two magnitudes, the Greeks developed the idea of various *means* between them. The **arithmetic mean** of two magnitudes a and b is half their sum, or $(a + b)/2$; the **geometric mean** is the square root of their product, or $\sqrt{ab}$; the **harmonic mean** is the reciprocal of the arithmetic mean of their reciprocals, which works out to $2ab/(a + b)$.

(a) Find the arithmetic, geometric, and harmonic mean of a and b, where $a = 1$ and $b = 2$. *Answer*: Arithmetic mean is $3/2 = 1.50$; geometric mean is $\sqrt{2} \approx 1.414$; and harmonic mean is $4/3 \approx 1.333$.

(b) Which mean between 1 and 2 measures the side of a square having the same area as a rectangle with sides 1 and 2?

(c) If a small plane travels into a strong headwind and averages one hundred miles per hour on its flight from point A to point B, then returns to A averaging two hundred miles per hour with a tailwind, which mean between 100 and 200 gives the average speed for the entire round-trip flight? *Hint*. The answer does not depend upon the distance from A to B. Take this distance to be 200 miles and see what happens then.

(d) Which mean between 1 and 2 should be used to measure a string whose frequency of vibration is the arithmetic mean of the frequencies produced by plucking strings of lengths 1 and 2? *Hint*. The fundamental frequency of vibration of a string, under constant tension, is inversely proportional to its length. If the longer of the strings vibrates at the frequency of middle C and the shorter is made half that length to emit high C, then this procedure should produce the G-string that splits the octave into intervals of a fifth and a fourth. (The musical interval from middle C to G is a fifth and from G to high C is a fourth.)

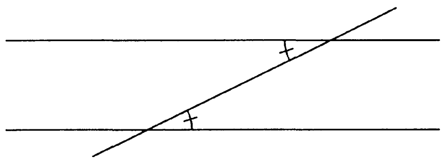
(e) The Greeks used the notion of harmony in a very general sense (the "harmony" of the soul, the "harmony" of the planets), but of course its fundamental meaning is musical. What is "harmonic" (in a musical sense) about the harmonic mean? *Hint*. See part (d) above.

(f) What is "geometric" about the geometric mean? *Hint*. Given two geometric magnitudes of lengths a and b, suppose we wish to find a magnitude x satisfying the geometric proportion $a : x :: x : b$. Show that x is geometric mean of a and b.

1.10. (*For use in exercise 3.13*) Calculate the following means:

(a) The harmonic mean between 3 and $2\sqrt{3}$. *Answer*: $12(2-\sqrt{3}) \approx 3.215390309$.
(b) The geometric mean between 3 and $12(2-\sqrt{3})$. *Answer*: $6\sqrt{2-\sqrt{3}} \approx 3.105828541$.
(c) The harmonic mean between 3.105828541 and 3.215390309. *Answer*: 3.159659942...
(d) The geometric mean between 3.105828541 and 3.159659942. *Answer*: 3.132628614...

1.11. The following proposition is typical of propositions in Euclid's *Elements*. Can you give a proof of this proposition?
If two parallel lines are cut by a transversal, then the alternate interior angles are equal.



(Prove that the indicated angles are equal. Look up a proof in a geometry book if you have trouble.)

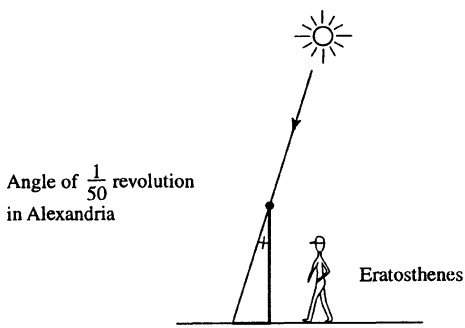
1.12. Euclid's proposition in exercise 1.11, like much of mathematics, may at first appear entirely "abstract", having no possible practical use. Before reading the next section, can you see any way at all that this proposition might ever be "useful"?

§2. Measuring the Earth

The root *geo-* means "earth" in Greek, and *geometry* literally means "earth-measurement". Eratosthenes (ca. 276–195 B.C.) did just that, with the aid of Euclid's proposition about alternate interior angles (see exercise 1.11 above). Eratosthenes convinced himself that the earth's circumference is about 50 times the distance from Alexandria to Aswan. (Aswan was known as "Syene" in the time of Eratosthenes, who worked in the great library at Alexandria, located in the delta of the Nile some 500 miles downstream from Aswan.)

We indicate part of Eratosthenes' reasoning, leaving the rest to the reader. Eratosthenes apparently proceeded upon the assumption that Alexandria was due north of Aswan. (It is not quite due north. Locate the two cities on a globe.) In Aswan there was a deep well that had an unusual feature. The sun shone straight down the well, casting no shadow at all, once every year: at the summer solstice. The sun, at noon on June

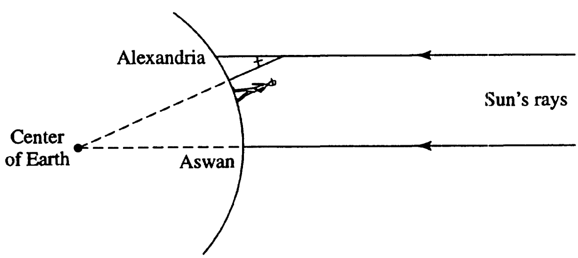
22, is directly overhead in Aswan, so the sunlight beaming down a vertical well is headed for the center of the earth. At the same time, in Alexandria, Eratosthenes observed the shadow cast by an upright pole, and measured an angle of slightly more than 7 degrees, or about one-fiftieth a complete revolution. (One-fiftieth a complete revolution is, in degrees, equal to $(\frac{1}{50})(360)$ degrees, or 7.2°.)



Exercises

2.1. Explain how, from the facts above, Eratosthenes might have argued that that the following proportion holds:

EARTH'S CIRCUMFERENCE : DISTANCE FROM ALEXANDRIA TO ASWAN :: 50:1.



2.2. Eratosthenes would have been able to judge the distance between Alexandria and Aswan by knowing how long it took soldiers to get from one city to the other at their standard marching pace. Given the modern measurement of 500 miles as the distance between these cities, explain how Eratosthenes would have immediately inferred that the circumference of the earth is about 25,000 miles.

2.3. Look up the circumference of the earth in an encyclopedia or almanac. How close is it to 25,000 miles?

2.4. In 1492 Columbus had his own idea of the earth's size. Did he think it larger or smaller than Eratosthenes had measured it? What would have happened if Columbus had believed Eratosthenes' calculation to be correct?

§3. Measuring the Circle

In Section 2 of Chapter 2 we introduced six ratios $r_1, r_2, \ldots, r_6$ of certain geometric magnitudes. Archimedes proved that r_3 and r_2 were the same ratio in disguise. To show $r_3 = r_2$ is to show that $A/r^2 = C/2r$, which—upon multiplying both sides by r^2—is seen to be equivalent to showing that

$$A = \tfrac{1}{2} rC, \tag{1}$$

i.e., that the area of a circle is equal to half the radius times the circumference. Archimedes proved equality (1) by showing that neither side is greater than the other. To do this he gave two *reductio ad absurdum* arguments, which are now among the most famous arguments in mathematics. The reader may work through the details of Archimedes' reasoning in some problems at the end of the chapter.

Taking equality (1) as established, we must do for r_2 and r_3 what we have already done for r_1. We have already given r_1 an appropriate name (by the Pythagorean theorem, r_1 must be $\sqrt{2}$) and we have learned in Section 6 of Chapter 2 how to approximate its numerical value as closely as we please.

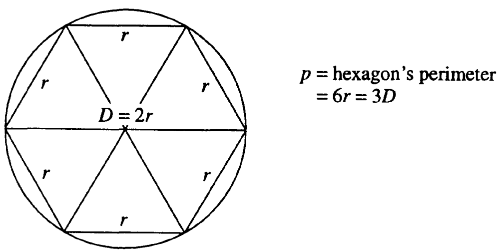
Surprisingly, it was not until the eighteenth century that it became conventional to use the Greek letter π to denote the common numerical value of r_2 and r_3. Presumably π is supposed to remind us of the first letter in *perimeter*, or *periphery*, which in turn is supposed to remind us that π fundamentally stands for our ratio r_2 of the circumference ("periphery") of a circle to its diameter. As a result of Archimedes' work we know that $r_3 = \pi = r_2$, which is to say that

$$\frac{A}{r^2} = \pi = \frac{C}{2r}.$$

Our familiar formulas for the area and circumference of a circle come directly from this. From the fact that $A/r^2 = \pi$ we see that $A = \pi r^2$ and from the fact that $\pi = C/2r$ we see that $C = 2\pi r$. Thus Archimedes' proof of equality (1), on which the equality of r_2 and r_3 depends, has quite significant consequences.

Calculating a numerical value for π is not as easy as calculating a numerical value for $\sqrt{2}$. The Babylonian method of approximating $\sqrt{2}$, discussed in Section 6 of Chapter 2, is easy to understand and simple to use. It can be proved that this method gives roughly twice as many correct decimal places with each succeeding guess. For example, if we have already guessed an approximation to $\sqrt{2}$ accurate to four decimal places, then by taking that guess as input in the Babylonian method we will get an output accurate to eight places. A succeeding step would give sixteen places, and we should have accuracy to thirty-two places with one more step. Thus we can easily generate a sequence of rational approximations tending very rapidly toward $\sqrt{2}$ as a limit.

There have been many attempts through the ages to find a ratio of integers equivalent to π. There are passages in the writings of ancient cultures enabling us to infer what those cultures took as numerical approximations to π. In the Old Testament, for example, there are a couple of passages alluding to a construction made about 1000 B.C. of a round container that is ten cubits from brim to brim, with a line of thirty cubits measuring its circumference. If we let a cubit take the role of our unit here, this might be interpreted as saying that $C : D :: 30 : 10$, so $C/D = 30/10 = 3$. It is easy to see, however, that in a circle the ratio $C : D$ certainly exceeds 3.



Here we have a regular hexagon inscribed in a circle, easily seen to be made up of six equilateral triangles. (*Regular* here means that all six sides have the same length.) Clearly, the perimeter p of the hexagon is less than the circumference C of the circle, so $p/D < C/D$. But $p/D = 3$ (as the reader is asked to show in exercise 3.9) and thus $3 < C/D$. Since $C/D = \pi$, this proves that

$$3 < \pi. \tag{2}$$

Inequality (2), coming from an inscribed hexagon, gives a *lower* bound for π. An *upper* bound can be gotten from studying a circumscribed hexagon. The reader is asked to do this in exercise 3.10 and deduce that

$$\pi < 2\sqrt{3} = 3.46\ldots \tag{3}$$

Thus, by using both an inscribed and a circumscribed regular polygon with 6 sides we see from inequalities (2) and (3) that

$$3.00 < \pi < 3.46. \tag{4}$$

By using inscribed and circumscribed regular polygons with 12 sides the reader can prove (in optional exercises 3.11 and 3.12) that

$$6\sqrt{2-\sqrt{3}} < \pi < 12(2-\sqrt{3}) \tag{5}$$

which, in decimal expression, takes the form

$$3.10 < \pi < 3.21. \tag{6}$$

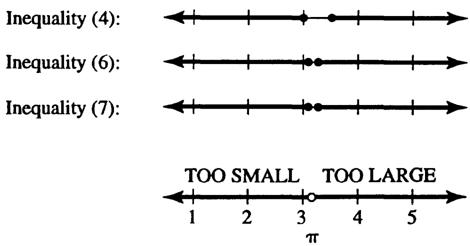
Archimedes went on to calculate what happens when 24, 48, and 96 sides are used. From the 96-sided inscribed and circumscribed polygon he concluded finally that

$$3\tfrac{10}{71} < \pi < 3\tfrac{10}{70}.$$

Expressed in decimal notation Archimedes' inequality implies that

$$3.1408 < \pi < 3.1429, \tag{7}$$

which justifies our writing $\pi = 3.14\ldots$ (note the ellipsis indicated by the three dots) to express in modern notation the information in (7). Let us picture the results so far obtained by our method of exhaustion:



Searching for all the wrong ratios:
Rational approximations to π by Archimedes' Method

Archimedes developed a systematic way (which we have not given here) to take the approximations given by a regular polygon with n sides and produce the approximations given by a polygon with $2n$ sides. At each stage only square roots are needed to calculate numerical results—as in going from inequality (4) to inequality (5) above. Thus, in principle, Archimedes' algorithm can be carried out to whatever accuracy one wishes. In practice, the calculations become tedious for large n, as they require one to multiply ever larger numbers by ever smaller ones, and Archimedes stopped with $n = 96$.

Some historians have found reason to speculate that Archimedes made a fresh attempt beginning with $n = 10$ and doubling consecutively until he reached $n = 640$. If so, he might have found π to an accuracy of four decimal places rather than two. This work, however, if ever done, is now lost. In any case, later mathematicians did such work and achieved closer approximations to π than Archimedes. The Chinese mathematician and astronomer Tsu Chung-Chi in the fifth century A.D. used polygons of several thousand sides to prove an inequality which we would express as

$$3.1415926 < \pi < 3.1415927. \tag{8}$$

For many centuries it seemed as if such refinements of Archimedes' idea offered the best way to approximate π. Methods of calculus, however, have suggested fresh, new approaches to this problem. None of these, however, suggest how far the sophisticated modern algorithms tailored for electronic computers have carried us. With their aid we now know π to an accuracy of several billion decimal places.

Exercises

3.1. From the formulas $A = \pi r^2$ and $C = 2\pi r$ you might be tempted to say that π is the area of a circle of radius one and π is also the circumference of a circle of diameter one. But this seems to say that *an area is a circumference*(?) Obviously, it is not quite correct to say such things. It is also not correct to say that π is 3.14. What is wrong? *Answer*: The number π is not an area or a length. It is the (dimensionless) ratio r_2, which Archimedes proved equivalent to r_3. The area of the circle of unit radius is π *square units*, and the length of circumference of a circle of unit diameter is π *units*. And it is simply false to say π is 3.14 because Archimedes proved $\pi > 3.14$ when he proved inequality (7).

3.2. There is an assumption implicit in our definition of π to which Eudoxus paid much attention. How do we know that the ratio of the area of a circle to the square on its radius is the same regardless of the size of the circle? *Hint*. This question is like the question (discussed in Section 4 of Chapter 2) about the ratio of the diagonal to the side of a square, regardless of the size of the square. Can you answer it in a similar way? Can you answer it in a different way by noting that Archimedes' method of approximating π results in the same approximations no matter what size the circle is?

3.3. (*Literal versus non-literal interpretations*) If an historical account of the ancient past speaks of a circle of circumference 30 cubits and diameter 10 cubits, should we interpret it literally to mean that $\pi = 30/10 = 3.0$? A modern scientist might interpret these data in terms of *significant digits*. Then to say the diameter D measures 10 cubits means simply that D lies somewhere between 9.5 and 10.5 cubits. Similarly, the circumference C is known only to lie somewhere between 29.5 and 30.5 cubits. Show that this implies that $59/21 < C/D < 61/19$, from which one may conclude that $2.81 < \pi < 3.22$. What do you think of this as an interpretation of *I Kings* 7:23?

3.4. Approximations can be confusing when combined with an "equals" sign. For each of the following equalities tell whether it is *true* or *false* and *why*.
(a) $\pi = 3.14\ldots$ *Answer*: True, because "3.14 ..." simply signifies a number located between 3.14 and 3.15, and inequality (7) shows that π is such a number.
(b) $\pi = 3.14$. *Hint*. See exercise 3.1.
(c) $\sqrt{2}+\sqrt{3} = 1.41\ \ldots + 1.73\ \ldots = 3.14\ \ldots$
(d) $3.14\ \ldots = \pi$.
(e) $\sqrt{2}+\sqrt{3} = 3.14\ \ldots = \pi$.
(f) $\sqrt{2}+\sqrt{3} = \pi$.
Hint. This has to do with our understanding of decimal representations, which is related to, but not identical with, the notion of significant digits. When we say, for example, that $\sqrt{2} = 1.41\ \ldots$ we mean, of course, only that $\sqrt{2}$ lies somewhere between 1.41 and 1.42. (When we say $\sqrt{2} \approx 1.41$ we mean $1.405 < \sqrt{2} < 1.415$.)

3.5. Archimedes showed that π lies somewhere in the range between $3\frac{10}{71}$ and $3\frac{10}{70}$.
(a) Does 355/113 lie in this range? This, as it turns out, is the closest ratio of "small" integers to π. The Chinese used 355/113 as early as the fifth century A.D. (To help remember 355/113, note that it is just 1,1,3,3,5,5 "upside-down".)
(b) Does 52163/16604 lie in this range?
(c) Does $\sqrt{2}+\sqrt{3}$ lie in this range? *Hint*. Use the results of exercises 6.1 and 6.5 of Chapter 2. (Make sure your answer to exercise 3.4(f) is consistent with your answer to this question.)

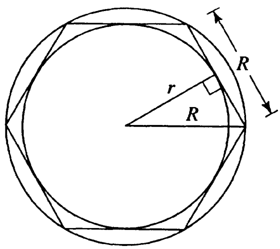
3.6. Are there infinitely many *rational* numbers between $3\frac{10}{71}$ and $3\frac{10}{70}$? Are there infinitely many *irrational* numbers between these two? Make a guess as to whether π is more likely to be *rational* or *irrational*. *Hint*. See exercise 5.17 of Chapter 2. (You are just asked to make a guess as to the question of the rationality of π. This difficult question was first settled by J.H. Lambert in a paper presented to the Berlin Academy in 1768.)

3.7. (*A modern way of approximating* π) Leaf ahead to find formula (13) in Section 6. Taking this formula for granted (it is given without proof), use the remark following it to arrive at a good approximation of π on your own.

3.8. (*Four guesses*) We have gotten to know r_1, r_2 and r_3. Make a guess about each of r_4, r_5, r_6, and r_7, introduced in Section 2 of Chapter 2, as to whether they are *rational* or *irrational*. (Archimedes knew the answer for each of these, but it will be easier for us to find these ratios after we learn some calculus.)

3.9. A regular hexagon inscribed in a circle is pictured just before inequality (2) above.
(a) Show that such a hexagon is made up of six equilateral triangles. *Hint*. If you divide 360 degrees by 6, you get the size of the central angle in each of these triangles. Then use exercise 3.1(a) of Chapter 2.
(b) Use the result of part (a) to prove inequality (2).

3.10. Consider a regular hexagon with an inscribed circle of radius r and a circumscribed circle of radius R. Let P denote the perimeter of the hexagon.
  (a) Show that P exceeds the circumference C of the smaller circle.
  (b) Show that $R/r = 2\sqrt{3}/3$ by applying the Pythagorean theorem to the right triangle in the figure here.
  (c) Use part (b) to show that $P/2r = 2\sqrt{3}$. *Hint.* $P/2r = (P/2R)(R/r)$ and you know $P/2R = 3$ from your work in exercise 3.9.
  (d) Use the results of parts (a) and (c) to prove inequality (3). *Hint.* $C/2r = \pi$.



3.11. (*Optional*) Archimedes next considered a regular polygon with 12 sides instead of 6. Since $360/12 = 30$ this means the polygon consists of 12 isosceles triangles whose legs are equal to the radius r of a circle, with a central angle of 30 degrees between the legs.
  (a) Show that if the legs of such an isosceles triangle have length r, then its base has length $r\sqrt{2-\sqrt{3}}$. *Hint.* You may need to find other lengths before finding the length of the base.
  (b) Show that the perimeter p of a 12-sided regular polygon inscribed within a circle of radius r is given by $p = 12r\sqrt{2-\sqrt{3}}$.
  (c) Use the result of part (b) to prove the left-hand inequality in (5). *Hint.* $p < C$, so $p/2r < C/2r$.

3.12. (*Optional*) Let P denote the perimeter of a 12-sided regular polygon circumscribed about a circle of radius r.
  (a) If p is the perimeter of the *inscribed* polygon as in exercise 3.11, show that $P/p = B/b$, where B and b are the respective lengths of bases of the isosceles triangles making up the circumscribed and inscribed polygons. *Hint.* This is easy: $P = 12B$ and $p = 12b$.
  (b) Show by similar triangles that $B/b = r/h$, where h is the height of the isosceles triangle with base b.
  (c) Show that $h = (r/2)\sqrt{2+\sqrt{3}}$.
  (d) Use (a), (b), and (c) to show $P/p = 2/\sqrt{2+\sqrt{3}}$.
  (e) Calculate the ratio $P/2r$ by multiplying P/p by $p/2r$. *Hint.* Remember that we know from exercise 3.11 that $p/2r = 6\sqrt{2-\sqrt{3}}$.
  (f) Simplify your answer to part (e) appropriately to prove the right-hand inequality in (5).

3.13. (*Gregory's amazing observation*) In the seventeenth century James Gregory observed that the inscribed and circumscribed perimeters calculated by successively doubling the number of sides can be found quite quickly by using harmonic and geometric means in a certain way.

(a) Describe how harmonic and geometric means seem to enter in the second line of the table here. *Hint.* Review exercise 1.10 (a), (b).

sides	$p/2r$	$P/2r$	
6	3	$2\sqrt{3}$	[inequality (4)]
12	$6\sqrt{2-\sqrt{3}}$	$12(2-\sqrt{3})$	[inequality (5)]
24	—	—	
48	—	—	
96	—	—	

(b) Assuming this pattern continues, calculate the numbers corresponding to 24 sides. *Hint.* Review exercise 1.10 (c), (d). Use decimal representations. (To *prove* the pattern continues, see Appendix 3, problem 10.)

(c) Assuming this pattern continues, calculate the numbers corresponding to 48 and to 96 sides. Is Archimedes' inequality (7) justified by your numbers for a 96-sided polygon?

(d) Assuming this pattern continues, keep doubling seven more times, until you have the numbers for a polygon of 12,288 sides. Is Tsu Chung-Chi's inequality (8) justified by your numbers for a 12,288-sided polygon? Is it justified by your numbers for a 24,576-sided polygon?

3.14. (*At last we have arrived!*) If you have access to an electronic spreadsheet, you can quickly extend the table you began in the previous problem to contain as many rows as you please. Keep on going until your computer prints the same decimal representation for both the lower and upper bounds for π. You are as close to π as your computer can get you.

§4. Measuring the Cone

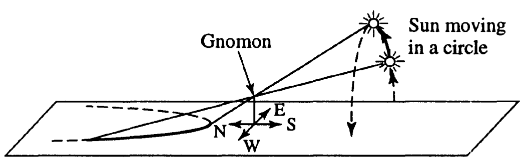
An ellipsis must accompany almost all decimal expansions (despite the confusion seen in exercise 3.4 that this practice can sometimes bring). An ellipsis is conventionally indicated in mathematics, as in rhetoric, by writing three dots to signify that something is being left out. *Ellipsis*, *hyperbole*, and *parable* are familiar terms from rhetoric that come from the same Greek roots as *ellipse*, *hyperbola*, and *parabola*, the names given to the three kinds of **conic sections** by Apollonius of Perga (250–175 B.C.).

Recall that an ellipsis involves abbreviating a longer statement; hyperbole is rhetorical exaggeration that overshoots the mark; while a parable is, of course, right on the mark. Their mathematical equivalents are *less than*, *greater than*, and *equal to*, and this is the way these Greek words entered Apollonius' long rhetorical sentences ("equations") describing the conditions that must be met in order for a point to lie on each of the three kinds of curves formed when a cone intersects a plane.

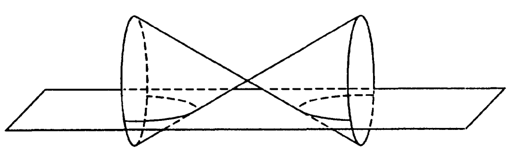
The easiest way to understand this is to refer back to the long statement of Apollonius quoted in Section 5 of Chapter 2 and imagine it written in its original Greek. The phrase *equal to* in "its square will be equal to the rectangle ..." would be expressed with a form of the Greek verb *paraballein* (literally "thrown alongside of"). Apollonius found that all three types of conic sections were described by virtually the same sentence except for the verb, and so—perhaps following Archimedes' advice—he named the three types of curves after the three verbs that differentiate them. Actually, parabolas, hyperbolas, and ellipses had been studied by the Greeks long before Apollonius gave them their names. No one knows for sure why they arose, though they were used in the fourth century B.C. in connection with the problem of the duplication of the cube.

It has been conjectured that the conics arose because of the sundial, which in its simplest form consists of an upright stick called a **gnomon** whose shadow moves with the sun. The word is related to the word *know*. The gnomon "knows" (i.e., measures) the time. The gnomon gradually became identified with anything vaguely in the shape of the Greek letter Γ (gamma), such as a carpenter's square to make right angles, or even the shapes of successive odd numbers out of which squares are made, as in exercise 1.4 of Chapter 2.

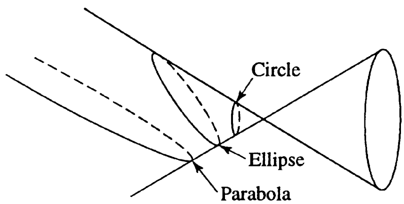
Since the sun is never directly overhead in Greece but always in the southern sky, one might model the sun's movement by a circle in a vertical plane offset to the south as indicated in the figure below. A branch of a hyperbola is traced out on the plane of the earth by the tip of the shadow of the gnomon as the sun rises and sets.



This picture naturally suggests a two-cusped cone, which is what Apollonius used. The curve pictured above is just one of two branches of a hyperbola, both branches being unbounded curves in the plane.



The other two types of conic sections occur when the plane cuts only one cusp of the cone. Usually this results in an ellipse, but we get a parabola if the plane happens to be parallel to one of the "sun's rays". Ellipses are always bounded curves in the plane, while a parabola is unbounded. A circle is a special case of an ellipse.

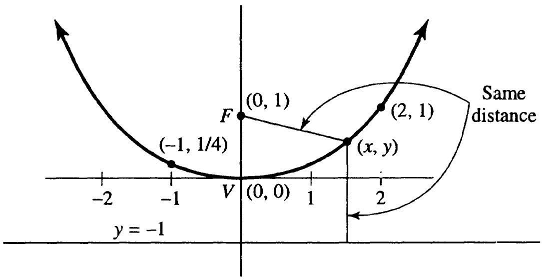


Johannes Kepler (1571–1630), the astronomer, found it fascinating to reflect how the shape of such curves is altered as the angle changes between the plane and the cone, everything else being fixed. Start with a circle made by intersecting a two-cusped cone with a plane and imagine the angle made by the plane with the cone being slowly increased. The circle evolves into oval-shaped ellipses that become more elongated, then a parabola appears at the instant the plane is parallel to the "side" of the second cusp, and suddenly a second branch appears as the plane catches the second cusp to make a hyperbola. Kepler thus saw a kind of continuous evolution of ellipse to hyperbola, with the parabola serving as intermediary, "trying to be like both". The parabola occurs, so to speak, at the instant when *less than* changes to *greater than* in Apollonius' rhetorical description of the curve.

We already know how to find an equation for a circle by specifying a center and radius—without thinking of the circle as a conic section one gets by intersecting a cone and a plane in a certain way. The Greeks, remarkably, proved analogous things about the other conic sections. We can find an equation for a parabola by specifying a *focus* and *directrix*—without thinking of the parabola as the intersection of a cone and a plane. The focus F is a point, the directrix $\mathbf{D}$ is a line not containing the focus. *The curve consisting of all points P equidistant from F and* $\mathbf{D}$ is a parabola, where by the distance from a point to a line is meant the perpendicular distance.

A whimsical description may be easier to grasp. Pretend the line $\mathbf{D}$ is the edge of a beach and the point F is the location of a (point-sized) boat at sea. Someone yells "Shark!" and each swimmer swims toward the nearest safe haven—either directly toward the beach $\mathbf{D}$ or toward the boat F. The parabola consists of the locations of all swimmers who are equally far from both safe havens (and perhaps hesitate before deciding which way to go).

It is obvious from this description that the point V halfway between F and $\mathbf{D}$ will lie on the parabola, and that the parabola will be symmetric with respect to its *axis* (the line through F and V). The point V is called the *vertex* of the parabola.



If we skip from the time of Apollonius to the seventeenth century we can use Cartesian coordinates to find an algebraic equation for the parabola whose focus F is (0, 1) and whose directrix $\mathbf{D}$ is the horizontal line given by the equation $y = -1$. The vertex V, halfway between F and $\mathbf{D}$, is then located at the origin (0, 0), and the axis of the parabola (the line through F and V) is the y-axis. To find such an equation is to translate the statement

> "(x, y) is on the parabola with focus $F = (0, 1)$ and whose directrix $\mathbf{D}$ is the line $y = -1$"

into an algebraic equation. This we do in the following steps, each of which is easily seen to be equivalent to the preceding:

$$\text{Dist } (x, y) \text{ to } F \text{ is the same as Dist } (x, y) \text{ to } \mathbf{D} \tag{9}$$

$$\text{Dist } (x, y) \text{ to } (0, 1) \text{ is the same as Dist } (x, y) \text{ to } (x, -1)$$

$$\sqrt{(x-0)^2 + (y-1)^2} = \sqrt{(x-x)^2 + (y-(-1))^2}$$

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

$$x^2 + y^2 - 2y + 1 = y^2 + 2y + 1$$

$$4y = x^2$$

$$y = \tfrac{1}{4}x^2. \tag{10}$$

Equation (10) gives y as a function of x. We get a function, of course, because a parabola, if placed on a Cartesian coordinate system with its axis parallel to the y-axis, is never cut twice by any vertical line. Moreover, as Descartes and Fermat discovered, the graph of any "quadratic function" of the form $y = ax^2 + bx + c$ is a parabola. We shall see later, in connection with a discussion of the quadratic formula, that any quadratic function falls into one of the six types pictured in Section 9 of Chapter 4.

Equation (10), if rewritten as $x^2 = 4y$, comes out rather as Apollonius said it would in the passage quoted in Section 5 of Chapter 2. The expression x^2, interpreted geometrically, is the area of a square whose sides have length x units, and the expression $4y$ is the area of a rectangle of length y units and width 4 units. Thus the equation $x^2 = 4y$ says that the square made on the line segment running parallel to the directrix from the point (x, y) to the parabola's axis has the same area as a rectangle made by a segment of length y and a certain fixed line of length 4 units.

Let us rewrite condition (9), which describes a parabola, in terms of the ratio r defined by

$$r = \frac{\text{Dist } (x, y) \text{ to } F}{\text{Dist } (x, y) \text{ to } \mathbf{D}}. \tag{11}$$

Equation (9) says (*why*?) that if r is set equal to 1 then the graph of equation (11) is a parabola. The Greeks discovered that an *ellipse* is described by equation (11) if we set r equal to any fixed positive number *less than* 1, while a *hyperbola* is described by equation (11) if we set r equal to any fixed positive number *greater than* 1. The ratio r is now known as the *eccentricity* of the conic section described by equation (11).

We cannot do justice to the conic sections here because the topic is far too large. Apollonius wrote a treatise on the conics containing eight books and still left many interesting things for others to say later. The most famous statement regarding conics is surely Kepler's observation that each planet goes around the sun in an elliptical orbit with the sun at a focus. The eccentricities of planetary orbits range from 0.01 to 0.09, except for Mercury and Pluto, whose eccentricities are about 0.21 and 0.25. The smaller the eccentricity, the more circular is the orbit.

Kepler, incidentally, was the first to refer to the crucial point defining a conic (the point where he happily located the sun) by the Latin word *focus*, which literally means "fireplace" or "hearth". Apollonius had described how to find this crucial point, given the cone and the cutting plane, and recognized how it interacted with the directrix in defining the curve, but never gave it a name. In fact, the ellipse and hyperbola each have two foci and two directrices (see exercises 4.4 through 4.7), and Apollonius knew how to describe these curves in terms of their foci without reference to their directrices. If the reader has studied these curves before, the chances are that they were introduced by giving Apollonius' description in terms of their two foci.

Newton (1642–1727) later combined calculus and physics to deduce, in essence, that a conic section with the sun at one focus is the only possible path for a planet or comet (or any relatively small heavenly body), so long as the only force acting upon it is the gravitational attraction of the sun. Since an ellipse is the only bounded conic section and since the path of a planet or a (returning) comet is bounded, such paths must be elliptical. Halley's comet, which returns every 76 years or so, has an elliptical orbit with an eccentricity of about 0.98. With its eccentricity very close to 1.00, the orbit of Halley's comet is an ellipse that very closely resembles a parabola. A comet with a parabolic (or hyperbolic) orbit would, of course, never return to the solar system.

Galileo (1564–1642) argued that a parabolic path should be taken by a cannonball in flight if the air friction that slows it down is ignored. Parabolic reflectors are commonly used today in everything from satellite dishes to headlights. The analytic geometry of Fermat and Descartes reveals that every conic section, when placed on a Cartesian coordinate system, has an algebraic equation in x and y of second degree. The study of second-degree algebraic equations is, in essence, the study of conic sections!

Pascal (1623–1662) and Desargues (1591–1661) invented a whole new geometry (projective geometry) by reflecting upon how conic sections are related to the various ways the shadow of a circular ring can be projected on a wall. The circle gives rise to the trigonometric functions and the most natural definition of logarithms involves the numerical calculation (by the method of exhaustion) of areas beneath portions of the hyperbola $y = 1/x$. The cone has proved to be misnamed. It is, both literally and figuratively, a cornucopia.

Exercises

4.1. Consider the parabola whose focus F is at (0, 2) and whose directrix **D** is the line with equation $y = -2$.
   (a) What are the coordinates of the vertex of this parabola?
   (b) What is the axis of this parabola?
   (c) Find an equation for this parabola. *Hint.* Begin with the defining sentence (9) and proceed in the direction of equation (10), making appropriate modifications.

4.2. Consider the parabola whose focus F is at $(0,p)$ and whose directrix is the line whose equation is $y = -p$. Find an equation for this parabola. *Answer*: $4py = x^2$.

4.3. In a sense, the "simplest" parabola is the one whose equation is $y = x^2$. Where is its focus? Where is its directrix? *Hint.* This is really quite easy, in view of exercise 4.2. The equation $y = x^2$ is the equation $4py = x^2$ provided $p = 1/4$. So where are the focus and directrix?

4.4. Consider the hyperbola whose focus F is (0, 1), whose directrix **D** is given by the line $y = -1$, and whose eccentricity is 2.
(a) Find an equation of this hyperbola. *Hint.* This means we should set r equal to 2 in condition (11), which implies

$$\text{Dist } (x, y) \text{ to } (0, 1) \text{ is } twice \text{ the Dist } (x, y) \text{ to } (x, -1).$$

Now proceed just as in getting from (9) to (10) above, except take account of the extra factor of 2. *Answer*: $x^2 - 3y^2 - 10y - 3 = 0$.
(b) Is $(0, -3)$ on this hyperbola?
(c) Is $(0, -1/3)$ on this hyperbola?
(d) Is the graph of this hyperbola the graph of a function?

4.5. Consider the ellipse whose focus F is (0, 1), whose directrix **D** is given by the line $y = -1$, and whose eccentricity is 1/2.
(a) Find an equation of this ellipse. *Answer*: $4x^2 + 3y^2 - 10y + 3 = 0$.
(b) Is (0, 3) on this ellipse?
(c) Is (0, 1/3) on this ellipse?
(d) Is the graph of this ellipse the graph of a function?

4.6. Find an equation of the ellipse whose focus F is (0,7/3), whose directrix **D** is given by the line $y = 13/3$, and whose eccentricity is 1/2. *Answer*: This, as it turns out, is the same ellipse as in exercise 4.5, but described in terms of its *other* focus and directrix. Condition (11) in this setting should work out to the same answer as in exercise 4.5.

4.7. Find an equation of the hyperbola whose focus F is $(0, -13/3)$, whose directrix **D** is given by the line $y = -7/3$, and whose eccentricity is 2. *Answer*: This has been set up to be the same hyperbola as in exercise 4.4, but described in terms of its *other* focus and directrix. Condition (11) in this setting should work out to the same answer as in exercise 4.4.

4.8. (*Curve sketching*) the hyperbola and ellipse of exercise 4.4 and 4.5 have the same focus and directrix as the parabola whose graph is already sketched in this section. Use the equations found in these exercises—or just use their descriptions in terms of equation (11)—to make rough sketches of the graphs of these two curves on the same coordinate system as the parabola. Notice how the ellipse is "inside" the parabola and the hyperbola is "outside". Label the three curves with their eccentricities of 1/2, 1, and 2. Does this give you a sense of what Kepler meant by the parabola "trying to be like both an ellipse and a hyperbola"?

4.9. (*The "simplest" hyperbola*) We have investigated the "simplest" parabola $y = x^2$ in exercise 4.3. Surely the simplest ellipse is the unit circle $x^2 + y^2 = 1$, which we have investigated in Chapter 2. We should not close this section without at least mentioning the simplest hyperbola, whose equation is $y = 1/x$. This curve plays an important role in constructing logarithms, as will be indicated in problems 21–22 of Chapter 6 and in problem 29 of Chapter 7. (It is not obvious from our definition in terms of equation (11) that the graph of $y = 1/x$ is a hyperbola, but it is. It has a focus at $(\sqrt{2}, \sqrt{2})$, and an eccentricity of $\sqrt{2}$.)

(a) Sketch a rough graph of the hyperbola given by $y = 1/x$. *Answer*: (Its graph is sketched in Section 1 of Chapter 5.)
(b) Is the graph of $y = 1/x$ the graph of a *function*? If so, what is its domain and range?

§5. The Spirit of Archimedes

Euclid's *Elements*, despite its achievements in rigor and synthesis, fails to reflect the nature of Greek mathematics in two important respects, both of which are vividly illustrated in the person of Archimedes of Syracuse (287–212 B.C.), the supreme mathematician of antiquity. In the first place, that spirit of delight in discovery—the spirit so associated with Archimedes—is nowhere to be found in Euclid. Second, the deepest work of Archimedes goes far beyond the elementary "straightedge and compass" constructions to which Euclid restricts himself in his *Elements*.

What did Archimedes do? He developed a significant part of the calculus. The fundamental notion of the calculus—that of a *limit*—was well understood by Archimedes, although he did not call it by name. His understanding of the essential idea is implicit in his work. What is even more surprising, Archimedes had a clearer grasp of this notion than the seventeenth-century mathematicians who invented the term.

In addition to his mathematics, which includes writings of unmistakable modernity in spirit (just glance for a moment at the appendix on Archimedes), Archimedes developed the theory of floating bodies into the science now known as hydrostatics. In the course of doing this, he effectively created mathematical physics. He was also an inventor of ingenious and useful devices such as a water pump, elaborate compound pulleys utilizing the law of the lever to remarkable advantage, and a mechanical contraption that described accurately the motions of the heavenly bodies. In spite of these accomplishments in applied mathematics, he is said to have regarded himself as the purest of pure mathematicians. Even today among mathematicians, only Newton and Gauss are mentioned in the same breath as Archimedes.

Rome discovered Archimedes the hard way. Attacking Syracuse in 214 B.C., the Roman general Marcellus had no way of guessing how formidable a foe he would encounter. Archimedes, in great old age, had invented and deployed all manner of weapons and techniques to repel the Roman legions. Plutarch's description of the campaign of Marcellus abruptly shifts to a description of Archimedes himself, who scared the pluperfect hell out of the Romans:

> In fine, when such terror had seized upon the Romans that, if they did but see a little rope or a piece of wood from the wall, instantly crying out, that there it was again, Archimedes was about to let fly some

> engine at them, they turned their backs and fled; Marcellus desisted from conflicts and assaults, putting all his hope in a long siege.
>
> Yet Archimedes possessed so high a spirit, so profound a soul, and such treasures of scientific knowledge, that though these inventions had now obtained him the renown of more than human sagacity, he yet would not deign to leave behind him any commentary or writing on such subjects; but, repudiating as sordid and ignoble the whole trade of engineering, and every sort of art that lends itself to mere use and profit, he placed his whole affection and ambition in those purer speculations where there can be no reference to the vulgar needs of life; studies the superiority of which to all others is unquestioned, and in which the only doubt can be whether the beauty and grandeur of the subjects examined, or the precision and cogency of the methods and means of proof, most deserve our admiration. It is not possible to find in all geometry more difficult and intricate questions, or more simple and lucid explanations. Some ascribe this to his natural genius; while others think that incredible effort and toil produced these, to all appearances, easy and unlabored results....
>
> And thus it ceases to be incredible that (as is commonly told of him) the charm of his familiar and domestic Siren made him forget his food and neglect his person, to that degree that when he was occasionally carried by absolute violence to bathe or have his body anointed, he used to trace geometrical figures in the ashes of the fire, and diagrams in the oll on his body, being in a state of entire preoccupation, and, in the truest sense, divine possession with his love and delight in science.
>
> —Plutarch's *Lives*, translated by John Dryden

When historians speak of a spirit—such as the spirit of democracy or the spirit of Rome—they usually refer to an abstract idea. Here, however, Plutarch describes a spirit that is the very opposite of an abstraction. It is so real, in fact, that it even cries out and is heard, although only on auspicious occasions. This spirit was never so much at home as when it resided within the body of Archimedes and raised the roofs of Syracuse with its colossal shouts of surprise and delight.

Archimedes' spirit has been described by phrases rendered into English both as a "raging Siren" and as a "familiar demon". The first is an apt description of the overflow of exuberance in the moment of light; but such moments come only to those who can stand the dark. The second seems more descriptive of that spirit of compelling total engagement whose charm kept Archimedes through the long nights and made him "trace geometrical figures in the ashes of the fire, and diagrams in the oil on his body, being in a state of entire preoccupation, and, in the truest sense, divine possession." All students of mathematics have known all manner of approximations to this kind of spirit. Without its aid one can do little; with its aid one can, like Archimedes, aspire to move the earth.

Indeed, this same spirit is almost as conspicuously present in the first modern mathematician powerful enough to be compared with Archimedes. Isaac Newton could also muster the self-discipline to keep his subject constantly before him for great periods of time in order to "wait till the first dawnings open little by little into the full light." Almost supernatural power can come to those who struggle with a problem so hard as to feel its presence within themselves. The Greek phrase *en theos*, "a god within"—from which *enthusiasm* comes—suggests the power that such an engagement can generate.

Archimedes' power of concentration, however, caused his death by sword at the age of 75. His military inventions had helped to keep the Romans at bay for some two years, but Marcellus' men finally broke through. According to one account, Archimedes was struck down when he ignored a Roman soldier's order and continued to study the lines and curves he had drawn in the sand. This was in 212 B.C. Eratosthenes and Apollonius died fairly soon thereafter, and the glory of Greek mathematics was soon to fade.

The long period of Roman domination followed, but as we have noted, the Greek *enkuklios paideia* lived on with its new Latin name, *artes liberales*. Its survival was aided by educated Greek slaves who tutored their Roman masters' children, an irony that did not go unnoticed by the Romans. *Graecia capta ferum victorem cepit*—"Conquered Greece took captive her barbarous conqueror"—wrote the Roman poet Horace around 15 B.C.

Cicero (106–43 B.C.) wrote in old age of his tenure in Sicily in 75 B.C. when he remembered having learned some verses describing a sphere within a cylinder marking the tomb of Archimedes, Syracuse's "one most ingenious citizen" yet an "obscure, insignificant person". After some search in the company of prominent Syracusans who knew nothing of the tomb, he spotted a small column of this description. Slaves were sent in to remove a dense overgrowth of brambles and thickets, and Cicero found the beginning words of these verses barely legible on the pedestal. He restored the tomb, although it has since disappeared.

This poignant story, it has been remarked, is almost the only contribution to the history of mathematics made by a Roman. Although Cicero venerated Plato, he ignored Plato's strong endorsement of the quadrivium and used the Latin term *humanitas* to translate the Greek *paideia*. He valued mathematics only for its uses in engineering or in everyday life:

> With the Greeks geometry was regarded with the utmost respect, and consequently none were held in greater honor than mathematicians, but we Romans have restricted this art to the practical purposes of measuring and reckoning.
>
> But on the other hand we speedily welcomed the orator....
>
> —Cicero, *Tusculan Disputations*

When the Romans allowed *oratio* to supersede *ratio* as the key element in the *artes liberales*, they shifted the center of education decisively toward the trivium, pushing mathematics far to the side to be studied only for its practical utility as an applied science. We should not be too quick to fault the Romans, however, for today we still fail to see how chimerical is the nature of utilitarianism. The immediately practical may soon become as useless as Roman numerals, while ultimately useful things, such as a liberal education, may never become immediately practical.

We are surely free, like the Romans, to try to embrace utilitarianism if we wish; but who is wise enough to say today what might be useful tomorrow? Consider the calculus, which was to grow out of seventeenth-century mathematics. It is still prized highly today for its utility, even by some who value relevance and applicability of knowledge more than knowledge itself. Calculus is indispensable to the modern engineer. Yet Archimedes, who "repudiat[ed] as sordid and ignoble the whole trade of engineering, and every sort of art that lends itself to mere use and profit," had unlocked some of the secrets of the calculus in the normal course of his studies.

Today the other secrets of the calculus seem to us not far from Archimedes. But Roman engineers added to mathematics little of value. For a thousand years the puzzling riddles lay right where Archimedes fell, in the reddening sand, amongst pebbles, lines, and curves.

Exercises

5.1 Archimedes is famous in mathematics for his deep results about areas, volumes, and centers of gravity, which are notoriously difficult. Yet his method of trisecting an angle is beautifully simple. Read about this method at the beginning of the appendix on Archimedes, then prove the method works by using the hint in problem 1 at the end of that appendix.

5.2 Alfred North Whitehead was fond of comparing the Greek and Roman attitudes toward mathematics (and toward abstract thought in general) by simply remarking that no Roman ever lost his life because he was engaged in the contemplation of a mathematical diagram. Explain in more detail the point that Whitehead was making.

5.3 Do a little outside reading about the Greeks, and particularly about Archimedes. For example, read pages 19–34 of E. T. Bell's *Men of Mathematics* (Simon and Schuster, New York, 1937).

(a) What did Archimedes mean when he said, "Give me a place to stand on, and I will move the earth!"

(b) What does E. T. Bell mean when he says that modern mathematics was born with Archimedes and died with him for over two thousand years? *Hint.* See problem 16 at the end of this chapter.

(c) "Eureka! Eureka!" shouted the streaking sage of Syracuse. Why?

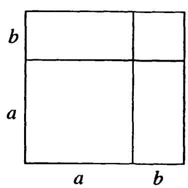
§6. What is the Rest of the Story?

We have just about finished our review of precalculus mathematics and should be well prepared to turn in the next chapter to the new ideas of calculus developed in the seventeenth century. Thus we shall be moving quickly from the time of Archimedes to the rise of calculus, two periods separated by nearly 2000 years. Mathematically, this transition is not difficult because of the astonishing modernity of Archimedes. Historically, however, nothing could be more wrenching. If this were a history of calculus, rather than a pedagogical approach making selective use of history, we should thoroughly discuss this long period. Instead, we must content ourselves with the briefest of summaries.

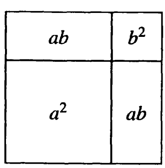
For whatever reasons (see problem 26 at the end of this chapter), western Europe added little to mathematics for many centuries following the death of Archimedes, except for the work of some Alexandrian Greeks such as Claudius Ptolemy (ca. 100–178) and Pappus (ca. 250–300). Chinese and Indian mathematicians cultivated the subject and made some innovations, but mathematics did not begin to steer out of the doldrums until it was buoyed up and carried forward by the rising tide of the Islamic movement in the eighth and ninth centuries.

Islamic scientists were particularly attracted to astronomy, to which they made many new contributions, and which inspired their development of new theorems in solid geometry and of remarkably detailed and accurate trigonometric tables. Ptolemy's book on Greek astronomy was highly esteemed by the Arabs, who called it *al-magistri*, "the greatest", and ever since it has been known as the *Almagest*. They preserved, translated into Arabic, studied, and in some cases expanded upon many of the Greek classics written by Euclid, Archimedes, and Apollonius—not to mention the great number of nonmathematical Greek and Roman classics that attracted them as well.

They also popularized the Hindu-Arabic numerals, now almost universally used, and they experimented with decimal fractions. In addition, Islamic mathematicians cultivated algebra, itself an Arabic word—*al-jabr*, meaning "restoring"—that comes from the title to al-Khowarizmi's famous book written around 825. With the Greeks, algebra had been tied to geometry, but now this connection was being severed. The Arabs had nothing like our modern notation, but they took an important step in this direction.



Square of "a plus b"



Greek "geometrical" algebra

$$(a+b)^2 = a^2 + 2ab + b^2$$

Modern equivalent expression

The Egyptian mathematician ibn al-Haytham (965–1039)—known in the West as Alhazen and famous also for his work in optics—developed a technique using figurate numbers by which one could, in principle, find a formula for the sums of powers of integers, generalizing the Greek formulas which stopped with sums of cubes. Thus, for example, he proved that

$$1^4 + 2^4 + 3^4 + \cdots + n^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}.$$

Ibn al-Haytham's approach is like that discussed in Section 1 of Appendix 2 in this book, which uses the same idea as rediscovered by Blaise Pascal (1623–1662). Such formulas, as we shall see in Chapter 7, are helpful in determining certain areas and volumes.

Mathematics is often rediscovered because no one can know everything that has been learned before, particularly if it is written in a foreign language. Ibn al-Haytham and other Islamic mathematicians did not know all of Archimedes' works and, in the course of adding original results relating to volumes of solids and their centers of gravity, also rediscovered some deep theorems of Archimedes. The Persian poet al-Khayyami—famous in the West as Omar Khayyam (1048–1131)—was also a mathematician, wrote a book on algebra, and worked hard and effectively on general methods of solving cubic equations, a topic that attracted many of his colleagues and to which Archimedes had also contributed.

Not too long ago it was commonly thought that Islamic mathematicians mainly preserved and refined existing material. Recent historical scholarship has revealed, however, that they made quite strikingly original contributions, as indicated above, and had the capacity to transmit to other cultures by the twelfth and thirteenth centuries a great deal more than they had received from the Greeks and Indians in the eighth and ninth centuries.

While Islamic science was declining, some amazing discoveries involving sums of infinite series were made by Indian mathematicians, for example,

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots \tag{12}$$

and

$$\frac{\pi}{4} = \frac{3}{4} + \frac{1}{3^3 - 3} - \frac{1}{5^3 - 5} + \frac{1}{7^3 - 7} - \cdots \tag{13}$$

These equalities appear in Sanskrit verse from an early sixteenth-century Indian work. Adding up just a few terms of the right-hand side of equality (13) and then multiplying by 4 yields an excellent approximation to π. Unfortunately, these and many other Islamic and Indian contributions did not make their way to the West and had to be rediscovered by European mathematicians on their way to the development of calculus.

There were also adumbrations of calculus in the work of Seki Kowa (1642–1708), "the Japanese Newton", but little is known for certain because of Japan's isolation at that time and the Samurai code of modesty to which Seki subscribed. There has been found, however, a 1683 manuscript of Seki's on determinants, which were independently introduced in Europe by Leibniz (1646–1716), who, incidentally, rediscovered formula (12) above.

The torch that first burned so brilliantly in Greece has thus been taken up by diverse peoples and cultures, and Archimedes' cry of surprise and delight now resounds in many tongues. Does this not suggest that the spirit behind Archimedes' Olympian quest for excellence belongs to all people of all cultures? Surely it behooves us to view the spirit of Archimedes as Plato would have us view an eternal form—not as an expression of the genius of one man or one culture, but as part of the essence of humanity itself, exerting its influence through time immemorial.

Is it not this quintessentially human spirit that makes us cry out in exaltation when we succeed in the face of overwhelming odds? Until the last eureka is shouted, the spirit of mathematics will live.

Problem Set for Chapter 3

1. Which aspects of the Pythagorean philosophy influenced Plato? Describe how.

2. Plato and Aristotle held differing views about the nature of mathematics and about the value of mathematics. Contrast their opinions on each of these issues.

3. (*Where did geometry come from?*) Aristotle said that geometry was cultivated by the priestly class in Egypt and Democritus presumably referred to surveyors when he spoke of Egyptian "rope-stretchers". Yet the Pythagoreans, in their analogy

   GEOMETRY : ASTRONOMY :: ARITHMETIC : MUSIC,

   associated geometry with the heavens, a source of fascination even in prehistoric times.
   (a) What is the literal meaning of the Greek word *geometry*?
   (b) Where do you think most likely that geometry originated? Did it come from the earth, from the heavens, or from somewhere else?

(c) Plato said that the word *geometry* is a misnomer, for the subject-matter of the geometry developed in Greece is not found in the earth—or in the heavens either, for that matter. What did Plato mean by this?

4. (*See exercise 1.9 for definitions of arithmetic, geometric, and harmonic means.*) An *arithmetic sequence* is a sequence like $1, 2, 3, 4, \ldots$, where each term (except, of course, the first and the last) is the arithmetic mean of the terms before and after it. A *harmonic sequence* is a sequence like $1, 1/2, 1/3, 1/4, \ldots$, where each term is the harmonic mean of the terms before and after. And a *geometric sequence* is defined similarly.
   (a) What kind of sequence is $1, x, x^2, x^3, \ldots$?
   (b) If $a_1, a_2, a_3, \ldots$ is an arithmetic sequence of nonzero numbers, what kind of sequence is $1/a_1, 1/a_2, 1/a_3, \ldots$?
   (c) Is the sequence $1, 3, 6, 10, \ldots$ of triangular numbers *arithmetic? harmonic? geometric?* How about the sequence of squares: 1, 4, 9, 16...?
   (d) Is it possible for a sequence to be arithmetic, harmonic, and geometric at the same time? If so, give an example of such a sequence. *Hint.* Only the dullest type of sequence can be all three at once.
   (e) The successive notes on a piano—C, C#, D, D#, E, F, F#, etc.—might be loosely described as an *arithmetic* sequence, since they progress from the *first* note to the *second* to the *third*, etc. The successive pitches (frequencies) of these notes is not an arithmetic sequence, however. What kind of sequence is it?

5. "Limits are as simple as pi." This pun contains a good deal of truth. Explain.

6. There is some reason to believe that Plato conjectured that $\sqrt{2}$ and $\sqrt{3}$, when added, should give π. Here Plato was thinking along philosophical—not mathematical—lines. Explain clearly how, in the century following Plato, Archimedes finally settled this conjecture. *Suggestion*: Re-work the ideas circulating around exercise 3.5 (c), explaining carefully in a couple of paragraphs the reasoning involved.

7. (*What is the sum of cubes?*) The Greeks noticed a pattern relating cubes and odd numbers: $1^3 = 1$, the first odd number; $2^3 = 8 = 3 + 5$, the sum of the next two odd numbers; $3^3 = 27 = 7 + 9 + 11$, the sum of the next three odd numbers; $4^3 = 64 = 13 + 15 + 17 + 19$, the sum of the next four odd numbers. Thus the sum of the first *four* cubes is equal to the sum of the first *ten* odd numbers (since $10 = 1 + 2 + 3 + 4$), which is in turn equal to the square of 10, or 100. That is,

$$\begin{aligned} 1^3 + 2^3 + 3^3 + 4^3 &= 1 + 3 + 5 + \cdots + 19 \\ &= \text{sum of first ten } (= 1 + 2 + 3 + 4) \text{ odd numbers} \\ &= \text{the square of ten [by exercise 1.4, Chapter 2]} \\ &= 100. \end{aligned}$$

   (a) The pattern continues, revealing a simple formula for the sum of cubes. By extending this pattern to the general case, show that the sum of the first n cubes is equal to the square of $n(n+1)/2$. *Suggestion*: Copy down the following steps and explain how the results of exercise 1.4 and 1.5 of Chapter 2 help justify them:

$$
\begin{aligned}
&1^3+2^3+3^3+4^3+\cdots+n^3\\
&\qquad= \text{the sum of the first } (1+2+\cdots+n) \text{ odd numbers}\\
&\qquad= \text{the sum of the first } \frac{n(n+1)}{2} \text{ odd numbers}\\
&\qquad= \text{the square of } \frac{n(n+1)}{2}\\
&\qquad= \frac{n^2(n+1)^2}{4}.
\end{aligned}
$$

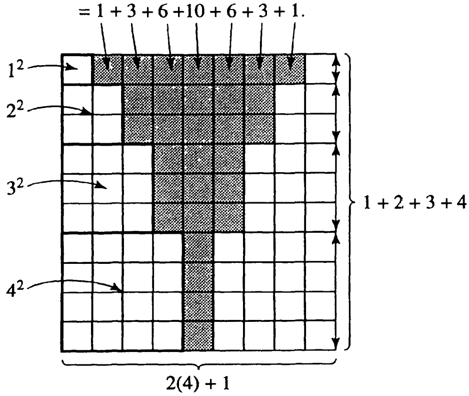
(b) Use the formula of part (a) to evaluate the sum $1^3+2^3+3^3+4^3+\cdots+100^3$.

(c) Does your work in this problem destroy the magic in problem 21 of Chapter 1?

8. (*What is the sum of squares?*) In the preceding problem we derived a formula for the sum of cubes. The formula for the sum of squares is more difficult, but was discovered early on. Archimedes explained how it could be discovered by reasoning somewhat along the following lines.

(a) Recall from exercise 1.7 of Chapter 2 that the triangular numbers are 1, 3, 6, 10, 15, etc., the n-th triangular number being $1+2+\cdots+n=n(n+1)/2$. It is obvious from consideration of a square as a figurate number that any square (greater than 1) is the sum of two consecutive triangular numbers. Thus $4=1+3$; $9=3+6$; $16=6+10$, etc. Hence *any sum of squares can always be rewritten as a sum of triangular numbers.* For example,

$$
\begin{aligned}
1^2+2^2+3^2+4^2 &= 1+(1+3)+(3+6)+(6+10)\\
&= 1+3+6+10+6+3+1.
\end{aligned}
$$



Since each of the three areas of the large rectangle above is equal to $1^2+2^2+3^2+4^2$ it follows that

$$\begin{aligned}3(1^2+2^2+3^2+4^2) &= \text{area of large rectangle}\\ &= (1+2+3+4)(2(4)+1)\\ &= \frac{4(4+1)}{2}(2(4)+1).\\ &\quad \text{[by exercise 1.5, Chapter 2]}\end{aligned}$$

Show by the same reasoning (using exercise 1.5 and 1.7 of Chapter 2) that

$$\begin{aligned}3(1^2+2^2+\cdots+n^2) &= (1+2+\cdots+n)(2n+1)\\ &= \frac{n(n+1)}{2}(2n+1),\end{aligned}$$

then divide both sides by 3 to get the formula for the sum of squares:

$$1^2+2^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6}.$$

9. (*The pyramid problem is now easy*) Imagine a huge "pyramid" of baseballs resting upon a square base. The "top vertex" of the pyramid is a single baseball which rests upon 4 baseballs in the shape of a 2×2 square. These in turn rest upon 9 baseballs in the shape of a 3×3 square. And so on, down to the bottom. If the base of the pyramid consists of a square of baseballs of size 100×100, how many balls are in the entire pyramid? *Hint.* The total number of balls is obviously given by $1^2 + 2^2 + \cdots + 100^2$, and this is quickly evaluated using the formula derived at the end of the preceding problem.

10. (*Approximating the ratio r_6*) Consider again the "pyramid" in the preceding problem and consider its size relative to a "cube" of baseballs also built on a base of 100×100 balls.
    (a) Find the ratio of the number of balls in the cube to the number of balls in the pyramid whose base is the same size. *Hint.* The number of balls in a cube built on a base of 100×100 balls is, of course, 100^3.
    (b) Explain why you should expect this ratio to be very close to, but not equal to, the ratio r_6 introduced in Chapter 2, Section 2.
    (c) Re-do the ratio of the cube to the pyramid if both have a base of 1000×1000 balls instead of 100×100. Explain why you should expect this ratio to be even closer to r_6 than your answer to part (a).
    (d) Re-do the ratio of the cube to the pyramid if both have a base of $10{,}000 \times 10{,}000$ balls. Explain why you should expect this ratio to be even closer to r_6 than your answer to part (c).

11. (*The ratio r_6 as a limit?*) Find the ratio of the cube to the pyramid described in the preceding problem if both have a base of $n \times n$ balls. Show that this ratio is given by

$$\frac{6n^2}{(n+1)(2n+1)}.$$

(a) Rewrite this ratio by dividing top and bottom by n^2 to see that it is equal to

$$\frac{6}{\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right)}.$$

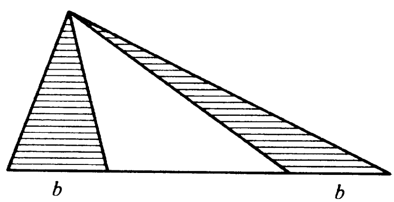
(b) Setting n equal to 100 in this expression (so that $1/n = 0.01$) should make it agree with your answer to problem 10(a). Is your answer to 10(a) equal to $6/(1.01)(2.01)$?

(c) Is your answer to 10(c) equal to $6/(1.001)(2.001)$? And is your answer to 10(d) equal to $6/(1.0001)(2.0001)$?

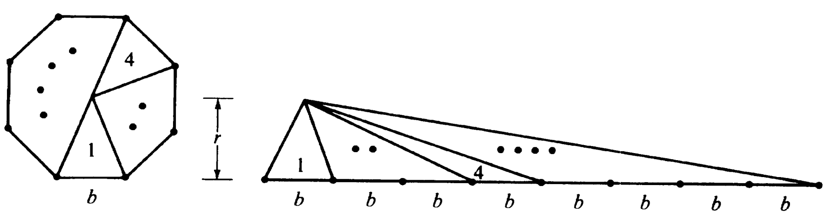
(d) The numbers $6/(1.01)(2.01)$, $6/(1.001)(2.001)$, $6/(1.0001)(2.0001)$, $\cdots$ seem to be getting progressively closer to the exact value of the ratio r_6. Can you now guess what the exact value of r_6 must be? According to Archimedes, Democritus (ca. 460–370 B.C.), the famous philosopher of atomism, was the first to discover the exact value of the ratio r_6. (No one knows how Democritus did it, but perhaps he counted atoms instead of baseballs.)

(e) Did you guess correctly about the rationality of r_6 in your answer to exercise 3.8?

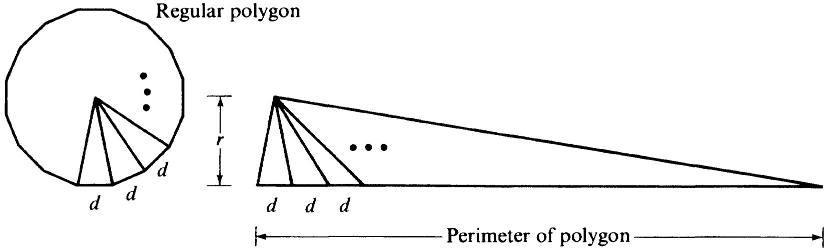
12. Consider the figure below, where the two triangles have a vertex in common and the lengths of their bases are equal. Prove that the triangles have the same area.



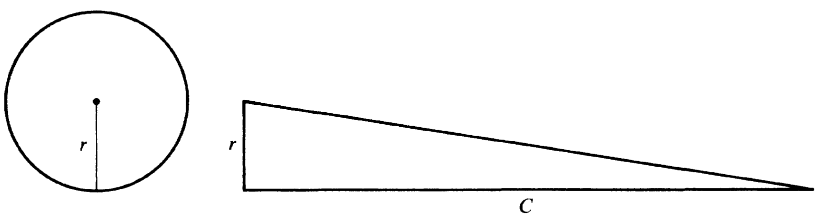
13. Here is a problem that intrigued the Greeks. Given a figure, construct a *triangle* whose area is the same as that of the figure. (For example, given the lunes of problem 18 of Chapter 2, Hippocrates found a triangle of the same size.) Do this for a *regular octagon*. (*Regular* means all sides have the same length and all angles made by adjacent sides are equal.) *Hint*. Stare at the figure below, and use the result of the preceding problem. (Both figures can be thought of as being made up of eight triangles.)



14. By a *regular polygon of n sides* is meant a figure in the plane bounded by n equal sides with n equal angles. (Problem 13 dealt with a regular polygon of 8 sides.) Let r denote the perpendicular distance from the center of a regular polygon to a side. Show that the area of a regular polygon is equal to the area of a triangle whose height is r and whose base is equal in length to the perimeter of the polygon. *Hint.* Stare at the figure below, and use the same reasoning as you did in the preceding problem.



15. (*Here the reader is asked simply to make a guess after considering the evidence.*) Keep in mind that the equality of areas of the regular polygon and the corresponding triangle pictured in problem 14 holds, no matter how many sides the polygon has. This equality of areas holds for a polygon of a billion sides, for instance. Keeping this in mind, stare at the two figures below. One is a circle of radius r, and the other is a triangle of height r whose base is equal in length to the circumference of the circle.



Now make a guess as to which of the following is true:
(a) The area of the circle exceeds the area of the triangle.
(b) The area of the triangle exceeds that of the circle.
(c) The area of the triangle equals the area of the circle.

16. (*For more ambitious students*) the amount and the type of reasoning which constitute an irrefutable argument in mathematics have never been fixed. Some things that the seventeenth century took as *obvious* (i.e., requiring no proof), the twentieth century and also the ancient Greeks accepted only after a careful demonstration from basic principles had been given. If you believe that the statement in part (c) of problem 15 is "obviously" true, then you are in the good company of some of the keenest minds of the seventeenth cen-

tury. They would reason that equality between areas of polygons and circles carries over "in the limit", a circle being regarded as the "geometric" limit of the polygons that approximate it ever more closely.

On the other hand, you may feel that statement (c) in problem 15 requires a clear proof because you have only made an "educated guess" that it is true. If so, then you are at home with Archimedes and with most twentieth-century mathematicians who would think so too. Archimedes proved 15(c) by showing that 15(a) leads to a contradiction, as does 15(b). Can you? (See the appendix on Archimedes for his proof.)

17. Write a paragraph explaining clearly in your own words how the truth of statement (c) in problem 15 leads us quickly to a proof of the familiar formulas $A = \pi r^2$ and $C = 2\pi r$.

18. The formula $A = \pi r^2$ can be thought of as defining A as a function of r. Plot a few points on the curve determined by this function. What is its domain? What is its range?

    Similarly, the formula $C = 2\pi r$ can be thought of as defining C as a function of r. What is its domain? What is its range? Plot a few points on the curve determined by this function. Is the "curve" really a straight line?

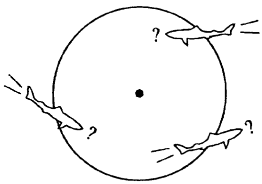
19. Consider a conic section whose focus is at (4,2) and whose directrix is the horizontal line with equation $y = -2$. Write an equation for the conic section if
    (a) it is a parabola.
    (b) it is an ellipse whose eccentricity is 1/2.
    (c) it is a hyperbola whose eccentricity is 2.

20. Define the terms *ellipsis*, *parable*, *hyperbole*, *ellipse*, *parabola*, *hyperbola*, and explain how the latter three terms from geometry are related to their counterparts in rhetoric.

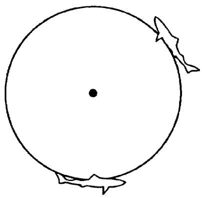
21. (*Messy, but illuminating*) Find an equation of the parabola with focus (600, 179,999.5), whose directrix is given by the line $y = 180{,}000.5$. After simplifying your answer, explain why *Lim: A Fable*, from Chapter 1, Sections 7–8, might be subtitled "a parable of a parabola". Make a guess as to the name of the curve f of Section 2, Chapter 1.

22. (*There is only one parabola!*) Although there are lots of circles, Plato thought of the Circle as a single "form". That is, any two circles are *similar* in the same sense that any two equilateral triangles are similar. Speaking very loosely, we consider two figures in the plane to be similar if a photograph of one, when properly enlarged, can be superimposed precisely upon the other. Thus, for example, a 3–4–5 right triangle is similar to a 6–8–10 right triangle by a magnification ratio of 2. (Not all right triangles are similar, however, because no matter how you magnify a 3–4–5 right triangle, you cannot put it in the form of a 5–12–13 right triangle.)
    (a) Whereas it is obvious that there is only one circle in the sense explained above, it is rarely observed that *there is only one parabola*. Explain why any two parabolas are similar.
    (b) What about ellipses? Are any two ellipses similar? What if they have the same eccentricity?
    (c) What about hyperbolas? Are any two hyperbolas similar?

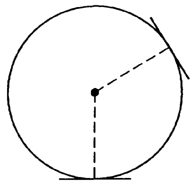
23. (*A whimsical introduction to tangents*) the definition of a *tangent line* to a curve at a point P lying on the curve will be given precisely in the next chapter. The intuitive idea can be given here, however. Suppose you are a shark, shrunk to the size of a very short line segment, and you are swimming in a perfect circle. That is, you are swimming in such a way as to keep your midpoint—the midpoint of that line segment—on a given circle. Make a *reductio ad absurdum* argument showing that the only way you can do this is to position yourself always so that your body—the line segment—is perpendicular to the radial line joining the center of the circle to your midpoint. (*Don't take this problem too seriously. Just give an informal argument suitable for presentation to pool sharks convincing them that if they don't position themselves in this manner they will either veer inside the circle or outside the circle.*)



Stupid sharks



Smart sharks



Tangent lines to a circle

24. (*More whimsy, more tangents*) Do you "get the idea" of a tangent line to a curve? Then consider the graph of the parabola $y = \frac{1}{4}x^2$. Suppose you are a shark—shrunk to the size of a short line segment—swimming so that your midpoint always lies on this parabola.
    (a) How must you position yourself in order not to veer off the curve? *Hint.* Look at the graph of this parabola, which is pictured in Section 4. Suppose your midpoint is located at the point (x, y) that is indicated in this picture. You had better position yourself so that you—the line segment—bisect a certain angle. What angle? Why?
    (b) In essence there is only one circle. Describe how you would draw a tangent line at any point P on a circle, by finishing this sentence: "The tangent line at the point P on a circle with center O is the line through P that is perpendicular to ..."
    (c) In essence there is only one parabola. Describe how you would draw a tangent line at any point P on a parabola by finishing this sentence: "The tangent line at the point P on a parabola with focus F and directrix $\mathbf{D}$ is the line through P that bisects the angle made by the lines ..."
    (d) Use your answer to part (c) to draw the tangent line to the graph of $y = \frac{1}{4}x^2$ (pictured in Section 4) at each of the following points: (2, 1), (0, 0) and $(-2, 1)$.

The Greeks could construct tangent lines to all the conic sections, and to many more curves with *geometric* descriptions. Calculus, as we shall see in the next chapter, provides a simple method of quickly finding tangent lines to the much more general class of curves arising in analytic geometry through *algebraic* descriptions.

25. Write a paragraph explaining clearly why the modern terminology in inequality (11) of Section 6 of Chapter 2 says much the same thing as the ancient language of the following statement (12) that avoids abbreviative symbolism.

26. (*To follow up on the preceding question*) Some historians attribute the eventual decline of the Greek mathematical tradition to the Greeks' failure to develop abbreviative symbolism. Others say the fairly sudden and steep decline, which begins shortly after 200 B.C., is principally due to the Greeks' overemphasizing geometry and thereby failing to embrace the numerical methods that are made almost second nature by the modern use of decimal expansions, which were not widely adopted before the seventeenth century. Still others point to a "horror of infinity" as the main cause, saying the Greeks were so wary of encountering paradoxes when they approached the infinite that they never did so without an exruciatingly rigorous circumspection that would intimidate most students. Finally, many historians point to the "cold breath of Rome" as the reason why the flame of mathematics died, alluding to the fact that during their long period of domination the "practical" Romans showed no interest in promoting research that was not immediately "useful". What do you think about the relative importance of these reasons? Rank these four reasons in order of their importance.

27. (*For poets: a long question with a short answer*) Here is a poem by Walt Whitman (1819–1892):

    When I heard the learn'd astronomer,
    When the proofs, the figures, were ranged in columns before me,
    When I was shown the charts and diagrams, to add, divide, and measure
      them,
    When I sitting heard the astronomer where he lectured
      with much applause in the lecture-room,
    How soon unaccountable I became tired and sick,
    Till rising and gliding out I wander'd off by myself,
    In the mystical moist night-air, and from time to time,
    Look'd up in perfect silence at the stars.

    A different perception of beauty in nature is found in the following words of the French mathematician Henri Poincaré (1854–1912):

    > The scientist does not study nature because it is useful: he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and if nature were not worth knowing, life would not be worth living. Of course I do not here speak of that beauty that strikes the senses, the beauty of qualities and appearances; not that I undervalue such beauty, far from it, but it has nothing to do with science; I mean that profounder beauty which comes from the harmonious order of the parts, and which a pure intelligence can grasp.

    Poincaré writes in prose, not in poetry, and his words are not so well known as Whitman's. Help to rectify this by expressing Poincaré's sentiments in a short poem of your own composition—the shorter, the better. Can you do any better than the two lines of Plato that are quoted in Section 1?

28. (*A long question with an even shorter answer*) Consider the following quotation from Alfred North Whitehead, on Roman civilization:

    > Rome itself stands for the impress of organization and unity upon diverse fermenting elements. Roman law embodies the secret of Roman greatness in

> its Stoic respect for intimate rights of human nature within an iron framework of empire. Europe is always flying apart because of the diverse explosive character of its inheritance, and coming together because it can never shake off that impress of unity it has received from Rome. The history of Europe is the history of Rome curbing the Hebrew and the Greek, with their various impulses of religion, and of science, and of art, and of quest for material comfort, and of lust of domination, which are all at daggers drawn with each other. The vision of Rome is the vision of the unity of civilization.

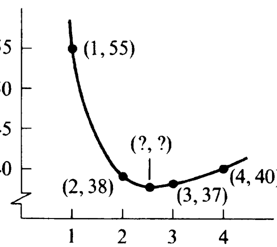
What did the Romans do for mathematics?

29. *(A sophistry)* Here is the sort of thing Plato abhorred, a slick argument calculated to deceive and demean the uneducated. Find the major fallacy in this sophistry.

> Don't you know, you silly mortal, that you won't live more than a hundred years or so at most? And don't you see how foolish and futile it is for you to think that you might do in your brief lifetime what only a god could do? Look at you, trying to find the least number in the range of costs given by $C = 7L + 48/L$, where L can take on infinitely many values. Don't you see that you, a pitiful mortal, can never know all the infinitely many possible costs? Only a god could know!
>
> How do you expect to find the least cost if you can't find all the costs? Someone could always come after you and find some new value of L that has a lower cost C than any cost you might have found during your pitifully brief life. How foolish you are to think you are like a god! Your hubris could fill the ocean. Why don't you go drown yourself?

4 CHAPTER Sherlock Holmes Meets Pierre de Fermat

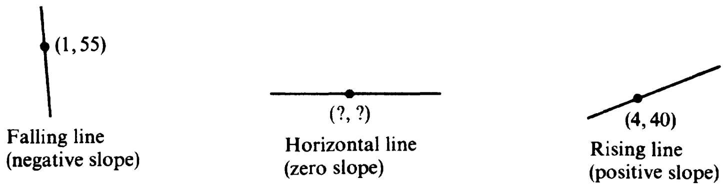
Given a curve, such as the one below, how can one locate its lowest point? This problem arose naturally in Chapter 1, along with the analogous problem of finding the highest point on a curve. Both problems can be solved by the same method, to which we now turn.



§1. Rising and Falling Lines

We must agree first how to use the words *rising* and *falling*, for there is danger of misunderstanding. Is the curve above *rising* or *falling* as it passes through the point (1, 55)? The answer depends upon whether one thinks of the curve as being traced out from left to right or in the reverse direction. So that we all speak the same language, let us agree to think of any function's curve as being traced out *from left to right* (or *from west to east*, if you prefer). The curve above is then falling as it passes through (1, 55), and rising as it passes through (4, 40).

Before going further, we had better mention the simplest curves of all: straight lines. Below are pictured *falling, horizontal,* and *rising* lines.



These three lines, if superimposed upon the curve pictured on the preceding page, may give the reader a hint as to the method we shall develop. At each point P on a curve, we shall seek *the line through P that most closely approximates the curve near P.* This line will be called the **tangent line** to the curve at P. The discussion of tangent lines begins in Section 3, and most of this chapter is devoted to their study.

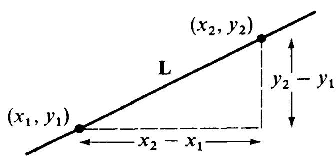
What does the study of tangent lines to curves have to do with the problem stated in the first sentence of this chapter? Look again at the curve on the preceding page. It is pretty clear, is it not, that the lowest point occurs where the tangent line is *horizontal,* that is, where the tangent line slopes neither up nor down.

We must give a precise meaning to the word *slope.*

Definition

The **slope** of the line joining (x_1, y_1) and (x_2, y_2) is given by

$$\frac{y_2 - y_1}{x_2 - x_1}, \quad \text{provided } x_1 \neq x_2.$$



For example, the slope of the line joining (4, 6) and (5, −3) is given by

$$\frac{-3 - 6}{5 - 4} = -9.$$

The slope of a line is a number that measures how fast the line *rises* (or, when the slope is negative, how fast the line *falls*). If **L** is the line joining (x_1, y_1) and (x_2, y_2), then

$$\frac{y_2 - y_1}{x_2 - x_1} = \text{Slope } \mathbf{L},$$

so that we have

$$y_2 - y_1 = (\text{Slope } \mathbf{L})(x_2 - x_1). \tag{1}$$

The relation expressed in equation (1) will be useful later in finding an equation of the line **L**.

Exercises

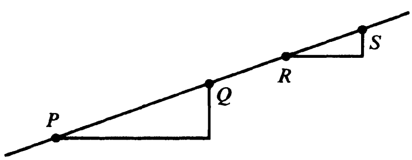
1.1. Find the slope of the line joining $(1, 2)$ and $(3, 7)$, and draw a picture of this line. *Partial answer*: Slope is $5/2$.

1.2. Find the slope and draw a picture of the line joining $(1, 2)$ and $(3, -2)$. Is this line *rising* or *falling*?

1.3. Find the slope and draw a picture of the line joining $(1, 2)$ and $(5, 2)$. Is this line *rising* or *falling*?

1.4. Find the slope and draw a picture of the line joining $(1, 2)$ and $(1, 5)$. *Partial answer*: Slope is undefined.

1.5. Using geometry, show that the slope of a line is independent of which pair of points is chosen to calculate the slope. That is, in the figure below, show that the slope from P to Q is equal to the slope from R to S. *Hint*. The slope is simply a ratio of two sides of a triangle. Prove that the triangles are similar.



1.6. (a) Find the slope of the line **L** joining $(1, 2)$ and $(3, 5)$.
(b) Find the slope of the line joining $(1, 2)$ and $(300, 450)$.
(c) Using your answers to (a) and (b), decide whether the point $(300, 450)$ lies *above, on*, or *below* the line **L** joining $(1, 2)$ and $(3, 5)$.
(d) Is the point $(301, 452)$ on this line **L**? How do you know?

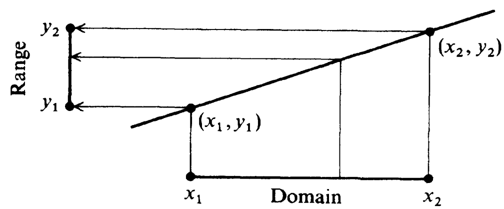
§2. Linear Functions

It is easy to see, as illustrated in exercises 1.1–1.3, that a line is

rising if its slope is *positive*,
falling if its slope is *negative*,
horizontal if its slope is *zero*.

Some curves (and we shall understand a line to be an especially simple kind of curve) determine functions, and some do not. Any *non-vertical* line does determine a function. (Why?) Such a function is called a **linear** function.

The slope of a line tells us something about the linear function it determines. It tells us how much the function "stretches". What does this mean? Look at the figure below, where a line of positive slope is pictured, and consider the function determined by this line. If the domain is the interval from x_1 to x_2 and if the corresponding range extends from y_1 to y_2, then by what factor is the domain stretched as it is sent into the range?

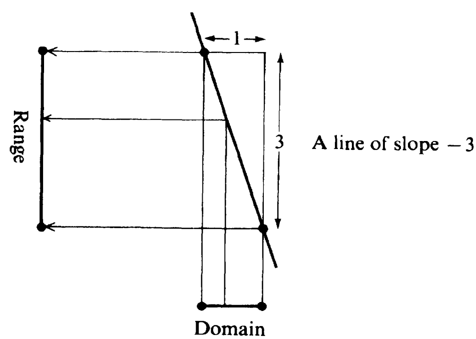


From the figure, the length of the range is $y_2 - y_1$, and the length of the domain is $x_2 - x_1$. Equation (1) above thus says:

$$\text{Length of range} = (\text{slope of line})(\text{length of domain}).$$

The slope of the line thus gives the factor by which a linear function stretches lengths. A line of slope 3, for instance, determines a linear function that sends any interval into an interval three times as long. A line of slope 3, considered as a function, has a "stretching factor" of 3.

The preceding discussion applies to lines of *positive* slope. Suppose the slope of a line is negative, say -3. Then the linear function determined by the line still has a stretching factor of 3, but intervals in the domain are "flipped upside down" before they land in the range.



The notion of slope makes it easy to write or to recognize equations of nonvertical lines. The following exercises illustrate this.

Exercises

2.1. Translate into words the algebraic equation $y - 2 = \frac{3}{2}(x - 1)$. *Answer*: By equation (1), this says "(x, y) lies on the line of slope $3/2$ passing through $(1, 2)$." [This is an equation, then, of the line described in exercise 1.6(a).]

2.2. Translate into words the algebraic equation $y - 4 = 3(x - 2)$, and sketch the line determined by this equation.

2.3. Translate into words the algebraic equation $y + 4 = -2(x - 3)$, and sketch the line determined. *Hint*. First rewrite the equation as $y - (-4) = -2(x - 3)$, then use equation (1).

2.4. Translate into words each of the following equations, and sketch the line determined.
   (a) $y = 2x + 4$. *Hint*. Rewrite as $y - 4 = 2(x - 0)$.
   (b) $3x + 4y = 6$. *Hint*. First solve for y. Then proceed as you did in part (a).
   (c) $y = \pi x + \sqrt{2}$.
   (d) $y = 5$. *Hint (if needed)*. Rewrite as $y - 5 = 0(x - 0)$, and use equation (1).

2.5. The **slope-intercept** form of the equation of a line is

$$y = bx + c,$$

where b and c are constants.
   (a) Rewrite this equation as $y - c = b(x - 0)$. Find the slope of the line determined by this equation, and find both coordinates of the point where the line meets the y-axis.
   (b) Describe the curve determined by the function given by $f(x) = 3x + 5$. *Answer*: The graph of f is a line passing through $(0, 5)$ with slope 3, since the algebraic rule $3x + 5$ is in slope-intercept form.
   (c) Describe the curve determined by each of the following rules:
      (i) $-2x - 5$.
      (ii) $x - 1$.
      (iii) $5 - x$.

2.6. Find an algebraic equation for the line of slope 3 passing through $(0, \pi)$. *Answer*: By equation (1), a point (x, y) lies on this line if and only if $y - \pi = 3(x - 0)$, or (simplifying) $y = 3x + \pi$.

2.7. Find an algebraic equation for the line of slope 3 passing through $(\pi, 0)$.

2.8. (a) Find the slope of the line joining $(4, 6)$ and $(3, 8)$.
   (b) Using your answer to (a), find an equation of the line joining $(4, 6)$ and $(3, 8)$.

2.9. Find an equation of the line joining

(a) $(0, 0)$ and $(1, -2)$.
(b) $(3, 4)$ and $(4, 7)$. *Answer*: $y = 3x - 5$.
(c) $(3, 4)$ and $(7, 4)$.
(d) $(3, 4)$ and $(3, 7)$. *Hint*. This is made simple, not hard, by the fact that the slope is undefined. Use common sense.

§3. The Principle of Elimination

In the preceding section we have made an essentially complete investigation of the simplest kind of function. We have learned that any function given by a rule of the form

$$bx + c$$

is a linear function. Its graph is a line of slope b passing through the point $(0, c)$ on the vertical axis.

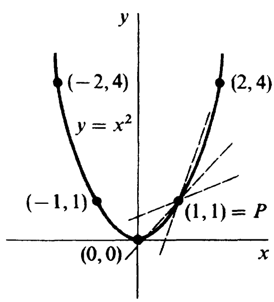
The next simplest kind of function is a *quadratic* function, arising when a linear expression $bx + c$ is modified by a term involving a square: A **quadratic function** is given by an algebraic rule of the form

$$ax^2 + bx + c, \quad \textit{where } a \neq 0.$$

The behavior of quadratic functions is not hard to study. To investigate that behavior, *and to learn at the same time how to find tangent lines to curves*, let us consider the simplest quadratic function of all. This is, of course, the **squaring function** given by

$$f(x) = x^2.$$

Plotting a lot of points on the graph of the squaring function shows that it looks something like this:



We are ready to move toward attacking the problem stated in the first sentence of this chapter. We have already hinted that the solution of that problem involves the study of tangent lines to curves. Our task now is to figure out exactly what a tangent line is. So far, we have only made the (rather vague) statement that the tangent line to a curve at a point P is the line through P that most closely resembles the curve near P. With this meager thread to hold on to, *how can one determine the slope of the tangent line to the squaring function at the point* $P = (1, 1)$?

This is a challenging question, even for the keen mind of a master sleuth. Let us therefore enlist the aid of the great detective:

Sherlock Holmes's Principle
When you have eliminated the impossible, whatever remains, however improbable, must be the truth.

The answer to a question is among what remains after wrong answers have been set aside. By this principle of elimination, the tangent line is the line left when all "nontangent" lines have been discarded. Will this be helpful to us? We shall see. Let us first look at some exercises to test whether this principle of elimination is well understood.

Exercises

Apply Sherlock Holmes's principle to each of the following situations.

3.1. Winnie the Pooh's honey is gone. Everyone but Tigger has a valid alibi that proves his innocence. *Answer*: By Holmes's principle, Tigger stole the honey, *provided it was stolen.*

3.2. A survey shows that Peter Pan is a citizen of no country questioned in the survey, and England is the only country not questioned. *Answer*: By Holmes's principle, Peter Pan is a citizen of England, *provided that Peter Pan is a citizen of some country.*

3.3. The county seat of Yoknapatawpha County, Mississippi, is none other than the city of Jefferson.

3.4. *1984* $+ h$ is not the title of a famous book, if h is not equal to 0. *Answer*: No famous book has a numerical title, except possibly *1984*

3.5. If $h \neq 0$, then the area of the unit circle is not $\pi + h$. *Answer*: The area of the unit circle is none other than π.

3.6. If $h \neq 0$, then $h + 2$ is not the answer to a certain problem in arithmetic.

3.7. If $h \neq 0$, then $(h^2 + 2h)/h$ is not the answer to a certain problem. *Hint.* $(h^2 + 2h)/h = h + 2$ if $h \neq 0$.

3.8. If $h \neq 0$, then $(h^2 + 4h)/h$ is not the answer to a certain problem. *Answer*: By Holmes's principle, the answer must be 4, *provided the problem has an answer* (and provided the answer is a number).

3.9. If $h \neq 0$, then $(h^2 + 9h)/h$ is not the answer to a certain problem.

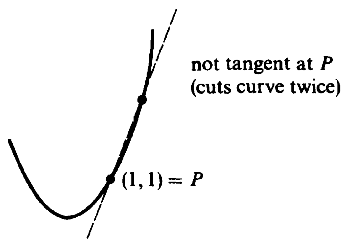
3.10. Is Holmes's principle a rediscovery of an ancient Greek method?

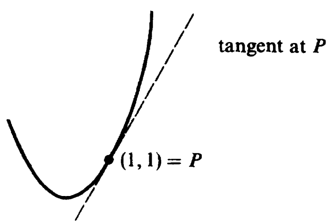
§4. The Slope of a Tangent Line

We are prepared to begin our detective work. To employ the principle of Sherlock Holmes, we must attain skill at finding wrong answers, in order to eliminate them. Let us recall the question:

What is the slope of the tangent line to the curve $y = x^2$ at the point $P = (1, 1)$? (2)

How can we get a wrong answer to this question? Look once again at the graph of the squaring function near P. A line through P that cuts the curve twice will *not* be tangent at P, it would seem. The tangent line at P will touch the curve only at P.

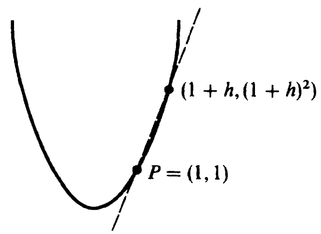




Now we have a clue. To obtain a *wrong* answer to question (2), we need only find the slope of a line joining P to another point on the graph of the squaring function. This graph consists of each point in the plane whose second coordinate is the square of its first coordinate. Another point on the curve, then, is $(1 + h, (1 + h)^2)$ *if h is not equal to zero*. (If h is 0, this "other" point would coincide with P.) The slope of the nontangent line joining $(1, 1)$ and $(1 + h, (1 + h)^2)$ is given by

$$\frac{(1+h)^2 - 1}{1 + h - 1} = \frac{1 + 2h + h^2 - 1}{h} = \frac{2h + h^2}{h}.$$

We now know a host of wrong answers to question (2), for if $h \neq 0$, then $(2h + h^2)/h$ is the slope of a line that is *not* tangent at P. Note that this expression simplifies to $2 + h$.



This line, not tangent at P if $h \neq 0$, has a slope equal to $2+h$

What is the answer to question (2), now that we know that $2+h$ is not the answer, if $h \neq 0$. The only number not eliminated is 2. By Sherlock Holmes's principle, the answer to question (2) must be 2, provided the question has an answer. That is,

$$\text{the slope of the tangent line to } y = x^2 \text{ at } (1,1) \text{ is } 2, \tag{3}$$

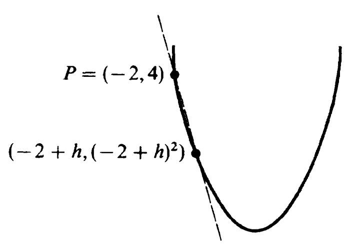
provided the curve $y = x^2$ has a tangent line at $(1,1)$. Elementary, dear Watson! □

Holmes's method illustrates the curious fact that it is possible to get the right answer by first considering how to get wrong answers. Let us try another question.

$$\text{What is the slope of the tangent line to the curve } y = x^2 \text{ at the point } (-2,4)? \tag{4}$$

Let us consider how to get wrong answers to question (4). A wrong answer is the slope of the line joining $(-2,4)$ and $(-2+h,(-2+h)^2)$ if $h \neq 0$. The slope of this nontangent (or *secant*, as a line cutting a curve twice is often called) is given by

$$\begin{aligned}\frac{(-2+h)^2-4}{-2+h+2} &= \frac{4-4h+h^2-4}{h}\\ &= \frac{-4h+h^2}{h}\\ &= -4+h \quad \text{if } h \neq 0.\end{aligned}$$



This secant has a slope equal to $-4+h$

What is the answer to question (4), now that we know $-4+h$ is not the answer if $h \neq 0$? The only number not eliminated is -4. By Holmes's principle, the answer must be -4, provided there is an answer. That is,

$$\begin{array}{l}\text{the slope of the tangent}\\ \text{line to } y = x^2 \text{ at } (-2,4) \text{ is } -4,\end{array} \tag{5}$$

provided the curve $y = x^2$ has a tangent line at $(-2, 4)$.

Exercises

Apply the principle of elimination to each of the following.

4.1. What is the slope of the tangent line to the curve $y = x^2$ at $(0, 0)$? *Answer*: It is 0, provided there is a tangent line.

4.2. What is the slope of the tangent line to the curve $y = x^2$ at the point $(2, 4)$?

4.3. What is the slope of the tangent line to $y = x^2$ at (π, π^2)? *Answer*: 2π, if there is a tangent line.

4.4. What is the slope of the tangent line to the curve $y = x^2 + 3$ at the point $(1, 4)$?

4.5. What is the slope of the tangent line to the curve $y = x^2 + 3x$ at the point $(1, 4)$? *Hint*. A wrong answer to this question is given by

$$\frac{(1+h)^2 + 3(1+h) - 4}{h} = 5 + h \quad \text{if } h \neq 0.$$

4.6. What is the slope of the tangent line to the curve $y = x^2 + 3x + 2$ at the point $(1, 6)$? *Answer*: 5, if there is one.

4.7. What is the slope of the tangent line to the curve $y = x^2 + 3x + 2$ at the point $(\pi, \pi^2 + 3\pi + 2)$?

§5. Fermat's Method and the Derivative

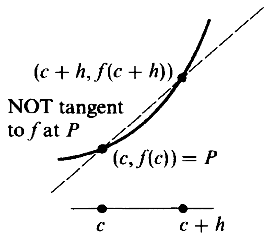
As clever as Holmes's method is, it has serious drawbacks, as illustrated in problem 17 at the end of this chapter. One worrisome thing about this method is that things are left hanging a bit at the end. How do we know whether a curve has a tangent line at a certain point? What is needed is a clear definition.

Pierre de Fermat pointed the way toward using the notion of *limit* to invent a workable definition of the *slope of a tangent line to a curve*. It is only a slight modification of the method we have just employed, but by it the drawbacks to Holmes's method are removed.

Fermat described the following method of finding the slope of the tangent line to a curve f at a given point $P = (c, f(c))$ on the curve. First find the slope of the line joining $(c, f(c))$ and $(c+h, f(c+h))$, where $h \neq 0$. Although this slope, which is given by

$$\frac{f(c+h) - f(c)}{h}, \tag{6}$$

is likely *not* the desired slope of the tangent line, it clearly approximates the desired slope as h is taken nearer to zero. It is natural, then, for us to *define* the slope of the tangent line at $(c, f(c))$ to be the number (if there is one) that expression (6) is trying to become as h approaches zero.



This nontangent line has slope given by expression (6)

Definition

The **slope of the tangent line** to the curve f at the point $(c, f(c))$ is defined to be

$$\operatorname*{Limit}_{h \to 0} \frac{f(c+h) - f(c)}{h}.$$

Fermat's idea is simple yet subtle. The "right answer" is the limiting value of wrong answers that approximate it ever so closely. Here are several examples to illustrate Fermat's method.

EXAMPLE 1

Find the slope of the tangent line to the curve $y = x^2$ at the point $(1, 1)$.

Here the function is given by $f(x) = x^2$, and the point P is $(1, f(1))$. According to Fermat's method, the slope of the tangent line at $(1, f(1))$ is given by

$$\begin{aligned}\operatorname*{Limit}_{h \to 0} \frac{f(1+h) - f(1)}{h} &= \operatorname*{Limit}_{h \to 0} \frac{(1+h)^2 - 1}{h} \\ &= \operatorname*{Limit}_{h \to 0} \frac{h^2 + 2h}{h} \\ &= \operatorname*{Limit}_{h \to 0}(h + 2) \\ &= 2.\end{aligned}$$

Note that nothing is left hanging at the end. Since the limit exists, there is a tangent line, and its slope is equal to that limit. By Fermat's definition, the existence of a tangent line is tantamount to the existence of the limit of expression (6). □

EXAMPLE 2
Find the slope of the tangent line to the curve $y = x^2$ at the point (π, π^2).

Here we have the squaring function again, given by $f(x) = x^2$, and the point P is $(\pi, f(\pi))$. By Fermat's method, the slope of the tangent line is

$$
\begin{aligned}
\operatorname*{Limit}_{h \to 0} \frac{f(\pi + h) - f(\pi)}{h} &= \operatorname*{Limit}_{h \to 0} \frac{(\pi + h)^2 - \pi^2}{h} \\
&= \operatorname*{Limit}_{h \to 0} \frac{\pi^2 + 2\pi h + h^2 - \pi^2}{h} \\
&= \operatorname*{Limit}_{h \to 0} \frac{2\pi h + h^2}{h} \\
&= \operatorname*{Limit}_{h \to 0} (2\pi + h) \\
&= 2\pi.
\end{aligned}
$$
□

EXAMPLE 3
Find the slope of the tangent line to the curve $y = x^2$ at the point (x, x^2).

This is so similar to Example 2 that the reader can probably guess the answer. The answer is $2x$, for the same reason that the answer to the preceding example is 2π. This is seen by a calculation identical to that of Example 2, with x replacing π:

$$
\begin{aligned}
\operatorname*{Limit}_{h \to 0} \frac{f(x + h) - f(x)}{h} &= \operatorname*{Limit}_{h \to 0} \frac{(x + h)^2 - x^2}{h} \\
&\ \ \vdots \\
&= \operatorname*{Limit}_{h \to 0} (2x + h) = 2x.
\end{aligned}
$$

(The reader is asked to fill in the missing steps in this calculation.) □

The work of Examples 1–3 may be summarized in a table:

x	$\xrightarrow{f}$ y	$\xrightarrow{f'}$ Slope of tangent line at (x, y)
1	1	2
π	π^2	2π
x	x^2	$2x$
-1	1	?

If we recall the definition of a function in terms of a pair of columns, then we see that the *first* and *third* columns above determine a *new* function. This new function, derived from the original function f, will be denoted by f' and called the **derivative** of f. From the third line of the table above, we see that the rule determining f' is simply the "doubling" rule, sending x to $2x$. That is, we see that

$$\text{if } f(x) = x^2, \quad \text{then } f'(x) = 2x. \tag{7}$$

Or, in words, *the derivative of the squaring function is the doubling function.*

EXAMPLE 4

Find the slope of the tangent line to the curve $y = x^2$ at the point $(-1, 1)$.

Now there is no need to go back to Fermat's method, because *the derivative gives us the general slope-predicting rule.* All that is asked here is that the question mark in the preceding table be filled in appropriately, and that is now easy. The answer is $f'(-1)$, which is equal to -2, since f' is the doubling function. □

EXAMPLE 5

Find the slope of the tangent line to the curve $y = x^2$ at the point $(4, 16)$.

The answer is $f'(4)$, which is equal to 8. □

The function f sends x into y. What does the function f' do? It is convenient to let y' stand for the long phrase "Slope of the tangent line at (x, y)". Then the function f' sends x into y'. Thus, equation (7) says exactly the same thing as

$$\text{if } y = x^2, \quad \text{then } y' = 2x.$$

Exercises

5.1. Fill in the missing steps in Example 3.

5.2. Find the slope of the tangent line to the curve $y = x^2$ at the point $(3, 9)$
   (a) by using Fermat's method, going through all the steps to find the limit of $((3 + h)^2 - 9)/h$ as h approaches 0.
   (b) by using the shortcut method of Example 5, knowing that the derivative of the squaring function is the doubling function.

5.3. What does statement (5) of Section 4 say in terms of y'? *Answer*: It says, "Given $y = x^2$, then y' is -4 when x is -2."

5.4. What does statement (3) of Section 4 say about y'?

5.5. What does the answer to exercise 4.5 say about y'? *Answer*: It says, "Given $y = x^2 + 3x$, then y' is 5 when x is 1 (assuming there is a tangent line)."

5.6. What does the answer to exercise 4.6 say about y'?

§6. The Interplay between a Function and Its Derivative

The derivative f' is useful for many reasons. One reason (we shall see others later) is that f' gives information about the behavior of the original function f. To illustrate this, let us continue to study the squaring function f, whose derivative, we have seen, is the doubling function.

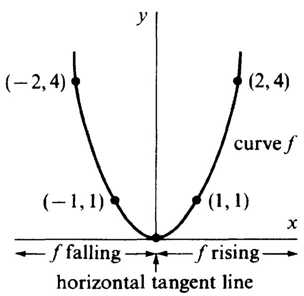
First, note that f' is just as "good" a function as f. The equation $y' = f'(x)$ determines a curve too! In this case the rule for f' is the *linear* expression

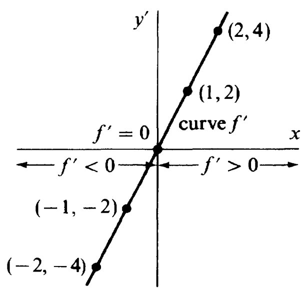
$$2x,$$

which we should recognize immediately to be pictured as a line of slope 2, passing through the origin $(0, 0)$ in the x–y' plane.

x	y	y'
1	1	2
x	x^2	$2x$
π	π^2	2π
s	s^2	$2s$

To see the interplay between f and f', it is convenient to picture the curve f' on a separate coordinate system (the x–y' plane) and to compare it with the curve f in the x–y plane.





At a point where the curve f is *falling*, the tangent line must have a *negative* slope. Hence, if f is falling at the point $(c, f(c))$, then $f'(c)$ must be negative. Similarly, when f is *rising*, then f' must be *positive*. And when the curve f has a *horizontal tangent line*, then f' must be zero.

Exercises

6.1. Find both coordinates of a point on the curve $y = x^2$ where the slope of the tangent line is 3. *Answer*: We are required to fill in the question marks correctly in the following table:

x	y	y'
?	?	3
x	x^2	$2x$

When $y' = 3$, we have $2x = 3$; so $x = \frac{3}{2}$. When $x = \frac{3}{2}$, $y = \left(\frac{3}{2}\right)^2 = \frac{9}{4}$. Therefore, the slope of the tangent line is 3 at the point $\left(\frac{3}{2}, \frac{9}{4}\right)$.

6.2. Find both coordinates of a point on the curve $y = x^2$ where the slope of the tangent line is
(a) -2.
(b) 0.
(c) 10.
(d) 5.

6.3. Suppose the slope of the tangent line to a curve is -1 at a certain point. Is the curve *rising* or *falling* as it passes through that point?

6.4. Find an equation of the tangent line to the curve $y = x^2$ at the point $(3, 9)$. *Answer*: When x is 3, y' is 6; so the slope of the tangent line is 6. An equation of the line of slope 6 through $(3, 9)$ is $y - 9 = 6(x - 3)$.

6.5. Find an equation of the tangent line to the curve $y = x^2$ at the point
(a) $(1, 1)$.
(b) $(-1, 1)$.
(c) (π, π^2).
Answer: (c) $y - \pi^2 = 2\pi(x - \pi)$.

§7. Solving Optimization Problems with Derivatives

Compare the equation $y = x^2$ with the equation $A = s^2$. Both equations determine *the same function*. Why? Because both equations define exactly the same rule, the *squaring rule*. The curve in the x-y plane of the equation $y = x^2$ is identical with the curve in the s-A plane of the equation $A = s^2$. Since the derivative of the squaring function is the doubling function, it is clear that

$$\text{if } A = s^2, \quad \text{then } A' = 2s. \tag{8}$$

By the same token, we know, for example, that

$$\text{if } y = L^2, \quad \text{then } y' = 2L.$$

Changing only the *names* of the variables doesn't alter the function, or its derivative, at all.

Let us find another quadratic function to play with. In Example 2 of Chapter 1 we encountered the personable function g given by the quadratic rule

$$-\frac{s^2}{2} + 600s.$$

What is the rule for g', the derivative of g?

s	A	A'
400	160,000	
700	175,000	
s	$-\frac{1}{2}s^2 + 600s$	?

Can you *guess* the rule for g', before we work it out below? *There is nothing wrong with guessing.* Consider the facts. Statement (8) tells us that from the expression s^2 in the second column we derive the expression $2s$ in the third column. On the basis of this, what would you guess to be derived from the expression $-\frac{1}{2}s^2$? As for the expression $600s$, that is easy. This is just a linear expression of slope 600, leading one to expect that from the expression $600s$ in the second column we would derive the expression 600 in the third. From these facts, what would you guess:

$$\text{If } g(s) = -\frac{s^2}{2} + 600s, \quad \text{then } g'(s) = ?$$

To verify your guess, go back to the definition of the derived function. By definition, $g'(s)$ is the slope of the tangent line to the curve g at the point $(s, g(s))$. Using Fermat's method to calculate that slope, we have

$$\begin{aligned}
g'(s) &= \underset{h\to 0}{\text{Limit}} \frac{g(s+h) - g(s)}{h} \\
&= \underset{h\to 0}{\text{Limit}} \frac{-\frac{1}{2}(s+h)^2 + 600(s+h) - (-\frac{1}{2}s^2 + 600s)}{h} \\
&= \underset{h\to 0}{\text{Limit}} \frac{-\frac{1}{2}s^2 - sh - \frac{1}{2}h^2 + 600s + 600h + \frac{1}{2}s^2 - 600s}{h} \\
&= \underset{h\to 0}{\text{Limit}} \frac{-sh - \frac{1}{2}h^2 + 600h}{h}
\end{aligned}$$

$$= \underset{h\to 0}{\text{Limit}}\left(-s - \frac{h}{2} + 600\right)$$

$$= -s + 600.$$

Thus we see that $g'(s) = -s + 600$. In other words,

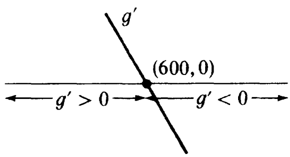
$$\text{if } A = -\frac{s^2}{2} + 600s, \quad \text{then } A' = -s + 600.$$

s	A	A'
?	?	0
s	$-\frac{1}{2}s^2 + 600s$	$-s + 600$

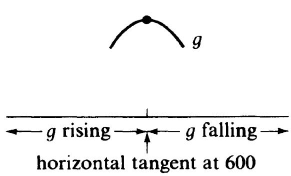
We now have enough information to determine the highest point (?, ?) on the curve g, for this point must occur where the slope of the tangent is zero. This is easy, for $A' = 0$ when

$$-s + 600 = 0,$$
$$s = 600.$$

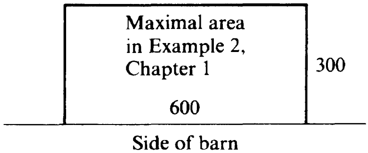
Thus, at the point (600, 180,000), the curve g has a horizontal tangent line. How do we know this is the *highest* point? Look at the derivative. The curve $A' = -s + 600$ is a linear curve of slope -1, and A' is 0 when s is 600. The derivative looks like this:



Therefore the curve g must be rising to the left of 600 and falling to the right of 600. This means that, at $s = 600$, the maximal A is attained.



The optimization problem arising in Example 2, Chapter 1, is now solved. The maximal area is 180,000 square feet, attained when the length s along the barn is 600 feet:



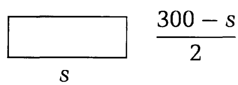
Let us attack a similar problem to show how easy an optimization problem can become when calculus is applied.

EXAMPLE 6

A farmer has 300 meters of fencing to enclose three sides of a rectangular area. The fourth side is bounded by a long barn and requires no fence. What is the largest area she can enclose?

We want to maximize the area A, which varies in terms of the length s along the barn. Letting G denote the function that arises, we have

$$A = G(s) = \text{area (in square meters) of}$$



$$= 150s - \frac{s^2}{2},$$

where the domain is specified by the inequality $0 < s < 300$.

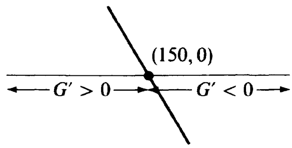
To solve this optimization problem, we must find the highest point (?, ?) on the curve G, which gives the area A as a function of s. Toward this end, we take the derivative:

$$\text{If } A = 150s - \frac{s^2}{2}, \quad \text{then } A' = 150 - s \quad \text{(why?).}$$

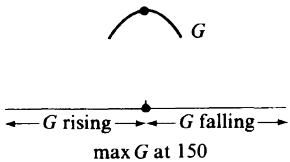
Thus A' is 0 when s is 150, and we have found the point on the curve G where the tangent line is horizontal:

$s \xrightarrow{G}$	A	A'
150	11,250	0
s	$150s - \frac{1}{2}s^2$	$150 - s$

In all likelihood, the point (150, 11,250) is the highest point on the curve G. To *prove* that it is, look at the sign of the derivative on either side of 150. The derivative G' is given by the linear rule $150 - s$, and thus looks like this:



Therefore, the curve G must be rising to the left of 150 and falling to the right. This shows that, at $s = 150$, G attains its maximum.



The maximal area is 11,250 square meters. □

Exercises

7.1. The derivative of the squaring function is the doubling function. The slope of the line $bx + c$ is b. Use these facts and try your hand at quickly *guessing* answers to the following:
   (a) If $y = x^2 + 3x$, what is y'?
   (b) If $y = x^2 - 6x$, what is y'?
   (c) If $A = 4s^2 + 60s$, what is A'?
   (d) If $y = 10x^2 + 4x + 20$, what is y'?
   (e) If $y = ax^2 + bx + c$, what is y'?

7.2. In each of (a) through (e) of exercise 7.1, use Fermat's method to verify the correctness of your guess. *Answer*: (a) Given $y = f(x) = x^2 + 3x$, by Fermat's method we have

$$\begin{aligned} y' = f'(x) &= \operatorname*{Limit}_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \operatorname*{Limit}_{h \to 0} \frac{(x+h)^2 + 3(x+h) - (x^2 + 3x)}{h} \\ &= \operatorname*{Limit}_{h \to 0} \frac{2xh + h^2 + 3h}{h} \\ &= \operatorname*{Limit}_{h \to 0} (2x + h + 3) \\ &= 2x + 3. \end{aligned}$$

7.3. In each of (a) through (e) of exercise 7.1, find both coordinates of the point on the quadratic where the tangent line is horizontal. *Answer*: (a) At the point $(-3/2, -9/4)$, y' is zero.

7.4. Find both coordinates of the highest point on the curve $A = 1200w - 2w^2$, with domain $0 < w < 600$. (This is the function which arose in problem 3 at the end of Chapter 1.)

7.5. A farmer has 4000 feet of fencing to enclose *three* sides of a rectangular area (the fourth side being bounded by a long fence already standing). Find the largest area that can be enclosed, and specify the dimensions that should be used to attain maximal area.

7.6. A farmer has 4000 feet of fencing to enclose *four* sides of a rectangular area. What dimensions should be used to maximize the area enclosed?

7.7. By working through the following steps in turn, *find a pair of positive numbers whose sum is 10 and whose product is as large as possible.*

(a) We want to maximize their product. Let P denote their product. What is P if the first number is 2? (First find the second number, using the fact that the sum of the two numbers must be 10.)

(b) What is P if the first number is π?

(c) What is P if the first number is x?

(d) Your answer to (c) yields a quadratic rule giving P as a function of x. What is P'?

(e) What is the domain of the function you found in part (c)? (Remember that *both* numbers must be positive.)

(f) Find both coordinates of the highest point on the graph of the quadratic function of part (c).

(g) Answer the question of problem 7.7 with a complete sentence.

7.8. Express the number 10 as the sum of two positive numbers in such a way that the sum of the *square* of the first and *three times* the second is as *small* as possible. *Hint*. This is similar to exercise 7.7.

7.9. Work through the following steps in turn, in order to answer the question at the end.

(a) In the x-y plane, draw the line $y = 3x + 2$. Also indicate the position of the point $(4, 0)$.

(b) Find the *square* of the distance between the point $(4, 0)$ and the point on the line $y = 3x + 2$ whose first coordinate is π. (First find the second coordinate, then find the square of the distance by the Pythagorean theorem.)

(c) Find the square of the distance between the point $(4, 0)$ and the point on the line $y = 3x + 2$ whose first coordinate is x. *Answer*: $10x^2 + 4x + 20$.

(d) The rule written down in the answer to part (c) is the quadratic function in exercise 7.3(d). Find the value of x that yields the minimum of this function.

(e) Find both coordinates of the point on the line $y = 3x + 2$ that is closest to the point $(4, 0)$. *Answer*: $(-1/5, 7/5)$.

7.10. Find both coordinates of the point on the line $y = 5 - 2x$ that is closest to the point $(0, 0)$.

§8. Definition of the Derivative

Calculus relies greatly upon derivatives. We therefore seek rules enabling us to write down quickly the derivative of any function we might meet. We have already found such a rule for writing down the derivative of any quadratic function:

$$\text{If } y = ax^2 + bx + c, \quad \text{then } y' = 2ax + b.$$

(Another way of expressing the same thing is, "If $f(x) = ax^2 + bx + c$, then $f'(x) = 2ax + b$.") By virtue of this simple rule, there is no need to go through all the details of Fermat's method in order to find the derivative of a quadratic. In the next chapter, however, we shall meet more complicated algebraic functions, such as are given by the rules $1/x$ (the *reciprocal function*), x^3 (the *cubing function*), $\sqrt{x}$ (the *square root function*), etc. To find their derivatives, we must be clear about the definition of the derivative.

If f is any function, the rule defining its derivative f' is given below. The derivative is defined so that, at a point x, the derivative f' gives the slope of the tangent line to the curve f at the point $(x, f(x))$. Since Fermat's method gives this slope, we have the following definition.

Definition
Given a function f, and a point x in its domain, the **derivative** f' is defined by the rule

$$f'(x) = \operatorname*{Limit}_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

Note that the definition of the derivative incorporates all three basic notions: *variable, function, limit.*

To calculate f' directly from this definition is sometimes tedious, requiring several lines of computation. However, as in Section 7, it is possible to guess and to verify shortcut rules of finding derivatives. This will be the business of Chapter 5. To understand that chapter, it is necessary to understand the preceding definition and to recognize a derivative when it is staring you in the face. That is the point of the following exercises.

Exercises

8.1. Consider each of the expressions below, and show that you recognize it as a derivative.

(a) $\displaystyle \operatorname*{Limit}_{h \to 0} \frac{f(\pi+h) - f(\pi)}{h}.$

(b) $\operatorname*{Limit}_{h\to 0}(1/h)(f(x+h)-f(x))$.
(c) $\operatorname*{Limit}_{h\to 0}(1/h)(g(1+h)-g(1))$.
(d) $\operatorname*{Limit}_{h\to 0}(1/h)(F(s+h)-F(s))$.

Partial answer: The expression (a) is equal to $f'(\pi)$, and (d) is equal to $F'(s)$.

8.2. Let $f(x)=x^3$, $F(x)=\sqrt{x}$, $g(x)=1/x$. Which of the following is equal to $f'(x)$? to $F'(x)$? to $g'(x)$?

$$\operatorname*{Limit}_{h\to 0}\frac{\sqrt{x+h}-\sqrt{x}}{h},\quad \operatorname*{Limit}_{h\to 0}\frac{1}{h}\left(\frac{1}{x+h}-\frac{1}{x}\right),\quad \operatorname*{Limit}_{h\to 0}\frac{(x+h)^3-x^3}{h}.$$

8.3. (*For more ambitious students*) Evaluate each of the limits in exercise 8.2. (Answers may be found in Chapter 5.)

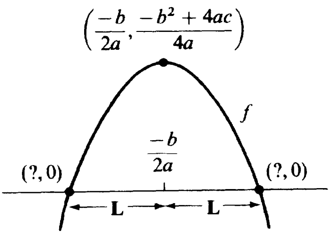
8.4. (*For more ambitious students*) Consider the function f given by $f(L)=7L+(48/L)$.
(a) What is $f(L+h)$?
(b) Simplify the expression $f(L+h)-f(L)$, as much as possible, by combining fractions with the use of a common denominator.
(c) Divide your answer to (b) by h, where $h\neq 0$. *Answer*: $7-(48/L(L+h))$.
(d) Find $f'(L)$, by taking the limit of your answer to (c), as h tends to 0.
(e) Solve Example 1 of Chapter 1 by using f' to find the minimal cost.

§9. Classifying Quadratics: the Quadratic Formula

The reader has probably heard of the *quadratic formula*, which is the answer to question (9) below. This formula was known long before calculus was developed, but our study in this chapter of the calculus of quadratics may cast a new light upon it. We have seen that the quadratic function given by $f(x)=ax^2+bx+c$ has a horizontal tangent line at the point

$$\left(\frac{-b}{2a},\frac{-b^2+4ac}{4a}\right).$$

This followed, as in exercise 7.3(e), from setting the derivative $f'(x)$ equal to zero and solving for x. The graph of the quadratic f might look like this:



x	$\xrightarrow{f}$ y	y'
$-b/2a$	$(-b^2+4ac)/4a$	0
x	ax^2+bx+c	$2ax+b$

For any quadratic f we have thus answered the question, *when is $f'(x)$ equal to zero?* (*Answer*: At the "critical point" $-b/2a$.) We now ask a different question.

$$x \xrightarrow{f} y$$

x	y
?	0
?	0
x	$ax^2 + bx + c$

$$\text{When is } f(x) \text{ equal to zero?} \tag{9}$$

The clue to answering question (9) lies in the apparent symmetry of the curve above. We are inclined to guess that the question has *two answers*, each lying the same distance $\mathbf{L}$ from the critical point $-b/2a$. There ought to be, then, some number $\mathbf{L}$ such that

$$f(x) = 0 \quad \text{when } x = \begin{cases} \dfrac{-b}{2a} + \mathbf{L}, \\ \dfrac{-b}{2a} - \mathbf{L}. \end{cases}$$

All that remains is to find this number $\mathbf{L}$. Since $f((-b/2a) + \mathbf{L}) = 0$, we have

$$a\left(\frac{-b}{2a} + \mathbf{L}\right)^2 + b\left(\frac{-b}{2a} + \mathbf{L}\right) + c = 0. \tag{10}$$

In (10), when the first term is squared out, a cancellation results (the reader is asked to perform the calculations), and eventually we get

$$4a^2\mathbf{L}^2 = b^2 - 4ac. \tag{11}$$

Equation (11) bears some scrutiny. We are trying to find $\mathbf{L}$, with a, b, and c being given. Note that the left-hand side of (11) is a square, since $4a^2\mathbf{L}^2 = (2a\mathbf{L})^2$, and therefore cannot be negative. If $b^2 - 4ac$ should be negative, then there is no number $\mathbf{L}$ satisfying (11). On the other hand, if $b^2 - 4ac$ is nonnegative, we can take its square root to solve for $\mathbf{L}$. From (11) there are two possible paths:

Case I. If $b^2 - 4ac < 0$, then there is no number $\mathbf{L}$ satisfying equation (11), and hence there is no number $\mathbf{L}$ satisfying (10).

Case II. If $b^2 - 4ac \geq 0$, then by taking square roots we get

$$2a\mathbf{L} = \sqrt{b^2 - 4ac},$$

$$\mathbf{L} = \frac{\sqrt{b^2 - 4ac}}{2a}.$$

We now know a formula for **L** and we know that the answer to question (9) is given by $x = (-b/2a) \pm \mathbf{L}$. Putting these facts together yields the *quadratic formula* in the theorem below.

Theorem on Quadratics
The equation

$$ax^2 + bx + c = 0 \quad (a \neq 0)$$

has solutions given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \text{(quadratic formula)}$$

provided that the discriminant $b^2 - 4ac$ *is not negative. The equation has no solution if the discriminant is negative.*

For example, consider the equation

$$-16x^2 - 50x + 200 = 0.$$

The discriminant here is $(-50)^2 - 4(-16)(200) = 15{,}300$, whose square root is approximately 123.7. The quadratic formula says the solutions are

$$x = \frac{-(-50) \pm \sqrt{15{,}300}}{-32} \approx \begin{cases} -5.43. \\ 2.30. \end{cases} \qquad \square$$

As another example, consider the equation

$$-16x^2 + 50x - 200 = 0.$$

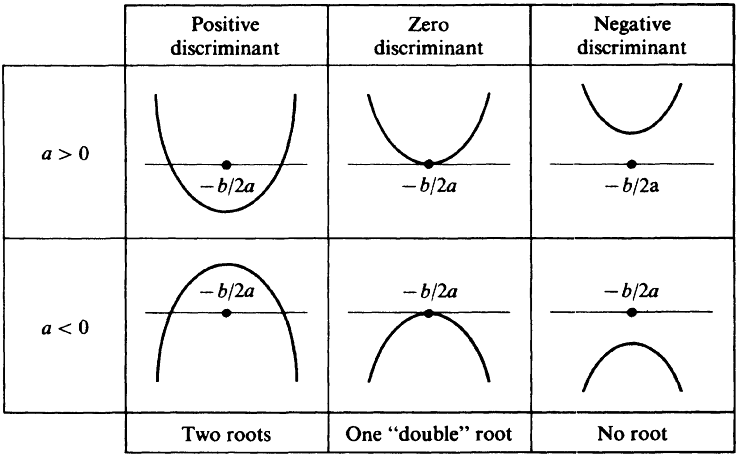
The discriminant here is $(50)^2 - 4(-16)(-200) = -10{,}300$, which is negative, showing that the equation has no solutions. $\square$

If the discriminant is equal to zero then the "two" solutions meld into one. In this case, **L** is zero, making the critical point $-b/2a$ into a "double root" of the quadratic equation. This happens in the equation $x^2 - 4x + 4 = 0$ where the discriminant is $(-4)^2 - 4(1)(4) = 0$, and the quadratic formula yields

$$x = \frac{-(-4) \pm 0}{2} = 2$$

as the only solution. $\square$

There is nothing mysterious going on here. A little reflection shows that any quadratic function falls into one of the following six classifications.



Exercises

9.1. Solve the equation $x^2 - 4x + 4 = 0$ by factoring.

9.2. Solve the equation $x^2 - x - 6 = 0$
(a) by the quadratic formula.
(b) by factoring $x^2 - x - 6$ into the product of $x + 2$ and $x - 3$.

9.3. Solve the equation $x^2 - x - 4 = 0$.

9.4. Factor the quadratic $x^2 - x - 4$ into the product of two linear expressions. *Answer*: $x^2 - x - 4 = (x - \frac{1}{2}(1 + \sqrt{17}))(x - \frac{1}{2}(1 - \sqrt{17}))$.

9.5. Solve the equation $x^2 - 6x + 13 = 0$.

9.6. Show all the steps of an algebraic derivation of equation (11), beginning with equation (10).

9.7. For each of the six categories of quadratics pictured above, give an example.

9.8. Consider once again Example 1 of Chapter 1. Show that it is impossible to build the fence described there for a cost C of \$35. *Hint*. If $C = 35$, then $7L + (48/L) = 35$, so upon multiplying through by L we get

$$7L^2 + 48 = 35L,$$

Use the theorem on quadratics to show that, no matter what the length L of the front fence is, this equation cannot be satisfied.

9.9. What is the least positive value of C for which the equation $7L^2 - CL + 48 = 0$ has a solution? What, therefore, is the least possible cost in Example 1 of Chapter 1?

9.10. The quadratic formula says that the roots of the equation $ax^2 + 5x - 3 = 0$ are given by $(-5 \pm \sqrt{25 + 12a})/2a$ if $a \neq 0$. If $a = 0$ this equation becomes simply $5x - 3 = 0$, whose single solution is given by $x = 3/5$. *How do the pair of roots occurring when a is near 0 manage to go away and leave us with only the single root of* $3/5$ *when* $a = 0$? This question leads us to investigate the limit of each of these roots as a tends to 0:

$$\operatorname*{Limit}_{a \to 0} \frac{-5 + \sqrt{25 + 12a}}{2a} \quad \text{and} \quad \operatorname*{Limit}_{a \to 0} \frac{-5 - \sqrt{25 + 12a}}{2a}.$$

(a) Do either of these limits exist? Does either limit seem to be equal to $3/5$? [Plug in a very small nonzero value of a and see how close you are to $3/5$.]

(b) On the same coordinate system graph the line $y = 5x - 3$ and graph the three quadratic curves whose equations are $y = ax^2 + 5x - 3$ for $a = 1, 1/10$, and $1/100$. Indicate the points where each of these quadratic curves crosses the x-axis. Referring to these graphs, write a short paragraph explaining why this bit of analytic geometry dispels any mystery in your answers to part (a) by visually showing why those answers should have been obvious all along.

9.11. Find $\text{Limit}_{a \to 0}(-3 + \sqrt{9 + 28a})/2a$. *Hint.* This is the limit of a root of $ax^2 + 3x - 7 = 0$ and you can guess the right answer by the technique used in part (b) of exercise 9.10. But the standard way of doing this is first to "rationalize the numerator" by rewriting the troublesome expression as follows:

$$\begin{aligned} \frac{-3 + \sqrt{9 + 28a}}{2a} &= \frac{-3 + \sqrt{9 + 28a}}{2a} \, \frac{-3 - \sqrt{9 + 28a}}{-3 - \sqrt{9 + 28a}} \\ &= \frac{9 - (9 + 28a)}{2a(-3 - \sqrt{9 + 28a})} = \frac{-14}{-3 - \sqrt{9 + 28a}}. \end{aligned}$$

The limit as a tends to 0 is now easy to take.

§10. Newton's Method: Using Derivatives to Solve Equations

Beginning with the equation $ax^2 + bx + c = 0$ we needed some fairly complicated algebra in Section 9 to derive the quadratic formula that solves this equation. A less complicated approach using calculus leads to a method that solves not only quadratics, but cubics, quartics, and even more complicated equations of the form $f(x) = 0$. The essence of this method is expressed in the simple formula (12) below, which was discovered and exploited in the seventeenth century by Isaac Newton and others, whose approach to this problem is similar in spirit to the ancient Babylonian method of approximating square roots, a topic discussed in Chapter 2.

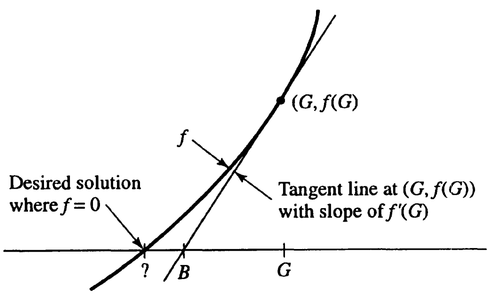
A method (or procedure, or "recipe") that describes steps to be taken toward attaining a desired result in mathematics is called an *algorithm*. The word, which originally came into English as *algorism*, is a corruption of the name of the ninth-century Islamic mathematician al-Khowarizmi, who was well known for his writings on algebra. Recall the main steps of the Babylonian recipe for cooking up the square root of 2:

BABYLONIAN ALGORITHM TO SOLVE EQUATION $x^2 = 2$

1. Input G.	[Make an initial guess at $\sqrt{2}$; call it G.]
2. Let $B = (G + 2/G)/2$.	[B is expected to be a better guess than G.]
3. Let $G = B$.	[Let the better guess serve as a new G.]
4. Go to Step 2.	[Repeat procedure, getting a new B.]

As we have seen in Section 6 of Chapter 2, this procedure will accept any positive number G as an initial guess and will produce succeeding values of G that approach $\sqrt{2}$ as a limit. In theory the algorithm above will run forever, producing ever closer approximations to $\sqrt{2}$, but in practice we cut it off when we are satisfied with the degree of approximation already obtained.

The Babylonian problem is to solve for x in the equation $x^2 = 2$, which may be rewritten as $x^2 - 2 = 0$. In modern terms the problem may be rephrased as seeking the positive solution to the equation $f(x) = 0$, where $f(x) = x^2 - 2$. That is, we want the x-coordinate of the point where the curve f crosses the x-axis. Given a guess G at this x-coordinate, the picture below indicates how we might use the idea of a tangent line to produce a better guess B. Note in this picture how much closer B is to the desired solution than G.



The picture tells us how to get a better guess B: *Given* G, *let* B *equal the* x*-intercept of the line tangent to* f *at* $(G, f(G))$. Since this tangent line is the line of slope $f'(G)$ passing through $(G, f(G))$ it has the equation

$$y - f(G) = f'(G)(x - G).$$

Our better guess B is then the value of x when $y = 0$. Thus we must arrange to choose B so the equation above is satisfied when $y = 0$ and $x = B$, i.e., we must have

$$0 - f(G) = f'(G)(B - G).$$

Our object is to find B if we are given G. Assuming $f'(G)$ is nonzero we first divide to get

$$-\frac{f(G)}{f'(G)} = B - G,$$

and then we have what we want:

$$B = G - \frac{f(G)}{f'(G)}. \tag{12}$$

Formula (12) expresses **Newton's method** (sometimes called the Newton-Raphson algorithm) of computing a better guess B for a given guess G at a solution to the equation $f(x) = 0$.

NEWTON'S METHOD TO SOLVE EQUATION $f(x) = 0$

1. Input G.	[Make an initial guess at solution; call it G.]
2. Let $B = G - f(G)/f'(G)$.	[Formula (12); B is expected to be better.]
3. Let $G = B$.	[Let the better guess serve as a new G.]
4. Go to Step 2.	[Repeat procedure, getting a new B.]

Newton's method can be applied, in principle, to any function f whose derivative you know. In practice it is prudent to make sure the equation $f(x) = 0$ that you are trying to solve does indeed have a solution by making a rough sketch of the curve f and assuring yourself that the curve does indeed cross the x-axis. It usually helps to take a little time to choose an initial guess G that is already fairly close to the solution you seek. (Use trial-and-error to choose your initial guess G so that $f(G)$ is already close to 0.) Then the method often works with astonishing quickness, as in the examples below.

EXAMPLE 7

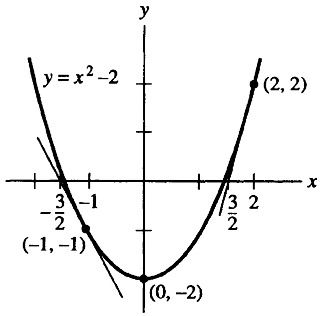
Apply Newton's method to solve the equation $x^2 - 2 = 0$.

Here we are solving $f(x) = 0$ where $f(x) = x^2 - 2$, so $f'(x) = 2x$. Thus $f(G) = G^2 - 2$ and $f'(G) = 2G$ so that formula (12) becomes

$$B = G - \frac{G^2 - 2}{2G}. \tag{13}$$

As we have already noted in exercise 6.2(b) of Chapter 2, formula (13) produces exactly the same results as the Babylonian formula $B = (G + 2/G)/2$: If $G = 2$, then $B = 3/2$; if $G = 3/2$, then $B = 17/12$. And so on. The successive values of 2, 3/2, 17/12, 577/408, etc. get successively closer to $\sqrt{2}$, the positive solution to the equation $f(x) = 0$.

What happens if we begin with a negative number, say $G = -1$, as our initial guess. Then, as the reader is asked to verify in exercise 10.1 below, the successive B's calculated by formula (13) will tend to $-\sqrt{2}$, the negative solution of the equation $f(x) = 0$. Studying the figure below makes it clear why any negative initial guess here will tend to the negative solution and any positive first guess will tend to the positive solution. The figure also makes clear why it is not a good idea in this situation to choose 0 as a first guess.



If $G = -1$, then $B = -\frac{3}{2}$.

If $G = 2$, then $B = \frac{3}{2}$.

Newton's method not only gives the same result as the Babylonian method in the problem of finding square roots; it gives the same result, in essence, as the quadratic formula in the problem of finding solutions to a quadratic equation.

EXAMPLE 8
In Section 9 we applied the quadratic formula to solve $-16x^2 - 50x + 200 = 0$ and worked out the approximate solutions of -5.43 and 2.30. Solve this equation by Newton's method.

Here we have the equation $f(x) = 0$ where $f(x) = -16x^2 - 50x + 200$, so $f'(x) = -32x - 50$. Thus $f(G) = -16G^2 - 50G + 200$ and $f'(G) = -32G - 50$, so formula (12) becomes

$$B = G - \frac{-16G^2 - 50G + 200}{-32G - 50}$$

$$= G - \frac{8G^2 + 25G - 100}{16G + 25}. \tag{14}$$

Let us make two tables of successive G's and B's, with $G = 0$ as initial guess in the first table and $G = -5$ in the second. Newton's method often takes only a few calculations to get as close as you please to a solution, but these calculations can be done even more quickly with the help of an electronic spreadsheet, if you have access to one.

G	B [by formula (14)]
0.00000000	4.00000000
4.00000000	2.56179775
2.56179775	2.31103680
2.31103680	2.30292005
2.30292005	2.30291152
2.30291152	2.30291152

G	B [by formula (14)]
−5.0000000	−5.4545455
−5.4545455	−5.4380027
−5.4380027	−5.4279115
−5.4279115	−5.4279115

There is no point in continuing either table further, as the bottom numbers would only repeat themselves unless we go to the trouble of calculating them to many more decimal places. When they repeat, then of course we have $B = G$ in formula (12), which in turn immediately implies (*why?*) that $f(G)/f'(G) = 0$. But this means that $f(G) = 0$, i.e., G is a solution to the equation $f(x) = 0$. Thus the solutions to the equation $-16x^2 - 50x + 200 = 0$ are given to great accuracy by the repeated numbers -5.4279115 and 2.30291152. This agrees, of course, with the answers worked out in the previous section using the quadratic formula.

Notice that starting with the guess $G = 0$, which is relatively far away from the root near 2.30, the method takes a couple of steps longer to achieve its goal than starting at $G = -5$, which is already rather close to the root near -5.43. □

EXAMPLE 9

Apply Newton's method to find the cube root of 2.

To find the cube root of 2 is to find the number x whose cube is 2, i.e., to solve the equation $x^3 = 2$. *This equation is not of the form* $f(x) = 0$, *but any equation can be put into this form*, simply by taking all the terms over to the left-hand side, here getting $x^3 - 2 = 0$. Thus we have $f(x) = x^3 - 2$. It is true that

$$\text{if } f(x) = x^3 - 2, \quad \text{then } f'(x) = 3x^2, \tag{15}$$

a result already obtained by the reader who has successfully done the relevant part of exercise 8.3. We will soon see how to get this result more easily (without going through Fermat's method) in exercise 6.1 of the next chapter, but let us go ahead and use it now.

The rest is simple. To find $\sqrt[3]{2}$ we apply Newton's method to the equation $f(x) = 0$ where $f(x) = x^3 - 2$ and $f'(x) = 3x^2$, so $f(G) = G^3 - 2$ and $f'(G) = 3G^2$. Formula (12) becomes

$$B = G - \frac{G^3 - 2}{3G^2} \tag{16}$$

and this leads to the following table of successive approximations, starting with the initial guess $G = 1$.

G	B[from formula (16)]
1.00000000	1.3333333
1.33333333	1.2812503
1.2812503	1.2602743
1.2602743	1.2599213
1.2599213	1.2599213

Thus $\sqrt[3]{2} \approx 1.2599213$. (Did your response to exercise 6.7 of Chapter 2, using a different formula, produce this same table of values? It should have.) □

Warning. Some equations, like $7x^2 - 35x + 48 = 0$, studied in exercise 9.8 in the previous section, do not have solutions. If you apply Newton's algorithm to this equation the successive numbers you get will show no tendency to repeat themselves, no matter what initial guess you make. If you are foolish enough to apply Newton's method to try to find non-existent solutions, you will get what you deserve.

Exercises

10.1. Start with any negative number, say $G = -1$, as your initial guess G, and use formula (13) to calculate successively better and better guesses at a solution to the equation $x^2 - 2 = 0$. Do they get closer and closer to the negative root of this equation? What happens if you choose $G = 0$ as your initial guess?

10.2. Find the two roots of $x^2 - x - 4 = 0$ by
(a) using the quadratic formula. (You may have already done this in 9.3 or 9.4.)
(b) using Newton's method. Hint. Here formula (12) becomes $B = G - (G^2 - G - 4)/(2G - 1)$. Make two tables analogous to those of Example 8. Take an initial guesses of $G = 0$ in one and of $G = 3$ in the other. Does Newton's method lead to the same answer as in part (a)?

10.3. Approximate the square root of 10 by Newton's method, i.e., solve the quadratic equation $x^2 - 10 = 0$ for its positive root. (If you take $G = 3$ as your initial guess you should get exactly the same sequence of approximations as in exercise 6.3 of Chapter 2.)

10.4. Approximate the square root of 48/7 by Newton's method. (If you take $G = 2$ as your initial guess you should get exactly the same sequence of approximations as in exercise 6.6 of Chapter 2.)

10.5. Applying Newton's method to a linear equation is rather heavy-handed, but instructive. Apply the method to solve $5x - 3 = 0$ by first writing down formula (12) with $f(x) = 5x - 3$ and $f'(x) = 5$. Then make an outrageous first guess, like $G = 100$. What happens? Now try $G = -1000$ as your first guess. You have just discovered the *rule of false position*, a method of solving linear equations that dates back to the ancient Egyptians. Explain geometrically what is going on here. *Hint.* In the picture from which formula (12) was derived, the curve f is straight if f is a linear function, so its tangent line is itself.

10.6. Approximate the cube root of 16 by Newton's method. *Hint.* In equation (16) replace "2" by "16" and make the resulting table, choosing a good initial guess at $\sqrt[3]{16}$.

10.7. In Newton's method to solve $x^3 - 27x + 27 = 0$, what does formula (12) become? *Answer*: Using the result of exercise 8.3 we see it is

$$B = G - \frac{G^3 - 27G + 27}{3G^2 - 27}.$$

10.8. In Newton's method to solve $x^3 - 5x^2 + 6x - 1 = 0$, what does formula (12) become? Use it to find one root of this equation. Can you find all three roots by making three different initial guesses? What are the three roots?

§11. Three Frenchmen

The influence of France was increasingly felt throughout Europe in the seventeenth century. This influence was particularly strong in mathematics. France nurtured no fewer than three mathematical minds of the first rank, in addition to many lesser lights.

Blaise Pascal (1623–1662), who at the age of eighteen invented the first calculating machine, might have been unsurpassed as a mathematician, had his other great talents not drawn him elsewhere. Even so, he helped give birth to projective geometry and to the theory of probability, and he came very close to discovering the fundamental theorem of calculus (to be discussed in Chapter 7). In fact, Leibniz hit upon the fundamental theorem while reading a mathematics paper by Pascal.

Little need be said here of René Descartes (1596–1650), for half the world already knows his name. We have noted earlier that he developed analytic geometry and made it widely known through his writings. Without analytic geometry the step up to the calculus would be formidable indeed. Isaac Newton was to say, "If I have seen further than Descartes, it is by standing on the shoulders of giants." One of those giants was, of course, Descartes himself.

Another giant was Pierre de Fermat (1601–1665). Fermat occupies a special place in the hearts of those who love mathematics. His appeal is that of the *amateur* who can outdo the professionals. Fermat developed analytic geometry in 1629, but did not publicize the fact, and Descartes got all the credit with a paper published in 1637. In correspondence with Pascal, Fermat was an equal partner in creating the theory of probability. He corrected mistakes that Descartes and Pascal made, in fields where they were acknowledged as masters, and was rarely himself in error. Fermat's real love was the theory of numbers, which was revolutionized by his accomplishments.

Unlike Descartes and Pascal, Fermat published little, and his work is known mainly through his letters and through the notes he was accustomed to make in the margins of books. Although Fermat never used the word *limit*, Laplace's appellation—"Fermat, the true inventor of the differential calculus"—can still be maintained. Isaac Newton, in a letter discovered only in 1934, stated that his own early ideas about calculus came directly from Fermat's way of drawing tangents, presumably because of Fermat's implicit use of limits.

Problem Set for Chapter 4

1. Use Fermat's method (not a shortcut rule) to show that the derivative of $x^2 - 6x + 13$ is given by $2x - 6$.

2. Consider the function defined by $f(x) = x^2 - 6x + 13$.
   (a) Fill in the question marks in the following table using the fact that $y' = 2x - 6$.

$x \xrightarrow{f} y$

x	y	y'
0	?	?
?	?	-4
5	?	?
π	?	?
?	?	0

   (b) Use the first line of the table to find the slope of the tangent line to the curve f at the point $(0, 13)$.
   (c) Is the curve f *rising* or *falling* as it passes through the point $(0, 13)$?
   (d) Write an equation of the tangent line to the curve f at the point $(0, 13)$.
   (e) For what values of x is $f'(x)$ positive?
   (f) For what values of x is the curve f rising?
   (g) For what values of x is the curve f falling?
   (h) Find both coordinates of the point where the tangent line to the curve f is horizontal.
   (i) Sketch the curve f in the x-y plane, making sure your sketch is in accordance with your answers to the preceding three questions.

(j) If the domain of f is taken to be all values of x satisfying the inequality $0 \le x \le 5$, what is the range?
(k) What is the range of f if the domain is given by the inequality $0 < x < 5$?
(l) What is the range if the domain is $0 \le x < 2$?
(m) What is the range if the domain is unrestricted?

3. Consider the function defined by $f(x) = 2x^2 - 6x$.
   (a) What is the slope of the tangent line to the curve f at the point $(0, 0)$?
   (b) Is the curve f *rising* or *falling* at $(0, 0)$?
   (c) For what values of x is the curve f falling?
   (d) Sketch the curve f.
   (e) What is the range of f if the domain is $0 \le x \le 4$?
   (f) What is the range of f if the domain is $0 < x < 1$?
   (g) What is the range of f if the domain is unrestricted?

4. Consider the function defined by $f(x) = 8 - x^2 + 3x$.
   (a) What is the range if the domain is $0 \le x < 2$?
   (b) What is the range if the domain is $-2 < x < 0$?

5. Consider the function defined by $f(x) = 8 - 3x$.
   (a) What is the range if the domain is $0 \le x \le 3$?
   (b) What is the range if the domain is $-2 < x < 5$?

6. Find both coordinates of the point on the line $y = 2x - 3$ that is closest to $(0, 0)$.

7. Find both coordinates of the point on the line $y = 6 - x$ that is closest to $(-2, -4)$.

8. (*This problem is like Example 1 of Chapter 1, except we have a specified amount of money instead of a specified area.*) A sum of \$56 has been allocated to pay for fencing four sides of a rectangular area. If the front fence costs \$5 per yard and each of the other three sides costs \$2 per yard, what is the maximum area that can be enclosed? *Hint.* If L is the length in yards of the front side, and A is the corresponding area in square yards, then A/L is the width of the rectangle, so $56 = 7L + 4A/L$. (*Do you see why?*) Solve this equation to get A in terms of L. Then get A' quickly and use it to find the maximum area.

9. A *Norman window* is in the shape of a rectangle surmounted by a semicircle. If the perimeter of the window is 16 feet, find the dimensions which allow the most light to pass through the window.

10. Although the point $(2, -1)$ is not on the quadratic $y = x^2 - 2x + 3$, there are two tangent lines to this quadratic that pass through this point. Find an equation of either one of these lines.

11. A wire 500 centimeters long is cut in two. The first part is bent into the circumference of a circle, and the second is bent into the perimeter of a square. How should the wire be cut in order that the combined areas of the circle and the square be as small as possible?

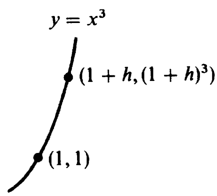
12. Use Fermat's method to show that if $f(x) = 7x - 9$, then $f'(x) = 7$.

13. Use Fermat's method to show that if $f(x) = 7$, then $f'(x) = 0$.

14. Consider the quadratic function given by $f(x) = x^2 + 2x + 7$. When $x = 0$, then $f(x) = 7$. What is $f'(x)$ when $x = 0$? Does your answer contradict the result of problem 13? Explain.

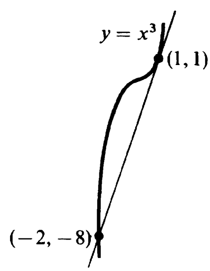
15. Reread problem 16 in the problem set at the end of Chapter 1. Find the maximal area that can be enclosed.

16. Carry out the following steps in order to accomplish the last step.
    (a) The points (1, 1) and $(1+h, (1+h)^3)$ lie on the curve $y = x^3$. Find the slope of the line joining these points (assuming, of course, that $h \neq 0$).
    (b) Simplify your answer to part (a) by using the fact that $(1+h)^3 = 1 + 3h + 3h^2 + h^3$.



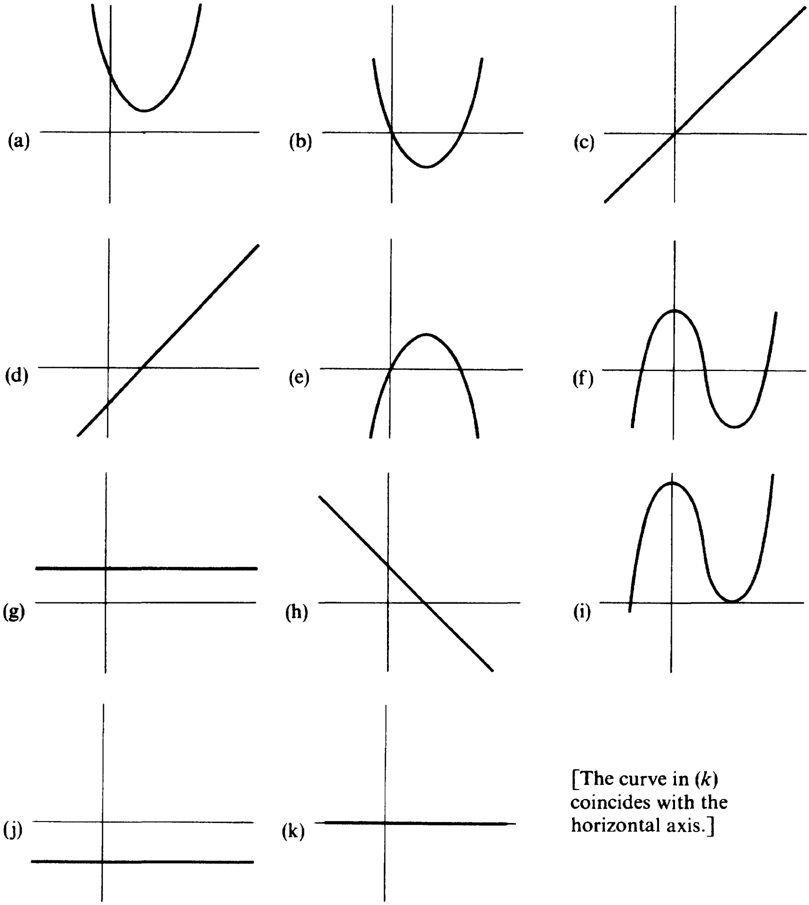
    (c) Take the limit, as $h \to 0$, of your answer to part (b), and thus show that the slope of the tangent line to the curve $y = x^3$ at $(1, 1)$ is 3.

17. *(This problem is supposed to show why the "Sherlock Holmes method" of finding tangent lines will not always work. Actually, Descartes proposed a closely related method, but it had to be discarded in favor of Fermat's approach.)* The Sherlock Holmes method of finding tangents rests upon the belief that a line joining two points on a curve cannot be tangent to the curve. (This happens to be true for quadratic curves.)
    (a) Using the result of problem 16(c), write an equation of the tangent line at $(1, 1)$ to the curve $y = x^3$.
    (b) Does the point $(-2, -8)$ lie on the line of part (a)?
    (c) Does the point $(-2, -8)$ lie on the curve $y = x^3$?



    (d) If the Sherlock Holmes method were applied to the curve $y = x^3$ at the point $(1, 1)$, would it work?

18. Consider each of the following functions, and match it with its derivative. [The derivative of the function pictured in (a), for example, is pictured in (d).]



[The curve in (k) coincides with the horizontal axis.]

19. If the page containing the curves of problem 18 is turned upside down, we see eleven new functions pictured. Match each of these with its derivative. [If the curve pictured in (a) is viewed upside down, its derivative is pictured in (h), viewed upside down.]

20. (*The general "Minerva problem"*) The method used in Section 7 to find the largest area enclosed by a rectangle where the total length of three sides is 1200 feet is easily generalized to cover any specified total length.

(a) A farmer has π kilometers of fencing to fence in three sides of a rectangular field for her cow Minerva, the fourth side being bounded by a long fence already standing. What is the maximal area that can be enclosed? Draw a picture, giving the dimensions, of the field with maximal area subject to these constraints.

(b) A farmer has a length P of fencing to fence in three sides of a rectangular field for her cow Minerva, the fourth side being bounded by a long fence already standing. What is the maximal area that can be enclosed? Draw a picture, giving the dimensions, of the field with maximal area subject to these constraints.

21. In the corner of a large courtyard a rectangular enclosure is to be built like the one pictured in problem 16 at the end of Chapter 1. To pay for the material, \$1000 has been allocated. This is to be used to pay for both the stone fence, which costs \$10 per meter, and the wooden fence, which costs \$5 per meter. What is the maximal area that can be enclosed? Draw a picture giving its dimensions.

22. (*The general "courtyard problem"*) This problem generalizes problem 21, just as problem 20 generalizes the Minerva problem.
    (a) In the corner of a large courtyard a rectangular enclosure is to be built like the one pictured in problem 16 of the problem set at the end of Chapter 1. To pay for the material, P dollars has been allocated. This is to be used to pay for both the stone fence, which costs A dollars per meter, and the wooden fence, which costs B dollars per meter. What is the maximal area that can be enclosed? Draw a picture giving its dimensions.
    (b) Show that the way to optimize the area in the courtyard problem is simply to put half the allocated money into wood and the other half into stone. *Hint.* You have solved the problem in part (a). Just explain why your solution allocates the same amount of money for stone and wood.

23. Use the algebraic technique of "rationalizing the numerator" illustrated in exercise 9.11 to find the following two limits:

    (a) $\operatorname*{Limit}_{x\to 0} \dfrac{-4+\sqrt{16-28x}}{2x}$. (b) $\operatorname*{Limit}_{x\to 0} \dfrac{3-\sqrt{9-40x}}{x}$.

    Knowing the limits of two expressions makes it easy to find the limit of their product or quotient. Using your answers to (a) and (b), find each of the following limits quickly, without any involved calculation:

    (c) $\operatorname*{Limit}_{x\to 0} \dfrac{(-4+\sqrt{16-28x})(3-\sqrt{9-40x})}{2x^2}$.

    (d) $\operatorname*{Limit}_{x\to 0} \dfrac{-4+\sqrt{16-28x}}{6-2\sqrt{9-40x}}$.

    (e) $\operatorname*{Limit}_{x\to 0} \left(\dfrac{-4+\sqrt{16-28x}}{6-2\sqrt{9-40x}}\right)^3$.

24. (*Do you understand Newton's method?*) Here is a simple test. Fill in the question mark in each table with the number Newton would give for his guess at the right answer.

(a)

x	y	y'
1	1	-27
?	0	

(b)

x	y	y'
-2	-1	10
?	0	

25. (*Newton's method*) Find the root of the equation $x^3 - 27x + 27 = 0$ that lies near 1 by taking $G = 1$ as your initial guess in Newton's method. Find this root to an accuracy of at least four decimal places. Then find the other two roots of this equation by taking different initial guesses in Newton's method. *Hint.* See exercise 10.7.

26. (*Fermat's "other" method*) We have seen how useful Fermat's method of tangents is in studying continuous curves. Fermat also developed a *discrete* method for use in *reductio ad absurdum* proofs in the theory of numbers. Learn this method by going through the steps below that lead to Fermat's strikingly modern proof of the irrationality of $\sqrt{2}$.
   (a) But first, explain in what sense "concrete" and "continuous" can each be regarded as antonyms of "discrete". If you don't know this already, you don't know as much about continuity as you might. (The discreet use of a dictionary will help.)
   (b) If $\sqrt{2} = m/n$, then $m^2 = 2n^2$ and thus $m^2 - mn = 2n^2 - mn$. It follows that $m(m - n) = n(2n - m)$. Show that this implies that the ratio m/n can be re-expressed as
   $$\frac{m}{n} = \frac{2n - m}{m - n}.$$
   (c) Show that if $\sqrt{2} = m/n$, where m and n are positive integers, then $0 < m - n < n$ and $0 < 2n - m < m$. *Hint.* Since $1 < \sqrt{2} < 2$, we know $1 < m/n < 2$. Multiplying each member of this inequality by n shows that $n < m < 2n$. Continue working with this last inequality to deduce the two inequalities required.
   (d) Notice that parts (b) and (c) together show that if $\sqrt{2}$ can be expressed as the quotient of positive integers m and n, then m/n can be re-expressed as the quotient of still smaller positive integers. By repeating the same argument we can re-express this second quotient as the quotient of even smaller positive integers. And so on. As we continually repeat the same argument we produce an "infinite descent" of ever-decreasing positive integers (in both numerator and denominator).
   (e) But just a moment's thought shows you that *there can be no such thing.* Think about it! You can have an "infinite ascent" of ever-increasing positive integers, but you certainly cannot have an infinite descent of integers without eventually having negative integers appear.* Thus *any*

* Fermat was not the first to make this simple observation (Euclid noted it as an obvious fact two thousand years earlier), but he was the first of the moderns to sense how this observation captures the most essential property of the Pythagorean world of numbers. No matter where you start in this world, even if you start with the number of grains of sand in the earth, you cannot count downwards forever using only positive integers. A consequence of this is the **well-ordering principle**: Any (nonempty) collection of positive integers contains a least integer. *Reason*: If no member of the collection is its least member,

supposition that leads to an infinite descent of positive integers must be false. This is the essence of **Fermat's method of infinite descent.** Since we have seen in part (d) that the supposition that $\sqrt{2}$ is rational leads to an infinite descent of positive integers, that supposition must be false, Q.E.D.

27. Use Fermat's method described in the preceding problem to prove that $\sqrt{6}$ is irrational. *Hint.* First show that if $m/n = \sqrt{6}$, then $m/n = (6n - 2m)/(m - 2n)$; and note that $2 < \sqrt{6} < 3$. Then go through steps analogous to steps (c) and (d) of the previous problem and obtain a contradiction if m and n are positive integers.

28. Prove that $\sqrt{10}$ is irrational. *Hint.* If $m/n = \sqrt{10}$, then $m/n = (10n - 3m)/(m - 3n)$; note that $3 < \sqrt{10} < 4$.

29. Something very much like Fermat's observation is already implicit in the Pythagorean proof of the irrationality of $\sqrt{2}$ sketched in Chapter 2, which begins by taking it as obvious that *any rational number can be re-expressed in "lowest terms"*. Write out a proof of this "obvious" fact. *Hint.* Observe that if it were not the case, then for some rational m/n we could continue to cancel out common factors in numerator and denominator forever. Explain how this would lead to an infinite descent of positive integers (in both numerator and denominator), and hence to a contradiction.

we could start by choosing any member we please and be able to find one that is smaller, thus counting downwards forever using only positive integers from the given collection.

A "keen sense of the obvious" is not to be derided. It is one of those childlike qualities often associated with genius. Do you sense the difference between the discrete world of positive integers in which you can never count downwards forever, and the continuous world of positive real numbers, in which you can? The ancient Babylonians showed us how to count downwards forever toward $\sqrt{2}$ by using the sequence beginning 3/2, 17/12, 577/408, etc. generated by their algorithm. The infinite set of positive real numbers generated here obviously does not contain a least member (because $\sqrt{2}$ is not a member of this set), whereas every (nonempty) set of positive integers obviously does.

To fail to sense such a difference is to confuse the continuous with the discrete. Although Fermat captured the most essential property of the *discrete* set of positive integers in the early seventeenth century, it was not until the late nineteenth century (see Section 6 of Chapter 2) that mathematicians were able to pin down the most essential property of the *continuous* set of real numbers.

5 CHAPTER Optimistic Steps

What is calculus? It is the study of the interplay between a function and its derivative. There are quite a few aspects to this interplay, some of which may be surprising. In this chapter we shall learn more about the use of derivatives in solving optimization problems. To do this efficiently, the major part of the chapter is concentrated upon the development of shortcut rules for finding derivatives.

§1. The Derivative of the Reciprocal Function

If f is a function, then $f'(x)$ is defined as the limit of the *difference quotient*

$$\frac{f(x+h)-f(x)}{h} \tag{1}$$

as h tends to zero. (*Do you see why expression* (1) *is a quotient of differences*?) In order to find this limit, it is often necessary to use a little algebra to write the difference quotient in a simple way.

Before proceeding, let us review very briefly the algebra of simplifying fractions by combining them with the use of a common denominator. For instance,

$$\frac{1}{5}-\frac{1}{4}=\frac{4}{5\cdot 4}-\frac{5}{5\cdot 4}=\frac{4-5}{5\cdot 4}=\frac{-1}{20},$$

$$\frac{1}{\pi+2}-\frac{1}{\pi}=\frac{\pi}{(\pi+2)\pi}-\frac{\pi+2}{(\pi+2)\pi}=\frac{\pi-(\pi+2)}{(\pi+2)\pi}=\frac{-2}{(\pi+2)\pi},$$

and, by the same token,

$$\frac{1}{x+h}-\frac{1}{x}=\frac{x}{(x+h)x}-\frac{x+h}{(x+h)x}=\frac{x-(x+h)}{(x+h)x}=\frac{-h}{(x+h)x}. \tag{2}$$

Whenever the rule for f involves division, the difference quotient (1) is often filled with fractions that need to be combined by using a common denominator.

EXAMPLE 1

Find $f'(x)$, if $f(x)=1/x$.

Here, the algebraic rule $1/x$ involves division (and, of course, is undefined when x is 0). Let us first use a common denominator to simplify the expression

$$f(x+h)-f(x)$$

before dividing by h and taking the limit. Since $f(x)=1/x$ and $f(x+h)=1/(x+h)$, equation (2) shows that

$$f(x+h)-f(x)=\frac{-h}{(x+h)x}.$$

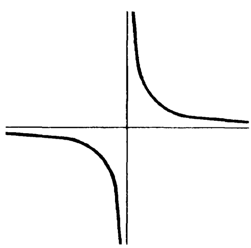
Dividing by nonzero h yields

$$\frac{f(x+h)-f(x)}{h}=\frac{-1}{(x+h)x}.$$

Using this simplified expression for the difference quotient makes it easy to take the limit:

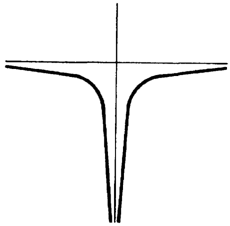
$$\begin{aligned} f'(x) &= \operatorname*{Limit}_{h\to 0}\frac{f(x+h)-f(x)}{h} \\ &= \operatorname*{Limit}_{h\to 0}\frac{-1}{(x+h)x} \\ &= \frac{-1}{(x+0)x} \\ &= \frac{-1}{x^2}. \end{aligned}$$

The derivative of $1/x$ is therefore $-1/x^2$. The expression $1/x$ is called the *reciprocal* of x. We now know the derivative of the reciprocal function.



$f(x)=\dfrac{1}{x}$

The domain of f does not include 0. Note that f' is always negative (because the curve f is always falling).



$f'(x) = \frac{-1}{x^2}$

Exercises

1.1. If $f(t) = 1/t$, what is $f'(t)$? *Answer*: $f'(t) = -1/t^2$, because f is simply the reciprocal function. *Remember that changing only the name of the variable does not alter the function.*

1.2. If $f(L) = 1/L$, what is $f'(L)$?

1.3. If $A = 1/s$, what is A'?

1.4. In each of the following, first make a guess as to the expression giving the derivative. Then verify your guess by Fermat's method.
(a) $y = 48/L$.
(b) $C = 7L + (48/L)$.
(c) $y = 5/(x - 7)$.

1.5. Consider the function given by $f(x) = 1(\pi x + 7)$. Carry out the following steps to find $f'(x)$.
(a) Simplify the expression $f(x+h) - f(x)$ by using a common denominator.
(b) Divide your answer to (a) by nonzero h.
(c) Find the limit, as h tends to zero, of your answer to (b). *Answer*: $f'(x) = -\pi/(\pi x + 7)^2$.

1.6. Suppose $f(x) = 1/g(x)$, where g is some given function. [For instance, exercise 1.5 dealt with the case where $g(x) = \pi x + 7$.] Since f is expressed in terms of g, the difference quotient of f can be expressed in terms of g as well. Show that the difference quotient of f can be expressed by

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

Hint. Begin by writing

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

and combine the fractions on the right. Then divide by h.

§2. General Rules for Reciprocals and for Constant Multiples

Now that we know the derivative of $1/x$, it is natural to ask about the derivatives of similar expressions involving reciprocals. For example, what is the derivative of $1/x^2$? Or of $1/(x^2+2x)$? Or, more generally, what is the derivative of $1/g(x)$, where g is some given function?

To answer this question, first recall that $g'(x)$ is the limit, as h approaches zero, of the difference quotient

$$\frac{g(x+h)-g(x)}{h} \tag{3}$$

In order that the quotient (3) tend to a limit, the numerator of the quotient must tend to 0 as h tends to 0. *Reason*: The denominator h tends to 0; if the numerator did not, then the quotient would "blow up" as h approached 0, and consequently the limit would not exist.

When the numerator in (3) tends to 0, we have

$$\operatorname*{Limit}_{h\to 0} g(x+h) = g(x) \tag{4}$$

This fact will be useful in just a moment.

We can now answer the question raised at the beginning of this section: *What is the derivative of* $1/g(x)$? The derivative of a function is the limit of its difference quotient. The difference quotient of $1/g(x)$ is simplified in exercise 1.6, showing that *the derivative of* $1/g(x)$ *is equal to*

$$\operatorname*{Limit}_{h\to 0} \frac{-1}{g(x+h)g(x)} \cdot \frac{g(x+h)-g(x)}{h} = \frac{-1}{g(x)g(x)}\, g'(x),$$

where (4) has been used (*how?*) in evaluating this limit. We have just proved the following rule to be valid.

General Rule for Reciprocals

The derivative of $1/g(x)$ *is*

$$\frac{-1}{(g(x))^2}\, g'(x)$$

if the function g has a derivative.

If we suppress writing the variable, this rule can be expressed in a very compact way. It says

$$\left(\frac{1}{g}\right)' = \frac{-1}{g^2}\, g'. \tag{5}$$

For example, *what is the derivative of* $1/(4-3x)$? By (5),

$$\left(\frac{1}{4-3x}\right)' = \frac{-1}{(4-3x)^2}(4-3x)' = \frac{3}{(4-3x)^2},$$

since $(4-3x)' = -3$. □

For another example, *what is the derivative of* $1/(\pi x+7)$? By (5),

$$\left(\frac{1}{\pi x+7}\right)' = \frac{-1}{(\pi x+7)^2}(\pi x+7)' = \frac{-\pi}{(\pi x+7)^2},$$

which agrees with the answer to exercise 1.5. □

A rule that is much easier to prove involves multiplication by constants. We have essentially guessed this rule already, in Chapter 4, when we guessed that the derivative of $\frac{1}{2}s^2$ ought to be one-half the derivative of s^2:

$$\begin{aligned}\left(\frac{s^2}{2}\right)' &= \frac{1}{2}(s^2)' \\ &= \frac{1}{2}(2s) = s.\end{aligned}$$

It takes little imagination to guess that there ought to be a general rule involving multiplication by constants, like $\frac{1}{2}$.

Rule for Constant Multiples
The derivative of $c \cdot g(x)$ *is* $c \cdot g'(x)$ *if* c *is a constant and the function* g *has a derivative.*

For example, *what is the derivative of* $100/(\pi x+7)$? By the rule for constant multiplies,

$$\left(\frac{100}{\pi x+7}\right)' = 100\left(\frac{1}{\pi x+7}\right)' = \frac{-100\pi}{(\pi x+7)^2},$$

where the second equality comes from the general reciprocal rule. □

The reader is asked to prove the rule for constant multiples, in a problem at the end of this chapter.

Exercises

2.1. Find y' if $y = 1/(7+3x)$. *Answer:* $y' = -3/(7+3x)^2$.

2.2. Use the general rule for reciprocals to find the derivatives of the following functions:
(a) $f(x) = 1/x^2$.
(b) $g(x) = 1/(x^2 + 2x)$.
(c) $F(x) = 1/(6 - 3x)$.
(d) $G(x) = 1/(2x^2 - 3x + 4)$.
Answer: (b) $g'(x) = (-2x - 2)/(x^2 + 2x)^2$.

2.3. Apply the general rule for reciprocals to find the derivative of $1/x$. Does your answer agree with the answer obtained in Section 1, where Fermat's method was used?

2.4. Find y' if $y = 1/t^2$. *Answer*: $y' = -2/t^3$.

2.5. Find C' if $C = 1/(L^2 + 4)$.

2.6. Use the general rule for reciprocals, together with the rule for constant multiples, to find the derivatives of the following functions, expressed as algebraic rules:
(a) $5/x^2$.
(b) $14/(x^2 + 2x)$.
(c) $\pi/(6 - 3x)$.
(d) $100/(2x^2 - 3x + 4)$.
(e) $5/(4 - 3s)$.
(f) $48/L$.
(g) $-1/t^2$.
(h) $6/(\pi - \theta)$.

§3. The Sum Rule and the Second Derivative

There is an easy rule involving the sum of two functions. We have essentially guessed this rule already, in Chapter 4, when we guessed that the derivative of $ax^2 + bx + c$ ought to be equal to the sum of the derivatives of ax^2 and of $bx + c$:

$$\begin{aligned}(ax^2 + bx + c)' &= (ax^2)' + (bx + c)' \\ &= 2ax + b.\end{aligned}$$

One would surely suspect that this is a special case of a general rule.

Rule for Sums
The derivative of $f(x) + g(x)$ is equal to $f'(x) + g'(x)$ if the functions f and g have derivatives.

This rule is true, but its proof is left to the reader as a problem at the end of this chapter (problem 37).

One often has to use several rules at once.

EXAMPLE 2
Find the derivative of the function given by the algebraic expression $6x^2 + (17/(x^2 + 3x))$.

Here we have a sum, and the sum rule says $(f + g)' = f' + g'$. Therefore,

$$\begin{aligned}
\left(6x^2 + \frac{17}{x^2 + 3x}\right)' &= (6x^2)' + \left(\frac{17}{x^2 + 3x}\right)' \quad \text{(by sum rule)} \\
&= 6(x^2)' + 17\left(\frac{1}{x^2 + 3x}\right)' \quad \text{(by rule for constant multiples)} \\
&= 6(2x) + 17\frac{-1}{(x^2 + 3x)^2}(2x + 3),
\end{aligned}$$

by the rule for quadratics, together with the reciprocal rule. The answer may be simplified, if desired, to $12x - ((34x + 51)/(x^2 + 3x)^2)$. □

The calculation in Example 2 required several steps. The reader will find that, with practice, it is easy to combine these steps into one:

The derivative of $5x^2 + (6/(x^2 - 2))$ *is* $5(2x) + 6(-1/(x^2 - 2)^2)(2x)$.
The derivative of $7L + (48/L)$ *is* $7 + 48(-1/L^2)(1)$.
The derivative of $(10/x) - (45/x^2) - 5x^2 + x - \pi$ *is* $10(-1/x^2)(1) - 45(-1/x^4)(2x) - 5(2x) + 1$. □

Having taken one derivative, we have nothing preventing us from taking a second derivative. The **second derivative** (the derivative of f') is denoted by f''. We now have $y = f(x), y' = f'(x)$, and $y'' = f''(x)$.

EXAMPLE 3
Find the first and second derivatives of the function given by

$$f(x) = 2x^2 + \frac{3}{x} - 4.$$

Here, the first derivative is given by

$$\begin{aligned}
f'(x) &= (2x^2 + \frac{3}{x} - 4)' \\
&= 4x - \frac{3}{x^2},
\end{aligned}$$

and the second derivative is given by

$$\begin{aligned}
f''(x) &= \left(4x - \frac{3}{x^2}\right)' \\
&= 4 - 3\frac{-1}{x^4}(2x) \\
&= 4 + \frac{6}{x^3}.
\end{aligned}$$

□

A function is given by a pair of columns, and its derivative adds a third column to consider. The second derivative gives us still another column to play with. For example, in the function f above, if we let x equal 1, we get

x	y	y'	y''
1	1	1	10

The first two columns tell us that the curve f goes through the point $P = (1, 1)$. The third column tells us that the tangent at P has a positive slope, so the curve f is rising as it goes through P. What does the fourth column tell us? As we shall see in the next section, the positive second derivative tells us that the curve f, as it passes through P, looks rather like a smile.

Exercises

3.1. For the function f of Example 3, fill in the question marks appropriately.

x	y	y'	y''
2	?	?	?
-1	?	?	?
$\frac{1}{2}$	?	?	?

3.2. Find the *second* derivative of each of the following functions.
(a) $f(x) = 2x^2 + 3x - 5$.
(b) $g(x) = -3x^2 + 4x - \sqrt{2}$. *Answer*: $g''(x) = -6$.
(c) $f(L) = 7L + (48/L)$.
(d) $g(s) = 600s - \frac{1}{2}s^2$.
(e) $G(t) = t^2 - 6t + (5/(t-3))$.
Answer: $G''(t) = 2 + ((10t - 30)/\,(t^2 - 6t + 9)^2)$.
(f) $F(x) = (10/x) - 5x^2 + x - \pi$.

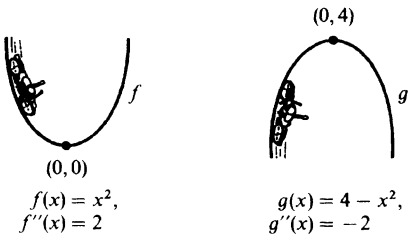
§4. The Second Derivative and Concavity

The second derivative f'' gives the same sort of information about f' as the first derivative f' gives about f. Indirectly, then, the second derivative says something about the behavior of the original function f. Let us try to find out exactly what f'' tells about f.

First we must agree on some terminology to describe how a curve is "curving". There are several terms in use for this (the phrases *concave upwards* and *concave downwards* are common descriptions), but they do not seem to be immediately suggestive of what they are intended to describe. To remedy this, let us depart from common terminology and make up our own way of describing how a curve curves.

We have already agreed, on the first page of Chapter 4, to think of a curve as being traced out from left to right. If we thought of the curve as describing a road on a road map, then the pencil point tracing out the curve moves in a generally eastward direction. Pretend that you, on your motorcycle, have been shrunk to the size of that pencil point tracing out the curve. As you journey eastward, there is a simple way you can describe how the road is curving. All that need be said is whether you are leaning to your *left* or to your *right*, in order to keep your motorcycle on the road.

The functions f and g below curve in opposite directions. Note that their second derivatives f'' and g'' have opposite signs.

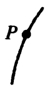


In the case of the curve $y = x^2$, you must always lean to your *left* to stay on the curve; you lean always to your *right* to stay on the curve $y = 4 - x^2$. The first curve has a *positive* second derivative; the second curve has a *negative* second derivative. These are simple examples, in that the second derivative is constant in both, whereas we shall see that generally the second derivative will change sign when the "road" starts to curve the other way. Nevertheless, these examples give us a clue to the truth: *When f'' is positive, the curve f is bending to the left; when f'' is negative, the curve f is bending to the right.*

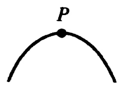
Why should it be this way? Focus attention on a particular point P lying on a curve f. Then $P = (c, f(c))$ for some number c. A little reflection shows that

$$\text{if } f''(c) < 0, \quad \begin{array}{l}\text{then the curve } f \text{ is bending to} \\ \text{the right as it passes through} \\ \text{the point } P = (c, f(c)).\end{array} \tag{6}$$

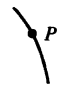
To see this, all one needs to recall is that when the derivative of a function is negative, then the values of the function are *decreasing* (because its curve is falling). Keeping this in mind, and remembering that f'' is the derivative of f', one can see the plausibility of statement (6), as follows. Suppose $f''(c) < 0$. Then the values of f' are decreasing, i.e., the slopes of the tangent lines to the curve f are decreasing, as the curve f passes through $(c, f(c))$. But decreasing slopes of the tangent lines near this point imply that the curve is bending to the right. Near P the curve f must look like one of the following if $f''(c)$ is negative:



(a) Rising, bending right

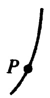


(b) "Flat" at P, bending right

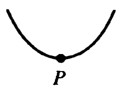


(c) Falling, bending right

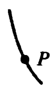
Similar reflection shows that, near P, the curve f must look like one of the following if $f''(c)$ is positive:



(d) Rising, bending left



(e) "Flat" at P, bending left



(f) Falling, bending left

The upshot of the preceding discussion is this. While the sign of the *first* derivative tells whether the curve is rising or falling, the sign of the *second* derivative tells whether it is bending to the left or to the right. The second derivative f'' tells which way the curve f is curving. □

When the second derivative is negative, as in the figures (a), (b), and (c) above, we have described the curve as "bending to the right". It is described as *concave down* in most books, and figures (d), (e), and (f) are described as *concave up*. The definition of concavity in these terms is given in a problem at the end of the chapter. There is little point in learning these terms, however, if you are interested in studying calculus only for a semester or so. In fact, it might be better to describe (d), (e), and (f) as "smiles", and call (a), (b), and (c) "frowns", and just remember that *a positive second derivative always draws a smile.*

Knowing both derivatives of a function at a point gives us a fairly good idea of what the curve looks like nearby. The word *local* (as opposed to *global*) is used in mathematics to describe this kind of information; it tells us what the road looks like only in a small neighborhood of a point as we roar through on our motorcycle. What adventures may lie elsewhere on the road remain to be seen.

EXAMPLE 4

Describe the local behavior of the curve $y = x^2 + (8/x)$ as the curve passes through

(a) $(1, 9)$
(b) $(2, 8)$.
(c) $(-2, 0)$.

It is intended that we sketch the curve locally near each of these points, so as to indicate whether the curve is rising or falling, "smiling" or "frowning", as it passes through. From the first and second derivatives

$$y' = 2x - \frac{8}{x^2} \quad \text{and} \quad y'' = 2 + \frac{16}{x^3},$$

we can fill in the following table.

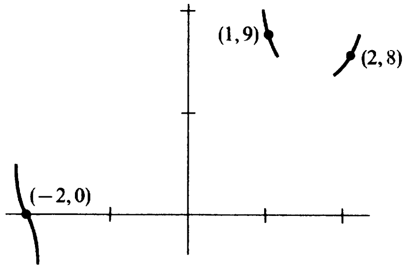
x	y	y'	y''
1	9	−6	18
2	8	2	4
−2	0	−6	0

It is really only the *sign* of y' and y'' that we need. From the first line of the table we see that y' is negative and y'' is positive. The curve is then falling and smiling as it goes through $(1, 9)$. It must resemble the curve sketched in figure (f) above, with $P = (1, 9)$.

From the second line of the table, with both derivatives positive, we see that the curve must resemble the one of figure (d), with $P = (2, 8)$.

The third line of the table, with y' negative, shows the curve falling as it passes through $(-2, 0)$, but how do we interpret the fact that y'' is zero? We must look at the sign of y'' for x just less than -2 and for x just greater than -2. Doing this reveals that the sign of y'' switches from positive to negative. This means that, at the point $(-2, 0)$, the curve stops bending left and starts bending right (or, if you prefer, the concavity switches from *up* to *down*). *Such a point, where the curve stops bending one way and starts bending the other way, is called a* **point of inflection**.

All the information gleaned above is in this picture:



The curve $y = x^2 + (8/x)$, pictured locally in neighborhoods of three points. The point $(-2, 0)$ is a point of inflection. □

At a point of inflection the second derivative must be zero (*why?*). However, the second derivative can be zero at points other than inflection points. A straight line has no inflection points, but the second derivative of a linear function is always zero.

Exercises

4.1. Consider the function given by $f(x) = x^2 - 3x + 2$. Proceed as in Example 4 to describe the local behavior of the curve f as it passes through
(a) $(0, 2)$.
(b) $(2, 0)$.
(c) $(1, 0)$.

4.2. Use the data collected in exercise 3.1 to describe the local behavior of the function f of Example 3 as it passes through $(2, \frac{11}{2})$, $(-1, -5)$, and $(\frac{1}{2}, \frac{5}{2})$.

4.3. Describe the local behavior of the curve $y = 10x - (4/x)$ near
(a) $(1, 6)$.
(b) $(2, 18)$.
(c) $(-\frac{1}{2}, 3)$.
Partial answer: The curve looks like figure (a), page 161, with $P = (1, 6)$.

4.4. Describe the local behavior of the curve $C = 7L + (48/L)$ near
(a) $(2, 38)$.
(b) $(2.5, 36.70)$.
(c) $(3, 37)$.

4.5. Describe the local behavior of the curve $y = x^2 + (8/x)$ near
(a) $(-1, -7)$.
(b) $(\frac{4}{3}, 7\frac{7}{9})$.
(*This is the curve of Example* 4.)

4.6. From the meager sketch in Example 4 above, we see that the curve appears to have a local minimum between $x = 1$ and $x = 2$. Find the x-coordinate of this local minimum. Is this a global minimum as well, i.e., is this the lowest point on the entire curve?

4.7. Does any linear function have an inflection point? Does any quadratic function have an inflection point?

4.8 Find both coordinates of an inflection point on the curve $y = x^2 + (1/x)$.

4.9. Suppose you have a function f and a point c where $f'(c) = 0$ and $f''(c)$ is positive. Have you found a local *minimum* or a local *maximum* of f?

§5. The Rule for Squares

There remain four shortcut rules to be discussed in this chapter. We need to find out how to take derivatives of *squares*, of *square roots*, of *products*, and of *quotients*. One might suspect that we shall therefore have to go through all the details of Fermat's method four more times. Fortunately, things can be arranged so that we have to do Fermat's method only once more, to find a rule for squares. The only algebraic trick we shall need is a simple one. The difference of two squares factors into the product of their sum and difference:

$$a^2 - b^2 = (a+b)(a-b).$$

Suppose we are presented with a function whose rule involves a square. For example, suppose we have a rule $f(x)$ given by $(x^2+3)^2$, or by $(5x - (7/x))^2$, or, more generally, by $(g(x))^2$, where g is a function whose derivative we know. By Fermat's method, if $f(x) = (g(x))^2$, then

$$\begin{aligned} f'(x) &= \operatorname*{Limit}_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \operatorname*{Limit}_{h\to 0} \frac{(g(x+h))^2 - (g(x))^2}{h}. \end{aligned}$$

How can this difference quotient be simplified, in order to find its limit? *Answer*: Since the numerator is the difference of two squares, we can factor it into the product of their sum and difference, to get

$$\begin{aligned} f'(x) &= \operatorname*{Limit}_{h\to 0} \frac{(g(x+h) + g(x))(g(x+h) - g(x))}{h} \\ &= \operatorname*{Limit}_{h\to 0} (g(x+h) + g(x)) \left(\frac{g(x+h) - g(x)}{h}\right) \\ &= (g(x) + g(x))g'(x) \quad \text{[by equation (4)]} \\ &= 2g(x)g'(x), \end{aligned}$$

provided, of course, that g has a derivative. We have our rule.

Rule for Squares
The derivative of $(g(x))^2$ *is* $2g(x)g'(x)$ *if the function* g *has a derivative.*

By this rule, for example,

the derivative of $(x^2+3)^2$ *is* $2(x^2+3)(2x)$,
the derivative of $(5x-(7/x))^2$ *is* $2(5x-(7/x))(5+(7/x^2))$.

Exercises

5.1. Find the derivative of $(3x+5)^2$ by using the rule for squares. *Answer*: Here we have $(g(x))^2$, where $g(x)=3x+5$. By the rule for squares, its derivative is $2gg'$, which is $2(3x+5)(3)$, or $18x+30$.
[*The reader should check that this is the same answer one obtains by first writing* $(3x+5)^2$ *as* $9x^2+30x+25$, *and then taking the derivative by the quadratic rule.*]

5.2. Find the derivative of $(6-7x)^3$
(a) by applying the rule for squares.
(b) by first squaring the expression $6-7x$ and then applying the rule for quadratics.

5.3. Find the derivative of each of the following functions, expressed as algebraic rules.
(a) $(x^2-5x)^2$.
(b) $(7L+(48/L))^2$.
(c) $(5x^2+(6/(x^2-2)))^2$.
(d) $(3x^3-5x+\sqrt{2})^2$.
(e) $(2x^2+(3/x)-4)^2$.
Answer: (b) $2(7L+(48/L))(7-(48/L^2))$.

5.4. True or false? *The derivative of a square is equal to the square of the derivative.*

5.5. Find the derivative of x^4 by regarding x^4 as $(x^2)^2$ and using the rule for squares. *Answer*: $4x^3$.

5.6. Find the derivative of $(1/x)^2$
(a) by applying the rule for squares.
(b) by first writing $(1/x)^2=1/x^2$ and then using the general reciprocal rule.

5.7. What is the derivative of
(a) $(x^2+5x)^2$?
(b) $(f(x)+g(x))^2$? *Answer*: $2(f(x)+g(x))(f'(x)+g'(x))$.

§6. The Product Rule and the Square Root Rule

It is *not true* that the derivative of a product is the product of the derivatives. The product rule is a little more complicated than that. It is easy to derive the product rule, though, because *a product can always be expressed in terms of squares*, and we already know the rule for squares.

To see the relationship between products and squares, begin with the familiar identity

$$(a+b)^2 = a^2 + 2ab + b^2$$

and "solve" this equation for the product ab. You get

$$ab = \frac{1}{2}((a+b)^2 - a^2 - b^2),$$

which expresses the product ab in terms of squares. The same thing holds, of course, for functions. If f and g are functions, then their product can be written as

$$fg = \frac{1}{2}((f+g)^2 - f^2 - g^2).$$

Taking derivatives by using the rule for squares, and also using the result of exercise 5.7, we find that

$$\begin{aligned}(fg)' &= \frac{1}{2}[2(f+g)(f'+g') - 2ff' - 2gg'] \\ &= (f+g)(f'+g') - ff' - gg' \\ &= ff' + fg' + gf' + gg' - ff' - gg' \\ &= fg' + gf'.\end{aligned}$$

Rule for Products
The derivative of $f(x)g(x)$ is $f(x)g'(x) + g(x)f'(x)$, provided that f and g have derivatives.

The reader may find it easier to remember the product rule by reading it in words: *The derivative of a product is equal to the first term times the derivative of the second, plus the "other way around".*

As an example, let us find the derivative of the product $(x+2)(x-3)$. By the product rule, it is

$$\begin{aligned}&(x+2)(x-3)' + (x-3)(x+2)' \\ &= (x+2)(1) + (x-3)(1) \\ &= 2x - 1.\end{aligned}$$

In this example we can check our answer by noting that the product $(x+2)(x-3)$ is equal to the quadratic expression $x^2 - x - 6$, whose derivative is indeed $2x - 1$. □

Here are some more examples, with the answers left in an unsimplified form.

The derivative of $(x^2 + x)(5x - 2)$ *is equal to*

$$(x^2 + x)(5) + (5x - 2)(2x + 1).$$

The derivative of $(7L + (48/L))(L^2 - \pi)$ *is equal to*

$$\left(7L + \frac{48}{L}\right)(2L) + (L^2 - \pi)\left(7 - \frac{48}{L^2}\right). \qquad \square$$

Now, what about square roots? How do we get the derivative of f, if the function f is given by $f(x) = \sqrt{1 + x^2}$? Or by $\sqrt{2x - 3}$? Or, more generally, by $\sqrt{g(x)}$, where g is some function whose derivative we already know? We can guess the answer to this question by using the rule for squares: If $f = \sqrt{g}$, then (by squaring both sides) we have

$$\begin{aligned} f^2 &= g, \\ 2ff' &= g' && \text{(by rule for squares)}, \\ f' &= \frac{1}{2f}\, g' && \text{(solving for } f'\text{)}, \\ f' &= \frac{1}{2\sqrt{g}}\, g' && \text{(since } f = \sqrt{g}\text{)}. \end{aligned}$$

Square Root Rule
The derivative of $\sqrt{g(x)}$ *is* $\dfrac{1}{2\sqrt{g(x)}}\, g'(x)$, *provided the function* g *has a derivative.*

The application of this rule is quite straightforward.

The derivative of $\sqrt{1 + x^2}$ *is*

$$\frac{1}{2\sqrt{1 + x^2}}(2x) = \frac{x}{\sqrt{1 + x^2}}.$$

The derivative of $\sqrt{2x - 3}$ *is*

$$\frac{1}{2\sqrt{2x - 3}}(2) = \frac{1}{\sqrt{2x - 3}}.$$

The derivative of $\sqrt{x}$ *is*

$$\frac{1}{2\sqrt{x}}(1) = \frac{1}{2\sqrt{x}}. \qquad \square$$

Exercises

6.1. Find the derivative of x^3, by regarding x^3 as the product of x^2 and x. *Answer*: $(x^3)' = (x^2x)' = x^2(1) + x(2x) = 3x^2$.

6.2. Find the derivative of x^4 by regarding x^4 as the product of x^3 and x. Does your answer agree with exercise 5.5?

6.3 Find the derivative of each of the following.
(a) x^5.
(b) x^6.
(c) x^7.
(d) x^n, where n is a positive integer.

6.4. *Any quotient can be expressed as a product.* Find the derivative of $x/(x+3)$, by regarding this quotient as the product of x and $1/(x+3)$.

Answer:
$$\begin{aligned}\left(\frac{x}{x+3}\right)' &= \left(x\frac{1}{x+3}\right)' \\ &= x\frac{-1}{(x+3)^2}+\frac{1}{x+3}(1) \\ &= \frac{-x}{(x+3)^2}+\frac{1}{x+3}.\end{aligned}$$

6.5. Find the derivative of $x^3/(5x+1)$ by regarding this quotient as the product of x^3 and $1/(5x+1)$.

6.6. Use the square root rule to find the derivatives of the following.
(a) $\sqrt{9+x^2}$.
(b) $17\sqrt{3x^2-2x}$.
(c) $\sqrt{\pi x}$.
(d) $\sqrt{2}$.

6.7 Find the derivative of $x^4\sqrt{1+x}$.

Answer:
$$\begin{aligned}(x^4\sqrt{1+x})' &= x^4(\sqrt{1+x})' + (\sqrt{1+x})(x^4)' \\ &= x^4\left(\frac{1}{2\sqrt{1+x}}\right) + (\sqrt{1+x})(4x^3) \\ &= \frac{x^4}{2\sqrt{1+x}} + 4x^3\sqrt{1+x}.\end{aligned}$$

6.8. Find the first derivatives of the following.
(a) $x^2\sqrt{1+x^2}$.
(b) $x\sqrt{x}$.
(c) $x^3\sqrt{2x-3}$.
(d) $x^6\sqrt{3x^2-2x}$.

6.9. Find the second derivative y'' if $y=\sqrt{2x+5}$.

§7. The Quotient Rule

Exercises 6.4 and 6.5 give the clue to finding a rule for obtaining the derivative of a quotient f/g: *Regard it as the product of* $1/g$ *and* f. Then we have

$$\begin{aligned}\left(\frac{f}{g}\right)' &= \left(\frac{1}{g}f\right)' \\ &= \left(\frac{1}{g}\right)f' + f\left(\frac{1}{g}\right)' \\ &= \left(\frac{g}{g^2}\right)f' + f\left(\frac{-1}{g^2}\right)g' \\ &= \frac{gf' - fg'}{g^2}.\end{aligned}$$

Rule for Quotients
The derivative of $f(x)/g(x)$ is

$$\frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}$$

if the functions f and g have derivatives.

The reader may find it easier to remember the quotient rule by reading it in words: *The derivative of a quotient is equal to the bottom times the derivative of the top,* minus *the other way around, over the bottom squared.*

For example,

the derivative of $x/(x+3)$ is

$$\frac{(x+3)(1) - (x)(1)}{(x+3)^2},$$

the derivative of $x^3/(5x+1)$ is

$$\frac{(5x+1)(3x^2) - (x^3)(5)}{(5x+1)^2}.$$

(The reader should check that these answers, with the help of a little algebra, may be seen to agree with the answers to exercises 6.4 and 6.5.) □

Exercises

7.1. If $y = 2x/(x^2-3)$, what is y'? *Answer*: $y' = (-2x^2-6)/(x^2-3)^2$.

7.2. If $y = 2x/(x^2-3)$, what is y''? *Hint.* Use the quotient rule to find the derivative of the answer to exercise 7.1. In the course of doing this, the rule for squares will come in handy in finding the derivative of the bottom. Don't take time to simplify your answer.

7.3. Describe the local behavior of the curve $y = 2x/(x^2 - 3)$ near the point $(2, 4)$ and near $(1, -1)$. Is $(0, 0)$ an inflection point?

7.4. Find the first derivative of each of the following. Do not simplify.
(a) $(x + 2)/(x - 2)$.
(b) $(x^2 - 3)/(x - 2)^2$.
(c) $x^3/\sqrt{1 + x^2}$.
(d) $\sqrt{(x + 2)/(x - 2)}$.
(e) $((3x - 1)/x^2)^2$. *Answer*: $2((3x - 1)/x^2)((x^2(3) - (3x - 1)(2x))/x^4)$.

7.5. Consider the function given by $f(x) = 5x^2 + (6/(x^2 - 2))$. What does the curve f look like, locally, near the point $(0, -3)$? Near $(2, 23)$?

7.6 What does the curve $y = \sqrt{16 + x^2}$ look like, locally, as it passes through $(0, 4)$?

§8. Solving Optimization Problems

Where are we now? We have just completed an unavoidable digression from our original theme, which was the solution of optimization problems. As we saw in Chapter 1, an optimization problem leads to the problem of finding the highest (or lowest) point on a certain curve. This, in turn, has led to the study of derivatives, because derivatives cast light on the behavior of a curve. And now, at last, we know how to bypass Fermat's method and use the following rules instead.

(1) $(cf)' = c \cdot f'$ (constant multiples).
(2) $(f + g)' = f' + g'$ (sums).
(3) $(1/g)' = (-1/g^2)g'$ (reciprocals).
(4) $(g^2)' = 2gg'$ (squares).
(5) $(\sqrt{g})' = \frac{1}{2}g'/\sqrt{g}$ (square roots).
(6) $(fg)' = fg' + gf'$ (products).
(7) $(f/g)' = (gf' - fg')/g^2$ (quotients).

There is also the useful *power rule* derived in exercise 6.3: $(x^n)' = nx^{n-1}$, where n can be any positive integer. The applicability of this rule when n is not a positive integer will be tested in problem 32 at the end of this chapter.

The reader should practice using these rules until they have been memorized. Then the taking of derivatives will be quite a routine matter, and the most important step in solving an optimization problem will have been mastered.

We can finally come to grips with the topic to which the title of this chapter alludes. What are the steps leading to the solution of an optimization problem? Basically, there are just two steps. First, translate the problem into the geometric problem of finding the highest (or lowest) point on a certain curve f; and second, find f' and use it as an aid in understanding how the curve f behaves.

The critical points to be found in sketching a curve f are those where the tangent line to the curve is horizontal. [That leads to a definition: *To say that x is a* **critical point** *of f is to say that* $f'(x) = 0$.] Usually, although not always, the function f will attain its optimal value at a critical point.

To verify whether the optimum has been found, make a rough sketch of the curve near each critical point (the second derivative is helpful here) and near each endpoint of the domain.

Remember the idea of seeking wrong answers in order to eliminate them? Here is an application that shows the power of calculus (and shows the major fallacy in the sophistry of problem 29 of Chapter 3).

Theorem on Optimization and Elimination
In searching for the largest and smallest values in the range of f on the domain $a \leq x \leq b$, we may eliminate from consideration those values of x where $f'(x)$ is nonzero, unless x is one of the endpoints (a or b). *Restated positively*: If f has a derivative, then in searching for the extreme values of f, we need only check the values of f at the endpoints of its domain and at critical points inside.

Proof
If $f'(x)$ is nonzero then the curve f is either rising or falling as it passes through the point $(x, f(x))$, so such a point (if it is not an endpoint of the curve) can be neither the highest nor lowest point on the curve and may be eliminated from a search for these points. □

As we have seen, some curves do not have a highest (or lowest) point. It can be proved, however, that a curve must have such points if it comes from a *continuous* function and if the domain is an interval *containing its endpoints*. This is a deep theorem of analysis, the modern branch of mathematics into which seventeenth-century calculus evolved, and cannot be proved here. The moral for us is to be aware of what a function is doing near the endpoints of its domain, particularly if the domain does not include endpoints. If a continuous curve fails to have a highest (or lowest) point, then by the theorem of analysis the trouble must lie in the behavior of the function near an endpoint missing from its domain.

EXAMPLE 5
Find the highest point on the curve f given by

$$f(x) = 2x + 3,$$

on the domain

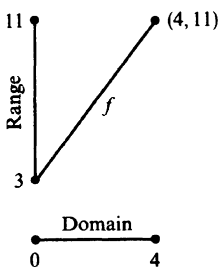
(a) $0 \leq x \leq 4$.
(b) $0 < x < 4$.

Let us look first for all critical points in the domain, that is, all values x for which $f'(x) = 0$. Here we have $f'(x) = 2$, which shows that there are no such values. Since f has no critical points, the principle of analysis mentioned above guarantees that the extreme values of f must occur at the endpoints of the domain. At the endpoint 0, the value of f is 3; at the endpoint 4, the value of f is 11. Therefore,

(a) if the domain is $0 \le x \le 4$, then $(4, 11)$ is the highest point on the curve f.
(b) if the domain is $0 < x < 4$, then the curve f contains no highest point.

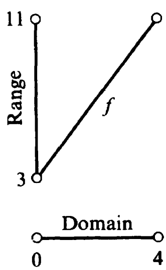
Note that, to draw the conclusions (a) and (b), *we did not have to draw a picture of the curve f*! The reader may wish to draw a picture anyway, to see better what is going on. The expression $2x + 3$ reveals f to be a linear function of slope 2:

(a) Domain: $0 \le x \le 4$
Range: $3 \le y \le 11$.



The highest point on the curve is $(4, 11)$. The greatest number in the range is 11; the least is 3.

(b) Domain: $0 < x < 4$
Range: $3 < y < 11$.



There is no highest (or lowest) point on the curve, because the range contains no greatest (or least) number

EXAMPLE 6
Let $C = 7L + (48/L)$, with domain $0 < L$. Find the least possible value of C.

Let us first find all critical points in the domain, that is, points where C' is zero. Since $C' = 7 - (48/L^2)$, C' is zero when

$$7 - \frac{48}{L^2} = 0,$$

$$7L^2 - 48 = 0 \quad \text{(multiplying through by } L^2\text{)},$$

$$7L^2 = 48,$$

$$L^2 = \frac{48}{7},$$

$$L = \pm\sqrt{\frac{48}{7}}.$$

L	C	C'
?		0

Because $-\sqrt{48/7}$ is not in the domain, the only critical point in the domain is

$$\sqrt{\frac{48}{7}} = 2.619\ldots. \quad \text{[by exercise 6.6 of Chapter 2]}$$

At the critical point, the corresponding value of C is given by

$$C = 7\sqrt{\frac{48}{7}} + \frac{48}{\sqrt{\frac{48}{7}}} = 36.661\ldots.$$

We must now show that this is the least possible value of C.

The second derivative helps here. It is given by

$$C'' = \frac{96}{L^3},$$

which is (obviously) *positive* for *all* values of L in the domain $0 < L$. Therefore, the curve $C = 7L + (48/L)$ is *always* bending to its left (or smiling). The point

$$(2.619\ldots, 36.661\ldots),$$

being the point on the curve where a tangent line is horizontal, must be the lowest point on the curve.

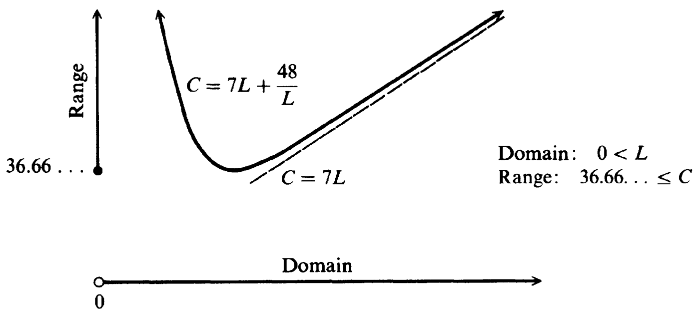
We can now answer the question raised in Example 1 of Chapter 1. The least amount of money that will pay for the fencing is \$36.66 (rounded off to the nearest cent). □

The preceding example was discussed rather thoroughly without ever drawing the curve. We found the lowest point and we discovered that the curve was always bending to the left. If it is desired to sketch the curve, what additional information is needed? *Answer*: Information about the curve's behavior near the "endpoints" of the domain, i.e., when L is very small and when L is very large.

To get this information, use common sense. Look at the two terms $7L$ and $48/L$, whose sum gives C. What happens to each of them when L is very small? The first term $7L$ is *negligible* (i.e., nearly zero), so the curve behaves essentially like the graph of $48/L$ when L is small. The expression $48/L$ increases without bound as $L \to 0^+$. (See Example 7, near the end of Chapter 1.)

What happens when L is very large? Then the expression $48/L$ is negligible, so the curve behaves essentially like the graph of $7L$ when L is large. The expression $7L$ produces a line of slope 7. For large L, the curve $C = 7L + (48/L)$ approximates a straight line of slope 7.

Putting these facts together produces the following sketch:



Exercises

8.1. For each of the following functions, find its maximum value, if it has one.
(a) $f(x) = 5 - 2x, 0 \le x < 3$. *Hint.* Proceed as in Example 5.
(b) $F(x) = x^2 - 2x, 0 \le x \le 4$.
(c) $g(x) = x - (1/x) + 6, 0 < x \le 8$.
(d) $G(x) = x + (1/x) + 6, 0 < x \le 8$.
Answers: (b) max F is 8. (d) max G does not exist.

8.2. For each of the functions in exercise 8.1, find its minimum value, if it has one. *Answers*: (a) min f does not exist. (b) min F is -1.

8.3. Sketch the curve $y = 4x + (36/x)$, with domain $0 < x$, indicating both coordinates of the lowest point. Also, indicate how the curve looks when x is very small and when x is very large.

8.4. A rectangular pen containing 48 square meters is to be fenced in. The front will cost \$5 per meter of fencing, while each of the other three sides will cost \$3 per meter. What is the least amount of money that will pay for the fencing?

8.5. In the problem set at the end of Chapter 1, read again problem 19. Find the greatest possible volume V. *Answer*: max $V = 40\sqrt{10}$, attained at critical point $L = \sqrt{40}$.

§9. Summary

Here, in detail, are the steps that have been illustrated above.

Step 1. Algebraic formulation:

(a) See the problem in terms of variables. (The quantity to be optimized is one variable, say y, and you have to find a second variable, say x, on which y depends.)

(b) Write down an algebraic rule f, giving y in terms of x.

(c) Specify the domain of the function f.

Step 2. Geometric analysis:

(a) See the problem as one of finding the highest (or lowest) point on the curve f.

(b) Find the derivative f'. (And find f'' too, if it can be done without much trouble.)

(c) Find the critical points, if any, that lie in the domain of f. (That is, find all values of x in the domain of f that satisfy the equation $f'(x) = 0$.)

(d) Check what happens near the endpoints of the domain.

(e) Using the information of steps 2(c) and 2(d), find the desired highest (or lowest) point on the curve f.

[The second derivative may be helpful in steps 2(d) and 2(e).]

Step 3. Back to everyday life:

(a) Read the problem again, to determine exactly what was called for. (Was it the *first* or *second* coordinate, or *both*, of the *highest* or *lowest* point of the curve that you were seeking?)

(b) Give a direct answer to the question raised in the problem, by writing a complete, concise sentence.

Step 1(c) is easy to forget, and thus deserves emphasis. The domain must be specified; otherwise, steps 2(c) and 2(d) cannot be carried out. Step 3 is also easy to forget. In concentrating on step 2, you can lose sight of your goal and, as a consequence, do unnecessary work. When a problem takes a long time to work, it is a good idea to remind yourself now and then what you are after.

Here is another example to illustrate these steps.

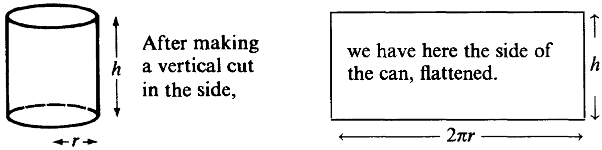
EXAMPLE 7

An ordinary metal can (shaped like a cylinder) is to be fashioned, using 54π square inches of metal. What choice of radius and height will maximize the volume of the can?

Here, we want to maximize the volume, so let V denote the volume, which is given in terms of the radius r and height h by the formula

$$V = (\text{area of base})\,(\text{height})$$
$$= \pi r^2 h. \tag{7}$$

The rule $V = \pi r^2 h$ gives V in terms of *two* variables. We need to get V in terms of only *one* variable, and this can be done, as follows, by finding a relation between r and h. The picture below shows that the area of the side of the can is given by $2\pi rh$:



The total amount of metal available, 54π square inches, must equal the amount in the side of the can, plus the amount in the circular top and bottom:

$$54\pi = 2\pi rh + 2\pi r^2.$$

This is a relation between r and h. It is easy to solve for h (the reader is asked to do it), and obtain

$$h = \frac{27 - r^2}{r}. \tag{8}$$

Putting equations (7) and (8) together gives

$$\begin{aligned} V &= \pi r^2 \left(\frac{27 - r^2}{r}\right) \\ &= \pi r(27 - r^2) \\ &= 27\pi r - \pi r^3, \end{aligned}$$

which expresses V in terms of r alone. The problem now is to find the value of r that yields the maximal volume V, where

$$V = 27\pi r - \pi r^3, \quad 0 < r < \sqrt{27}.$$

[The radius r must be less than $\sqrt{27}$. *Reason*: The height h must be positive, so, by equation (8), $27 - r^2$ must be positive.]

r	V	V'
?		0
r	$27\pi r - \pi r^3$	$27\pi - 3\pi r^2$

Let us find critical points. The derivative is given by

$$V' = 27\pi - 3\pi r^2,$$

which is zero when (dividing by 3π)

$$0 = 9 - r^2,$$
$$r^2 = 9,$$
$$r = \pm 3.$$

Since -3 is not in the domain, the only critical point is 3.

We now show that when r is 3, the volume V is maximal. This is easy to see, for the second derivative is given by

$$V'' = -6\pi r,$$

which is (obviously) negative *throughout the domain*. The curve is therefore always bending to its right (or frowning), and hence it must reach its highest point at the place where it has a horizontal tangent line. (At both endpoints of the domain, V tends to zero.)

To maximize the volume, the radius should be 3 inches, and the corresponding height, by equation (8), should be 6 inches. □

A Final Remark. As in Examples 5, 6, and 7, it is not really necessary to sketch the curve in order to do the problem. If f has a derivative, then the *extreme values* (maximum and minimum) can be located by checking among the endpoints and the critical points. Curve sketching is to be encouraged, because pictures say more than words, but the principles of analysis are valid regardless of how well one draws.

Problem Set for Chapter 5

1. A rectangular pen bordering a road is to be fenced in. The fence along the road will cost \$7 per meter, while each of the other three sides will cost \$3 per meter.
   (a) What is the minimum cost of the fencing if the pen is to contain 36 square meters?
   (b) What is the maximum area that can be enclosed by spending \$120 on fencing?

2. A book company wants to put 60 square inches of type on a rectangular page, leaving margins of 1 inch on the sides and bottom and of 2 inches at the top. What should be the dimensions of the page in order to minimize the amount of paper used?

3. Write an equation of the tangent line to the curve $y = 5/x^2$ at the point $(1, 5)$.

4. Tell whether the curve $y = 5/x^2$ is bending to the *right* or to the *left* as it passes through
   (a) $(1, 5)$.
   (b) $(-1, 5)$.

5. Consider the function f defined by $f(x) = 5x\sqrt{1-3x}$.
   (a) Is the curve f *rising* or *falling* as it passes through the point $(0, 0)$?
   (b) Is the curve f bending to the *left* or to the *right* as it passes through the point $(0, 0)$?

6. Consider the function given by $f(x) = ax^2 + bx + c$, where $a, b,$ and c are constants. Which way does the curve f bend if
   (a) $a > 0$?
   (b) $a < 0$?
   (c) $a = 0$?

7. Consider the curve $C = 3L + (27/L), 0 < L$.
   (a) Find the first coordinate of a point on this curve where the tangent line is horizontal.
   (b) This curve always bends the same way on the domain $0 < L$. Which way?
   (c) From your answer to part (b), you know the point found in part (a) must be the *highest* or *lowest* point on the curve?

8. Consider the curve $C = 3L + (27/L), L < 0$.
   (a) This curve always bend the same way on the domain $L < 0$. Which way?
   (b) Find both coordinates of the highest point on the curve.
   (c) Does this curve have a lowest point?

9. What is the range of f, where $f(L) = 3L + (27/L)$, with domain $L \neq 0$? *Hint.* First sketch the curve f, using the information obtained in problems 7 and 8.

10. (*A problem in curve sketching.*) Consider the cubic equation $y = x^3 - 3x + 2$.
    (a) The derivative $y' = 3x^2 - 3$ is a simple quadratic function. Plot the graph of this quadratic, and, on a different coordinate system, plot the graph of the simple linear function $y'' = 6x$.
    (b) The two graphs just sketched give much information about the original cubic. Use these two graphs to specify on what interval(s) the cubic is
        (i) rising.
        (ii) falling.
        (iii) bending to the left.
        (iv) bending to the right.
    (c) Find both coordinates of an inflection point of the cubic.
    (d) There are two points on this cubic where there is a horizontal tangent line. Find both coordinates of both points.
    (e) Sketch the curve $y = x^3 - 3x + 2$, using all the information just obtained.
    (f) Specify the range of the cubic if the domain is
        (i) $0 \leq x \leq 3$.
        (ii) $-2 \leq x < 0$.
        (iii) $-2 < x \leq 0$.
        (iv) unrestricted.

11. Consider the cubic equation $y = x^3 + 3x + 2$.
    (a) Sketch the graph of this cubic, after first investigating its first and second derivatives, as in the preceding problem.
    (b) Find both coordinates of an inflection point.
    (c) Specify the range of this cubic if its domain is given by $-1 \le x \le 4$.

12. As in the preceding two problems, carry out an analysis of the cubic $y = x^3 + 2$, sketch its graph, and find its point of inflection. What is its range if its domain is given by $-1 \le x < 2$?

13. Suppose $y = (t^2 - 3)/(t + 2)$. Find y' and y'' when t is 0, and use this information to sketch the curve locally, near the point $P = (0, -\frac{3}{2})$.

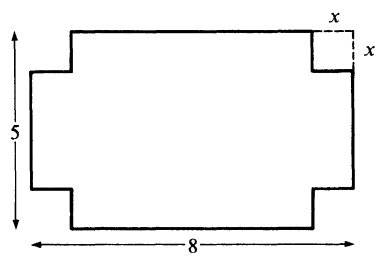
14. Sketch the curve $y = x^2 + (8/x), x \neq 0$, a portion of which has already been sketched in Section 4.

15. Express the number 10 as the sum of two positive numbers in such a way that the sum of the cube of the first and the square of the second is as small as possible.

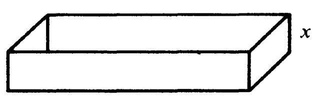
16. Find the point on the graph of $y^2 = 4x$ that is nearest the point $(2, 1)$.

17. What is the smallest slope that a tangent line to the curve $y = x^3 + 3x + 2$ could possible have?

18. Identical squares are to be cut out of each corner of a piece of metal that is shaped like a rectangle of dimensions 5 feet by 8 feet. The four squares are then discarded, and the sides folded upwards to make a large box, with open top. Let x be the length of the sides of the squares cut out, and let V be the corresponding volume of the box. Find the value of x that maximizes V.



Discard squares, fold up sides:



19. Identical squares are to be cut out of each corner of a rectangular piece of metal measuring 10 meters by 4 meters. Then the squares are to be dis-

carded, and the sides folded up to make a water trough for thirsty horses. What size squares should be cut out in order to maximize the volume of the trough?

20. Out of 100 square centimeters of metal, the sides, top, and bottom of a cylindrical can are to be fashioned. What should be the radius of the base of the can in order to maximize the amount of chicken soup that the can will hold?

21. In exercise 4.11 of Chapter 2, a certain cost C was given in terms of a length x by the equation

$$C = 7\sqrt{9+x^2} - 2x + 26, \quad 0 \leq x \leq 13.$$

(a) Find C', then fill in the table below.

x	C	C'
0	47	?
4	53	?

(b) Tell whether the curve is *rising* or *falling* as it passes through $(0, 47)$ and through $(4, 53)$. Can you conclude that the lowest point on the curve lies somewhere between these two points?
(c) Find the value of x that yields the least cost C.
(d) Read again exercise 4.11 of Chapter 2. Then draw a picture of how the cable should be built in order to minimize the cost of the cable.

22. A lighthouse is located 4 miles offshore. The nearest town is 5 miles downshore. Whenever she goes into town, the lighthouse keeper must take a motorboat containing her motorcycle, dock at a point somewhere downshore, then ride the rest of the way by motorcycle. Where should the boat be docked in order to minimize the time of the trip to town if
(a) the motorboat goes 20 miles per hour and the motorcycle goes 40 miles per hour?
(b) the motorboat and the motorcycle travel at the same speed?
(c) the motorboat goes A miles per hour and the motorcycle goes B miles per hour?

23. In problem 18 at the end of Chapter 1, a certain cost C was given in terms of a length L by the equation

$$C = 3L^2 + \frac{96}{L}, \quad 0 < L.$$

(a) Find the value of L that minimizes the cost.
(b) Read again problem 18 of Chapter 1, and draw a picture indicating the dimensions of the metal container that will minimize its cost.

24. Find the dimensions of the cheapest possible trash can with square base and rectangular sides, subject to the following specifications. The volume of the can is to be 3 cubic meters, the material for the sides costs \$0.30 per square meter, and the material for the base costs \$0.50 per square meter.

25. In the preceding problem, suppose it is decided to add a top to the can, made out of light metal costing only \$0.10 per square meter. With this addition, what are the dimensions of the cheapest can?

26. The metal used in making the top and bottom of a *cylindrical* can will cost \$0.03 per square centimeter, while the metal used in the side of the can will cost \$0.02 per square centimeter. If the volume of the can is to be 100 cubic centimeters, what should be the dimensions of the can in order to minimize the cost?

27. Out of 160 square feet of material, a container is to be made. What dimensions will maximize the volume of the container if the container is to be shaped like
    (a) a rectangular figure with square base and open top?
    (b) a rectangular figure with square base and with a top?
    (c) a cylindrical can without a top?
    (d) a cylindrical can with a top?

28. A Norman window is in the shape of a rectangle surmounted by a semicircle. Find the dimensions of the window that will allow the most light to pass, provided that the perimeter of the window is 8 meters.

29. A wire is to be cut in two. The first part is to be bent into the circumference of a circle, and the second part into the perimeter of a square. How should the wire be cut in order to minimize the combined area of the circle and square if
    (a) the wire is 100 centimeters long?
    (b) the wire is A centimeters long?

30. The definition of *concave upward* is as follows: *A curve f is* **concave upward** *if it lies above each tangent line (with the obvious exception of the point of tangency).*
    (a) If f'' is always positive, is the curve f concave upward?
    (b) If the area in the plane lying up above the curve f forms a concave figure, is the curve f concave upward?

31. Find the first derivatives of the following. Do not simplify you answers.
    (a) $(x/(x-6))^2$
    (b) $\sqrt{x/(x-6)}$.
    (c) $(4x^5 - 3x)(x^2 - x + \sqrt{2})^2$.
    (d) $x^4/(x^3 + x - 3)^2$.
    (e) $(x^5 - x)\sqrt{2 + 7x}$.
    (f) $\sqrt{x^5}$.

32. Mathematics shares some characteristics with experimental science. One notices a pattern developing, and then one tries to guess a general rule. The rule must then be tested for its applicability to new situations. One hopes that a widely applicable rule can be derived logically from simpler principles that are already accepted. Consider the rule

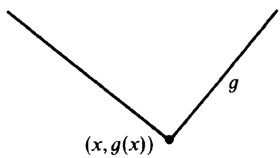
$$(x^n)' = nx^{n-1},$$

which ought to have been guessed in exercise 6.3 of this chapter. This rule applies where n is a *positive integer*. Let us test this rule for wider applicability.

(a) Apply the rule above to find $(x^{-1})'$. Does it result in the correct answer? (We already know that the derivative of x^{-1} is $-1/x^2$, from our work in Section 1.)
(b) Apply the rule to fine $(x^{-2})'$. Does it give the derivative of $1/x^2$?
(c) Apply the rule to find $(x^{1/2})'$. Does it give the derivative of $\sqrt{x}$?
(d) Apply the rule to find $(x^{3/2})'$. Does it give the derivative of $x\sqrt{x}$? (The derivative of $x\sqrt{x}$ can be taken by the product rule.)
(e) Apply your rule to find $(x^{5/2})'$. Does it give the derivative of $\sqrt{x^5}$?
(f) Apply your rule to find $(x^{-7/2})'$. Does it give the derivative of $1/\sqrt{x^7}$?
(g) If it made any sense to speak of raising a number to the power π, what would you guess is the derivative of x^π?

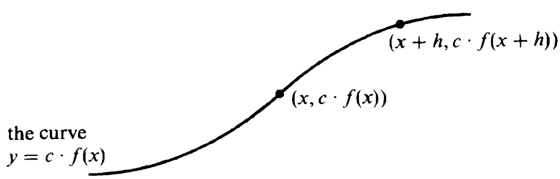
33. Suppose that a function g has a derivative at a point x. Does it necessarily follow that g is *continuous* at x? That is, if $g'(x)$ exists, does it follow that $g(x) = \operatorname{Lim} g$ at x? *Hint.* In Section 2 we saw that if $g'(x)$ exists, then equation (4) necessarily follows.

34. Suppose that a function g is continuous at a point x. Does it necessarily follow that g has a derivative at x? *Hint.* Consider a function whose curve has a "corner", as pictured below, at the point $(x, g(x))$.

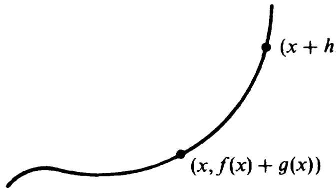


35. Use Fermat's method to show that the derivative of a constant function is zero. [This is so easy that it is easy to miss. You are to show that if $g(x) = c$, where c is a constant, then $g'(x) = 0$ for all x.]

36. Give two proofs of the rule for constant multiples, which states that $(c \cdot f)' = c \cdot f'$.
(a) First, by applying the product rule to the product $f \cdot g$, where $g(x) = c$, and using the result of problem 35.
(b) Secondly, by applying Fermat's method to the situation pictured below:



37. By applying Fermat's method to the situation pictured below, prove the rule for sums, which states that $(f+g)' = f' + g'$.



38. Use Fermat's method (not a shortcut rule) to show that if $f(x) = \sqrt{x}$, then $f'(x) = 1/2\sqrt{x}$. *Hint.* Simplify the difference quotient of f by multiplying both the top and the bottom by the expression $\sqrt{x+h} + \sqrt{x}$.

39. Use Fermat's method (not a shortcut rule) to prove directly the product rule, which states that $(fg)' = fg' + gf'$. *Hint.* First find the difference quotient of the product function given by $y = f(x)g(x)$. Simplify it by inserting the expression $f(x+h)g(x) - f(x+h)g(x)$ into the numerator to get

$$f(x+h)\frac{g(x+h)-g(x)}{h} + g(x)\frac{f(x+h)-f(x)}{h}.$$

Then find the limit as h tends to 0.

40. Derive the reciprocal rule from the product rule, by proceeding as follows. Assuming that the expression $1/g$ has a derivative, begin with the obvious equality

$$g\left(\frac{1}{g}\right) = 1,$$

and use the product rule, together with the result of problem 35, to write

$$g\left(\frac{1}{g}\right)' + \left(\frac{1}{g}\right)g' = 0$$

then solve for the derivative of $1/g$.

41. (*Newton's method again*) Now that we know the derivative of x^{12}, we can use it to calculate the decimal expansion of the twelfth root of 2, which plays a crucial role in the theory of music. Carry this out by applying Newton's method (Chapter 4, Section 10) to solve the equation $x^{12} - 2 = 0$, taking $G = 1$ as your initial guess. Do you get the same succession of approximations to $\sqrt[12]{2}$ that you did in exercise 6.8 of Chapter 2?

42. (*A cube-root rule?*) Given $f = \sqrt[3]{g}$, is there a "cube-root rule" enabling us to find f' quickly? The way we came upon the square-root rule in Section 6 can be modified to do this.
   (a) If $f = \sqrt[3]{g}$, then of course, $f^3 = g$, so $g = f^3 = f^2 f$. What is g'? *Hint.* Since $g = f^2 f$, then $g' = (f^2 f)'$. Use the product rule to find $(f^2 f)'$, but be careful; you'll have to use the rule for squares in the course of carrying out the product rule.

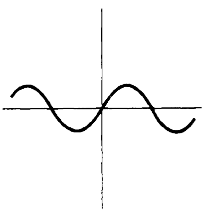
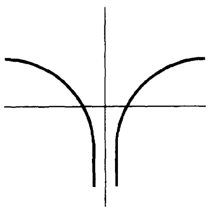
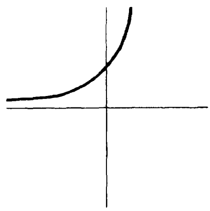
(b) We are after f' here. Take your equation $g' = 3f^2f'$ derived in part (a) that gives g' in terms of f and f' and "solve" it for f'.

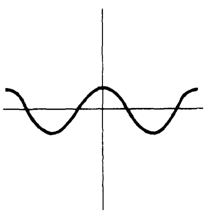
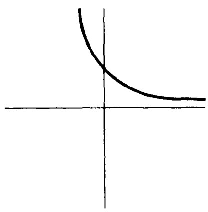
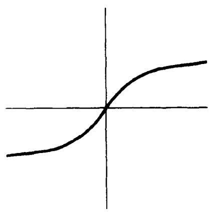
(c) From part (b) you should have an equation expressing f' in terms of g' and f. If in this equation you replace f by $\sqrt[3]{g}$, then you have f' expressed in terms of g and g'. This gives you the cube-root rule. Write it in a complete sentence: "If $f = \sqrt[3]{g}$, then $f' = \ldots$."

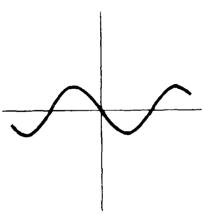
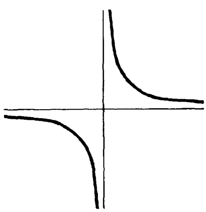
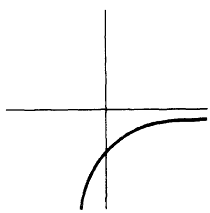
(d) Apply your rule derived in part (c) to the simple case when $g(x) = x$: If $f(x) = \sqrt[3]{x}$, then $f'(x) = ???$

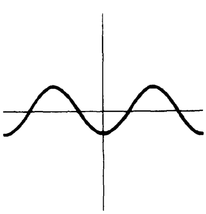
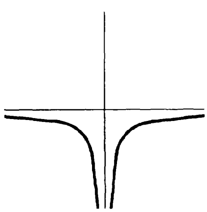
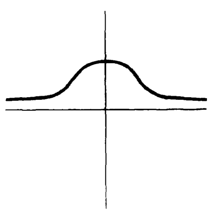
(e) In the spirit of problem 32, *make a guess*: If $f(x) = x^{1/3}$, then $f'(x) = ???$ Does your guess agree with your answer to (d)?

43. Match each of the following functions (a) through (j) with its derivative. [The derivatives of (k) and (l) are not pictured.]

(a) 
(b) 
(c) 

(d) 
(e) 
(f) 

(g) 
(h) 
(i) 

(j) 
(k) 
(l) 

44. The rules for sums, products, etc. enable us to find derivatives of complicated functions made up out of simpler functions whose derivatives we know. Suppose we are given the following data concerning functions f and g.

f

x	y	y'	y''
1	2	3	4
2	3	5	−2
3	1	−2	1

g

x	y	y'	y''
1	1	−2	3
2	3	0	5
3	2	−1	0

Apply these rules properly to the data above to fill in the question marks in the tables for the function f^2, $\sqrt{g}$, fg, and f/g.

f^2

x	y	y'	y''
1	4	12	34
2	?	?	?
3	?	?	?

(a)

$\sqrt{g}$

x	y	y'	y''
1	?	?	?
2	?	?	?
3	$\sqrt{2}$	$\frac{-1}{2\sqrt{2}}$	$\frac{-1}{8\sqrt{2}}$

(b)

fg

x	y	y'	y''
1	?	?	?
2	9	15	9
3	?	?	?

(c)

f/g

x	y	y'	y''
1	2	7	26
2	?	?	?
3	?	?	?

(d)

45. Using the data given in problem 44, describe the local behavior of the curves f, g, and fg when x is 1, 2, and 3. (Sketch nine rough pictures describing each of these three curves near (x, y) for each of these three values of x, indicating whether the curves are rising or falling, "smiling" or "frowning".)

46. Using the data given in problem 44, describe the local behavior of the curves f^2, $\sqrt{g}$, and f/g when x is 1, 2, and 3. (Sketch nine rough pictures, as in the preceding problem.)

47. Using the data given in problem 44, describe the local behavior of the curves $1/f$ and $1/g$ when x is 1, 2, and 3. (Sketch six rough pictures.)

48. (*We still need one more rule.*) Given two functions f and g, we can *compose* the functions by first applying g to get $g(x)$ and then applying f to get $f(g(x))$. The new function sending x to $f(g(x))$ is denoted $f \circ g$ and called the composition of f with g. Notice that if we take the composition in the other order, using $g \circ f$ to send x to $g(f(x))$, we don't necessarily get the same result. Given the information on f and g in problem 44, fill in the question marks below. *Do not yet attempt to fill in the blanks giving values of the derivatives. You will be asked to do this after learning the chain rule in Chapter 6.*

$f \circ g$

x	y	y'	y''
1	?	—	—
2	?	—	—
3	3	-5	-2

$g \circ f$

x	y	y'	y''
1	?	—	—
2	?	—	—
3	?	—	—

49. Write a thoughtful response to the sophistry presented in the final problem in Chapter 3, explaining clearly how the theorem on optimization and elimination of Section 8 helps bring to light the major fallacy in this argument.

CHAPTER 6

Chains and Change

Things change. The world is in flux. How can one understand a world in which change plays so great a role? The seventeenth-century answer given by Leibniz and Newton is simplicity itself:

Study *change.*

To study change is to study the way things vary. We have done a little of this in the preceding chapters, but we have not yet taken up this study in earnest. The derivative has a remarkable ability to capture the dynamics of change. A main point of this chapter is that *the derivative may be viewed as measuring the instantaneous rate of change.*

How can this be? Before answering, we need to develop symbolism that is suggestive of the ideas involved. The symbolism of *primes* (as in y' or f') to denote the derivative offers no aid to our new endeavor. In fact, the main advantage of denoting the derivative by f' is that this notation suggests that the derivative is a *function*. Once this important fact has been hammered home, the use of primes to denote derivatives offers no special advantage, and may be discarded in the presence of a superior system of symbolism.

§1. Leibniz's Notation: Mathematics and Poetry

A superior system of notation for the calculus was developed by Leibniz. If y is a function of x, Leibniz denoted the derivative by

$$\frac{dy}{dx},$$

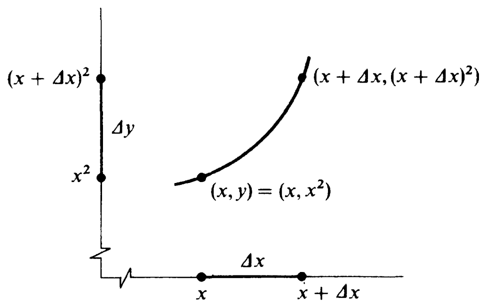
instead of by y'. Or, if A is a function of s, Leibniz called the derivative

$$\frac{dA}{ds},$$

instead of A'.

At this point the reader is doubtless mystified as to why this symbolism is supposed to be more helpful than the perfectly good notation already developed. It is helpful only if one views the derivative the way Leibniz did. Let us illustrate how Leibniz would go about showing that *the derivative of* x^2 *is* $2x$.

Consider a fixed point (x, y) on the curve $y = x^2$.



A portion of the curve $y = x^2$

Let Δx be a small change in the variable x. ("Δ" is the Greek letter *delta*. The expression Δx is to be taken as a whole, and not to be confused with a product. The change Δx may be either positive or negative.) *What is the corresponding change* Δy *in the variable* y? From the figure, it is clearly given by

$$\begin{aligned} \Delta y &= (x + \Delta x)^2 - x^2 \\ &= x^2 + 2x(\Delta x) + (\Delta x)^2 - x^2 \\ &= 2x(\Delta x) + (\Delta x)^2. \end{aligned}$$

Therefore, the ratio of the change (or *increase*) in y to the increase in x which caused it is given by

$$\frac{\Delta y}{\Delta x} = 2x + \Delta x. \tag{1}$$

As mentioned above, the increase Δx may be either *positive* or *negative* (a *negative increase* of course represents a decrease), but may not be zero.

What happens in equation (1) as Δx tends to 0? Then the fraction $\Delta y/\Delta x$ approaches what might be termed the *instantaneous rate of increase of y with respect to x*, which, using equation (1), is equal to

$$\operatorname*{Limit}_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \operatorname*{Limit}_{\Delta x \to 0} (2x + \Delta x) = 2x.$$

(The reason that we get the derivative of x^2 should be plain. As seen in the figure above, the ratio of changes $\Delta y/\Delta x$ is also the slope of a line that approaches the tangent line at (x, y) when Δx approaches zero.)

The ratio $\Delta y/\Delta x$ of changes taking place over an interval of length Δx is *not* of primary interest here. Leibniz wanted the "ultimate ratio", or the *instantaneous* rate of increase taking place at the point x. This is what happens as the length Δx shrinks to zero, and this Leibniz called dy/dx. That is, the symbol dy/dx is defined as follows:

$$\frac{dy}{dx} = \operatorname*{Limit}_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

Why did Leibniz choose to denote the derivative this way? What is in the symbol dy/dx that is not in the name *derivative*? A lot, as it turns out. First, dy/dx reminds us that the derivative is the limit of ratios of changes. Secondly, because the symbol dy/dx looks like a fraction, it reminds us that the derivative is a limit of fractions, of "quotients of differences". The symbol dy/dx, by its very form, gives a hint that *the derivative might be expected to exhibit some of the familiar properties of fractions.* In a lighter vein, the reason Leibniz chose this symbolism is that, by the seventeenth century, the ancient Greek letter Δ had evolved "in the limit" to the modern d. What could be more natural than to denote the limit of $\Delta y/\Delta x$ by dy/dx?

It is hard to overestimate the value of appropriate symbolism. Of all creatures, only human beings have much ability to name things and to coin phrases. Poets like Shakespeare do this best of all.

> ... as imagination bodies forth
> The forms of things unknown, the poet's pen
> Turns them to shapes and gives to airy nothing
> A local habitation and a name.
>
> *A Midsummer-Night's Dream*, Act V

It can be contended that Leibniz's way of writing the calculus approaches the poetic. One can be borne up and carried along purely by his symbolism, while his symbols themselves may appear to take on a life all their own. Mathematics and poetry are different, but they are not so far apart as one might think.

The reader who is skeptical of the remarks just made is asked to suspend a final judgment until this chapter and the next are completed. Any skepticism that still remains may be lessened by reading Chapter 8. In the meantime, just to show that the remarks above are not especially radical, here is a well-known quotation from a man who won the Nobel Prize in literature:

> Mathematics, rightly viewed, possesses not only truth, but supreme beauty—a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show. The true spirit of delight, the exaltation, the sense of being more than man, which is the touchstone of the highest excellence, is to be found in mathematics as surely as in poetry.
>
> Bertrand Russell

Before we can see anything in Leibniz's notation, we must learn how to use it. There is no poetry in the examples that follow, only illustrations of how things are said in the language of Leibniz.

$$\text{If } y = x^2, \quad \text{then } \frac{dy}{dx} = \frac{d}{dx}(x^2) = 2x.$$

$$\text{If } A = s^2, \quad \text{then } \frac{dA}{ds} = \frac{d}{ds}(s^2) = 2s.$$

$$\text{If } f(x) = x^3, \quad \text{then } \frac{df}{dx} = 3x^2.$$

$$\frac{d}{dt}(\sqrt{1+3t^2}) = \frac{6t}{2\sqrt{1+3t^2}}.$$

$$\text{If } y = f \cdot g, \quad \text{then } \frac{dy}{dx} = f\frac{dg}{dx} + g\frac{df}{dx} \quad \text{(the product rule)}.$$

$$\frac{d}{dx}\left(\frac{1}{g}\right) = \frac{-1}{g^2}\frac{dg}{dx} \quad \text{(the reciprocal rule)}.$$

Exercises

1.0. What similarities and differences do *you* see between mathematics and poetry? (You might be interested in comparing your views with those expressed in W.M. Priestley, Mathematics and Poetry: How Wide the Gap? *Mathematical Intelligencer* Vol. 12, No. 1, 14–19 (1990) and in Vol. 12, No. 3, 5–6.)

1.1. Write the *square root rule* in Leibniz's notation: If $y = \sqrt{g}$, then $dy/dx = ?$

1.2. Write the *quotient rule*: $\frac{d}{dx}(f/g) = ?$

1.3. Find dC/dL if $C = 7L + (48/L)$.

1.4. What is $\frac{d}{dt}(t^2 + 3t + \pi)$?

1.5. Use the product rule to find $\frac{d}{dw}(w^4\sqrt{3+5w})$.

1.6. Find the derivatives of each of the following, expressing your answer in Leibniz's notation.
(a) $L = 12/W$.
(b) $C = (84/W) + 4W$.
(c) $C = 2\pi r$.
(d) $A = \pi r^2$.
Answers: (b) $dC/dW = (-84/W^2) + 4$. (c) $dC/dr = 2\pi$.

1.7. (*This question is not entirely frivolous, as will be seen in Chapter 7.*) The ancient Greek letter Δ has by the seventeenth century evolved "in the limit" to the letter d. What about the Greek letter Σ (sigma)? What is the "seventeenth-century limit" of Σ?

1.8. Is dy/dx the quotient of "dy" and "dx"? *Answer*: No. The derivative is denoted by the entire symbol dy/dx. Just as one understands the word *rainbow* without feeling any need to know what *ra* and *inbow* might mean, so one can understand dy/dx without ascribing meaning to dy and dx.

1.9. A familiar rule for fractions is $(A/B)(B/C) = A/C$. If derivatives behaved like fractions, what would the product

$$\frac{dy}{dx}\frac{dx}{dt}$$

be equal to? *Answer*: dy/dt.

1.10. If derivatives behaved like fractions, what would the following products of derivatives be equal to?
(a) $(dC/dL)(dL/dW)$.
(b) $(dA/dr)(dr/dt)$.
(c) $(dL/dW)(dW/dL)$.

§2. The Derivative as Instantaneous Speed

Suppose a rock is thrown directly upward, and suppose that, at time t seconds after it is released, its height h (in feet) is given by the equation

$$h = -16t^2 + 64t.$$

To illustrate the ideas just introduced concerning change, let us try to answer the following questions.

(a) During its first second of flight, what is the rock's *average* speed?
(b) What is the rock's *instantaneous* speed when $t = 1$ (i.e., 1 second after release)?

(c) When $t = 3$ (i.e., 3 seconds after release), is the rock going *up* or *down*?
(d) When does the rock attain its maximum height?
(e) What is the rock's *initial* velocity, i.e., what speed was given the rock at the instant of release?

To get hold of this situation, let us set up the usual table. Since $h = -16t^2 + 64t$, the derivative is given by

$$\frac{dh}{dt} = -32t + 64.$$

Plugging in a few numbers gives rise to this table.

t (in seconds)	h (in feet)	dh/dt (in feet per second)
$\left.\begin{matrix}0\\1\end{matrix}\right]\Delta t$	$\left.\begin{matrix}0\\48\end{matrix}\right]\Delta h$	32
2	64	0
3	48	-32
t	$-16t^2 + 64t$	$-32t + 64$

Why is dh/dt in the units of feet per second (ft/sec)? Because the change Δh in h is in feet and the change Δt is in seconds, so $\Delta h/\Delta t$ is in feet per second, showing that dh/dt is the limit of numbers $\Delta h/\Delta t$ that are in units of feet per second. Let us answer the questions raised in turn.

(a) During the initial 1-second interval, we have $\Delta t = 1$. The corresponding distance traveled is the change in height from 0 to 48 feet. That is, $\Delta h = 48$. The *average* speed for the first second is then given by

$$\begin{aligned}\frac{\text{distance traveled}}{\text{time taken}} &= \frac{\Delta h}{\Delta t}\\ &= \frac{48 \text{ feet}}{1 \text{ second}}\\ &= 48 \text{ ft/sec}.\end{aligned}$$

(b) The derivative dh/dt gives the instantaneous rate of increase of height with respect to time. When $t = 1, dh/dt = 32$ ft/sec. If the rock had a speedometer inside it to measure the upward speed, the speedometer should read 32 when t is 1.
(c) When t is 3, then dh/dt is -32, so that dh/dt, which measures the rate of *increase* of height, is *negative*. Since the rate of increase of height is

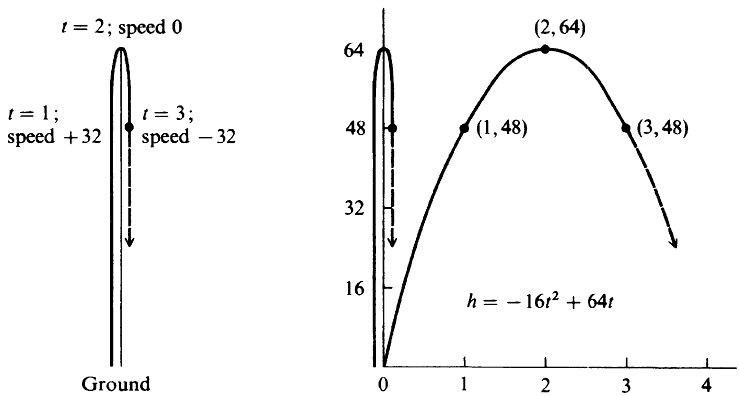
negative, the rock is *falling* when t is 3. (The instantaneous speed is 32 ft/sec *downward* when t is 3.)

(d) The rock is going up when the upward speed dh/dt is positive and is going down when dh/dt is negative. The rock must therefore attain its maximum height when dh/dt is zero. This occurs when t is 2.

(e) It is easy to be confused about speeds at the moment of release and at the moment of impact with the ground. But there is no question that at any intermediate time t, the speed is given by the expression $-32t + 64$. To avoid confusion, let us agree that the initial speed is the limit of this expression as t tends to zero from the right:

$$\operatorname*{Limit}_{t \to 0^+} (-32t + 64) = 64 \text{ ft/sec.} \qquad \square$$

It is easy to sketch the quadratic curve $h = -16t^2 + 64t$, and thus it is easy to picture the situation described above. *Avoid the mistake of thinking that the rock travels along the curve, however.* The rock moves straight up (until $t = 2$), then straight down, along the vertical axis.



The rock hits the ground when $t = 4$. *What is its speed at the moment of impact?* God only knows. We can know if we interpret the question as requiring us to find the speed the rock is *approaching* as t tends to the moment of impact. Then the answer is easy. The upward speed at the moment of impact is approaching

$$\operatorname*{Limit}_{t \to 4^-} (-32t + 64) = -64 \text{ ft/sec.}$$

(The negative sign occurs because $-32t + 64$ gives the *upward* speed). At the moment of impact, the rock is approaching a *downward* speed of 64 ft/sec.

Exercises

2.1. A rock is thrown directly upward. It height h (in feet) at time t seconds after release is given by

$$h = -16t^2 + 128t.$$

(a) What is the rock's *average* speed during its first second of flight?
(b) What is the rock's *instantaneous* speed when $t = 1$?
(c) Is the rock going *up* or *down* when $t = 3$?
(d) When is the maximum height attained?
(e) What is the average speed of the rock during the time interval between $t = 1$ and $t = 3$?
(f) What is the instantaneous speed when $t = 2$?
(g) What is the rock's initial speed?
(h) When does the rock hit the ground, i.e., when is the height h equal to zero? *Answer*: When $t = 8$.
(i) What is the speed of the rock when it hits the ground? (See the discussion above for a proper interpretation of this question.)

2.2. A rocket travels directly upward. At time t seconds after it is launched, its height h in feet is given by

$$h = 50t^3 + 80t.$$

(a) What is the rocket's average speed during its first 2 seconds of flight? *Answer*: 280 ft/sec.
(b) What should the rocket's speedometer read when $t = 1$?
(c) Let v stand for the rocket's speedometer reading (so that $v = dh/dt$). Then v, like h, is a function of t. Fill in the question marks appropriately in the following table.

t	h	v	dv/dt
1	?	230	?
2	560	680	?
π	?	?	?
t	$50t^3 + 80t$	?	?

(d) Think about the units in which things are measured. Here we have t in seconds, h in feet, so v is in feet per second. What units is dv/dt measured in? *Answer*: ft/sec per second.
(e) When t is 2, is the speedometer reading v *increasing* or *decreasing*? *Hint.* This is the same question as, "Is dv/dt *positive* or *negative*?" It is also the same question as, "Is the rocket *accelerating* or *decelerating* in its upward movement?"

(f) Is the rocket *accelerating* or *decelerating* when $t = 1$?
(g) **Acceleration** *is defined as the rate of increase of speed.* What is the rocket's instantaneous acceleration when $t = 1$? *Answer*: The rate of increase of speed, dv/dt, is equal to 300 ft/sec per second, when $t = 1$.
(h) What is the rocket's instantaneous acceleration when $t = 2$?

2.3. Go back to the situation described in exercise 2.1.
(a) Fill in the following table.

t	h	v	dv/dt
1	112	96	?
4	?	?	?
6	?	?	?
t	$-16t^2 + 128t$	?	?

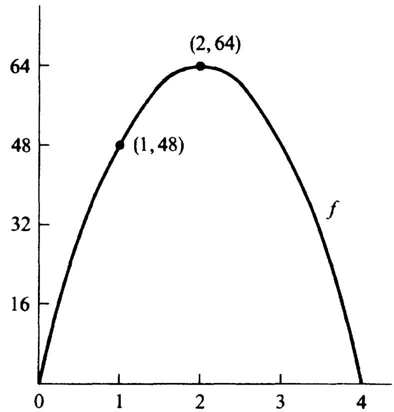
(b) Since v is the *upward* speed, and since dv/dt measures the rate of *increase* of v, it follows that dv/dt measures the *upward* acceleration. In this case the upward acceleration is constant. What is it? *Answer*: It is -32 ft/sec per second. (This is what gravity does, near the earth's surface. Each second the effect of gravity is to decrease the upward speed of a freely falling body by 32 ft/sec.)
(c) If a freely falling body is given an initial speed of $+128$ ft/sec, how many seconds will gravity take to change the speed to
(i) 64 ft/sec?
(ii) -64 ft/sec?
(d) In the rocket problem of exercise 2.2, why isn't the acceleration -32 ft/sec per second, since this is the acceleration due to gravity?

§3. Continuity and Nature

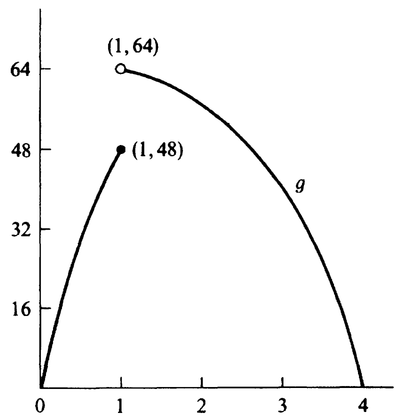
Laws that govern nature command our interest. Leibniz believed that all such laws are subject to the following basic principle.

Leibniz's Principle of Continuity
Nature must behave in a continuous fashion.

What does this mean? It is best to look at a concrete example, so consider again the motion of the rock discussed at length in the preceding section. This could be regarded as a simple experiment in physics, out of which arise the variables h and t, related by the function f pictured here.



Notice that f is continuous at each point in its domain $0 \leq t \leq 4$. According to Leibniz's principle, it could be no other way, for f describes a process that actually takes place in nature. It would be impossible, for example, for the rock to behave as described by the function g pictured below.



This describes the rock climbing steadily to 48 feet, then instantly leaping to a height of almost 64 feet. Only a miracle (i.e., something that disregards the laws of nature) could accomplish this! Leibniz's principle says that nature simply cannot allow the discontinuity of the function g at the point $t = 1$. If there are laws of nature, then these laws determine an underlying purpose, and the action of nature must agree with that purpose. Thus only *continuous* functions can arise out of this experiment, or any experiment, in physics. Or so the philosopher thought.

> Nothing happens all at once, and it is one of my great maxims, and among the most completely verified, that *nature never makes leaps*: which I called the *Law of Continuity*. . . .
>
> Leibniz

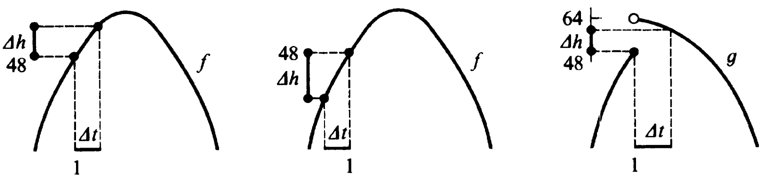
Let us go into this a bit further. Consider the instant when t is 1. What happens *naturally* (i.e., in the course of nature) is supposed to be continuously related both to the past and to the future. What does this mean in terms of *change*? Does continuity mean that a small change Δt in time will produce only a small change Δh in height? Certainly *not*, because anyone can think of occasions where nature allows large changes in little time. Instead, continuity means that, as Δt is taken nearer and nearer to zero, then the corresponding change Δh must also tend to zero:

$$\Delta t \to 0 \quad \text{implies} \quad \Delta h \to 0.$$

In other words, to say that h is a continuous function of t is to say

$$\operatorname*{Limit}_{\Delta t \to 0} \Delta h = 0. \tag{2}$$

To illustrate this, notice the difference in the behavior of f and g near the point $t = 1$.



The first two figures show that as $\Delta t \to 0$, either through positive or negative values, $\Delta h \to 0$. Thus, $\text{Limit}_{\Delta t \to 0}\, \Delta h = 0$ and f is continuous at 1. The third figure shows that as $\Delta t \to 0$ through positive values, $\Delta h \to 16$. Thus, $\text{Limit}_{\Delta t \to 0^+}\, \Delta h \neq 0$ and g is discontinuous at 1.

Condition (2) expresses the definition of continuity in terms of change. We should check to see that the definition of continuity by condition (2) agrees with the definition of continuity given in Chapter 1. To see this, assume that $h = f(t)$ satisfies condition (2). Because $\Delta h = f(t + \Delta t) - f(t)$, this condition tells us that

$$\operatorname*{Limit}_{\Delta t \to 0} f(t + \Delta t) - f(t) = 0,$$

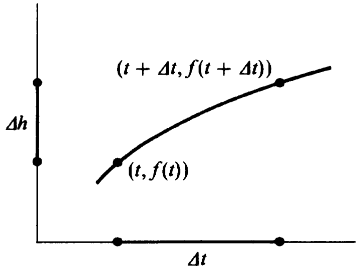
which means

$$\operatorname*{Limit}_{\Delta t \to 0} f(t + \Delta t) = f(t),$$

which says that

$$f(t) = \text{Limit } f \text{ at } t,$$

showing that f satisfies the definition of continuity given in Chapter 1.



Think of t as being fixed, with Δt tending to zero

Exercises

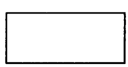
3.1. Suppose that, at a certain point, the derivative dh/dt exists. Prove that condition (2) must then be satisfied, showing that continuity follows from the existence of the derivative. *Hint.* As $\Delta t \to 0, \Delta h/\Delta t$ tends to the limit dh/dt. Why does this imply that $\Delta h \to 0$?

3.2. Does the existence of a derivative follow from continuity? That is, if $\Delta t \to 0$ implies $\Delta h \to 0$, does it automatically follow that the limit of $\Delta h/\Delta t$ exists?

3.3. If derivatives behaved like fractions, what would you expect the following products of derivatives to be equal to?
(a) $(dA/dC)(dC/dr)$.
(b) $(dy/dx)(dx/dy)$.
(c) $(dV/dh)(dh/dt)$.

§4. A Chain Rule?

Believe it or not, there is still something to be learned from the example given on the second page of Chapter 1. Three variables arise from that example, related in the following way.

$C =$ cost of  W, where $W \cdot L = 12$.

L

The variable C can be expressed either in terms of L alone or of W alone, while the variables W and L are themselves related by the fact that their product must be 12, the area of the rectangle. This leads to the following relations.

$$C = 7L + \frac{48}{L}, \tag{3}$$

$$L = \frac{12}{W}, \tag{4}$$

$$C = \frac{84}{W} + 4W. \tag{5}$$

Equations (3) and (4) might be thought of as links in a *chain* of relations which together produce equation (5). That is, the first two equations show how C is a function of L and L is a function of W. This chain of relations forces C to be a function of W, namely, the function specified in equation (5). One feels that there ought to be a rule governing derivatives in the presence of such a chain. The derivatives that arise from (3), (4), and (5) are as follows.

$$\frac{dC}{dL} = 7 - \frac{48}{L^2}. \tag{3'}$$

$$\frac{dL}{dW} = \frac{-12}{W^2}. \tag{4'}$$

$$\frac{dC}{dW} = \frac{-84}{W^2} + 4. \tag{5'}$$

Is there a "chain rule", as one feels there ought to be? Leibniz's notation suggests one to us. The notation suggests that derivatives might act like fractions, in which case we might expect that the product $(dC/dL)(dL/dW)$ is equal to dC/dW. Let us see if this is so.

$$\begin{aligned}
\frac{dC}{dL}\frac{dL}{dW} &= \left(7 - \frac{48}{L^2}\right)\left(\frac{-12}{W^2}\right) \quad [\text{from}(3')\text{and}(4')] \\
&= \frac{-84}{W^2} + \frac{48 \cdot 12}{L^2 W^2} \\
&= \frac{-84}{W^2} + \frac{48 \cdot 12}{(LW)^2} \\
&= \frac{-84}{W^2} + 4 \quad (\text{since } LW = 12) \\
&= \frac{dC}{dW} \quad [\text{from}(5')].
\end{aligned}$$

It is so! Leibniz's notation has suggested a *chain rule* for derivatives. Has any magician's trick ever been so delightful as this?

Exercises

4.1. From equation (4) we get $W = 12/L$.
 (a) Find dW/dL.
 (b) If derivatives behaved like fractions, one might expect that the product of dW/dL and dL/dW is 1. Is it? Multiply your answer to part (a) by dL/dW as given in equation (4') and use equation (4) to simplify.

4.2. Suppose $y = 1/x$.
 (a) Find dy/dx.
 (b) Suppose, in addition, that $x = t^2 + 3t$. Find dx/dt.
 (c) The chain of relations $y = 1/x$ and $x = t^2 + 3t$ tells us that $y = 1/(t^2 + 3t)$. Find dy/dt by using the general rule for reciprocals.
 (d) Using your answers to parts (a) and (b), find the product of dy/dx and dx/dt. Does your answer agree with dy/dt as found in part (c)? *Hint.* After finding the product, get your answer entirely in terms of t by replacing x with $t^2 + 3t$.

4.3. Suppose $y = \sqrt{u}$.
 (a) Find dy/du.
 (b) Suppose, in addition, that $u = x^2 + 9$. Find du/dx.
 (c) The chain of relations $y = \sqrt{u}$ and $u = x^2 + 9$ tells us that $y = \sqrt{x^2 + 9}$. Find dy/dx by using the square root rule.
 (d) Using your answers to parts (a) and (b), find the product of dy/du and du/dx. Does your answer agree with dy/dx as found in part (c)? It should.

4.4. Consider the chain of relations $y = u^5$ and $u = 3x^2 + 7x$. What does this tell us about the dependence of y upon x? *Answer*: The dependence of y upon x is expressed by the rule $y = (3x^2 + 7x)^5$.

4.5. Consider each of the following chains of relations. What does it tell us about the dependence of y upon x?
 (a) $y = u^3, u = 7x - 13$.
 (b) $y = 5/(2 - t), t = 2 - x$. *Answer*: $y = 5/x$.
 (c) $y = u^2, u = x^3 - 3x + \pi$.

4.6. A complicated dependence can often be regarded as made up of a chain of simpler dependences. For each of the following, specify such a chain.
 (a) $y = (4x^2 - 6x)^7$. *Answer*: This can be regarded as the result of the chain $y = u^7, u = 4x^2 - 6x$.
 (b) $y = \sqrt{3 - 2x + x^2}$.
 (c) $y = (19x - 4)^5$.
 (d) $y = (5x + (1/x))^4$. *Answer*: This is $y = u^4$, where $u = 5x + (1/x)$.

§5. The Chain Rule

Suppose that we have two functions that form a chain of relations, and suppose that each has a derivative. That is the setting for the chain rule.

Chain Rule

If y is a function of u and u is a function of x, then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

This is the rule that Leibniz's notation enabled us to guess. It is also a rule that Leibniz's notation enables us to remember, since it says, essentially, that derivatives multiply just like fractions, provided that Leibniz's notation is used.

Why is this rule true? First note that u is a continuous function of x, since du/dx exists. This means that

$$\Delta u \to 0 \quad \text{as } \Delta x \to 0. \tag{6}$$

(See exercise 3.1.) This fact will be useful in a moment.

To see the plausibility of the chain rule, consider what is produced by a nonzero change Δx in x. First, a change Δu in u occurs (since u is a function of x), and then the change Δu in turn produces a change Δy in y (since y is a function of u). By ordinary multiplication of fractions,

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u}\frac{\Delta u}{\Delta x}, \tag{7}$$

provided $\Delta u \neq 0$. As $\Delta x \to 0$, equation (7) becomes "in the limit"

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx},$$

by virtue of condition (6). □

What has just been given is more of a "plausibility argument" than a real proof to justify the chain rule. The trouble is that equation (7) does not hold if $\Delta u = 0$, i.e., if a nonzero Δx should produce no change in u. A more careful proof, taking account of this troublesome case, may be found (by those rare readers blessed with both skepticism and patience) in any book on real analysis. Let us for the time being accept the chain rule as true, and learn how to use it. It is the most important rule governing derivatives.

EXAMPLE 1

What is the derivative of $(3x^2 + 7x)^5$?

Here we want dy/dx, where $y = (3x^2 + 7x)^5$. As in exercise 4.4, we may regard y as being given in terms of x by the chain of relations

$$y = u^5 \quad \text{and} \quad u = 3x^2 + 7x.$$

By the chain rule,

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du}\frac{du}{dx}\\ &= 5u^4(6x+7)\\ &= 5(3x^2+7x)^4(6x+7).\end{aligned}$$ □

EXAMPLE 2
What is the derivative of $(x^3 - 3x + \pi)^2$?

Here we want dy/dx, where $y = (x^3 - 3x + \pi)^2 = u^2$, if we set u equal to $x^3 - 3x + \pi$. We then have the chain

$$y = u^2 \quad \text{and} \quad u = x^3 - 3x + \pi.$$

By the chain rule,

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du}\frac{du}{dx}\\ &= 2u(3x^2-3)\\ &= 2(x^3-3x+\pi)(3x^2-3).\end{aligned}$$

(Note that Example 2 can be done by the rule for squares, to get the same answer. The rule for squares is simply the special case of the chain rule that arises when a chain of relations involves a square.) □

Exercises

5.1. What is the derivative of $(7x - 13)^3$? *Hint*. The chain of relations in exercise 4.5(a) arises here.

5.2. Find the derivative of $(4x^2 - 6x)^7$. *Hint*. We want dy/dx, where $y = (4x^2 - 6x)^7 = u^7$, if we set u equal to $4x^2 - 6x$.

5.3. Regard each of the following as being given by an appropriate chain of relations, and use the chain rule to obtain the derivative.
(a) $(5x + (1/x))^4$.
(b) $(x^2 - 2x + 1)^3$.
(c) $(t^2 + t)^3$.
(d) $(5L - 16\pi L^2)^4$.

5.4. The area, radius, and circumference of a circle are related by the chain $A = \pi r^2$ and $r = (1/2\pi)C$. Find dA/dC by the chain rule. *Answer*: $dA/dC = (dA/dr)(dr/dC) = (2\pi r)(1/2\pi) = r = (1/2\pi)C$.

5.5. On the basis of the answer to exercise 5.4, and with nothing but Leibniz's notation to guide your intuition, guess what dC/dA is. If $dA/dC = r$, then dC/dA ought to be ...?

5.6. From the equations $A = \pi r^2$ and $C = 2\pi r$,
  (a) find an algebraic rule giving C in terms of A.
  (b) find dC/dA from your answer to part (a), and see if it agrees with the guess made in exercise 5.5.
  Partial answer: $C = \sqrt{4\pi A}$. (Find dC/dA by the square root rule.)

5.7. If $y = x^2, x > 0$, then it follows that $x = \sqrt{y}$, $y > 0$.
  (a) Find dy/dx from the equation $y = x^2$.
  (b) Find dx/dy from the equation $x = \sqrt{y}$.
  (c) The expression dy/dx "looks like" the reciprocal of dx/dy. Is it?

§6. Related Rates

Problems involving *related rates* are tailor-made for the chain rule. Let us do an easy example, then a harder example, in order to make some observations on how such problems may be handled.

EXAMPLE 3

A pebble is dropped in still water, forming a circular ripple whose radius is expanding at a rate of 3 inches per second. When the radius is 7 inches, how fast is the area A of the ripple increasing?

In rate problems it is important to get straight exactly what we are required to find, as well as what we are given to start off with. One must remember that the derivative measures the instantaneous rate of increase. The rate of increase of area A (with respect to time) is then dA/dt. The goal of Example 3 is then to find dA/dt when r is 7. This may be abbreviated by

$$\left.\frac{dA}{dt}\right|_{r=7} \tag{8}$$

(read "dA/dt, evaluated when r is 7"). The expression (8) is the rate we are required to find.

What are we given to work with? The first sentence of Example 3 tells us a related rate:

$$\frac{dr}{dt} = 3, \tag{9}$$

and we know that there is a chain of relations connecting the variables A, r, and t:

$$A = \pi r^2 \quad \text{and} \quad r \text{ is a function of } t.$$

By the chain rule, using (9),

$$\frac{dA}{dt} = \frac{dA}{dr}\frac{dr}{dt} = 2\pi r(3).$$

Therefore,

$$\frac{dA}{dt} = 6\pi r.$$

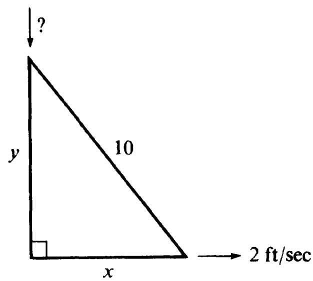
To evaluate expression (8), plug in $r = 7$:

$$\begin{aligned}\left.\frac{dA}{dt}\right|_{r=7} &= 6\pi r\big|_{r=7}\\ &= 6\pi(7)\\ &= 42\pi \text{ in}^2/\text{sec}.\end{aligned}$$

(The expression "in^2/sec" abbreviates the phrase "square inches per second". Why must dA/dt come out in these units?) □

EXAMPLE 4

The bottom end of a 10-foot ladder resting against a wall is pulled away from the wall at a rate of 2 ft/sec. At what rate is the top end falling at the instant when the bottom end is 6 feet from the wall?



There are two ways to work this problem, and we shall look at both of them. As with virtually every calculus problem, the first step is to *see the problem in terms of variables*. Time is certainly one variable, and the others are x and y, the legs of a right triangle formed by the ladder, the wall, and the floor. As time t increases, it is evident that x increases and y decreases. The derivative dy/dt gives the rate of *increase* of y. The rate at which the top of the ladder *falls* is the rate of *decrease* of y, which is the negative of dy/dt. Thus we are required to find

$$-\left.\frac{dy}{dt}\right|_{x=6}. \tag{10}$$

What are we given to work with? We know a related rate:

$$\frac{dx}{dt} = 2, \tag{11}$$

and we known an age-old relation that connects the variables x and y:

$$x^2 + y^2 = 100. \tag{12}$$

Solving this equation for y shows that we have the following chain of relations connecting y, x, and t:

$$y = \sqrt{100 - x^2} \quad \text{and} \quad x \text{ is a function of } t.$$

By the chain rule, with help from equation (11),

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = \frac{-2x}{2\sqrt{100 - x^2}}(2),$$

so that

$$-\frac{dy}{dt} = \frac{2x}{\sqrt{100 - x^2}}.$$

To find (10), as desired, plug in $x = 6$.

$$\begin{aligned} -\frac{dy}{dt}\bigg|_{x=6} &= \frac{12}{\sqrt{100 - 36}} \\ &= \frac{3}{2} \text{ ft/sec.} \end{aligned}$$

□

An alternate way to finish this problem is as follows. In equation (12), x and y both depend upon t. Taking the derivative with respect to t yields

$$\begin{aligned} \frac{d}{dt}(x^2 + y^2) &= \frac{d}{dt}(100), \\ 2x\frac{dx}{dt} + 2y\frac{dy}{dt} &= 0 \quad \text{(by rule for squares)}, \\ 4x + 2y\frac{dy}{dt} &= 0 \quad [\text{by}(11)], \end{aligned}$$

so that

$$-\frac{dy}{dt} = \frac{2x}{y}.$$

Therefore,

$$-\frac{dy}{dt}\bigg|_{x=6} = \frac{2x}{y}\bigg|_{x=6} = \frac{2(6)}{8} = \frac{3}{2} \text{ ft/sec},$$

since (by the Pythagorean theorem) $y = 8$ when $x = 6$. □

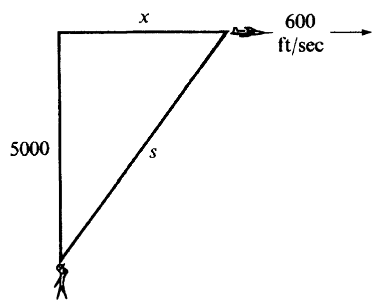
Related rates problems may seem difficult at first because everything in them seems to be changing at once. But this is only an invitation to see the problem in terms of variables and to use the derivative's magic power to measure change. Then adopt the philosopher's point of view. Seek that which does not change, that "holds sway above the flux". Search for a relation between the variables that always holds. This relation may be as simple as the Pythagorean theorem (in Example 4) or the formula for the area of a circle (in Example 3). Finally, express yourself in the language of Leibniz. It will lead you to the truth.

Exercises

6.1. A pebble is dropped in still water, forming a circular ripple whose radius is increasing at a rate of 5 inches per second. When the radius is 3 inches, how fast is the area of the ripple increasing?

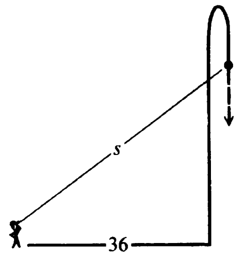
6.2. The bottom end of a 13-foot ladder resting against a wall is pulled away from the wall at a rate of 3 feet per second. How fast is the top end falling when
(a) the bottom end is 5 feet from the wall?
(b) the top end is 5 feet from the floor?
Answer: (a) $\frac{5}{4}$ ft/sec. (b) $\frac{36}{5}$ ft/sec.

6.3. An airplane is flying horizontally at 5000 feet, with speed 600 ft/sec, and an observer is on the ground. Let s be the distance from the observer to the airplane.



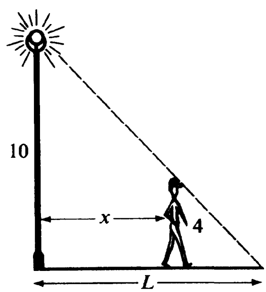
Find ds/dt, the rate of increase of s, at each of the following instants.
(a) Two seconds after the plane passes directly above the observer.
(b) One second after the plane is directly overhead.
(c) At the instant the plane is overhead.
(d) Three seconds before the plane is directly overhead.
Hint. (c) At this instant the plane is closest to the observer, so s assumes its minimum. What value does the derivative take when a minimum is attained?
Hint. (d) You want $(ds/dt)|_{x=-1800}$, and you should expect a negative answer, since the distance s is decreasing at this instant.

6.4. Consider again the rock whose motion is described at length in Section 2. Suppose there is an observer at ground level, 36 feet from the point where the rock is released. Let s be the distance from the observer to the rock.



Find ds/dt when
(a) $t = 1$ (and $h = 48$).
(b) $t = 2$ (and $h = 64$).
(c) $t = 3$ (and $h = 48$).
Hint. From the table in Section 2, dh/dt is 32, 0, and -32 at times $t = 1, 2,$ and 3, respectively.

6.5. A child 4 feet tall walks directly away from a street light that is 10 feet above the ground. She walks at a rate of 5 feet per second. How fast does the tip of her shadow move? *Hint*. See figure below. You want dL/dt, knowing that $dx/dt = 5$, and knowing, from similar triangles, that the relation $L/10 = (L - x)/4$ always holds. Proceed as in the alternate solution to Example 4.



§7. Antiderivatives

We have so far been mainly concerned with the following operation: getting a function, forming with it a quotient of differences, and taking a limit in order to get its derivative. This operation is called *differentiation*.

To *differentiate* a function is to take its derivative. For example, we get

$$\begin{aligned} 1 &\quad \text{by differentiating} \quad t, \\ t &\quad \text{by differentiating} \quad \frac{t^2}{2}, \\ t^2 &\quad \text{by differentiating} \quad \frac{t^3}{3}, \\ t^3 &\quad \text{by differentiating} \quad \frac{t^4}{4}. \end{aligned} \tag{13}$$

In general, we get

$$t^n \quad \textit{by differentiating} \quad \frac{t^{n+1}}{n+1}.$$

What we have been studying so far is called the *differential* calculus. The name is due to Leibniz who, writing in Latin, spoke of "calculus differentialis". An important concern of the differential calculus is simply to fill in the question mark, given the following table.

t	y	dy/dt
t	$f(t)$	?

The differential calculus is concerned with how to get from the second column above, to the third. It is done, of course, by means of Fermat's method. We shall now consider the reverse problem: *how to get from the third column back to the second.* This is a principal concern of "calculus integralis", as Leibniz called it, writing in 1696.

t	y	dy/dt
t	?	$f(t)$

This is the problem of finding an *antiderivative.* Fortunately, we already know a little about antiderivatives. From the formulas (13) it is completely obvious that

$$\begin{aligned} t &\quad \text{is an antiderivative of} \quad 1, \\ \frac{t^2}{2} &\quad \text{is an antiderivative of} \quad t, \\ \frac{t^3}{3} &\quad \text{is an antiderivative of} \quad t^2, \\ \frac{t^4}{4} &\quad \text{is an antiderivative of} \quad t^3. \end{aligned}$$

In general,

$$\frac{t^{n+1}}{n+1} \quad \textit{is an antiderivative of} \quad t^n.$$

Knowing this enables one to find easily antiderivatives of many functions whose rules involve powers only:

An antiderivative of -32 *is* $-32t$.
An antiderivative of $-32t$ *is* $-16t^2$.
An antiderivative of $64 - 32t$ *is* $64 - 16t^2$.
An antiderivative of $1 + 4t - 9t^2$ *is* $t + 2t^2 - 3t^3$.

We generally say "an", rather than "the", in speaking of antiderivatives, because there is generally more than one antiderivative of a given function. Having found an antiderivative, you can easily find another, simply by adding any constant to the one already in hand. *Reason*: If F is an antiderivative of f (i.e., if $F' = f$), then $F + C$ is too [since $(F + C)' = F'$, the derivative of a constant C being 0]. Thus, for example,

$$-32t, \quad -32t - 7, \quad -32t + \pi, \quad -32t + C,$$

where C can be any constant, are all antiderivatives of -32. We cannot speak of "the" antiderivative of -32, unless we specify, by giving additional information, exactly which antiderivative we mean.

EXAMPLE 5
Consider the function given by $f(t) = -32$, with domain $0 \leq t$. Find

(a) *an* antiderivative F of f.
(b) *the* antiderivative F of f that takes the value 64 when t is 0.
(c) *the* antiderivative F of f that takes the value -40 when t is 5.

We have already answered part (a). Any function whose rule is of the form $-32t + C$ will do, where C can be any constant (including, of course, 0). □

To answer (b), note that what is required is to fill in properly the following table.

t	$F(t)$	$f(t)$
0	64	
t	?	-32

The first line of the table gives enough information (we hope) to make the antiderivative unique. By part (a) we expect

$$F(t) = -32t + C, \tag{14}$$

and by the first line of the table we must have

$$F(0) = 64. \tag{15}$$

From (14), with $t = 0$, we get

$$F(0) = -32(0) + C = C.$$

This equation, together with (15), shows that C must be 64. Thus, in equation (14), not just any constant C will do; we must have $F(t) = -32t + 64$. □

To answer (c) we must satisfy the condition

$$F(5) = -40, \tag{16}$$

in addition to equation (14), which implies

$$F(5) = -32(5) + C. \tag{17}$$

Putting (16) and (17) together determines C:

$$-40 = -160 + C,$$

$$C = 120.$$

The answer to part (c) is then given by $F(t) = -32t + 120$. □

An antiderivative is generally determined, not uniquely, but only "up to an additive constant". Additional information, as in parts (b) and (c) of Example 5, is required to specify a unique antiderivative. How can we be *sure* of uniqueness, though? Our procedure here will be justified in the next section.

Exercises

7.1. In each of the following, specify an antiderivative F of the given function f.
   (a) $f(t) = 3t + 2$. *Answer*: $F(t) = \frac{3}{2}t^2 + 2t + C$.
   (b) $f(t) = -32t + 96$.
   (c) $f(t) = 1 + t + t^2$.
   (d) $f(t) = \pi t^3$. *Answer*: $F(t) = \frac{1}{4}\pi t^4 + C$.

7.2. In each of parts (a) through (d) of exercise 7.1, find the antiderivative F that takes the value 6 when $t = 2$.
   Answers: (a) $F(t) = \frac{3}{2}t^2 + 2t - 4$. (d) $F(t) = \frac{1}{4}\pi t^4 + 6 - 4\pi$.

7.3. In each of the following the derivative of h is specified. Use antiderivatives to find h itself.
   (a) $dh/dt = 96 - 32t$.
   (b) $dh/dt = -40 - 32t$.
   (c) $dh/dt = -32t$.

7.4. In each of parts (a) through (c) of exercise 7.3, find the antiderivative h that takes the value 100 when $t = 0$. *Answer*: (a) $h = 96t - 16t^2 + 100$.

7.5. In each of parts (a) through (c) of exercise 7.3, find the antiderivative h that takes the value 100 when $t = 1$. *Answer*: (a) $h = 96t - 16t^2 + 20$.

§8. A Fundamental Principle and Freely Falling Bodies

Taking antiderivatives points us in a direction exactly opposite the direction of differential calculus, and leads to the study of *integral* calculus. The reason for the use of the word *integral* will be explained in Chapter 7.

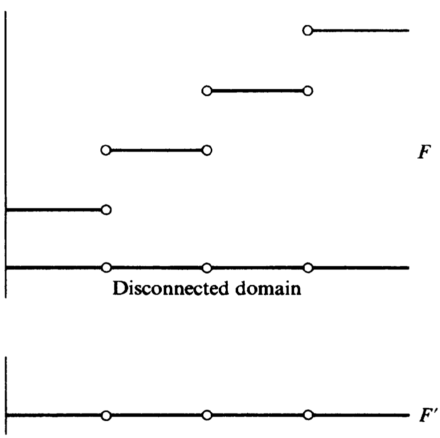
Let us examine more carefully the notion of an antiderivative. Because we get 0 by differentiating a constant function $F(t) = C$, it seems plausible that

$$\text{any antiderivative of 0 is a constant function.} \tag{18}$$

Can we be sure of this? Another way of saying the same thing is as follows:

$$\text{If } F'(t) = 0, \quad \text{then } F(t) = C \quad \text{for some constant } C. \tag{19}$$

Statements (18) and (19) are true, but only with the additional understanding that the domain in question is *connected*, that is, has no holes in it. The example pictured below shows that statement (19) can fail with holes in the domain. In this example $F'(t) = 0$, yet the function F is not constant.



In analysis it is shown that (18) and (19) are true, *provided the domain is connected.* The reader is asked to accept this as intuitively obvious. It has an important consequence, which will provide the basis for much that will follow.

A Fundamental Principle of Integral Calculus
Let F and A be functions defined on the same connected domain, and assume that $dA/dt = dF/dt$. Then,

$$A(t) = F(t) + C$$

for some constant C.

Proof
First note that

$$\frac{d}{dt}(A - F) = \frac{dA}{dt} - \frac{dF}{dt} = 0, \tag{20}$$

since $dA/dt = dF/dt$, by assumption. Therefore,

$$\begin{aligned} A - F &= \text{an antiderivative of } 0 \quad &&[\text{by } (20)] \\ &= \text{a constant function} &&[\text{by } (18)], \end{aligned}$$

since the domain is assumed connected. Thus, for some constant C,

$$\begin{aligned} A(t) - F(t) &= C, \\ A(t) &= F(t) + C. \end{aligned}$$

□

The fundamental principle just established is sometimes phrased this way: *Two antiderivatives of the same function differ by a constant.* The reader is cautioned to remember that zero is a perfectly good constant.

Intuition often runs ahead of reason. A good example of this is found in Section 7, where we expected equation (14) to hold and to justify what followed there. Now we know that we were right. The fundamental principle guarantees equation (14), for it says that any antiderivative *whatsoever* of -32 differs by a constant from $-32t$, on a connected domain. Thus, parts (b) and (c) of Example 5 do indeed have unique answers, for the domain of Example 5 is connected.

Though the fundamental principle may appear abstract, it has quite practical uses. It comes into play whenever we know the rate of change of a quantity and want to know the quantity itself. An example of this is furnished by the study of *freely falling bodies.* This refers to the vertical movement of objects thrown in the air near the earth's surface. If gravity is the only force acting on the body (which means that the body is not self-propelled and that the effect of air friction is ignored), the body is said to be *freely falling.*

Near the earth's surface, the effect of gravity is very simple to describe. Each second gravity *decreases* the upward speed of a freely falling body by 32 ft/sec. That is, if v is the upward speed of a freely falling body, then the effect of gravity is specified by the equation

$$\frac{dv}{dt} = -32. \tag{21}$$

This equation gives the rate of increase of v. If we want to know v itself, the fundamental principle says that

$$v = -32t + C$$

for some constant C. We need additional information to determine C. If, for example, the initial speed was known to be 64 ft/sec, then $v = -32t + 64$, as in part (b) of Example 5. If, as in part (c) of Example 5, the speed is known to be downwards at 40 ft/sec when $t = 5$, then we must have $v = -32t + 120$.

EXAMPLE 6

A rock is thrown upward from ground level with an initial speed of 64 ft/sec. Treating the rock as a freely falling body, answer the following:

(a) What is the maximum height attained by the rock?
(b) Where is the rock 3 seconds after it is released?
(c) When, and with what speed, will it hit the ground?

Here, we know that the upward speed v is given by $v = -32t + 64$, by the remarks preceding the example. But the upward speed v is equal to dh/dt, the rate of increase of height. Hence.

$$\frac{dh}{dt} = -32t + 64.$$

Therefore, by the fundamental principle,

$$h = -16t^2 + 64t + C \tag{22}$$

for some constant C. What is C? Since the rock is thrown from ground level, we must have $h = 0$ when $t = 0$, so that, from (22),

$$0 = -16(0)^2 + 64(0) + C.$$

Therefore, $C = 0$ and (22) becomes

$$h = -16t^2 + 64t.$$

But this is the height formula that was discussed at length in Section 2. From that section we know that the maximum height of the rock is 64 feet, the rock is at 48 feet and falling when $t = 3$, and it hits the ground when $t = 4$ with a downward speed approaching 64 ft/sec. □

The way the preceding example was begun involved two steps that can be schematized as follows:

t (time since release)	h (height)	$v\ (=dh/dt)$ (upward speed)	dv/dt (upward acceleration due to gravity)
0	0	$+64$	
t	(Step 2)	(Step 1)	-32

Example 6 involves two antiderivatives. In step 1, when we "pull back" from the fourth column to the third, we must adjust the constant so that the initial speed is 64, as required. When we pull back from the third column to the second, another constant must be adjusted to be in accordance with the given initial height.

EXAMPLE 7

From a building 200 feet high, a ball is thrown downward at an initial speed of 50 ft/sec. Find an algebraic expression for the height of the ball in terms of the time since release, treating the ball as a freely falling object.

Here, we begin with the following information, and we want to fill in the question mark giving h in terms of t. To do this we must first fill in the other question mark properly.

t	h	$v = dh/dt$	dv/dt
0	200	-50	
t	?	?	-32

We must simply pull back twice by taking antiderivatives, each time adjusting the constant in accordance with the initial conditions. Pulling back to the third column yields $v = -32t - 50$, and thus in the second column we must have $h = -16t^2 - 50t + 200$. This equation gives h in terms of t, so long as gravity is the only force that acts upon the ball, that is, until $h = 0$ when the ball hits the ground. This occurs when $t \approx 2.30$, according to our work in Example 8 of Chapter 4. □

The method outlined in Examples 6 and 7 develops a mathematical "model" predicting the motion of a freely falling body. Given the initial speed and the initial location, we can find the formulas governing the speed and the height at *any* time, so long as gravity acts. Thus the dynamics of a freely falling body can be worked out with ease, through the help of our fundamental principle of integral calculus.

Our model is hardly perfect, however, for the notion of a *freely falling body* is an idealization of what actually happens when a rock is tossed up in the air. Air friction has its effect, particularly at high speeds, and a more complex model is required to account for this and other factors.

Exercises

8.1. A ball is thrown vertically from a cliff. Find its upward speed v in terms of the time t since release if
  (a) it is initially thrown upward at 96 ft/sec.
  (b) it is initially thrown downward at 60 ft/sec.

8.2. A rock is thrown upward from ground level at an initial speed of 96 ft/sec.
  (a) What is its maximum height?
  (b) Where is the rock 4 seconds after release?
  (c) When will the rock hit the ground?
  (d) What is the speed of the rock at the moment of impact? *Answer*: $\text{Limit}_{t\to 6^-}\ v = -96$ ft/sec.

8.3. From a tower 256 feet high, a ball is thrown upward at an initial speed of 96 ft/sec. When, and with what speed, will it hit the ground? *Hint*.

t	h	v	dv/dt
0	256	96	
t	(Step 2)	(Step 1)	-32
?	0	?	

Partial answer: At impact the downward speed approaches 160 ft/sec.

8.4. Suppose, in exercise 8.3, the ball is thrown *downward* initially at 96 ft/sec. When, and with what speed, will it hit the ground? *Hint*. This is done like exercise 8.3, except the initial speed is -96 instead of 96.

8.5. A rifle is supposed to have a muzzle velocity of 1000 ft/sec. If it is fired straight up, how high will the bullet go?

8.6. A certain rifle, when fired straight up, will send a bullet to a height of 2000 feet. What is the muzzle velocity of the rifle? *Hint*. Letting v_0 be the muzzle velocity, we have

t	h	v
0	0	v_0
	2000	0

Find v_0, beginning with equation (21).

8.7. A boy hurls a ball directly upwards. It hits the ground 8 seconds later. What was the ball's initial speed? *Hint.*

t	h	v
0	0	v_0
8	0	

Find v_0.

§9. Antiderivatives and Distance

The method of freely falling bodies does not apply, of course, to self-propelled objects like motorcycles, cars, and rockets. Nevertheless, antiderivatives come into play with self-propelled objects when it is desired to convert speedometer readings into distance traveled. Suppose a navigator charts his speedometer reading as it varies over the span of an hour. How can the navigator determine from his chart the distance traveled during this hour? The answer involves antiderivatives.

EXAMPLE 8

A rocket ship blasts through the firmament on a journey directly away from the earth. At noon on a certain day the navigator becomes interested in the ship's speedometer reading as a function of time, and finds that it is given by $100t^3 - 400t^2 + 800t$, where t is the time in hours since noon. If this function f gives the speedometer reading in km/hr (kilometers per hour), find the distance traveled by the rocket

(a) between noon and two o'clock.
(b) between one and four o'clock.

The speedometer reading is the instantaneous rate of change of distance from the earth. If we let s be the distance from the earth, we then have

$$\frac{ds}{dt} = 100t^3 - 400t^2 + 800t = f(t).$$

The distance s must be in kilometers, since the speedometer reading is given in km/hr. Schematically, the situation we are faced with can be pictured as follows.

t (hours since noon)	s (distance from earth)	ds/dt (speed)
0	?	
1	?	
2	?	
4	?	
t	$F(t)$	$f(t)$

We know the expression for $f(t)$ and we need to fill in the question marks correctly to answer (a) and (b) above.

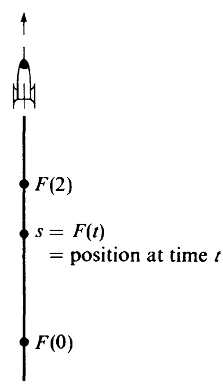
We first find an antiderivative F of f:

$$s = F(t) = \frac{100t^4}{4} - \frac{400t^3}{3} + \frac{800t^2}{2} + C.$$

We know the *position function* F must be of this form by the fundamental principle, but we are not given enough information to determine C. Nevertheless, by plugging in the values 0, 1, 2, and 4 into this expression for F, we can easily figure out what was required:

(a) The distance traveled between time $t = 0$ and $t = 2$ is equal to

$$\begin{aligned}(\text{position at } t = 2) \text{ minus } (\text{position at } t = 0)& \\ &= F(2) - F(0) \\ &= 933.7 + C - C \\ &= 933.7\,\text{km}.\end{aligned}$$



(b) The distance traveled between time $t = 1$ and $t = 4$ is equal to

$$\begin{aligned}F(4) - F(1) &= 4266.7 + C - (291.7 + C) \\ &= 3975\,\text{km}.\end{aligned}$$

□

To get the distance traveled, given the speed function f, is then a job for antiderivatives. If f is an antiderivative of f, then F gives the position at time t, so that the initial and final positions are readily determined. The distance traveled is simply the distance between the initial and final position, if the ship does not reverse course. (In Example 8 the speed function is always positive, so the direction of travel is always away from the earth.)

What happens if the speed function changes sign in the midst of the journey, so that the course of travel is reversed? Then the distance traveled must be calculated in two steps, as illustrated in the next example.

EXAMPLE 9
A rock is thrown upward at an initial speed of 64 ft/sec. How far does the rock travel during the first 3 seconds of its flight?

Here the speed function f is given by $f(t) = 64 - 32t$, because of the influence of gravity. When $t = 2$, the sign of the speed function changes from positive to negative, showing that the rock's motion changes from up to down. We must calculate separately the distance traveled by the rock during its upward and downward journey. An antiderivative F is given by

$$F(t) = -16t^2 + 64t + C$$

for some constant C. We are not given enough information to determine C, for we do not know the initial height of the rock.

Nevertheless, the distance traveled upward is

$$F(2) - F(0) = 64 + C - C = 64\,\text{ft}.$$

The distance traveled downward from $t = 2$ to $t = 3$ is

$$\begin{aligned} F(2) - F(3) &= 64 + C - (48 + C) \\ &= 16\,\text{ft}. \end{aligned}$$

The total distance traveled is then 80 *feet*, even though the distance between the rock's final and initial positions, given by $F(3) - F(0)$, is only 48 feet. (Example 9 is, of course, essentially the same situation that we have met twice before, in Section 2 and in Example 6, Section 7.) □

Suppose we have two continuous functions f and g defined on the domain $0 \leq t$, and suppose $f(t) \leq g(t)$, i.e., suppose the first never exceeds the second. Let us think of these functions as giving the upward speeds of two particles moving on the vertical axis, whose "heights" are then given by antiderivatives $F(t)$ and $G(t)$. What can we conclude from the fact that the rate of upward motion of the first particle never exceeds that of the second? *Answer*: It is intuitively clear that after any time t, the net increase in height of the first cannot exceed that of the second, i.e.,

$F(t) - F(0)$ cannot exceed $G(t) - G(0)$. *The reader should reflect for a moment to see that this conclusion follows even if the speeds f and g should sometimes be negative, producing downward motion.*

Let us state our conclusion as a theorem, even though we have supported it here by a (perhaps questionable) appeal to our intuition. In problem 27 of Chapter 7 the reader will be asked to justify this conclusion by an appeal to the fundamental theorem of calculus.

Theorem on Antiderivatives and Inequalities

If $f(t) \leq g(t)$ on the domain $0 \leq t$, then

$$F(t) - F(0) \leq G(t) - G(0), \quad \text{if } 0 \leq t,$$

provided that F and G are antiderivatives of f and g, respectively.

This theorem tells us how to go from inequalities on functions to inequalities on their antiderivatives. Or, equivalently, it tells us how to go from inequalities on derivatives of functions to inequalities on the functions themselves.

EXAMPLE 10

Suppose $f(t) \leq 5 - 4t + 9t^2$ if $0 \leq t$. What is the corresponding inequality forced upon $F(t)$, where F is an antiderivative of f?

Here we are given $f(t) \leq g(t)$, where $g(t) = 5 - 4 + 9t^2$. Using the theorem above with $G(t) = 5t - 2t^2 + 3t^3$, we may deduce that

$$F(t) - F(0) \leq 5t - 2t^2 + 3t^3, \quad \text{if } 0 \leq t.$$

To get a bound on $F(t)$ itself we add the constant $F(0)$ to each member of this inequality:

$$F(t) \leq F(0) + 5t - 2t^2 + 3t^3, \quad \text{if } 0 \leq t.$$

EXAMPLE 11

Suppose $h(t)$ gives the height of a particle at time t and suppose $h'(t)$ is bounded as follows:

$$5 - 4t + 8t^2 \leq h'(t) \leq 5 - 4t + 9t^2, \quad \text{if } 0 \leq t.$$

Knowing these bounds on the values of its derivative, what can we say about bounds on $h(t)$ itself?

Applying the theorem above to each inequality separately we infer, just as in Example 10, that

$$5t - 2t^2 + \frac{8}{3}t^3 \leq h(t) - h(0) \leq 5t - 2t^2 + 3t^3, \quad \text{if } 0 \leq t.$$

Adding $h(0)$ to each of the three members of this inequality shows that $h(t)$ must satisfy the inequality

$$h(0) + 5t - 2t^2 + \frac{8}{3}t^3 \leq h(t) \leq h(0) + 5t - 2t^2 + 3t^3, \quad \text{if } 0 \leq t. \qquad \square$$

In the previous section we learned that if $dA/dt = dF/dt$ on a connected domain, then $A(t)$ and $F(t)$ differ by a constant. It follows that $A(t) - A(0) = F(t) - F(0)$ if 0 lies in this domain. Now we know that if $dA/dt \leq dF/dt$ when $0 \leq t$, then $A(t) - A(0) \leq F(t) - F(0)$. The surprising power of this seemingly simple observation about antiderivatives and inequalities may be seen in Appendix 5, where a brief account of the theory of Taylor approximations is given.

Exercises

9.1. In Example 8, find the distance traveled by the rocket ship between
   (a) $t = 0$ and $t = 3$.
   (b) $t = 1$ and $t = 3$.

9.2. If ds/dt is the upward speed, then its rate of increase, which is $d(ds/dt)/dt$, is the upward acceleration. In Leibniz's notation, the symbol $d(ds/dt)/dt$ is abbreviated to d^2s/dt^2. Find the upward acceleration in Example 8, and then answer the following:
   (a) Is the rocket *accelerating* or *decelerating* in its upward movement when $t = 3$?
   (b) Is the rocket *accelerating* or *decelerating* when $t = 0$? *Answer*: Since $d^2s/dt^2|_{t=0} = 800$ km/hr per hour, which is *positive*, the rocket is accelerating.

9.3. In Example 9, how far does the rock travel between
   (a) $t = 1$ and $t = 3$?
   (b) $t = 0$ and $t = 4$? *Answer*: 128 feet.

9.4. A stone is thrown upward from a tower window at an initial speed of 48 ft/sec. Find the distance traveled by the stone during its first 3 seconds of flight, treating it as a freely falling body.

9.5. Do exercise 9.4 with the modification that the stone is thrown downward instead of upward.

9.6. The speed function f of a ship stays constant at 30 km/hr, i.e., $f(t) = 30$. Find how far the ship travels between $t = 1$ and $t = 4$,
   (a) by the method of antiderivatives, as in Example 8.
   (b) by common sense.

9.7. A ship moves in a straight line. Its speed function f is unknown but is bounded by the inequality

$$30 + 2t \leq f(t) \leq 30 + 4t, \quad \text{when } 0 \leq t.$$

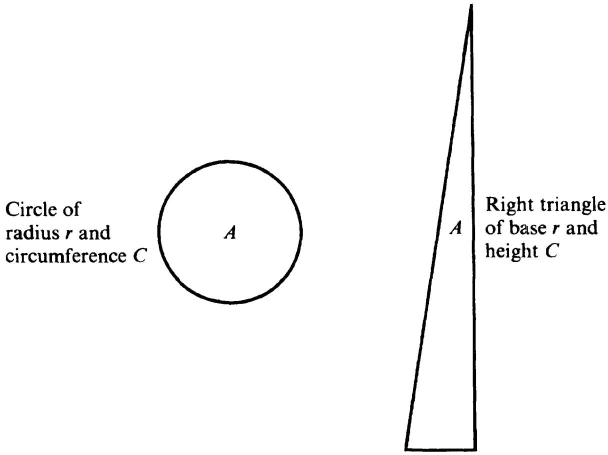
Use the theorem on antiderivatives and inequalities to find bounds on how far the ship travels between the times $t = 0$ and $t = 3$. *Hint*. The distance traveled is given by $F(3) - F(0)$, where F is the position function of the ship.

§10. A Token

There is a lot to be learned in the simple pastime of contemplating a circle. (All of trigonometry, in fact, arises this way.) In Chapter 3 it was established that the area A of a circle is given by

$$A = \pi r^2, \tag{23}$$

where r is the radius. This followed from Archimedes' demonstration that the two figures below have the same area.



The equality of areas produces the equation $\pi r^2 = \frac{1}{2}Cr$, from which we get the formula for the circumference,

$$C = 2\pi r. \tag{24}$$

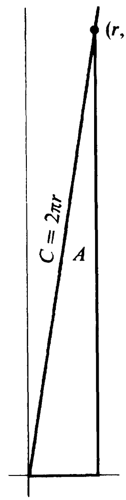
From (23) we derive the equation $dA/dr = 2\pi r$, so that by (24) we have

$$\frac{dA}{dr} = C. \tag{25}$$

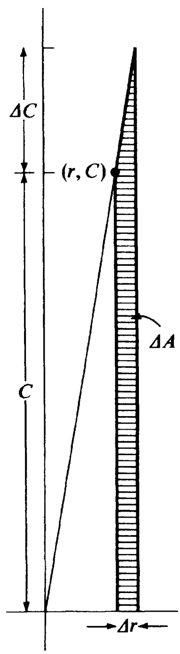
Thus the derivative (with respect to r) of the area of a circle is equal to the circumference! It takes only a little sensitivity to recognize that there must be here some sort of underlying harmony that has so far gone unnoticed. Equation (25) is a token from the gods. It is up to us to figure out what it really means. Remember the words of Xenophanes and Heraclitus!

Is equation (25) just an accident? Or should we have realized, by adopting the proper point of view, that this equation was bound to be true? *Let us set about trying to derive equation* (25) *directly from fundamental considerations*. We may discover something worth knowing in the process.

The equation $C = 2\pi r$ defines, of course, a straight line of slope 2π passing through the origin. The variables r, C, and A are then related as indicated in the figure.



If r is changed by a small amount Δr, what are the corresponding changes ΔC and ΔA? They are as indicated in the figure below.



To calculate the change ΔA in area, regard it as being made up of a rectangle surmounted by a triangle. The rectangle has base Δr and height C, and the triangle has base Δr and height ΔC. Therefore,

$$\Delta A = C(\Delta r) + \frac{1}{2}(\Delta r)(\Delta C).$$

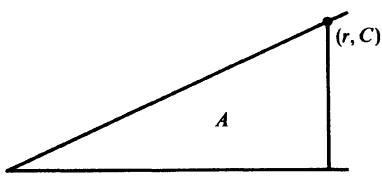
Dividing by Δr produces

$$\frac{\Delta A}{\Delta r} = C + \frac{1}{2}(\Delta C).$$

As $\Delta r \to 0$, we must have $\Delta C \to 0$ also, because C is a continuous function of r. Therefore,

$$\begin{aligned}\frac{dA}{dr} &= \operatorname*{Limit}_{\Delta r \to 0} \frac{\Delta A}{\Delta r} \\ &= \operatorname*{Limit}_{\Delta r \to 0} \left[C + \frac{1}{2}(\Delta C)\right] \\ &= C.\end{aligned}$$

What have we learned? We have learned that the equation $dA/dr = C$ is simply a consequence of the fact that A is the area beneath the curve giving C as a continuous function of r. That is, given the picture



it must follow that $dA/dr = C$.

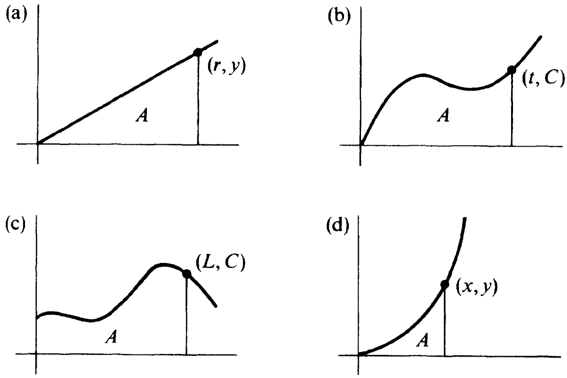
Are areas beneath continuous functions always related to the functions in this way? What is the secret that is still eluding us?

It turns our that what is behind all this is the fundamental theorem of calculus. Leibniz guessed it, probably sometime in the 1670s. Actually, Isaac Newton had come upon it in 1666 (at the age of twenty-three), but kept it a secret.

The fundamental theorem is discussed in Chapter 7. The reader may possibly be able to guess what it says beforehand, however, after doing the following exercises. The many uses of this theorem will still bring surprise.

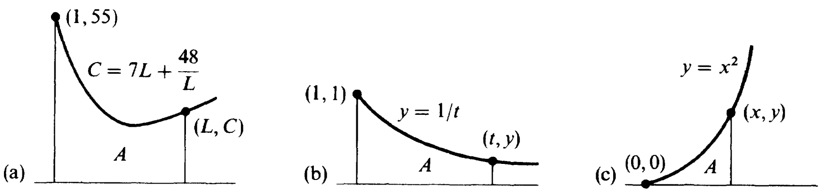
Exercises

10.1. Make a guess about a relation between the three variables that occur in each of the following pictures.



Answer: (c) $dA/dL = C$. (d) $dA/dx = y$.

10.2. Guess again, as in exercise 10.1, but utilize the fact that equations for the curves are furnished.



Answer: (c) $dA/dx = x^2$.

§11. Leibniz

Leibniz, like Descartes, is one of several mathematicians who were also distinguished philosophers. He said that we live in "the best of all possible worlds". Voltaire admired Leibniz, but could not accept this conclusion and satirized Leibniz's "optimism" in *Candide*. Relatively recent developments in physics have shown, however, that profound truth can be found in Leibniz's seemingly naive belief:

MENSIS OCTOBRIS A. M DC LXXXIV. 467

NOVA METHODUS PRO MAXIMIS ET MINIMIS, itemque tangentibus, quæ nec fractas, nec irrationales quantitates moratur, & singulare pro illis calculi genus, per G. G. L.

Sit axis AX, & curvæ plures, ut VV, WW, YY, ZZ, quarum ordinatæ, ad axem normales, VX, WX, YX, ZX, quæ vocentur respective, v, w, y, z; & ipsa AX abscissa ab axe, vocetur x. Tangentes sint VB, WC, YD, ZE axi occurrentes respective in punctis B, C, D, E. Jam recta aliqua pro arbitrio assumta vocetur dx, & recta quæ sit ad dx, ut v (vel w, vel y, vel z) est ad VB (vel WC, vel YD, vel ZE) vocetur dv (vel dw, vel dy vel dz) sive differentia ipsarum v (vel ipsarum w, aut y, aut z) His positis calculi regulæ erunt tales: TAB. XII

Sit a quantitas data constans, erit da æqualis 0, & d $\overline{ax}$ erit æqu. a dx: si sit y æqu v (seu ordinata quævis curvæ YY, æqualis cuivis ordinatæ respondenti curvæ VV) erit dy æqu. dv. Jam *Additio & Subtractio*: si sit z − y + w + x æqu. v, erit d $\overline{z-y+w+x}$ seu dv, æqu dz − dy + dw + dx. *Multiplicatio*, d $\overline{xv}$ æqu. x dv + v dx, seu posito y æqu. xv, fiet dy æqu x dv + v dx. In arbitrio enim est vel formulam, ut xv, vel compendio pro ea literam, ut y, adhibere. Notandum & x & dx eodem modo in hoc calculo tractari, ut y & dy, vel aliam literam indeterminatam cum sua differentiali. Notandum etiam non dari semper regressum a differentiali Æquatione, nisi cum quadam cautione, de quo alibi. Porro *Divisio*, d $\frac{v}{y}$ vel (posito z æqu. $\frac{v}{y}$) dz æqu. $\frac{\pm v\,dy \mp y\,dv}{yy}$

Quoad *Signa* hoc probe notandum, cum in calculo pro litera substituitur simpliciter ejus differentialis, servari quidem eadem signa, & pro + z scribi + dz, pro − z scribi − dz, ut ex additione & subtractione paulo ante posita apparet; sed quando ad exegesin valorum venitur, seu cum consideratur ipsius z relatio ad x, tunc apparere, an valor ipsius dz sit quantitas affirmativa, an nihilo minor seu negativa: quod posterius cum fit, tunc tangens ZE ducitur a puncto Z non versus A, sed in partes contrarias seu infra X. id est tunc cum ipsæ ordinatæ

Nnn 3 decre-

Figure 1. First page of the first paper published on the calculus. Leibniz wrote this short account—only six pages long—in 1684. The long title reads "A new method for maxima and minima, as well as for tangents, which is not obstructed by fractional and irrational quantities, and a unique calculus for them".

> Ours, according to Leibniz, is the best of all possible worlds, and the laws of nature can therefore be described in terms of extremal principles.
>
> C.L. Siegel and J.K. Moser*

Modern books on celestial mechanics show that the course actually chosen as the path of a heavenly body is the optimum among all possible courses. It should be added that the *optimum path* must be defined quite carefully, and in a way that Fermat was more likely to have foreseen than Leibniz. Nevertheless, the optimization techniques of Leibniz's calculus enter the picture in an essential way.

Few men have been more gifted than Leibniz. He invented a calculating machine that could multiply, divide, and take roots. He organized the Berlin Academy of Sciences and was its first president. He knew many languages, was an historian and a diplomat, with interests in economics, and a pioneer in the field of international law.

But when he died in 1716, little notice was taken. Only one mourner attended the funeral of Gottfried Wilhelm von Leibniz, and an observer said that "he was buried more like a robber than what he really was, the ornament of his country."

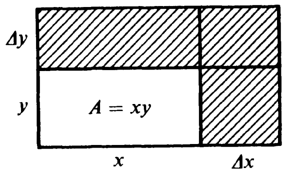
Problem Set for Chapter 6

1. A motorcycle travels on a straight road leading directly away from a city. At time t hours past noon its distance from the city is $10t^3 - 40t^2 + 80t$ miles.
   (a) How far does the motorcycle go between one o'clock and three o'clock?
   (b) What is its average speed over the time interval between one o'clock and three o'clock?
   (c) What is the speedometer reading at two o'clock?
   (d) At two o'clock, is the motorcycle *accelerating* or *decelerating*?
   (e) At one o'clock, is the motorcycle *accelerating* or *decelerating*?

2. The height of a rock at time t is given by $h = -4.9t^2 + 20t$, where h is in meters and t is in seconds.
   (a) Is the rock *rising* or *falling* when $t = 3$?
   (b) How fast is the rock going when $t = 3$?
   (c) When does the rock attain its maximal height?
   (d) What is the acceleration of the rock?

3. Suppose x and y are each functions of t. Let A denote their product. (If x and y are positive, A can be pictured as the area of a rectangle whose sides vary in length as t increases.)

* Siegel/Moser, *Lectures on Celestial Mechanics* (New York: Springer-Verlag, 1971) p.1.

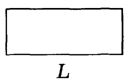


The shaded area is ΔA

(a) In time Δt, x becomes $x + \Delta x$ and y becomes $y + \Delta y$. Hence $A + \Delta A = (x + \Delta x)(y + \Delta y)$. Use this equation to find ΔA in terms of $x, y, \Delta x$, and Δy.
(b) Suppose dx/dt and dy/dt exist. What must happen to Δx and Δy as $\Delta t \to 0$?
(c) Derive the *product rule*, by carrying out the following steps.
   (i) Take your answer to (a), and divide both sides of the equation by Δt.
   (ii) Take the limit as $\Delta t \to 0$, using your answer to part (b), to show that

$$\frac{dA}{dt} = x\,\frac{dy}{dt} + y\,\frac{dx}{dt}.$$

4. Back in Chapter 1, we encountered this situation linking the three variables A, s, and w:

A = area of  w, where $2w + s = 1200$.

This leads to the chain

$$A = 600s - \frac{1}{2}\,s^2, \quad \text{where } s = 1200 - 2w,$$

which produces the equation $A = 1200w - 2w^2$.
(a) Using the three equations above, find $dA/ds, ds/dw$, and dA/dw.
(b) Multiply dA/ds by ds/dw. Is your answer equal to dA/dw?
(c) Find dw/dx after solving for w in the equation $2w + s = 1200$. Is it equal to the reciprocal of ds/dw?

5. Use the chain rule to find the derivative of each of the following:
(a) $(x^2 + 7x)^4$.
(b) $(x^3 - (1/x))^6$.
(c) $((x - 2)/(x + 2))^3$.
(d) $(x^4 - 3x + \pi)^5$.

6. Find the *second* derivative of the expression given in 5(a), paying attention to the fact that since its first derivative is expressed as a product by the chain rule, you must use the product rule to work out its second derivative. (And in the course of implementing the product rule you will have to use the chain rule again.) Now go back to the problem 48 of the previous chapter and give it a try.
(a) (*Simple*) In problem 48 of Chapter 5, fill in the blanks properly giving the first derivatives. These are simple products by the chain rule.

(b) (*Challenging*) Fill in the blanks properly giving the second derivatives in this problem. You must correctly use the product rule and the chain rule (again) here.

7. Suppose the radius r of a circle is increasing at the rate of 7 in/sec at the instant when r is 5 inches.
   (a) (*Easy*) How fast is the area A of the circle increasing at this instant?
   (b) (*Harder*) Find d^2A/dt^2 at this instant if $d^2r/dt^2 = 3$ in/sec per second.

8. A ladder 20 feet long rests against a wall. If its bottom end is pulled away from the wall at a constant rate of 5 ft/sec, how fast is the top of the ladder descending
   (a) when the bottom end is 12 feet from the wall?
   (b) when the top end is 12 feet from the floor?

9. A man 5 feet tall walks directly away from the base of a street light at a rate of 3 ft/sec. How fast does the length of his shadow increase if the street light is 12 feet tall?

10. An observer is 80 feet from a railroad track when a train passes at a rate of 50 ft/sec. How fast is the train's engineer moving away from the observer at the instant they are
    (a) 80 feet apart?
    (b) 100 feet apart?

11. Syrup is poured on a pancake at a constant rate so that the circular area covered by the syrup is increasing at a rate of 3 in^2/sec. How fast is the radius of this circular area increasing at the instant when the radius is 2 inches?

12. Find antiderivatives of each of the following.
    (a) $3t^2 + 12t + \pi$.
    (b) $1/t^2$.
    (c) $t^3 - 5t + 3$.
    (d) $t^5 + 4t^3 - 16t^2$.

13. In each of parts (a) through (d) of problem 12, find an antiderivative that takes the value 0 when $t = 1$. Is there a unique answer in each case?

14. From a window 276 feet high, a rock is thrown upward at an initial speed of 50 ft/sec. Answer the following questions, treating the rock as a freely falling body.
    (a) When will the rock attain its maximal height?
    (b) When will it hit the ground?
    (c) What will be the speed of the rock when it hits the ground?

15. A baseball is thrown straight up. What was its initial speed if
    (a) it reaches a maximum height of 100 feet?
    (b) it hits the ground 5 seconds after it is released?
    (c) it is at a height of 60 feet 2 seconds after it is released?

16. Since 32 feet is about 9.8 meters, equation (21) of Section 8 becomes $dv/dt = -9.8$ m/sec per sec. By taking antiderivatives twice, show that the height h in

meters of a freely falling body is $-4.9t^2 + v_0t + h_0$, where v_0 is the initial upward speed in m/sec, h_0 is the initial height in meters, and t is the time in seconds after the body is released.

17. In Example 7 of Section 8, find the speed at which the ball hits the ground.

18. The derivative of a certain function f is given by $f'(x) = 10 - 6x$. It is also known that $f(2) = 3$. Find the largest number in the range of f.

19. Suppose that the least number in the range of a certain function g is 2. Suppose also that $g'(x) = 2x - 4$. Find $g(3)$.

20. Think about a tennis ball just as it lands on the ground after being dropped. It bounces up. The upward speed is negative just before impact, and positive just after. Does this mean that the speed function must be discontinuous at the instant of impact? Is Leibniz's principle of continuity violated? Or can you see a way to save this principle by a more careful examination of what actually happens at the moment of impact?

21. (*For ambitious students only*) Although we know the derivative of the reciprocal function, we do not yet know an *antiderivative* of it. Nevertheless, suppose that we have somehow found the antiderivative A of the reciprocal function that takes the value 0 at the point 1. That is, we have a function A satisfying the following:

t	$A(t)$	$A'(t)$
1	0	
t	$A(t)$	$1/t$

Although we do not yet have any sort of formula by which to express the rule for the function A, we can nevertheless deduce some interesting things about it.

(a) To begin with, we know that if $L = A(t)$, then $dL/dt = 1/t$. This makes it unlikely that the domain of the function A includes the point 0. Why?

(b) Let $y = A(\pi t)$. This may be regarded as the chain $y = A(u)$, where $u = \pi t$. Use the chain rule to find dy/dt. *Hint.* $dy/du = 1/u$.

(c) The work in parts (a) and (b) shows that $dL/dt = dy/dt$. By the fundamental principle of integral calculus, there must be some constant C such that $y = L + C$, i.e., $A(\pi t) = A(t) + C$, on a connected domain. Show that the constant C must be $A(\pi)$. *Hint.* $A(1) = 0$.

(d) We now know that $A(\pi t) = A(\pi) + A(t)$, since $C = A(\pi)$. Assuming that the domain of A is the connected set of all positive numbers, show that, for any $s > 0$ and $t > 0$, we have

$$A(st) = A(s) + A(t).$$

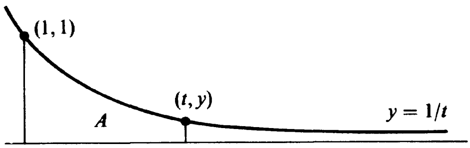
Hint. Use the same reasoning as before. Just consider s instead of π.

(e) The equation in (d) shows that the function A "converts multiplication into addition" in a sense. That is, the action of A on a *product* st is the *sum* of the action on each term. By letting $t = s$ in this equation, prove that $A(s^2) = 2A(s)$ if $s > 0$.

(f) Prove that $A(s^3) = 3A(s)$ if $s > 0$.
(g) In the equation in (d), let $t = 1/s$, and show that $A(1/s) = -A(s)$ if $s > 0$.
(h) In the equation in (d), let $s = t = \sqrt{x}$ and prove that $A(\sqrt{x}) = \frac{1}{2}A(x)$.
(i) If $L = A(f)$, then $dL/df = 1/f$. What would you guess is the formula for df/dL?

(This has been a preview of the logarithmic function.)

22. Consider the area A as indicated below:

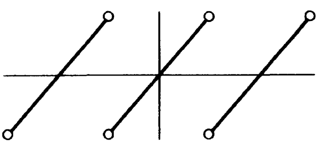


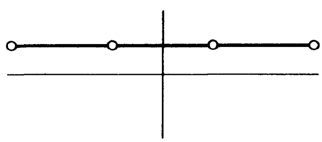
(a) What is A when $t = 1$?
(b) What would you guess dA/dt to be?
(c) Here, A is a function of t. Does it satisfy the table set up at the beginning of problem 21?

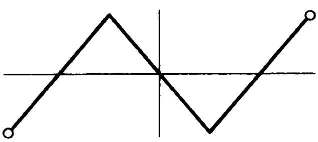
23. A rock is thrown up at an initial speed of 96 ft/sec. How far does the rock travel during
(a) the first 2 seconds of flight?
(b) the first 5 seconds of flight?

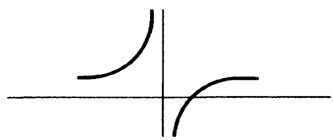
24. A small, tired bug is climbing up the y-axis. At time $t = 1$, the bug is at the origin and, from that time on, her speed is given by $f(t) = 4/t^2$.
(a) How far does the bug go between times $t = 1$ and $t = 2$?
(b) At what time t will the bug be at position $y = 3$?
(c) At what time t will the bug be at position $y = 3.75$?
(d) How far does the bug go between times $t = 1$ and $t = 1000$?
(e) Will the bug ever reach the position $y = 4$?

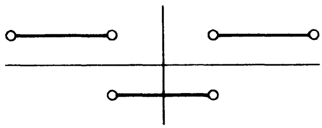
25. Match each of the following functions (a) through (g) with its derivative. [The derivative of (h) is not pictured.]

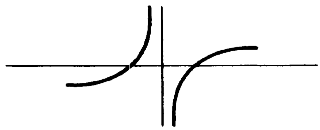
(a)


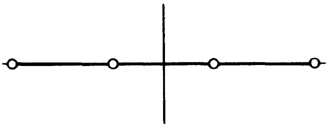
(b)


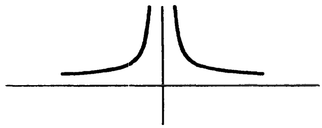
(c)


(d)


(e)


(f)


(g)


(h)


[The curve in (g) lies on the horizontal axis, but has holes in it. The left branch of the curve in (f) is identical with that of (d), but translated downwards.]

26. In problem 25, the curves (d) and (f) have the same derivative, but do not differ by a constant. Doesn't this contradict the fundamental principle of integral calculus? Explain why not.

27. The acceleration due to gravity near the earth's surface is often denoted in physics books by g. Thus g here denotes a constant that is equal to -32 ft/sec per sec, if we use seconds and feet to measure time and height. If we use meters instead of feet, then $g = -9.8$ m/sec per sec.
    (a) Find $h(t)$ if the initial height of a freely falling body is h_0 and its initial upward speed is v_0, i.e., fill in properly the blank space below, where g is the (constant) acceleration due to gravity:

t	h	dh/dt	d^2h/dt^2
0	h_0	v_0	g
t	—		g

    The only difference between this problem and problem 16 is that here we use g to denote the upward acceleration instead of a definite number like -9.8 or -32. The point of this is to see exactly how the second derivative of h enters into the expression giving $h(t)$.
    (b) The numerical value of the constant g varies with the choice of units for time and height. What is the value of g in centimeters per second per second? What is its value in centimeters per minute per minute? Can you make up units of distance and time so that the value of g will be -1?

28. (*What if gravity is not constant?*) In fact, the acceleration of gravity is not constant, but varies with height. If a body begins its free fall at a great distance above the earth, its acceleration due to gravity might at first be something like -30 ft/sec per sec, tending to -32 ft/sec per sec as it approaches the ground. In this case the body will not be falling as rapidly as the motion

described in problem 27 so its actual height $h(t)$ will in fact be higher than the height h calculated there.

(a) Explain why this means that $h_0 + v_0 t + \frac{1}{2} g t^2 \leq h(t)$ for all $t \geq 0$ during which the body is in free fall. *Hint.* Use your answer to problem 27.

(b) On the other hand, if we compute h in problem 27 assuming that the acceleration due to gravity is constant during the entire trip at $G = -30$ ft/sec per second we must come out with a larger height than the actual height $h(t)$. Show that this, together with the result of part (a), implies that while the body is in free fall,

$$h_0 + v_0 t + \frac{1}{2} g t^2 \leq h(t) \leq h_0 + v_0 t + \frac{1}{2} G t^2, \quad \text{if } 0 \leq t.$$

(c) Suppose the initial height of a freely falling object is 700,000 ft, at which height the acceleration due to gravity is -30 ft/sec per second. Suppose its initial speed is *downwards* at 1500 ft/sec. Use the inequality of part (b) to get upper and lower bounds on the height of the object 150 seconds after its release.

The reader impatient to know what the third, fourth, and higher derivatives are good for may turn to Appendix 5 after working problem 28.

7 CHAPTER The Integrity of Ancient and Modern Mathematics

When minds of first order meet, sparks fly, even across the centuries. The fundamental theorem of calculus, to be discussed in this chapter, is the result of such a pyrotechnic fusion of ideas. When Leibniz and Newton met Eudoxus and Archimedes, the calculus was rounded out into a whole. By the end of the seventeenth century it was becoming evident that calculus was not a bag of unrelated tricks but was an entity complete unto itself.

The point of this chapter is to see our subject as a unified whole, and the fundamental theorem is what really ties it together. Before coming to this theorem, let us recall briefly what we have seen so far. Calculus is largely the study of the interplay between a function and its derivative. In Chapters 4 and 5 we saw the *geometric* aspect of this interplay, which gives insight into the study of curves lying in a plane. As a by-product, the solution of *optimization* problems was effected. In Chapter 6, a *dynamic* aspect of this interplay revealed itself, throwing light upon the study of *change*. Previously vague terms, like *instantaneous velocity*, *acceleration*, and *rate of growth*, were seen to have natural and precise meanings couched in calculus. In addition, the fundamental notion of *continuity* has been clarified in terms of *limits*, and we have learned to solve equations by *Newton's method*.

We have seen by now that the interplay between a function and its antiderivative is signally rich. In this chapter we study still another aspect of this interplay. Calculus permits the easy calculation of the *area* of a figure bounded by curves in the plane.

§1. Areas and Antiderivatives?

Why should there be any connection between the calculus and the calculation of area? Isaac Newton saw the connection at an early age, having learned something, no doubt, from studying at Cambridge University under the tutelage of Isaac Barrow. While Newton was keeping his secrets to himself, the light came to Leibniz upon studying a mathematics paper by Pascal. The connection is a secret no longer.

Let us try to *guess* the connection first and put off until later an attempt to prove that our guess is correct. The key is to work through several simple examples and to observe that two seemingly different approaches yield the same result.

To see the landscape clearly, a motorcycle ride will help, if the reader will put up with just one more trip. Suppose you are watching the speedometer and therefore know the function f giving the speed of the motorcycle in terms of time. What method(s) can be applied to the speed function f, in order to calculate the distance traveled between, say, the times $t = 1$ and $t = 4$?

EXAMPLE 1

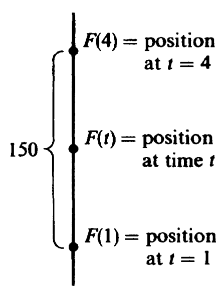
Suppose the speed is constant at 50 km/hr, i.e., the speed function is given by $f(t) = 50$. What is the distance traveled between $t = 1$ and $t = 4$?

One way the distance traveled can be found is by the antiderivative method illustrated in Section 9 of Chapter 6. Since the speed is always positive in this example, the distance traveled is just the distance between the motorcycle's initial and final positions. The position function F is an antiderivative of the speed function f, so

$$F(t) = 50t + C,$$

where C is some constant. The distance traveled is then

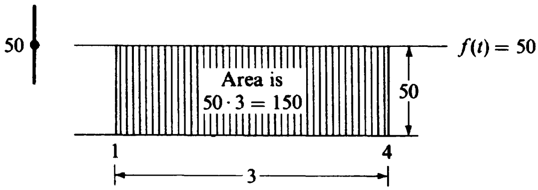
$$\begin{aligned} F(4) - F(1) &= 200 + C - (50 + C) \\ &= 150 \text{ kilometers.} \end{aligned}$$



Common sense reveals a simpler way to do this problem, however, for the speed is *constant* at 50 km/hr. Traveling at 50 km/hr for 3 hours, the motorcycle covers a distance of

$$50 \cdot 3 = 150 \text{ kilometers.}$$

The product $50 \cdot 3$ has a striking significance if we look at the graph of the speed function f, which is simply a horizontal line. One cannot help but notice that *the distance traveled between times $t = 1$ and $t = 4$ is numerically equal to the area beneath the curve f, between $t = 1$ and $t = 4$:*



$$\text{Area beneath } f, \text{ from } t = 1 \text{ to } t = 4, \text{is } 150 = F(4) - F(1),$$

where F is an antiderivative of f. Could it be that the area beneath *any* curve is so simply related to an antiderivative? □

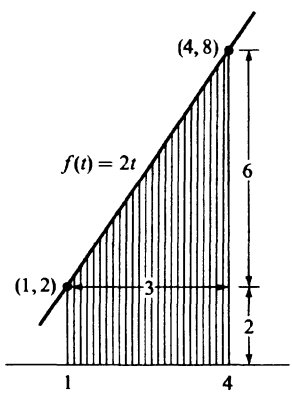
EXAMPLE 2

Suppose the speed is given by $f(t) = 2t$. What is the distance traveled between $t = 1$ and $t = 4$?

An antiderivative F is given by $F(t) = t^2 + C$. Since the speed $2t$ is always positive between $t = 1$ and $t = 4$, the distance traveled is

$$F(4) - F(1) = 16 + C - (1 + C) = 15 \text{ units.}$$

Let us check to see if this is equal to the area beneath the curve f. Since the graph of $f(t) = 2t$ is simply a line of slope 2, the area in question looks like this:



The area is made up of a rectangle of area $3 \cdot 2 = 6$, surmounted by a right triangle of area $\frac{1}{2}(3)(6) = 9$. The area beneath the curve f is then

$$6 + 9 = 15. \qquad \square$$

The two methods agree once again! The area beneath the graph of the positive function f again turns out to be the same number as that calculated by the antiderivative method, i.e.,

$$F(4) - F(1). \tag{1}$$

We shall see a lot of such expressions as (1), and it will be convenient to have an abbreviation for them. The notation $F|_a^b$ or $[F(t)]_a^b$ is defined to do this:

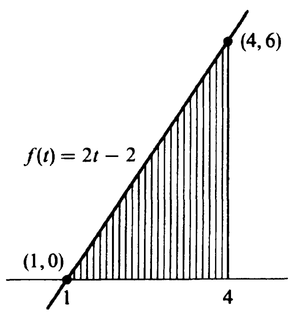
$$F|_a^b = [F(t)]_a^b = F(b) - F(a).$$

For example,

$$\begin{aligned} 50t|_1^4 &= 50(4) - 50(1) = 150, \\ [t^2]_1^4 &= 4^2 - 1^2 = 15, \\ [t^2 - 2t]_1^4 &= (16 - 8) - (1 - 2) = 9. \end{aligned} \tag{2}$$

EXAMPLE 3

Consider the area beneath the curve given by $f(t) = 2t - 2$, between $t = 1$ and $t = 4$. Sketch this area and see if it is equal to that calculated by the antiderivative method.



The area is easily seen to be a right triangle of base 3 and height 6, having an area of

$$\frac{1}{2}(3)(6) = 9,$$

which agrees with the number calculated by the antiderivative method in equation (2) preceding the example. $\square$

Exercises

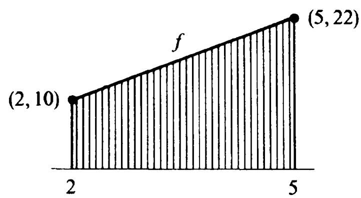
(*Remember that the phrase* "beneath the curve" *means* "below the curve and above the horizontal axis".)

1.1. Sketch the graphs of each of the following linear functions f and find the area beneath f, between $t = 1$ and $t = 4$, by splitting the area into a rectangle surmounted by a triangle, as in Examples 2 and 3.
(a) $f(t) = 10 - 2t$.
(b) $f(t) = t$.
(c) $f(t) = 4t - 3$.

1.2. For each of the three linear functions of exercise 1.1, apply the antiderivative method. That is, find an antiderivative F and calculate the expression (1). *Answer*: (b) $\frac{1}{2}t^2|_1^4 = \frac{1}{2}(16) - \frac{1}{2}(1) = \frac{15}{2}$.

1.3. Apply the method of antiderivatives to each of the following.
(a) $f(t) = 4t + 2$, from $t = 2$ to $t = 5$.
(b) $f(t) = 4t + 2$, from $t = 1$ to $t = 4$.
(c) $f(t) = t$, from $t = 0$ to $t = 1$.
(d) $f(t) = 5 - t$, from $t = 0$ to $t = 2$.
Answer: (a) $[2t^2 + 2t]_2^5 = 60 - 12 = 48$.

1.4. The answer to each of the four parts of exercise 1.3 ought to be equal to a certain area. In each case, sketch the area. *Answer*: (a) The area of 48 is that lying beneath the curve $f(t) = 4t + 2$, $2 \leq t \leq 5$.

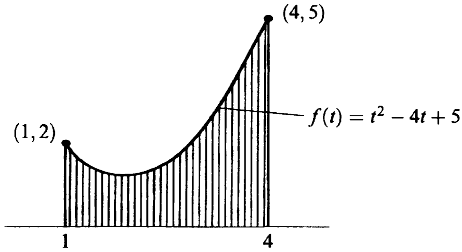


1.5. Apply the method of antiderivatives to each of the following.
(a) $f(t) = 1/t^2$, from $t = 1$ to $t = 4$.
(b) $f(t) = 1/t^2$, from $t = 2$ to $t = 6$.
(c) $f(t) = t^2$, from $t = 0$ to $t = 1$.
(d) $f(t) = t^2 - 4t + 5$, from $t = 1$ to $t = 4$.
Answer: (a) $[-1/t]_1^4 = -\frac{1}{4} - (-1) = \frac{3}{4}$.

§2. Areas Bounded by Curves

Consider the area beneath the quadratic curve given by

$$f(t) = t^2 - 4t + 5, \quad 1 \leq t \leq 4.$$



An antiderivative F of f is given by

$$F(t) = \frac{1}{3}t^3 - 2t^2 + 5t.$$

In view of the way things have turned out up to now, one might *guess* that this area is equal to

$$\left[\frac{1}{3}t^3 - 2t^2 + 5t\right]_1^4 = \left(\frac{64}{3} - 32 + 20\right) - \left(\frac{1}{3} - 2 + 5\right)$$

$$= 6.$$

The answer of 6 square units is surely easy to calculate by the antiderivative method. But how can we be sure that this method gives the correct area? We must first have a clear definition of *area*.

The importance of the role of definitions (in any subject, but particularly in mathematics) is not often noticed. At first we generally have only an intuitive conception of some notion that seems of interest. However, we can deal with intuitive notions, like *tangent line* and *area*, only in a superficial way until we assign these notions a precise significance, showing how they are related to ideas with which we are quite at home. Even more important (in any subject) is the choice of *what* terms to define, for that choice will determine one's language and consequently will ease—or hinder—one's way. When Fermat chose to think in terms of the intuitive notion of a *limit*, he rendered invaluable service to all who would enter mathematics.

Fermat pointed us toward a definition that clarified the idea of a *tangent line* and enabled us to travel in this book as far as we have. To travel much further with security, we must seek clarification of the notion of *area*. What does it mean to assert that the area pictured above is 6 square units? The figure is bounded by a curve on one side! Is it nonsense to speak of the "area" inside a curved figure?

This question was profoundly considered long ago by Archimedes, who became the master of a method introduced still earlier by Eudoxus.

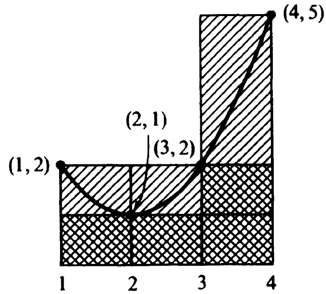
Archimedes, of course, had no notion of antiderivatives, but he could calculate areas (and volumes!) enclosed by curved figures. He used the method of Eudoxus, coupled with his own awesome technique.

The exercises below may suggest the essence of Eudoxus' method, but the discussion in depth of this method is postponed until Section 5. There we shall again seek out Eudoxus and Archimedes, who knew what they were talking about.

Exercises

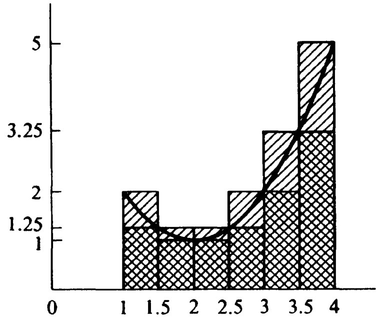
2.1. Review problems 8 through 16 in the problem set at the end of Chapter 3.

2.2. Consider the two "stairstep" figures superimposed on the curve $f(t) = t^2 - 4t + 5$, $1 \le t \le 4$.



Use them to convince yourself that the area beneath the curve exceeds 4 square units but is less than 9 square units.

2.3. What can you deduce by considering twice as many steps?



Answer: The area beneath the curve exceeds 4.875 square units but is less than 7.375 square units.

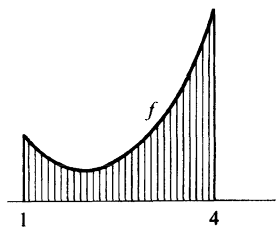
2.4. By putting in a few more steps, convince yourself that the area beneath the curve exceeds 5 square units but is less than 7 square units.

2.5. (*A question for speculation*) Make up a definition of the *area* enclosed by a curved figure lying in the plane. There are several ways this might be defined. Can you think of a way to define the area as a number that is the limit of other numbers that approximate it ever so closely? This is the way we shall proceed in Section 6, but an alternative approach may be found in problem 30 at the end of the chapter.

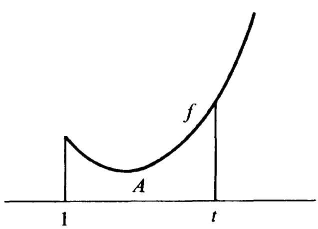
§3. Areas and Antiderivatives

The exercises in Section 2 point the way toward a definition of the notion of *area*. The definition will be stated precisely in Section 5. Right now, let us take for granted the fact that the notion of area dates from antiquity, and ask a seventeenth-century question: *What have areas got to do with antiderivatives?*

The answer to this question was given independently by Newton and Leibniz, and runs somewhat as follows. The key step in most calculus problems is to see the problem in terms of variables. How can we see the problem of calculating this area, for example, as a problem involving variables?



The answer is to consider the way the indicated area A varies in terms of t in the picture below.



t	A	dA/dt
1	0	
4	?	

(Do you see why we set $A = 0$ when $t = 1$?) We want to find the area A when $t = 4$. We have made a guess that the antiderivative method will probably give it to us:

$$\textit{When } t = 4, \textit{ then } A = F(4) - F(1), \tag{3}$$

where F is an antiderivative of f.

So far, equation (3) is only an educated guess. To prove that it is correct, let us try to find a formula expressing A in terms of t, in order to plug in $t = 4$. From the picture given above, we may expect that

$$\frac{dA}{dt} = f(t). \tag{4}$$

(See Section 10 of Chapter 6.) A proof of (4) will be forthcoming shortly, but first note that (4) says that A, like F, is an antiderivative of f. By the fundamental principle of integral calculus, A and F differ by some constant C, i.e.,

$$A = F(t) + C. \tag{5}$$

What is C? Since $A = 0$ when $t = 1$, equation (5) shows

$$0 = F(1) + C,$$

so that $C = -F(1)$ and (5) becomes

$$A = F(t) - F(1). \tag{6}$$

Statement (3), which we were trying to prove, is now an obvious consequence of (6)! □

A proof of (4), on which the preceding argument hangs, will be given below, but the style of argument just seen will be valuable later and ought to be remembered. It consists of three steps, culminating in a proof of (3):

Step 1. By (4), we have

t	A	dA/dt
1	0	
t		$f(t)$

Step 2. By the fundamental principle, since $F' = f$, we have

t	A	dA/dt
1	0	
t	$F(t) + C$	$f(t)$

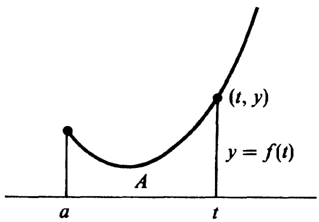
Step 3. Adjusting C so that $A = 0$ when $t = 1$ yields this information from which (3) follows easily.

t	A	dA/dt
1	0	
t	$F(t) - F(1)$	$f(t)$
4	?	

To make things complete, we must prove (4), which shows the connection between areas and antiderivatives. Note that here we are dealing with a function whose graph lies above the axis, i.e., a *positive* function. A more general case is treated in Section 4.

Theorem on Areas and Antiderivatives

Let f be a positive, continuous function, and let A be the area beneath the curve f from $x = a$ to $x = t$.

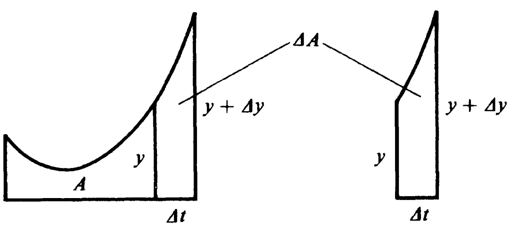


Then A is an antiderivative of f:

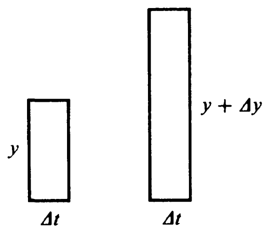
$$\frac{dA}{dt} = f(t).$$

Proof

Let t be fixed, and let $y = f(t)$. To find dA/dt, we first consider the change ΔA in area produced by a nonzero change Δt:



Compare the size of ΔA with that of rectangles built upon the same base of length Δt.



Since the area of these rectangles are equal, respectively, to $y \cdot \Delta t$ and $(y + \Delta y) \cdot \Delta t$, it follows that ΔA *lies between* $y \cdot \Delta t$ *and* $(y + \Delta y) \cdot \Delta t$. Dividing by Δt then shows that

$$\frac{\Delta A}{\Delta t} \text{ lies between } y \text{ and } y + \Delta y. \tag{7}$$

From (7) it is easy to determine the limit of $\Delta A/\Delta t$ as $\Delta t \to 0$, for Δy must tend to zero as well (since y is a continuous function of t). Thus $\Delta A/\Delta t$, being sandwiched between y and $y + \Delta y$, must tend to y, i.e.,

$$\frac{dA}{dt} = \operatorname*{Limit}_{\Delta t \to 0} \frac{\Delta A}{\Delta t} = y = f(t). \qquad \square$$

The proof just given may seem to rely on the picture that shows the curve f *rising* as it passes through (t, y) and also shows the change Δt as being *positive*. If the curve is falling, or if Δt is negative, the pictures have to be redrawn, but the proof has been worded in such a way as to require no change. If the function "wiggles" violently near (t, y) so that the curve is neither rising nor falling there, our proof is invalid, but the theorem is still true, as shown in a more careful demonstration better deferred to a course in analysis.

Exercises

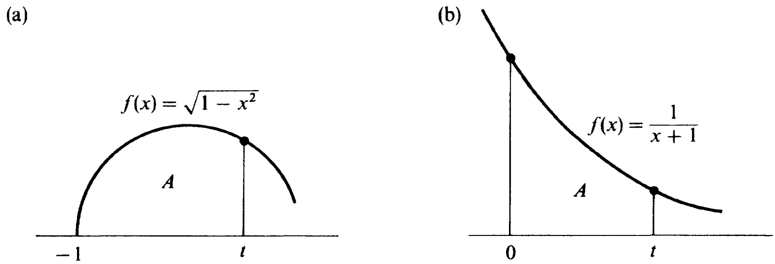
3.1. Find the area beneath the graph of each of the following equations, from 1 to 4.

(a) $f(t) = t^2 - 2t + 6$. (b) $f(t) = 1/t^2$.
(c) $f(x) = x^2$. (d) $f(x) = x^3$.
(e) $f(x) = x^2 + x^3$. (f) $y = 3t^2 + 5$.
(g) $y = 4x^3 - 3x^2$. (h) $y = \pi$.
(i) $h = -16t^2 + 64t$. (j) $g(s) = 600s - \frac{1}{2}s^2$.

Answers: (b) $\frac{3}{4}$ square units. [See exercise 1.5(a).] (d) $x^4/4|_1^4 = (256/4) - (1/4) = 255/4 = 63\frac{3}{4}$ square units. (f) 78 square units. (h) 3π square units.

3.2. In Section 10 of Chapter 6 you were asked to do some problems by guesswork. With the aid of the theorem on areas and antiderivatives, go back and do exercises 10.1 and 10.2 without guessing.

3.3. With the aid of the theorem on areas and antiderivatives, find dA/dt in each of the following situations. *Specify your answer in terms of t.*



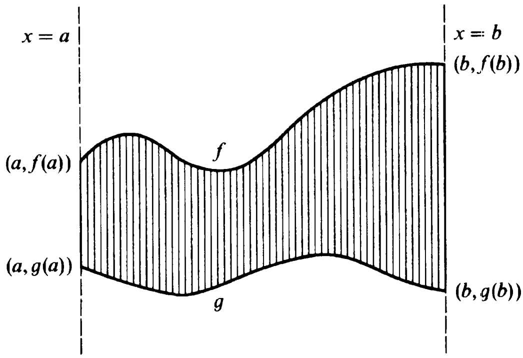
Answer: (a) Since $f(x) = \sqrt{1 - x^2}$, we have $f(t) = \sqrt{1 - t^2}$. By the theorem, $dA/dt = f(t) = \sqrt{1 - t^2}$. (Note that this problem really has nothing to do with "x". The answer would be the same if the function f had been expressed by writing $f(s) = \sqrt{1 - s^2}$, or by writing the equation $y = \sqrt{1 - L^2}$ to specify the curve f. In this problem x is a *dummy variable*, in the sense that the answer is unchanged if "x" is renamed as "s" or "L".)

3.4. The algebraic rule $\sqrt{1 - t^2}$ has domain $-1 \leq t \leq 1$. Find an antiderivative of this function. *Answer*: Let A be the function of t specified by the picture in exercise 3.3(a). (This function is specified in words, not as an algebraic rule, but it is a perfectly good function, and the theorem on areas and antiderivatives shows that it answers this question.)

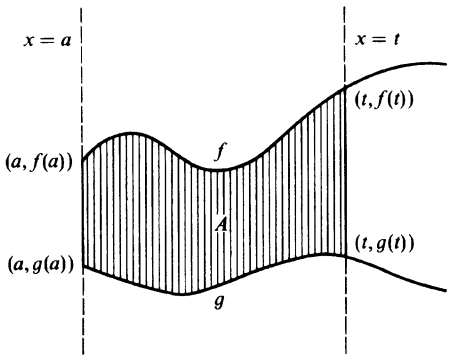
3.5. Find an antiderivative of each of the following functions, expressed as algebraic rules.
(a) $1/(t + 1)$, $0 \leq t$.
(b) $1/t$, $1 \leq t$.
(c) $\sqrt{4 - t^2}$, $-2 \leq t \leq 2$.
(d) $1/(t^2 + 1)$, $0 \leq t$.
Answers: (a) Let A be the function of t specified in the picture in exercise 3.3(b). (b) Let A be the function of t specified in the picture in problem 22 at the end of Chapter 6.

§4. Areas between Curves

The preceding section studied how to find the area between a curve and a certain straight line (the horizontal axis). It is just as easy to find the area between a curve and another curve.



Finding the area between two curves is a problem that can be approached by the method of Newton and Leibniz outlined in Section 3. The key is to see the problem in terms of variables. Let A be the area indicated below, so that A is a function of t.



If both f and g are continuous, *with f lying above g*, it follows that

$$\frac{dA}{dt} = f(t) - g(t). \tag{8}$$

The proof of (8) will not be given, because the idea of the proof is so similar to that of the theorem of Section 3. [The only basic difference is this. In the theorem of Section 3, ΔA was seen to be roughly equal to the product $f(t)\,\Delta t$; whereas here ΔA is roughly equal to $(f(t) - g(t))\,\Delta t$.] From (8) it is easy to deduce, as explained below, a more general area principle.

General Area Principle

Let f and g be continuous curves, with f lying above g. Then the area between f and g, from $x = a$ to $x = b$, is given by

$$[F - G]_a^b,$$

where F is an antiderivative of f and G is an antiderivative of g.

Proof

The proof follows exactly the pattern of the three steps described in Section 3. Using (8), we have the following information about the dependence of A upon t:

t	A	dA/dt
a	0	
t	?	$f(t) - g(t)$
b	?	

Equation (8) says that A is an antiderivative of $f - g$. Since $F - G$ is too (*why?*), the fundamental principle of integral calculus says that

$$A = (F(t) - G(t)) + C \tag{9}$$

for some constant C. What is C? Because $A = 0$ when $t = a$, we get from equation (9) that

$$0 = (F(a) - G(a)) + C.$$

This shows that $C = -(F(a) - G(a))$, so that (9) becomes

$$A = (F(t) - G(t)) - (F(a) - G(a)) = [F - G]_a^t.$$

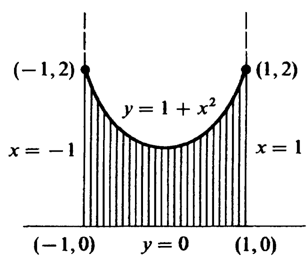
Thus the formula $A = [F - G]_a^t$ expresses A in terms of t. When $t = b$, then A becomes the desired area, as pictured at the beginning of this section. From the formula, when $t = b$,

$$A = [F - G]_a^b. \qquad \square$$

In applying the general area principle to areas bounded by curves, it is essential to note which curve is on top. If the curves cross one or more times, several applications of the area principle may be required. (See Example 8.)

EXAMPLE 4

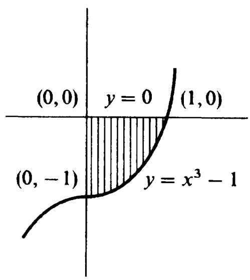
Find the indicated area.



This is the area between the curves f and g, where $f(x) = 1 + x^2$, $g(x) = 0$, $a = -1$, $b = 1$. Antiderivatives of f and g are given by $F(x) = x + x^3/3$ and $G(x) = 0$. By the general area principle, the area is

$$\begin{aligned}[F - G]_{-1}^{1} &= \left[x + \frac{x^3}{3}\right]_{-1}^{1} \\ &= \frac{4}{3} - \left(\frac{-4}{3}\right) = \frac{8}{3} \text{ square units.}\end{aligned}$$ □

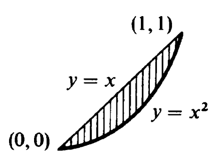
EXAMPLE 5
Find the indicated area.



The curve lying on top here is given by $f(x) = 0$, while the bottom curve's equation is $g(x) = x^3 - 1$. Here, $a = 0$ and $b = 1$. Antiderivatives are given by $F(x) = 0$ and $G(x) = x^4/4 - x$. By the general area principle, the area is equal to

$$\begin{aligned}[F - G]_0^1 &= \left[0 - \left(\frac{x^4}{4} - x\right)\right]_0^1 \\ &= \frac{3}{4} \text{ square units.}\end{aligned}$$ □

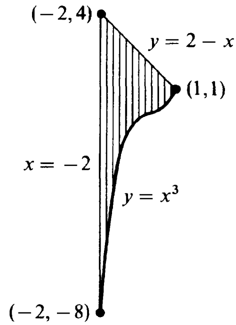
EXAMPLE 6
Find the indicated area.



Let $f(x) = x$, $g(x) = x^2$, $a = 0$, $b = 1$. The area is equal to

$$[F - G]_0^1 = \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1$$
$$= \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \text{ square units.} \qquad \square$$

EXAMPLE 7
Find the indicated area.

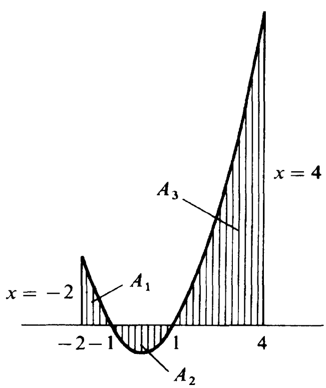


Let $f(x) = 2 - x$, $g(x) = x^3$, $a = -2$, $b = 1$. The area is given by

$$\left[2x - \frac{x^2}{2} - \frac{x^4}{4}\right]_{-2}^1,$$

which, when evaluated, is seen to be $\frac{45}{4}$ square units. $\square$

EXAMPLE 8
Consider the curve given by $f(x) = x^2 - 1$, with domain $-2 \le x \le 4$.



Find the area between the curve f and the x-axis.

Since the curve crosses the x-axis twice, the required area splits into three pieces, A_1, A_2, and A_3, as indicated. In each piece the area principle may be applied, taking account as to which of the curves $y = x^2 - 1$ and $y = 0$ is on top. We get

$$A_1 = \left[\frac{x^3}{3} - x\right]_{-2}^{-1} = \frac{4}{3},$$

$$A_2 = \left[0 - \left(\frac{x^3}{3} - x\right)\right]_{-1}^{1} = \left[x - \frac{x^3}{3}\right]_{-1}^{1} = \frac{4}{3},$$

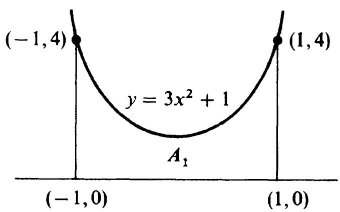
$$A_3 = \left[\frac{x^3}{3} - x\right]_{1}^{4} = 18.$$

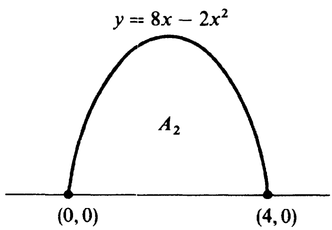
The total area between the curve f and the x-axis is then

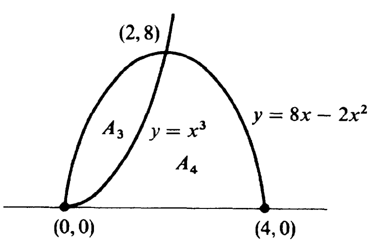
$$A_1 + A_2 + A_3 = 20\tfrac{2}{3} \text{ square units.} \qquad \square$$

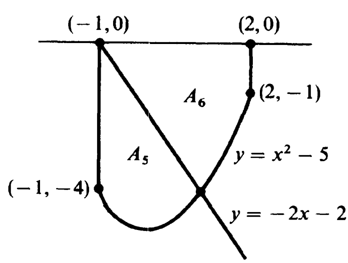
Exercises

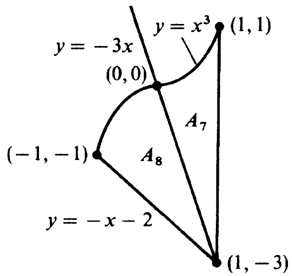
4.1. In each of the following, find the indicated area. *Hint for* A_4. First find A_3. $A_4 = A_2 - A_3$ (*why?*).

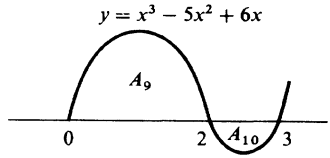






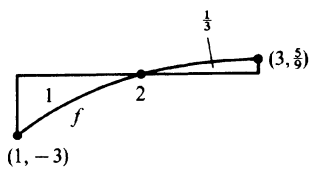






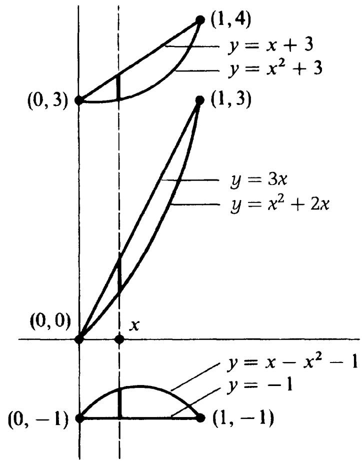
Partial answer: $A_2 = \frac{64}{3}$. $A_5 = \frac{16}{3}$. $A_7 = \frac{7}{4}$. $A_9 = \frac{8}{3}$.

4.2. Find the area between each of the following curves and the x-axis, as illustrated in Example 8.
(a) $f(x) = 4 - x^2, -3 \leq x \leq 4$.
(b) $f(x) = x^3 - 5x^2 + 6x, -2 \leq x \leq 4$.
(c) $f(x) = 1 - (4/x^2), 1 \leq x \leq 3$.
Answer: (c) $\frac{4}{3}$ square units:



§5. Eudoxus' Method and the Integral

Integrity, integer, integration, integral—these words have the same root meaning, that of "wholeness". To *integrate* is to collect into a whole. What we are now studying is called *integral calculus*, and it is high time to explain why Leibniz chose to call it that. The reason may stem from an observation made by Leibniz [and, before him, by Cavalieri (1598–1647) and others], an observation that may be confirmed by the general area principle. *If, when figures in the plane are set one above the other, they are seen to be made up of equal* "vertical segments", *then the areas of the figures must be equal.* This delightful observation is generally known as **Cavalieri's principle**. A concrete illustration is below.



Here, a vertical line through any point x on the horizontal axis shows a vertical segment of length $x - x^2$ in all three figures. Cavalieri's principle says the three figures must have the same area. This is confirmed by the general area principle, which says that all three figures have an area of $[(x^2/2) - (x^3/3)]_0^1 = \frac{1}{6}$ square units.

A rough* statement of Cavalieri's principle is that the area of a figure is determined by the vertical line segments that make it up. The area of a figure would thus seem to be the result of collecting into a whole, or *integrating*, all its vertical line segments. Leibniz toyed with the idea of regarding any area as an *integral* of (infinitely many) line segments.

This idea raises serious questions. The area of each vertical line segment is of course zero, since a line segment has no width. Yet somehow Leibniz would have us believe that *infinitely many* zeros integrate into a nonzero total area! The paradoxical nature of this idea was recognized by Leibniz, who nevertheless persisted in believing the idea valuable, at least on an intuitive level. Leibniz was never able to describe clearly this intuitive perception, and it has generally been regarded with suspicion.

Nonetheless, by thinking on this intuitive level Leibniz was able to make important discoveries. Justification for some of these discoveries often had to wait for later mathematicians, as Leibniz sometimes had difficulty in saying what he meant. The difficulty is understandable, for it is related to one of the old paradoxes of Zeno (ca. 495–435 B.C.), but further discussion of this is postponed until Chapter 8.

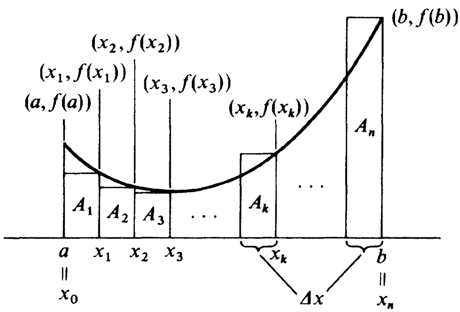
* And, as it stands, quite inaccurate. See problem 21 at the end of this chapter.

The point of the preceding discussion was to explain how the word *integral* entered the calculus. Leibniz wanted to refer to an area as an integral, and out of respect for Leibniz we shall do likewise. However, we discard his fuzzy notion about an "integral of zeros" collecting together to yield a nonzero number. We seek a slight modification of Leibniz's notion of an *integral* to bring things into clear focus. How can this be done?

Once again we turn for help to the notion of a limit, which has already done more than its share to clarify the idea of a tangent to a curve and the idea of continuity. As we shall see, the integral has a natural definition in terms of a limit, by means of a modification of a method introduced by Eudoxus over 2000 years ago.

Eudoxus, of course, never spoke of limits, nor did Archimedes. The Greeks never called limits by name, but could sometimes manage to get the same job done by using the method of elimination. (In modern terms, this amounts to finding an area A by somehow eliminating all numbers larger than A, together with all numbers smaller, leaving the desired number A as the only number left.*) Our experience in Chapter 4 suggests that the use of limits may be preferable to the use of the principle of elimination.

Here, then, is **Eudoxus' method** (in modern dress), defining the **integral** of a function f on the domain $a \le x \le b$.



Idea: *As n gets larger, the area beneath the staircase with n steps approaches the area beneath the curve f.*

Consider any large positive integer n, and divide the interval $a \le x \le b$ into n subintervals,† each of the same length Δx, so that

$$\Delta x = \frac{b-a}{n}. \tag{10}$$

* An example of this method may be found in the appendix on Archimedes.

† Think of n as being the current size of the national debt.

Thus (see the picture) we have $x_0 = a$, $x_1 = a + \Delta x$, $x_2 = a + 2\,\Delta x$, $x_3 = a + 3\,\Delta x$, and so on. Finally, at the last, we have

$$\begin{aligned} x_n &= a + n\,\Delta x \\ &= a + \frac{n(b-a)}{n} = b. \end{aligned}$$

There is a convenient way to abbreviate the preceding two sentences, namely,

$$x_k = a + k\,\Delta x, \quad \text{for } k = 0, 1, 2, \ldots, n. \tag{11}$$

The area A_k of the k-th rectangle (see the picture) is simply the product of its height and width:

$$A_k = f(x_k)\,\Delta x. \tag{12}$$

Therefore, the total area S_n beneath the staircase figure with n steps is the sum $A_1 + A_2 + \cdots + A_n$. We abbreviate this by writing

$$S_n = \sum_{k=1}^{n} A_k$$

(read "S_n equals the sum, as k runs from 1 to n, of A_k"). Substituting the expression (12) for A_k shows

$$S_n = \sum_{k=1}^{n} f(x_k)\,\Delta x. \tag{13}$$

Now S_n is the area beneath the staircase figure with n steps, and it is *not* likely to be equal to the area beneath the curve f. However, as n is taken larger and larger, S_n clearly approximates the area beneath f to great accuracy. The area beneath f is the *limit* of S_n as n increases without bound,* so we define the integral of f to be this limit. That is, the integral of f from a to b is defined as

$$\text{Limit } S_n = \underset{\Delta x \to 0}{\text{Limit}} \sum_{k=1}^{n} f(x_k)\,\Delta x \quad [\text{by (13)}], \tag{14}$$

since (10) shows that $\Delta x \to 0$ as n increases without bound.

The right-hand side of (14) suggests that the integral of f from a to b might be denoted by

$$\int_a^b f(x)\,dx, \tag{15}$$

since this is the symbol that results from replacing the Greek Δ by

* Think of the national debt.

the letter d, the Greek $\sum$ by the letter $\int$ (a seventeenth-century S) and replacing the discrete points x_k by the continuous variable x, which runs from a to b.

Definition

Let f be a function with domain $a \leq x \leq b$. **The integral of f from a to b** is denoted by

$$\int_a^b f(x)\,dx \quad \left(\text{or, for short, by } \int_a^b f\right)$$

and is defined to be the number calculated by Eudoxus' method:

$$\int_a^b f(x)\,dx = \operatorname*{Limit}_{\Delta x \to 0} \sum_{k=1}^{n} f(x_k)\,\Delta x = \text{Limit } S_n,$$

where S_n is defined by equation (13).

The idea of Eudoxus' method is not unlike the idea behind Fermat's method. The "right answer" for the integral $\int_a^b f$ is the limit of "wrong answers" S_n that come quite close to the integral when n is quite large. In integral calculus Eudoxus' method assumes a role of importance parallel to the role played by Fermat's method in differential calculus. It defines the basic notion to be studied.

Just as we have found shortcuts to Fermat's method, so we can find shortcuts to the method of Eudoxus. We can sometimes guess the value of an integral by interpreting the integral as an area. For instance, from exercise 3.1(d) we may expect that

$$\int_1^4 f(x)\,dx = 63\frac{3}{4}, \quad \text{where } f(x) = x^3,$$

or, more briefly,

$$\int_1^4 x^3\,dx = 63\frac{3}{4}.$$

If the variable is called t instead of x, we modify our notation accordingly. From exercise 1.1, without using Eudoxus' method, we expect to have

$$\int_1^4 (10-2t)\,dt = 15, \quad \int_1^4 t\,dt = \frac{15}{2}, \quad \int_1^4 (4t-3)\,dt = 21.$$

These examples should suggest that integrals, defined by Eudoxus' method, can be calculated by the method of antiderivatives. This is true, and it is essentially the content of the fundamental theorem. Before the fundamental theorem can be appreciated, however, we must learn to be at home with

(a) interpreting integrals as areas (being careful, because an integral is not always an area).
(b) calculating integrals by Eudoxus' method (as a limit of sums).

The first of these should be accomplished by the following exercises. Section 6 deals with the second.

Exercises

5.1. In exercise 3.1, ten areas were found. Express each of these areas as integrals, and express the answers you found in exercise 3.1 in integral notation. *Answers*: (b) $\int_1^4 (1/t^2)\,dt = \frac{3}{4}$. (c) $\int_1^4 x^2\,dx = 21$. (h) $\int_1^4 \pi\,dt = \int_1^4 \pi\,dx = \int_1^4 \pi\,ds = 3\pi$.

5.2. Interpret each of the following integrals as an area.

(a) $\int_0^4 (8x - 2x^2)\,dx$. (b) $\int_{-1}^1 (3x^2 + 1)\,dx$.
(c) $\int_0^2 (x^3 - 5x^2 + 6x)\,dx$. (d) $\int_0^5 \pi\,dt$.
(e) $\int_{-3}^3 \sqrt{9 - t^2}\,dt$. (f) $\int_{-2}^2 \sqrt{4 - t^2}\,dt$.
(g) $\int_0^5 \sqrt{25 - t^2}\,dt$. (h) $\int_{-3}^0 \sqrt{9 - x^2}\,dx$.

Answers: (a) This integral is equal to the area A_2 of exercise 4.1. (d) This integral is equal to the area beneath the curve $y = \pi$, from $t = 0$ to $t = 5$. (e) This integral is equal to the area beneath the curve $y = \sqrt{9 - t^2}$, a semicircle (*why?*), from $t = -3$ to $t = 3$.

5.3. Evaluate each of the integrals in 5.2 by some means *other than* Eudoxus' method. Answer: (e) The area of a semicircle of radius 3 is equal to half the area of the full circle, or $9\pi/2$. Therefore, $\int_{-3}^3 \sqrt{9 - t^2}\,dt = 9\pi/2$.

5.4. If a function f is negative, i.e., if its graph lies below the horizontal axis, then all the A_k's of equation (12) are negative. Use this to explain why, *when f is negative*, then $\int_a^b f(x)\,dx$ will *not* be an area. *Hint*. No area is negative.

§6. The Integral as a Limit of Sums

It takes time and patience to carry out Eudoxus' method of calculating an integral. Since the antiderivative method is shorter, one may ask why time should be spent studying Eudoxus. There are several reasons.

(a) $\int_a^b f$ cannot be calculated by antiderivatives unless an antiderivative of f is known. [There are many functions, such as $1/x$ and $1/(1 + x^2)$, whose antiderivatives we have not yet met.]
(b) Eudoxus' method leads to a clear understanding of what is meant by the *area beneath a curve*.

(c) Eudoxus' method emphasizes that an integral is a limit of sums. Areas are not the only quantities that are limits of sums. As we shall see, *volumes* can also be regarded as limits of sums, and they can be expressed by integrals. Integrals are of use in expressing other quantities as well, such as the quantity of *work* required to put a satellite in orbit.

Eudoxus' method involves the sum of n numbers, where n is a large integer. Such a sum must be simplified before the limit can be taken in order to find the integral. There is one such sum which, thanks to the Pythagoreans, we know how to calculate already.

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}. \tag{16}$$

(See Chapter 2, exercise 1.5.) The formula for the sum of the *squares* of the first n positive integers has also been known since antiquity:

$$1 + 4 + 9 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}. \tag{17}$$

Archimedes' proof of formula (17) was outlined in problem 8 at the end of Chapter 3 and a modern (seventeenth-century) proof may be found in Section 1 of Appendix 2.

To deal efficiently with sums, an efficient system of notation must be developed. The symbol

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

is used as an indication to sum up n numbers that are to be indexed by k. We refer to k as the *index of summation*. For instance,

$$\sum_{k=1}^{3} k = 1 + 2 + 3 = 6,$$

because $\sum_{k=1}^{3} k$ indicates the sum of 3 numbers, the numbers being expressed by k, where k runs from 1 to 3. Similarly,

$$\sum_{k=1}^{3} 5k = 5 \cdot 1 + 5 \cdot 2 + 5 \cdot 3 = 5(1 + 2 + 3). \tag{18}$$

This simplifies, of course, to $5 \cdot 6 = 30$, but it is more important to note that equation (18) shows that

$$\sum_{k=1}^{3} 5k = 5 \cdot \sum_{k=1}^{3} k. \tag{19}$$

What is the "real reason" that the number 5 can be brought out in front of the summation sign, as in (19)? It is that 5 is an expression which is independent of the index k. It therefore occurs in each of the summands and can be factored out in front, as seen in (18).

What has just been illustrated in the simple example given above is most important to remember when trying to simplify sums. *Note when an expression can be brought out in front of the summation sign.* We can see, for example, that

$$\sum_{k=1}^{n}\left(\frac{1}{n^2}\right)k=\left(\frac{1}{n^2}\right)\cdot\sum_{k=1}^{n}k, \tag{20}$$

simply because the expression $1/n^2$ is independent of the index k. In fact, beginning with equation (20), we can carry out a complete simplification as follows. We already know how to simplify the sum that occurs in the right-hand side of (20). Equation (16) says

$$\sum_{k=1}^{n}k=\frac{n(n+1)}{2}. \tag{21}$$

[*Do you see why equation* (16) *says exactly this*?] When this is used in (20) we get

$$\sum_{k=1}^{n}\left(\frac{1}{n^2}\right)k=\left(\frac{1}{n^2}\right)\frac{n(n+1)}{2}=\frac{1}{2}\left(\frac{n}{n}\right)\left(\frac{n+1}{n}\right)=\frac{1}{2}\left(1+\frac{1}{n}\right). \tag{22}$$

As a first example of Eudoxus' method, let us calculate an integral whose value we already know by other means.

Example 9

Calculate $\int_0^1 x\,dx$ directly from its definition by applying the method of Eudoxus.

Here, we must apply Eudoxus' method to the function given by $f(x)=x$ on the domain $0\le x\le 1$. Thus we have $a=0$, $b=1$, and

$$\Delta x=\frac{b-a}{n}=\frac{1-0}{n}=\frac{1}{n}\quad\text{[from (10)]},$$

$$x_k=a+k\,\Delta x=0+k\left(\frac{1}{n}\right)=\frac{k}{n}\quad\text{[from (11)]},$$

$$f(x_k)=x_k=\frac{k}{n}.$$

Using these, we first find an approximation S_n to the desired integral. From (13),

$$S_n=\sum_{k=1}^{n}f(x_k)\,\Delta x=\sum_{k=1}^{n}\left(\frac{k}{n}\right)\left(\frac{1}{n}\right)=\sum_{k=1}^{n}\left(\frac{1}{n^2}\right)k.$$

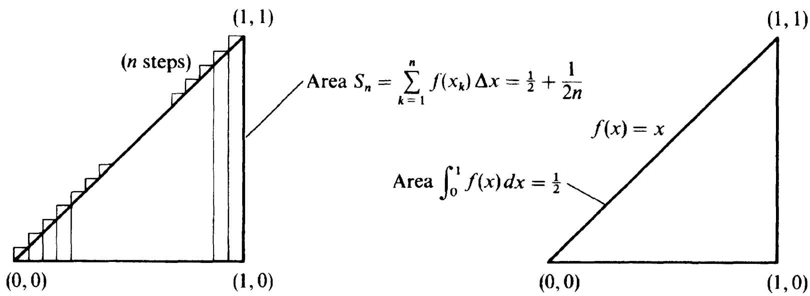
The integral, by definition, is the limit of S_n as n increases without bound. In order to find that limit we must first simplify the expression for S_n. This has already been carried out in (22), so we have

$$S_n = \frac{1}{2}\left(1 + \frac{1}{n}\right). \tag{23}$$

From (23), it is easy to find Limit S_n, for it is obvious that, as n grows increasingly larger, $1/n \to 0$. Therefore,

$$\int_0^1 x\,dx = \text{Limit } S_n = \text{Limit } \frac{1}{2}\left(1 + \frac{1}{n}\right) = \frac{1}{2}(1 + 0) = \frac{1}{2}. \qquad \square$$

Example 9 shows that Eudoxus' method, like Fermat's method, can be carried out without ever drawing a geometric picture to describe what goes on. A picture aids the understanding, however, so let us draw one. What was shown in (23) is that the area S_n of the staircase figured with n steps is equal to $\frac{1}{2} + 1/2n$ square units. As n gets larger (or, equivalently, as $\Delta x \to 0$), the jagged figure on the left approximates more and more the area on the right.



The integral of Example 9 is, of course, calculated much more quickly by simply using the formula for the area of a triangle. Or, by the anti-derivative method,

$$\int_0^1 x\,dx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}. \qquad \square$$

Before doing a second example it might be well to make a small point about summations. Here is a question that is easy to miss because it is too simple. *What is*

$$\sum_{k=1}^{3} 1$$

equal to? To answer this question, remember that "$\sum_{k=1}^{3}$" indicates that 3 numbers are to be summed. For instance,

$$\sum_{k=1}^{3} A_k = A_1 + A_2 + A_3.$$

If $A_k = 1$ (that is, if $A_1 = 1, A_2 = 1, A_3 = 1$), this becomes

$$\sum_{k=1}^{3} 1 = 1 + 1 + 1 = 3,$$

answering our question. By the same token we see that

$$\sum_{k=1}^{n} 1 = \underbrace{1 + 1 + 1 + \cdots + 1}_{n \text{ summands}} = n. \tag{24}$$

EXAMPLE 10

Calculate $\int_1^4 2x\,dx$ directly from its definition by applying the method of Eudoxus.

We apply Eudoxus' method to the function given by $f(x) = 2x$ on the domain $1 \le x \le 4$. Thus we have $a = 1$, $b = 4$, and

$$\Delta x = \frac{b-a}{n} = \frac{4-1}{n} = \frac{3}{n} \quad \text{[from (10)]},$$

$$x_k = a + k\,\Delta x = 1 + k\left(\frac{3}{n}\right) = 1 + \frac{3k}{n} \quad \text{[from (11)]},$$

$$f(x_k) = 2x_k = 2 + \frac{6k}{n}.$$

Hence,

$$\begin{aligned}
S_n = \sum_{k=1}^{n} f(x_k)\,\Delta x &= \sum_{k=1}^{n}\left(2 + \frac{6k}{n}\right)\left(\frac{3}{n}\right)\\
&= \sum_{k=1}^{n}\left(\frac{6}{n} + \frac{18k}{n^2}\right)\\
&= \sum_{k=1}^{n}\frac{6}{n} + \sum_{k=1}^{n}\frac{18k}{n^2} \quad (\textit{why?})\\
&= \frac{6}{n}\sum_{k=1}^{n} 1 + \frac{18}{n^2}\sum_{k=1}^{n} k \quad (\textit{why?})\\
&= \frac{6}{n}(n) + \frac{18}{n^2}\,\frac{n(n+1)}{2} \quad \text{[by (21) and (24)]}\\
&= 6\left(\frac{n}{n}\right) + \left(\frac{18}{2}\right)\left(\frac{n}{n}\right)\left(\frac{n+1}{n}\right) = 6 + 9\left(1 + \frac{1}{n}\right).
\end{aligned}$$

Therefore,

$$\int_1^4 2x\,dx = \text{Limit } S_n = \text{Limit } 6 + 9\left(1 + \frac{1}{n}\right) = 6 + 9(1+0) = 15. \qquad \square$$

By comparison with Eudoxus' method, antiderivatives evaluate integrals like lightning:

$$\int_1^4 2x\,dx = x^2\big|_1^4 = 16 - 1 = 15.$$

The point of these examples, however, has nothing to do with speed of calculation. Only an electronic computer would regard Eudoxus' method as speedy. The point is to emphasize that the integral is a limit of sums and can be calculated without reference to any geometric figure and without any knowledge whatever of derivatives or antiderivatives.

The integral $\int_a^b f$ does have a geometric interpretation, however, as the area beneath the curve f, *if f is not negative*. By another stroke of good fortune, the integral enjoys a connection with antiderivatives, to be stated precisely in the fundamental theorem. Since such delightful connections can be proved to be true, the most intellectual of minds might regard them as unsurprising, being merely part of the nature of things. Some of the rest of us, who know the meaning of serendipity, happily find it here.

Exercises

(Be willing to put in a little time practicing the use of summation notation. It is quite efficient, once learned. The appendix on sums and limits may be helpful.)

6.1. Go through the following steps to calculate the integral $\int_0^7 (3x+2)\,dx$.

(a) Use formula (10) of Section 5 to find Δx. *Answer*: $\Delta x = (7-0)/n = 7/n$.

(b) Use formula (11) to find x_k. *Answer*: $x_k = 0 + k\,\Delta x = 0 + k(7/n) = 7k/n$.

(c) Find $f(x_k)$. *Answer*: Here we have $f(x) = 3x + 2$, so $f(x_k) = 3x_k + 2 = (21k/n) + 2$.

(d) Use formula (13) with your answers to (a) and (c) to find S_n.

Answer: $S_n = \sum_{k=1}^n f(x_k)\,\Delta x = \sum_{k=1}^n \left(\frac{21k}{n} + 2\right)\left(\frac{7}{n}\right)$.

(e) (*The hard part*) Simplify S_n by using (21) and (24), as illustrated in Example 10.

Answer: $S_n = (147/2)(1 + 1/n) + 14$.

(f) The integral is defined as the limit of S_n. Find this limit by using your answer to (e).

Answer: $\int_0^7 (3x+2)\,dx = \text{Limit } (147/2)(1+1/n)+14 = (147/2)(1+0)+14$
$= 175/2.$

(g) Check your answer by interpreting the integral as an area as in exercise 5.2.
Answer: $\int_0^7 (3x+2)\,dx = [\text{area beneath } 3x+2 \text{ from 0 to 7}]$
$= \left[\frac{3x^2}{2}+2x\right]_0^7 = \frac{175}{2}.$

6.2. Calculate each of these integrals by going through steps (a)–(f) in exercise 6.1.
(a) $\int_0^4 (5x+1)\,dx.$
(b) $\int_0^2 (5x+1)\,dx.$
(c) $\int_2^4 (5x+1)\,dx.$
(d) $\int_2^4 (1-5x)\,dx.$
Answer: (d) -28. (The integral here is not an area, since the function $1-5x$ is negative on the domain $2 \leq x \leq 4$.)

6.3. Explain why, in exercise 6.2, it is to be expected that the sum of your answers to parts (b) and (c) is equal to the answer to part (a).

6.4. Write formula (17) in summation notation. *Answer*: $\sum_{k=1}^n k^2 = n(n+1) \times (2n+1)/6.$

6.5. Use your answer to exercise 6.4 to help calculate $\int_0^1 x^2\,dx$ directly from its definition by Eudoxus' method. *Answer*: $\frac{1}{3}$. (See Appendix 2, Section 2.)

§7. Some Properties of the Integral

In Section 2 we guessed that a certain area was equal to 6 square units without knowing, at that time, what was meant by the area within a figure bounded by a curve. We now have Eudoxus' method of determining such an area, and we can therefore check our guess of Section 2. Let us do that, with an eye out for noticing some properties of the integral.

Consider, then, the function given by

$$f(x) = x^2 - 4x + 5, \quad 1 \leq x \leq 4.$$

Applying the method of Eudoxus to find the area beneath f, we have $\Delta x = 3/n$, $x_k = 1 + 3k/n$, and

$$S_n = \sum_{k=1}^n f(x_k)\Delta x = \sum_{k=1}^n [x_k^2 - 4x_k + 5]\Delta x \tag{25}$$
$$= \sum_{k=1}^n \left[\left(1+\frac{3k}{n}\right)^2 - 4\left(1+\frac{3k}{n}\right)+5\right]\left(\frac{3}{n}\right)$$

$$= \sum_{k=1}^{n} \frac{27k^2}{n^3} - \frac{18k}{n^2} + \frac{6}{n} \quad \text{(by collecting terms)}$$

$$= \frac{27}{n^3} \sum_{k=1}^{n} k^2 - \frac{18}{n^2} \sum_{k=1}^{n} k + \frac{6}{n} \sum_{k=1}^{n} 1$$

$$= \frac{27}{6}\left(1 + \frac{1}{n}\right)\left(2 + \frac{1}{n}\right) - 9\left(1 + \frac{1}{n}\right) + 6,$$

by (16), (17), and (24). (The reader is asked to fill in the missing steps in this calculation.) Since $1/n \to 0$ as n gets larger, it is easy to take the limit of S_n, which gives the area beneath f. The area beneath f is then equal to

$$\begin{aligned} \text{Limit } S_n &= \text{Limit}\left[\frac{27}{6}\left(1 + \frac{1}{n}\right)\left(2 + \frac{1}{n}\right) - 9\left(1 + \frac{1}{n}\right) + 6\right] \\ &= \frac{27}{6}(1)(2) - 9(1) + 6 \\ &= 6 \text{ square units.} \end{aligned}$$

This confirms our guess, and shows that

$$\int_1^4 (x^2 - 4x + 5)\, dx = 6.$$

Looking back over the calculation given above, we can notice an important property of the integral. From line (25), we see that

$$S_n = \sum_{k=1}^{n} x_k^2 \,\Delta x - \sum_{k=1}^{n} 4x_k \,\Delta x + \sum_{k=1}^{n} 5 \,\Delta x. \tag{26}$$

What happens to this equation "in the limit"? As n increases without bound, equation (26) becomes (*do you see why?*)

$$\int_1^4 (x^2 - 4x + 5)\, dx = \int_1^4 x^2\, dx - \int_1^4 4x\, dx + \int_1^4 5\, dx.$$

This suggests that *the integral of a sum of functions is equal to the sum of their integrals*. This is true:

Sum Rule for Integrals

If the functions f and g have integrals on the domain $a \le x \le b$, then

$$\int_a^b (f(x) + g(x))\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx.$$

Proof

$$\int_a^b (f(x)+g(x))\,dx = \text{Limit} \sum_{k=1}^{n} (f(x_k)+g(x_k))\Delta x$$
$$= \text{Limit} \sum_{k=1}^{n} f(x_k)\,\Delta x + \text{Limit} \sum_{k=1}^{n} g(x_k)\,\Delta x$$
$$= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx. \qquad \square$$

What about a rule for constant multiples? Is it true that $\int_1^4 4x\,dx$ is equal to $4\int_1^4 x\,dx$? Sure it is:

Constant-Multiple Rule for Integrals
If f has an integral on the domain $a \le x \le b$, then for any constant c,

$$\int_a^b c\cdot f(x)\,dx = c\int_a^b f(x)\,dx.$$

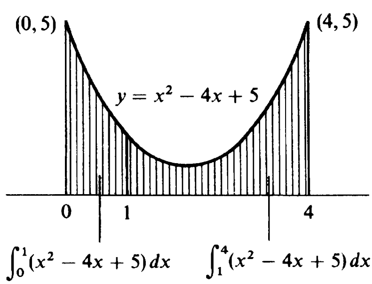
Proof
We know that

$$\sum_{k=1}^{n} c\cdot f(x_k)\,\Delta x = c\sum_{k=1}^{n} f(x_k)\,\Delta x, \tag{27}$$

since the constant c is independent of the index of summation. Since equation (27) holds for each n, no matter how large, we get, in the limit,

$$\int_a^b c\cdot f(x)\,dx = c\int_a^b f(x)\,dx. \qquad \square$$

Another property of the integral is suggested by this figure.



Since the total shaded area is equal to $\int_0^4 (x^2 - 4x + 5)\,dx$, we know that $\int_0^1 (x^2 - 4x + 5)\,dx + \int_1^4 (x^2 - 4x + 5)\,dx = \int_0^4 (x^2 - 4x + 5)\,dx$. We are led to suspect that, in general, if $\int_0^4 f$ exists, then so does $\int_0^1 f$ and $\int_1^4 f$, and we have the *additivity property*

$$\int_0^1 f + \int_1^4 f = \int_0^4 f.$$

We can incorporate this into an existence theorem.

Existence Theorem for Integrals

If f is a continuous function throughout the domain $a \leq x \leq b$, then the integral

$$\int_a^b f = \int_a^b f(x)\,dx$$

exists. Moreover, if t is between a and b, then

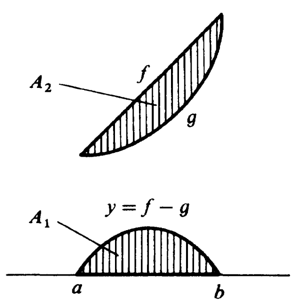
$$\int_a^b f = \int_a^t f + \int_t^b f.$$

What does it mean to say that an integral "exists"? It means, simply, that Limit S_n exists, where S_n is the approximating sum from Eudoxus' method. The limit will exist, according to the theorem given above, if f is continuous. But the proof of the existence theorem is better left to a course in analysis. Let us take it for granted that f *has an integral from a to b if f is continuous on $a \leq x \leq b$.*

Exercises

7.1. Fill in the missing steps in the calculation of S_n that begins with equation (25).

7.2. Consider the areas A_1 and A_2 in the figure.



Cavalieri's principle says that $A_1 = A_2$. Prove that this is true by showing, in order,

(a) $A_1 = \int_a^b (f - g)$.

(b) $A_2 = \int_a^b f - \int_a^b g$. [Draw pictures of the areas measured by these integrals.]

(c) $A_1 = A_2$. [Use (a), (b), and the rule for sums and constant multiples.]

7.3. Is the integral of a product equal to the product of the integrals?

7.4. Use Eudoxus' method to calculate

(a) $\int_0^1 (x^2 - 4x + 5)\,dx$.

(b) $\int_0^4 (x^2 - 4x + 5)\,dx$.

7.5. Is it true that $\int_1^4 \frac{1}{x}\,dx + \int_4^{4\pi} \frac{1}{x}\,dx = \int_1^{4\pi} \frac{1}{x}\,dx$?
Hint. Use the additivity property discussed in this section.

7.6. Attempt to calculate $\int_0^1 (1/x^2)\,dx$ by Eudoxus' method.

(a) Show that $S_n = \sum_{k=1}^{n} n/k^2$.

(b) What is S_1? S_2? S_3? *Partial answer*: $S_2 = 2.5$.

(c) Show that S_n is never less than n. *Hint*. Show $S_n = n(1 + \cdots)$.

(d) Does Limit S_n exist? *Answer*: In view of part (c), S_n cannot tend to a limit, since it grows arbitrarily large as n increases.

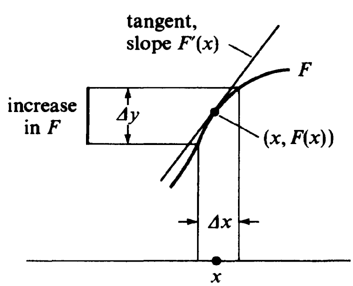
(e) Does $\int_0^1 (1/x^2)\,dx$ exist? *Hint*. By definition, the integral is equal to Limit S_n. Use part (d).

(f) Does your answer to part (e) contradict the existence theorem for integrals? Why not?

(g) Draw a picture of the area that the integral $\int_0^1 (1/x^2)\,dx$ is "trying" to measure. Why can it not be measured?

§8. The Fundamental Theorem

The fundamental theorem shows the connection between the two branches of calculus, *differential* and *integral*. The connection is really between Fermat's method and Eudoxus' method, of course. To prepare the way for the fundamental theorem, let us review Fermat's method, using the notation of Leibniz.



Consider a function F and a point x, and let Δx be length of a small interval that contains x. Then, by Fermat's method, it follows that

$$F'(x) = \underset{\Delta x \to 0}{\text{Limit}} \frac{\Delta y}{\Delta x}.$$

This means, roughly speaking, that the number $\Delta y/\Delta x$ is very close to $F'(x)$ when Δx is very close to (but not equal to) zero. In symbols,

$$F'(x) \approx \frac{\Delta y}{\Delta x}, \quad \text{provided } \Delta x \approx 0,$$

where "$\approx$" stands for "approximate equality". It will serve our purpose to rewrite this as

$$F'(x) - \frac{\Delta y}{\Delta x} \approx 0, \quad \text{provided } \Delta x \approx 0. \tag{28}$$

Here, x is fixed, and the expression $F'(x) - (\Delta y/\Delta x)$ varies in terms of Δx. Let the letter o stand for this expression:

$$F'(x) - \frac{\Delta y}{\Delta x} = o. \tag{29}$$

By (28), we know something about the variable o:

$$o \approx 0, \quad \text{provided } \Delta x \approx 0. \tag{30}$$

From (29), upon multiplying through by Δx,

$$F'(x)\,\Delta x - \Delta y = o\,\Delta x,$$

$$F'(x)\,\Delta x = \Delta y + o\,\Delta x. \tag{31}$$

Equation (31), in connection with the information given in (30), is the key to the proof of the fundamental theorem. Note that (31) is, so to speak, what one gets by beginning with Fermat's method and "undoing it". Roughly speaking, (31) says that when the derivative is multiplied by Δx, a small change in x, you get a close approximation to Δy, the corresponding change in y. Remember that o is not 0, but rather a variable tending to 0 as $\Delta x \to 0$.

The Fundamental Theorem of Calculus
If f is a continuous function with domain $a \le x \le b$, then

$$\int_a^b f(x)\,dx = F(b) - F(a),$$

where F is any antiderivative of f.

Proof
(Didn't we prove this already when we proved the area principle? *Answer*: No. The integral of f is not always an area. The fundamental

theorem asserts that the antiderivative method works even when the function f is not always positive.)

Since we know that F is an antiderivative of f, equation (31) says

$$f(x)\,\Delta x = \Delta y + o\,\Delta x,$$

where Δy is the change in F corresponding to the change Δx in x. Applying this to the k-th subinterval in Eudoxus' method, we have

$$f(x_k)\Delta x = \text{change in } F \text{ on } k\text{-th subinterval} + o_k\,\Delta x.$$

Hence the approximating sum to the integral $\int_a^b f$ can be expressed as

$$\sum_{k=1}^{n} f(x_k)\,\Delta x = \sum_{k=1}^{n} \text{change in } F \text{ on } k\text{-th subinterval} + \sum_{k=1}^{n} o_k\,\Delta x. \tag{32}$$

It is obvious that

$$\begin{aligned}\sum_{k=1}^{n} \begin{matrix}\text{change in } F \text{ on}\\ k\text{-th subinterval}\end{matrix} &= \begin{matrix}\text{change in } F \text{ on the}\\ \text{entire interval } a \le x \le b\end{matrix}\\ &= F(b) - F(a).\end{aligned}$$

Therefore, from (32),

$$\sum_{k=1}^{n} f(x_k)\,\Delta x = F(b) - F(a) + \sum_{k=1}^{n} o_k\,\Delta x,$$

and, taking the limit as $\Delta x \to 0$, we get

$$\begin{aligned}\int_a^b f(x)\,dx &= F(b) - F(a) + \int_a^b 0\,dx \quad \text{[by (30)]}\\ &= F(b) - F(a).\end{aligned}$$

□

The careful reader may feel that the last step in the proof given above does not justify adequately the fact that

$$\operatorname*{Limit}_{\Delta x \to 0} \sum_{k=1}^{n} o_k\,\Delta x = \int_a^b 0\,dx. \tag{33}$$

The careful reader is right. Although (33) is surely made plausible by (30), it has not been justified rigorously in the above proof. Rigorous proof of (33) is better deferred to a course in analysis.

Exercises

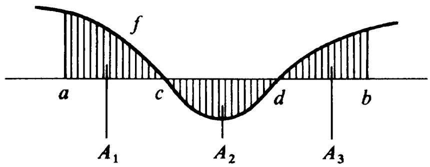
8.1. (a) Evaluate the integral $\int_{-1}^{1} x\,dx$ by the antiderivative method.
(b) Evaluate $\int_{-1}^{1} x\,dx$ by Eudoxus' method. *Answer*: 0.

(c) Do your answers to (a) and (b) agree, as the fundamental theorem asserts?

(d) Can the integral in question be regarded as an *area*?

8.2. Prove that if the continuous curve f crosses the x-axis, then the integral $\int_a^b f$ gives the *algebraic* sum of the areas between the curve f and the axis, counting area above as positive and below as negative. *Hint.* In the picture below, you want to show that

$$\int_a^b f = A_1 - A_2 + A_3.$$



Prove this by giving three reasons—one for each of the following equalities:

$$\int_a^b f = \int_a^c f + \int_c^d f + \int_d^b f$$

$$= \int_a^c (f - 0) - \int_c^d (0 - f) + \int_d^b (f - 0) = A_1 - A_2 + A_3.$$

8.3. Evaluate each of the following integrals by using the fundamental theorem.

(a) $\int_0^2 (1 - x^3)\, dx$. (b) $\int_0^2 3\, dx$.

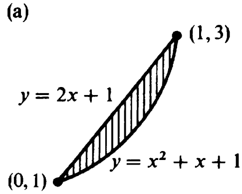
(c) $\int_{-1}^4 3x^2\, dx$. (d) $\int_{-1}^1 (\pi - \pi x^2)\, dx$.

(e) $\int_1^{10} (1/x^2)\, dx$. (f) $\int_{-10}^{-1} (1/x^2)\, dx$.

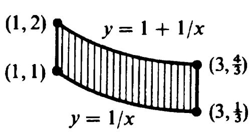
Answers: (d) $4\pi/3$. (e) $\frac{9}{10}$. (f) $\frac{9}{10}$.

8.4. First express each of the following areas as an integral. Then evaluate the integral, using the fundamental theorem.

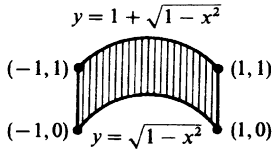
(a)



(b)



(c)



Answer: (a) $\int_0^1 (x - x^2)\,dx = [(x^2/2) - (x^3/3)]_0^1 = \frac{1}{6}$ square units.

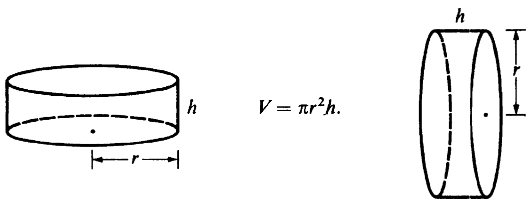
8.5. (*For careful readers*) What is wrong with the following "calculation"? $\int_{-1}^{1} (1/x^2)\,dx = -1/x|_{-1}^{1} = -2.$

8.6. Consider the integral $\int_0^7 \pi \frac{25}{49} x^2\,dx$.
   (a) Evaluate this integral, using the fundamental theorem.
   (b) Draw a picture of an area that is represented by this integral. (On the following pages, we shall see that this same integral also represents a *volume*.)

§9. Integrals and Volumes

Integrals, defined by Eudoxus' method, arise naturally in many contexts having nothing to do with area. Yet the fundamental theorem can still be used to evaluate the integral, provided an appropriate antiderivative can be found. This is why the fundamental theorem is of much more significance than the area principle. Many illustrations of this may be seen in Chapter 8.

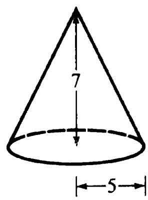
One illustration is readily at hand. Let us consider the problem of finding volumes of *solids of revolution*. The only thing we need to know at the outset is the formula for the volume of a cylinder. It is given by the product of the area of the circular base and the height.



Working carefully through an example will enable us to see a shortcut way of working many similar examples. The key is to try to express the volume desired as an integral. The integral can then be evaluated by the fundamental theorem.

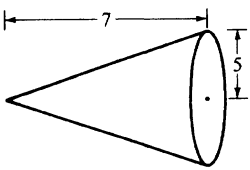
EXAMPLE 11

Determine the volume of a cone if its height is 7 feet and if the radius of its base is 5 feet.

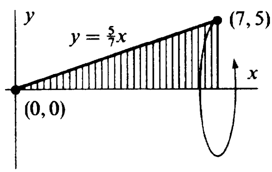


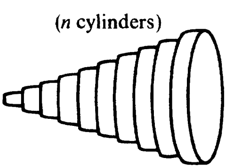
What is meant by the *volume* of a solid figure? This question is easily answered by means of the notion of a *limit*. We can get the volume by approximating it ever more closely and then obtaining it exactly as the limit of our approximations. The desired volume, we shall see, will turn out to be the limit of a sum, just as in Eudoxus' method. That is, the desired volume will turn out to be an integral.

Let us carry out this procedure. If we turn our given cone on its side, we see that it could be regarded as the solid figure obtained by revolving the indicated area 360 degrees about the horizontal axis. Such a solid figure is called a *solid of revolution*. The volume of any solid of revolution is easy to obtain by the method described below.

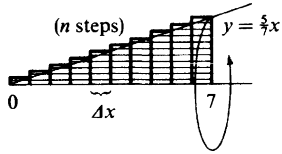


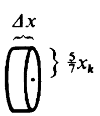
Our cone is the solid obtained by revolving the area beneath the line $y = \frac{5}{7}x$, between 0 and 7.



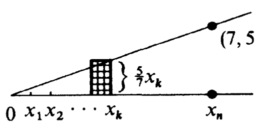


A "jagged cone" that approximates ours is obtained by revolving the staircase figure.





The kth cylinder comes from revolving the kth step.



From the formula for the volume of a cylinder, the volume of the k-th cylinder is clearly given by

$$\pi\left(\frac{5}{7}x_k\right)^2 \Delta x = \pi\frac{25}{49}x_k^2\,\Delta x.$$

The jagged cone is made up of n cylinders. Its volume is the sum of the volumes of these cylinders:

$$\begin{matrix}\text{Volume of}\\ \text{jagged cone}\end{matrix} = \sum_{k=1}^{n} \begin{matrix}\text{volume of}\\ k\text{-th cylinder}\end{matrix} = \sum_{k=1}^{n} \pi\frac{25}{49}x_k^2\,\Delta x. \tag{34}$$

As $\Delta x \to 0$, the jagged cone approximates our given cone ever more closely. Therefore,

$$\begin{aligned}\text{Volume of given cone} &= \operatorname*{Limit}_{\Delta x\to 0}(\text{volume of jagged cone})\\ &= \operatorname*{Limit}_{\Delta x\to 0}\sum_{k=1}^{n}\pi\frac{25}{49}x_k^2\,\Delta x \quad [\text{by } (34)].\end{aligned}$$

This says that the volume of our given cone is equal to the limit of a sum, i.e., to an *integral*. What integral is it? The domain is surely $0 \le x \le 7$, because the points x_k subdivide that domain. Clearly, then,

$$\begin{aligned}\operatorname*{Limit}_{\Delta x\to 0}\sum_{k=1}^{n}\pi\frac{25}{49}x_k^2\,\Delta x &= \int_0^7 \pi\frac{25}{49}x^2\,dx\\ &= \pi\frac{25}{49}\frac{x^3}{3}\bigg|_0^7 \approx 183.26.\end{aligned}$$

The volume of a cone, with $h = 7$ feet and $r = 5$ feet, is then given by

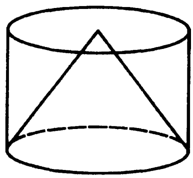
$$\frac{\pi(25)(7)}{3} \approx 183.26 \text{ cubic feet.} \qquad \square$$

One might conjecture that the volume of a cone of height h and radius r is given by $\pi r^2 h/3$. This seems to be what the answer to Example 11 is trying to tell us. The reader is asked to verify this conjecture in an exercise to follow.

Exercises

9.1. Determine the volume of a cone of height h and radius r. (Just work through each step of Example 11, but with h in place of 7 and with r in place of 5.)

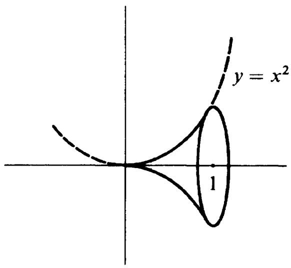
9.2. Compare a cone with a cylinder of the same base and height. Using your answer to exercise 9.1, find the ratio of the volume of the cylinder to the volume of the inscribed cone.



Answer: The ratio is 3:1, first proved by the man himself, Eudoxus of Cnidus, as an application of the method that now bears his name.

9.3. Suppose it is desired to cut a cone parallel to its base, in such a way that the two resulting pieces have the same volume. Where should the cut be made?

9.4. In Example 11, the area beneath the line $y = \frac{5}{7}x$, $0 \leq x \leq 7$, was revolved about the x-axis, and the volume of the resulting solid of revolution found. Suppose instead we revolve the area beneath the quadratic $y = x^2$, $0 \leq x \leq 1$.



Let V be the volume of the resulting solid, and let the points x_k subdivide the interval $0 \leq x \leq 1$, as in Eudoxus' method. For each of the equalities that follow, give a reason to justify it.

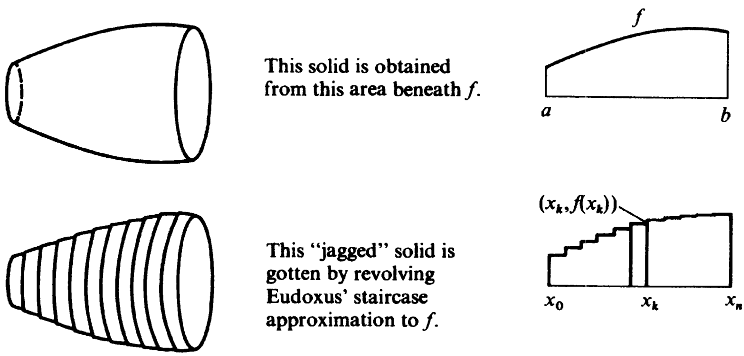
$$V = \operatorname*{Limit}_{\Delta x \to 0} \sum_{k=1}^{n} \pi x_k^4 \, \Delta x$$

$$= \int_0^1 \pi x^4 \, dx = \pi \int_0^1 x^4 \, dx = \pi \left[\frac{x^5}{5}\right]_0^1 = \frac{\pi}{5} \text{ cubic units.}$$

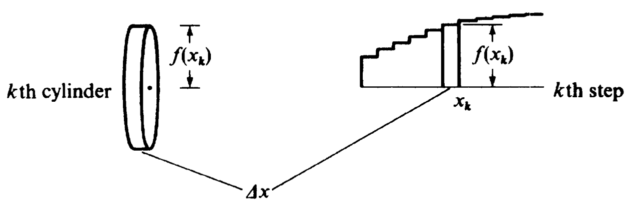
§10. The Volume of a Solid of Revolution

We found the volume of a cone in Section 9 by regarding the cone as a solid of revolution. Thus its volume could be approximated by a "jagged" solid of revolution, then calculated exactly as an integral. Exactly the same procedure will give us an integral formula for the volume of any solid of revolution.

Consider the solid of revolution obtained by revolving the area beneath a continuous curve f, with domain $a \le x \le b$.



The jagged solid is made up of n cylinders, if the staircase has n steps. The k-th cylinder comes from revolving the k-th step:



Since the volume of the k-th cylinder is $\pi(f(x_k))^2\,\Delta x$, the volume of the jagged solid is

$$\sum_{k=1}^{n} \pi(f(x_k))^2\,\Delta x.$$

As $\Delta x \to 0$, the jagged solid's volume tends to

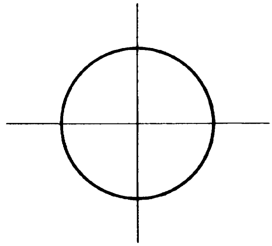
$$\int_a^b \pi(f(x))^2\,dx. \tag{35}$$

Formula (35) *then gives the volume of the solid of revolution obtained by revolving the area beneath the curve f, from $x = a$ to $x = b$, about the x-axis.*

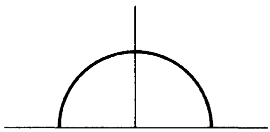
EXAMPLE 12
Find the volume of a sphere whose radius is 7 meters.

A sphere can be regarded as the solid of revolution obtained by revolving a semicircle about its diameter.

$x^2 + y^2 = 49$



Circle, radius 7



Semicircle, radius 7

The equation

$$y = \sqrt{49 - x^2}, \quad -7 \leq x \leq 7,$$

describes a semicircle of radius 7 whose diameter lies on the x-axis. By formula (35), the volume of a sphere of radius 7 meters is given by

$$\begin{aligned} \int_{-7}^{7} \pi(\sqrt{49 - x^2})^2 \, dx &= \pi \int_{-7}^{7} (49 - x^2) \, dx \\ &= \pi \left[49x - \frac{x^3}{3} \right]_{-7}^{7} \\ &= \pi 7^3 \left(\frac{4}{3} \right) \text{cubic meters.} \end{aligned}$$ □

One might conjecture that the volume of a sphere of radius r is given by $4\pi r^3/3$. That could be what the answer to Example 12, where $r = 7$, is trying to tell us. The reader is asked to verify this conjecture in an exercise to follow.

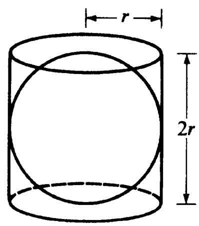
Exercises

10.1. Use formula (35) to find the volumes of the solids of revolution obtained by revolving the areas under each of the following curves.
(a) $f(x) = x^2, 0 \leq x \leq 5$.
(b) $f(x) = x^2, -4 \leq x \leq 4$.
(c) $f(x) = x + 1, 0 \leq x \leq 3$.
(d) $f(x) = \sqrt{1 + x^2}, -1 \leq x \leq 2$.
Answers: (a) 625π cubic units. (c) 21π cubic units.

10.2. Consider the integral $\int_1^3 (\pi/x^2)\,dx$. Draw a picture of
(a) a figure in the plane whose area is given by this integral.
(b) a solid of revolution whose volume is given by this integral.

10.3. Determine the volume of a sphere of radius r. (Just work through the steps of Example 12 with r in place of 7.)

10.4. Consider a sphere in comparison with a cylinder in which the sphere is inscribed. Using your answer to exercise 10.3, find the ratio of the volume of the cylinder to the volume of the sphere.



Answer: The ratio is 3 : 2, as first proved by Archimedes in the third century B.C. (See the appendix on Archimedes.)

§11. Isaac Newton

The Plague, in 1664–1665, had at least one fortunate consequence. Cambridge University was forced to shut down. Newton, having just received his B.A. degree, moved back to the English countryside where he had been born on Christmas Day of 1642. Newton delighted in privacy. The next two years of secluded life by an apple grove produced astonishing results. Newton came into possession of ideas that would enable him to create modern physics virtually by himself, with the help of calculus, which he also created at the same time.

Newton not only had ideas of his own. He could see new features hidden in the ideas of others. Whereas Fermat had developed his method only to find tangent lines to curves, Newton saw in this same method the means of defining a derived function that would measure instantaneous rates of change. Newton used this to study the physics of motion, of which he postulated certain "universal laws". When he put these laws together with his calculus and with the law of gravitation (also discovered on the farm), Newton derived the equations governing the motion of the planets about the sun.

Johannes Kepler (1571–1630) had earlier made the significant discovery that the planets travel in elliptical orbits, but Kepler could not explain why. Newton knew how to explain that this was no more mysterious than the fall of an apple from the tree. When Newton overcame his secretive nature and finally revealed in 1687 the magnitude of his work, the effect was overwhelming. Newton, it was said, had "explained the universe". That was, of course, an overstatement, whose repetition finally prompted the playful couplet of Alexander Pope,

> Nature and Nature's laws lay hid in night;
> God said, "Let Newton be!" and all was light.

Nevertheless, it is generally conceded that Newton's *Mathematical Principles of Natural Philosophy* (1687) remains the greatest single work in the history of science. Perhaps never before or since has so much been uncovered at a single stroke. Aided by 20 years of thought, Newton wrote it, start to finish, in 18 months.

An aura of mystery still surrounds the man:

> ... Newton with his prism and silent face,
> The marble index of a mind forever
> Voyaging through strange seas of thought, alone.

So wrote William Wordsworth near the dawn of the nineteenth century, upon marveling at a statue of Newton celebrating his work in optics.

Newton's highest compliment came from his only rival, who published (in 1684) the first paper on the calculus. There would be bitter years of controversy over who first discovered the calculus. But,

> Taking mathematics from the beginning of the world to the time of Newton, what he has done is much the better half.
>
> Leibniz

Although others owned bits and pieces of the calculus, Newton was the first to have the whole subject at his command. He rose like Archimedes above the age in which he lived, moved by a spirit impervious to time.

> I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.
>
> Newton

[1]

PHILOSOPHIÆ NATURALIS Principia MATHEMATICA

Definitiones.

Def. I.

Quantitas Materiæ eſt menſura ejuſdem orta ex illius Denſitate & Magnitudine conjunctim.

AEr duplo denſior in duplo ſpatio quadruplus eſt. Idem intellige de Nive et Pulveribus per compreſſionem vel liquefactionem condenſatis. Et par eſt ratio corporum omnium, quæ per cauſas quaſcunq; diverſimode condenſantur. Medii interea, ſi quod fuerit, interſtitia partium libere pervadentis, hic nullam rationem habeo. Hanc autem quantitatem ſub nomine corporis vel Maſſæ in ſequentibus paſſim intelligo. Innoteſcit ea per corporis cujuſq; pondus. Nam ponderi proportionalem eſſe reperi per experimenta pendulorum accuratiſſime inſtituta, uti poſthac docebitur.

B Def.

Figure 2. The first page of Newton's *Principia* (1687)

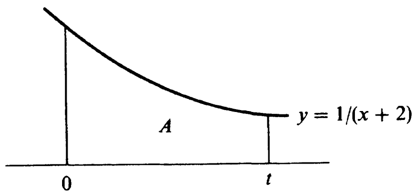
The transport of calculus from the seashore to the stars: that was Newton's accomplishment. A dream of old Pythagoras had been realized at last.

Problem Set for Chapter 7

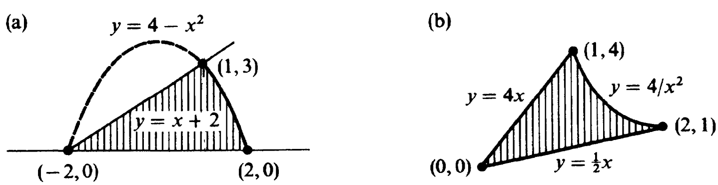
1. A boat travels along a straight course. At time t hours past noon, its speed is $t^2 - 4t + 10$ km/hr. How far does the boat travel between three o'clock and six o'clock?

2. Sketch the quadratic curve $y = t^2 - 4t + 10$, and find the area beneath this curve, between $t = 3$ and $t = 6$.

3. If A is the area indicated in the figure below, find dA/dt. (Your answer should be expressed in terms of t, of course.)



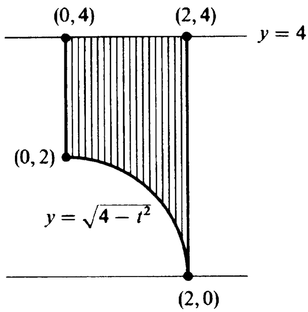
4. Find the indicated area. (In each case, split the area into two pieces by drawing an appropriate vertical line, find the area of each piece separately, and add.)



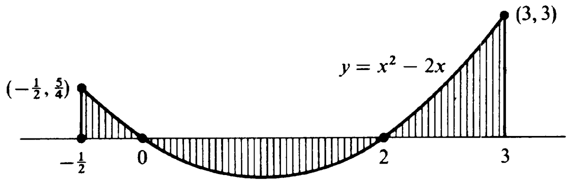
5. Draw a picture of the area represented by the integral $\int_0^2 4\,dt$, and evaluate the integral by finding the area of your picture.

6. Evaluate the integral $\int_0^2 \sqrt{4 - t^2}\,dt$, after first drawing a picture of the area it represents.

7. Find the indicated area by using your answers to problems 5 and 6. Why is the general area principle of no use to you here?



8. Find the indicated area by splitting it into three parts, finding the area of each part, and adding.



9. Calculate the following integrals directly from their definition by Eudoxus' method.
   (a) $\int_0^2 4x\,dx$.
   (b) $\int_1^4 (2x - 1)\,dx$.
   (c) $\int_a^b x\,dx$.

10. Consider the integral $\int_{-1/2}^{3} (x^2 - 2x)\,dx$.
    (a) Illustrate Eudoxus' method in calculating this integral.
    (b) Illustrate the fundamental theorem in calculating this integral.
    (c) Explain why it is to be expected that your answers to parts (a) and (b) agree with each other, yet disagree with your answer to problem 8.

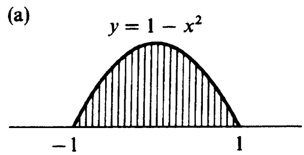
11. (a) By using the appropriate rules for derivatives, and simplifying your answer, show that

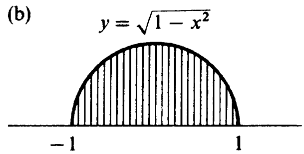
$$\text{if } F(x) = \frac{3x^2}{2\sqrt{1 + x^3}}, \qquad \text{then } F'(x) = \frac{3x(4 + x^3)}{4(1 + x^3)^{3/2}}.$$

    (b) Evaluate the integral

$$\int_0^2 \frac{3x(4 + x^3)}{4(1 + x^3)^{3/2}}\,dx.$$

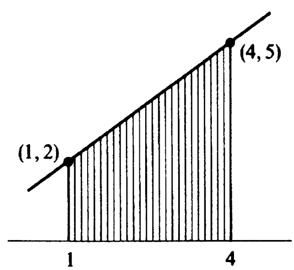
12. Find the indicated areas, by any means.



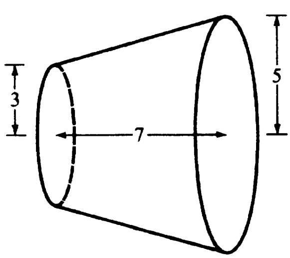


13. Find the volumes of the solids of revolution obtained by revolving the areas of problem 12 about the x-axis.

14. (a) Find an equation of the line joining (1,2) and (4,5).
    (b) Find the volume of the frustum of a cone obtained by revolving the indicated area about the x-axis.

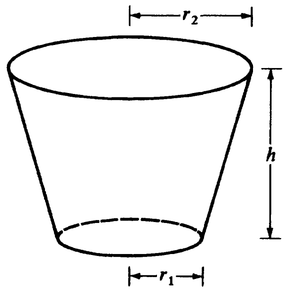


15. Find the volume of the indicated flower pot, shaped like the frustum of a cone.

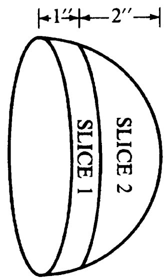


Hint. The volume is that obtained from revolving the area beneath the line joining (0, 3) and (7, 5).

16. Find a formula (in terms of r_1, r_2, and h) for the volume of a frustum of a cone with the indicated dimensions.



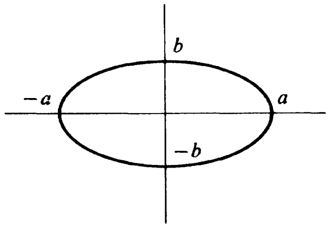
17. A grapefruit half, shaped like a hemisphere of radius 3 inches, is sliced in two, as indicated.



   (a) Find the volume of each slice.
   (b) Which slice has greater volume?
   (c) Where should the slice be made in order to divide the grapefruit into pieces of equal volume? (Your answer may be expressed as the solution to a certain cubic. Solve the cubic by Newton's method.)

18. Consider the integral $\int_0^5 \pi x^2\, dx$.
   (a) Draw a picture of a figure whose area is given by this integral.
   (b) Draw a picture of a solid of revolution whose volume is given by this integral.

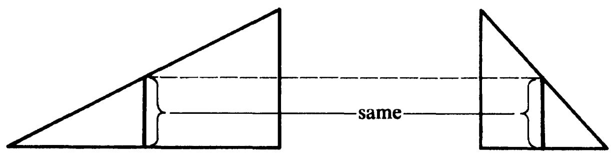
19. The equation $(x^2/a^2) + (y^2/b^2) = 1$ has an *ellipse* as its graph. Let A be the area inside the ellipse. Justify each of the following equalities in the calculation of A.



$$A = 2\int_{-a}^{a} \sqrt{b^2 - \frac{b^2x^2}{a^2}}dx$$
$$= 2\frac{b}{a}\int_{-a}^{a} \sqrt{a^2 - x^2}\, dx = 2\frac{b}{a}\left(\frac{1}{2}\pi a^2\right) = \pi ab.$$

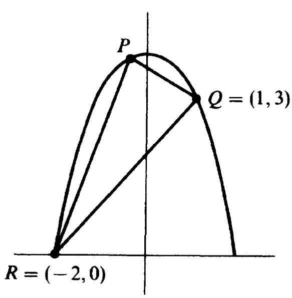
20. If the ellipse of problem 19 is revolved about the x-axis, an *ellipsoid of revolution* (watermelon) results. Find the volume inside it.

21. The two right triangles below might be regarded as being made up of "identical vertical segments", if their segments are made to correspond as indicated. But it is clear that the two triangles do not have the same area. Does this violate Cavalieri's principle? Why not?

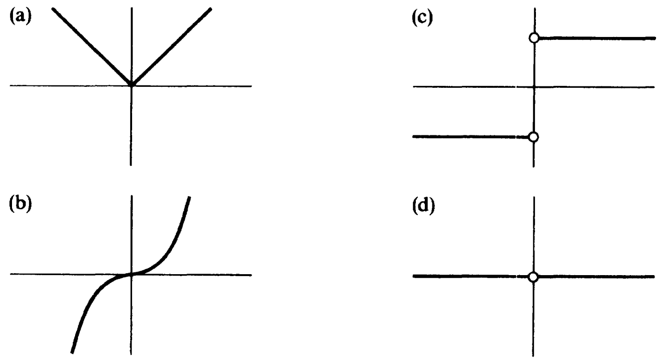


22. In the problem set at the end of Chapter 3, work problem 12 by applying Cavalieri's principle. *Hint.* Go to problem 12, Chapter 3. Turn your head sideways as far as you can. Don't strain your neck.

23. Find the coordinates of the point P lying on the curve $y = 4 - x^2$ that maximizes the area of the indicated triangle PQR. (P must lie between R and Q.)



24. Match each of the curves below with its derivative.



[The curve in (d) coincides with the horizontal axis, but has a hole in it.]

25. Sometimes we can evaluate an integral most easily if we notice that its value measures the size of an area. Draw a picture of the areas represented by each of the following integrals, then evaluate the integrals by splitting up each of the areas into two pieces—a right triangle and a "slice of pie". Then add together the areas of these two pieces, each of which is easily found using basic geometric formulas. Why is the fundamental theorem of calculus of no use to you here?
    (a) $\int_0^{\sqrt{2}/2} \sqrt{1-x^2}\,dx$. (b) $\int_0^{\sqrt{3}/2} \sqrt{1-x^2}\,dx$.

26. (*Proving the fundamental principle of integral calculus*) This principle states that if $F'(t) = 0$ for all t in a connected domain, then F is a constant function, i.e., F takes the same value at any two points a and b. To prove $F(a) = F(b)$, we need only show that $F(b) - F(a) = 0$. Do this by justifying each of the following steps.
    (a) For any two points in a connected domain, explain why the domain contains all points lying between them. *Hint.* What does *connected* mean? (See Chapter 6, Section 8.)
    (b) If $F'(t) = 0$ for each t in a connected domain and if a and b lie in this domain with $a < b$, then $\int_a^b F'(t)\,dt = 0$. *Hint.* By part (a), $F'(t) = 0$ if $a \le t \le b$.
    (c) $\int_a^b F'(t)\,dt = F(b) - F(a)$. *Hint.* Use the fundamental theorem of calculus.
    (d) $F(b) - F(a) = 0$. *Hint.* Put together the results of parts (b) and (c).

27. (*Antiderivatives and inequalities*) Given two functions f and g with $f(x) \le g(x)$ on the domain $a \le x \le b$, it follows, of course, that $f(x_k) \le g(x_k)$ for any point x_k in this domain.
    (a) Explain why, in Eudoxus' method, we then have $\sum_{k=1}^{n} f(x_k)\,\Delta x \le \sum_{k=1}^{n} g(x_k)\,\Delta x$.
    (b) Prove that $\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx$ if each of these integrals exists. *Hint.* Each integral is the limit of its appropriate sum in Eudoxus' method. Use the result of part (a).

(c) Show that the inequality of part (b) implies that $F(b) - F(a) \leq G(b) - G(a)$, where F and G are antiderivatives of f and g, respectively. *Hint.* Use the fundamental theorem of calculus.

(d) In Section 10 of Chapter 6 we stated without proof a theorem on antiderivatives and inequalities. Can you prove this theorem now? *Hint.* In part (c) let the roles of a and b be played by 0 and t.

28. If $f(t) = t^n$, an antiderivative is generally given by $F(t) = t^{n+1}/(n+1)$. The reciprocal function given by $f(t) = t^{-1} = 1/t$ does not come under this rule. (*Why not? What goes wrong in this rule when $n = -1$?*) Since we have no antiderivative of the reciprocal function we must calculate integrals of it by Eudoxus' method. In 1647 Gregory of St. Vincent made a remarkable observation about certain integrals involving this function. Gregory's observation was really about the areas represented by the integrals several decades before integrals (or antiderivatives) were invented.

(a) Show that the approximating sum S_n of the integral $\int_1^{\pi} \frac{1}{t} dt$ is given by $S_n = \sum_{k=1}^{n} (\pi - 1)/(n + k\pi - k)$. (Here we have $\Delta t = (\pi - 1)/n$, $t_k = 1 + k(\pi - 1)/n$, etc.)

(b) Show that the approximating sum S_n of the integral $\int_4^{4\pi} \frac{1}{t}\, dt$ is also given by $S_n = \sum_{k=1}^{n} (\pi - 1)/(n + k\pi - k)$. (Now we have $\Delta t = (4\pi - 4)/n$, $t_k = 4 + k(4\pi - 4)/n$, etc.)

(c) Gregory could now conclude that $\int_1^{\pi} \frac{1}{t} dt = \int_4^{4\pi} \frac{1}{t} dt$ because the result of parts (a) and (b) shows that both these integrals are limits of identical S_n's. Draw a picture of the areas represented by these two integrals and see whether you can understand geometrically why this is true. (*Can you see that the first area can be transformed into the second if it is scaled down vertically by a factor of $1/4$ and then scaled up horizontally by a factor of 4?*)

29. (*Looking backwards and forwards with logarithms*) It is obvious from the remarks at the end of Section 7 that

$$\int_1^{4\pi} \frac{1}{t} dt = \int_1^{4} \frac{1}{t} dt + \int_4^{4\pi} \frac{1}{t} dt.$$

(a) Use this equality together with the result of part (c) of problem 28 to show that $\int_1^{4\pi} \frac{1}{t} dt = \int_1^{4} \frac{1}{t} dt + \int_1^{\pi} \frac{1}{t} dt$, and then note that there is nothing special here about 4 and π. Show, as Gregory of St. Vincent did, that for any numbers a and b exceeding 1, it is true that

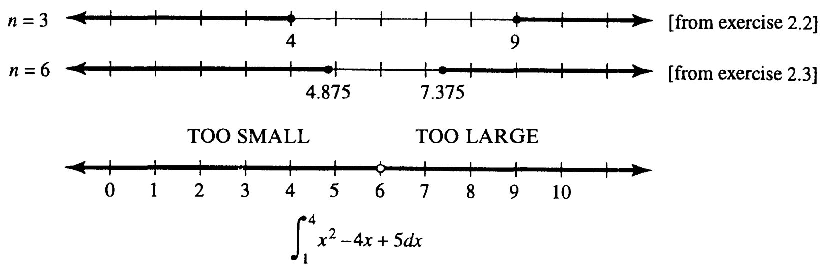
$$\int_1^{ab} \frac{1}{t} dt = \int_1^{a} \frac{1}{t} dt + \int_1^{b} \frac{1}{t} dt.$$

(b) Explain how the result of part (a) is the same result—just expressed in different language—as that obtained in problems 21 and 22 at the end of Chapter 6.

(c) Finally, explain the statement made at the end of Section 4 of Chapter 3 about the connection between hyperbolas and the numerical calculation of logarithms.

Gregory's approach to logarithms superseded the work of Napier (1550–1617) and Briggs (1561–1631), who calculated the first accurate logarithmic tables. Napier took quite a different "inverse" approach that we shall not discuss, except to say that it is related to the idea of changing arithmetic movement to geometric movement—an idea we met briefly in problem 14 of Chapter 2.

30. (*The Riemann integral*) While we feel very much at home with the number denoted by $\sqrt{2}$ and with the number denoted by π, we may still be less comfortable with the number denoted by $\int_a^b f(x)\,dx$. Thanks to G.F.B. Riemann (1826–1866), we now know that the numbers behind all three of these symbols are equally easy to comprehend.
    (a) Study the figure from Section 6 of Chapter 2 entitled "searching for wrong ratios"; then study the analogous figure from Section 3 of Chapter 3.
    (b) Let's take a particular integral, say $\int_1^4 x^2 - 4x + 5\,dx$, in hopes of seeing how to handle all integrals in the same way. The following figure should be self-explanatory:



We need to have a name for the numbers that are "too big", such as the number 9 calculated in exercise 2.2 after dividing the domain into $n = 3$ pieces, or the number 7.375 calculated in exercise 2.3 by dividing the domain into $n = 6$ pieces. They are called *upper sums* for this integral. Similarly, the numbers 4 and 4.875 are *lower sums* corresponding to partitions of the domain when $n = 3$ and $n = 6$, respectively. For each positive integer n (even if n is as large as the national debt), we get numbers that are too big and numbers that are too small by taking upper sums and lower sums, respectively. Complete the following sentence the way you think Riemann might have: "The integral $\int_a^b f(x)\,dx$ can be simply defined as the number, if there is only one such number, that lies between every ____ and every ____."

CHAPTER 8 Romance in Reason

Having reached the fundamental theorem of calculus, we have come to a natural place to pause and take stock of our accomplishments and aspirations. Let us close this volume by discussing the way calculus was viewed in the seventeenth century, looking both backward and forward in time to put this remarkable century's mathematical thought in true perspective—and perhaps to learn something about the true nature of mathematics itself.

How can we speak of such inscrutable things as these? Learning to tell the truth may have been the original purpose of a liberal arts education in ancient Greece, but times have changed, and truth today no longer seems to be that bright sun to which Plato said we are all so naturally drawn. The twentieth-century mathematician John von Neumann remarked that truth is much too complicated to allow anything but approximations. The best we can do, most of the time, is to glimpse truth as a kind of limit that we cannot attain, but only strive toward.

In this spirit we may say that the emphasis placed in Chapter 2 upon the concurrence of the rise of mathematics and of rational thought reveals much about the nature of mathematics, but gives us only an approximation to the whole truth. Just as number has both rational and irrational elements, so does mathematics itself. What we shall see in this chapter is that, although the seventeenth century is justly called the Age of Reason because of its progress in science, much of its "reasoning" in mathematics is more characteristic of the later Romantic Movement, as manifested in its opposition to the glorification of rational thought. This chapter, then, is concerned with "romantic" elements in supposedly "rational" mathematics.

So far as calculus itself is concerned, incidentally, its "age of reason" is really the nineteenth century. That is when analysis had to be developed to make rational sense out of the paradoxes and general confusion that arose out of the huge mass of seventeenth- and eighteenth-century mathematical research largely due to—but overly dependent upon—an unbridled trust in mathematical intuition. The reader can get a hint of what analysis is like by re-reading the end of Section 6 of Chapter 2, but the full story is best left to another volume.

Let us review very briefly the historical movements leading up to the Age of Reason. We have already mentioned at the end of Chapter 3 how the torch of mathematics was taken up by diverse peoples—among them the Chinese, Indians, and Arabs—who preserved or rediscovered much of Greek mathematics and developed much that was new. The discovery by western Europe of many of these works helped rekindle the flame of mathematics in the Renaissance. Neoclassicism, a movement to revive or to adapt the classical style, arose in mathematics just as it arose in European literature, art, and music.

As every student of history knows, all this led in time to the Age of Reason, the Enlightenment, and eventually to the Romantic Movement. And every student of the liberal arts will know something of the way these historical movements are reflected in literature, art, and music.

Let us not leave mathematics out of the liberal arts. How does mathematics enter into this scheme of things? It is obvious that the development of calculus helped bring about the rise of modern science, to which the Enlightenment pointed with such pride. However, let us not be content with such an obvious remark. While we may not learn "the true position of mathematics as an element in the history of thought", we may yet learn something by musing about the nature of mathematics.

The key word of the discussion in this chapter is *tension*. It has been contended that the life in any work of art derives from the creation and resolution of tension, where "tension" is understood in a rather broad sense. Certainly the vitality of mathematics springs from a kind of tension. Mathematics itself, being in residence between the humanities and the sciences, is stretched in many directions: toward beauty, form, and vision on the one hand; and toward utility, function, and rationality on the other. And these are only a few of the struggles taking place within mathematics:

> Mathematics as an expression of the human mind reflects the active will, the contemplative reason, and the desire for aesthetic perfection. Its basic elements are logic and intuition, analysis and construction, generality and individuality. Though different traditions may emphasize different aspects, it is only the interplay of these antithetic forces and the struggle for their synthesis that constitute the life, usefulness, and supreme value of mathematical science.
>
> R. Courant and H. Robbins*

* *What is Mathematics?* Oxford University Press, New York, 1941, p. xv.

§1. Guessing versus Reasoning

Mathematics has always been associated with reason, or rational thought. The word *rational* still carries the connotation of "measurement" or "calculation". A rational man measures, or calculates, the effect of his activity.

The mathematician, in the act of making a discovery, is hardly "rational", however. The view that mathematicians employ only a cold, unexcited, strictly logical approach to their calling is somewhat distorted. Archimedes, struggling with a perplexing problem, once had such an exciting idea that he jumped from his bath to run naked and screaming down the streets of Syracuse. And Newton's greatness as a mathematician seems not to have been due primarily to his ability to reason correctly:

> I fancy his pre-eminence is due to his muscles of intuition being the strongest and most enduring with which a man has ever been gifted. Anyone who has ever attempted pure scientific or philosophical thought knows how one can hold a problem momentarily in one's mind and apply all one's powers of concentration to piercing through it, and how it will dissolve and escape and you will find that what you are surveying is a blank. I believe that Newton could hold a problem in his mind for hours and days and weeks until it surrendered to him its secret. Then being a supreme mathematical technician he could dress it up, how you will, for purposes of exposition, but it was his intuition which was pre-eminently extraordinary—'so happy in his conjectures', said de Morgan, 'as to seem to know more than he could possibly have any means of proving'....
>
> John Maynard Keynes*

Conjectures, or guesses, play a largely unrecognized role in mathematics. We have seen in this book some instances of how they work. Early in Chapter 4 we guessed that the slope of a certain tangent line was 2, yet it took a while to find a reason why. Early in Chapter 7 we guessed that a certain area was 6 square units, but reasoned justification for that guess could come only much later. "Humble thyself, impotent reason!" exhorted Pascal, who almost discovered the calculus himself, before Newton. While reason may demonstrate the truth of a guess, reason alone rarely discovers anything of significance.

> Mathematics is regarded as a demonstrative science. Yet this is only one of its aspects. Finished mathematics presented in a finished form appears as purely demonstrative, consisting of proofs only. Yet mathematics in the making resembles any other human knowledge in the making. You have to guess a mathematical theorem before you prove it;

* "Newton, the Man" from *Essays in Biography*, Horizon Press Inc., New York, 1951, p. 312. (This essay appears also in *Newton Tercentenary Celebrations*, Cambridge University Press, 1947.)

> you have to guess the idea of the proof before you carry through the details. You have to combine observations and follow analogies; you have to try and try again. The result of a mathematician's creative work is demonstrative reasoning, a proof; but the proof is discovered by plausible reasoning, by guessing.
>
> G. Pólya*

Let there be no doubt of the existence, in mathematics, of knowledge acquired in nonrational ways:

> I have had my solutions for a long time, but I do not yet know how I am to arrive at them.
>
> Gauss

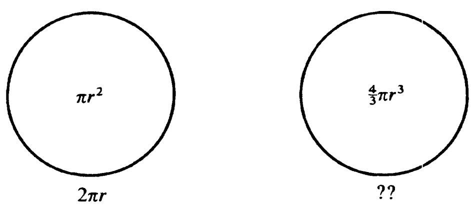
> It's plain to me by the fountain I draw from, though I will not undertake to prove it to others.
>
> Newton

> Certain things first became clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards, because investigation by the said method did not furnish an actual demonstration.
>
> Archimedes

Gauss†, Newton, and Archimedes stand in a class above all other mathematicians. We thus have it on the highest authority that imagination plays a role in mathematics at least rivaling, and perhaps surpassing, the role of reason. Great mathematicians have both gifts, in great degree.

Exercises

1.1. Make a guess as to the formula for the surface area S of a sphere of radius r. Reason by analogy with a circle, which is a "sphere" in the plane: for a circle, $A = \pi r^2$ and $C = 2\pi r$; for a sphere, $V = \frac{4}{3}\pi r^3$ and $S = ?$



* *Induction and Analogy in Mathematics*, Princeton University Press, 1954, p. vi.
† Carl Friedrich Gauss (1777–1855), preeminent German mathematician.

1.2. Give the approximate dates and general characteristics of the Renaissance, the Age of Reason, and the Romantic Movement. For help, consult an encyclopedia or a history of Western civilization.

1.3. Leaf through Volume 1 of *Mathematics and Plausible Reasoning*, G. Pólya, Princeton University Press, Princeton, N.J., 1954, for a fuller understanding of the art of guessing.

§2. Atomism versus Common Sense

How is nonrational, or intuitive, knowledge possible? What sorts of tricks were used by the developers of the calculus to tell them which way to go? Many tricks stem from the "atoms of Democritus". The Greek Democritus (ca. 460–370 B.C.) supported the doctrine of *atomism*, which holds that bodies are made up of *atoms*, or indivisible units. An atomist would raise no objection to thinking of a line as the sum of its points, or of an area as the sum of its vertical line segments. Atomism regards time as being made up of instants, an instant being a "point" in time.

Like most philosophical doctrines, atomism has its drawbacks. Common sense seems to tell us that time, like a pencil point moving smoothly along a line, is a "flowing", or "continuous", kind of thing. How can a *continuous* entity like time be made up of *discrete* instants? How can an atomist answer the Arrow Paradox of Zeno (ca. 495–435 B.C.)?

> Consider an arrow flying through the air. At each instant the arrow is motionless. How can the arrow move if it is motionless at each instant?

The same general sort of "paradox" is not uncommon in mathematics:

> How can a line segment have nonzero length, if each of its points has length zero?
>
> How can a planar figure have nonzero area, if each of its vertical line segments has area zero?

The inadequacy of atomism is evident. The atoms of a body apparently need not reflect all the properties of that body: whereas a line has length, its points do not. The whole may be something more than the sum of its atoms.

What good, then, is atomism? In mathematics it is often an aid to the intuition. Democritus used it to make an inspired guess about the proper formula for the volume of a cone or pyramid (one-third the area of the base times the height, in either case). Democritus had the imagination to guess the correct answer, but was never able to offer any rational justification for that answer. It was Eudoxus whose method provided the demonstrative proof. Both deserve credit: Democritus as seer, and Eudoxus as sage.

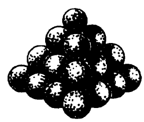
> ... in the case of the theorems the proof of which Eudoxus was the first to discover, namely that the cone is a third part of the cylinder, and the pyramid of the prism, having the same base and equal height, we should give no small share of the credit to Democritus who was the first to make the assertion ... but did not prove it.
>
> Archimedes

The passage above, as well as the quotation from Archimedes given earlier, is taken from a letter addressed to Eratosthenes. Archimedes goes on in this letter to describe *discovery* and *proof* as complementary aspects of mathematics. He then describes how he used atomism, in a novel way, to conjecture the truth of some of his most celebrated theorems, which he proved later by a masterful use of Eudoxus' method. The means of discovery and the means of proof were completely different.

Exercises

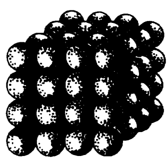
2.1. Archimedes' letter to Eratosthenes is discussed briefly in an appendix to this book. Read the appendix on Archimedes.

2.2. (*The purpose of this exercise is to give a clue as to how Democritus might have used atomism to guess that the volume of a pyramid is one-third the area of the base times the height.*) Consider a bunch of cannonballs (to be thought of as large atoms).



(a) Find the number of cannonballs in a pyramid if the base is
  (i) two by two.
  (ii) three by three.
  (iii) four by four. *Answer*: 30.
  (iv) n by n. *Hint*. See appendix on sums.

(b) Find the number of cannonballs in a *cube* if the base is
  (i) two by two.
  (ii) three by three.
  (iii) four by four. *Answer*: 64.
  (iv) n by n.



(c) Find the ratio of the number of cannonballs in a pyramid with square base to the number of cannonballs in a cube with the same base, if the base is n by n. *Answer*: $n(n+1)(2n+1)/6n^3$.
(d) The ratio in part (c) has 1/3 as a limit. How might this have helped Democritus?

§3. Seer versus Sage

It has never been a secret that the pursuit of mathematics requires more than the power of deductive reasoning. Even a rationalist allows this possibility.

> There are only two ways open to man for attaining certain knowledge of truth: clear intuition and necessary deduction.
>
> Descartes

The *seer* who discovers is just as much a mathematician as the *sage* who proves. Some mathematicians, like Archimedes, are coequally seer and sage, and no mathematician is wholly one or the other. Nevertheless the distinction is useful. The seer has the gift of vision—intuition, divination, or imaginative insight. Whereas the sage is blessed with wisdom—sound judgment, good taste, and reason. The seer points his hand to the sky while his eye darts around the heavens as if to see everything at a single instant. The sage, however, plants his feet squarely on the ground, his gaze fixed upon his object, and marks his world with a steady eye.

The distinction between seer and sage is of interest when one examines the philosophies of Plato and Aristotle, insofar as they pertain to mathematics. Platonism has often been seen to animate speculation, the searching and re-searching for undiscovered truths lying just beyond our ken. Seers are often disciples of Plato.

If Plato animates speculation, says Whitehead, then Aristotle animates scholarship. Aristotle emphasized the consolidation of knowledge, through reason, into a coherent system. Aristotle's influence may be seen in the form in which Euclid's *Elements* was cast, even though the content of the *Elements* owes its existence to the spirit of speculation. Aristotle's influence was great, and Greek mathematics appears almost always in finished form, cold, unexcited, and with strict logic, as if it might have been written by a sage alone.

The fact that classical Greek texts presented only proofs became a source of some annoyance later. One might think that the Greeks, in a wondrous plot, had all agreed to conspire against the seventeenth century by refusing to divulge their means of discovery.

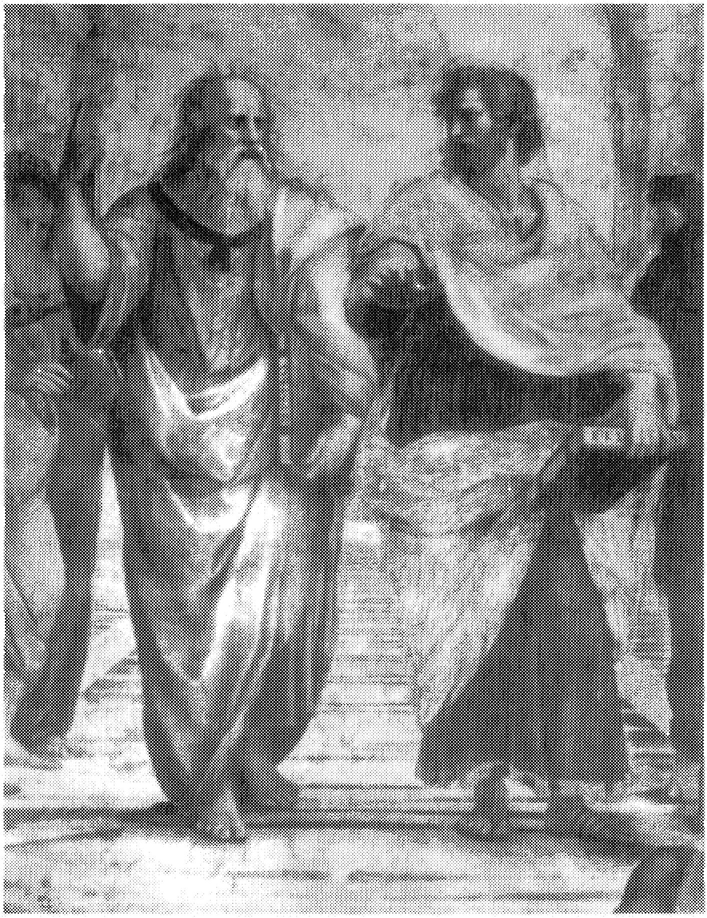


Figure 3.* Plato and Aristotle in Raphael's "School of Athens". *Even now there is a very wavering grasp of the true position of mathematics as an element in the history of thought.* —Whitehead.

* Reproduced, with permission, from *Raphael*, by Oskar Fischel, translated by Bernard Rackham, Routledge & Kegan Paul Ltd., London, 1948.

> . . . like some artisans who conceal their secret, they feared, perhaps, that the ease and simplicity of their [hidden] method, if become popular, would diminish its importance, and they preferred to make themselves admired by leaving to us, as the product of their art, certain barren truths deduced with subtlety, rather than to teach us that art itself, the knowledge of which would end our admiration.
>
> Descartes

Descartes could not have known that the "art itself", the seer's vision, had been freely given by Archimedes in his letter* to Eratosthenes mentioned earlier. And, having revealed his secret method (discussed in an appendix to this book), Archimedes wrote,

> I am persuaded that this method will be of no little service to mathematics. For I foresee that this method, once understood, will be used to discover other theorems which have not yet occurred to me, by other mathematicians, now living or yet unborn.

Unfortunately for Descartes, the contents of Archimedes' letter had been lost for centuries and were found again only in 1906, in Turkey, by the Danish philologist J.L. Heiberg.

The preference expressed by Descartes for the seer as opposed to the sage was typical of seventeenth- (and eighteenth-)century thought in mathematics. The influence of Aristotle was at an ebb, and Platonism was once again ascendent. It is curious that, in the Age of Reason, the climate was such as to permit a lapse of rigor in the reasoning used by mathematicians. The "idea itself", the means of discovery, became more important than the rigorous logical demonstration. The happy acceptance of vague, but intuitively suggestive remarks as a valid proof was not unusual. A reaction set in against the "over-precise" manner of the Greeks, which could only impede the progress of seventeenth-century mathematics.

An illustration of this is seen in Newton's *Principia*, written in 1687. In composing this greatest of works Newton attempted to emulate the rigorous Archimedean style; but, by doing so, he only made the *Principia* more difficult for modern minds to comprehend:

> The ponderous instrument of synthesis (Archimedism), so effective in his hands, has never since been grasped by one who could use it for such purposes; and we gaze at it with admiring curiosity, as on some gigantic implement of war, which stands idle among the memorials of ancient days, and makes us wonder what manner of man he was who could wield as a weapon what we can hardly lift as a burden.
>
> William Whewell

* See *The Method of Archimedes*, a Supplement to *The Works of Archimedes*, edited by T.L. Heath, Cambridge University Press, 1912 (also available in paperback by Dover Publications).

The passage above is from a nineteenth-century book on the history of science. In the seventeenth century, one might well have heard of Newton's *Principia* what King James had earlier said of Francis Bacon's *Novum Organum*, that "it was like the peace of God, which passeth all understanding."

Exercises

3.1. "If logic is the hygiene of the mathematician, it is not his source of food." The twentieth-century mathematician André Weil said this. What does Weil mean?

3.2. On the whole, has the spirit of Plato or of Aristotle been more conducive to the progress of science?

3.3. Find Newton's *Principia* in a library. The full title in English is *Mathematical Principles of Natural Philosophy*.
  (a) What did Newton mean by "natural philosophy"?
  (b) Why did Newton write in Latin?
  (c) Newton's book is generally acknowledged as the greatest single work ever written on science. Why?
  (d) Why is Newton's book so little read today?

3.4. (*For more ambitious students*) "The world will again sink into the boredom of a drab detail of rational thought, unless we retain in the sky some reflection of light from the sun of Hellenism." Read Chapter 7, "Laws of Nature", from *Adventures of Ideas*, A.N. Whitehead, Macmillan, New York, 1933, then tell what is meant by Whitehead's warning.

§4. The Discrete versus the Continuous

The attraction of atomism is probably the emphasis it places upon the *discrete*, a notion that seems quite transparent to the intuition. Certainly the discrete is easier to comprehend than the continuous. It is easier to think about a stationary pebble than about flowing sand. It is easier to think about a stationary instant than about flowing time.

The temptation is great to attempt to explain the continuous in terms of the discrete. Newton tried to explain light this way. Although common sense tells us that light is a "continuous" phenomenon, Newton spoke of "particles of light", as if a light ray was made up of a huge number of discrete units. Newton's description was not really taken seriously until the development of quantum theory in the twentieth century.

Within mathematics itself, the tension between the discrete and the continuous is profound.

> The whole of mathematical history may be interpreted as a battle for supremacy between these two concepts. This conflict may be but an echo of the older strife so prominent in early Greek philosophy, the struggle of the One to subdue the Many. But the image of a battle is not wholly appropriate, in mathematics at least, as the continuous and the discrete have frequently helped one another to progress.
>
> E.T. Bell*

The development of calculus given in this book has been based upon the notion of *limit*, which is intimately related to the idea of *continuity*. Thus we have placed much more emphasis upon the continuous than upon the discrete. It must now be confessed that the seventeenth century attempted a discrete approach to the calculus as well, the description of which makes up a remarkable chapter in the history of ideas.

An attempt will now be made to describe this discrete approach. If the reader finds this approach slightly incomprehensible, there is a good reason for it. The description is kept at a very intuitive level in order to ignore any difficulties that might be seen by close logical scrutiny. This is the way some things were done in the mathematics of the seventeenth century, and much of the great progress made then is undoubtedly due to this approach. Had there not been a lapse in emphasis upon logical rigor, Newton and Leibniz might have feared to put some of their speculations into print. It is helpful to remember that the tone of seventeenth-century mathematics contrasts greatly with the classical Greek. Seventeenth-century mathematics often reads as if written by seer alone.

With our eyes "in a fine frenzy rolling", let us seek the seer's vision. Consider the following question:

> What does the difference Δx become, as Δx tends to zero, but is never allowed to equal zero?

The answer given by Leibniz might run something like the following:

> The difference Δx becomes a quantity of infinitesimal size, to be denoted by "dx" and called the *differential of* x. To say that dx is an infinitesimal is to say that dx is not zero, but is smaller than any positive number.

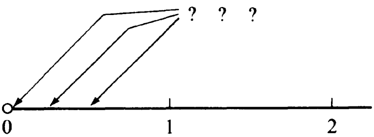
The differential of y, where y is a function of x, is "defined" in a similar way: the differential dy is what the difference Δy becomes, as Δx tends to zero, but is not allowed to equal zero. Leibniz thought of the derivative as an actual quotient of the differentials dy and dx. Thinking this way leads one to discover the chain rule. Newton thought in an intuitive way along much the same lines, but his terminology was different. He spoke of *fluents, fluxions*, and their *moments*, because he thought of a variable as a flowing quantity.

In reply to the question given above, many of us would say that the

* E.T. Bell, *The Development of Mathematics*, McGraw-Hill, 1945, p. 13.

question is simply ill-posed and has no answer. The difference Δx does not "become" anything, because it is never zero. It just "keeps on changing". This would be the reply of a sage. Calculus was not discovered by a sage.

It is easy to criticize the notion of an infinitesimal, if it is regarded as a fixed quantity somehow squeezed between zero and the positive numbers. Where could it be on the number line? There is no place for it:



If there is such a thing as an infinitesimal, it must be a new kind of quantity, for it cannot be pictured as a point on the number line. The notion of the *infinitesimal* is one of the most elusive ideas ever conceived. Attempts to describe it, as with the adjectives *nascent* and *evanescent*, bordered upon the the comic. The first adjective means "just born", the second means "just vanishing".

However, we should not laugh at this seventeenth-century version of atomism called *infinitesimal analysis*. Though it all seems so vague, it was really a noble attempt to reconcile, through a rather mystical notion, the two great cooperating opposites of mathematics. An infinitesimal was supposed to be a discrete entity that retained qualities of the continuous.

Exercises

4.1. Prove that there is no positive number lying "next" to zero. That is, show that between any positive number and zero lies another number. *Hint.* Make a *reductio ad absurdum* argument, using the fact that halving a nonzero quantity always results in a new quantity that is closer to zero.

4.2. Criticize the following statement. "The tangent line to a curve at a point is the line through the point and the next point on the curve."

4.3. (*For more ambitious students*) Read in a philosophy book about Leibniz's theory of monads. Write a paper explaining this theory, and explaining how it may be related to Leibniz's theory of differentials.

§5. The Infinitesimal Calculus

Let us continue in the spirit of the preceding section, agreeing to pretend that we know what an infinitesimal is. Let us also agree to accept the romantic notion that logic is unimportant, that something is true as soon as it is felt. This is the setting for discussing the remarkable theory of the infinitesimal calculus, in the spirit of seventeenth-century mathematical thought.

To start off with a bang, let us consider the fundamental theorem of calculus:

$$\int_a^b f(x)\,dx = F(b) - F(a) \quad \text{if } F' = f.$$

Infinitesimal calculus is better done in Leibniz's notation, which does not name functions, only variables. To put the fundamental theorem in this notation, let $y = F(x)$, so that $dy/dx = F'(x) = f(x)$, $y(b) = F(b)$, and $y(a) = F(a)$.

The Fundamental Theorem of Calculus
$\int_a^b (dy/dx)\,dx = y(b) - y(a)$.

"Proof"!
Canceling the differential dx we have

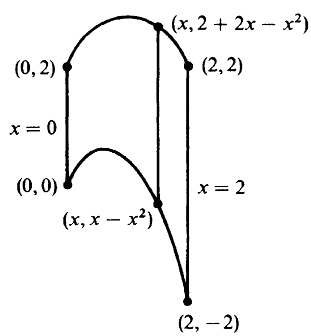
$$\int_a^b \frac{dy}{dx}\,dx = \int_a^b dy. \tag{1}$$

Now $\int_a^b dy$ is simply the sum, from a to b, of all the infinitesimal changes in y! This will obviously add up to the *total* change in the function y, as x runs from a to b:

$$\int_a^b dy = y(b) - y(a).$$

The fundamental theorem is simply the result of putting equations (1) and (2) together! (?)

EXAMPLE 1
Use infinitesimal calculus to find the area between the curves $y = 2 + 2x - x^2$ and $y = x - x^2$, between $x = 0$ and $x = 2$.



Let x be any number between 0 and 2. Consider the vertical segment through x whose width is infinitesimal! The length of this segment is

$$2 + 2x - x^2 - (x - x^2) = 2 + x.$$

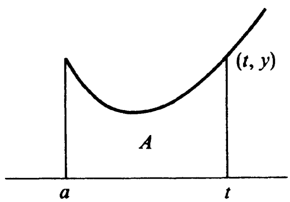
The width of this segment is the infinitesimal dx! Its area is therefore $(2 + x)\,dx$! The entire area between the curves is the sum of these infinitesimal areas $(2 + x)\,dx$, as x runs from 0 to 2! The entire area is then

$$\int_0^2 (2 + x)\,dx = 6 \text{ square units},$$

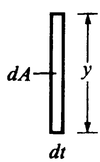
by the fundamental theorem of calculus! (?)

EXAMPLE 2

Use infinitesimal calculus to find dA/dt, where A is the area beneath the curve y, between $x = a$ and $x = t$.



This is easy! As any transcendental eye can see, the infinitesimal change dA in area is clearly given by a rectangle of height y and width dt!

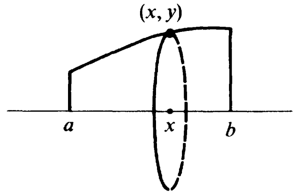


$$dA = y\,dt.$$

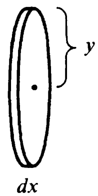
Dividing by dt, we see that $dA/dt = y$. Thus the area beneath a curve y yields an antiderivative of y. (?)

EXAMPLE 3

Use infinitesimal calculus to find the volume of a solid generated by revolving the area beneath a curve about the x-axis.



Let us consider the solid of revolution generated by the curve y from $x = a$ to $x = b$. If x is any number between a and b, consider the volume of the indicated slice through x of infinitesimal width!



Its infinitesimal volume dV is given by

$$dV = \pi y^2\, dx,$$

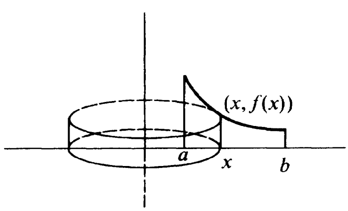
by the well-known formula for the volume of an infinitesimal cylinder! The total volume V of the solid is the sum of all these infinitesimal dV's, as x runs from a to b!

$$V = \int_{x=a}^{x=b} dV = \int_a^b \pi y^2\, dx.$$

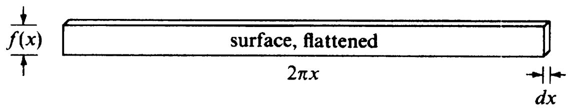
This is the formula for the volume of a solid of revolution! (?)

EXAMPLE 4

Suppose a solid of revolution is generated by revolving the area beneath a curve about the y-axis. Use infinitesimal calculus to find the formula for the solid.



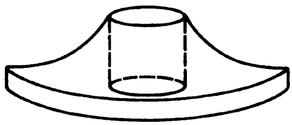
Let x be any number between a and b, and consider what happens to the vertical segment through x, as it is revolved about the y-axis. The surface of a cylinder is obtained, the cylinder being of height $f(x)$ and radius x. Let us find the infinitesimal volume of this surface, whose thickness is infinitesimal! When the surface is flattened out, a rectangular solid is obtained, dimensions $2\pi x$, $f(x)$, and dx.



Its infinitesimal volume dV is then given by

$$dV = 2\pi x f(x)\,dx.$$

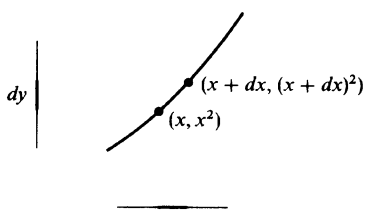
The total volume V is the sum of these infinitesimals dV, as x runs from a to b!



$$V = \int_{x=a}^{x=b} dV = \int_a^b 2\pi x f(x)\,dx$$

is the required formula! (?)

EXAMPLE 5
Use infinitesimal calculus to find the derivative of the squaring function.



The two points virtually coincide

On the curve $y = x^2$ consider the point (x, x^2). If x is changed by an infinitesimal amount dx, then

$$dy = (x + dx)^2 - x^2 = 2x\,dx + (dx)^2.$$

Figure 4.* Engraving from Euler's *Introductio in Analysin Infinitorum* (1748), the first great treatise on analysis. Every textbook on calculus today borrows, more or less, from Euler.

* This illustration serves as frontispiece for Abraham Robinson's historic *Non-standard Analysis*, North-Holland, Amsterdam, 1966. Reproduced by kind permission of Elsevier-North Holland.

Dividing by dx, we obtain

$$\frac{dy}{dx} = 2x + dx. \tag{3}$$

Since dx is infinitesimal, equation (3) becomes

$$\frac{dy}{dx} = 2x. \quad (?) \tag{4}$$

Example 5 leads to an interesting question for anyone who professes to understand the "reasoning" in it. *How can* $2x + dx$ *be equal to* $2x$ *unless* dx *is zero? And if* dx *is zero then how do you justify dividing by it to arrive at equation* (3)? One way to avoid embarrassment is to discard the notion that an infinitesimal is fixed and to think of it instead as a variable tending to zero. In that sense equation (3) does "become" equation (4). This leads to the discarding of fixed infinitesimals in favor of the notion of a *limit*.

Exercises

5.1. Discuss the following quotations of Bertrand Russell.

(a) "It is peculiar fact about the genesis and growth of new disciplines that too much rigour too early imposed stifles the imagination and stultifies invention. A certain freedom from the strictures of sustained formality tends to promote the development of a subject in its early stages, even if this means the risk of a certain amount of error." (*Wisdom of the West*, Rathbone Books Limited, London, 1959, p.280.)

(b) "Instinct, intuition, or insight is what first leads to the beliefs which subsequent reason confirms or confutes ... Reason is a harmonising, controlling force rather than a creative one. Even in the most purely logical realms, it is insight that first arrives at what is new." (*Our Knowledge of the External World*, George Allen & Unwin Ltd., London, 1949, p.22.)

5.2. In the spirit of this section write out a "proof" of each of the following, using infinitesimals. If ever you feel yourself getting into logical difficulties, adopt the visionary's style of reasoning by making exclamations instead of statements.

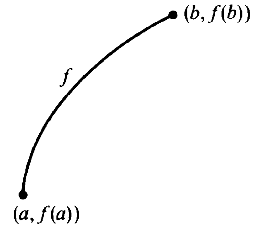
(a) The distance traveled is the integral of the speed function.

(b) $dy/dt = (dy/dx)(dx/dt)$.

(c) $ds/dt = (dt/ds)^{-1}$.

Answer: (c) *What else could it be!*

5.3. Consider a curve f. What is its *length* from the point $(a, f(a))$ to $(b, f(b))$?

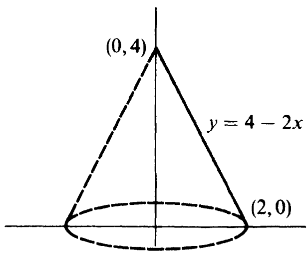


Use infinitesimals to try to guess a formula for the length of a curve. (Try to express it as an integral.) After you have made your guess, check to see if it works when the curve f is a straight line.

5.4. By carrying out the following steps, check to see if the formula of Example 4 works in this simple case considered here.

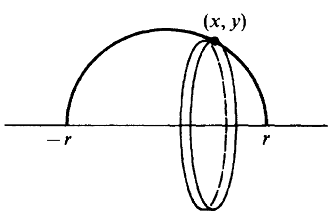
(a) Use the known formula (derived in Chapter 7) for the volume of a cone to find the volume of a cone of radius 2, height 4.

(b) Consider the cone generated by revolving the area beneath the line $y = 4 - 2x$, $0 \leq x \leq 2$, about the y-axis.



Apply the formula of Example 4 to this situation. Does it give the proper answer as found in part (a)?

5.5. (*How do you measure surface area?*) In exercise 1.1 you were asked to make a guess as to the formula for the surface area S of a sphere of radius r. Consider the following way one might reason by the use of infinitesimals. The sphere of radius r is obtained by revolving the graph of $y = \sqrt{r^2 - x^2}$, $r \leq x \leq r$, about the x-axis.



When the point (x, y) is revolved we get the surface of an infinitesimal cylinder of radius y and height dx! This surface, when flattened out into a rectangle, is seen to have an area of $2\pi y\,dx$. Therefore (!), the total surface area is given by

$$S = \int_{-r}^{r} 2\pi y\,dx = 2\pi \int_{-r}^{r} y\,dx = 2\pi \begin{pmatrix} \text{the area beneath the curve } y \\ \text{from } -r \text{ to } r \end{pmatrix}.$$

Since the area beneath the semicircle is $\pi r^2/2$, it follows that $S = 2\pi(\pi r^2/2) = \pi^2 r^2$. (?)

(a) This probably disagrees with the guess you made in exercise 1.1. Do you still have more confidence in your guess, or do you accept the infinitesimal analysis just given? *Why*?

(b) Are you willing to accept all the infinitesimal analysis given in this section? If not, how do you decide what to accept and what to reject as invalid?

Hint. Archimedes showed that the surface area of a sphere of radius r is given by $S = 4\pi r^2$. (See Section 2 of the appendix on Archimedes.)

5.6. Archimedes was the first to find the correct numerical value of the ratio r_7 introduced in exercise 2.6 of Chapter 2. Use his formula $S = 4\pi r^2$ to find this value. Find r_4 also.

§6. Analysis versus Modern Developments

As Leibniz grew older, he began to move away from infinitesimals and toward the notion of *limit*. The following excerpt from a letter written in 1702 is seen as evidence for this.

> One must remember ... that incomparably small quantities ... are by no means constant and determined. On the contrary, since they may be made as small as we like, they play the same part in geometric reasoning as the infinitely small in the strict sense. For if an antagonist denies the correctness of our theorems, our calculations show that the error is smaller than any given quantity, since it is in our power to decrease the incomparably small ... as much as is necessary for our purpose.
>
> Leibniz*

This passage says, in effect, that the use of infinitesimals can be regarded as a shortcut means of taking a limit. The objectionable reasoning in Example 5, Leibniz seems to say, is merely a quick way of getting the result that was derived with a little more care in Chapter 6, Section 1.

It was doubly difficult for Leibniz to discard the infinitesimal, for this conception had inspired both his mathematics and, in his theory of

* From a letter to Varignon, as given in *Ways of Thought of Great Mathematicians*, by Herbert Meschkowski, Holden-Day, 1964, p. 58.

monads, his philosophy. In deciding whether to disown his brainchild Leibniz must have experienced quite a struggle between mind and heart. His indecision is understandable. It was easier for Newton, who renounced the infinitely small in his later work, in favor of an intuitive understanding of the limit notion. Analysis in mathematics became based upon limits.

Atomism continues to survive today, though, and so do infinitesimals. Though sometimes, as in exercise 5.5. infinitesimals lead one astray, generally they point one in the proper direction like magic. Their success led Leo Tolstoy to seek an historical adaptation.

> [Infinitesimal calculus], unknown to the ancients, when dealing with problems of motion admits the conception of the infinitely small, and so conforms to the chief condition of motion (absolute continuity) and thereby corrects the inevitable error which the human mind cannot avoid when it deals with separate elements of motion instead of examining continuous motion.
>
> In seeking the laws of historical movement just the same thing happens. The movement of humanity, arising as it does from innumerable arbitrary human wills, is continuous....
>
> Only by taking infinitesimally small units for observation (the differential of history, that is, the individual tendencies of men) and attaining to the art of integrating them (that is, finding the sum of these infinitesimals) can we hope to arrive at the laws of history.
>
> Tolstoy*

These words were written in the middle of the nineteenth century, showing that infinitesimals were alive and kicking then. Even in the mid-twentieth century infinitesimals were used in calculus, though only, it was supposed to be emphasized, as a shortcut means of deriving what was done more rigorously by means of limits.

Analysis, a branch of mathematics growing out of calculus to develop a precise notion of limit, had in the late nineteenth century given calculus a firm foundation. Analysis tried to do for the calculus what Euclid had attempted to do for geometry: base the entire structure upon a few simple general principles. The "ε-δ" definition of a limit (the discussion of which we defer) has given a precise meaning to that notion. (The reader will have observed that the discussion of limits so far offered in this book has been completely intuitive in character.)

Quite recently mathematics has seen the exciting development of *nonstandard analysis*, which makes real sense out of infinitesimals. This work was pioneered in the 1960s by Abraham Robinson (1918–1974), who used sophisticated modern mathematical ideas to capture the intuitive notion of an infinitesimal. The discrete approach to the calculus, thought for so long to have been a heroic failure, may yet be a success after all.

* *War and Peace*, translated by Louise and Aylmer Maude, Oxford University Press, London, 1970, Book XI, Chapter I.

At the same time, a movement in quite the opposite direction has been born. Errett Bishop (1928–1983) and his followers have developed an approach to the calculus that is more down-to-earth and constructive in nature than traditional analysis. It will be interesting to see how calculus looks in the year 2015, on its 350-th birthday.

Exercises

6.1. Read Chapter 7, "The Beginning of Modern Mathematics, 1637–1687", in *The Development of Mathematics*, E.T. Bell, McGraw-Hill, New York, 1945.

6.2. Read the first section of Book Eleven of *War and Peace*, from which Tolstoy's quotation above is taken.

6.3. Read pp. 1–2 of *Non-standard Analysis*, Abraham Robinson, North-Holland, Amsterdam, 1966. Compare it with "A Constructivist Manifesto", pp. 1–10 of *Foundations of Constructive Analysis*, Errett Bishop, McGraw-Hill, New York, 1967.

§7. Faith versus Reason

Having had a very brief view of what has happened to calculus recently, let us get back to the early eighteenth century, where the old conflict between faith and reason still raged. A minor, but revealing incident in this conflict concerns infinitesimals, which became the ammunition for a skirmish between Edmund Halley, the astronomer, and George Berkeley (pronounced BARK-ly), the philosopher.

Halley, so the story goes, had persuaded a friend of Berkeley's to become skeptical about his religious beliefs, whereupon they were rejected on the grounds that theologians' claims could not be justified so soundly as the claims of mathematicians. This infuriated the Irishman Berkeley, who was about to be made a bishop in the Church of England. His outrage was so great that he sought not to shore up the foundations of theology, but to undermine those of mathematics. The result was an extraordinary essay, *The Analyst*, "a discourse addressed to an infidel mathematician".

> Whereas then it is supposed that you apprehend more distinctly, consider more closely, infer more justly, and conclude more accurately than other men, and that you are therefore less religious because more judicious, I shall claim the privilege of a Freethinker; and take the liberty to inquire into the object, principles, and method of demonstration admitted by the mathematicians of the present age, with the same freedom that you presume to treat the principles and mysteries of Religion; to the end that all men may see what right you have to lead, or what encouragement others have to follow you....

THE

ANALYST;

OR, A

DISCOURSE

Addressed to an

Infidel MATHEMATICIAN.

WHEREIN

It is examined whether the Object, Principles, and Inferences of the modern Analysis are more distinctly conceived, or more evidently deduced, than Religious Mysteries and Points of Faith.

By the AUTHOR of *The Minute Philosopher.*

First cast out the beam out of thine own Eye; and then shalt thou see clearly to cast out the mote out of thy brother's eye. S. Matt. c. vii. v. 5.

LONDON:

Printed for J. TONSON in the *Strand*. 1734.

Figure 5. Title page of Berkeley's "The Analyst".
[H]e who can digest a second or third fluxion ... need not, methinks, be squeamish about any point in divinity.

Berkeley wrote *The Analyst* in 1734, not too many years after the deaths of Newton and Leibniz. In his essay Berkeley forcefully made the argument given at the end of Section 5, in criticism of the logic used in infinitesimal calculus. He pointed out quite rightly that the seventeenth century was content to accept arguments that the ancient Greeks would have discarded as inadequate. The implication was that seventeenth-century mathematicians were accepting arguments on faith, not on reason. Berkeley then went on drily to inquire whether infinitesimals were not "ghosts of departed quantities", implying that nothing in theology could be more ghostlike than the basic notion of infinitesimal calculus.

Attempts by mathematicians to answer Berkeley's splendid philippic sometimes became too verbose to be effectual. Today it is admitted by virtually every student of mathematics that some of Berkeley's objections to the calculus were unanswerable until the nineteenth century, when analysis at last produced a precise definition of the notion of a *limit*.

Exercises

7.1. Read one of the following chapters from *Mathematics in Western Culture*, Morris Kline, Oxford University Press, New York, 1953.
   (a) Chapter XVI, "The Newtonian Influence: Science and Philosophy".
   (b) Chapter XVII, "The Newtonian Influence: Religion".
   (c) Chapter XVIII, "The Newtonian Influence: Literature and Aesthetics".

7.2. Read the excerpts and commentary arising from Berkeley's *Analyst* given on pp. 286–293 of *World of Mathematics*, edited by James R. Newman, Simon and Schuster, New York, 1956.

7.3. Read Chapter 13, "From Intuition to Absolute Rigor, 1700–1900", in *The Development of Mathematics*, E.T. Bell, McGraw-Hill, New York, 1945.

§8. Conclusion

It is curious that, despite his fulminations, Bishop Berkeley accepted the calculus on faith. He believed, he said, that correct results in the calculus were the product of some "compensation of errors" in reasoning.

Mathematicians soon became aware of the shaky ground on which the calculus was erected, as indicated by the admonition of d'Alembert (1717–1783): "Go forward, and faith will follow!" In the conflict between faith and reason, mathematics had the potential for use by either side. Among the founders of the calculus both Leibniz and Newton exhibited strong interest in theology. Leibniz made a serious attempt to reunite the

Protestant and Catholic churches, and Newton thought of his work as helping to prove the existence of God. The study of things eternal may tend to heighten one's awareness of religion.

The first edition (1771) of the *Encyclopaedia Britannica* quoted with approval the sentiments expressed a century earlier by Issac Barrow. Barrow was Newton's teacher, who resigned from Cambridge in order that Newton be given his professorship. The words could have been written by Plato himself:

> The mathematics ... effectually exercise, not vainly delude, nor vexatiously torment, studious minds with obscure subtilties; but plainly demonstrate every thing within their reach, draw certain conclusions, instruct by profitable rules, and unfold pleasant questions. These disciplines likewise enure and corroborate the mind to a constant diligence in study; they wholly deliver us from a credulous simplicity, most strongly fortify us against the vanity of scepticism, effectually restrain us from a rash presumption, most easily incline us to a due assent, perfectly subject us to the government of right reason. While the mind is abstracted and elevated from sensible matter, distinctly views pure forms, conceives the beauty of ideas, and investigates the harmony of proportions; the manners themselves are sensibly corrected and improved, the affections composed and rectified, the fancy calmed and settled, and the understanding raised and excited to more divine contemplations.*

Nevertheless, mathematics appears generally seen as allied with reason in opposition to faith. Voltaire pointed to the spectacular achievements of "rational" science and mathematics, and demanded the right to examine everything under the authority of reason. Voltaire won out, for a time, and it was little noted that mathematicians of the Enlightenment were using their instinct more than their intellect:

> They defined their terms vaguely and used their methods loosely, and the logic of their arguments was made to fit the dictates of their intuition. In short, they broke all the laws of rigor and of mathematical decorum.
>
> The veritable orgy which followed the introduction of the infinitesimals ... was but a natural reaction. Intuition had too long been held imprisoned by the severe rigor of the Greeks. Now it broke loose, and there were no Euclids to keep its romantic flight in check.
>
> Tobias Dantzig†

The Enlightenment, emphasizing intellect, was to be washed aside by the Romantic Movement that declared, with Rousseau, the primary nature of instinct. Romanticism can be, and has been, described in a variety of ways. But if it is marked by a reaction against Neoclassicism, an emphasis upon imagination, a disregard for decorum, and a predi-

* "Mathematics", *Encyclopaedia Britannica*, Vol. III, 1771, pp. 30–31.

† *Number, the Language of Science*, Macmillan, New York, 1939, p. 130.

lection for the seer to the sage, then the Romantic Movement was already rampant in mathematics at the very height of the Age of Reason.

Is this a paradox? Or is it a blunt reminder to us that mathematics has always incorporated the elements of both movements? Mathematics is romance in reason.

Problem Set for Chapter 8

1. Much of our knowledge of Greek mathematics comes from a commentary written on the first book of Euclid's *Elements* by Proclus, who lived in the fifth century. In his commentary Proclus says that mathematics

   > arouses our innate knowledge, awakens our intellect, purges our understanding, brings to light the concepts that belong essentially to us, takes away the forget-fulness and ignorance that we have from birth, sets us free from the bonds of unreason; and all this by the favor of the god [Hermes?] who is truly the patron of this science, who brings our intellectual endowments to light, fills everything with divine reason, moves our souls towards Nous [the highest form of knowledge], awakens us as it were from our heavy slumber, through our searching turns us back upon ourselves, through our birthpangs perfects us, and through the discovery of pure Nous leads us to the blessed life. (*Proclus: A Commentary on the First Book of Euclid's Elements*, translated by Glenn R. Morrow, Princeton University Press, 1970, p.38)

   (a) Compare this passage with the passage quoted in Chapter 8, Section 8, written in the *Encyclopaedia Britannica* some 1300 years later.
   (b) From the quotation above it is obvious whom Proclus thought to be the supreme philosopher. Was it Plato or Aristotle?

2. Name at least five noted philosophers who have also been mathematicians. Is it just an accident that some of the most eminent philosophers have also been mathematicians?

3. The Greek writings of Plato, the French of Pascal, Descartes, and Poincaré, the English of Russell and Whitehead have all been acclaimed as models of prose style by students of literature. Is it just an accident that some of the most eminent writers have also devoted themselves to mathematics?

4. Consider each of the following relatively recent statements. Which of them is virtually a restatement of the principle of continuity? Which of them remind you of Pythagoras? of Plato? of Archimedes? of Leibniz?
   (a) Remote from human passions, remote even from the pitiful facts of nature, the generations have gradually created an ordered cosmos, where pure thought can dwell as in its natural home and where one, at least, of our nobler impulses can escape from the dreary exile of the actual world (Bertrand Russell, twentieth century).
   (b) Since the fabric of the world is the most perfect and was established by the wisest Creator, nothing happens in this world in which some reason of maximum or minimum would not come to light (Leonhard Euler, eighteenth century).

(c) Apart from a certain smoothness in the nature of things, there can be no knowledge, no useful method, no intelligent purpose (Alfred North Whitehead, twentieth century).
(d) The great book of Nature lies ever open before our eyes and the true philosophy is written in it ... But we cannot read it unless we have first learned the language and the characters in which it is written ... It is written in mathematical language (Galileo, seventeen century).
(e) I have never done anything 'useful'.... The case for my life ... is this: that I have added something to knowledge, and helped others to add more; and that these somethings have a value which differs in degree only, and not in kind, from that of the creations of the great mathematicians, or of any of the artists, great or small, who have left some kind of memorial behind them (G.H. Hardy, twentieth century).

5. Isaac Newton wrote, "By a Number we understand not so much a Multitude of Unities, as the abstracted Ratio of any Quantity to another Quantity of the same Kind, which we take for Unity." Contrast this modern notion of number with the ancient Greek idea, and explain the importance of this new point of view to the development of mathematics and science in modern times.

6. In his notebook Voltaire wrote, "Before Kepler all men were blind. Kepler had one eye, Newton had two." Compare Voltaire's extravagant praise of the scientific spirit with the sentiments expressed in the following poem by the English poet and mystic William Blake (1757–1827). Note that Blake expects his reader to be familiar with the "atoms of Democritus" and Newton's "particles of light"—topics we met briefly in Chapter 8.

   Mock on, Mock on Voltaire, Rousseau:
   Mock on, Mock on: 'tis all in vain!
   You throw the sand against the wind,
   And the wind blows it back again.

   And every sand becomes a Gem
   Reflected in the beams divine;
   Blown back they blind the mocking Eye,
   But still in Israel's paths they shine.

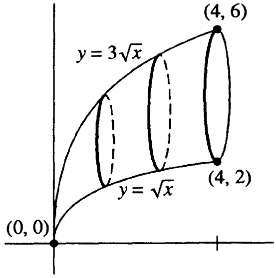
   The Atoms of Democritus
   And Newton's Particles of light
   Are sands upon the Red sea shore,
   Where Israel's tents do shine so bright.

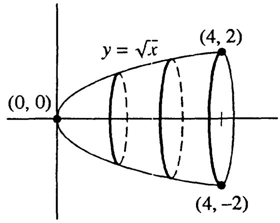
   What does Blake think of the eighteenth-century French *philosophes* who felt that the modern age began with the publication of Newton's *Principia*?

7. In Chapter 7, Cavalieri's Principle was discussed for figures lying in the plane. Actually, Cavalieri formulated an analogous principle for figures lying in three-dimensional space. Can you? "Two solid figures have the same volume if ..."

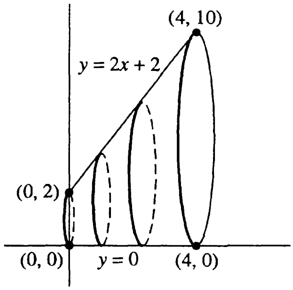
8. Consider the figure on the left below that looks like a powder-horn, where each vertical cross-section is a circle.

(a) Explain, using your answer to problem 7, why you should expect the powder-horn's volume to be equal to the volume of the solid of revolution pictured on the right.
(b) Find the volume of the powder-horn, using the result of part (a) and the integral formula for the volume of this solid of revolution.





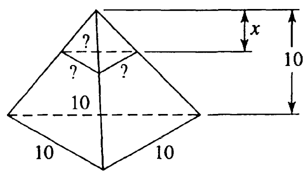
9. Find the volume of the solid figure below, where each vertical cross-section is a circle. *Hint.* You can handle this solid just as you handled the powder-horn above.



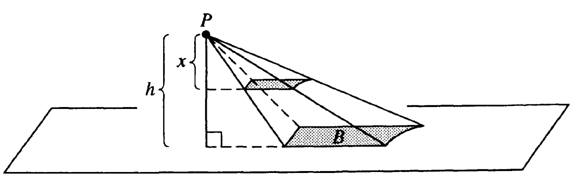
10. An intuitive idea behind infinitesimal calculus is that you can get the area of a figure in the plane by "integrating the lengths of its vertical cross-sections". Can you extend this idea to three dimensions? "You get the volume of a figure in three-dimensional space by integrating the ____ of its vertical cross-sections." *Hint.* In three dimensions the vertical cross-sections are two-dimensional, so it no longer makes sense to speak of their "length". Look at the integrals you have as answers to the preceding two problems and note what measurement of the cross-sections you are integrating.

11. Consider the figure below, which is a pyramid whose base is an equilateral triangle with sides of 10 units each, and whose height perpendicular to this base is also 10 units.

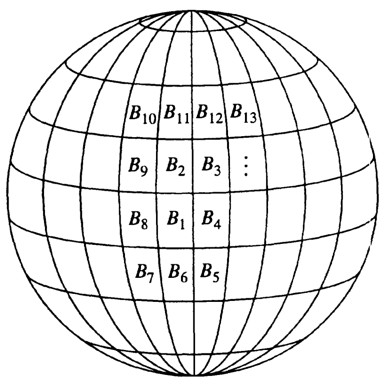
(a) Show, using similar triangles, that the horizontal cross-section lying a distance of x units down from the top vertex is an equilateral triangle with sides of $x/10$ units.
(b) Find the area of an equilateral triangle, each of whose sides has length $x/10$.
(c) Use your answers to problems 10, 11(a), and 11(b) to write down an integral giving the volume of this pyramid. (You may consider the horizontal cross-sections to be vertical cross-sections, of course, just by rotating the figure ninety degrees.)
(d) Use the fundamental theorem of calculus to evaluate the integral of part (c).
(e) Does it turn out, from your answer to part (d), that the volume of the pyramid is equal to one-third the area of its base times its height?



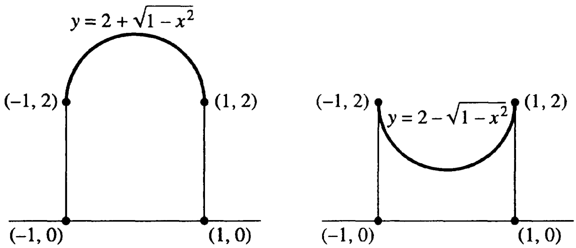
12. (*Much easier than it may at first appear*) Take any familiar figure in the plane, such as a triangle, a circle, an ellipse, or even something more complicated as indicated in the drawing below. Let the point P be situated a distance h above the plane and consider the solid—which we might call the (generalized) cone built upon the given planar figure—made up of all lines joining P to the given planar area. Go through the following steps to show that if B is the numerical measure of the given planar area, then the volume V of the cone built upon it is given by $V = \frac{1}{3}Bh$.
(a) Show that the indicated cross-sectional area located a distance of x units down from P is similar to the area at the base of the cone, with the ratio x/h being the contraction factor involved in this similarity.
(b) Explain why, if the contraction factor between similar planar figures is x/h, then the numerical ratio of their areas is x^2/h^2.
(c) Deduce from parts (a) and (b) that the numerical measure of the cross-sectional area at height x above the plane is Bx^2/h^2.
(d) Use part (c) and the result of problem 10 to show that the required volume V is given by the integral $\int_0^h (B/h^2)x^2\,dx$.
(e) Remembering that B and h are constants here, use the fundamental theorem to evaluate the integral in part (d).



13. (*The surface area of a sphere*) Let us assume that if you own land on the surface of the earth, then you also own all the land beneath it, down to the point P at the center of the earth. In other words, you own the generalized cone built on your land with vertex P.
    (a) Find the volume of earth that you own, expressed in terms of B and r, where B is the numerical measure of the area owned by you on the surface of the earth and r is the radius of the earth. *Hint.* Use the result of problem 12, assuming your land—which lies on the (curved) surface of the earth—is in fact "flat".
    (b) Show that the volume of earth owned by n small flatland-owners is given by $(B_1 + B_2 + \cdots + B_n)r/3$, where B_j is the surface area of the j-th landowner.
    (c) Assume the entire surface S of the earth is owned by n small flatland-owners, so that $S = B_1 + B_2 + \cdots + B_n$. Use the result of part (b) to show that the volume V of the earth is given by $V = Sr/3$.
    (d) Argue that the formula $V = Sr/3$ must hold for a perfect sphere. *Hint.* Use limits, noting that by part (c) this formula holds for arbitrarily many "flat" subdivisions of the sphere, each of which can be arbitrarily small.
    (e) We know that $V = 4\pi r^3/3$ from exercise 10.3 of Chapter 7. Combine this equation with the equation $V = Sr/3$ to get the surface area S in terms of r.
    (f) Compare your answer to part (e) with your answers to exercises 1.1 and 5.5.



14. (*The volume of a doughnut*) One way to find the volume of a doughnut is to find the volume of the "solid" doughnut and then subtract the volume of the "hole".
    (a) Find the volume of the solid of revolution obtained by revolving each of the areas below about the x-axis. *Hint.* In the course of doing this, you will be faced with evaluating the integral $\int_{-1}^{1} \sqrt{1 - x^2}\,dx$. This, of course, is $\pi/2$. (*Why?*)



(b) Find the volume of the "doughnut" obtained by revolving about the x-axis the area within the circle with equation $x^2 + (y - 2)^2 = 1$. *Hint.* All you have to do is find the difference between the two volumes you calculated in part (a). (*Why?*)

(c) Use the technique of part (b) to find the volume of the "doughnut" obtained by revolving about the x-axis the area within the circle with equation $x^2 + (y - 4)^2 = 4$. *Comment*: There is a shorter way to get the answer, however, for the doughnut of part (c) is similar to the doughnut of part (b) by a "stretching factor" of two in each dimension. (Draw a picture of the two doughnuts to see this.) What must then be the ratio of the volumes of these similar solids?

15. Pappus of Alexandria, who lived in the fourth century, stated that the volume of a solid of revolution is equal to the numerical value of the area revolved, multiplied by the distance through which its center of gravity revolves. The center of gravity of a circle is, of course, its ordinary center. Pappus' proof, however, has been lost. Let us test whether Pappus is correct in the situations described in problem 14.
    (a) Multiply the area of the circle in part (b) of problem 14 by 4π, the distance through which the point (0, 2) moves as it revolves 360 degrees about the x-axis. Is the result equal to the volume you calculated for the doughnut described in that problem?
    (b) Multiply the area of the circle of part (c) of problem 14 by the distance through which (0, 4) revolves. Does the result agree with the volume you calculated in part (c) of problem 14?

16. Consider the doughnut made by revolving about the x-axis the area within the circle whose equation is $x^2 + (y - 3)^2 = 4$.
    (a) Find the volume of this doughnut using the technique of problem 14(b).
    (b) Find the volume of this doughnut using the technique of problem 15.

17. (*The most beautiful proof in mathematics?*) Recall our simple *reductio ad absurdum* proof from Chapter 2 showing that there is no largest integer. (Suppose there is, call it N, then quickly derive a contradiction by considering the integer $N + 1$.) By following the steps below, show likewise that *there is no largest prime*.
    (a) Suppose there is a largest prime and call it P. Then 2, 3, 5, 7, 11, . . . , P is a complete list of all the primes. Consider the integer N made by adding 1 to the product of all these primes: $N = 1 + 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdots P$. Show

that N is not divisible by any of the primes in our complete list. *Hint.* Show that N is equal to 1 plus a multiple of 2, so N has a remainder of 1 when divided by 2; then show that N is equal to 1 plus a multiple of 3, so N has a remainder of 1 when divided by 3; and so on.

(b) Show that N is not prime. *Hint.* We are supposing that P is the largest prime. Is N larger than P?

(c) Show that N must be divisible by some prime. *Hint.* Since we have just established that N is not prime, it must be equal to the product of two smaller integers: $N = A \cdot B$, where A and B are both less than N. If either A or B is prime we have shown what is required; if not, then A can be factored into the product of two smaller integers: $A = C \cdot D$. If one of these is prime we have shown what is required; if not, then C can be factored into the product of still smaller integers. And so on. Show that eventually we must get a prime factor of N this way. *Further hint.* Show that if we never get a prime this way, then we get an infinite descent $N, A, C, \ldots$ of positive integers, and recall Fermat's method of infinite descent described in problem 26 of Chapter 4.

(d) Show that the supposition that there is a largest prime leads to a contradiction. *Hint.* The supposition that there is a largest prime leads to the result of part (a) and the result of part (c).

This is essentially the proof given in Euclid's *Elements*, showing that the primes march on forever. It is as close to poetry, perhaps, as elementary mathematics can come.

18. The American poet Wallace Stevens (1879–1955) took a train trip through the South in 1918 and spent some time in Tennessee. Shortly thereafter he published *Anecdote of the Jar*, which is reprinted as an epigraph to this book just before page 1. What do you think is the main theme of *Calculus: A Liberal Art*, and in what ways does Stevens's poem reflect this theme? (You may wish to compare your answer with the answer given in "Mathematics and Poetry: How Wide the Gap?", *The Mathematical Intelligencer*, Vol. 12, No. 1, 1990, pp. 14–19.)

APPENDIX 1

Writings "About" Mathematics

The complaint of some humanists that mathematicians make no attempt to describe to others their function is unjustified. On the contrary, a list of articles and books published with this purpose in mind is extensive, owing to the efforts of a number of writers, many of whom have been distinguished mathematicians.

This appendix calls attention to some of these writings about mathematics, which are by and large nontechnical in nature and addressed to the general reader. It also serves as a way of acknowledging, however inadequately, the debt owed by the author of this book to the writings of others. The list of works mentioned reflects to some degree, of course, the taste of the author.

Quite an extensive list of such writings is included among those cited by Matthew P. Gaffney and Lynn Arthur Steen in their *Annotated Bibliography of Expository Writings in the Mathematical Sciences* (Mathematical Association of America, Washington, D.C., 1976). Since then, continual updatings have appeared in the *Mathematics Magazine* and, more recently, in the *College Mathematics Journal.* All of these are published by the Mathematical Association of America, which may soon organize this large collection and make it available on the World Wide Web.

§1. The Nature of Mathematics

The quotation just inside the title page of this text is taken from Alfred North Whitehead's chapter on mathematics in *Science and the Modern*

World [W7].* This essay, "Mathematics as an Element in the History of Thought", is a modern classic. Quite a different kind of article, but equally celebrated, is Henri Poincaré's "Mathematical Creation". The fascinating story told here by the great French mathematician is often mentioned in discussions of creativity. Poincaré's article may be found in [G2], for example, as well as in [P3]. It inspired Hadamard's *The Psychology of Invention in the Mathematical Field* [H1].

It is easy to justify mathematics in terms of its utility, but G.H. Hardy's brief and controversial defense [H5] of mathematics rests largely upon aesthetic grounds. *A Mathematician's Apology* has been reprinted, with a foreword by C.P. Snow. Of this unique book, Graham Greene has said, "There is nothing here which the layman cannot understand except possibly one theorem, and I know no writing—except perhaps Henry James's introductory essays—which conveys so clearly and with such an absence of fuss the excitement of the creative artist." Davis and Hersh [D4] paint quite a different picture of mathematics, which they see as a social enterprise helping to mold our collective consciousness—not as a part of the Platonic realm envisaged by Hardy.

The attractiveness of mathematics as an activity similar to creative endeavors in the fine arts is the theme of an article [H2] by Paul Halmos, which can also be found in [C3], as well as in the collection [H3] of Halmos's expository articles. In Rothstein's *Emblems of Mind* [R4] the Pythagorean callings of mathematics and music are seen to be animated by the pursuit of beauty and truth in subtly analogous ways. Mathematics is tied to poetry in [B7] and [P5], to all the humanities in [W6], and to beauty itself in [L1]. This theme is given its most eloquent expression, perhaps, by Bertrand Russell: "Mathematics, rightly viewed, possesses not only truth, but supreme beauty ..." (See Chapter 6, Section 1, for the rest.) It is only fair to add here that Russell later partially repudiated some of these sentiments in [R5], owing to certain developments in twentieth-century philosophy.

Douglas Hofstadter's *Gödel, Escher, Bach* [H7] winds around the paradoxes of self-referentiality in mathematics, painting, and music. Roger Penrose challenges Hofstadter's views on computers and thinking in [P1] and [P2]. For more on how minds may work, see the intriguing book by Marvin Minsky [M6].

Hardy says that mathematics is not a contemplative, but a creative subject. One cannot know the nature of mathematics without doing some mathematics. The books by George Pólya [P4] and by Courant and Robbins [C8] will help the reader learn what Hardy means. Anyone who wishes to know what mathematics is like as a profession should read Paul Halmos's "automathography" [H4].

* Expressions is square brackets refer to the listing in the bibliography at the end of this appendix.

§2. About Mathematicians

E.T. Bell's *Men of Mathematics* [B2] is probably the most popular book about mathematicians. In his inimitable, crusty style, Bell relates the lives and achievements of some thirty mathematicians, the latest being Georg Cantor, who died in 1918. Bell's occasional penchant for telling a good story without thoroughly checking the facts, however, should be borne in mind. Dauben's definitive biography [D2] of the beleaguered Cantor, for example, paints quite a different picture from Bell's in several respects.

Despite its title, *Men of Mathematics* devotes some space to women, such as Sofya Kovalevskaya (1850–1891), the first woman to earn a doctorate in mathematics. Kovalevskaya was not only a star student of Weierstrass, but an accomplished Russian poet and novelist, who wrote about her early life in [K7]. According to Reid [R3], *Men of Mathematics* was instrumental in Julia Robinson's becoming a mathematician. It was only a matter of time, of course, before there should appear a book entitled *Women of Mathematics* [G7]; see also [O2].

Only a few mathematicians have written autobiographies; one of some literary merit is by Bertrand Russell, who won the Nobel Prize for Literature in 1950. Since Russell did little in mathematics after World War I, only Volume I of his book [R6] is a mathematician's autobiography. Norbert Wiener, who received his Ph.D. degree at the age of eighteen and studied under Russell and Hardy, has written a two-volume autobiography [W8], [W9]. Wiener is the father of cybernetics.

Two Polish mathematicians, Mark Kac and Stanislaw Ulam, who collaborated on the excellent but demanding essay *Mathematics and Logic* [K2], have written lively autobiographies [K1], [U1], as has the celebrated French mathematician André Weil [W2]. All three of these moved from their native countries to continue their work in the United States. One might have expected Weil in [W2] to give us more than just a glimpse of his famous sister, Simone, but he has written about her elsewhere.

Robert Kanigel has written for the general reader an absorbing biography [K3] of the brilliant, self-taught Indian mathematician Srinivasa Ramanujan, whom G.H. Hardy "discovered". Oystein Ore's biography [O1] of the great Norwegian mathematician Niels Henrik Abel is well worth reading. Abel died at the age of twenty-six. Michael Mahoney has described Fermat's mathematical career [M2], and Leopold Infeld [I1] has written movingly of another remarkable French mathematician, Evariste Galois, who died at twenty. Isaac Newton continues to receive much attention [W3], [F1].

A thoughtful biography [R1] of David Hilbert is the work of Constance Reid, who followed it with the life story [R2] of Richard Courant and a personal account [R3] of her mathematician sister, Julia Robinson.

Hilbert, long associated with Göttingen, helped chart the course of modern mathematics by outlining in 1900 twenty-three unsolved problems of central importance for the future; and Courant, in moving from Göttingen to New York, was instrumental in raising the level of mathematical research in the United States. Julia Robinson helped solve one of Hilbert's problems.

The multivolumed *Dictionary of Scientific Biography* [G4] is an invaluable source of information on mathematicians and scientists of the past. A good impression of the personalities of contemporary mathematicians may be found in the lively interviews collected in [A1] and [A2].

As a rule, of course, only the celebrated mathematician is written about. The ordinary teacher of mathematics, upon whom real inspiration descends grudgingly and fleetingly, is usually thought undeserving of attention. But Donald Weidman has written a brief, sympathetic account [W1], now reprinted in [C3], of the run-of-the-mill mathematician's fate.

§3. History and Development of Mathematics

D.J. Struik's *A Concise History of Mathematics* [S7] is most handy for the general reader. Four good textbooks used by undergraduate courses in the history of mathematics are Eves [E2], Boyer [B5], Burton [B9], and Katz [K4]. Eves gives a fuller treatment of geometry, whereas Boyer leans toward analysis and Burton toward number theory. Katz gives more consideration to the development of mathematics by non-Western cultures, often driven by the needs of astronomy. Ivor Grattan-Guinness's recent *History of the Mathematical Sciences* [G6] is large, comprehensive, lively, and witty.

Boyer wrote an earlier book on the development of calculus, but on this subject one may prefer the later investigation of Henry Edwards [E1] or, for quite a different kind of treatment, the concise genetic approach of Otto Toeplitz [T1]. Simmons [S2] and Dunham [D7] offer calculus students rare mathematical gems and engaging historical sketches; Dudley [D6] provides readings for a course in calculus.

Dantzig's book [D1], dealing mainly with analysis, is easy to read. Meschkowski's little book [M5] is good, as is Kline's big book [K5]. Bell [B3] is unique. For samples of little gems by ancient and classical mathematicians such as Archimedes, Pascal, and Leibniz, see the collections by Coolidge [C7] or Meschkowski [M4]. Extensive source books are Smith [S3] and Struik [S8]; Calinger [C1] is most engaging. Collections of historical readings, such as [F2], [C2], and [B1], may be utilized by instructors in diverse ways to help integrate the history of mathematics with its pedagogy.

If Greek mathematics is your interest, then you may find van der Waerden [V1] most rewarding. Here is an exciting account by a distinguished mathematician who writes especially for the interested general reader. Fowler [F3] provides a new look at the mathematics of Plato's Academy. The puzzling question of how much Plato has to do with our conception of Pythagoras is discussed in Burkert's scholarly book [B8]. Neugebauer [N3] writes about Babylonian and Egyptian mathematics, and Gillings [G3] offers more on Egypt.

Historians continue to uncover interesting mathematical activities in all ancient civilizations, and even to discern "proto-mathematical" activities in less well-developed cultures all over the globe. Frank Swetz's collection [S9] shows how diverse is the cultural base of mathematics, but opinion still varies as to what constitutes "real mathematics" and what, therefore, the history of mathematics should really be about. Do we make too much of the character impressed upon mathematics by the development of the axiomatic method? Do we praise the Greeks too highly for the consequent central role played by mathematics in the rise of liberal education? A multicultural view of mathematical ideas is given by Ascher [A4], and non-European roots of mathematics are discussed further in Joseph's interesting and provocative book, *The Crest of the Peacock* [J1].

§4. Philosophy of Mathematics

Is mathematics created or is it discovered? What are the foundations of mathematics? Answers to such questions may vary according to whether one subscribes to the philosophy of intuitionism, formalism, logicism, or Platonism. DeLong's much-praised *Profile of Mathematical Logic* [D5], which helped to inspire Hofstadter's book [H7], considers such topics. A briefer treatment may be found in Körner [K6]. Anglin and Lambek offer a concise textbook [A3] on the history, philosophy, and foundations of mathematics.

Hermann Weyl, who was an intuitionist, wrote a deep book [W4], *Philosophy of Mathematics and Natural Science*. It demands deliberate reading. Poincaré is articulate on this subject, as on every other subject collected in his essays [P3]. The intuitionist position has been revived and modified, with the coming into prominence of constructive methods. See the review [S6] of Bishop's *Foundations of Constructive Analysis*, which is reprinted, although much abridged, in [C3].

Hilbert was the foremost proponent of formalism, and passages in [R1] attempt to capture the spirit of the early twentieth-century debate over the validity of the formalist thesis. The celebrated theorem proved by Kurt Gödel in 1931, which was initially seen as a great setback for Hilbert's hopes, has been outlined for the general reader by Nagel and Newman [N1].

Russell and Whitehead promulgated the logistic thesis. See Henkin [H6], also reprinted in [C3], for some comments on its current state.

Platonism remains a lively issue, receiving hostility in [S2], dismissal in [D4], indifference in [H7], sympathy in [H4] and [G1], qualified endorsement in [P1], efforts at rehabilitation in [M1], and unabashed affection in [R4]. It is quite possible, however, that the proving of mathematical theorems will increasingly rely upon a symbiotic relationship between human beings and computers. Does this mean that formalism will win out?

Many important issues in the philosophy of mathematics turn on one's answer to a seemingly simple question: *What is a real number?* One might think that all mathematicians would agree upon real numbers by now, since numbers have been around for a long time, if not eternally. But see Steen [S4]. For more on nonstandard analysis, see Dauben's biography [D3] of its creator, Abraham Robinson.

§5. Collections of Expository Articles

If a guide to writings about mathematics listed only one entry, it would have to be the four-volume set *World of Mathematics* [N4], edited by James R. Newman. Some of the articles mentioned elsewhere here are reprinted, although often abridged, in this superb collection. Anyone interested in mathematics is probably already familiar with this work. The collection by Campbell and Higgins [C3] is a worthy supplement of more recent writings.

Shortly after World War II there appeared in France an ambitious collection of expository essays about mathematics. Written mainly by French mathematicians, these cover a great range of topics, even including the relationship between mathematics and music, aesthetics, philosophic idealism, social change, and Marxism. Although some are quite dated by now, many retain their original striking quality. It is good to have them available in English [L2].

An even greater range of topics is covered in the encyclopedia edited by Grattan-Guinness [G5], which offers, in addition, a rich source of references to works about mathematics at the ends of appropriate articles.

§6. Miscellaneous Writings

The book by Littlewood [L3], Hardy's great collaborator, surely goes under the heading of miscellaneous writings. Hermann Weyl's *Symmetry* [W5] and Hugo Steinhaus's *Mathematical Snapshots* [S5] have been widely admired. J.D. Williams has written for the layman a delightful book on game theory, *The Compleat Strategyst* [W10].

Mathematics plays an important role in Jacob Bronowski's *Ascent of Man* [B6]. Menninger's *Number Words and Number Symbols* [M3] must be on the coffee table of every modern Pythagorean, and a more recent view of numbers is given by Conway and Guy [C6]. Vilenkin's book [V2] is an attractive elementary introduction to infinity as a mathematical concept.

Roger Nelson's "wordless proofs" [N2] are often striking. Recreational mathematics and its history is discussed in [G8]. Osserman [O3] attempts to explain to the general reader the curved space-time of the cosmos. Bochner's collection of essays [B4] contains a delightfully idiosyncratic appendix. Here, for example, Moritz Cantor's huge tome [C4], once considered the definitive history of mathematics (up to 1800), is characterized as "one of those large-scale works by bearded gaslight-Victorians which the 20th century does not quite know how to supersede with whatever it might try to supersede them with."

Mathematical Gems [H8] is the first of a series of brief books, all with similar titles, written by Ross Honsberger containing imaginative discussions of a variety of intriguing problems. Eves's books [E3], [E4], also the first of a series of like titles, help preserve the folklore of mathematics, as does Moritz's classic collection [M7] of "witty, profound, amusing passages about mathematics and mathematicians". Quotable passages of more recent vintage may be found in Schmalz [S1].

Martin Gardner, whose column on mathematics in *Scientific American* was widely read for twenty-five years, has recently given us a broad selection of his essays [G1], including many articles of interest about mathematics. One of Gardner's most popular publications has been his annotated edition [C5] of "Alice in Wonderland", the brainchild of that dour Oxford professor of mathematics, the Reverend Charles Lutwidge Dodgson.

Bibliography

[A1] Albers, D. and Alexanderson, G. *Mathematical People*, Birkhäuser, Boston, 1985.

[A2] Albers, D., Alexanderson, G. and Reid, C. *More Mathematical People*, Harcourt Brace Jovanovich, Orlando FL, 1990.

[A3] Anglin, W. and Lambek, J. *The Heritage of Thales*, Springer-Verlag, New York, 1995.

[A4] Ascher, M. *Ethnomathematics: A Multicultural View of Mathematical Ideas*, Brooks Cole, Pacific Grove CA, 1991.

[B1] Bekken, O., Fauvel, J., Johansson, B., Katz, V. and Swetz, F. *Learn from the Masters*, Mathematical Association of America, Washington DC, 1995.

[B2] Bell, E. *Men of Mathematics*, Simon and Schuster, New York, 1937.

[B3] Bell, E. *The Development of Mathematics*, McGraw-Hill, New York, 1945.

[B4] Bochner, S. *The Role of Mathematics in the Rise of Science*, Princeton University Press, Princeton, 1966.

[B5] Boyer, C. *A History of Mathematics*, rev. by Uta Merzbach, Wiley, New York, 1991.

[B6] Bronowski, J. *The Ascent of Man*, Little, Brown, New York, 1973.

[B7] Buchanan, S. *Poetry and Mathematics*, John Day, New York, 1929.

[B8] Burkert, W. *Lore and Science in Ancient Pythagoreanism*, Harvard University Press, Cambridge MA, 1972.

[B9] Burton, D. *History of Mathematics: An Introduction*, Brown, Dubuque IA, 1995.

[C1] Calinger, R., ed. *Classics of Mathematics*, Prentice Hall, Englewood Cliffs NJ, 1995.

[C2] Calinger, R., ed. *Vita Mathematica: Historical Research and Integration with Teaching*, Mathematical Association of America, Washington DC, 1996.

[C3] Campbell, D. and Higgins, J., eds. *Mathematics: People, Problems, Results*, 3 vols., Wadsworth, Belmont CA, 1984.

[C4] Cantor, M., ed. *Vorlesungen über Geschichte der Mathematik*, 4 vols., Johnson reprint, New York, 1965.

[C5] Carroll, Lewis, *The Annotated Alice: Alice's Adventures in Wonderland & Through the Looking Glass*, notes by Martin Gardner, World Publishing Company, New York, 1963.

[C6] Conway, J. and Guy, R. *The Book of Numbers*, Springer-Verlag, New York, 1996.

[C7] Coolidge, J. *The Mathematics of Great Amateurs*, Dover, New York, 1963.

[C8] Courant, R. and Robbins, H. *What is Mathematics?* Oxford, New York, 1941.

[D1] Dantzig, T. *Number, the Language of Science*, Doubleday Anchor, New York, 1956.

[D2] Dauben, J. *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge MA, 1979.

[D3] Dauben, J. *Abraham Robinson: The Creation of Nonstandard Analysis, a Personal and Mathematical Odyssey*, Princeton University Press, Princeton, 1995.

[D4] Davis, P. and Hersh, R. *The Mathematical Experience*, Birkhäuser, Boston, 1980.

[D5] DeLong, H. *A Profile of Mathematical Logic*, Addison-Wesley, New York, 1970.

[D6] Dudley, U. *Readings for Calculus*, Mathematical Association of America, Washington DC, 1993.

[D7] Dunham, W. *Journey through Genius*, Wiley, New York, 1990.

[E1] Edwards, C.H. *The Historical Development of the Calculus*, Springer-Verlag, New York, 1979.

[E2] Eves, H. *An Introduction to the History of Mathematics*, Holt Rinehart, and Winston, New York, 1969.

[E3] Eves, H. *In Mathematical Circles*, Prindle, Weber, and Schmidt, Boston, 1969.

[E4] Eves, H. *Mathematical Circles Revisited*, Prindle, Weber, and Schmidt, Boston, 1971.

[F1] Fauvel, J., Flood, R., Shortland, M. and Wilson, R., eds. *Let Newton Be!* Oxford, New York, 1988.

[F2] Fauvel, J. and Gray, J., eds. *The History of Mathematics: A Reader*, Macmillan, London, 1987.

[F3] Fowler, D. *The Mathematics of Plato's Academy: A New Reconstruction*, Oxford, New York, 1990.

[G1] Gardner, M. *The Night is Large: Collected Essays, 1938–1995*, St. Martin's Press, New York, 1996.

[G2] Ghiselin, B., ed. *The Creative Process*, Mentor, New York, 1952.

[G3] Gillings, R. *Mathematics in the Time of the Pharaohs*, M.I.T. Press, Cambridge, 1972.

[G4] Gillispie, C., ed. *The Dictionary of Scientific Biography*, 18 vols., Scribners, New York, 1970–1997.

[G5] Grattan-Guinness, I., ed. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, 2 vols., Routledge, New York, 1994.

[G6] Grattan-Guinness, I. *The Fontana History of the Mathematical Sciences: The Rainbow of Mathematics*, Fontana Press, London, 1997.

[G7] Grinstein, L. and Campbell, P. *Women of Mathematics: A Biobibliographic Sourcebook*, Greenwood Press, New York, 1987.

[G8] Guy, R. and Woodrow, R., eds. *The Lighter Side of Mathematics*, Mathematical Association of America, Washington DC, 1994.

[H1] Hadamard, J. *The Psychology of Invention in the Mathematical Field*, Dover, New York, 1954.

[H2] Halmos, P. Mathematics as a creative art, *The American Scientist* **56,** 375–389 (1968).

[H3] Halmos, P. *Selecta: Expository Articles*, Springer-Verlag, New York, 1983.

[H4] Halmos, P. *I Want to be a Mathematician: An Automathography*, Springer-Verlag, New York, 1985.

[H5] Hardy, G.H. *A Mathematician's Apology*, Cambridge, New York, 1967.

[H6] Henkin, L. Are logic and mathematics identical? *Science* **138,** 788–794 (1962).

[H7] Hofstadter, D. *Gödel, Escher, Bach*, Basic Books, New York, 1979.

[H8] Honsberger, R. *Mathematical Gems*, Mathematical Association of America, Washington DC, 1973.

[I1] Infeld, L. *Whom the Gods Love*, Whittlesey House, New York, 1948.

[J1] Joseph, G. *The Crest of the Peacock: Non-European Roots of Mathematics*, Penguin, London, 1991.

[K1] Kac, M. *Enigmas of Chance: An Autobiography*, U. of California Press, Berkeley, 1987.

[K2] Kac, M. and Ulam, S. *Mathematics and Logic: Retrospect and Prospects*, Praeger, New York, 1968.

[K3] Kanigel, R. *The Man Who Knew Infinity: A Life of the Genius Ramanujan*, Macmillan, New York, 1991.

[K4] Katz, V. *A History of Mathematics: An Introduction*, Harper Collins, New York, 1993.

[K5] Kline, M. *Mathematical Thought from Ancient to Modern Times*, Oxford, New York, 1972.

[K6] Körner, Stephan, *The Philosophy of Mathematics: An Introduction*, Harper Torchbooks, New York, 1962.

[K7] Kovalevskaya, S. *A Russian Childhood*, ed. by B. Stillman, Springer, New York, 1978.

[L1] Lang, S. *The Beauty of Doing Mathematics*, Springer-Verlag, New York, 1985.

[L2] Le Lionnais, F., ed. *Great Currents of Mathematical Thought*, 2 vols., Dover, New York, 1971.

[L3] Littlewood, J.E. *A Mathematician's Miscellany*, Methuen, London, 1953.

[M1] Maddy, P. *Realism in Mathematics*, Oxford, New York, 1990.

[M2] Mahoney, M. *The Mathematical Career of Pierre de Fermat, 1601–1665*, 2nd ed., Princeton University Press, Princeton, 1994.

[M3] Menninger, K. *Number Words and Number Symbols: A Cultural History of Numbers*, M.I.T. Press, Cambridge, 1969.

[M4] Meschkowski, H. *Ways of Thought of Great Mathematicians*, Holden-Day, San Francisco, 1964.

[M5] Meschkowski, H. *Evolution of Mathematical Thought*, Holden-Day, San Francisco, 1965.

[M6] Minsky, M. *The Society of Mind*, Simon and Schuster, New York, 1985.

[M7] Moritz, R. *Memorabilia Mathematica: The Philomath's Quotation Book*, Mathematical Association of America, Washington DC, 1993.

[N1] Nagel, E. and Newman, J.R. *Gödel's Proof*, N.Y.U. Press, New York, 1958.

[N2] Nelson, R. *Proofs Without Words: Exercises in Visual Thinking*, Mathematical Association of America, Washington DC, 1993.

[N3] Neugebauer, O. *The Exact Sciences in Antiquity*, Dover, New York, 1969.

[N4] Newman, J.R., ed. *The World of Mathematics*, 4 vols., Simon and Schuster, New York, 1956.

[O1] Ore, O. *Niels Henrik Abel: Mathematician Extraordinary*, U. Minn. Press, Minneapolis, 1957.

[O2] Osen, L. *Women in Mathematics*, M.I.T. Press, Cambridge, 1974.

[O3] Osserman, R. *Poetry of the Universe: A Mathematical Exploration of the Cosmos*, Anchor, New York, 1995.

[P1] Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford, New York, 1989.

[P2] Penrose, R. *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford, New York, 1996.

[P3] Poincaré, H. *The Foundations of Science*, The Science Press, New York, 1929.

[P4] Pólya, G. *Mathematical Discovery*, Wiley, New York, 1962.

[P5] Priestley, W. Mathematics and poetry: How wide the gap? *Mathematical Intelligencer* **12,** 14–19 (1990).

[R1] Reid, C. *Hilbert*, Springer-Verlag, New York, 1970.

[R2] Reid, C. *Courant in Göttingen and New York*, Springer-Verlag, New York, 1976.

[R3] Reid, C. *Julia: A Life in Mathematics*, Mathematical Association of America, Washington DC, 1996.

[R4] Rothstein, E. *Emblems of Mind: The Inner Life of Music and Mathematics*, Times Books, New York, 1995.

[R5] Russell, B. The retreat from Pythagoras, *The Basic Writings of Bertrand Russell*, Simon and Schuster, New York, 252–256 (1961).

[R6] Russell, B. *The Autobiography of Bertrand Russell*, 3 vols., Atlantic-Little Brown, Boston, 1967.

[S1] Schmalz, R. *Out of the Mouths of Mathematicians: A Quotation Book for Philomaths*, Mathematical Association of America, Washington DC, 1993.

[S2] Simmons, G. *Calculus Gems*, McGraw-Hill, New York, 1992.

[S3] Smith, D.E. *A Source Book in Mathematics*, 2 vols., Dover, New York, 1959.

[S4] Steen, L. New models of the real-number line, *Scientific American* **225,** 92–99 (1971).

[S5] Steinhaus, H. *Mathematical Snapshots*, Oxford, New York, 1969.

[S6] Stolzenberg, G. Review of Errett Bishop's *Foundations of Constructive Analysis, Bulletin of the American Mathematical Society* **76,** 301–323 (1970).

[S7] Struik, D.J. *A Concise History of Mathematics*, Dover, New York, 1967.

[S8] Struik, D.J. *A Source Book in Mathematics, 1200–1800*, Harvard, Cambridge, 1969.

[S9] Swetz, F.J., ed. *From Five Fingers to Infinity*, Open Court, Chicago, 1994.

[T1] Toeplitz, O. *The Calculus: A Genetic Approach*, University of Chicago Press, Chicago, 1963.

[U1] Ulam, S. *Adventures of a Mathematician*, Scribners, New York, 1983.

[V1] Van der Waerden, B.L. *Science Awakening*, Oxford, New York, 1961.

[V2] Vilenkin, N. *In Search of Infinity*, trans. by Abe Shenitzer, Birkhäuser, Boston, 1995.

[W1] Weidman, D. Emotional perils of mathematics, *Science* **149**, 1048 (1965).

[W2] Weil, A. *The Apprenticeship of a Mathematician*, trans. by Jennifer Gage, Birkhäuser, Boston, 1992.

[W3] Westfall, R. *Never at Rest: A Biography of Isaac Newton*, Cambridge, New York, 1980.

[W4] Weyl, H. *Philosophy of Mathematics and Natural Science*, Princeton University Press, Princeton, 1949.

[W5] Weyl, H. *Symmetry*, Princeton University Press, Princeton, 1952.

[W6] White, A., ed. *Essays in Humanistic Mathematics*, MAA Notes #32, Mathematical Association of America, Washington DC, 1993.

[W7] Whitehead, A. *Science and the Modern World*, Macmillan, New York, 1925.

[W8] Wiener, N. *Ex-prodigy*, M.I.T. Press, Cambridge, 1964.

[W9] Wiener, N. *I am a Mathematician*, M.I.T. Press, Cambridge, 1964.

[W10] Williams, J.D. *The Compleat Strategyst*, McGraw-Hill, New York, 1954.

APPENDIX 2

Sums and Their Limits

This appendix is intended to supplement the discussion of Eudoxus' method given in Chapter 7 by presenting proofs of several summation formulas and by offering some more examples of summation techniques. What follows may be read immediately after Section 6 of Chapter 7.

Let us begin by discussing a problem which may be no further away than the nearest supermarket. We shall solve it two ways. The first solution makes no use of summation techniques.

Problem A

Oranges are stacked in the form of a pyramid whose base is rectangular, with 6 oranges along one side and 10 oranges along the other. How many oranges are in the pyramid?

Solution 1

The bottom level, of dimensions 6 by 10, has 4 more oranges along one side than along the other. As any experienced stacker of oranges knows, it follows that every level will have 4 more oranges in one dimension than in the other. The size of the top level has to be 1 orange by 5 oranges, the next level must be of size 2 by 6, followed by a level of size 3 by 7, and so on. Adding the oranges in each level—beginning at the top level—we see that the total number of oranges is given by

$$1 \cdot 5 + 2 \cdot 6 + 3 \cdot 7 + 4 \cdot 8 + 5 \cdot 9 + 6 \cdot 10 = 175. \qquad \square$$

Solution 2

The k-th level (counting from the top level down) will have k oranges along one side and $k+4$ oranges along the other. Hence the k-th level will contain $k(k+4) = k^2 + 4k$ oranges. There are obviously 6 levels in all, so the total number of oranges is equal to

$$\sum_{k=1}^{6} \text{oranges on } k\text{-th level} = \sum_{k=1}^{6}(k^2+4k) = \sum_{k=1}^{6} k^2 + 4\sum_{k=1}^{6} k. \qquad (*)$$

There are easy formulas for $\sum k$, for $\sum k^2$, and for $\sum k^3$, given as follows:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \qquad (1)$$

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

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

(We shall prove these formulas shortly.) When $n = 6$, formulas (1) and (2) become

$$\sum_{k=1}^{6} k = \frac{6(7)}{2} = 21; \quad \sum_{k=1}^{6} k^2 = \frac{6(7)(13)}{6} = 91.$$

These, when put together with equation (*) above, show that the total number of oranges is equal to $91 + 4(21) = 175$. □

In Problem A we needed to add together only six quantities. With such a small number of summands the use of summation techniques saves little time. Solution 2 will probably consume as much time as Solution 1. With a large number of summands, however, the use of summation techniques is almost indispensable. The reader will find it helpful to memorize formulas (1), (2), and (3).

Problem A$^+$

Baseballs are stacked in the form of a huge pyramid, with a rectangular base of 60 balls by 50 balls. How many baseballs does the pyramid contain?

Solution 1

Every level will have 10 more balls in one dimension than in the other, and there will be 50 levels in all. Counting from the top level down, we see that the total number of balls is given by

$$1 \cdot 11 + 2 \cdot 12 + 3 \cdot 13 + \cdots + 50 \cdot 60 = ??!$$ □

Solution 2

The k-th level from the top will have k balls along one side and $k+10$ balls along the other. Hence the k-th level contains $k(k+10) = k^2 + 10k$ balls, and the total number of balls in the entire pyramid is given by

$$\begin{aligned}\sum_{k=1}^{50}(k^2+10k) &= \sum_{k=1}^{50}k^2 + 10\cdot\sum_{k=1}^{50}k \\ &= \frac{50(51)(101)}{6} + 10\,\frac{50\cdot 51}{2} \qquad \text{[by (1) and (2)]} \\ &= 55{,}675. \end{aligned}$$

□

§1. Collapsing Sums; Proofs of Formulas (1), (2) and (3)

Here are some examples of a simple but important type of summation. The sum depends only upon the first and last terms, since the intermediate terms cancel. The sum "collapses", making it quite easy to add up. The following equalities are obvious.

$$\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\cdots+\left(\frac{1}{n}-\frac{1}{n+1}\right) = 1-\frac{1}{n+1}, \tag{4}$$

$$[1^2-0^2]+[2^2-1^2]+[3^2-2^2]+\cdots+[n^2-(n-1)^2] = n^2, \tag{5}$$

$$[1^3-0^3]+[2^3-1^3]+[3^3-2^3]+\cdots+[n^3-(n-1)^3] = n^3. \tag{6}$$

In summation notation, formulas (4), (5), and (6) are expressed as follows.

$$\sum_{k=1}^{n}\left(\frac{1}{k}-\frac{1}{k+1}\right) = 1-\frac{1}{n+1}; \tag{4'}$$

$$\sum_{k=1}^{n}(k^2-(k-1)^2) = n^2; \tag{5'}$$

$$\sum_{k=1}^{n}(k^3-(k-1)^3) = n^3. \tag{6'}$$

What is this good for? It seems too obvious to lead to anything interesting. Yet interesting results are immediately at hand. Since $(1/k) - (1/(k+1)) = 1/(k^2+k)$, the obvious formula (4') immediately yields the interesting summation formula

$$\sum_{k=1}^{n} 1/(k^2+k) = n/(n+1),$$

a result which may not be obvious. And the obvious formulas (5') and (6') lead immediately to proofs of formulas (1) and (2).

Proof of Formula (1)
First note that

$$k^2 - (k-1)^2 = k^2 - (k^2 - 2k + 1) = 2k - 1,$$

so that formula (5') becomes

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

Therefore,

$$2 \cdot \sum_{k=1}^{n} k - \sum_{k=1}^{n} 1 = n^2,$$

$$2 \cdot \sum_{k=1}^{n} k - n = n^2,$$

$$2 \cdot \sum_{k=1}^{n} k = n^2 + n,$$

$$\sum_{k=1}^{n} k = \frac{n^2+n}{2}. \qquad \square$$

Proof of Formula (2)
First note that

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

so that formula (6') becomes

$$\sum_{k=1}^{n}(3k^2 - 3k + 1) = n^3.$$

Therefore,

$$3 \cdot \sum_{k=1}^{n} k^2 - 3 \cdot \sum_{k=1}^{n} k + \sum_{k=1}^{n} 1 = n^3,$$

$$3 \cdot \sum_{k=1}^{n} k^2 - 3\,\frac{n(n+1)}{2} + n = n^3.$$

We want to derive formula (2) by solving this equation for $\sum k^2$. This is easier to do if first we multiply through by 2:

$$6 \cdot \sum_{k=1}^{n} k^2 - 3n(n+1) + 2n = 2n^3.$$

Therefore (the reader is asked to supply the missing steps),

$$\begin{aligned} 6 \cdot \sum_{k=1}^{n} k^2 &= 2n^3 + 3n^2 + n \\ &= n(2n^2 + 3n + 1) \\ &= n(n+1)(2n+1), \end{aligned}$$

and formula (2) is obtained upon dividing by 6. □

Proof of Formula (3)

Since the idea of these proofs should be familiar by now, only the main steps are given. The reader is asked to fill in the details. Beginning with the collapsing sum

$$\sum_{k=1}^{n} (k^4 - (k-1)^4) = n^4,$$

and noting that $k^4 - (k-1)^4 = 4k^3 - 6k^2 + 4k - 1$, we get

$$4 \cdot \sum k^3 - 6 \cdot \sum k^2 + 4 \cdot \sum k - \sum 1 = n^4,$$

where all the summations run from $k = 1$ to $k = n$. Using formulas already derived for $\sum k^2$, $\sum k$, and $\sum 1$, we obtain the equation

$$4 \cdot \sum k^3 - (2n^3 + 3n^2 + n) + (2n^2 + 2n) - n = n^4.$$

Therefore,

$$\begin{aligned} 4 \cdot \sum k^3 &= n^4 + 2n^3 + n^2 \\ &= n^2(n^2 + 2n + 1) \\ &= n^2(n+1)^2, \end{aligned}$$

and formula (3) is obtained upon dividing by 4. □

The summation formulas (1), (2), and (3) were known to the Greeks, but the proofs presented here are modern. The modern approach, using summation notation, has the advantage of applying equally well to the determination of formulas for $\sum k^4$, $\sum k^5$, etc.—sums which the Greeks apparently did not consider. The determination of such formulas is left to the reader as an exercise.

§2. Integrals of Quadratics and Cubics

Many examples of Eudoxus' method of calculating integrals are given in Chapter 7, but most of them deal with relatively simple linear functions. Here are some slightly more complicated applications of the method.

EXAMPLE 1
Calculate the integral $\int_0^1 x^2\,dx$, directly from its definition as a limit of sums.

Solution
We are asked to calculate $\int_a^b f(x)\,dx$, where $a = 0$, $b = 1$, and $f(x) = x^2$. Here we have $\Delta x = 1/n$, $x_k = k/n$, and $f(x_k) = k^2/n^2$. An approximating sum for the desired integral is then given by

$$\begin{aligned} S_n &= \sum_{k=1}^{n} f(x_k)\Delta x \\ &= \sum_{k=1}^{n} \left(\frac{k^2}{n^2}\right)\left(\frac{1}{n}\right) \\ &= \frac{1}{n^3}\sum_{k=1}^{n} k^2 \\ &= \frac{n(n+1)(2n+1)}{6n^3} \qquad \text{[by (2)]} \\ &= \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right). \end{aligned}$$

Therefore,

$$\int_0^1 x^2\,dx = \text{Limit } S_n = \text{Limit } \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) = \frac{1}{6}(1)(2) = \frac{1}{3}. \qquad \square$$

EXAMPLE 2
Calculate the integral $\int_0^\pi ax^2\,dx$ directly from its definition as a limit of sums.

Solution
This is only a slight modification of the preceding example. Here we have $\Delta x = \pi/n$, $x_k = k\pi/n$, and $f(x_k) = ak^2\pi^2/n^2$.

$$S_n = \sum_{k=1}^{n}\left(\frac{ak^2\pi^2}{n^2}\right)\left(\frac{\pi}{n}\right) = \frac{a\pi^3}{n^3}\sum_{k=1}^{n} k^2 = \frac{a\pi^3 n(n+1)(2n+1)}{6n^3}.$$

Therefore,

$$\int_0^\pi ax^2\,dx = \text{Limit } S_n = \text{Limit } \frac{a\pi^3}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) = \frac{a\pi^3}{3}. \qquad \square$$

EXAMPLE 3
Calculate the integral $\int_0^t ax^2\,dx$ by Eudoxus' method.

Solution
This is treated just like the preceding example except that we have t in place of π. It is therefore obvious that the bottom line will read

$$\int_0^t ax^2\,dx = \text{Limit } S_n = \text{Limit } \frac{at^3}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) = \frac{at^3}{3}. \qquad \square$$

EXAMPLE 4
Calculate the integral $\int_0^1 x^3\,dx$ by Eudoxus' method.

Solution
Here we have

$$\begin{aligned} S_n &= \sum_{k=1}^{n}\left(\frac{k^3}{n^3}\right)\left(\frac{1}{n}\right) \\ &= \frac{1}{n^4}\sum_{k=1}^{n} k^3 \\ &= \frac{n^2(n+1)^2}{4n^4} \text{ [by (3)]} \\ &= \frac{1}{4}\left(1+\frac{1}{n}\right)^2. \end{aligned}$$

Therefore,

$$\int_0^1 x^3\,dx = \text{Limit } S_n = \text{Limit } \frac{1}{4}\left(1+\frac{1}{n}\right)^2 = \frac{1}{4}(1) = \frac{1}{4}. \qquad \square$$

EXAMPLE 5
Calculate the integral $\int_a^b x^3\,dx$ by Eudoxus' method.

Solution
Here are the main steps. The details are left to the reader. To save space the index k is suppressed in the summations below. The summations are understood to run from $k = 1$ to $k = n$.

Since $x_k = a + k\,\Delta x$, we have

$$\begin{aligned} S_n &= \sum x_k^3\,\Delta x \\ &= \sum (a + k\,\Delta x)^3\,\Delta x \\ &= \sum [a^3 + 3a^2k(\Delta x) + 3ak^2(\Delta x)^2 + k^3(\Delta x)^3]\Delta x \\ &= a^3(\Delta x)\sum 1 + 3a^2(\Delta x)^2 \sum k + 3a(\Delta x)^3 \sum k^2 + (\Delta)^4 \sum k^3. \end{aligned}$$

Using the fact that $\Delta x = (b-a)/n$ and using the summation formulas for $\sum 1$, $\sum k$, $\sum k^2$, and $\sum k^3$, we get

$$\begin{aligned} S_n = a^3(b-a) &+ \frac{3}{2}a^2(b-a)^2\left(1+\frac{1}{n}\right) + \frac{1}{2}a(b-a)^3\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) \\ &+ \frac{1}{4}(b-a)^4\left(1+\frac{1}{n}\right)^2. \end{aligned}$$

Taking the limit of S_n as n increases without bound, we see that

$$\begin{aligned} \int_a^b x^3\,dx &= a^3(b-a) + \frac{3}{2}a^2(b-a)^2 + a(b-a)^3 + \frac{1}{4}(b-a)^4 \\ &= \frac{b^4}{4} - \frac{a^4}{4}. \end{aligned}$$

□

§3. Geometric Series and Applications

The equation $(1+x)(1-x) = 1 - x^2$ is such a simple algebraic identity that the most interesting thing about it is rarely noticed. The interesting thing is that it comes from a collapsing sum.

$$(1+x)(1-x) = (1)(1-x) + (x)(1-x) = (1-x) + (x - x^2) = 1 - x^2.$$

This is about the simplest possible example of a collapsing sum. We should ask whether this example generalizes readily and whether the generalization is even more interesting. The answer is *yes* to both questions.

The immediate generalization is this:

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

And the far-reaching generalization is one of the most important identities in mathematics:

$$\begin{aligned} (1 + x + x^2 + \cdots + x^n)(1-x) &= (1-x) + (x-x^2) + \cdots + (x^n - x^{n+1}) \\ &= 1 - x^{n+1}. \end{aligned}$$

The identity is important because from it we get the following summation formula (by dividing both sides of the identity by $1 - x$):

$$1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x}, \quad \text{if } x \neq 1. \tag{7}$$

The series on the left is called a *geometric* series and its sum is given in equation (7). The result just obtained is useful enough to be called a theorem. (The "geometric" nature of $1, x, x^2, \ldots$ is revealed in problem 4 of Chapter 3.)

Theorem on Geometric Series

For a geometric series the following summation formula is valid, provided $x \neq 1$:

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

(*where* x^0 *is understood to be* 1).

The reader should note the difference between the type of series now being considered and the type that was considered in Section 1. In Section 1 we found, for example, that

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6};$$

whereas by the theorem just proved (with $x = 2$) we have

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

The reader should be careful not to confuse $\sum 2^k$ with $\sum k^2$. One is a geometric series while the other is not. Note also that the formula for the geometric series above is for the sum beginning with index $k = 0$. By subtracting 1 from both sides of equation (7) we get an analogous formula where the index k begins at 1:

$$\sum_{k=1}^{n} x^k = \frac{x - x^{n+1}}{1 - x}, \quad x \neq 1.$$

EXAMPLE 6

Evaluate the sum

$$1 + \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \cdots + \left(\frac{1}{4}\right)^n.$$

Solution

The sum in question comes from the geometric series

$$\sum_{k=0}^{n} \left(\frac{1}{4}\right)^k.$$

By (7) its sum is

$$\frac{1-\left(\frac{1}{4}\right)^{n+1}}{1-\frac{1}{4}} = \frac{4}{3} - \frac{1}{3}\left(\frac{1}{4}\right)^n \approx \frac{4}{3}$$

if n is a large positive integer. □

EXAMPLE 7

Consider the function given by $f(x) = 1/(1-x)$. Find an approximation of this function given by nonnegative powers of x.

Solution

Equation (7) says

$$1 + x + x^2 + x^3 + \cdots + x^n = \frac{1}{1-x} - \frac{x^{n+1}}{1-x} \approx \frac{1}{1-x}$$

if $x \approx 0$. Therefore,

$$\frac{1}{1-x} \approx \sum_{k=0}^{n} x^k, \quad \text{if } x \approx 0.$$

This answer should be compared to the result of problem 11(i) at the end of Appendix 5. □

Here is a dazzling application of the use of a geometric series. It is due to Fermat. While the theorem is easy to prove using the fundamental theorem of calculus, Fermat was able to prove it before the fundamental theorem was known. The proof demands careful attention, as some details of it are left to the reader.

Theorem (Fermat)

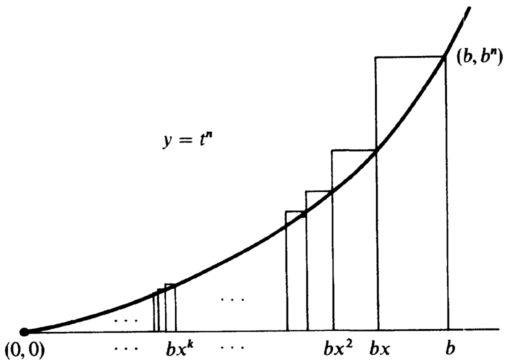
Let A be the area beneath the graph of the curve $y = t^n$ (where n is a positive integer) between $t = 0$ and $t = b$. Then $A = b^{n+1}/(n+1)$.

Proof

Let x be a positive number just less than 1 and consider the (infinite) sequence of numbers

$$b, bx, bx^2, bx^3, \ldots, bx^k, \ldots$$

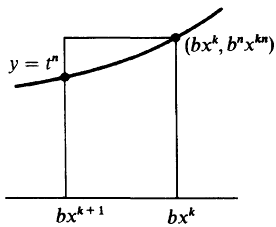
which subdivide the interval from $t = 0$ to $t = b$ into infinitely many subintervals. For fixed x, let A_x be the area beneath the staircase built upon these subintervals.



As $x \to 1^-$ the staircase approximates the curve ever more closely. Hence the area A beneath the curve is given by

$$A = \operatorname*{Limit}_{x \to 1^-} A_x.$$

We shall first calculate A_x. To do this it is convenient to start at the top step and go down, with the top step counted as the 0-th step. Then the k-th step looks like this.



height: $b^n x^{kn}$
width: $b(1 - x)x^k$

The area beneath the k-th step is then $b^{n+1}(1 - x)(x^{n+1})^k$, and the total area A_x (where $x < 1$) is given by summing, beginning at $k = 0$:

$$\begin{aligned}
A_x &= b^{n+1}(1 - x)[1 + x^{n+1} + (x^{n+1})^2 + (x^{n+1})^3 + \cdots] \\
&= b^{n+1}(1 - x)\left[\frac{1}{1 - x^{n+1}}\right] \quad \text{(why?)} \\
&= b^{n+1}\left[\frac{1 - x}{1 - x^{n+1}}\right] \\
&= b^{n+1}/(1 + x + x^2 + \cdots + x^n) \quad \text{[by (7)]}.
\end{aligned}$$

Therefore,

$$A = \operatorname*{Limit}_{x\to 1^-} A_x = \operatorname*{Limit}_{x\to 1^-} b^{n+1}/(1+x+x^2+\cdots+x^n)$$
$$= b^{n+1}/(n+1). \qquad \square$$

Problems (Optional)

1. Baseballs are stacked in the form of a pyramid with a rectangular base of 40 balls by 36 balls. How many balls are in the pyramid?

2. Evaluate each of the following sums.
   (a) $1\cdot 8 + 2\cdot 9 + 3\cdot 10 + 4\cdot 11 + \cdots + 50\cdot 57$.
   (b) $1/(1\cdot 2) + 1/(2\cdot 3) + 1/(3\cdot 4) + \cdots + 1/(99\cdot 100)$.
   (c) $\sum_{k=1}^{100} k^3$.
   (d) $\sum_{k=1}^{100} 3^k$.
   (e) $\sum_{k=0}^{100} (\frac{1}{2})^k$.

3. (a) Prove that $\sum_{k=1}^{n} k^4 = n(n+1)(2n+1)(3n^2+3n-1)/30$.
   (b) Use the result of part (a) to calculate $\int_0^1 x^4\,dx$ by Eudoxus' method.

4. Find a formula for $\sum k^5$ and use it to calculate $\int_0^1 x^5\,dx$. *Caution*: Keep a cool head. This problem can cause nervous breakdowns.

5. Apply Eudoxus' method to calculate each of the following integrals.
   (a) $\int_1^4 (3x^3 - 2x^2)\,dx$.
   (b) $\int_{-2}^{2} (2x^3 - 7x)\,dx$.
   (c) $\int_{-1}^{5} (x^2 - x + 4)\,dx$.

6. *(For those who think they understand infinity)* This book has avoided mentioning the symbol for infinity until now, because the symbol is so easily misunderstood. Test whether you understand it or not, by explaining why it is natural to write

$$\sum_{k=0}^{\infty} x^k = 1/(1-x), \quad \text{if } -1 < x < 1;$$

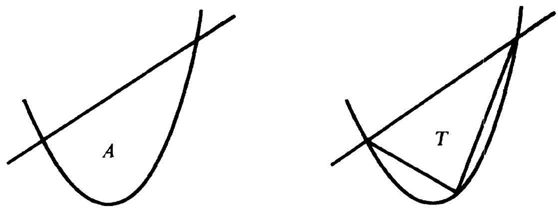
   and yet at the same time to write

$$\sum_{k=0}^{\infty} x^k \neq 1/(1-x), \quad \text{if } x < -1 \text{ or } x > 1.$$

7. (a) When did Fermat die?
   (b) When did Newton discover the fundamental theorem of calculus?
   (c) Look again at the theorem of Fermat's proved in the last section. Give a one-line proof of this theorem by making use of the fundamental theorem of calculus.

8. (*An ambitious project*) Geometric series continue to find many surprising applications, even to the present day. Yet none could be more charming than the application made by Archimedes to effect a quadrature of the parabola. He proved that the area of a segment of a parabola is equal to four-thirds the area of its largest inscribed triangle. (The factor 4/3 comes from the fact that $\sum_{k=0}^{\infty}\left(\frac{1}{4}\right)^k = 4/3$, as seen in Example 6.) In modern terminology we may state this result as follows:

   Theorem (Archimedes)
   Let A be the area enclosed between the graph of a linear function and the graph of a quadratic function, and let T be the area of the largest triangle that can be inscribed in A. Then $A = (4/3)T$.



   Either (a) prove this result through the use of calculus; or (b) look up Archimedes' original proof and write a paper on it, sketching the main points.

9. Prove that for each positive integer n,

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

APPENDIX 3 Archimedes

The mind of Archimedes is modern. Though he was born about 287 B.C., one may expect to have difficulty understanding his work unless one knows something of the developments in mathematics that took place two thousand years later.

In his published papers Archimedes characteristically put together his ideas with such tight logic that the adjective *archimedean* has come to refer to any logical demonstration meeting the very highest standards of rigor. The reader interested in seeing truly archimedean demonstrations is invited to consult T.L. Heath, *The Works of Archimedes*, Cambridge, 1912 (also available in paperback by Dover Publications). This appendix outlines only a few of his ideas, and these are given presentations that may be described as casual if compared with archimedean standards.

§1. Archimedes and the Classical Problems

The three so-called "classical problems of antiquity" are as follows.

(*The Trisection Problem*) Given an angle, devise a method for constructing another angle one-third as large.

(*The Quadrature of the Circle*) Given a circle, devise a method for constructing a square having the same area.

(*The Duplication of the Cube*) Given a cube, devise a method for constructing another cube whose volume is twice as large.

It is probably safe to say that every Greek mathematician worked seriously on at least one of these famous problems. It ought to be surprising, therefore, that Euclid's *Elements* gives no account of them.

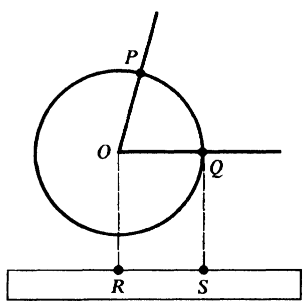
Why did Euclid not discuss these easily stated, natural problems? The reason is simple. *Euclid did not know how to do them. Nor did anyone else.* The construction of the required trisection, quadrature, or duplication eluded the efforts of the greatest mathematicians.

It is important to understand what Euclid meant by a "construction". In Euclidean geometry a construction may use a ruler and a compass, *but nothing else.* And the "ruler" can have no distance markings on it, its use being only as a straightedge to draw straight lines through points already constructed. Using Euclidean constructions, no one was able to solve any of the three problems above.

Archimedes somehow recognized the futility of Euclidean methods of attacking these problems, and reacted in a thoroughly modern way. If traditional theory proves inadequate to handle the type of thing for which it was designed, then something new is needed.

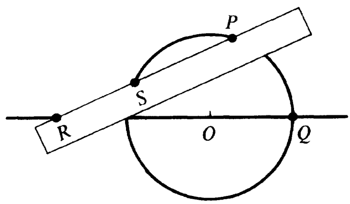
In a sense, it is "obvious" that each of the problems above has a solution. For example, it is obvious that there exists a trisection of a sixty-degree angle. (An angle of twenty degrees, of course, does the trick.) The whole problem is in *constructing* an angle of twenty degrees from a given angle of sixty degrees *through the use of ruler and compass alone.* Archimedes devised the following construction of striking simplicity.

Given an angle, we may construct a circle whose center O lies at the angle's vertex:



On a ruler, or straightedge, mark off two points R and S the distance between which is equal to the radius OQ. Now perform the following trick with the straightedge. Keeping the point R on the line through OQ

and keeping the point S on the circle, manipulate the straightedge until it touches the point P:



The angle PRQ is the required trisection of the given angle POQ. The proof of this is easy and is left to the reader. *Hint.* Begin by drawing the triangles *RSO* and *SOP* and note that they are isosceles triangles.

As Archimedes pointed out, the construction just outlined is done with ruler and compass, but it is not a Euclidean construction. Why is it not a Euclidean construction?

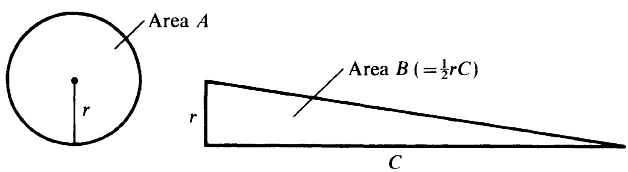
Archimedes' answer to the problem of squaring the circle resulted in one of the most important papers in mathematics. Instead of finding a square of the same area as the circle, Archimedes found a triangle, which is just as good.

Archimedes' Quadrature of the Circle

A circle has the same area as a triangle whose base is equal to the circumference of the circle and whose height is equal to the radius.

Proof

Let A be the area of the circle of radius r and let B be the area of the right triangle with legs of lengths r and C.



There are clearly three logical possibilities:

(a) $A > B$,
(b) $A < B$,
(c) $A = B$.

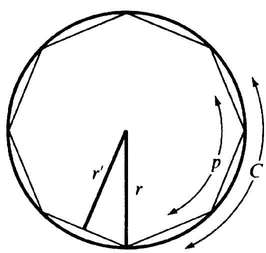
To prove (c) Archimedes used the principle of elimination. He proved that neither possibility (a) nor possibility (b) could be true. This leaves (c) as the only case not eliminated.

Proof that possibility (a) is false

We use the method of *reductio ad absurdum*. Suppose $A > B$. Then the number B is not equal to A but is only an approximation. How can we get a better approximation? That is easy. We can approximate the circle as closely as we please by a regular polygon inscribed inside, and therefore there is such a polygon whose area P is a better approximation to A. Then we have

$$A > P > B. \tag{1}$$

But this leads quickly to a contradiction. Let p denote the length of the polygon's perimeter and let r' denote the polygon's "radius" (see the figure below).



It follows that

$$\begin{aligned} P &= \tfrac{1}{2} r'p \quad \text{(by problem 14 of Chapter 3)} \\ &< \tfrac{1}{2} rC \quad \text{(since } r' < r \text{ and } p < C\text{)}. \end{aligned}$$

Therefore, $P < \frac{1}{2}rC$. But $\frac{1}{2}rC = B$, so

$$P < B. \tag{2}$$

Statements (1) and (2) contradict each other. This contradiction arises from the supposition that $A > B$. This supposition is therefore false. Possibility (a) has been proved false.

Proof that possibility (b) is false

This proof follows closely the lines of the proof above, except that a *circumscribed* polygon approximating the circle is brought into play. Suppose that $A < B$. Then there is a circumscribed regular polygon satisfying condition (1) *but with the inequalities reversed*. This leads quickly to a contradiction (the demonstration of which is left to the reader), which shows that the supposition $A < B$ must be false.

By "double elimination" it follows that $A = B$. □

In the theorem above, Archimedes tells us how to construct a triangle equal in area to a given circle. However, the construction given is not a Euclidean construction. (*Why not?*)

In fact, this theorem occupies only a small (though essential) place in Archimedes' celebrated paper on the quadrature of the circle. The main body of the paper is concerned with estimating the numerical value of the ratio of the circumference of a circle to its diameter. (Today this ratio is always denoted by the Greek letter π—the first letter in the Greek word for *perimeter*—but this notation was popularized only in the eighteenth century by Euler.) Archimedes found this ratio to be between $3\frac{10}{71}$ and $3\frac{1}{7}$, as we have seen in Section 3 of Chapter 3.

Actually, there existed before Archimedes other successful (but not Euclidean) methods of trisecting angles and squaring circles. In fact, a single curve—aptly called the quadratrix—could be employed in solving both problems. Hippias and Dinostratus had shown how to do this, but at the expense of a considerable departure from traditional methods.

Like Hippias and Dinostratus, Archimedes did not hesitate to break with tradition when tradition prevented him from attending his calling. But when he broke, he evidently did not like to go further away than he had to. When Archimedes found Euclidean constructions inadequate he tried to develop adequate constructions that were almost Euclidean. Happily he found them, as we have seen above, even though he also found a single curve—the spiral—that could be used to do the same job as the quadratrix.

In the case of the third of the classical problems, the duplication of the cube, Archimedes offered no new solution. Solutions (using non-Euclidean constructions) had already been given by Archytas, Eudoxus, Eratosthenes, Apollonius, and others. In modern terms the problem can be stated as follows. Given a cube whose sides have length s (yielding a volume of s^3), construct a cube with sides x whose volume is $2s^3$, or twice as large. This means one must construct a length x satisfying the equation

$$x^3 = 2s^3, \tag{3}$$

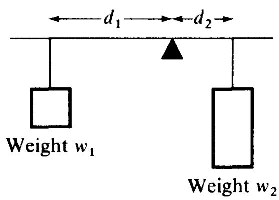
where s is given. There are many (non-Euclidean) ways of doing this.

Archimedes did something harder. Instead of posing for himself the problem of solving the simple cubic equation (3), Archimedes tackled the analysis of cubic equations in general. Since the Greeks couched all their algebra in geometric terms, and since they did not consider negative numbers, it would not be said today that Archimedes gave the first complete analysis of the general cubic equation. But if it were said, it would not be far wrong.

As we have seen, Archimedes failed in his attempts to solve the three classical problems of antiquity through the exclusive use of Euclidean methods. These are among his few failures, but we now know that they are nothing to be ashamed of. No Euclidean method, no matter how ingenious, will solve any of these problems. The inadequacy of Euclidean methods in this regard was conclusively demonstrated in the nineteenth century. This demonstration may be found in many undergraduate texts on modern algebra, but it is beyond the scope of the present text.

§2. Archimedes' Method

When does a seesaw balance? *Answer*: when the moments on each side are equal, the *moment* being defined as the product of a weight with its distance from the lever's fulcrum. This principle is a cornerstone of *statics*, a branch of physics that studies conditions of equilibrium.

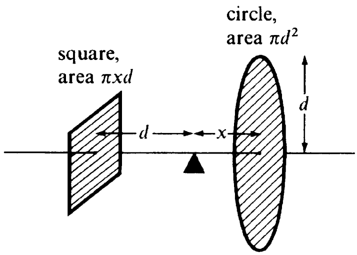


Law of the lever:
The lever is in equilibrium if $w_1 d_1 = w_2 d_2$

This principle was known to the Greeks before Archimedes was born. Yet Archimedes was the one to see how this tool could be used to open the way toward mathematical physics. He postulated simple axioms about statics, from which he proceeded to deduce the law of the lever and much, much more. He began investigating, with great success, the problem of finding the centroid, or center of gravity, of a solid figure. When he incorporated into all this his famous *principle of buoyancy* (the upward force on an object submerged in water is exactly equal to the weight of the water displaced), he invented the science of *hydrostatics*. Though Archimedes is said to have deplored "the whole trade of engineering", he could not have failed to know that his work would have practical applications to engineering. Everything from the design of more efficient compound pulleys to the design of more stable floating vessels is connected with it.

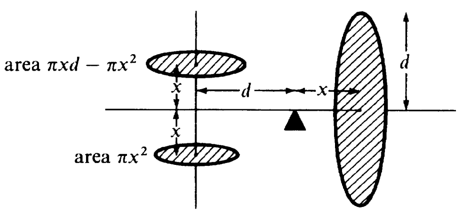
As impressive as all this might be, there is yet another application of the law of the lever that is even more surprising. Archimedes perceived—in what must be described as a flash of genius—that the lever can be brought into play with problems of pure mathematics. While a physical principle cannot, of course, be admitted into an archimedean demonstration of pure mathematics, a physical principle (or anything else, for that matter) can certainly be used to make guesses. And an archimedean guess precedes an archimedean demonstration.

The law of the lever applies only to the physical world, of course. But it occurred to Archimedes that there ought to be an analogous law in the realm of geometry! What should such a law say? Archimedes began to play with the idea of balancing geometric objects against each other. He reasoned, for example, that equilibrium would hold in the following situation.

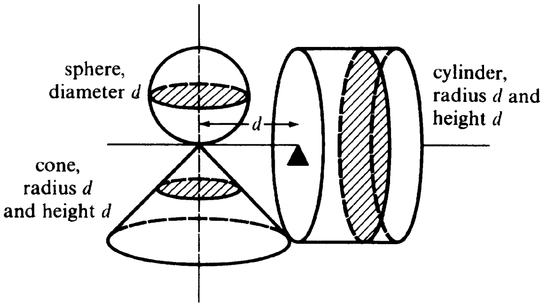


Since $\pi xd \cdot d = \pi d^2 \cdot x$, equilibrium obtains

Since a body behaves as if all its weight is concentrated at its center of gravity, the configuration below is essentially the same as the one above, and is therefore in equilibrium. (The combined area of the two circles below is equal to the area of the square in the picture above.)



Reasoning somewhat as Cavalieri was to do centuries later (see Chapter 7), Archimedes concluded that we must have equilibrium in the following figure—for each vertical slice through the cylinder is exactly balanced by a corresponding pair of horizontal slices in the sphere and cone.



What is all this good for? Archimedes used it to guess the volume of a sphere. The volumes of cones and cylinders were well known already. Since the configuration above balances, the law of the lever says that

$$(\text{volume of sphere and cone}) \cdot d = (\text{volume of cylinder}) \cdot \frac{d}{2}.$$

(The length $d/2$ is the moment arm of the center of gravity of the cylinder.) Dividing by d in this equation shows that

$$\text{volume of sphere} + \text{volume of cone} = \tfrac{1}{2}(\text{volume of cylinder}).$$

Let V denote the volume of the sphere of diameter d. Then by known formulas for the volumes of cones and cylinders the above equation becomes

$$V + \frac{1}{3}\pi d^3 = \frac{1}{2}\pi d^3,$$

from which it follows that

$$V = \frac{1}{6}\pi d^3 = \frac{4}{3}\pi r^3, \tag{4}$$

where r $(=\frac{1}{2}d)$ is the radius. The volume of a sphere is then given by equation (4).

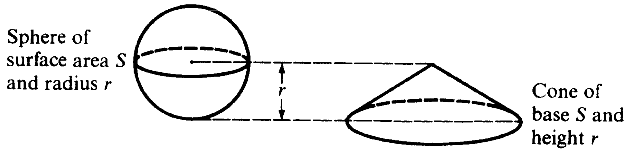
This is one of several extraordinary balancing acts that Archimedes was able to perform. They are all examples of his so-called "method", described in his famous letter to Eratosthenes. He emphasized that his method was used only to make guesses at what seemed to be plausible. Once he knew the likely truth he could prove it by rigorous means, such as the principle of double elimination illustrated in Section 1.

Let us look at just one more example of what a genius can see. In his letter to Eratosthenes, Archimedes says

> . . . judging from the fact that any circle is equal to a triangle with base equal to the circumference and height equal to the radius of the circle, I apprehended that, in like manner, any sphere is equal to a cone with base equal to the surface of the sphere and height equal to the radius.*

* From *The Method of Archimedes*, pp. 20–21 of the supplement to *The Works of Archimedes*, edited by T. L. Heath, Cambridge, 1912.

Archimedes has guessed this *cubature* of the sphere:



If this guess is correct it follows that

$$\frac{4}{3}\pi r^3 = \frac{1}{3}Sr$$

from the formulas for the volumes of spheres and cones. Solving this equation for the surface area S yields

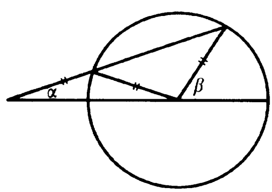
$$S = 4\pi r^2. \tag{5}$$

In this way Archimedes guessed the correct formula (5) for the surface area of a sphere. The surface area is exactly four times as large as any great circle in the sphere, according to Archimedes. Having guessed the right answer he then proved it by completely different means, giving a rigorous demonstration to meet his standards.

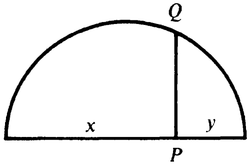
As Archimedes once noted on a different occasion, a light touch—if properly applied—can move the earth.

Problems (Optional)

1. Prove that Archimedes' trisection technique actually works, as follows: in the figure of Section 1 illustrating this technique, let $\alpha =$ angle PRQ, let $\beta =$ angle POQ, and prove that $\beta = 3\alpha$.



2. (a) Given lengths x and y, outline a Euclidean construction that produces the length $\sqrt{xy}$. *Hint.* Ponder the figure below. What is the length PQ?



Semicircle with diameter $x + y$

(b) Devise a way of effecting a quadrature of a rectangle, using only Euclidean methods. That is, given a rectangle of sides x and y, construct a square having the same area as the rectangle. *Hint.* Use the result of part (a).

(c) Devise a way of effecting a quadrature of a triangle, using only Euclidean methods. *Hint.* First construct by Euclidean methods a rectangle having the same area as the given triangle. Then use the result of part (b).

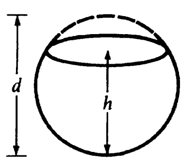
3. Many references are listed at the end of Appendix 1. Several of them discuss the quadratrix of Hippias and Dinostratus.
   (a) Find a book that discusses the quadratrix (also called the trisectrix).
   (b) Be prepared to illustrate in class how the quadratrix can be employed to trisect an angle and to square a circle.

4. Find a book that discusses Archimedes' spiral and be prepared to illustrate in class how the spiral can be employed to trisect an angle and to square a circle.

5. It is impossible to duplicate the cube using only Euclidean constructions. Find a way to do it that uses constructions that are "almost" Euclidean. If you need to, find and use a book that discusses the Greek attempts to solve the *Delian problem* (as the problem of duplicating the cube was known).

6. We have already noted (in exercise 10.4 of Chapter 7) that Archimedes found the ratio of the volume of a cylinder to the volume of an inscribed sphere. Archimedes also found the ratio of the surface area of the cylinder (including its base and top) to the surface area of the inscribed sphere. What is this ratio?

7. Archimedes' "balancing act" described in Section 2 works not only for a sphere but also (as Archimedes pointed out) for a segment of a sphere. Only a slight modification of the method described in Section 2 is needed to determine the volume of this segment of a sphere:



A segment with height *h* of a sphere with diameter *d*

Guess what this volume is by using the method of Archimedes. Then verify your result by calculus. *Hint.* Let V denote the volume of the segment of height h pictured above. By Archimedes' method of balancing, derive the relation

$$\left(V + \frac{1}{3}\pi h^3\right)d = (\pi d^2 h)\frac{h}{2}.$$

Then solve for V. Check your answer to see that it is the same as the integral

$$\int_{-r}^{-r+h} \pi(r^2 - x^2)\,dx,$$

which gives V as the volume of a solid of revolution.

8. The proof given in Section 1 of Archimedes' quadrature of the circle is left unfinished. The proof that possibility (b) is false is left to the reader. Write out this proof in detail.

9. Write a short essay either defending or attacking Voltaire's assertion that Archimedes is superior in imagination to Homer.

10. *(Only for those who have studied trigonometry)* Familiarity with trigonometric identities involving double-angle (or half-angle) formulas makes it easy to prove that the pattern observed by the Scottish mathematician James Gregory (1638–1675) must continue in the successive rows of the table begun in exercise 3.13 of Chapter 3. There is no evidence that Archimedes noticed this pattern, even though his work, done before our modern sine and cosine functions were defined, shows an ease in using formulas equivalent to our double-angle formulas for the sine and cosine.
    (a) Using familiar trigonometric identities, show that $2\tan(\theta/2)$ is the harmonic mean of $\sin(\theta)$ and $\tan(\theta)$ if θ is an acute angle.
    (b) Show that for any acute angle θ, $2\sin(\theta/2)$ is the geometric mean of $\sin(\theta)$ and $2\tan(\theta/2)$. *Hint.* Show that $4\sin^2(\theta/2) = 2\tan(\theta/2)\sin(\theta)$.
    (c) Using an appropriate figure, show that a regular polygon with n sides inscribed in a circle of radius r has a perimeter p given by $p = 2nr\sin(\theta)$, where θ is $180/n$ degrees. Then show that it has a perimeter P given by $P = 2nr\tan(\theta)$ if the polygon is circumscribed instead of inscribed.
    (d) Deduce from part (c) that the table below must produce exactly the same numbers as the table in exercise 3.13 of Chapter 3.

n	$n\sin(180/n)$	$n\tan(180/n)$
6	$6\sin(30°)$	$6\tan(30°)$
12	$12\sin(15°)$	$12\tan(15°)$
$\vdots$	$\vdots$	$\vdots$
n	$n\sin(\theta)$	$n\tan(\theta)$
$2n$	$2n\sin(\theta/2)$	$2n\tan(\theta/2)$
$\vdots$	$\vdots$	$\vdots$

    (e) Deduce from parts (a) and (b) that the pattern observed by Gregory must continue.

APPENDIX 4 Clean Writing in Mathematics

Style is like good manners. Its lambent presence is barely noticeable, but its absence is conspicuous. Taken in a broad sense, style can be discerned almost everywhere. One can speak of style (or its absence) in playing basketball, in hosting a dinner party, in presiding over a meeting, in teaching a class, or even—the subject of this appendix—in writing out the solution to a problem in calculus.

In such activities style is characterized by the light touch that draws harmony out of imminent disorder and makes difficult things seem easy. Everyone hates the burden of unnecessary fuss and bother; the grace that comes from easing this burden is the hallmark of style. In any purposeful activity it is style that eases the way.

Style must be natural because it cannot be affected. Affectation will draw attention only to itself, while style would draw attention straightway to the goal at hand.

Style is an outgrowth of education, not a product of it, for style cannot be readily taught or learned. It is acquired almost incidentally, like good manners, by those who want to please. Yet the final aim of education may be the cultivation of a sense for style.

> Finally [out of education], there should grow the most austere of all mental qualities; I mean the sense for style. It is an aesthetic sense, based on admiration for the direct attainment of a foreseen end, simply and without waste. Style in art, style in literature, style in logic, style in practical execution have fundamentally the same aesthetic qualities, namely, attainment and restraint. The love of a subject in itself and for itself, where it is not the sleepy pleasure of pacing a mental quarterdeck, is the love of style as manifested in that study.

> ... Style, in its finest sense, is the last acquirement of the educated mind; it is also the most useful. It pervades the whole being. The administrator with a sense for style hates waste; the engineer with a sense for style economises his material; the artisan with a sense for style prefers good work. Style is the ultimate morality of mind.
>
> But above style, and above knowledge, there is something, a vague shape like fate above the Greek gods. That something is Power. Style is the fashioning of power, the restraining of power. But, after all, the power of attainment of the desired end is fundamental. The first thing to do is to get there. Do not bother about your style, but solve your problem, justify the ways of God to man, administer your province, or do whatever else is set before you.
>
> Where, then, does style help? In this, with style the end is attained without side issues, without raising undesirable inflammations. With style you attain your end and nothing but your end. With style the effect of your activity is calculable, and foresight is the last gift of gods to men.
>
> Alfred North Whitehead*

§1. What to Do After Solving a Problem

Much of this text aims at aiding the reader to acquire the power to solve problems. This appendix is not about solving problems, but about what to do afterwards. Unless a problem is so easy that its answer is virtually apparent at the outset, one should not be content with merely finding the answer. One ought to develop a style of justifying what one believes to be true.

The tone of that justification should be geared to the expectations of those to whom it is addressed. Archimedes aimed at satisfying the highest expectations of his most critical fellow mathematicians.

> [Archimedes' deliberate style] suggests the tactics of some great strategist who foresees everything, eliminates everything not immediately conducive to the execution of his plan, masters every position in its order, and then suddenly (when the very elaboration of the scheme has almost obscured, in the mind of the spectator, its ultimate object) strikes the final blow. Thus we read in Archimedes proposition after proposition the bearing of which is not immediately obvious but which we find infallibly used later on; and we are led by such easy stages that the difficulty of the original problem, as presented at the outset, is scarcely appreciated.
>
> T. L. Heath†

Plutarch must have been right in suggesting that it was only by means of the greatest labor that Archimedes' works appear so unlabored. Archi-

* Presidential address to the Mathematical Association of England, 1916. (Reprinted in *The Aims of Education*, by A. N. Whitehead, Macmillan, 1929, p. 24.)

† Preface to *The Works of Archimedes*. Cambridge. 1897, p. vi.

medes was willing to put forth any amount of time and effort in his work. He was baffled for years, he tells us, before he was able to write some of his papers.

No one (including the instructor) in an introductory calculus course should be expected to meet archimedean standards. But some standards of clean exposition can be developed and maintained. Here are a few rules that might be considered.

(1) Do not slavishly follow any set of rules, even these.
(2) State clearly what information has been given at the outset, and make each succeeding step in your reasoning follow from what has gone before.
(3) If you introduce a symbol such as "x", be sure to indicate what it stands for. Your reader may not guess. Similarly, never introduce the pronoun "it" unless there can be no confusion as to what "it" stands for.
(4) Say exactly what you mean. Do not, for example, put an "equals" sign between unequal quantities.
(5) Write complete sentences and punctuate them correctly. Remember that an equation is (usually) a sentence.
(6) By being as concise and as natural as you can, disguise whatever effort it may have cost you to attain your goal. Be serious but not solemn.
(7) When you have completed your argument and have led your reader to the end, state your full conclusion in a complete sentence. Then stop writing.
(8) Review what you have written and delete anything irrelevant.

All of these rules may be condensed into one short Latin phrase:

Respice finem!*

It takes thought and time to produce a clear and concise piece of writing. The story is told that Pascal—a master of French prose—once apologized at the end of a long letter, saying that he simply had not had time to write a short letter. The great mathematician C. F. Gauss told a friend:

> You know that I write slowly. This is chiefly because I am never satisfied until I have said as much as possible in a few words, and writing briefly takes far more time than writing at length.†

But one can write too little just as easily as one can write too much. A proper balance must be struck.

* Literally, "Respect your goal!" or "Have a high regard for the final result!" The phrase is often understood in its broadest sense, where it expresses a philosophy of life.

† From a letter by Gauss, as quoted in *Ways of Thought of Great Mathematicians*, by Herbert Meschkowski. Holden-Day, 1964, p. 62.

EXAMPLE
Consider the function given by $f(x) = x^2 - 4x + 2$. Find the coordinates of the highest point on the graph of f if the domain of f is specified by the inequality $0 \le x \le 3$.

"Solution" by Student D

$$y = x^2 - 4x + 2 = 2x - 4 = 0$$
$$2x = 4$$
$$x = 2$$

Remark
Student D appears to be slavishly following rule 1, for he has broken rules 4, 5, and 7. His statement

$$x^2 - 4x + 2 = 2x - 4$$

adds a touch of algebraic humor to this brief comedy of errors.

"Solution" by Student C
If $y = x^2 - 4x + 2$, then $y' = 2x - 4$. The derivative is then equal to zero when $2x - 4 = 0$, or $x = 2$. The point $x = 2$ is then the highest point on the graph of f.

Remark
Although Student C demonstrates knowledge of a nonalgebraic language—and thus appears to be better educated than Student D—his attempted solution is still inadequate. For one thing, the point $x = 2$ is on the x-axis and not on the graph of f.

"Solution" by Student B
To find the highest point, we set the derivative $2x - 4$ equal to zero. We get $x = 2$. Since $f(2) = -2$ we have a horizontal tangent line to the graph of f at the point $(2, -2)$. The highest point is therefore $(2, -2)$.

Remark
Student B has favored us with four informative sentences indicating much knowledge of calculus. But the fourth sentence does not follow from the third, and this breaks rule 2.

Solution by Student A
The only critical point occurs when $x = 2$. Since the largest value attained by a continuous function must occur at a critical point or at an endpoint, we need only glance at the following table to see that $(0, 2)$ is the highest point on the graph of f.

x	y
0	2
2	-2
3	-1

Remark
Student A has style.

Solution by Student A$^+$
Since the second derivative (given by $f''(x) = 2$) is always positive, the curve f is always concave upwards. Every such curve, like every smile, reaches its highest point at one end, and the endpoints here are (0, 2) and $(3, -1)$. The highest point on the curve is then (0, 2).

§2. Rewriting

"This first thing to do is to get there. Do not bother about your style, but solve your problem ..." Whitehead's point is well taken. Virtually any means of solving a problem is legitimate, whether by a calculated method, by eliminating wrong answers, or by pure guesswork. If you are like the author of this book, you will make a big mess. You will fill up pages with hastily scrawled, illegible handwriting (half of which will be crossed out, being irrelevant), you will sketch badly drawn figures (which will not be improved when you spill coffee on them), and you will lose your pencil (the one that still had a good eraser). You will begin to believe those who say that scientific research is the purest example of an essentially comic activity.

But you learn, after all, through play; comic activity serves a serious purpose. Almost miraculously, your playful attempts may begin to give form to something new, however dimly conceived. Then your work is really cut out for you. What is becoming clear to you must be shown related to things familiar to all. It is here, with your end already in mind, that you begin to worry about style.

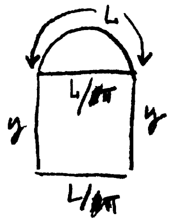
The chances are that you must rethink your whole project. First you must decide for whom you are writing. Are you addressing your instructor and classmates, or some wider circle? It is well to keep in mind some real or imaginary audience.

What is your goal? Is it to impress your reader with your knowledge, or is it to lead your reader to that knowledge? Or do you see your task as offering the most direct possible justification of some assertion? Your goal will determine your style.

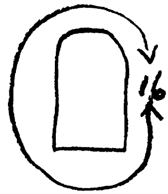
EXAMPLE

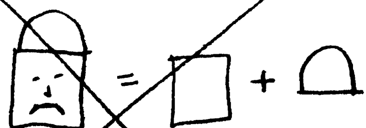
A Norman window is in the shape of a rectangle surmounted by a semicircle. If the perimeter of the window is 16 feet, find the dimensions maximizing the area.

Comic activity.



$$2y + 2(L/\cancel{2}\pi) = 16$$
$$y + L/\cancel{2}\pi = 8$$
$$y = 8 - L/\cancel{2}\pi$$

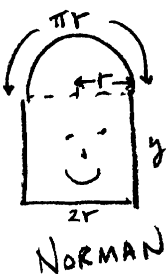


A =  =

$= y(L/\cancel{2}\pi)$ + area semicircle

$= (8 - L/\cancel{2}\pi)(L/\cancel{2}\pi) + ?$

must be easier way



$$2y + \pi r + 2r = 16$$
$$y = \frac{16 - \pi r - 2r}{2}$$
$$A = \frac{1}{2}\pi r^2 + 2ry$$
$$= \frac{1}{2}\pi r^2 + \cancel{2}r\left(\frac{16 - \pi r - 2r}{\cancel{2}}\right)$$
$$= \frac{1}{2}\pi r^2 + 16r - \pi r^2 - 2r^2$$

$$A' = \pi r + 16 - 2\pi r - 4r = 0 \text{ when}$$

$$-\pi r - 4r = -16$$

$$(-\pi - 4)\, r = -16$$

$$r = \frac{16}{4+\pi} \approx 2 \quad \underline{\text{EUREKA}}$$

$$A'' = \pi - 2\pi - 4 = -\pi - 4 < 0$$

so $\underline{\text{from}}$

$\underline{\text{MAX}}$

Solution A

Let r be the radius of the semicircle making up the top of a Norman window whose perimeter is 16 feet. It follows easily that the rectangular portion of the window must be of dimensions $2r$ by $(16 - \pi r - 2r)/2$ feet. The area A of the window is the sum of the areas of the semicircle and the rectangle:

$$\begin{aligned} A &= \frac{1}{2}\pi r^2 + 2r(16 - \pi r - 2r)/2 \\ &= \left(-2 - \frac{1}{2}\pi\right) r^2 + 16r. \end{aligned}$$

This is just a simple quadratic function whose leading term is negative, so it attains a maximum at its critical point. To find the critical point we set the derivative A' equal to zero:

$$\begin{aligned} 2\left(-2 - \frac{1}{2}\pi\right) r + 16 &= 0, \\ (4 + \pi) r &= 16, \\ r &= 16/(4 + \pi) \\ &\approx 2.24 \text{ feet.} \end{aligned}$$

To maximize the area, the rectangular portion of the window must be of dimensions 4.48 by 2.24 feet, approximately.

Solution A$^+$

It is no harder to consider the more general problem where a fixed perimeter P is specified, and to prove the following theorem.

Norman Window Theorem

Let P be the perimeter of a Norman window, made up of a rectangle surmounted by a semicircle. Then the area of the window is maximized if the rectangle has base $2P/(4+\pi)$ and height $P/(4+\pi)$.

Proof

If r is the radius of the semicircle, it follows by easy algebra that the rectangle has these dimensions:

$$\text{base} = 2r, \tag{$*$}$$

$$\text{height} = \frac{P - \pi r - 2r}{2}. \tag{$**$}$$

The area A of the window is the sum of the areas of the semicircle and the rectangle. When these are calculated and combined we get

$$\begin{aligned} A &= -\left(\frac{4+\pi}{2}\right)r^2 + Pr, \\ A' &= -(4+\pi)r + P, \\ A'' &= -(4+\pi). \end{aligned}$$

Setting A' equal to zero immediately yields $r = P/(4+\pi)$. This gives the only critical point, which is a maximum since A'' is negative. Substituting this value of r into equations (*) and (**) shows that the area A is maximized when the rectangle is of dimensions $2P/(4+\pi)$ by $P/(4+\pi)$.

$\square$

§3. Summary

Like virtually every course taught in the liberal arts, a course in mathematics is in part a course in writing. A student cannot learn to think like a mathematician without learning to write like a mathematician.

Style in writing is of little use, however, unless you first have something to say. To find something new you must strike out on your own, with a willingness to make mistakes, to learn from them, and to laugh at yourself. By playing the fool in a comedy of errors, you may find the means to climb up to a more serious level.

What has been discussed in this appendix is the classical sense of style that is so well depicted by Whitehead. This sense derives to some extent from classical mathematics. In modern times other notions about style have arisen, the most notable being that style is virtually synonymous with self-expression. Since everyone agrees that it is fatal to imitate the style of another, many believe that style is acquired only by writing away after one's own fashion with a proud indifference to any discipline imposed from outside.

Mathematics was born, however, in a less (or perhaps more) sophisticated time when self-expression was not so important. The purpose of education was not then to learn to express yourself; it was to learn to tell the truth. And even today, when writing in the discipline of mathematics, you may find yourself most squarely standing between the reader and the truth. In this austere place there is little room for self-expression. You must get yourself out of the way. It is only good manners to bow.

Problems (Optional)

1. Consider the function given by $f(x) = 3x^2 + 6x + 7$. Find the coordinates of the highest point on the graph of f if the domain of f is specified by the inequality $0 \le x \le 3$. Write up your solution like Student A in Section 1. Or be less methodical and more creative like Student A$^+$.

2. A Norman window is in the shape of a rectangle surmounted by a semicircle. If the perimeter of the window is 16 feet, find the dimensions maximizing the area of the *rectangular portion* of the window.

3. An athletic field is to be built roughly in the shape of an oval, with a 400-meter track as its perimeter. The field is to consist of a rectangle with a semicircle at each end. Find the dimensions of the field maximizing the area of the rectangular portion.

4. In problem 3, find the dimensions of the field maximizing its total area.

5. (*Make a guess.*) If you did the preceding problem correctly, you can probably guess the correct answer to "Dido's problem": Given a piece of string of fixed length, say 400 meters, what curve should you make from it to enclose the largest possible area? You are free to mold the string into a triangle, a rectangle, a square, a hexagon, an ellipse, or any other curve that encloses an area inside it. What curve would you use?

   (According to legend, Dido was given a challenge similar to this by a local chieftain who derisively told her she could have all the land she could enclose with a bull's hide. She then cut the hide up into razor-thin strips and tied them together to make a very long strand. After shaping the strand in such a way as to maximize the area inside, Dido claimed her new kingdom of Carthage and became its fabled queen.)

6. Antique mantel pieces from the Renaissance sometimes have RESPICE FINEM carved upon them. Here the phrase we met in Section 1 expresses a philosophy of life. Describe this philosophy by contrasting it with that expressed by another familiar Latin phrase, *carpe diem* ("seize the day").

7. In the passage quoted at the beginning of this appendix Alfred North Whitehead, one of the twentieth century's greatest philosophers, implies that the final aim of education is the cultivation of a sense for style. What do you think is the final aim of education and what, if anything, does the study of mathematics contribute to its attainment?

5 APPENDIX From Freely Falling Bodies to Taylor Series

What are the higher-order derivatives good for? Studying them leads us into a realm of striking mathematical ideas developed in the eighteenth and nineteenth century. One of the things we see in this appendix, for example, is that if you have imprecise, or "approximate", information about the general size of some higher derivative of a function—say, you know that its tenth derivative is never smaller than -1 nor greater than $+1$—then you can deduce quite surprising information about the size of the function itself (see Example 10 below). This way of gleaning information about how the size of a higher derivative of a function affects the size of the function itself is called the theory of *Taylor approximation*.

The term comes from the name of Brook Taylor (1685–1731), although Taylor himself never considered the question about how closely his "Taylor series" approximates the function from which it arises. This question was taken up later by others, notably the French analyst Joseph Louis Lagrange (1736–1813), who came close to the heart of the matter. In this appendix, however, we shall not follow the historical development because, in retrospect, it may be seen that the central question is more easily asked and more quickly answered by reconsidering the simple theory of freely falling bodies discussed in Chapter 6.

§1. Freely Falling Bodies and Quadratic Approximations

If the acceleration due to gravity of a freely falling body is not precisely constant, but instead varies (in an unknown way) between two constants

g and G (where $g < G$), then the height h of the body cannot be exactly predicted. Nevertheless, it is easy to see, as in problem 28 of Chapter 6, that the height $h(t)$ at time t must satisfy the inequality

$$h_0 + v_0 t + \tfrac{1}{2} g t^2 \leq h(t) \leq h_0 + v_0 t + \tfrac{1}{2} G t^2, \tag{1}$$

where h_0 is the initial height and v_0 is the initial upward speed, that is, $h_0 = h(0)$ and $v_0 = h'(0)$. Beginning with time $t = 0$, inequality (1) will continue to hold for positive values of t until such time as forces other than gravity come into play. Thus, if air resistance is ignored, inequality (1) will usually be expected to hold until the body hits the ground.

We have stumbled here upon a small miracle. We have noted that a bound upon the second derivative $h''(t)$ between g and G forces a major constraint, given by inequality (1), upon the size of $h(t)$ itself. If we are given the initial values $h(0)$ and $h'(0)$ and we are further told that $h''(t)$ always lies between two constants g and G, then we may invoke inequality (1) to predict the approximate size of $h(t)$ when $t > 0$.

EXAMPLE 1

Let h be a function satisfying $h(0) = 200$, $h'(0) = 100$, and suppose the second derivative of h is not known exactly, but is bounded between -32 and -30,

$$-32 \leq h''(t) \leq -30.$$

What can we say about the approximate size of $h(1)$? What about $h(2)$? What about $h(4)$ and $h(8)$?

According to inequality (1), we can say that for all positive t for which the inequality $-32 \leq h''(t) \leq -30$ obtains, we are entitled to write

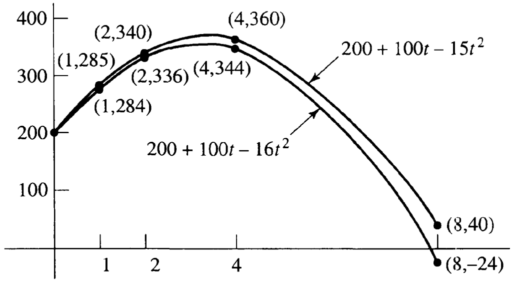
$$200 + 100t + \tfrac{1}{2}(-32)t^2 \leq h(t) \leq 200 + 100t + \tfrac{1}{2}(-30)t^2.$$

Substituting 1, 2, 4, and 8 for t in this inequality shows that

$$\begin{aligned} 284 &\leq h(1) \leq 285, \\ 336 &\leq h(2) \leq 340, \\ 344 &\leq h(4) \leq 360, \\ -24 &\leq h(8) \leq 40. \end{aligned}$$

□

Here we see that the upper and lower bounds in these inequalities are close together for $h(1)$ but grow farther apart in the inequalities for $h(2)$, $h(4)$, and $h(8)$. This will turn out to be a general feature of Taylor approximation. The closer t is to the point where we have the most precise data—the point 0, in this case—the better able we are to approximate $h(t)$ by using the upper and lower bounds in inequality (1).



The function $h = h(t)$ *is bounded between two quadratic functions if* $0 \le t$. *The bounding quadratic curves grow farther apart as t increases.*

Is there anything special about freely falling bodies in this regard? Shouldn't we expect the *quadratic inequality* (1) to apply generally to any function $h = h(t)$ whose second derivative is bounded between g and G? The answer is *yes*, and a quick proof of this general principle of **second-degree Taylor approximation** is outlined in problem 1 at the end of this appendix. Let us see what sort of information this principle gives.

EXAMPLE 2

(*Applying inequality (1) to general functions*) Suppose $h(0) = 10$, $h'(0) = 1/20$, and $-1/4000 \le h''(t) \le 0$ if $0 \le t$. What can we say about the approximate size of $h(1)$? What about $h(2)$? What about $h(4)$, $h(8)$, and $h(20)$?

By inequality (1), with $g = -1/4000$ and $G = 0$, we can say that

$$10 + (1/20)t - (1/8000)t^2 \le h(t) \le 10 + (1/20)t, \quad \text{if } 0 \le t. \tag{2}$$

Substituting 1, 2, 4, 8, and 20 for t in inequality (2) shows that

$$10.049875 \le h(1) \le 10.050000, \tag{3}$$

$$10.0995 \le h(2) \le 10.1000, \tag{4}$$

$$10.198 \le h(4) \le 10.200, \tag{5}$$

$$10.392 \le h(8) \le 10.400, \tag{6}$$

$$10.95 \le h(20) \le 11.00. \tag{7}$$

□

It turns out, as the reader is asked to demonstrate in problem 2 of this appendix, that the function given by $h(t) = \sqrt{t + 100}$ actually satisfies the conditions given in Example 2, so that these inequalities must hold when $h(1) = \sqrt{101}$, $h(2) = \sqrt{102}$, etc. Our last two inequalities, (6) and (7) above, for example, then tell us that

$$10.392 \le \sqrt{108} \le 10.400,$$

$$10.95 \le \sqrt{120} \le 11.00,$$

which, of course, are easily seen to be true.

§2. Cubic Inequalities from Bounds on Third Derivatives

In the preceding section we have seen how a bound on the second derivative of a function leads to a quadratic inequality restricting the size of the function itself. Now suppose we have a bound on the *third* derivative of a function h, that is, suppose that for some constants m and M,

$$m \le h'''(t) \le M, \quad \text{if } 0 \le t. \tag{8}$$

To deal with this, we need only remind ourselves that the third derivative of h is, of course, the second derivative of h'. Thus we may apply inequality (1) to this situation, where the role of h is now played by h', and the roles of the constants g and G are played by m and M. We get

$$h'(0) + h''(0)t + \tfrac{1}{2}mt^2 \le h'(t) \le h'(0) + h''(0)t + \tfrac{1}{2}Mt^2, \quad \text{if } 0 \le t.$$

Now we face the situation introduced in Example 11 of Chapter 6, and we proceed in the same way. The theorem on antiderivatives and inequalities (Chapter 6, Section 9) says that from this inequality we may take antiderivatives to deduce that

$$\begin{aligned} h'(0)t + h''(0)\frac{t^2}{2} + m\frac{t^3}{6} &\le h(t) - h(0) \\ &\le h'(0)t + h''(0)\frac{t^2}{2} + M\frac{t^3}{6}, \quad \text{if } 0 \le t. \end{aligned}$$

Here, $h'(0)$ and $h''(0)$ are merely constants, just like m and M, so the business of taking antiderivatives could hardly have been simpler. Adding the quantity $h(0)$ to each of the three members of this inequality then proves that $h(t)$ is sandwiched between two cubic functions if its third derivative satisfies condition (8):

$$\begin{aligned} h(0) + h'(0)t + \tfrac{1}{2}h''(0)t^2 + \tfrac{1}{6}mt^3 &\le h(t) \\ &\le h(0) + h'(0)t + \tfrac{1}{2}h''(0)t^2 + \tfrac{1}{6}Mt^3, \quad \text{if } 0 \le t. \end{aligned} \tag{9}$$

Let us put this general principle of **third-degree Taylor approximation** to use.

EXAMPLE 3
Suppose we know that $3 \leq h'''(t) \leq 6$, where the function $h = h(t)$ satisfies the initial conditions given in the first row of the table below.

t	h	h'	h''
0	7	5	−3
1	?		
2	?		
4	?		

(10)

What can be said about the action of the function $h = h(t)$ when $t = 1$, $t = 2$, and $t = 4$?

We simply use inequality (9) with $h(0) = 7$, $h'(0) = 5$, $h''(0) = -3$, $m = 3$, and $M = 6$ to write

$$7 + 5t - \tfrac{3}{2}t^2 + \tfrac{1}{2}t^3 \leq h(t) \leq 7 + 5t - \tfrac{3}{2}t^2 + t^3, \quad \text{if } 0 \leq t. \tag{11}$$

Substituting 1, 2, and 4 for t in the cubic inequality (11) gives us the required information:

$$11 \leq h(1) \leq 11.5,$$
$$15 \leq h(2) \leq 19,$$
$$35 \leq h(4) \leq 67.$$

Note that we know the approximate value of the function at 1 much more precisely than at 4, because 1 is closer to the point 0 at which we have precise knowledge about the function. □

EXAMPLE 4
Can we approximate integrals as well? Given the information about the function $h = h(t)$ in the first row of the table in Example 3 and the bound $3 \leq h'''(t) \leq 6$, what can we say about the integral $\int_0^1 h(t)\,dt$?

Here we haven't enough information to determine the integral exactly, but we do have enough to write down inequality (11), from which it follows easily (*why?*) that

$$\int_0^1 (7 + 5t - \tfrac{3}{2}t^2 + \tfrac{1}{2}t^3)\,dt \leq \int_0^1 h(t)\,dt \leq \int_0^1 (7 + 5t - \tfrac{3}{2}t^2 + t^3)\,dt. \tag{12}$$

By using the fundamental theorem to evaluate the integrals on the left and right, we deduce that

$$9.125 \leq \int_0^1 h(t)\,dt \leq 9.25. \tag{13}$$

Thus we have determined the value of the integral from 0 to 1 with fairly good accuracy. If we had tried to estimate the size of the integral from 0 to 4, however, we would have found its value to be much more indeterminate. (See problem 3.) □

The reader may recall the binomial theorem from algebra, which gives a formula for raising the sum of two quantities to any positive integral power. In the case of the third power, the formula is given by

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3. \tag{14}$$

Surprisingly, this result for third powers can be obtained as a consequence of the Taylor theory of cubic approximations, as seen in Example 5.

EXAMPLE 5
Apply the result (9) on cubic inequalities when $h(t) = (\pi + t)^3$.

We must take the first three derivatives of $h(t)$, find a bound on the third derivative, then write inequality (9) using the data we have found. This is easy:

$$\begin{aligned}
h(t) &= (\pi + t)^3 && [\text{so that } h(0) = \pi^3] \\
h'(t) &= 3(\pi + t)^2 && [\text{so that } h'(0) = 3\pi^2] \\
h''(t) &= 6(\pi + t) && [\text{so that } h''(0) = 6\pi] \\
h'''(t) &= 6 && [\text{so that } 6 \le h'''(t) \le 6].
\end{aligned}$$

Using these initial conditions in inequality (9), with $m = M = 6$, we find that

$$\begin{aligned}
\pi^3 + 3\pi^2 t + \tfrac{1}{2}(6\pi)t^2 + \tfrac{1}{6}(6)t^3 &\le (\pi + t)^3 \\
&\le \pi^3 + 3\pi^2 t + \tfrac{1}{2}(6\pi)t^2 + \tfrac{1}{6}(6)t^3, \quad 0 \le t.
\end{aligned}$$

In this inequality the extreme left and right sides are identical. The expression sandwiched in between must therefore be identically equal to either, or

$$(\pi + t)^3 = \pi^3 + 3\pi^2 t + 3\pi t^2 + t^3.$$ □

We have thus used third-degree Taylor approximation to discover the binomial theorem (14) for third powers. The ambitious reader may wish to follow in the footsteps of Isaac Newton and use Taylor approximation of nth degree to discover the *general binomial theorem*. (See problem 11.) Newton did this, he later said, shortly before obtaining his bachelor's degree from Cambridge in 1664—in essence, by using "Taylor" series before Brook Taylor had even been born. His success inspired him to devote himself further to mathematics.

§3. Translation of Variables and Taylor Polynomials

It should be intuitively clear that the role of zero is not essential and that initial conditions given at any other point would do just as well by a simple "translation" of variables. Example 6 shows how to move things over.

EXAMPLE 6
Suppose $3 \le f'''(x) \le 6$, and suppose f satisfies the following initial conditions when $x = 10$.

x	$f(x)$	$f'(x)$	$f''(x)$
10	7	5	-3

(15)

Estimate the size of $f(11)$, $f(12)$, and $f(14)$, and the size of the integral $\int_{10}^{11} f(x)dx$.

The trouble here is that we are given the "initial data" not at $x = 0$, but at $x = 10$. To rectify this, all we have to do is introduce a new variable t defined by $t = x - 10$. Think of t as a "translation" of x by 10 units so as to move the point $x = 10$ back to the point $t = 0$. To say $t = x - 10$ is to say $x = t + 10$, so that (obviously) $dx/dt = 1$. Now consider the simple chain given by $h = f(x)$, where $x = t + 10$, that is, $h = f(t + 10)$. By the chain rule,

$$h'(t) = \frac{dh}{dt} = \frac{dh}{dx}\frac{dx}{dt} = f'(x) \quad \left[\text{since } \frac{dh}{dx} = f'(x) \text{ and } \frac{dx}{dt} = 1\right].$$

Thus, $h'(t) = f'(x)$, where $x = t + 10$. By using the chain rule again on this chain of relations, we get $h''(t) = f''(x)$, where $x = t + 10$, and one more application of the chain rule shows that $h'''(t) = f'''(x)$. Thus the information in the table in Example 6 in terms of x and $f(x)$ translates exactly into the information (10) in the first row of the table in Example 3 in terms of t and $h(t)$—from which inequality (11) follows. By substituting $x - 10$ for t in inequality (11), so that $h(t)$ becomes $f(x)$, we see that if

$0 \le x - 10$, that is, if $10 \le x$, then

$$7 + 5(x-10) - \tfrac{3}{2}(x-10)^2 + \tfrac{1}{2}(x-10)^3$$
$$\le f(x) \le 7 + 5(x-10) - \tfrac{3}{2}(x-10)^2 + (x-10)^3 \qquad (16)$$

Moreover, $f(11) = h(1)$, $f(12) = h(2)$, $f(14) = h(4)$, and $\int_{10}^{11} f(x)dx = \int_0^1 h(t)\,dt$, and estimates of the sizes of all these quantities are worked out in Example 3. Thus the answers for Example 6 can be read off the answers for Example 3 by a simple translation of the variable by 10 units. In particular, $11 \le f(11) \le 11.5$, and

$$9.125 \le \int_{10}^{11} f(x)\,dx \le 9.25. \qquad (17) \quad \square$$

To discuss the general theory of Taylor approximation efficiently, we must develop some abbreviative language. The first thing to introduce is the notion of **factorials**. The factorial of n, where n is a positive integer, is denoted by $n!$ and defined as the product of all positive integers up to, and including, n. Thus, $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, and so on. The factorials grow quite rapidly, so that 10! exceeds three million and 15! exceeds one trillion. For most of us, the national debt of the United States is a colossally large number, yet—at the time of this writing—it is still far less than 16! dollars. As we shall see in Section 4, it is the huge size of the factorials of relatively small integers that often makes it possible for Taylor approximation to be surprisingly effective.

We also need to have a more compact notation for higher derivatives. It is awkward to denote the fifth derivative of f by f''''', so it is conventionally denoted by $f^{(5)}$ instead. In general, we denote the kth derivative of a function f by $f^{(k)}$. Thus, instead of writing h''', we may write $h^{(3)}$. The reason for the parentheses here, of course, is to enable us to distinguish the nth derivative from the nth power. Thus f^2 denotes the square of the function f, whereas $f^{(2)}$ denotes the second derivative of f.

Finally, it is convenient to have a name for the row of numbers that tabulate the initial conditions we wish to use for our Taylor approximations. Let us borrow the word *signature* for this. Thus the row of four numbers displayed in (10) is the second-order signature of the function h at 0, and the row of four numbers in (15) is the second-order signature of the function f at 10. Note that it takes $n + 2$ numbers to specify the nth-order signature of a function.

Definition

Let f be a function having n derivatives, and let a be a fixed point in the domain of f. Then by the ***n*th-order signature of *f* at the point *a***, we mean the row of constants displayed in the second line of the table below:

x	$f(x)$	$f^{(1)}(x)$	$f^{(2)}(x)$	$f^{(3)}(x)$	$\cdots$	$f^{(n)}(x)$
a	$f(a)$	$f^{(1)}(a)$	$f^{(2)}(a)$	$f^{(3)}(a)$	$\cdots$	$f^{(n)}(a)$

Definition

The nth-degree **Taylor polynomial** of f, expanded at the point a, is the polynomial p_n of nth degree having the same nth-order signature as f at the point a.

How can we write down quickly the nth-degree Taylor polynomial from the nth-order signature? As we shall see in our next example, the first-degree Taylor polynomial p_1 of a function f expanded at the point a turns out to be nothing other than the linear function that is tangent to the curve f at the point $(a, f(a))$, and the higher degree polynomials are rather obvious generalizations of this notion.

EXAMPLE 7

Write the Taylor polynomials of degree 0, 1, 2, 3, 4, and 5, arising from the signature given below:

x	$f(x)$	$f^{(1)}(x)$	$f^{(2)}(x)$	$f^{(3)}(x)$	$f^{(4)}(x)$	$f^{(5)}(x)$
10	7	5	−3	1	1	1

(18)

Here we have $a = 10$ as the point of expansion, and therefore,

$$p_0(x) = 7 \qquad \text{[constant function; horizontal line through } (10,7)]$$

$$p_1(x) = 7 + 5(x-10) \qquad \text{[best linear approximation to } f \text{ at } (10,7)]$$

$$p_2(x) = 7 + 5(x-10) - \tfrac{3}{2}(x-10)^2$$

$$\text{[best quadratic approximation to } f \text{ at } (10,7)]$$

$$p_3(x) = 7 + 5(x-10) - \tfrac{3}{2}(x-10)^2 + \tfrac{1}{6}(x-10)^3 \qquad \text{[best cubic at } (10,7)]$$

$$p_4(x) = 7 + 5(x-10) - \tfrac{3}{2}(x-10)^2 + \tfrac{1}{6}(x-10)^3 + \tfrac{1}{24}(x-10)^4$$

$$\text{[best quartic]}$$

$$p_5(x) = 7 + 5(x-10) - \tfrac{3}{2}(x-10)^2 + \tfrac{1}{6}(x-10)^3 + \tfrac{1}{24}(x-10)^4$$
$$+ \tfrac{1}{120}(x-10)^5 \qquad \text{[best quintic]}$$

How do we know these are correct? We must verify that the nth-order signature of p_n at 10 agrees with the given nth-order signature of f at 10. To verify that our expression for $p_2(x)$ is correct, for example, we simply take two derivatives and evaluate them when $x = 10$:

$$p_2(x) = 7 + 5(x-10) - \tfrac{3}{2}(x-10)^2 \quad [\text{so that } p_2(10) = 7]$$
$$p_2'(x) = 5 - 3(x-10) \quad [\text{so that } p_2'(10) = 5]$$
$$p_2''(x) = -3 \quad [\text{so that } p_2''(10) = -3]$$

Thus our formula for p_2 is correct because the signature of p_2 at 10 agrees with the signature (18) of f at 10 through second order. It also agrees with (15), and the reader should note that this second-degree Taylor polynomial $p_2(x)$ has already arisen as a prominent part of (16) in connection with second-degree Taylor approximation. When we notice that the denominators 1, 2, 6, 24, and 120 in the expression for $p_5(x)$ are simply the successive factorials of the first five positive integers, we have discovered Taylor's secret of writing down quickly any Taylor polynomial:

Taylor's Theorem

Let f be a function with n derivatives and let a be a fixed point in the domain of f. Let the signature of f at a be given by

x	$f(x)$	$f^{(1)}(x)$	$f^{(2)}(x)$	$f^{(3)}(x)$	$\cdots$	$f^{(n)}(x)$
a	a_0	a_1	a_2	a_3	$\cdots$	a_n

(19)

Then the nth-degree Taylor polynomial of f at a is given by the formula

$$p_n(x) = a_0 + a_1(x-a) + \frac{a_2}{2!}(x-a)^2 + \frac{a_3}{3!}(x-a)^3 + \cdots + \frac{a_n}{n!}(x-a)^n. \quad (20)$$

Proof

It is left to the reader to check that the nth-order signature of p_n at a is given by (19). (It is just as easy to check this as to check the work in Example 7 above.) □

We expect—or rather, we hope—that $p_n(x)$ will be close to $f(x)$ if n is large:

$$f(x) \approx p_n(x). \quad (21)$$

Writing the approximation (21) is straightforward if we are given, or if we can calculate, the nth-order signature of f. The most interesting signature is a string of "ones" going on forever. Let us see what happens then.

Example 8

Suppose f is a function having the following signature at 0. Use approximation (21) to fill in the question mark with your best guess at the value of $f(1)$.

x	$f(x)$	$f^{(1)}(x)$	$f^{(2)}(x)$	$f^{(3)}(x)$	$\cdots$	$f^{(n)}(x)$
0	1	1	1	1	$\cdots$	1
1	?					

By using equation (20) with $a = 0$ and $1 = a_0 = a_1 = \cdots = a_n$, we have

$$p_n(x) = 1 + x + \tfrac{1}{2}x^2 + \tfrac{1}{6}x^3 + \tfrac{1}{24}x^4 + \tfrac{1}{120}x^5 + \tfrac{1}{720}x^6 + \cdots + \tfrac{1}{n!}x^n. \qquad (22)$$

Using equation (21) and setting x equal to 1 in the expression (22) for $p_n(x)$ gives us what we want:

$$f(1) \approx p_n(1) = 1 + 1 + \tfrac{1}{2} + \tfrac{1}{6} + \tfrac{1}{24} + \tfrac{1}{120} + \tfrac{1}{720} + \cdots + \tfrac{1}{n!}, \qquad (23)$$

where we hope this approximation will become more accurate as we take n larger and larger. Remembering how fast factorials grow ($10! = 3{,}628{,}800$), we should expect that the sum of only a few terms will give us very good accuracy. Taking $n = 10$, we get $f(1) \approx p_{10}(1) = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} + \frac{1}{5040} + \frac{1}{40320} + \frac{1}{362880} + \frac{1}{3628800} \approx 2.71828$. The number 2.71828... calculated this way—as the limit of $p_n(1)$ as n increases without bound—is called **Euler's number** (named for the eighteenth-century Swiss mathematician Leonhard Euler—pronounced "oiler") and is denoted by e. (See problem 16.)

§4. Taylor's Theorem with Remainder

The good thing about Taylor's theorem is that it gives us an answer fairly quickly, as in Example 8. The bad thing is that it often leaves an important question unanswered. If we know, for example, that $f(10) = 7$, $f'(10) = 5$, and $f''(10) = -3$, then by (21) we may quickly write $f(x) \approx p_2(x) = 7 + 5(x - 10) - \frac{3}{2}(x - 10)^2$. If we wanted to estimate $f(9)$, say, then we get $f(9) \approx p_2(9) = 7 + 5(9 - 10) - \frac{3}{2}(9 - 10)^2 = 1/2$. Thus, from Taylor's theorem we quickly estimate that $f(9) \approx 1/2$. The more important question, however, is *what kind of confidence can we have in this estimate?* What "error tolerance" must we allow for? That is, what can we say about the "remainder" in Taylor's approximation?* We must

* Approximations are worth little unless a corresponding estimate of the error tolerance to be allowed is also given (or understood). To say, for example, that Earth is "about five billion years old" cannot be taken seriously as a scientific statement because virtually any large number (it may be argued) is "about five billion". How could we know (unless the speaker has the courtesy to tell us, or unless we share common understanding about the use of significant digits) what numbers are *not*, in his view, "about equal to five billion"?

On the other hand, it is significant when a geologist says that Earth is "4.5 billion years old, with a possible error of at most 0.5 billion." From this we may infer that the geologist is prepared to present scientific data to show us that Earth is at least four billion years old and, further, to convince us that the age of the earth does not exceed five billion years. Whenever we use rough approximations instead of exact statements, we should be careful to include—if possible—a statement of the maximum possible error we might be committing.

make something more precise out of the rough approximation (21) before we can take it seriously.

Taylor's Theorem with Remainder
Let f be a function with $n+1$ derivatives and let a be a point in the domain of f. Let x be any other point in this domain and let m and M be, respectively, lower and upper bounds for the range of $f^{(n+1)}$ on the interval between a and x. Then, if $p_n(x)$ is given by equation (20),

$$f(x) = p_n(x) + R_n \tag{24}$$

where the "remainder" (or "error") R_n lies between $m((x-a)^{n+1}/(n+1)!)$ and $M((x-a)^{n+1}/(n+1)!)$.

Proof
First note that another way of stating the conclusion is that

$$p_n(x) + m\frac{(x-a)^{n+1}}{(n+1)!} \le f(x) \le p_n(x) + M\frac{(x-a)^{n+1}}{(n+1)!}, \tag{25}$$

but with an important, and perhaps surprising, provision: *The inequalities in (25) must be reversed if n is even and $x < a$.* (The inequality fails as written in this case because if n is even and $x < a$, then the left-hand side cannot be smaller than the right-hand side.)

The proof is made by repeatedly applying the theorem on antiderivatives and inequalities, beginning with the given information that for all t between a and x we have

$$m \le f^{(n+1)}(t) \le M.$$

Where we go from here depends upon whether $a < x$ or $x < a$. If $a < x$, the theorem on antiderivatives and inequalities says that for any number t between a and x we have

$$m(t-a) \le f^{(n)}(t) - a_n \le M(t-a), \quad [\text{since } a < t] \tag{26}$$

that is,

$$a_n + m(t-a) \le f^{(n)}(t) \le a_n + M(t-a), \quad \text{if } a \le t \le x.$$

Applying the theorem on antiderivatives and inequalities again, we see that if $a \le t \le x$,

$$a_n(t-a) + \frac{m}{2}(t-a)^2 \le f^{(n-1)}(t) - a_{n-1} \le a_n(t-a) + \frac{M}{2}(t-a)^2, \tag{27}$$

that is,

$$a_{n-1} + a_n(t-a) + \frac{m}{2}(t-a)^2 \le f^{(n-1)}(t) \le a_{n-1} + a_n(t-a) + \frac{M}{2}(t-a)^2.$$

Applying the theorem $n-1$ more times, using (20), then setting t equal to x, proves (25).

It remains to describe the proof in case $x < a$. Then the theorem on antiderivatives and inequalities would give us inequality (26), *but with the inequalities reversed*, since $t < a$. Applying the theorem again with $t < a$, however, will reverse the inequalities once more, and result in inequality (27) as written. A third application will reverse the inequalities, and a fourth will restore them. And so on. The final result is inequality (25) as written if n is odd, but with the inequalities reversed if $x < a$ and n is even. □

EXAMPLE 9
Suppose f is a function whose third derivative is bounded between $m = 3$ and $M = 6$, and which satisfies $f(10) = 7$, $f'(10) = 5$, and $f''(10) = -3$. At the beginning of this section, we have used Taylor's theorem to get quickly a "sloppy" estimate at the value of $f(9)$. Now use Taylor's theorem with remainder to make a *careful* estimate of $f(9)$.

Here we are given the second-order signature of f at 10. Taylor's theorem with $n = 2$ says

$$f(x) = p_2(x) + R = 7 + 5(x - 10) - \tfrac{3}{2}(x - 10)^2 + R, \tag{28}$$

where R $(= R_2)$ is bounded between

$$\frac{3}{3!}(x - 10)^3 \quad \text{and} \quad \frac{6}{3!}(x - 10)^3, \tag{29}$$

that is, R lies between $\frac{1}{2}(x - 10)^3$ and $(x - 10)^3$. Letting $x = 9$ in (28) and (29) shows that $f(9) = p_2(9) + R = 0.5 + R$, where R lies between -0.5 and -1.0. Thus, by Taylor's theorem with remainder, we know $f(9)$ lies between -0.5 and 0.

Another way to arrive at the same conclusion is to write inequality (25), remembering that the inequalities must be reversed (since $n = 2$ is even here, and since $9 < 10$). Inequality (25) here becomes inequality (16), which, when x is set equal to 9 (and the inequalities reversed) shows that $-0.5 \le f(9) \le 0$. □

Notice the difference between the vagueness of Taylor's rough approximation (21) and the precision of Taylor's theorem with remainder. In the setting of the previous example, approximation (21) says only that the value of $f(9)$ is approximately equal to $1/2$—but it *might* in fact be 37 or -11, for all the assurance we can have in using approximation (21) by itself. On the other hand, Taylor's theorem with remainder says we can be sure that $f(9)$ is between $-1/2$ and 0 if we know that the third derivative of f takes values only between 3 and 6 on the interval between 9 and 10.

EXAMPLE 10
Given a certain function $S = S(x)$, suppose we know that the tenth derivative $S^{(10)}(x)$ is bounded between -1 and 1 for all x. Suppose further

that its ninth-order signature at 0 is given as specified in the table below. Use Taylor's theorem with remainder to estimate $S(\pi/6)$, $S(-\pi/2)$, and $S(\pi/3)$.

x	S	$S^{(1)}$	$S^{(2)}$	$S^{(3)}$	$S^{(4)}$	$S^{(5)}$	$S^{(6)}$	$S^{(7)}$	$S^{(8)}$	$S^{(9)}$
0	0	1	0	-1	0	1	0	-1	0	1
$\pi/6$	?									
$-\pi/2$	?									
$\pi/3$	?									

Here, because of the zeroes at each even order in the signature, the Taylor polynomial $p_9(x)$ contains only odd powers of x. By (20), with $a = 0$ and $n = 9$, we have

$$p_9(x) = x - \tfrac{1}{6}x^3 + \tfrac{1}{120}x^5 - \tfrac{1}{5040}x^7 + \tfrac{1}{362880}x^9, \tag{30}$$

and

$$S(x) = p_9(x) + R, \tag{31}$$

where R lies between $\frac{-1}{3628800}\,x^{10}$ and $\frac{1}{3628800}\,x^{10}$. Letting $x = \frac{\pi}{6}$ in (30) and (31) shows that $S(\frac{\pi}{6}) = p_9(\frac{\pi}{6}) + R = 0.50000000 + R$, where R is exceedingly small, being less in absolute value than $\frac{1}{3628800}\left(\frac{\pi}{6}\right)^{10}$. Hence, we know that $S(\frac{\pi}{6}) = 0.50000000\ldots$.

Letting $x = \frac{-\pi}{2}$ in (30) and (31) shows that $S(\frac{-\pi}{2}) = p_9(\frac{-\pi}{2}) + R = -1.0000 + R$, where R is less in absolute value than $\frac{1}{3628800}\left(\frac{\pi}{2}\right)^{10}$. Hence, $S(\frac{-\pi}{2}) = -1.0000\ldots$.

Letting $x = \frac{\pi}{3}$ in (30) and (31) shows that $S(\frac{\pi}{3}) = p_9(\frac{\pi}{3}) + R = 0.866025 + R$, where R is less in absolute value than $\frac{1}{3628800}\left(\frac{\pi}{3}\right)^{10}$. Hence, $S(\frac{\pi}{3}) = 0.866025\ldots$. □

§5. L'Hôpital's Rule

As seen in our last example, Taylor's theorem with remainder can predict relatively accurately the value of a function f at a point x relatively far from the point a at which its signature is known. Of course, the prediction becomes increasingly accurate as the point x is taken increasingly close to a. This makes the theorem invaluable, when restated in the form below, in studying limits at a of expressions involving f.

Taylor's Theorem with Remainder (Restatement)
Under the hypotheses of Taylor's theorem with remainder, we may write $f(x) = p_n(x) + R_n$, where the remainder $R_n = R_n(x)$ tends to zero so fast, as x tends to a, that

$$\operatorname*{Limit}_{x \to a} \frac{R_n(x)}{(x-a)^n} = 0.$$

Proof

(*Easy*) We know that $R_n(x)$ lies between $m((x-a)^{n+1}/(n+1)!)$ and $M((x-a)^{n+1}/(n+1)!)$, and both these quantities clearly tend to 0 after being divided by $(x-a)^n$. □

EXAMPLE 11

Find the limit, as x tends to 0, of $(3-\sqrt{9+x})/(2-\sqrt{4-x})$.

The idea is that, since we are taking the limit at 0, things should become simpler if we approximate every complicated expression by its Taylor polynomial at 0. We could use any Taylor polynomial, but we save time by picking the one of lowest degree that will get the job done. Here, first-degree Taylor approximation will do. When we replace $\sqrt{9+x}$ by $3+\frac{1}{6}x+R_1(x)$ and $\sqrt{4-x}$ by $2-\frac{1}{4}x+Q_1(x)$, we see that

$$\frac{3-\sqrt{9+x}}{2-\sqrt{4-x}}=\frac{-\frac{1}{6}x-R_1(x)}{+\frac{1}{4}x-Q_1(x)}, \quad \text{if } x\neq 0.$$

Dividing numerator and denominator of the expression on the right by x, we get

$$\frac{3-\sqrt{9+x}}{2-\sqrt{4-x}}=\frac{-\frac{1}{6}-\frac{R_1(x)}{x}}{+\frac{1}{4}-\frac{Q_1(x)}{x}}, \quad \text{if } x\neq 0.$$

Now the limit is obvious, because Taylor's theorem says that both $R_1(x)/x$ and $Q_1(x)/x$ tend to zero, as x tends to zero. Therefore,

$$\operatorname*{Limit}_{x\to 0}\frac{3-\sqrt{9+x}}{2-\sqrt{4-x}}=\operatorname*{Limit}_{x\to 0}\frac{-\frac{1}{6}-\frac{R_1(x)}{x}}{+\frac{1}{4}-\frac{Q_1(x)}{x}}=\frac{-\frac{1}{6}}{+\frac{1}{4}}=-\frac{2}{3}. \qquad \square$$

When we must take the limit of a quotient at a point a, where both numerator and denominator tend to zero, all we have to do is to write the numerator in terms of an appropriate Taylor polynomial expanded at a, do the same for the denominator, and then use Taylor's theorem with remainder to see clearly what is going on. In the most interesting case the resulting limit is given by the strikingly simple formula (32) below.

L'Hôpital's Rule

Suppose f and g both satisfy the hypotheses of Taylor's theorem and suppose that the nth-order signatures of both f and g at a begin with a string of n zeroes, followed at last, in the case of g, by a nonzero number b_n. That is, suppose we have the following signatures at the point a:

x	$f(x)$	$f^{(1)}(x)$	$f^{(2)}(x)$	$f^{(3)}(x)$	$\cdots\cdots$	$f^{(n)}(x)$
a	0	0	0	0	$\cdots$ (all zeroes) $\cdots$	a_n

x	$g(x)$	$g^{(1)}(x)$	$g^{(2)}(x)$	$g^{(3)}(x)$	$\cdots\cdots$	$g^{(n)}(x)$
a	0	0	0	0	$\cdots$ (all zeroes) $\cdots$	$b_n(\neq 0)$

Suppose further that locally, near the point a, $g(x)$ is nonzero except at a itself. Then,

$$\operatorname*{Limit}_{x\to a} \frac{f(x)}{g(x)} = \frac{a_n}{b_n} = \frac{f^{(n)}(a)}{g^{(n)}(a)}. \tag{32}$$

Proof

Because of all the zeroes in the signature of f at a, the nth-degree Taylor polynomial of f consists of only one term, namely, $a_n(x-a)^n/n!$. Similarly, the nth-degree Taylor polynomial of g is $b_n(x-a)^n/n!$. By Taylor's theorem, as restated above, we have

$$\frac{f(x)}{g(x)} = \frac{\frac{a_n(x-a)^n}{n!} + R_n(x)}{\frac{b_n(x-a)^n}{n!} + Q_n(x)} = \frac{\frac{a_n}{n!} + \frac{R_n(x)}{(x-a)^n}}{\frac{b_n}{n!} + \frac{Q_n(x)}{(x-a)^n}} \to \frac{a_n}{b_n} = \frac{f^{(n)}(a)}{g^{(n)}(a)}, \quad \text{as } x \to a. \quad \square$$

EXAMPLE 12

(*Reworking Example 11*) Use L'Hôpital's Rule to evaluate quickly $\text{Limit}_{x\to 0}\,(3-\sqrt{9+x})/(2-\sqrt{4-x})$.

Here we have the limit (32), where $f(x) = 3-\sqrt{9+x}$ and $g(x) = 2-\sqrt{4-x}$. The signatures of f and g at 0 are quickly worked out to first order:

x	$f(x)$	$f'(x)$
0	0	$-1/6$

and

x	$g(x)$	$g'(x)$
0	0	$1/4$

so the answer, by (32), is the quotient of $-1/6$ by $1/4$, or $-2/3$. $\square$

EXAMPLE 13

Use L'Hôpital's Rule to find $\text{Limit}_{x\to 1}(x^4-6x^3+13x^2-12x+4)/(x^4-2x^3+x^2)$.

Here we have the limit (32), where $f(x) = x^4-6x^3+13x^2-12x+4$ and $g(x) = x^4-2x^3+x^2$. The signatures of f and g at 1 must be worked out to second order before we get a nonzero entry in the signature for g. We find that

x	$f(x)$	$f'(x)$	$f''(x)$
1	0	0	2

and

x	$g(x)$	$g'(x)$	$g''(x)$
1	0	0	2

so the answer, by (32), is

$$\operatorname*{Limit}_{x \to 1} \frac{x^4 - 6x^3 + 13x^2 - 12x + 4}{x^4 - 2x^3 + x^2} = \frac{2}{2} = 1. \qquad \square$$

Just as Taylor did not discover Taylor series, L'Hôpital did not discover L'Hôpital's Rule. The name comes from its appearance in the first calculus textbook, published in France in 1696 by the Marquis de L'Hôpital, a wealthy nobleman. L'Hôpital used notes provided him by Johann Bernoulli (1667–1748), a young Swiss mathematician who discovered the rule and who received a generous monthly stipend in exchange for giving L'Hôpital exclusive use of the lecture notes he was writing. Even after L'Hôpital's death in 1704, Bernoulli never asserted his right to be credited with the discovery of the rule that made his patron famous.

§6. Summary

Lagrange's notion that it would be a good idea to try to express a function as the sum of its (infinite) Taylor series would prove to have fruitful consequences, leading to an easily obtained precision in discussing such nonalgebraic functions as the sine, cosine, exponential, and logarithmic functions. In the problem set that follows, the reader can glimpse how this might be done. Developments in series also lead to a surprisingly natural understanding of the rich field of analysis of functions whose domains and ranges are complex numbers instead of real numbers. It is hoped that the reader will be inclined to pursue these ideas further in other courses. In any case, bon voyage!

Problems (Optional)

1. Suppose that the inequality $g \leq h''(t) \leq G$ holds if $0 \leq t$. Explain why the theorem in Chapter 6 on antiderivatives and inequalities implies that the inequality given below in part (a) follows, if $0 \leq t$. Then explain why each of the succeeding chain of statements (b)–(d) follows from the one before it.
   (a) $gt \leq h'(t) - h'(0) \leq Gt$.
   (b) $h'(0) + gt \leq h'(t) \leq h'(0) + Gt$.
   (c) $h'(0)t + \frac{1}{2}gt^2 \leq h(t) - h(0) \leq h'(0)t + \frac{1}{2}Gt^2$.
   (d) $h(0) + h'(0)t + \frac{1}{2}gt^2 \leq h(t) \leq h(0) + h'(0)t + \frac{1}{2}Gt^2$.
   (e) Explain how the result of the chain of reasoning (a)–(d) justifies the general principle of second-degree Taylor approximation—that is, inequality (1)—used in Example 2.

2. Consider the function given by $h(t) = \sqrt{t+100}$.
   (a) Find $h'(t)$, $h''(t)$, and $h'''(t)$.
   (b) Recall from Chapter 5 how to use the derivative of a function to help find its range. Then use your answer to part (a) to show that on the domain $0 \leq t$, the range of $h''(t)$ is given by $-1/4000 \leq h''(t) < 0$. (The range of h'' can be found by simply applying common sense to your expression giving $h''(t)$, but you will get the same answer if you use $h'''(t)$ to study the behavior of $h''(t)$.)
   (c) Explain how, by using the results of parts (a) and (b), we may infer that $10.049875 \leq \sqrt{101} \leq 10.050000$. *Hint.* See Example 2, equation (3).
   (d) Write the inequalities corresponding to $\sqrt{102}$ and $\sqrt{104}$ that come out of inequalities (4) and (5) in Example 2.
   (e) Approximate $\sqrt{110}$ by substituting $t = 10$ in the quadratic inequality (2) of Example 2.
   (f) We know that $\sqrt{121} = 11$, of course. Nevertheless, substitute $t = 21$ in inequality (2) and get bounds on $\sqrt{121}$. Does 11 lie between these bounds?

3. Suppose we have a function $h = h(t)$ and we know that $3 \leq h'''(t) \leq 6$, $h(0) = 7$, $h'(0) = 5$, and $h''(0) = 3$. These are the initial data given in Example 3.
   (a) Use inequality (11) to estimate $h(3)$.
   (b) Use inequality (11) to estimate $h(5)$.
   (c) Estimate the size of $\int_0^1 h(t)\,dt$ by first explaining how equation (12) is justified in Example 4; then do the easy calculation to show how equation (13) follows from equation (12).
   (d) Estimate the size of $\int_0^2 h(t)\,dt$.
   (e) Estimate the size of $\int_0^4 h(t)\,dt$. Explain why its value is so much more indeterminate than the value of the integrals approximated in parts (c) and (d).

4. Suppose we have a function $C = C(t)$ and we know that $0 \leq C'''(t) \leq 1$. Suppose further we have the initial conditions $C(0) = 1, C'(0) = 0$, and $C''(0) = -1$.
   (a) Use the cubic inequality (9) to estimate the value of $C(1/10)$.
   (b) Apply Taylor's third-degree treatment—inequality (9)—to estimate $C(1)$.
   (c) The reader is not supposed to know this yet, but the **cosine** function $C(t) = \cos(t)$ satisfies the initial conditions given here (and satisfies the given bound on its second derivative if $0 \leq t \leq 1$). Use a calculator to find $\cos(1/10)$. Is your calculator's value of $\cos(0.1)$ within the bounds you found in part (a)?
   (d) Is your calculator's value of $\cos(1.0)$ within the bounds you found in part (b)?

5. Suppose we have a function $S = S(t)$ and we know that $-1 \leq S'''(t) \leq 0$. Suppose further we have the initial conditions $S(0) = 0$, $S'(0) = 1$, and $S''(0) = 0$.
   (a) Use the cubic inequality (9) to estimate the value of $S(1/10)$.
   (b) Apply Taylor's third-degree treatment—inequality (9)—to estimate $S(1)$.
   (c) The reader is not supposed to know this yet, but the **sine** function $S(t) = \sin(t)$ satisfies the initial conditions given here (and satisfies the

given bound on its second derivative if $0 \le t \le 1$). Use a calculator to find $\sin(1/10)$. Is your calculator's value of $\sin(0.1)$ within the bounds you found in part (a)?

(d) Is your calculator's value of $\sin(1.0)$ within the bounds you found in part (b)?

(*The sine function's ninth-degree Taylor approximation is discussed in Example 10.*)

6. Consider the function given by $h(t) = \sqrt{t+25}$.

(a) Write out carefully the steps analogous to those of problem 2 to prove the inequality

$$5 + (1/10)t - (1/1000)t^2 \le \sqrt{t+25} \le 5 + (1/10)t, \quad \text{if } 0 \le t.$$

(b) Plot carefully the graph of the quadratic curve $y = 5 + (1/10)t - (1/1000)t^2$ and the graph of the line $y = 5 + (1/10)t$, both on the domain $0 \le t \le 3$. Do your graphs make it look as if the line is tangent to the quadratic curve at (0,5)? Is it?

(c) In view of the inequality proved in part (a), where must the curve $y = \sqrt{t+25}$ be situated relative to the graphs of the quadratic curve and its tangent line plotted in part (a)?

(d) Let $t = 3$ in the inequality of part (a) to get $\sqrt{28}$ bounded closely from below and above. Take the average of these two close bounds as a reasonable guess at $\sqrt{28}$. How good a guess is it? (To see how good a guess at $\sqrt{28}$ you have, square your guess and see how close to 28 you come. If this "quadratic guess" is disappointingly inaccurate, see the next problem for a better "cubic guess".)

7. We do not get great accuracy in problem 6 where we used second-degree approximations to try to estimate $\sqrt{28}$. Let us try third-degree approximations instead. If $h(t) = (t+25)^{1/2}$, then $h'(t) = (1/2)(t+25)^{-1/2}$, $h''(t) = -(1/4)(t+25)^{-3/2}$, and $h'''(t) = (3/8)(t+25)^{-5/2}$. Therefore we have $h(0) = 5$, $h'(0) = 1/10$, $h''(0) = -1/500$, and $h'''(0) = 3/25{,}000$. It is easy to see that $h'''(t)$ is bounded between 0 and $3/25{,}000$ if $0 \le t$. Thus we may take $m = 0$ and $M = 3/25{,}000$ in inequality (9) to write

$$\begin{aligned} 5 + (1/10)t - (1/1000)t^2 &\le \sqrt{t+25} \\ &\le 5 + (1/10)t - (1/1000)t^2 + (1/50{,}000)t^3, \quad \text{if } 0 \le t. \end{aligned}$$

(a) Let $t = 3$ in this inequality to get $\sqrt{28}$ bounded very closely from below and above. Then take the average of these bounds as a reasonable guess at $\sqrt{28}$. How good a guess is it? (How much better is it than the guess you made in the preceding problem?)

(b) Let $t = 1$ in this inequality to get $\sqrt{26}$ bounded very closely from below and above. Then take the average of these bounds as a reasonable guess at $\sqrt{26}$. How good a guess is it? (Square your guess and see how close you come to 26.)

8. (*Cube roots*) Suppose we want to find the cube root of 10. Let us use both quadratic and cubic approximations to compare their accuracy. Since 8 is the largest perfect cube smaller than 10, it is natural (*why?*) to begin with

$h(t) = (t+8)^{1/3}$. Then (see problem 32 at the end of Chapter 5) we have $h'(t) = (1/3)(t+8)^{-2/3}$ and $h''(t) = -(2/9)(t+8)^{-5/3}$, so $h(0) = 2$, $h'(0) = 1/12$, and $h''(0) = -1/144$. Moreover, $h''(t)$ lies between $-1/144$ and 0 if $0 \le t$.

(a) Show that these calculations, when used in inequality (1) of second-degree Taylor theory, imply that

$$2 + \tfrac{1}{12}t - \tfrac{1}{288}t^2 \le \sqrt[3]{t+8} \le 2 + \tfrac{1}{12}t, \quad \text{if } 0 \le t.$$

Let $t = 2$ in this inequality and conclude that $\sqrt[3]{10}$ lies between 2.1527 and 2.1667. The average of these, which is 2.1597, then ought to be a fairly good guess at $\sqrt[3]{10}$.

(b) Now let us investigate the size of $h'''(t)$. Since $h'''(t) = (12/27)(t+8)^{-8/3}$, we see $h'''(0) = 1/576$ and we see $h'''(t)$ is bounded between $m = 0$ and $M = 1/576$ for $0 \le t$. Show that inequality (9) from third-degree Taylor theory implies that

$$2 + \tfrac{1}{12}t - \tfrac{1}{288}t^2 \le \sqrt[3]{t+8} \le 2 + \tfrac{1}{12}t - \tfrac{1}{288}t^2 + \tfrac{1}{3456}t^3, \quad \text{if } 0 \le t.$$

Now let $t = 2$ and conclude that $\sqrt[3]{10}$ lies between 2.15278 and 2.15509. The average of these, which is 2.15394, ought to be a better guess at $\sqrt[3]{10}$ than was the guess of part (a).

(c) Let $t = 1$ in the inequality of part (b) and get very close bounds on $\sqrt[3]{9}$, worked out to at least five decimal places. Take the average of these very close numbers as a reasonable guess at a five-place approximation to $\sqrt[3]{9}$. How good is your guess? (Cube your guess and see how close to 9 it is.)

9. Approximate $\sqrt[3]{9}$ by applying Newton's method to solve the equation $x^3 - 9 = 0$, taking $G = 2$ as your initial guess. Compare Newton's method with Taylor's method of part (c) of the preceding problem. Which method is easier to understand? Which is easier to carry out? Which gives you a more satisfactory answer? Which is more valuable for a calculus student to learn?

10. In Example 5, Taylor's theorem with remainder—in the simple case when $a = 0$ and $n = 2$—is applied to $h(t) = (\pi + t)^3$. The result proves that $(\pi + t)^3 = \pi^3 + 3\pi^2 t + 3\pi t^2 + t^3$.

(a) In like manner, apply Taylor's theorem with remainder when $a = 0$ and $n = 3$ to the function given by $h(t) = (\pi + t)^4$. (Your first step is to find bounds m and M on the *fourth* derivative.) The result should be that $(\pi + t)^4 = \pi^4 + 4\pi^3 t + 6\pi^2 t^2 + 4\pi t^3 + t^4$.

(b) Apply Taylor's theorem with remainder when $n = 4$ to the function given by $h(t) = (\pi + t)^5$. What is the result?

(c) Apply Taylor's theorem with remainder for general n to the function given by $h(t) = (\pi + t)^n$. Write out enough of the first few terms so that the pattern for succeeding terms is evident. (If you do this correctly, you will have, in essence, proved the binomial theorem in the case where n can be any positive integer.)

11. (*For ambitious students only*) In problem 10 we remain in the realm of algebra because we consider only positive integral powers. Isaac Newton helped open up a new realm of analysis by considering fractional and negative powers.

(a) Consider the function $h(x) = (1 + x)^p$, where p is a constant, *but not a positive integer*. Write out enough of the first few terms of the Taylor polynomial approximations at $a = 0$ so that the pattern for succeeding terms is evident.
(b) Explain why the signature at 0 of the function $h(x)$ in part (a) "goes on forever", whereas the signature of the function in part (c) of the previous problem eventually becomes a string of zeroes from some point on.
(c) Use your answer to part (b) to explain why your answer to part (a) is an "infinite" series, whereas the binomial series from ordinary algebra is "finite".
(d) Let $p = 1/2$ in your answer to part (a) and write out the first three or four terms.
(e) Set x equal to $t/100$ in your answer to part (d).
(f) Multiply both sides of your equation in part (e) by 10.
(g) Explain how your answer to part (f) is related to the approximations to $\sqrt{t+100}$ calculated in problem 2. *Hint.* $\sqrt{t+100} = \sqrt{(100)(1+(t/100))} = 10(1+(t/100))^{1/2}$.
(h) Let $p = 1/3$ in your answer to part (a) and write out the first three or four terms. Then set x equal to $t/8$, and, finally, multiply both sides of your equation by 2. Explain how the result is related to the approximations to $\sqrt[3]{t+8}$ calculated in problem 8. *Hint.* $\sqrt[3]{t+8} = \sqrt[3]{8(1+(t/8))} = 2((1+(t/8))^{1/3}$.
(i) Let $p = -1$ in your answer to part (a) and write out the first three or four terms; then set x equal to $-t$. Explain how the result is related to the formula for the sum of a geometric series discussed in problem 6 of Appendix 2.
(j) Read Isaac Newton's short article "On the Binomial Theorem for Fractional and Negative Exponents" in *The World of Mathematics*, Volume 1, edited by James R. Newman, Simon & Schuster, pp. 521–524. Why do you suppose Newton did not consider exponents that are arbitrary real numbers, such as $\sqrt{2}$ or π?

12. Use the method of Example 11 to find each of the following limits:
(a) $\displaystyle \text{Limit}_{x \to 0} \frac{10 - \sqrt{x+100}}{5 - \sqrt{x+25}}$.
Hint. The appropriate Taylor polynomials are already worked out in problems 2 and 6.
(b) $\displaystyle \text{Limit}_{x \to 0} \frac{10 - \sqrt{x+100}}{2 - \sqrt[3]{x+8}}$.
Hint. The appropriate Taylor polynomials are already worked out in problems 2 and 8.

13. Use the method of Example 12 (L'Hôpital's Rule) to find each of the limits in the preceding problem.

14. The limits in problem 22 of Chapter 1 are easily found by the methods of Chapter 1. Show how L'Hôpital's Rule results in the same answer for each of these six limits.

15. Use the method of Example 12 (L'Hôpital's Rule) to find each of the three limits in problem 23, parts (a), (b), and (d), of Chapter 3.

16. Find each of the following limits. Use the method of Example 11, 12, or 13.

(a) $\displaystyle \operatorname*{Limit}_{x\to 2} \frac{x^2-4}{-3+\sqrt{x^2+5}}$.

(b) $\displaystyle \operatorname*{Limit}_{x\to 0} \frac{5x^2-5x^3}{24+x-12\sqrt[3]{x+8}}$.

(c) $\displaystyle \operatorname*{Limit}_{x\to 1} \frac{x^4+2x^3-3x^2-4x+4}{x^3+3x^2-4}$.

(d) $\displaystyle \operatorname*{Limit}_{x\to -2} \frac{x^4+2x^3-3x^2-4x+4}{x^3+3x^2-4}$.

17. (*Newton's method from Taylor series?*) The reasoning behind Newton's method becomes transparent if one studies the first figure drawn in Section 10 of Chapter 4. Newton himself, however, found $B = G - f(G)/f'(G)$ without referring to a figure at all.

(a) Given the equation $f(x) = 0$, and knowing $f(G)$ and $f'(G)$, just replace $f(x)$ by its Taylor polynomial $p_1(x)$ of first-degree obtained from the first-order signature of f at G. The resulting equation $p_1(x) = 0$ becomes

$$f(G) + f'(G)(x - G) = 0.$$

Do you get $G - f(G)/f'(G)$ as your solution when you solve this equation for x?

(b) (*Newton's method of second order?*) A more sophisticated method to solve $f(x) = 0$, knowing $f(G)$, $f'(G)$, and $f''(G)$, is to replace $f(x)$ by its Taylor polynomial $p_2(x)$ of second-degree obtained from the second-order signature of f at G, and solve for x in the equation $p_2(x) = 0$, that is, solve

$$f(G) + f'(G)(x - G) + \tfrac{1}{2}f''(G)(x - G)^2 = 0.$$

Our better guess B will be a value of x that is a solution to this equation, which may be regarded as a quadratic equation not in x, but in the expression $x - G$. Apply the quadratic formula to solve for this expression, then find a formula for B in terms of G after deciding which of the two roots you must take. Using the formula from this "second-order" Newton's method should result in even swifter convergence to the root of the given equation, once you get close to the root.

18. (*The central function of calculus*) A central role in the calculus should be played by a function f that is its own derivative. Of course, the "zero function" (the function given by $f(x) = 0$ for all x) is such a function. To avoid the trivial zero function, let us require not only that $f'(x) = f(x)$, but also that $f(0) = 1$. We can use Taylor series to calculate $f(x)$.

(a) If $f(x) = f'(x)$, show that $f'(x) = f''(x)$. (*This is so easy it might be difficult.*)

(b) If $f(x) = f'(x)$, show that $f(x) = f'(x) = f''(x) = \cdots = f^{(n)}(x) = \cdots$.

(c) Suppose $f(x) = f'(x)$ and $f(0) = 1$. Show, using part (b), that the signature of f at 0 is given by the "string of ones" in Example 8, where we see that $f(1) = e = 2.71828\ldots$.

(d) Using the method of Example 8, show that $f(1/2) = 1.64872\ldots$. (Notice that $1.64872\ldots$ seems to be suspiciously close to $\sqrt{2.71828\ldots}$).

(e) Using the method of Example 8, show that $f(-1) = 0.367879\ldots$ and notice that $0.36789\ldots$ seems to be suspiciously close to $1/(2.71828\ldots)$.

(f) (*Make a guess.*) From parts (e), (d), and (c), we should be very suspicious that we might have $f(-1) = e^{-1}$, $f(1/2) = e^{1/2}$, and $f(1) = e^1$, where e

is Euler's number, defined in Section 3. What would you guess is the formula for $f(x)$?

(g) Among the graphs plotted in problem 43 of Chapter 5, find the one that looks like it might be the graph of $f(x) = e^x$. Is it the one you picked in that problem to be its own derivative?

19. *(Calculating logarithms)* In problem 21 of Chapter 6 we found that a function $L = A(t)$ satisfying $A'(t) = 1/t$ and $A(1) = 0$ would have the desirable logarithmic property of changing multiplication into addition, that is, $A(st) = A(s) + A(t)$. At that time, however, we had no way of calculating values of this logarithmic function precisely.

(a) Since $A'(t) = 1/t = t^{-1}$, we have $A^{(2)}(t) = -t^{-2}$, $A^{(3)}(t) = 2t^{-3}$, $A^{(4)}(t) = -6t^{-4}$, $A^{(5)}(t) = 120t^{-5}$, and so on. *(Notice how the factorials pop up here.)* Hence, $A(1) = 0$, $A'(1) = 1$, $A''(1) = -1$, $A^{(3)}(1) = 2$, and so on. Show that at the point $a = 1$, the nth-degree Taylor polynomial of $A(t)$ is given by

$$p_n(t) = (t-1) - \tfrac{1}{2}(t-1)^2 + \tfrac{1}{3}(t-1)^3 - \tfrac{1}{4}(t-1)^4 + \cdots + \tfrac{(-1)^{n+1}}{n}(t-1)^n.$$

(b) In problem 18 we found that $e^{1/2} = 1.64872\ldots$. To approximate the logarithm of this number, let t equal $1.64872\ldots$ in the expression for $p_n(t)$ found in part (a). Does it appear that the logarithm of $e^{1/2}$ is equal to $1/2$?

(c) In problem 18 we found that $e^{-1} = 0.367879\ldots$. To approximate the logarithm of this number, let t equal $0.367879\ldots$ in the expression for $p_n(t)$. Does it appear that the logarithm of e^{-1} is equal to -1?

(d) *(Make a guess.)* On the basis of your answers to parts (b) and (c), can you guess what the logarithm of e^x will be? *If so, you should be able to fill in the blanks in the following sentence.* "If $t = e^x$, then the logarithm of t will be ____, and if x is the logarithm of t, then t must be ____."

(e) Explain why the logarithm of a number is simply the power to which Euler's number e must be raised in order to reach that number. (In other words, the logarithm function we have defined here is the logarithm "to the base e".)

This problem set has offered only the briefest of introductions to trigonometric functions, logarithms, and exponentials. There are many details to be filled in that would be presented much more leisurely in a calculus course at the next level.

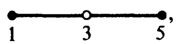
Answers to Selected Problems

Chapter 1

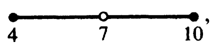
1. (a) $C = (84/W) + 4W$. (b) $0 < W$.

3. (a) $A = 1200w - 2w^2$. (b) $0 < w < 600$.

9. (a) domain F is



or $1 \leq x \leq 5$, $x \neq 3$. range F is



or $4 \leq y \leq 10$, $y \neq 7$.

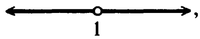
(b) domain f is

or $3 < x \leq 8$. range f is

or $y = 4$ or 7.

11. (a) No. (b) No.

13. (a) range F is



or $y \neq 1$.

17. (b) $A = 20x - (1/12)x^2$, $0 < x < 240$.

19. (c) $0 < L < \sqrt{120}$.

CHAPTER 2

1. *Hint.* Read Section 1.

5. (a) $(x+3)^2 + (y-4)^2 = 25$.
(b) $(x-3)^2 + y^2 = 5$.
(c) $(x-a)^2 + (y-b)^2 = r^2$.

7. (d) *Hint.* See exercise 5.11.

11. *Suggestion*: Given the fact that x is irrational, suppose that the equation $\sqrt{x} = m/n$ holds for some integers m and n. Explain why this supposition leads to an absurd conclusion as soon as you square both sides of this equation.

13. (b) *Suggestion*: Suppose the equation $\sqrt[12]{2} = m/n$ holds for some integers m and n. Explain why this supposition leads to an absurd conclusion as soon as you raise both sides of this equation to the sixth power.

15. (a) If $G = 5/4$, then $B = 378/300 = 63/50$; if $G = 63/50$, then $B = 375047/297675$.

17. *Hint.* The area of the largest semicircle is $\frac{1}{2}\pi(c/2)^2 = \frac{1}{8}\pi c^2 = \frac{1}{8}\pi(a^2+b^2)$, since $c^2 = a^2 + b^2$. Now calculate the sum of the areas of the two smaller semicircles and see whether you get the same thing.

19. *Hint.* Use the Pythagorean Theorem to find the length of the diagonal of a unit square. Can the distance between two points lying in the unit square exceed this length?

21. *Hint.* Use the fact that there are more mice in the world than there are hairs on any mouse.

CHAPTER 3

3. (a) (See the beginning of Section 2.)
(c) (See exercise 1.5.)

5. *Hint.* The geometric meaning of π is given by the ratio r_2 (or r_3), as defined in Section 2 of Chapter 2. Explain how, when we attempt to give a numerical meaning to π, as in inequalities (2) through (8) of this chapter—or as in exercise 3.14—the intuitive idea of a limit seems to be forced upon us.

9. $1^2 + 2^2 + 3^2 + \cdots + 100^2 = (100)(101)(201)/6 = 338,350$ balls.

13. *Hint.* Problem 12 shows that the triangle marked "4" is equal in area to the triangle marked "1".

15. The reader is asked only to make "an educated guess". (Problem 16 describes two ways that mathematicians have imaginatively responded to the facts presented here.)

17. *Suggestion*: First explain how, having proved statement (c), Archimedes knew that $A = \frac{1}{2}\, Cr$. Then divide both sides of this equality by r^2 to show that $A/r^2 = C/(2r)$. Explain why this means that the ratios r_2 and r_3 defined in Chapter 2 are equal (so that either can then be taken as the definition of the real number π). Thus, we know that $A/r^2 = \pi$ and $C/2r = \pi$, and the familiar formulas for the area and circumference of a circle follow immediately.

19. From equation (11) we see that the point (x, y) is on the parabola if the ratio r of the distance between (x, y) and F to the distance between (x, y) and $\mathbf{D}$ is 1; it is on the ellipse if this ratio r is $1/2$; and it is on the hyperbola if r is 2. That is, (x, y) is on the conic section with eccentricity r and with the given focus and directrix if

$$r = \frac{\sqrt{(x-4)^2 + (y-2)^2}}{\sqrt{(x-x)^2 + (y+2)^2}}.$$

Squaring both sides and simplifying shows that $r^2(y+2)^2 = (x-4)^2 + (y-2)^2$. Further simplification yields $x^2 - 8x + y^2 - r^2y^2 - 4y - 4r^2y + 20 - 4r^2 = 0$. Setting r equal to 1, $1/2$, and 2, respectively, gives the equations called for in parts (a), (b), and (c).

21. The equation you derive should show you that g is, in fact, a parabola. His friend f turns out to be (one branch of) a hyperbola.

25. *Hint.* Choose your unit to be the length of one side of the square mentioned in statement (12) and find the length of the diagonal of the square by the Pythagorean theorem. Then explain why inequality (11) and statement (12) say, in essence, the same thing.

29. *Hint.* Read the beginning of the next chapter.

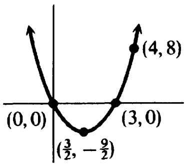
CHAPTER 4

1. If $f(x) = x^2 - 6x + 13$, then

$$\begin{aligned} f'(x) &= \operatorname*{Limit}_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \operatorname*{Limit}_{h\to 0} \frac{(x+h)^2 - 6(x+h) + 13 - x^2 + 6x - 13}{h} \\ &= \operatorname*{Limit}_{h\to 0} \frac{2xh + h^2 - 6h}{h} \\ &= \operatorname*{Limit}_{h\to 0}(2x + h - 6) \\ &= 2x - 6. \end{aligned}$$

3. (a) -6.
   (b) falling.

(c) $x < \frac{3}{2}$.
(d)



(e) $-4.5 \leq y \leq 8$.
(f) $-4 < y < 0$.
(g) $-4.5 \leq y$.

5. (a) $-1 \leq y \leq 8$. (b) $-7 < y < 14$.

7. $(4, 2)$.

9. (This problem is discussed in detail in Appendix 4.)

11. The wire should be cut so as to make the first part $500\pi/(4+\pi) \approx 220$ centimeters long in order to minimize the combined areas.

13. $f'(x) = \text{Limit}_{h\to 0}((f(x+h) - f(x))/h) = \text{Limit}_{h\to 0}((7-7)/h) = \text{Limit}_{h\to 0}\, 0 = 0$.

15. 1200 square meters.

17. (a) $y - 1 = 3(x-1)$, or $y = 3x - 2$. (b) Yes. (c) Yes. (d) No, the tangent line of slope 3 through $(1, 1)$ would be eliminated from consideration in "Holmes's method", because it cuts the curve twice.

18. The derivative of (a) is pictured in (d); the derivative of (b) is (d); of (c) is (g); of (d) is (g); of (e) is (h); of (f) is (b); of (g) is (k); of (h) is (j); of (i) is (b); of (j) is (k); of (k) is (k).

Chapter 5

1. (a) \$92.95. (b) 60 square meters.

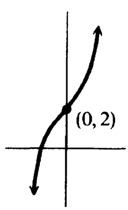
3. $y - 5 = -10(x - 1)$.

5. (a) rising. (b) to the right.

7. (a) 3. (b) to the left. (c) lowest.

9. range f is

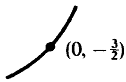
11. (a)



(b) $(0,2)$
(c) $-2 \le y \le 78$.

13.

t	y	y'	y''
0	$-\frac{3}{2}$	$\frac{3}{4}$	$\frac{1}{4}$



15. The first number should be $(-1+\sqrt{61})/3 \approx 2.27$, and the second should be ≈ 7.73.

17. 3.

19. The side of the cut-out square should be of length $(14-\sqrt{76})/6 \approx 0.88$ meters, in order to maximize the volume.

21. (d) Distance PQ should be $\sqrt{4/5} \approx 0.8944$ miles.

23. (a) min C occurs when $L = \sqrt[3]{16} \approx 2.52$ meters.

25. $\sqrt[3]{3}$ by $\sqrt[3]{3}$ by $\sqrt[3]{3}$ meters.

27. (a) approximately 7.30 by 7.30 by 3.65 feet.
(b) approximately 5.16 by 5.16 by 5.16 feet.
(c) both radius and height should be $\sqrt{160/3\pi} \approx 4.12$ feet.
(d) radius should be approximately 2.91 feet and height 5.82 feet.

29. (b) The first part should be $\pi A/(4+\pi)$ centimeters long.

31. (a) $2(x/(x-6))((x-6-x)/(x-6)^2)$. (f) $(1/2\sqrt{x^5})(5x^4)$.

43. (Partial answer) the derivative of (a) is (d); the derivative of (j) is (a).

Chapter 6

1. (a) 100 miles. (b) 50 mi/hr. (c) 40 mi/hr. (d) accelerating.
(e) decelerating.

3. (a) $\Delta A = x(\Delta y) + y(\Delta x) + (\Delta x)(\Delta y)$. (b) Δx and Δy must tend to zero since x and y are differentiable, hence continuous, functions of t.

5. (a) $4(x^2+7x)^3(2x+7)$. (c) $3((x-2)/(x+2))^2(4/(x+2)^2)$.

7. (a) $dA/dt = 2\pi r(dr/dt) = 2\pi(5)(7) = 70\pi\,\text{in}^2/\text{sec}$ at the instant in question.
   (b) *Hint.* Since $dA/dt = 2\pi r(dr/dt)$ and since r and dr/dt are both functions of t, the second derivative must be taken by applying the product rule to $2\pi r(dr/dt)$. (Note that d^2A/dt^2 is not equal to $(d^2A/dr^2)(d^2r/dt^2)$.)

9. $\frac{15}{7}$ ft/sec.

11. $(dr/dt)|_{r=2} = 3/4\pi\,\text{in/sec}$.

13. (b) $F(t) = 1 - (1/t)$. (This curve F is pictured in (d) of problem 25.) The answer is not unique, because the function pictured in figure (f) of problem 25 is also an antiderivative of $1/t^2$ that takes the value 0 when $t = 1$.

15. (a) 80 ft/sec. (b) 80 ft/sec. (c) 62 ft/sec.

17. The upward speed at impact is $\text{Limit}_{t\to 2.3^-}(-32t - 50) = -123.6$ ft/sec.

19. $g(3) = 3$.

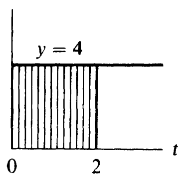
23. (b) 208 ft.

CHAPTER 7

1. 39 km.

3. $dA/dt = 1/(t+2)$.

5.



The area is $2 \cdot 4 = 8$.

7. $8 - \pi$ square units. (We do not yet know an antiderivative of $\sqrt{4 - t^2}$.)

9. (a)
$$\int_0^2 4x\,dx = \text{Limit} \sum_{k=1}^{n} 4(2k/n)(2/n) = \text{Limit}(16/n^2) \sum_{k=1}^{n} k$$
$$= \text{Limit}\, 8(1 + 1/n) = 8.$$

(c)
$$S_n = \sum_{k=1}^{n} x_k \Delta x = \sum_{k=1}^{n} (a + k\Delta x)\Delta x = a(\Delta x) \sum_{k=1}^{n} 1 + (\Delta x)^2 \sum_{k=1}^{n} k$$
$$= a\left(\frac{b-a}{n}\right)n + \left(\frac{b-a}{n}\right)^2 \frac{n(n+1)}{2}$$
$$= a(b-a) + (b-a)^2\left(\frac{1}{2}\right)\left(1 + \frac{1}{n}\right).$$
$$\int_a^b x\,dx = \text{Limit}\, S_n = a(b-a) + (b-a)^2\left(\frac{1}{2}\right) = \frac{b^2}{2} - \frac{a^2}{2}.$$

11. (b) 2.

13. (a) $16\pi/15$ cubic units. (b) $4\pi/3$ cubic units.

15. $343\pi/3$ cubic units.

17. (b) Slice 2.
    (c) (The cubic equation arising is the one whose solution is required in problem 25 of Chapter 4.)

23. $(-1/2, 15/4)$.

Chapter 8

2. *It is one of the rarest gifts to be able to hold a view with conviction and detachment at the same time. Philosophers and scientists more than other men strive to train themselves to achieve it, though in the end they are usually no more successful than the layman. Mathematics is admirably suited to foster this kind of attitude. It is by no means accidental that many of the greatest philosophers have also been mathematicians.* —Bertrand Russell.

7. Two solid figures have the same volume if, from some perspective, they can be seen to be made up of "cross-sections" with equal areas. (Think of a deck of cards. It takes up the same volume—doesn't it?—whether it is neatly stacked or is listing to one side.)

9. *Hint.* The vertical cross-section through a point x on the x-axis—where x ranges from 0 to 4—is a circle of radius $x+1$ and therefore has an area of $\pi(x+1)^2$. By Cavalieri's principle the total volume is given by the integral $\int_0^4 \pi(x+1)^2\,dx$. Evaluate this integral by the fundamental theorem.

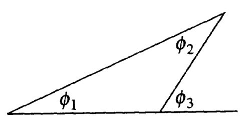
Appendix 2

1. $\sum_{k=1}^{36}(k^2+4k) = \sum_{k=1}^{36} k^2 + 4\sum_{k=1}^{36} k = \dfrac{(36)(37)(73)}{6} + 4\,\dfrac{(36)(37)}{2} = 18{,}870$ balls.

3. (b) *Hint.* $S_n = \sum_{k=1}^{n}\left(\dfrac{k^4}{n^4}\right)\left(\dfrac{1}{n}\right) = \left(\dfrac{1}{n^5}\right)\sum_{k=1}^{n} k^4 = \dfrac{n(n+1)(2n+1)(3n^2+3n-1)}{30n^5}$,
   from the result of part (a). To calculate the integral $\int_0^1 x^4\,dx$, find the limit of S_n as n increases without bound.

Appendix 3

1. *Hint.* Note that there are a pair of isosceles triangles in the figure, and recall that an "exterior" angle of a triangle is equal to the sum of the two "opposite" angles. That is, in the figure below we must have $\phi_1 + \phi_2 = \phi_3$. (*Why?*)



APPENDIX 4

3. The area of the rectangular portion of the field is maximized when its length is 100 meters and its width is $200/\pi$ meters.

APPENDIX 5

3. (a) $22 \leq h(3) \leq 35.5$.
   (b) $57 \leq h(5) \leq 119.5$.
   (d) $22 \leq \int_0^2 h(t)\,dt \leq 24$.
   (e) $64 \leq \int_0^4 h(t)\,dt \leq 100$.

5. *Partial answer*: Formula (9) tells us here that $t - \frac{1}{6}t^3 \leq S(t) \leq t$. Letting $t = 0.1$ shows that $0.9983 \leq S(0.1) \leq 0.1000$; letting $t = 1$ shows that $0.833 \leq S(1.0) \leq 1.000$.

Index

(continued from page ii)

James: Topological and Uniform Spaces.
Jänich: Linear Algebra.
Jänich: Topology.
Kemeny/Snell: Finite Markov Chains.
Kinsey: Topology of Surfaces.
Klambauer: Aspects of Calculus.
Lang: A First Course in Calculus. Fifth edition.
Lang: Calculus of Several Variables. Third edition.
Lang: Introduction to Linear Algebra. Second edition.
Lang: Linear Algebra. Third edition.
Lang: Undergraduate Algebra. Second edition.
Lang: Undergraduate Analysis.
Lax/Burstein/Lax: Calculus with Applications and Computing. Volume 1.
LeCuyer: College Mathematics with APL.
Lidl/Pilz: Applied Abstract Algebra. Second edition.
Logan: Applied Partial Differential Equations.
Macki-Strauss: Introduction to Optimal Control Theory.
Malitz: Introduction to Mathematical Logic.
Marsden/Weinstein: Calculus I, II, III. Second edition.
Martin: The Foundations of Geometry and the Non-Euclidean Plane.
Martin: Geometric Constructions.
Martin: Transformation Geometry: An Introduction to Symmetry.
Millman/Parker: Geometry: A Metric Approach with Models. Second edition.
Moschovakis: Notes on Set Theory.
Owen: A First Course in the Mathematical Foundations of Thermodynamics.
Palka: An Introduction to Complex Function Theory.
Pedrick: A First Course in Analysis.
Peressini/Sullivan/Uhl: The Mathematics of Nonlinear Programming.
Prenowitz/Jantosciak: Join Geometries.
Priestley: Calculus: A Liberal Art. Second edition.
Protter/Morrey: A First Course in Real Analysis. Second edition.
Protter/Morrey: Intermediate Calculus. Second edition.
Roman: An Introduction to Coding and Information Theory.
Ross: Elementary Analysis: The Theory of Calculus.
Samuel: Projective Geometry. *Readings in Mathematics.*
Scharlau/Opolka: From Fermat to Minkowski.
Sethuraman: Rings, Fields, and Vector Spaces: An Approach to Geometric Constructability.
Sigler: Algebra.
Silverman/Tate: Rational Points on Elliptic Curves.
Simmonds: A Brief on Tensor Analysis. Second edition.
Singer: Geometry: Plane and Fancy.
Singer/Thorpe: Lecture Notes on Elementary Topology and Geometry.
Smith: Linear Algebra. Third edition.
Smith: Primer of Modern Analysis. Second edition.
Stanton/White: Constructive Combinatorics.
Stillwell: Elements of Algebra: Geometry, Numbers, Equations.
Stillwell: Mathematics and Its History.
Stillwell: Numbers and Geometry. *Readings in Mathematics.*
Strayer: Linear Programming and Its Applications.
Thorpe: Elementary Topics in Differential Geometry.
Toth: Glimpses of Algebra and Geometry.

Troutman: Variational Calculus and Optimal Control. Second edition.
Valenza: Linear Algebra: An Introduction to Abstract Mathematics.
Whyburn/Duda: Dynamic Topology.
Wilson: Much Ado About Calculus.